

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ADMINISTRAÇÃO, CONTABILIDADE E ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

Peng Yaohao

***Support Vector Regression* aplicado à
previsão de taxas de câmbio**

Peng Yaohao

***Support Vector Regression* aplicado à
previsão de taxas de câmbio**

Área de concentração:
Finanças e métodos quantitativos

Orientador:
Professor Doutor Pedro Henrique Melo
Albuquerque

Brasília
2016

FICHA CATALOGRÁFICA

PP398s Peng, Yaohao
Support Vector Regression aplicado à previsão de
taxas de câmbio / Yaohao Peng; orientador Pedro
Albuquerque. -- Brasília, 2016.
129 p.

Dissertação (Mestrado - Mestrado em Administração)
- Universidade de Brasília, 2016.

1. Aprendizado de máquinas. 2. Métodos Kernel.
3. Mercado FOREX. 4. Fundamentos macroeconômicos. 5.
Capacidade preditiva. I. Albuquerque, Pedro, orient.
II. Título.

Soneto da dissertação⁰

Amanhece o elmo da lucidez,
Blindando a laxa fronteira mental.
Beligerante trilha bienal
À brisa do Noroeste outra vez.

Torres de cera e espelhos sem leis
Reverberam no lúgubre abissal.
O impregnável colóide incremental
Dissipa a névoa da insensatez.

Pérolas do éter, feixes de luz.
Com versos vítreos o infinito emerge,
E ao silêncio solipsista reduz.

Tênue reflexo do sublime véu,
Permeia mil sinapses e converge
A mais um baricentro de um papel.

⁰Quem nasceu para ser “heróico” dispensa “redondilhas”!

AGRADECIMENTOS

Agradeço imensamente ao Professor Pedro Henrique Melo Albuquerque pela orientação, pela dedicação e pelo companheirismo, por ser um profissional engajado com a construção do saber: uma pessoa exemplar que tenho a honra de chamar de mentor e o prazer de chamar de amigo.

Agradeço ao Professor Ivan Ricardo Gartner por todos os conselhos e críticas que me alertam constantemente para não me conformar com a simplicidade, por ter-me ensinado que um erro não é um acerto parcial, e por ter-me servido como uma referência no mundo acadêmico.

Agradeço aos examinadores da banca, Professores André Luiz Fernandes Cançado e Ivan Ricardo Gartner, pelas críticas e sugestões construtivas e pertinentes, os quais contribuíram decisivamente para a melhoria da qualidade deste trabalho.

Agradeço aos Professores Herbert Kimura, Vinícius Amorim Sobreiro e Otávio Ribeiro de Medeiros pela consideração e incentivos, por ter me mostrado que ainda há mundos a serem explorados e que o caminho para as conquistas está debaixo dos pés.

Agradeço aos meus irmãos (em ordem alfabética para não gerar discórdias...) Afonso Salustiano Neri, Christian Maciel Machado Rocha, João Lucas Magalini Zago, Matheus Gonçalves de Souza e Sérgio Thadeu Tavares da Silva Júnior. Obrigado por sempre estarem do meu lado e terem confiado em mim desde o começo. Meu sucesso é de vocês, pois vocês são o meu sucesso.

Agradeço aos docentes e colegas do PPGA/UnB e do grupo de pesquisa LAMFO pelo carinho e apoio durante essa jornada: Professor Tomas de Aquino Guimarães, Professor Edgar Reyes Junior, Professora Gisela Demo Fiúza, Professor Francisco Antonio Coelho Junior, Professor Ricardo Corrêa Gomes; Thiago Raymon, Sarah Sabino, João Gabriel Moraes, Pedro Correia, Leonardo Bosque, Emmanuel Abreu, Alexandre Leite, Bruno Miranda, Monique Azevedo, Mariana Montenegro, Ana Júlia Akaishi, Fernanda Amorim, Jader Martins, Gustavo Monteiro, Marcelo Felix, Matheus Facure, Luiz Medeiros, José Rômulo Vieira, Manoel Fonseca, Raphael Pereira, Leonardo Magno, Marcelo Coutinho, Felipe Calainho, Marina Uchôa, Gustavo Basso, Raphael Brocchi, Wanderson Bittencourt, Marina Garcia, Jorge Barbosa, Fábio Medina,

Mário Salimon, Juliana Moro, Ricardo Ken, Silvia Mori, Natália de Melo, David Bouças, Pablo Pessôa, Junia Falqueto, Oscar Oliveira, Everton Verga, Alexsander Dauzeley, André Porfírio, Marilú Castro, José Nilton, Alex Fabiane, Leovanir Richter, Sérgio Freitas, Ladilucy Armond, Vivian Carolino, Emília Faria, Eduardo Lafetá, Ana Carolina Costa, Isabela Ferraz, Bernardo Buta, Natasha Fogaça, Marcelo Finazzi, Mariana Rêgo, Rodrigo Montalvão, Walter Faiad, Nazareno Marques, Carolina Sgaraboto, Ana Paula Lopes, Marcelo Cardoso, Wanderson Lacerda, Tiago Silva, Cristiano Lúcio, Paulo Daltro, Luiz Fernando Pinto, Renata Telles, Leonel Cerqueira. Muito obrigado a todos por tudo.

Agradecimentos especiais à Professora Olinda Maria Gomes Lesses, pela oportunidade concedida para três períodos de docência, que me permitiu perceber, em definitivo, que “estar do outro lado” propicia um aprendizado sem igual; e a Pedro Alexandre Moura Barros Henrique, cuja disposição e gentileza foram determinantes para a finalização deste trabalho.

Fiquemos por aqui, pois se eu fosse dedicar a cada um que já contribuiu na minha caminhada nessa vida as palavras de agradecimento que merecem, **certamente** essa seção iria ficar maior que a dissertação em si...

RESUMO

O presente estudo realizou a previsão da taxa *spot* de 15 pares de câmbio mediante a aplicação de um algoritmo de aprendizado de máquinas – *Support Vector Regression* – com base em um modelo fundamentalista composto por 13 variáveis explicativas. Para a estimação das previsões, foram consideradas 9 funções *Kernel* extraídas da literatura científica, totalizando assim 135 modelos verificados. As previsões foram comparadas com o *benchmark Random Walk* e avaliadas em relação à taxa de acerto direcional do câmbio e às métricas de erro RMSE (raiz quadrada do erro quadrático médio) e MAE (erro absoluto médio). A significância estatística do incremento de poder explicativo dos modelos SVR em relação ao *Random Walk* foi verificada mediante a aplicação do *Reality Check Test* de White (2000). Os resultados mostram que os modelos SVR obtiveram desempenho preditivo satisfatório em relação ao *benchmark*, com vários dos modelos propostos apresentando forte significância estatística de superioridade preditiva. Por outro lado, observou-se que várias funções *Kernel* comumente utilizadas na literatura científica não lograram êxito em superar o *Random Walk*, apontando para uma possível lacuna no estado da arte de aprendizado de máquinas aplicada à previsão de taxas de câmbio. Por fim, discutiu-se acerca das implicações dos resultados obtidos para o desenvolvimento futuro da agenda de pesquisa correlata.

Palavras-chave: Aprendizado de máquinas, Métodos *Kernel*, Mercado FOREX, Fundamentos macroeconômicos, Capacidade preditiva.

ABSTRACT

This paper aims to forecast the spot exchange rate of 15 currency pairs by applying a machine learning algorithm – *Support Vector Regression* – based on a fundamentalist model composed of 13 explanatory variables. The predictions' estimation were obtained by applying 9 different *Kernel* functions extracted from the scientific literature, resulting in a total of 135 models verified. The predictions were compared to the Random Walk benchmark and evaluated for directional accuracy rate of exchange predictions and error performance indices RMSE (root mean square error) and MAE (mean absolute error). The statistical significance of the explanatory power gain via SVR models with respect to the Random Walk was checked by applying White (2000)'s Reality Check Test. The results show that SVR models achieved satisfactory predictive performance relative to the benchmark, with several of the proposed models showing strong statistical significance of predictive superiority. Furthermore, the results showed that mainstream Kernel functions commonly used in the scientific literature failed to outperform the Random Walk, indicating a possible gap in the state of art of machine learning methods applications to exchange rates forecasting. Finally, the paper presents a discussion about the implications of the obtained results for the future development of related research agendas.

Keywords: Machine Learning, Kernel Methods, Foreign Exchange Market, Macroeconomic Fundamentals, Predictive Ability.

LISTA DE FIGURAS

4.1	Representação do problema do SVM em dados linearmente separáveis.	57
4.2	SVM para classificação não-linear.	59
4.3	Parâmetros e função de decisão do ϵ - SVR	64

LISTA DE TABELAS

3.1	Variáveis independentes do modelo fundamentalista	55
4.1	Funções <i>Kernel</i> utilizadas no estudo	68
5.1	Intervalos de busca utilizados para os parâmetros de treinamento	90
7.1	Parâmetros ótimos para o câmbio USD-EUR após treinamento e validação . . .	106
7.2	Parâmetros ótimos para o câmbio USD-GBP após treinamento e validação . . .	106
7.3	Parâmetros ótimos para o câmbio USD-JPY após treinamento e validação . . .	107
7.4	Parâmetros ótimos para o câmbio USD-CNY após treinamento e validação . . .	107
7.5	Parâmetros ótimos para o câmbio USD-BRL após treinamento e validação . . .	108
7.6	Parâmetros ótimos para o câmbio EUR-GBP após treinamento e validação . . .	108
7.7	Parâmetros ótimos para o câmbio EUR-JPY após treinamento e validação . . .	108
7.8	Parâmetros ótimos para o câmbio EUR-CNY após treinamento e validação . . .	109
7.9	Parâmetros ótimos para o câmbio EUR-BRL após treinamento e validação . . .	109
7.10	Parâmetros ótimos para o câmbio GBP-JPY após treinamento e validação . . .	109
7.11	Parâmetros ótimos para o câmbio GBP-CNY após treinamento e validação . . .	110
7.12	Parâmetros ótimos para o câmbio GBP-BRL após treinamento e validação . . .	110
7.13	Parâmetros ótimos para o câmbio CNY-JPY após treinamento e validação . . .	111
7.14	Parâmetros ótimos para o câmbio BRL-JPY após treinamento e validação . . .	111
7.15	Parâmetros ótimos para o câmbio BRL-CNY após treinamento e validação . . .	112
7.16	Avaliação das previsões no conjunto de teste para o câmbio USD-EUR	112
7.17	Avaliação das previsões no conjunto de teste para o câmbio USD-GBP	113
7.18	Avaliação das previsões no conjunto de teste para o câmbio USD-JPY	113
7.19	Avaliação das previsões no conjunto de teste para o câmbio USD-CNY	113

7.20 Avaliação das previsões no conjunto de teste para o câmbio USD-BRL	114
7.21 Avaliação das previsões no conjunto de teste para o câmbio EUR-GBP	114
7.22 Avaliação das previsões no conjunto de teste para o câmbio EUR-JPY	114
7.23 Avaliação das previsões no conjunto de teste para o câmbio EUR-CNY	115
7.24 Avaliação das previsões no conjunto de teste para o câmbio EUR-BRL	115
7.25 Avaliação das previsões no conjunto de teste para o câmbio GBP-JPY	116
7.26 Avaliação das previsões no conjunto de teste para o câmbio GBP-CNY	116
7.27 Avaliação das previsões no conjunto de teste para o câmbio GBP-BRL	117
7.28 Avaliação das previsões no conjunto de teste para o câmbio CNY-JPY	117
7.29 Avaliação das previsões no conjunto de teste para o câmbio BRL-JPY	117
7.30 Avaliação das previsões no conjunto de teste para o câmbio BRL-CNY	118
7.31 P-valores do teste de White – Parte 1	118
7.32 P-valores do teste de White – Parte 2	119
7.33 P-valores do teste de White – Parte 3	119

LISTA DE ABREVIATURAS

ARMA *Autoregressive Moving Average*

ARIMA *Autoregressive Integrated Moving Average*

BPNN *Back-Propagation Neural Network*

BRL *Brazilian Real* – Real brasileiro

CNY *Chinese Yuan* – Yuan chinês (renminbi)

DBN *Deep Belief Network*

EUR Euro

FFNN *Feed Forward Neural Network*

FOREX *Foreign Exchange*

GARCH *Generalized Autoregressive Conditional Heteroskedasticity*

GBP *British Pound* – Libra esterlina

JPY *Japanese Yen* – Iene japonês

MAE *Mean Absolute Error* – Erro Médio Absoluto

MAPE *Mean Absolute Percentage Error* – Erro Médio Absoluto Percentual

MLP *Multilayer Perceptron*

MSE *Mean Square Error* – Erro Quadrático Médio

RMSE *Root Mean Square Error* – Raiz do Erro Quadrático Médio

RKHS *Reproducing Kernel Hilbert Space*

SDR *Special Drawing Rights*

SVM *Support Vector Machine*

SVR *Support Vector Regression*

USD *United States Dollar* – Dólar norte-americano

SUMÁRIO

1	Introdução	16
1.1	Formulação do problema	18
1.2	Justificativa e contribuições	20
2	Referencial teórico	24
2.1	Caracterização do mercado FOREX	24
2.2	Análise técnica e fundamentalista	25
2.3	Aprendizado de máquinas: aplicações em finanças	28
2.4	Previsão da taxa de câmbio: Estado da arte	29
2.4.1	Aprendizado de máquinas na previsão da taxa de câmbio	35
3	Seleção das variáveis	45
3.1	Variáveis levantadas	45
3.1.1	Câmbio defasado	46
3.1.2	Taxa de juros	46
3.1.3	Oferta monetária	47
3.1.4	Taxa de inflação	47
3.1.5	Índice de produção industrial	48
3.1.6	Balança comercial	49
3.1.7	Variação do índice da bolsa de valores	49
3.1.8	Déficit público	50
3.1.9	Taxa de inflação esperada	50
3.1.10	Movimento de capitais	51

3.1.11	Volume de reservas internacionais	51
3.1.12	Preço de <i>commodities</i>	52
3.2	Variáveis expurgadas	53
4	Metodologia	56
4.1	<i>Support Vector Machine</i>	56
4.1.1	SVM para classificação linear	56
4.1.2	SVM para classificação não linear	59
4.1.3	Forma dual do SVM	60
4.2	<i>Support Vector Regression</i>	62
4.3	Funções <i>Kernel</i>	66
4.3.1	Condição de Mercer	66
4.3.2	<i>Kernel</i> polinomial	68
4.3.3	<i>Kernel</i> Gaussiano	81
4.3.4	<i>Kernel</i> multiquádrico inverso	84
4.3.5	<i>Kernel</i> Log	85
4.3.6	<i>Kernel</i> Cauchy	85
4.3.7	<i>Kernel</i> Spline Estável	86
5	Análise empírica	88
5.1	Definição do <i>benchmark</i>	88
5.2	Coleta e tratamento dos dados	89
6	Resultados e discussão	93
6.1	Treinamento e validação	93
6.2	Avaliação das previsões no conjunto de teste	93
6.3	<i>Reality Check Test</i> de White	95
7	Conclusão e recomendações	101

7.1	Limitações e desenvolvimentos futuros	102
Anexo A:	Tabelas	106
A.1:	Parâmetros ótimos para os conjuntos de treinamento e validação	106
A.2:	Avaliação das previsões nos conjuntos de teste	112
A.3:	P-valores do <i>Reality Check Test</i> de White	118
Anexo B:	Código de programação em R	120
Referências Bibliográficas		127

1 INTRODUÇÃO

A taxa de câmbio é um instrumento essencial para se compreender a realidade econômico-financeira de uma economia, notadamente em um contexto contemporâneo de globalização financeira e interdependência crescente e consolidada entre os mercados a nível internacional (CHIȚU; EICHENGREEN; MEHL, 2014). Nesse sentido, a predição da taxa de câmbio mostra-se relevante para empresas, investidores e demais participantes do mercado financeiro internacional, na medida em que uma predição acurada dessa variável pode – a nível micro – melhor orientar as decisões estratégicas de financiamento e investimento; e – a nível macro – determinar encaminhamento das políticas econômicas de um país (comerciais, fiscais ou monetárias) (ROSSI, 2013; RODRIK, 2014), bem como pode ter seus efeitos estendidos para a determinação das diretrizes de política externa do país, refletindo-se em aspectos como o eixo de cálculo estratégico a ser adotado pelas ações diplomáticas no país em negociações internacionais ou foros internacionais de constituição da agenda política mundial (COOPER, 1972; GARDNER, 1980; HOLSTI, 1986).

Porém, a taxa de câmbio é uma variável que carrega consigo um elevado grau de instabilidade, dada a vasta gama de elementos passíveis a exercer influência sobre as mudanças dessa variável. Na literatura científica que diz respeito à predição da taxa de câmbio, os trabalhos de Goodman (1979), Cumby e Obstfeld (1982) e o artigo seminal de Meese e Rogoff (1983) preconizam que modelos lineares de previsão usando fundamentos macroeconômicos como preditores possuem baixo poder explicativo. Desse modo, Meese e Rogoff (1983) apresentam a conclusão de que o comportamento das taxas de câmbio segue um passeio aleatório (*Random Walk*), um resultado conhecido na literatura como *Meese-Rogoff puzzle*, o qual afirma que nenhum modelo estrutural fundamentalista supera o modelo de passeio aleatório para a predição da taxa de câmbio. A hipótese do *Random Walk* (FAMA, 1965)¹ implica que num mercado eficiente qualquer informação nova no mercado será imediatamente incorporada no novo nível dos preços dos ativos financeiros, de modo que, em média, qualquer estratégia no mercado financeiro – seja grafista, seja fundamentalista – não deve obter retornos consistentemente

¹As primeiras teorizações do *Random Walk* aplicado ao mercado financeiro remontam a Regnault (1863) e Bachelier (1900).

superiores a uma estratégia baseada puramente no acaso. Esse resultado vai ao encontro da chamada hipótese dos mercados eficientes (HME), tal qual enunciada na sua forma fraca² no trabalho clássico de Fama (1970).

No entanto, a produção acadêmica em relação a modelos que tentam prever o comportamento das taxas de câmbio tem se mostrado bastante ativa ao longo das décadas (BECKMANN; SCHÜSSLER, 2016), constituindo-se num tópico de pesquisa altamente debatido e comumente observado em artigos científicos publicados em periódicos de alto fator de impacto e relevância no desenvolvimento do conhecimento científico (ROSSI, 2013). Não se limitando apenas ao âmbito das finanças, a constante busca por modelos que objetivam relativizar o *Meese-Rogoff puzzle* se estende para campos como a macroeconomia e a computação, à medida que novas modelagens e metodologias mais arrojadas são desenvolvidas para investigar nuances que possam elevar o poder preditivo das taxas de câmbio.

O trabalho de Mark (1995) lança mão de um modelo fundamentalista que considera como variáveis independentes o diferencial³ da oferta monetária e da inflação, posteriormente conhecido na literatura como *monetary model*. A principal contribuição desse estudo se dá pelo achado de evidências que reforçam a existência de um componente previsível nas taxas de câmbio em horizontes temporais relativamente longos, levando à implicação de que a premissa de Meese e Rogoff (1983) pode não se observar – ou seja, as taxas de câmbio não seguem um *Random Walk* e modelos fundamentalistas podem apresentar significativo poder de previsão para essa variável. Evidências de que a predição da taxa de câmbio pode superar um *Random Walk* também são encontrados em Kuan e Liu (1995) e Kodogiannis e Lolis (2002).

O artigo de Mark (1995) exerceu uma grande influência no desenvolvimento da literatura científica em relação à possibilidade de predição das taxas de câmbio, uma vez que abriu o caminho para uma série de estudos que questionam o *Meese-Rogoff puzzle*: Faust, Rogers e Wright (2003) testam o modelo de Mark (1995) e identificam evidências de que este aumenta a previsibilidade das taxas de câmbio para horizontes temporais relativamente longos (superior a dois anos), de modo que para o curto prazo não há evidências que corroborem a robustez do modelo em relação a superar o *Random Walk*. Estudos como Engel e West (2005), Engel, Mark e West (2015) e Balke, Ma e Wohar (2013) chegam a conclusões parecidas ao artigo de Mark (1995) usando variáveis macroeconômicas similares (e algumas iguais, como por exemplo

²A forma fraca da HME postula que os níveis de preço presentes refletem toda a informação do passado. Ou seja, a variação do preço em t para o preço em $t + 1$ é totalmente aleatória.

³No presente estudo, o “diferencial” de uma variável diz respeito à diferença numérica entre as observações da referida variável realizadas nos mercados emissores das duas moedas consideradas no par de câmbio analisado. Por exemplo, para o câmbio USD-BRL, o diferencial da taxa de inflação será a diferença entre as inflações observadas nos mercados norte-americano e brasileiro.

a oferta de moeda) ao *monetary model*. Vários estudos recentes também indicam fortes evidências de fundamentos macroeconômicos possuem poder explicativo para a predição da taxa de câmbio (BACCHETTA; WINCOOP, 2013), bem como verificou-se estatisticamente (via teste de Granger) a evidências de uma forte relação de causalidade entre fundamentos macroeconômicos e taxas de câmbio (DĄBROWSKI; PAPIEŻ; ŚMIECH, 2014).

No contexto do aprendizado de máquinas, o estudo de Bissoondeal et al. (2008) mede a acurácia de predição das taxas de câmbio da lira turca em relação ao dólar e ao euro, comparando o *monetary model* de Mark (1995) com uma rede neural artificial. Os resultados desse artigo apontam evidências de que a rede neural apresenta desempenho superior, fornecendo ferramentas adicionais para contestar o *Meese-Rogoff puzzle*. Conforme elucidado nas seções posteriores, o aprendizado de máquinas é uma abordagem com uso crescente na literatura científica de finanças, apresentando resultados satisfatórios e fornecendo agendas de pesquisa promissoras; ademais, a literatura indica que a abordagem *Support Vector Machine* (doravante SVM) e suas extensões (como o *Support Vector Regression* – doravante SVR –, a ser utilizado no presente estudo) superam o desempenho de redes neurais, justificando a investigação da temática de previsão do câmbio com essa metodologia.

1.1 Formulação do problema

A literatura científica basicamente segue duas abordagens distintas para o tratamento da temática de previsão de taxas de câmbio: modelos de análise técnica e modelos de análise fundamentalista (NASSIRTOUSSI; WAH; LING, 2011, p. 8323). A abordagem técnica (ou “grafista”) assume que os preços do mercado financeiro se movem segundo tendências que são detectáveis *ex-ante*, uma vez que é assumida a premissa de que a história tende a se repetir, permitindo a identificação de padrões que irão subsidiar a previsão da tendência futura. Já a abordagem fundamentalista associa as causas das oscilações no mercado a fatores macroeconômicos⁴, tais como: relatórios econômicos, níveis da taxa de juros, política monetária e comércio internacional (OZTURK; TOROSLU; FIDAN, 2016, p. 172).

A produção acadêmica em relação à previsão da taxa de câmbio tem priorizado mais trabalhos de cunho técnico, em detrimento dos de cunho fundamentalista (KOROL, 2014, p. 50), devido à sofisticação crescente das metodologias utilizadas para a análise técnica e ao resul-

⁴No âmbito da predição da taxa de câmbio, assume-se que esta variável é exógena a nível microeconômico (*i.e.*, para indivíduos e empresas). Assim, a análise fundamentalista para essa temática não diz respeito a demonstrativos contábeis de empresas, todas as quais são assumidas como *price-takers* em relação às taxas de câmbio.

tado do *Meese-Rogoff puzzle*, o qual desfavorece o uso de modelos fundamentalistas ao criticar seu poder de captar o comportamento das taxas de câmbio. Porém, estudos recentes apontam que modelos fundamentalistas também podem apresentar alto poder preditivo, indicando inclusive um equilíbrio a longo prazo entre os taxas de câmbio e fundamentos macroeconômicos (KATUSIIME; SHAMSUDDIN; AGBOLA, 2015). Nesse sentido, o presente estudo busca aliar as vantagens de ambas as abordagens técnica e fundamentalista para formular um modelo que logre êxito em fornecer insumos adequados para a melhor tomada de decisão dos agentes participantes do mercado internacional de câmbio (*Foreign Exchange Market*, doravante FOREX) em relação às oscilações da taxa de câmbio.

O presente estudo apresenta a formulação de um modelo fundamentalista baseado em variáveis explicativas macroeconômicas embasadas na literatura científica; após feita a modelagem, os dados são trabalhados pela abordagem do aprendizado de máquinas – especificamente, pelo método *Support Vector Regression* (SVR) – para realizar a predição da taxa de câmbio nominal de seis moedas – dólar norte-americano, euro, libra esterlina, renminbi chinês, iene japonês e real brasileiro – combinados dois a dois, totalizando quinze pares de câmbio. Estudos que adotam estrutura similar incluem Eng et al. (2008) e Nassirtoussi, Wah e Ling (2011), os quais fizeram uso de uma rede neural artificial do tipo *Multilayer Perceptron* (MLP)⁵ após definidas as variáveis fundamentalistas. Ambos os estudos apresentaram resultados que reforçam a capacidade de variáveis macroeconômicas em prever o movimento das taxas de câmbio, e sugerem que pesquisas que façam a aplicação de outras variáveis explicativas e/ou metodologias de aprendizado de máquinas possuem grande margem para elevar o poder preditivo dos modelos que deles derivem.

Resultados da literatura indicam um “descompasso temporal” no tocante à previsibilidade das taxas de câmbio: por um lado, o poder explicativo de variáveis fundamentalistas tende a ser significativo em horizontes temporais maiores, em geral acima de 3 ou 4 anos (KILIAN; TAYLOR, 2003; ROSSI, 2013). Por outro lado, estudos de análise técnica, em especial estudos que fazem uso de aprendizado de máquinas, apontam que a previsibilidade da série está em horizontes temporais curtos, em geral abaixo de um mês (SANTOS; COSTA; COELHO, 2007; GALESHCHUK, 2016). Dessa maneira, o presente estudo considerará frequência mensal para os dados coletados, buscando verificar o poder explicativo de variáveis independentes fundamentalistas (definidas com o aporte teórico da produção acadêmica) a curto prazo fazendo uso de um algoritmo de aprendizado de máquinas (SVR), o qual tem demonstrado desempenho satisfatório em estudos científicos recentes.

⁵Um caso específico de rede neural com propagação retroativa capaz de classificar dados não separáveis linearmente (CYBENKO, 1992).

Após a especificação do modelo fundamentalista, os dados serão analisados sob a ótica do aprendizado de máquinas, de modo que a predição será feita mediante a aplicação do algoritmo *Support Vector Regression*; os resultados foram então comparados ao *benchmark* do modelo *Random Walk* mediante a aplicação do *Reality Check Test* de White (2000) a fim de verificar a significância estatística do poder preditivo dos modelos propostos.

Um conceito chave nos métodos SVM e SVR é a **função Kernel** – a ser abordado com detalhes em seção posterior – a qual é um dos insumos necessários para a execução do algoritmo, cuja escolha exerce influência decisiva para o desempenho preditivo do modelo, de modo que a própria escolha da função *Kernel* constitui em uma das principais perguntas em aberto no campo de aprendizado de máquinas. Constatou-se que grande parte dos estudos científicos levou em consideração um conjunto reduzido de funções *Kernel*, fenômeno que também se verifica na literatura específica de previsão da taxa de câmbio. No presente estudo, essa lacuna é levada em consideração, de modo que serão incorporadas diversas funções *Kernel* pouco utilizadas nos artigos científicos e verificar se a introdução dessas funções contribui significativamente para o incremento do poder explicativo do SVR para prever taxas de câmbio. Os resultados serão então comparados ao *Random Walk* a fim de verificar o poder preditivo do modelo proposto.

1.2 Justificativa e contribuições

A previsão da taxa de câmbio tem se configurado como um relevante tema de pesquisa nas finanças nos últimos anos, com grande número de publicações sendo produzidas sobre o assunto, dada a pertinência do tema não apenas para acadêmicos, mas também para organizações governamentais e demais participantes do mercado. (AHMED; LIU; VALENTE, 2016, p. 96). Ademais, pelo fato de as taxas de câmbio serem os fatores de conversão entre a moeda doméstica e as moedas estrangeiras, seu valor impacta diretamente nos ganhos auferidos por um país mediante sua atividade comercial no ambiente internacional, ganhos estes que condicionam suas decisões de política comercial, afetando conseqüentemente as políticas de investimento, a atratividade do mercado doméstico para investidores estrangeiros e, em última instância, a solidez macroeconômica de um país. O trabalho de Wieland, Wolters et al. (2013) aborda especificamente a relevância de boas estimativas das taxas de câmbio para a boa formulação de políticas econômicas, fornecendo evidência empírica de bancos centrais dos Estados Unidos e da Europa dependem fortemente de previsões de indicadores macroeconômicos (tais como câmbio, juros e inflação) para a tomada de decisões de política econômica. Ademais, sendo o comércio internacional ainda uma significativa força-motriz para impulsionar o crescimento dos

países – com destaque para os mercados emergentes –, o estudo das taxas de câmbio torna-se especialmente relevante (HOOY; SIONG-HOOK; TZE-HAW, 2015; BALDWIN, 2016).

A principal contribuição do presente trabalho é verificar a previsão de taxas de câmbio considerando várias funções *Kernel*, comparando o resultado preditivo entre essas diferentes funções e visando à identificação de funções *Kernel* que apresentem bom desempenho, mas que são pouco utilizadas na literatura. Além das funções *Kernel* polinomial e Gaussiano, comumente utilizadas em estudos de aprendizado de máquinas, serão considerados os *Kernels* Log, multiquádrico inverso, Cauchy e Spline estável, os quais têm sido utilizados em aplicações diversas de algoritmos de aprendizado de máquinas. A seção 4.3 apresenta uma discussão sobre cada função *Kernel* utilizada e seus respectivos embasamentos em produções acadêmicas.

Dado que a literatura científica ainda não apresentou consenso ou critérios acerca da escolha da função *Kernel* ideal (PILLONETTO et al., 2014) – especialmente no âmbito de finanças, onde há diversos fatos estilizados que elevam a complexidade das previsões – o presente estudo, ao testar vários *Kernels* pouco utilizados, pretende comparar não apenas modelos SVR com o *benchmark Random Walk*, mas também analisar de que forma a alteração da função *Kernel* utilizada no algoritmo acarreta mudanças significativas no poder explicativo de cada modelo. Dessa forma, o estudo irá descrever brevemente a importância das funções *Kernel* e apresentar algumas discussões em relação às suas formas funcionais e implicações na qualidade das previsões fornecidas: particularmente, será realizada um detalhamento acerca da capacidade de as funções *Kernel* generalizar interações não-lineares em dimensões elevadas, abarcando assim formas funcionais que capturam estruturas de dependência não-linear, enriquecendo a análise em relação ao *mainstream* da Academia, a qual prioriza modelos com forma funcional linear.

Ademais, o presente trabalho irá aliar as óticas de análise técnica e fundamentalista para a predição das taxas de câmbio, baseado nas evidências apresentadas pela literatura científica de que fundamentos macroeconômicos possuem poder explicativo para a previsão do câmbio; e de que metodologias de *machine learning* têm sido aplicados com sucesso em diversos trabalhos correlatos a modelagens em finanças e previsão de variáveis econômicas, algumas das quais estão listadas nas seções 2.3 e 2.4.1. A escolha das variáveis fundamentalistas explicativas será embasada em estudos encontrados na literatura, em especial no estudo bibliométrico de Rossi (2013), que realiza um mapeamento detalhado acerca dos trabalhos recentes de previsão de taxas de câmbio, além de verificar alguns pontos que definem o estado da arte da predição da taxa de câmbio, listados a seguir:

- (i) A maior parte dos estudos de previsão do câmbio de cunho fundamentalista envolve a

estimação de modelos lineares;

- (ii) O *benchmark* mais difícil de ser superado é o *Random Walk*;
- (iii) Grande parte dos estudos apresentaram capacidade de previsão satisfatória, porém frequentemente condicionados à amostra selecionada e ao horizonte temporal, de tal modo que não foram encontradas evidências estatísticas fortes o bastantes para refutar o *Meese-Rogoff Puzzle*;
- (iv) A grande maioria dos estudos fundamentalistas para a previsão do câmbio adotou frequência dos dados mensal ou trimestral;
- (v) Grande parte dos estudos utilizaram de métricas de erro (tais como RMSE e MAE) ou o teste de Diebold-Mariano (DIEBOLD; MARIANO, 1995) para a robustez das previsões, enquanto que a aplicação de testes de significância que avaliam a superioridade do modelo em relação ao *benchmark*, como o *Reality Check Test* de White (2000) ou o *Superior Predictive Ability Test* de Hansen (2005), não foi identificada.

Sendo assim, o presente estudo também adotou a periodicidade mensal para os dados, dada a restrição de que fundamentos macroeconômicos são, em geral, divulgados em frequência menor. O trabalho também adotou o *Random Walk* como modelo *benchmark* com o qual serão comparados as previsões dos modelos SVR. Em contrapartida ao padrão observado na literatura, foram testados modelos que incorporam interações não-lineares entre as variáveis explicativas, cujas previsões serão analisadas à luz do teste de White (2000), a fim de verificar a existência ou ausência de algum modelo SVR que seja “significativamente melhor” que o *Random Walk*.

A escolha das moedas teve como critério as moedas que são levadas em consideração para a valoração dos direitos especiais de saque (*Special Drawing Rights – SDR*)⁶ do Fundo Monetário Internacional (FMI): Dólar norte-americano, euro, libra esterlina, iene japonês e renminbi chinês - este último adicionado em novembro de 2015 e será adicionado efetivamente à cesta de moedas em outubro de 2016. A incorporação da moeda chinesa se justifica pela sua grande relevância para a análise das relações econômicas internacionais, haja vista o massivo volume de exportações do mercado chinês e a influência deste no contexto global (EICHEN-GREEN, 2011). Por fim, o câmbio do real brasileiro em relação a essas cinco moedas mais

⁶O SDR, criado pelo FMI em 1969, funciona como um mecanismo para equilibrar a liquidez da economia mundial. Uma alocação de SDR é atribuído a cada país – a depender do tamanho de sua economia em relação ao mundo – de modo que cada país pode converter SDRs em divisas internacionais e ser utilizado para a realização de transações internacionais, bem como comercializar SDRs com outros países. Para uma discussão mais aprofundada sobre a evolução histórica do SDR e sua importância no sistema financeiro internacional, ver Griffith-Jones e Kimmis (2001).

fortes também foi levado em consideração, dada a relevância político-econômica que o Brasil tem ganhado no ambiente geopolítico internacional na última década e a elevada quantidade de estudos científicos que trataram de mercados de países emergentes e em desenvolvimento (BODART; CANDELON; CARPANTIER, 2015; FERRARO; ROGOFF; ROSSI, 2015; KATUSIIME; SHAMSUDDIN; AGBOLA, 2015; PIERDZIOCH; RÜLKE, 2015; BASHER; HAUG; SADORSKY, 2016) e que consideraram taxas de câmbio em relação ao real brasileiro.

Como objetivo específico, o presente trabalho realizou uma revisão da literatura em relação à produção acadêmica recente relativa à predição da taxa de câmbio, englobando trabalhos de cunho tanto técnico quanto fundamentalista, a fim de mapear o estado da arte do tema estudado, propiciando a identificação de lacunas na construção do conhecimento científico e motivando a produção de pesquisas futuras no referido tema. Particularmente, foi realizado um levantamento de estudos nesse tema que fazem aplicações de técnicas de aprendizado de máquinas, com ênfase em métodos de redes neurais artificiais e SVM (e suas extensões, como o SVR), os quais constituem nas duas principais categorias de metodologias em aprendizado de máquinas vigentes nas produções acadêmicas recentes.

2 REFERENCIAL TEÓRICO

2.1 Caracterização do mercado FOREX

O mercado internacional de câmbios (FOREX, algumas vezes abreviado como FX) é o mercado financeiro com maior liquidez no mundo (SAGER; TAYLOR, 2006; TALEBI; HOANG; GAVRILOVA, 2014; BEKIROS, 2015), com um volume de transação diário que excede 5 trilhões de dólares (OZTURK; TOROSLU; FIDAN, 2016, p. 170). No mercado FOREX, as moedas são negociadas simultaneamente nos dois sentidos (ou seja, operações de compra e venda da mesma moeda são concomitantes), e seus participantes incluem bancos centrais, bancos comerciais, fundos de *hedge*, corporações e indivíduos. O mercado FOREX é notoriamente conhecido por sua alta volatilidade, dado que é um mercado descentralizado: inexistente uma contraparte central, de modo que em geral as transações são realizadas com interação direta entre os participantes do acordo. Essa estrutura traz consigo duas implicações: um elevado grau de fragmentação desse mercado – ou seja, há transações paralelas no mesmo instante temporal fechadas a níveis de preço distintos; e um reduzido grau de transparência – no sentido de haver um entrave adicional para que os preços assimilem as informações, dada a ausência de um ambiente físico para a realização das transações, muito embora haja centros de *trading* físicos que registram as transações realizadas (sem que nenhuma das partes esteja fisicamente presente no local), e que se concentram geograficamente a depender da conveniência para as partes envolvidas¹. Os pares de câmbio de maior liquidez são: dólar norte-americano e euro, dólar norte-americano e libra esterlina, dólar norte-americano e iene japonês, dólar norte-americano e dólar canadense, dólar norte-americano e dólar australiano, e dólar norte-americano e franco suíço (TALEBI; HOANG; GAVRILOVA, 2014, p. 2065).

Para um agente especulador, a estratégia no mercado FOREX é basicamente “comprar barato, vender caro”. Sendo assim, caso o agente possua expectativa que uma moeda **A** irá se valorizar em relação a uma outra moeda **B**, sua decisão seria comprar quantidade da moeda **A**

¹Os centros de *trading* do mercado FOREX com maior volume de transação são Londres, Nova York e Tóquio (SAGER; TAYLOR, 2006, p. 83)

e vender após a apreciação de **A** em relação a **B**, auferindo um ganho no processo. Dada a alta volatilidade e difícil previsão das taxas de câmbio, aliado ao alto custo envolvido em monitorar as variações desse mercado de grande liquidez e à estrutura peculiar do mercado FOREX, a literatura científica tem se motivado especialmente para estudos relativos a desenvolvimento de sistemas de *trading* automático, os quais são algoritmos de transação operados por inteligência artificial e aprendizado de máquina, formulados com o propósito de identificar padrões e tendências no mercado FOREX e com isso encontrar a melhor estratégia para maximizar o ganho (NI; YIN, 2009; BRITO; OLIVEIRA, 2012; EVANS; PAPPAS; XHAFA, 2013; BEKIROS, 2015; TALEBI; HOANG; GAVRILOVA, 2014; DYMOVA; SEVASTJANOV; KACZMAREK, 2016; GERLEIN et al., 2016).

Os fatos estilizados para o mercado FOREX podem ser sintetizados em quatro pontos (BEKIROS, 2015, p. 36): as taxas de câmbio são processos não-Gaussianos, apresentam cauda pesada e autocorrelação, e os momentos de suas distribuições são variantes com o tempo. Enquanto que alguns estudos da literatura têm proposto algumas aproximações das distribuições do câmbio (tais como Pareto-estável, t de Student-não-estável), outros autores descartam a possibilidade de adaptar essas séries a uma única distribuição, abordagem esta que abre espaço para o desenvolvimento de técnicas não paramétricas, como é o caso do aprendizado de máquinas e o SVR.

2.2 Análise técnica e fundamentalista

Estudos relativos a predição da taxa de câmbio que seguem a ótica da análise técnica vêm gerando uma quantidade maior de publicações nos últimos anos em relação a estudos de caráter fundamentalista (KOROL, 2014, p. 50). O embasamento básico da literatura de predição de câmbio é relativo a testar se de fato o *Meese-Rogoff puzzle* se aplica – ou seja, se é possível encontrar uma metodologia que supere robustamente um *Random Walk*, de modo que a essência da análise técnica é tentar identificar padrões que se repetem no passado que possam refletir tendências futuras. Há estudos na literatura que indicam que o comportamento das taxas de câmbio pode não ser completamente aleatório, dentre os quais se destaca Caraiani e Haven (2015), artigo no qual os autores analisam os câmbios diários do euro em relação a sete moedas do leste europeu e dos bálticos (kuna croata, koruna tcheco, florim húngaro, zlot polonês, leu romeno, koruna eslovaco e tolar esloveno) entre 1999 e 2010. O estudo observa evidências de multifractalidade entre as séries temporais, indicando a possibilidade de que as taxas de câmbio podem não ser tão imprevisíveis quanto o preconizado pelos estudos clássicos.

Zhang e Zhao (2009) utilizam SVM para identificar padrões de séries passadas e prever a

direção de variação do câmbio dólar/euro usando indicadores de análise técnica como variáveis independentes. Pierdzioch e Rülke (2015) realizam estudo de previsão da direção futura (valorização ou desvalorização) de taxas de câmbio de mercados emergentes usando curvas ROC (*relative operating characteristic curves*) como medida de performance; Yao et al. (2015) analisam o câmbio do Renminbi em relação a quatro moedas estrangeiras (dólar norte-americano, libra esterlina, euro e iene japonês) usando análise topológica de redes complexas. Há também numerosos estudos relativos à análise técnica incorporando diversas extensões paradigmáticas: a lógica *fuzzy* é utilizada por Gençay e Gradojevic (2010), os quais analisam o contágio à luz da entropia a longo prazo de inseguranças do mercado europeu de opções baseado na crise do “*black monday*” de 1987; Lento (2008) realiza análise técnica do índice S&P 500 à luz das finanças comportamentais. Para discussão mais detalhada acerca da produção científica recente de análise técnica no campo de finanças, ver Bekiros (2015) e Ozturk, Toroslu e Fidan (2016). Apesar de possuir uma popularidade relativamente menor em relação a trabalhos que propõem inovações metodológicas para a identificação de padrões técnicos, estudos de caráter fundamentalista continuam apresentando elevado número de publicações dentro da literatura relativa à previsão da taxa de câmbio. A conclusão de Meese e Rogoff (1983) de que modelos fundamentalistas de previsão ao câmbio não superam o *Random Walk* fez com que a produção científica no tema passasse por um período de relativo “pessimismo” em relação a modelos que incluíam fundamentos macroeconômicos (DĄBROWSKI; PAPIEŻ; ŚMIECH, 2014, p. 148). Porém, vários trabalhos constataram que, mediante uma cuidadosa seleção de parâmetros, modelos fundamentalistas para a previsão de taxas de câmbio fornecem poder explicativo suficiente para contestar o *Meese-Rogoff puzzle* (BECKMANN; SCHÜSSLER, 2016), dentre os quais se destacam os seguintes:

Yin e Li (2014) realizam estudo a respeito da relação entre indicadores macroeconômicos e a dinâmica da taxa de câmbio dólar/euro, encontrando evidências de que esses indicadores de fato ajudam a captar grande parte das variações no câmbio. Ademais, os autores realizam um teste de robustez do modelo, testando a significância para as taxas de câmbio do dólar em relação à libra esterlina, ao iene japonês e ao marco alemão (como *proxy* ao euro para o período anterior da introdução deste). Os resultados indicam que o poder explicativo dos fundamentos macroeconômicos mantiveram-se, reforçando a hipótese de que a introdução destes em modelos de previsão do câmbio ajudam a incrementar seu poder preditivo.

Dąbrowski, Papież e Śmiech (2014) fazem uso da análise de dados em painel para analisar a relação entre câmbio e fundamentos monetários levantados pelo *monetary model* de Mark (1995), este um dos primeiros trabalhos a apresentar questionamentos diretos ao *Meese-Rogoff puzzle*. O estudo abarcou 8 países da Europa oriental e central, e encontrou evidências (verifi-

cada mediante teste de Granger) de que as variáveis fundamentalistas levantadas apresentam uma relação de causalidade para com as taxas de câmbio, além de observar uma evidência de cointegração entre as duas partes. Dick, MacDonald e Menkhoff (2015) realizam estudo similar com câmbios do euro em relação ao dólar, à libra esterlina e ao iene japonês, constatando performance de previsão satisfatória para os fundamentos macroeconômicos, em especial a taxa de juros; além disso, o estudo encontra evidências de que a previsibilidade do câmbio, mesmo em horizontes temporais relativamente menores (dados mensais) é satisfatória, enquanto que o poder explicativo cai à medida que o lapso temporal aumenta. (RIVERA-PIZA, 1996; LAWS; THOMPSON, 2004; CHEUNG; CHINN; PASCUAL, 2005)²

Explorando mais a fundo o potencial de modelos fundamentalistas para superar o *Random Walk*, Byrne, Korobilis e Ribeiro (2016) lançam a ideia de levar em consideração o caráter também altamente dinâmico e imprevisível dos próprios fundamentos macroeconômicos, introduzindo a estes parâmetros variantes com o tempo via estimação Bayesiana. Ao considerar que as condições macroeconômicas também sofrem constantes e inesperadas mudanças, o modelo busca incorporar mais fatores que aumentem o poder preditivo de modelos fundamentalistas; porém, os resultados obtidos não apresentaram elevação significativa desse poder em relação a modelos com parâmetros constantes.

Fazendo uma confluência da abordagem fundamentalista em relação a estudos de análise técnica, Buncic e Piras (2016) formulam um modelo definido como uma média ponderada entre um modelo de previsão puramente técnico (“grafista”) e um outro puramente fundamentalista, e verificar a performance desse modelo combinado para prever o movimento de seis pares de câmbio (dólar norte-americano em relação a euro, libra esterlina, iene japonês, dólar australiano, dólar canadense e franco suíço). Mediante análise de métricas utilizadas para avaliar o erro de previsão (MFSE (*Mean Squared Forecast Error*), coeficiente de determinação (R^2) fora da amostra e estatística do teste CW (CLARK; WEST, 2007)) , o estudo mostra que a abordagem fundamentalista apresentou melhor capacidade preditiva que a abordagem grafista. Ademais, o artigo evidencia que o modelo combinado obteve performance significativamente melhor que o modelo *Random Walk* para todos os seis pares de câmbio, um resultado que fornece suporte para a abordagem do presente trabalho, que também irá aliar elementos da análise fundamentalista com a metodologia de aprendizado de máquina SVR.

²Laws e Thompson (2004) flexibilizam esse prazo para até três meses. Para horizontes maiores, não há evidências de ganhos em relação ao *Random Walk*

2.3 Aprendizado de máquinas: aplicações em finanças

Séries financeiras tendem a apresentar comportamento de difícil previsão, verificado pela constatação de alguns fatos estilizados, tais como distribuições probabilísticas de cauda pesada, conglomerados de volatilidade, etc. (SEWELL, 2011). Dessa forma, abordagens de *data mining*, que buscam extrair o máximo desses dados “caóticos” são de grande atratividade para o pesquisador dessa área, motivando o uso crescente de metodologias associadas ao aprendizado de máquina. Ravisankar et al. (2011), por exemplo, utilizam do *data mining* por aprendizado de máquinas para identificar fraudes em demonstrativos financeiros com base em uma amostra de 202 empresas listadas nas bolsas de valores da China. O estudo fez uso de métodos como redes neurais MLP, SVM, algoritmo genético e regressão logística, e chega à conclusão de que o uso do aprendizado de máquinas contribui para a melhor identificação de padrões – cujo nível de complexidade tende a extrapolar consideravelmente àqueles vistos mediante uma análise descritiva simplista –, de modo a aumentar a acurácia das previsões (nesse caso, classificação binária da empresa como fraudulenta ou não fraudulenta).

Lin, Hu e Tsai (2012) apresentam um estudo bibliométrico mapeando 130 publicações entre 1995 e 2010 sobre previsão de crises financeiras usando grande variedade de metodologias de aprendizado de máquinas, evidenciando o notório crescimento dessa abordagem na literatura científica contemporânea. Ademais, nos últimos anos o aprendizado de máquinas tem motivado vastas aplicações na previsão de séries temporais financeiras, apresentando resultados satisfatórios e justificando o desenvolvimento dessa agenda de pesquisa na área de finanças (ATSALAKIS; VALAVANIS, 2009; KROLLNER; VANSTONE; FINNIE, 2010; ABIDIN; JAAFAR, 2012).

Sun e Li (2012) usam o SVM para prever situações de *financial distress* com uma amostra de 135 empresas listadas nas bolsas de valores chinesas. O artigo comparou a performance de SVM de quatro *Kernels* distintos (linear, polinomial, gaussiano e sigmoidal) com a ponderação entre eles via análise multi-discriminante e análise de componentes principais. Os resultados mostram que o classificador com ponderação de *Kernels* possui performance significativamente melhor que classificadores com *Kernels* individuais, porém o resultado oscila com o número de *Kernels* utilizado para a ponderação; concomitantemente, o estudo aponta que, para o contexto específico de predição do *financial distress*, o *Kernel* Gaussiano é a melhor escolha para classificadores com apenas um *Kernel*.

No âmbito da análise técnica de séries temporais financeiras, aplicação do SVM é realizada pelo artigo de Gong et al. (2016), que faz uso de um SVM com *Kernel* Gaussiano, acoplado com o algoritmo de busca “extended UCR Suite”, para prever o comportamento do índice HSI (Heng-

Seng Index) entre 2003 e 2013. Os resultados também mostram que a vantagem que o SVM possui em reconhecer padrões durante seu treinamento potencializa a capacidade preditiva da variável dependente em questão.

A extensão SVR também tem sido aplicada com êxito no campo de finanças, em especial para a predição de séries temporais, dado que retorna uma estimativa numérica em vez de simplesmente classificar a variável dependente em torno de uma classe binária. No tocante à eficiência do SVR em relação a arquiteturas de aprendizado de máquinas baseadas em redes neurais, Tay e Cao (2001) realizaram estudo aplicando os dois modelos para a predição do retorno de títulos de dívida pública de quatro países (Estados Unidos, Canadá, Alemanha e França), além do índice S&P 500. Os modelos propostos foram comparados segundo as métricas Erro Quadrático Médio Normalizado (NMSE), Erro Médio Absoluto (MAE), Simetria Direcional (DS) e a Simetria Direcional Ponderada (WDS), os quais favoreceram o modelo SVR em detrimento do modelo de rede neural.

Resultado similar é encontrado por Beltrami, Loch e Silva (2011), cujo trabalho compara o desempenho de duas abordagens de aprendizado de máquinas – uma rede neural do tipo *Back-Propagation Neural Network* (BPNN)³ e um SVR – com o modelo clássico de precificação de Black e Scholes (1973). O estudo chega à conclusão de que o modelo SVR apresentou acurácia superior em comparação com os demais. Enquanto o resultado encontrado não refuta a validade do modelo de *Black & Scholes*, demonstra a relevância do desenvolvimento de estudos de aprendizado de máquinas aplicados em finanças, estes que possuem potencial de realizar extensões construtivas a trabalhos seminais que sedimentam este campo de estudos.

2.4 Previsão da taxa de câmbio: Estado da arte

Conforme descrito na seção anterior, a produção científica recente apresenta indícios de que o poder explicativo de modelos fundamentalistas pode de fato superar um *Random Walk*, contrariando a conclusão do *Meese-Rogoff puzzle*. A própria literatura identifica uma possível causa para esse fato: o incremento do poder explicativo está associado à introdução da não linearidade (CLARIDA; WALDMAN, 2008; CONRAD; LAMLA, 2010; CHAO; SHEN; ZHAO, 2011). Dado que o estudo de Meese e Rogoff (1983) testa apenas modelos estruturais com forma funcional linear, a incorporação de modelos não lineares se mostra como uma extensão natural para se justificar o uso de fundamentos macroeconômicos e o desempenho desejável evidenciado nos

³A arquitetura do BPNN consiste basicamente em um treinamento supervisionado de dois estágios, tal que os resultados calculados baseados nos *inputs* são ponderados retroativamente segundo algum critério de correção de erros (BELTRAMI; LOCH; SILVA, 2011, p. 576)

artigos supracitados (BURNS; MOOSA, 2015). Dessa forma, a premissa postulada por Meese e Rogoff (1983) pode ser interpretada como um reflexo da imprevisibilidade linear das taxas de câmbio, de modo que a conveniência matemática dos modelos lineares acaba culminando numa perda de poder explicativo, visto que insistir na linearidade constituiria num equívoco, quando há diversas evidências que esta não se verifica. (SHEN; CHAO; ZHAO, 2015, p. 243; p. 251).

Os resultados do *Meese-Rogoff puzzle*, ao identificar as fragilidades e o pobre poder explicativo que os modelos lineares oferecem, motivaram a literatura científica em relação à busca de modelos que incorporaram não-linearidades a fim de verificar se o poder preditivo para as taxas de câmbio apresenta melhorias. Kilian e Taylor (2003) desenvolvem estudo focando nesse ponto e lançam a hipótese de que o comportamento das séries de taxas de câmbio seguem um modelo autorregressivo de transição exponencial suavizada (*exponential smooth transition autoregressive* – ESTAR). A conclusão do artigo aponta evidências de que a previsibilidade dessa variável supera um *Random Walk* para horizontes temporais acima de 3 anos, fornecendo embasamento para o uso de modelos fundamentalistas. Ahmed, Liu e Valente (2016) apontam que os modelos lineares, mesmo os desenvolvidos recentemente (GREENAWAY et al., 2014; ENGEL; MARK; WEST, 2015), apresentam sérias limitações no tocante à previsão das taxas de câmbio; em especial, fatores de risco que derivam de modelos lineares – como o Índice de Sharpe – também são variáveis que não contribuem para o aumento da capacidade preditiva, e modelos que os utilizam tendem a oferecer quase nenhum poder explicativo adicional ao modelo do *Random Walk*.

Dado que há evidências de que podem existir estruturas de dependência não lineares para as taxas de câmbio, a introdução de modelos de estimação não-lineares torna-se uma opção atrativa pra tentar lograr êxito em superar um modelo *Random Walk* (BISSEONDEAL et al., 2008). Santos, Costa e Coelho (2007) mostram que modelos lineares possuem acurácia de previsão inferior a redes neurais artificiais, ao analisar a taxa de câmbio real/dólar entre 2000 e 2004 para dados de alta frequência (intervalos de 15, 60 e 120 minutos). López-Suárez e Rodríguez-López (2011) apresentam um modelo com termo de correção de erro não linear para avaliar a previsibilidade da taxa de câmbio, chegando à conclusão de que essa previsibilidade existe de fato ao se considerar estruturas de associação não lineares entre o câmbio e as variáveis independentes. Junttila e Korhonen (2011) realizam procedimento similar, corroborando a existência de relações não lineares entre taxa de câmbio e fundamentos macroeconômicos. Por conseguinte, o estudo sustenta a validade de modelos fundamentalistas especificadas com não linearidades, indicando potencial dessa classe de modelos em predizer a taxa de câmbio para horizontes temporais de curto prazo, em especial pelo fato de a produção acadêmica de cunho fundamentalista relativa ao tema ainda ser constituída majoritariamente por modelos lineares

de previsão (ROSSI, 2013, p. 1077–1081).

Lei (2013) realiza um estudo de predição das taxas de câmbio do dólar norte-americano em relação ao euro e à libra esterlina, fazendo uso da análise espectral – nas versões uni e multivariada – incorporando não linearidades. Os resultados obtidos apontam que a introdução da não linearidade contribui para que a precisão das estimativas superasse consistentemente o *Random Walk* nas séries estudadas. Pavlidis, Paya e Peel (2015) testam a existência de causalidade entre taxas de câmbio real e nível de consumo para 14 países-membro da Organização para a Cooperação e Desenvolvimento Econômico (*Organisation for Economic Co-operation and Development* – OECD), constatando que as evidências de causalidade aumentam significativamente ao se considerar relações não lineares entre as variáveis, revelando associações entre estas que não seriam captadas fazendo uso de uma métrica linear.

Alguns estudos na literatura de taxa de câmbios também exploraram a possibilidade de modelos multilaterais, modelando não apenas a relação de uma moeda com outra, mas também como e em que medida a cotação de duas moedas apresenta um “efeito contágio” para uma terceira moeda. Klein e Ku (2006) apresentam o argumento de que a apreciação ou depreciação de uma moeda cujo país emissor exerce um certo grau de liderança regional irá influenciar diretamente nas moedas dos países que estejam sob a sua zona de influência (efeito denominado “elasticidade zonal”). Greenaway et al. (2014) e Berg e Mark (2015) apresentam proposta parecida.

Dado que o mercado FOREX é notório pela sua grande liquidez e imprevisibilidade, é atrativa a predição não apenas do valor em si das taxas de câmbio, mas também da volatilidade das taxas de câmbio, tema este que também tem sido motivação de vários estudos ao longo dos últimos anos (BAYOUMI; EICHENGREEN, 1998; ANDERSEN et al., 2001; CORTE; SARNO; TSIKAS, 2011; KAMAL, 2012). Trabalhos relativos à modelagem da volatilidade no mercado de câmbios incluem abordagens que fazem uso do modelo clássico do GARCH (*Generalised Autoregressive Conditional Heteroskedasticity*) de Bollerslev (1986), bem de construtos relativos à volatilidade estocástica (CHOWDHURY; SARNO, 2004) e MS-GARCH (*Markov Switching GARCH*), acoplando o modelo GARCH à lógica da transição de regimes *markov-switching* (ABOUNOORI; ELMI; NADEMI, 2016). Especificamente em relação à análise dos fatores de risco que determinam a volatilidade no mercado FOREX, destacam-se e Lustig e Verdelhan (2007) Bansal e Shaliastovich (2013).

Haja vista a constatação de que estudos tem apontado que o mercado FOREX também apresenta o fato estilizado de *clusters* de volatilidade (LUX; MARCHESI, 2000), a análise de conglomerados (“clusters”) é uma metodologia que tem sido empregada com frequência na literatura. Baasher e Fakhr (2012) utilizam análise de conglomerados para prever a tendência de

variação da taxa de câmbio diária; Abraham (2002) realiza estudo similar para taxas de câmbio médias mensais; Addam et al. (2016) realizam uma seleção de *clusters* hierárquica para construir um sistema de *trading* automático, acoplado com um *framework* que permite a obtenção de dados do mercado FOREX em tempo real e a rápida tomada de decisão do agente participante.

O critério da liquidez é recorrentemente utilizado para delimitar o objeto de estudo (*i.e.*: quais pares de câmbio figurar-se-ão como as variáveis dependentes a serem preditas), um resultado natural, visto que a liquidez por si só representa uma das principais características do mercado FOREX; dessa forma, pares de câmbio envolvendo o dólar norte-americano e o euro aparecem com grande frequência nos estudos que têm sido desenvolvidos no tema. Não obstante, moedas emitidas por países da categoria “emergentes” também gozam de grande popularidade nas recentes publicações científicas, muito em razão das recentes transformações político-econômicas mundiais, acompanhada de uma ascensão de atores internacionais previamente marginalizados do centro decisório da agenda de governança global (EICHENGREEN, 2011; KOSE, 2011; HURRELL, 2013), notadamente em relação à agenda do comércio internacional e da constituição de cadeias globais de valor (FIORENTINO; CRAWFORD; TOQUEBOEUF, 2009; BALDWIN, 2012) e evidenciado em transformações na configuração de poder do sistema financeiro internacional, haja vista a recente inserção do renminbi chinês na cesta de valoração do SDR.

Estudos correlatos à predição da taxa de câmbio envolvendo diretamente mercados emergentes como objeto de estudo incluem os trabalhos de Kohlscheen (2014), Bodart, Candelon e Carpentier (2015) e Pierdzioch e Rülke (2015). Em relação a estudos que trazem resultados específicos a mercados emergentes, destacam-se o de Slavtcheva (2015), que aborda as influências da política de regime de câmbio adotada por um país no seu crescimento da produtividade e desenvolvimento financeiro; e o de Cheung e Rime (2014), o qual analisa microestruturas do mercado FOREX para analisar as conexões entre os câmbios *onshore* (renminbi da China continental) e *offshore* (dólar cantonês) do mercado chinês.

A abordagem de estudo de eventos também é utilizada em estudos relativos a taxas de câmbio, dado que estas são fortemente influenciadas por decisões políticas e vicissitudes psicológicas dos agentes econômicos, cujas nuances são de difícil modelagem e alta instabilidade. Assim, modelos de estudos de eventos analisam o efeito que um acontecimento (“evento”) exerce sobre determinada(s) variável(is), verificando a existência ou não de uma quebra estrutural na chamada “janela pós-evento” (CORRADO, 2011). Estudos que analisam a eficiência e assimetria informacionais aliados a essa metodologia também podem ser encontrados na literatura (FATUM; HUTCHISON; WU, 2012; CHABOUD et al., 2014). Evans e Lyons (2005) consideram o impacto de

notícias relativas a mudanças de fundamentos econômicos no mercado cambial para períodos temporais mais longos, constatando uma espécie de “persistência” de um mesmo anúncio que condiciona o comportamento do câmbio durante vários dias. Menkhoff (2010) e Li, Wong e Genev (2015) apresentam estudos bibliométricos acerca da produção científica recente em relação à análise dos efeitos da divulgação de notícias macroeconômicas no mercado cambial, destacando a pertinência desse tema de pesquisa e encaminhamentos de possíveis trabalhos futuros, com ênfase na vinculação da abordagem “*big data analysis*”.

Ehrmann e Fratzscher (2005) estudam o efeito de transmissão de notícias macroeconômicas para as taxas de câmbio dólar/euro, mostrando que a sensibilidade do câmbio frente às notícias é maior em períodos de alta volatilidade no mercado ou quando se observa um choque negativo de grande magnitude. Usando um modelo de regressão linear, o estudo explica aproximadamente 75% da variação direcional da taxa de câmbio estudada (sem fornecer estimativas da magnitude dessa variação). Dessa forma, extensões a esse estudo com modelos que incorporam não linearidades e que considerem estimativas numéricas – como o SVR – podem ser considerados em pesquisas futuras, considerando também o fato de o tamanho escolhido para as janelas de estimação influenciar decisivamente na performance do modelo (GIACOMINI; ROSSI, 2010).

Redl (2015) combina essa abordagem com um modelo de vetores autorregressivos estruturais para analisar o efeito da divulgação de notícias no movimento da taxa de câmbio dólar/libra esterlina. A conclusão do estudo reforça a importância dos fatores políticos no comportamento das taxas de câmbio, o qual está inerentemente atrelado a choques exógenos de difícil modelagem. AbuHamad, Mohd e Salim (2013) realizam estudo de eventos para mensurar o efeito da divulgação de dados macroeconômicos nos Estados Unidos nas flutuações do mercado FOREX, aliando à abordagem *business intelligence*, uma modificação de sistema de apoio à decisão baseado em inteligência artificial. Arghyrou e Pourpourides (2016) realizam estudo similar, analisando o efeito do anúncio das taxas de inflação na variação da taxa de câmbio, e em que medida há respostas assimétricas a depender da direção da variação inflacionária em relação à expectativa dessa variação. Nassirtoussi et al. (2015) agregam a abordagem do aprendizado de máquinas a essa ótica de investigação, analisando o efeito que as manchetes de jornais exercem sobre o mercado FOREX mediante técnicas de *text-mining*.

A incorporação do argumento de transição de regimes para o estudo de taxas de câmbio também ganhou popularidade nos últimos anos. A ideia básica é assumir que as variáveis possam apresentar comportamentos distintos a depender de qual “regime” essas variáveis estão situadas – ou seja, a cada instante temporal, as variáveis podem estar situadas em um estado

distinto de uma cadeia de Markov, cada qual desses estados possuindo um processo gerador de dados distinto que irá condicionar o comportamento das variáveis. Essa abordagem, denominada modelos de transição de Markov (*markov-switching models*), é particularmente atrativa para a modelagem de ciclos econômicos e cenários de crise macroeconômica sistêmica, coadunando com o cenário observado na economia mundial na última década (ISMAL; ISA, 2006; DUEKER; NEELY, 2007; HUANG; HUANG, 2007). A ideia do *markov-switching* vem sendo combinada com outras abordagens como processos de memória longa (SHI; HO, 2015) e aprendizado de máquina, ampliando ainda mais seu horizonte de aplicabilidade (BEKIROU, 2015).

Especificamente em relação à previsão de taxas de câmbio, a abordagem *markov-switching* é utilizada por Basher, Haug e Sadosky (2016) para verificar a significância de choques no nível de preço do petróleo no mercado internacional na predição do câmbio, concluindo que essa variável de fato pode ajudar na acurácia dessa predição; ademais, o trabalho também enfatiza a melhor performance de modelos não lineares em relação a construções lineares para tentar prever o câmbio.

Como extensão natural do modelo de *markov-switching*, os modelos de Cadeias de Markov ocultas (HMM – *Hidden Markov Models*) apresentam o raciocínio reverso: as probabilidades de transição entre os estados (“regimes”), assim como o número de estados, não são conhecidas *a priori*; assim, os modelos HMM tomam como base uma série histórica de estados observados para estimar os estados ocultos (“*hidden states*”) e sua matriz de transições. O artigo de Nootyaskool e Choengtong (2014) utiliza essa abordagem para prever o câmbio do baht tailandês em relação ao dólar norte-americano, obtendo resultados satisfatórios de acordo com a métrica de Erro Médio Absoluto Percentual (*Mean Absolute Percentage Error* – MAPE) total de 0.167%.

Outra frente de inovação em trabalhos relativos à previsão de taxas de câmbio é o uso da teoria da lógica *fuzzy*, a qual constitui um contraponto à lógica binária, ao presumir que a relação de um determinado elemento em relação a um conjunto de interesse “fuzzy” possa seguir uma escala gradativa (XIA; XU; ZHU, 2012). Em outras palavras, em vez de um elemento ser classificado simplesmente como “sim” ou “não” em relação a algum atributo de interesse, qualquer extensão de ordinalidade entre os extremos “sim” e “não” é possível. Korol (2014) elabora um estudo de previsão das taxas de câmbio de três moedas (iêne japonês, libra esterlina britânica e franco suíço) em relação ao dólar norte-americano, com a lógica *fuzzy* para a definição das variáveis explicativas, cujos valores implicam em gradações distintas a serem levadas em consideração na predição dos câmbios em si. Dymova, Sevastjanov e Kaczmarek (2016) propõe uma extensão a essa abordagem, permitindo que as gradações associadas às variáveis possam ter

interseções, elevando a complexidade da definição das classes das variáveis. Os autores então desenvolvem um sistema de *trading* baseado nessa “lógica *fuzzy* estendida” e aplicam-na em dados reais do mercado FOREX.

Aplicações à previsão das taxas de câmbio também foram verificadas em estudos relacionados à econofísica, cuja proposta central é aliar métodos atrelados à modelagem de problemas da física a problemas de economia e finanças. Por exemplo, a hipótese de que os preços de ações seguem um *Random Walk* pode ser modelada por um movimento Browniano padrão clássico, como visto em aplicações notórias de econofísica em finanças tais quais o trabalho seminal de Black e Scholes (1973). No contexto das taxas de câmbio, Muriel (2004) apresenta um modelo de equação cinética utilizada para modelar turbulências em sistemas físicos para prever a taxa de câmbio do franco suíço em relação ao dólar norte-americano e ao euro; Pinčak (2013) realiza uma aplicação da teoria das cordas da física quântica à predição do câmbio dólar/euro, analisando cenários com ou sem custos de transação. O modelo SVM foi utilizado como *benchmark*, e o estudo indica que a abordagem da teoria das cordas pode apresentar performance igual ou superior ao SVM para previsões de um período a frente.

2.4.1 Aprendizado de máquinas na previsão da taxa de câmbio

Em relação à literatura específica para a predição da taxa de câmbio, a aplicação do aprendizado de máquinas pode ser dividido em duas frentes principais: redes neurais artificiais e máquinas de suporte vetorial (LEE; OW; LING, 2014, p. 2) . O uso das redes neurais para a predição das taxas de câmbio remonta ao trabalho de Kuan e Liu (1995), o qual analisa cinco pares de câmbio com uma rede neural do tipo FFNN (*Feed Forward Neural Network*)⁴. Os resultados indicam que a introdução da abordagem do aprendizado de máquinas pode melhorar a precisão das estimativas. Posteriormente, Rivera-Piza (1996) mostra que previsões realizadas com uma rede neural artificial conseguem identificar tendências não lineares em dados históricos, obtendo estimativas satisfatórias para horizontes temporais menores que 30 dias. O estudo de Gradojevic (2002) introduz não linearidades à modelagem de microestruturas de mercado para as taxas de câmbio, mostrando que o uso de redes neurais artificiais produz resultados superiores a modelos lineares. Os resultados dos três trabalhos supracitados, além de mais uma vez fornecer embasamentos para questionar o *Meese-Rogoff puzzle*, dá subsídios para constatar os ganhos que o aprendizado de máquinas pode propiciar para a capacidade de previsão das

⁴A arquitetura FFNN está entre as mais simples dentre as redes neurais, dado que a informação é processada em sentido único, sem *loops* ou retroalimentações (AUER; BURGSTEINER; MAASS, 2008)

taxas de câmbio⁵.

A metodologia de redes neurais ainda possui presença bastante ativa em publicações científicas, conforme evidenciado pelo artigo recente de Galeshchuk (2016), o qual utiliza uma rede neural do tipo *Multilayer Preceptron* (MLP)⁶. A autora analisa três modelos para três frequências temporais distintas – com dados diários, mensais e trimestrais. O desempenho dos modelos é analisado pela métrica *relative prediction error* (GALESHCHUK, 2016, p. 448). Os resultados indicam que a predição das taxas de câmbio diária e mensal possuem boa precisão para um período temporal à frente, enquanto que, no horizonte trimestral, a previsão fica comprometida; Rossi (2013) chega a conclusão similar, constatando que a acurácia dos modelos de taxas de câmbio é variante a depender da escala temporal considerada, de modo que a performance preditiva dos modelos encontrados na literatura são altamente instáveis.

Comparando a performance de redes neurais frente a modelos econométricos clássicos como o ARMA (*Autoregressive Moving Average*) e o GARCH, Bissoondeal, Karoglou e Gazely (2011) mostram que as redes neurais apresentam melhor desempenho preditivo em relação a esses modelos tradicionais de séries temporais, muito embora a acurácia dependa fortemente da seleção dos parâmetros para o algoritmo, processo de seleção este que não possui critérios ou heurísticas que convirjam para uma escolha ótima, de modo que abordagens de *grid search* ainda são recorrentes em estudos de aprendizado de máquinas.

Ao longo dos anos, o uso de redes neurais artificiais vem aparecendo com frequência na literatura relativa a taxas de câmbio: He e Shen (2007) propõem uma ponderação ótima dos resultados de vários modelos de redes neurais, os quais são construídos por métodos de *bootstrap*. Os modelos foram avaliados com dados das taxas de câmbio diárias do dólar norte-americano em relação a 6 moedas distintas (dólar australiano, libra esterlina, dólar canadense, euro, franco suíço e iene japonês), entre 2003 e 2006. A performance foi medida pela métrica Erro Médio Quadrático Normalizado (*Normalized Mean Square Error* – NMSE).

Majhi, Panda e Sahoo (2009) usam dados simulados para testar a performance de dois modelos baseados em redes neurais (*functional link artificial neural network* – FLANN; e *cascaded functional link artificial neural network* – CFLANN) para a predição das taxas de câmbio da libra esterlina, do iene japonês e da rúpia indiana, comparando os modelos com o *benchmark* do algoritmo de filtro adaptativo por mínimos quadrados médios (*Least Mean Squares* – LMS)⁷. O estudo é expandido por Majhi et al. (2012), que introduzem mais duas construções de redes

⁵Outros estudos que fazem uso do aprendizado de máquinas e que apresentam performance superior a um *Random Walk* incluem Kodogiannis e Lolis (2002), Cao (2003), Chen e Leung (2004)

⁶A abordagem MLP foi primeiramente aplicada ao contexto de previsão de taxas de câmbio por Lisi e Schiavo (1999) e Dunis e Huang (2002)

⁷Para maiores detalhes, ver artigo de Widrow et al. (1977)

neurais: *Wilcoxon artificial neural network* (WANN) e *Wilcoxon functional link artificial neural network* (WFLANN), formulados levando-se em consideração a função perda norma de Wilcoxon⁸. Hua, Zhang e Leung (2010), por sua vez, realizam uma replicação do estudo de Majhi, Panda e Sahoo (2009) agregando o método de suavização exponencial adaptativa, logrando êxito em aprimorar a capacidade de previsão dos pares de câmbio em questão.

Ainda em relação a aplicações das redes neurais, Božić, Vukotić e Babić (2011) alia essa abordagem com a decomposição por ondaletas para a previsão da taxa de câmbio do RSD, mediante um procedimento de remoção de ruídos e posterior reconstrução do sinal. Philip, Taofiki e Bidemi (2011) realizam estudo de previsão do comportamento da taxa de câmbio da moeda nigeriana em relação ao euro, ao dólar norte-americano, à libra esterlina e ao iene japonês, comparando o desempenho de uma rede neural MLP com um modelo de cadeia de Markov oculta (*Hidden Markov Model*). Resultados indicam que a rede neural apresentou 81.2% de precisão, superando a abordagem HMM. Estudo similar é produzido por Badea (2013), que usa a mesma construção de rede neural de Philip, Taofiki e Bidemi (2011) para o câmbio dólar/leu romeno. Para uma síntese das aplicações recentes de redes neurais na modelagem de problemas de economia e de finanças, ver o trabalho de Falat e Pancikova (2015), que também tratam de aplicações de algoritmos de programação genética para prever o câmbio dólar norte-americano/dólar canadense.

Agregando a abordagem do aprendizado de máquinas com a análise fundamentalista, Nas-sirtoussi, Wah e Ling (2011) trabalham com uma rede neural MLP de três camadas para prever a taxa de câmbio mensal dólar/libra entre 1996 e 2010. O estudo considerou 3 conjuntos de variáveis fundamentalistas: reservas internacionais, vendas no varejo e comércio internacional de bens e serviços, cada um dos quais composto por diversos indicadores usados para mensurá-los. Cada conjunto fundamentalista forma um *neuron* de entrada para a rede neural, que realiza o processamento numa “camada oculta” e fornece os resultados numa terceira camada de *outputs*. A conclusão do artigo afirma que o uso da rede neural permite identificar relações complexas de modo a prever melhor o comportamento do mercado; ademais, reforça a importância de se considerar variáveis fundamentalistas, cuja inserção de fato contribui para a elevação do poder de previsão do modelo.

Com a boa performance e o conseqüente crescimento do uso do aprendizado de máquinas em finanças e temáticas correlatas, extensões mais sofisticadas de redes neurais artificiais também têm sido apresentadas com frequência nos últimos anos. Estudos com essa vocação incluem o de Rout, Majhi e Mohapatra (2012), o qual testa a viabilidade da predição da taxa de

⁸Para maiores detalhes, ver artigo de Hsieh, Lin e Jeng (2008)

câmbio para horizontes temporais longos – períodos de 6, 9 e 12 meses. Os autores lançam mão da RBFNN (*Radial Basis Function Neural Network*) para tentar prever o câmbio do iene japonês e da rúpia indiana em relação ao dólar em lapsos temporais longos. Os resultados indicam que o RBFNN apresenta melhor desempenho em comparação com as modelagens MLANN (*Multilayer artificial neural network*) e FLANN (*functional link artificial neural network*)⁹. Outro desenvolvimento da abordagem das redes neurais é apresentado por Sermpinis et al. (2012), o qual estuda a previsão da taxa de câmbio dólar/euro entre 2002 e 2010 com uma arquitetura de rede neural “Psi Sigma” (*Psi Sigma Neural Network – PSN*), cujas estimativas são ajustadas por uma série de métricas, tais como filtro de Kalman, média simples, média bayesiana e operador LASSO (*Least Absolute Shrinkage and Selection Operator*). O modelo proposto é comparado com um modelo ARMA de séries temporais, além de duas outras arquiteturas de rede neural (*Multi-Layer Perceptron – MLP*; e *Recurrent Neural Network – RNN*¹⁰), ambos de uso recorrente na literatura científica (HONG, 2006). O estudo indica que todas as abordagens de redes neurais apresentaram melhorias significativas no poder de previsão da taxa de câmbio estudada (resultado não observado para o modelo ARMA), de modo que o PSN apresentou a melhor acurácia; por outro lado, o filtro de Kalman mostrou-se ser a melhor abordagem de ajuste a ser combinada às estimativas das redes neurais.

Khan, Anwer e Banik (2013) aliam a lógica *fuzzy* com a introdução de não linearidades advindas do modelo de redes neurais, culminando num sistema de inferências *fuzzy* com redes neurais artificiais (*Artificial Neural Network Fuzzy Inference System – ANFIS*). Esse modelo é comparado ao modelo autorregressivo Markov Switch (MS-AR), o qual incorpora a possibilidade de transição de regimes para os dados. Os dois modelos são avaliados quanto à sua capacidade de previsão das taxas de câmbio diárias do dólar norte-americano em relação ao dólar canadense e ao taka bengali, no período de 1996 a 2013. As medidas de diagnóstico da acurácia foram correlação e raiz do erro quadrático médio (*Root Mean Square Error – RMSE*). Os autores chegam à conclusão de que a abordagem ANFIS supera o MS-AR para a predição das taxas estudadas, chamando atenção para o fato de ambas as abordagens considerarem não linearidades, e propõem possíveis estudos futuros com a introdução da lógica *fuzzy* para algoritmos de programação genética e a modelagem *Markov-Switching* também para a volatilidade (MS-GARCH).

Evans, Pappas e Xhafa (2013) realizam estudo de predição da taxa diária dos três pares de câmbio mais transacionados no mercado FOREX (dólar-euro, dólar-libra e euro-libra) entre

⁹Hua, Zhang e Leung (2010) mostra que o modelo FLANN é superior ao modelo MLANN, apesar de sua arquitetura de camada única.

¹⁰Arquitetura de rede neural que permite recursividade e condicionamentos a processamentos anteriores; são a contrapartida para redes neurais do tipo FFNN (GOLLER; KUCHLER, 1996)

2010 e 2012, integrando a lógica dos algoritmos genéticos numa rede neural FFNN para definir estratégias de *trading*. Foi realizada comparação com modelos de séries temporais (médias móveis e ARMA), e novamente a abordagem do aprendizado de máquinas apresentou melhor acurácia, logrando uma taxa de acerto superior a 70% em todas as estratégias nos três pares de câmbio. Durante a discussão dos resultados, o autor sugere a replicação do estudo considerando o classificador SVM, evidenciando o fato de a literatura científica ter constatado em diversos estudos as maiores robustez e precisão da abordagem SVM em relação a arquiteturas de redes neurais (EVANS; PAPPAS; XHAFA, 2013, p. 1265). De fato, a produção acadêmica aplicando o SVM e suas extensões para o contexto da predição do câmbio é numerosa e crescente, conforme elucidado a seguir:

Entre os primeiros trabalhos que fizeram uso do SVM está o de Kim (2003), que utilizou um SVM para prever a direção futura dos preços diários (entre 1989 e 1998) de fechamento de ações que compõem o índice de mercado da Coreia do Sul, comparando com uma rede neural do tipo BPNN. O estudo considerou 12 indicadores de análise técnica como variáveis independentes para a classificação, chegando ao resultado de que o SVM obteve um desempenho superior em relação ao BPNN, com 57,83% de acurácia. Porém, o modelo BPNN apresentou 54,73% de acurácia, mostrando ser também desejável em relação a um modelo *Random Walk*, evidenciando que a abordagem do aprendizado de máquinas pode ajudar a relativizar o *Meese-Rogoff puzzle*.

Kamruzzaman, Sarker e Ahmad (2003) testam o desempenho do SVM para a predição da taxa de câmbio do dólar australiano em relação a seis moedas estrangeiras (dólar norte-americano, libra esterlina, iene japonês, dólar de cingapura, dólar neozelandês e franco suíço) no mercado FOREX, testando para quatro *Kernels* distintos (Gaussiano, polinomial, linear e *spline*). As medidas de Erro Quadrático Normalizado Médio (*Normalized Mean Square Error – NMSE*) e Erro Absoluto Médio (*Mean Absolute Error – MAE*) foram utilizadas para avaliar a acurácia das máquinas, mostrando que os *Kernels* Gaussiano e polinomial obtiveram os melhores resultados. Ademais, o estudo conclui que o modelo baseado no SVM apresentou resultados melhores em termos de previsão do que modelos baseados em redes neurais¹¹. Esse resultado é ratificado posteriormente por Cao, Pang e Bai (2005), os quais utilizam um SVM com *Kernel* Gaussiano para analisar o câmbio diário dólar/libra entre 2003 e 2005, fazendo previsões um período à frente. Os resultados mostram que o SVM de fato acrescenta poder preditivo, via métrica de diagnóstico Erro Quadrático Médio (*Mean Square Error – MSE*).

Com a popularização em pesquisas de ponta no campo de aprendizado de máquinas e

¹¹Kamruzzaman e Sarker (2004) apresentam resultado similar.

seu consequente desenvolvimento nos últimos anos, várias extensões ao SVM foram propostos para a predição do câmbio. Yuan (2013), por exemplo, estudou a previsão da direção da taxa diária de câmbio dólar/renminbi chinês entre janeiro e outubro de 2006, fazendo uso de um novo modelo denominado “SVM suavizado polinomial” (*polynomial smooth support vector machine – PSSVM*). Os resultados mostram que a taxa de acerto para todos os três conjuntos de teste superou 85%, apontando que, mesmo sendo uma variável de alta volatilidade e passível de ser influenciada por diversos fatores de difícil mensuração ou modelagem, o uso do SVM pode aumentar significativamente a previsibilidade da taxa de câmbio.

Conforme apresentado nas seções anteriores, dado que o SVM é um algoritmo de classificação, sua extensão natural ao se considerar a existência de “infinitas classes” é o SVR, de modo que o algoritmo fornece um valor numérico para a variável dependente, exatamente como numa regressão. No contexto das taxas de câmbio, é certamente atrativo para o investidor e demais participantes do mercado ter em mãos uma estimativa numérica dessa variável, ampliando ainda mais os insumos para uma boa tomada de decisões frente a um ambiente de alta incerteza em relação a usar o SVM, o qual fornece um indicativo apenas de “sobe” ou “desce”. Mais que a direção, a magnitude dessa variação pode ser igualmente – senão mais importante – que saber a direção dessa variação, notadamente em cenários onde a direção é razoavelmente previsível.

Em relação ao uso do SVR na predição de taxas de câmbio, o artigo de Ince e Trafalis (2006) analisa os câmbios diários do dólar norte-americano em relação ao euro, à libra esterlina, ao dólar australiano e ao iene japonês entre 2000 e 2004. Os autores adotaram um processo em dois estágios para o modelo: primeiramente, usaram técnicas paramétricas (ARIMA e VAR – Vetor Autorregressivo) para selecionar os parâmetros a serem utilizados no segundo estágio, no qual são testadas duas abordagens não paramétricas (SVR e rede neural do tipo MLP). As métricas utilizadas para a mensuração da performance foram Erro Quadrático Médio (MSE) e Erro Absoluto Médio (MAE). Os autores chegaram à conclusão de que, tanto para os parâmetros selecionados pelo ARIMA quanto pelo VAR, as previsões apresentadas pelo SVR foram mais precisas que as fornecidas pelo MLP¹²; por outro lado, o estudo também indica que o critério de seleção dos melhores parâmetros varia entre os algoritmos, e que essa seleção faz diferença na performance do modelo, assim como a escolha adequada da função *Kernel*.

Para a modelagem da volatilidade no contexto da previsão do câmbio, a força do SVR também merece destaque, conforme elucidam Premanode e Toumazou (2013), os quais usam SVR combinado com o método de decomposição de sinais por modo empírico diferencial (*Differential*

¹²Were et al. (2015), apesar de não abordarem a temática de taxa de câmbios, comparam o SVR a um modelo de rede neural, com o primeiro sendo apontado como melhor.

Empirical Mode Decomposition) para modelar a volatilidade do mercado FOREX, considerando dados entre 1998 e 2010 de vários pares de câmbio (dólar norte-americano/euro, dólar norte-americano/iene japonês, dólar norte-americano/dólar australiano, e euro/franco suíço), além de várias variáveis macroeconômicas: taxas de juros de seis mercados distintos, preços de sete *commodities* no mercado internacional e índice de cinco grandes bolsas de valores (PREMANODE; TOUMAZOU, 2013, p. 379). Após simulação, o modelo é comparado ao MS-GARCH (*Markov Switching GARCH*) por sete métricas de diagnóstico: Erro Quadrático Médio (MSE), Erro Absoluto Médio (MAE), Erro Percentual Absoluto Médio (MAPE), coeficiente de determinação (R^2), critério de informação de Akaike (AIC), critério de informação Bayesiano (BIC) e diferencial relacionado ao preço (PRD). O trabalho conclui que o modelo SVR/EMD diferencial superou a performance do MS-GARCH – assim como do SVR individual – no sentido de demonstrar uma maior capacidade de suavizar e reduzir os ruídos dos dados, permitindo uma maior previsibilidade para a volatilidade. Ademais, o estudo integra com sucesso a metodologia do aprendizado de máquinas com variáveis macroeconômicas, ideia que também consta no presente estudo.

Constatado o sucesso das aplicações do SVR em estudos de finanças, o modelo tem sido aprimorado com a adição de extensões. Ni e Yin (2009) apresentam um modelo “híbrido” de misturas entre o SVR e SOMs (*Self-Organising Maps*) temporais, culminando num algoritmo que constrói um sistema virtual de *trading* para o mercado FOREX¹³. O modelo proposto é testado para a predição do câmbio diário dólar/libra, e seu desempenho é comparado com quatro indicadores de análise técnica – Média móvel exponencial (EMA), convergência/divergência de médias móveis (MACD), índice de força relativa (RSI) e %R de Williams. Os autores concluem que o modelo híbrido levantado obteve rentabilidade maior que os indicadores tradicionais, além de ter logrado êxito em integrar a análise de dados tanto quantitativos quanto qualitativos para a predição do câmbio. Brito e Oliveira (2012) realizam uma extensão desse estudo, aliando o SVR com SOMs de hierarquia crescente (*Growing Hierarchical Self-Organizing Maps* – GH-SOMs) e programação genética, a fim de encontrar a ponderação ótima entre indicadores de análise técnica para maximizar o ganho no mercado FOREX.

A ideia de modelo híbrido também está presente no artigo de Liu (2010), o qual estuda a taxa de câmbio diário dólar/renminbi entre 2005 e 2009. A ideia é combinar o SVR com a transformada discreta de ondaletas, a qual realiza uma decomposição do conjunto de informações e reconstitui as variáveis de interesse com base nas informações mais relevantes. Testando o modelo para os *Kernels* Gaussiano e Spline, o estudo conclui que a abordagem híbrida proposta supera o modelo de SVM individual, segundo avaliação de performance considerando a métrica raiz do erro quadrático médio (RMSE).

¹³O estudo pioneiro que alia SOMs à predição de câmbio é o de Hann e Steurer (1996)

Em relação a extensões ao SVR aplicadas na predição do câmbio, Li e Suohai (2013) apresentam três abordagens derivadas do SVR tradicional para avaliar a previsão da taxa de câmbio diária do renminbi chinês em relação a nove outras moedas (dólar norte-americano, euro, libra esterlina, iene japonês, dólar cantonês, ringuite malaio, rublo russo, dólar canadense e pataca macauense), entre janeiro e dezembro de 2012. As extensões apresentadas foram: *genetic algorithm* (GASVR), *particle swarm optimization algorithm* (PSOSVR) e *artificial fish swarm algorithm* (AFSASVR). A função *Kernel* utilizada para todas as quatro variantes do SVR foi o *Kernel* Gaussiano. Os resultados obtidos pelo estudo mostram que todos os quatro SVRs obtiveram acurácia satisfatória, dentre os quais o AFSASVR obteve as melhores previsões, além de apresentar melhoria significativa na velocidade de convergência computacional à solução ótima.

Proposta semelhante é vista no artigo de Sermpinis et al. (2015), o qual avalia a capacidade preditiva de um modelo híbrido de SVR combinado com *Rolling Genetic Algorithm* (RG-SVR) para taxas de câmbio diárias do euro em relação ao dólar, à libra e ao iene japonês entre 1999 e 2012. O RG-SVR proposto é comparado com 7 modelos de aprendizado de máquinas *benchmark* de uso frequente na literatura correlata: rede neural híbrida que combina funções Gaussianas adaptativas com *Particle Swarm Optimization* (ARBF-PSO), 2 versões do ϵ -SVR (cálculo dos parâmetros simplificado ou por *grid search*), mais 2 versões para o ν -SVR (pelo mesmo critério de diferenciação), e 2 abordagens integradas com algoritmos genéticos (GA-SVM e GA- ϵ -SVR)¹⁴. A métrica RMSE foi utilizada para avaliar a acurácia dos modelos, dentre os quais o RG-SVR proposto pelos autores obteve melhor precisão nas estimativas. Além disso, o estudo conclui de que os modelos SVR *benchmark* possuem alta sensibilidade em relação aos parâmetros escolhidos para a execução, e que estudos futuros relativos a essa agenda de pesquisa têm grande potencial de desenvolver ainda mais a eficiência dessa abordagem.

Extensões do SVR que incorporam a modelagem do caos também estão presentes na literatura científica, visto em artigos como o de Huang et al. (2010), os quais investigam se a inserção de um componente caótico incrementa o poder preditivo dos algoritmos SVR e rede neural. Foram analisadas séries de taxas de câmbio diárias do dólar norte-americano em relação a seis moedas (euro, libra esterlina, iene japonês, dólar australiano, dólar neozelandês e rublo russo) entre 2005 e 2007, comparando quatro modelos: rede neural do tipo BPNN e SVR, cada uma com suas versões com e sem modelagem de caos determinístico. A performance dos modelos foi avaliada com as métricas Erro Quadrático Médio (MSE), raiz do Erro Quadrático Médio (RMSE) e Erro Médio Absoluto (MAE). O modelo que apresentou melhor de-

¹⁴Rehman, Khan e Mahmud (2014) e Deng et al. (2015) apresentam discussões mais aprofundadas acerca da abordagem combinada de programação genética e métodos *Kernel*

sempenho consistente em todas as métricas foi o SVR com abordagem de caos determinístico proposta pelos autores.

Com o rápido desenvolvimento de métodos de aprendizado de máquinas, outras propostas além de redes neurais artificiais e máquinas de suporte vetorial têm emergido na literatura de previsão de taxas de câmbio. Talebi, Hoang e Gavrilova (2014) realizam uma extração de características em escalas múltiplas (*multi-scale feature extraction*) utilizando um classificador Gaussiano multivariado e *Bayesian Voting* para a classificação da tendência futura do câmbio euro/dólar entre 2000 e 2013. Os dados utilizados foram de alta frequência (várias observações *intraday*). O modelo apresenta um sistema de *trading* automatizado com boa precisão para identificar tendências de baixa ou de alta no mercado FOREX relativo ao par de câmbio analisado.

Estudo similar foi elaborado por Zhang e Zhao (2009), os quais fizeram uso do SVM para prever a direção de mudanças na taxa de câmbio diária dólar/euro entre 2007 e 2009. Neste estudo, os inputs para o modelo foram indicadores técnicos, sendo os dados oriundos do sistema Bloomberg no intervalo de 10 de julho de 2007 a 9 de julho de 2009, considerando 5 indicadores técnicos: Média móvel linear (MA), convergência/divergência de médias móveis (MACD), índice de força relativa (RSI), *Random Index* (RSV) e razão *BIAS*. O estudo constata que o SVM apresentou acurácia de previsão superior a 60%, enfatizando na força explicativa dessa abordagem mesmo no mercado cambial, um segmento de alta instabilidade e complexidade do mercado financeiro.

Nos últimos anos, um novo ramo do aprendizado de máquinas denominado "*Deep Learning*" têm ganhado atenção progressiva nas produções acadêmicas recentes, figurando como a nova fronteira do conhecimento, não apenas para a área finanças em específico, mas para todo o desenvolvimento do conhecimento científico em geral. O artigo recente de LeCun, Bengio e Hinton (2015), publicada na conceituada revista "Nature", aponta o *deep learning* como um dos tópicos mais promissores para pesquisas acadêmicas nos próximos anos, na medida em que possuem uma arquitetura de processamento que engloba grande número de camadas. O diferencial dessa abordagem advém do fato de que, a cada camada que o algoritmo processa, o problema de aprendizagem tende a ficar mais simples em termos de complexidade computacional, o que permite a otimização de estruturas bastante complexas com tempo de processamento altamente reduzido, evoluindo significativamente em relação a algoritmos de *grid search*, os quais dependem basicamente de "força bruta" (MONTAVON; BRAUN; MÜLLER, 2011).

Estudos que utilizam *deep learning* também já se fazem presentes na literatura de predição de câmbio. Chao, Shen e Zhao (2011) apresentam um *Deep Belief Network* (DBN) baseado

no algoritmo de aprendizado *greedy* de Hinton, Osindero e Teh (2006) A máquina proposta foi utilizada para prever as taxas de câmbio semanais do dólar em relação à libra esterlina e à rúpia indiana usando dois conjuntos de dados: dados entre 1976 e 1993 e entre 1994 e 2003. O estudo considerou seis medidas de performance dos modelos: raiz do erro quadrático médio (RMSE), erro absoluto médio (MAE), erro percentual absoluto médio (MAPE), precisão da direção (DA), correlação de Pearson e variância. O DBN proposto foi comparado com uma rede neural FFNN e obteve melhor performance preditiva; das seis medidas adotadas para o diagnóstico, o DBN superou o FFNN em todas delas, demonstrando não apenas uma precisão superior das estimativas, mas também uma estabilidade maior.

Os mesmos autores, com um estudo mais recente (SHEN; CHAO; ZHAO, 2015), aplicaram a arquitetura do DBN de Chao, Shen e Zhao (2011) para máquinas de aprendizado de Boltzmann restritos condicionalmente (*Continuous Restricted Boltzmann Machine – CRBM*), um modelo probabilístico que permite trabalhar com dados contínuos, e que conseqüentemente poderia fornecer maior poder explicativo. O modelo foi testado para a análise de três conjuntos de dados distintos, relativos às taxas de câmbio semanais do dólar em relação à libra esterlina (de 1976 a 1993), à rúpia indiana (de 1994 a 2003) e ao real brasileiro (de 2000 a 2004). O DBN proposto foi comparado com um FFNN, baseado nas mesmas métricas de mensuração da performance adotadas por Chao, Shen e Zhao (2011). As conclusões do artigo convergem para o estudo anterior, evidenciando a superioridade do DBN em relação ao FFNN, especialmente em relação em termos de estabilidade (menor variância).

Enquanto Shen, Chao e Zhao (2015) sugerem testar o poder preditivo da abordagem utilizada para dados de taxas de câmbio de frequência diária, o presente estudo almeja incorporar a ideia do *deep learning* para analisar se esta também pode fornecer melhorias para estimativas de horizonte mensal, um prazo que ainda pode ser previsto com relativa acurácia (GALESHCHUK, 2016, p. 448). Dado que a abordagem FFNN constitui numa das principais variantes da análise por redes neurais, os resultados mostrados pelos dois estudos supracitados mostram que arquiteturas do tipo *deep learning* tornam-se especialmente relevantes para explorar melhor o potencial do aprendizado de máquinas, principalmente para extensão em outras áreas do conhecimento, como é o caso do campo de finanças.

3 SELEÇÃO DAS VARIÁVEIS

3.1 Variáveis levantadas

A taxa de câmbio é uma variável de difícil previsão, haja vista o fato de que é, por natureza, vinculada a fatores de difícil mensuração e previsão como decisões políticas, catástrofes naturais, conflitos armados, etc. Assim, qualquer variação desses fatores “latentes” irá impactar decisivamente no comportamento do câmbio, a curto e a longo prazo (YUAN, 2013, p. 942). Sendo assim, é natural que qualquer modelo, seja grafista ou fundamentalista, tenha limitações inerentes à sua especificação, tendo como consequência imediata a correspondente perda na capacidade explicativa oriunda destas limitações. Dessa forma, a seleção das variáveis utilizadas no presente modelo não possui a vocação de ser exaustiva, buscando assim incluir variáveis identificadas na literatura científica como relevantes para a previsão das taxas de câmbio, as quais servirão como *inputs* para a execução do algoritmo SVR.

A seleção das variáveis explicativas fundamentalistas para a previsão da taxa de câmbio foi realizada analisando artigos clássicos da área e de produções científicas recentes correlatas ao tema. A referência principal foi o estudo bibliométrico de Rossi (2013), a qual realiza um levantamento detalhado acerca das principais publicações relativas à previsão do câmbio por análise fundamentalista, cobrindo a evolução da temática desde a introdução do *Meese-Rogoff Puzzle* (MEESE; ROGOFF, 1983). O estudo ainda apresenta uma tabela comparativa explicitando as variáveis fundamentalistas adotadas em cada estudo, bem como a forma funcional e a metodologia consideradas, além da frequência dos dados, a delimitação da amostra, as métricas de avaliação dos erros e o desempenho preditivo (ROSSI, 2013, p. 1076–1081). Para estudos publicados posteriormente ao ano de 2013, procedeu-se para a análise individual dos artigos dessa temática.

Sendo assim, a investigação da produção científica de trabalhos de cunho fundamentalista para a previsão da taxa de câmbio, culminando na determinação de 13 variáveis explicativas, apresentadas na tabela 3.1, juntamente com as principais referências científicas que embasam o seu uso. A variável dependente é a taxa de câmbio nominal *spot*.

3.1.1 Câmbio defasado

A hipótese do *Random Walk* preconiza que o valor esperado para o câmbio em $t + 1$ é simplesmente o câmbio observado em t ; ou seja, a melhor previsão para o período seguinte da variável, para esse tipo de processo, é o valor observado do período imediatamente anterior. Dessa forma, vários autores têm colocado a taxa de câmbio defasada em um período como regressor para o câmbio futuro, apresentando significância em diversos estudos (ENGEL, 1994; CARRIERO; KAPETANIOS; MARCELLINO, 2009), e portanto também foi adotado no presente estudo. No entanto, caso o processo gerador dos dados das taxas de câmbio não siga de fato um *Random Walk*, espera-se que se possa realizar previsões melhores para $t + 1$ do que simplesmente utilizar o valor observado em t com a introdução de outras variáveis explicativas, listadas nas subseções seguintes.

3.1.2 Taxa de juros

A taxa de juros, sendo a remuneração pela renúncia do consumo imediato, é um dos mais importantes instrumentos de política monetária. A taxa de juros é um fator determinante para o fluxo líquido de investimentos estrangeiros, o que por sua vez se traduz em entradas ou saídas líquidas de reservas internacionais, cujo volume irá definir a “oferta” e a “demanda” por moeda estrangeira no ambiente doméstico. Caso haja saída líquida de moeda estrangeira, sua quantidade disponível no mercado doméstico diminuirá, de modo que o câmbio doméstico tende a depreciar, pois observa-se um comportamento análogo a definição do preço de equilíbrio quando, *cæteris paribus*, a quantidade ofertada de um bem diminui; pelo mesmo raciocínio, em caso de entrada líquida de moeda estrangeira, a moeda doméstica tende a valorizar-se. Assim, ao observar o diferencial da taxa de juros entre os emissores de duas moedas consideradas num par de câmbio, caso a taxa de juros para o emissor da moeda **A** apresente um aumento relativo em relação ao juros no emissor da moeda **B**, detentores da moeda **B** tenderão depositar poupanças com uma taxa de remuneração maior, causando uma diminuição no nível de moeda **B** em circulação, e fazendo com que **A** se valorize frente a **B**.

As taxas de juros consideradas no presente estudo foram: *Federal Funds Rate* para o mercado norte-americano, *Bank of England Official Rate* para o mercado britânico, *European Central Bank Refi Rate* para a zona do Euro, *People's Bank Base Rate* para o mercado chinês, *Bank of Japan Rate* para o mercado japonês, e taxa *Selic* para o mercado brasileiro. A taxa de juros tem sido uma das principais variáveis utilizadas na literatura de previsão do câmbio, sendo frequentemente revisitada em estudos correlatos, ainda que sua significância seja oscilante –

isto é, grande número de estudos concluem que a taxa de juros não fornece poder explicativo adicional em relação ao *Random Walk*, assim como há numerosos estudos que apontam o contrário¹.

3.1.3 Oferta monetária

A grosso modo, a taxa de câmbio entre duas moedas é o fator de conversão que traduz o preço relativo entre elas. Dessa forma, é natural que essa variável dependa de fatores que influenciem na quantidade ofertada das moedas envolvidas. Pela lógica da oferta e da demanda, a quantidade ofertada de moeda de ambas as moedas doméstica e externa podem indicar uma tendência da direção de crescimento da taxa de câmbio associada a essas moedas: tomando a variação relativa da oferta monetária de duas moedas, é possível observar se a disponibilidade relativa da moeda **A** aumentou ou diminuiu em relação à disponibilidade da moeda **B**; caso a oferta relativa da moeda **A** tenha aumentado em relação à moeda **B**, **A** tende a desvalorizar-se perante a **B**. O diferencial da oferta monetária tem sido utilizado sistematicamente como um dos principais preditores na literatura de taxas de câmbio, sendo testada com recorrência tanto em estudos que confirmam quanto em estudos que refutam o *Meese-Rogoff Puzzle*. Para o presente estudo, utilizou-se a base monetária (meio de pagamento M1), que diz respeito ao montante total de moeda com alta liquidez, constituído pelo total de papel-moeda em poder do público mais o total de depósitos à vista nos bancos.

3.1.4 Taxa de inflação

A taxa de inflação, além de fornecer o nível de preços de uma economia, reflete o poder aquisitivo da moeda dessa economia. Via de regra, quanto maior é a inflação, maior é a corrosão exercida no poder aquisitivo de uma moeda, dada que a inflação degenera sua reserva de valor; assim, caso a inflação de um país **A** aumente em termos relativos em relação à inflação de uma país **B**, sua moeda tende a se desvalorizar em relação à moeda que – comparativamente – sofreu menos perda de seu poder aquisitivo. Juntamente com a taxa de juros e a oferta monetária, a taxa de inflação foi uma das variáveis testadas no estudo de Meese e Rogoff (1983), e tem sido constantemente considerada em estudos de cunho fundamentalista de previsão do câmbio desde então.

¹A listagem completa se encontra em (ROSSI, 2013, p. 1076–1081)

3.1.5 Índice de produção industrial

Um dos principais preditores para a taxa de câmbio levantados pela literatura é o diferencial do nível de renda, frequentemente mensurado pelo Produto Interno Bruto (PIB) ou alguma medida de produtividade que dependa dessa variável (*e.g.*: PIB per capita ou razão PIB–população economicamente ativa). Porém, o presente estudo não considerou o PIB como variável explicativa, dado que esta é uma estatística disponibilizada trimestralmente. De fato, observa-se que parcela significativa dos estudos de cunho fundamentalista adotaram periodicidade trimestral para as previsões face a esse entrave, dado que se trata de uma variável consagrada na literatura e utilizada pelos principais estudos seminais da área, tais como Meese e Rogoff (1983), Wolff (1987) e Mark (1995). Porém, dados os elevados número de transações e volume de cifras movimentadas no mercado FOREX, frequentemente utiliza-se o índice de produção industrial como variável *proxy* ao PIB e ao nível de renda de uma economia.

Mitchell et al. (2005) propuseram um estimador que faça a interpolação do PIB para frequência mensal; para tal, os autores levantaram 5 dimensões de indicadores para fazer a estimação: produção industrial, agricultura, construção, serviços públicos e serviços privados. No entanto, os autores constataram que apenas o indicador de produção industrial estava consistentemente disponível na frequência mensal², de modo que, para se obter os indicadores das outras 4 dimensões, recorreu-se a um novo processo de interpolação, tornando a mineração do dado bastante tediosa, além de prejudicar a inteligibilidade econômica da construção do indicador interpolado. Ademais, como o estudo de Mitchell et al. (2005) focou especificamente no mercado norte-americano, a maioria dos indicadores das outras 4 dimensões também não estavam disponíveis para as seis economias consideradas no presente estudo (Estados Unidos, Reino Unido, Zona do Euro, China, Japão e Brasil). Sendo assim, optou-se por utilizar apenas o índice de produção industrial como *proxy* mensal para o PIB, visto que, além de ser disponibilizada mensalmente para as seis economias estudadas, vários estudos apontam que esta variável é uma aproximação razoável para o PIB; mais especificamente no contexto de previsão de taxas de câmbio, o índice de produção industrial foi utilizado em diversos estudos recentes, (DICK; MACDONALD; MENKHOFF, 2015; BERG; MARK, 2015; DONAYRE; PANOVSKA, 2016), embasando o seu uso no presente estudo como substituto ao PIB trimestral.

²Para as dimensões da agricultura e dos serviços públicos, por exemplo, não havia nenhum indicador disponível.

3.1.6 Balança comercial

A balança comercial reflete a posição de um país em relação às suas transações econômicas internacionais, informando se um país é exportador ou importador líquido. Caso a balança comercial de um país **A** sofra uma variação superavitária superior relativo a um país **B**, implica que **A** está exportando mais que **B**, fazendo com que maior volume de divisas internacionais esteja entrando na economia de **A**, e assim fazendo com que a moeda emitida por **A** se valorize em relação à moeda emitida por **B**. Em contrapartida, um déficit relativo da balança comercial de **A** implica que **A** está pegando moeda estrangeira emprestada para cobrir o déficit, de modo que esse país demanda quantidade maior de divisas internacionais do que está recebendo pelas transações comerciais, e implicitamente “oferta” mais moeda doméstica do que os outros países “demandam” suas exportações, o que faz com que a moeda de **A** se desvalorize.

Estudos como Schinasi e Swamy (1989) e Meese e Rose (1991) fizeram uso dessa variável para modelos de previsão fundamentalistas, atestando para a relevância de sua inclusão para a elevação do poder preditivo em relação ao *Random Walk*. Como alternativa à balança comercial, Wright (2008) utilizou a conta de transações correntes do balanço de pagamentos como variável explicativa; porém o presente estudo desconsiderou essa última variável, dado que possui definição conceitual similar à balança comercial, além de figurar em poucos estudos correlatos e ter demonstrado significância estatística tímida em relação a estudos que utilizaram a balança comercial.

3.1.7 Variação do índice da bolsa de valores

Os índices de bolsas de valores servem como um indicativo do porte do mercado financeiro de um país, além de fornecer um reflexo geral de sua solidez macroeconômica. A uma mesma taxa de retorno, um investidor internacional tende a depositar seu dinheiro em mercados considerados “seguros”, preterindo mercados financeiros desordenados e com má reputação. Dentro do contexto da taxa de câmbio, a saúde do mercado financeiro de um país é um fator determinante para o nível de investimento externo que este irá receber; a depender do quão atrativo é esse mercado para investidores estrangeiros, a taxa de câmbio do país irá sofrer alterações: caso o país seja considerado um investimento desejável, maior volume de investimentos tende a fluir em direção ao seu mercado, fazendo com que a moeda local se valorize. Foram consideradas seis bolsas de valores, uma para cada moeda estudada (dólar, euro, libra, iene, renminbi e real). O critério para a escolha das bolsas de valores consideradas para o presente estudo foi o *market cap*: *New York Stock Exchange* e *NASDAQ* para o mercado norte-americano (*Stan-*

ard & Poor's 500 Index), *London Stock Exchange* para o mercado britânico (*FTSE 100 Index*), *Euronext NV* para a zona do Euro (*EURO STOXX 50*), *Shanghai Stock Exchange* para o mercado chinês (*SSE Composite Index*), *Tokyo Stock Exchange* para o mercado japonês (*Nikkei 225 Index*), e *BM&F Bovespa* para o mercado brasileiro (*Ibovespa*).

3.1.8 Déficit público

O déficit público diz respeito ao montante gasto pelo governo em relação às suas receitas tributárias; dessa forma, o gasto governamental exerce uma influência direta sobre a quantidade absoluta de moeda doméstica em circulação, o que afeta na taxa de câmbio da moeda do país, dado que, com maior volume de moeda doméstica em circulação em relação a moedas estrangeiras, a moeda doméstica tende a desvalorizar-se. Adicionalmente, gastos maiores do governo tendem a tornar o país menos atrativo a investimentos estrangeiros a curto prazo, pois um grande déficit público alimenta o crescimento da inflação, aumentando a pressão de desvalorização sobre a moeda doméstica. Além desses efeito mais “imediate”, Bouakez e Eyquem (2015) aponta que o gasto governamental exerce também um componente “indireto”, na medida em que um aumento inesperado nos gastos do governo (*i.e.*: um aumento que contrarie as expectativas do mercado) pode fazer com que o diferencial da taxa de juros a longo prazo em relação ao resto do mundo decresça, fazendo com que a moeda doméstica se desvalorize ainda mais.

3.1.9 Taxa de inflação esperada

Assim como a taxa de inflação observada, a expectativa da taxa de inflação também pode exercer influências na economia de um país, e conseqüentemente na taxa de câmbio de sua moeda. A literatura (CLARIDA; WALDMAN, 2008; CONRAD; LAMLA, 2010) aponta para um “efeito de sinalização” entre a inflação esperada e o câmbio, de modo que quando há uma “surpresa positiva” na inflação observada (*i.e.*: quando a inflação observada é superior à predita), espera-se que a moeda doméstica irá se valorizar (analogamente, quando a inflação observada é inferior à predita, a moeda doméstica se desvaloriza), um resultado aparentemente contraditório, mas que se explica pelo fato de uma “surpresa positiva” acarretar um aumento na taxa de juros de curto prazo³. Ademais, Arghyrou e Pourpourides (2016) apontam que esse efeito de sinalização pode ser assimétrico – ou seja, os efeitos de uma desvalorização da moeda oriunda de uma “surpresa negativa” possui magnitude e significância estatística maiores que os efeitos de

³Para uma discussão mais aprofundada sobre esse efeito, ver Neely (2010).

uma valorização da moeda oriunda de uma “surpresa positiva”, o que sugere a existência de interações não lineares entre a inflação esperada e o câmbio que não seriam eficientemente capturadas por modelos explicativos lineares, indo ao encontro da lacuna descrita pelo estado da arte do tema.

3.1.10 Movimento de capitais

A conta movimento de capitais do balanço de pagamentos reflete essencialmente o montante líquido de divisas internacionais que entraram na economia de um país. Notadamente numa era de internacionalização do capital e de crescente integração dos mercados financeiros, as taxas de câmbio se tornam especialmente sensíveis aos fluxos internacionais de capital (GELMAN et al., 2015). Caso o país emissor da moeda **A** apresente um aumento relativo na conta movimento de capitais superior ao país emissor da moeda **B**, há mais entrada de investimentos estrangeiros no país da moeda **A** do que saída de investimentos para outros países, o que faz com que a moeda **A** se valorize em relação a **B**.

3.1.11 Volume de reservas internacionais

A conta de reservas internacionais do balanço de pagamentos contabiliza o montante de ativos financeiros líquidos mantidos em moedas estrangeiras e em Direitos Especiais de Saque (SDRs), montante que pode ser utilizado para financiar déficits ou para influenciar a política monetária. No contexto de câmbios, a autoridade monetária de um país pode estocar moedas estrangeiras para fazer a moeda doméstica se desvalorizar, assim como pode utilizar as reservas para comprar ativos domésticos e valorizar o câmbio da moeda nacional. Os SDRs do FMI podem ser usados para finalidade similar, dado que podem ser convertidos em moedas *de facto*. Caso a variação do estoque de reservas internacionais de um país **A** seja superior em termos relativos à variação de um país **B**, implica que o país **A** está segurando mais ativos em moedas estrangeiras do que os outros países estão segurando ativos na moeda do país **A**; assim, proporcionalmente, a quantidade de ativos em moedas estrangeiras que saiu de circulação é maior que a quantidade de ativos na moeda do país **A**, de modo que a maior disponibilidade relativa da moeda do país **A** no mercado fará com que ela se desvalorize.

3.1.12 Preço de *commodities*

Estudos que tratam da “doença holandesa”⁴ – em especial focados em mercados emergentes e em desenvolvimento (BODART; CANDELON; CARPANTIER, 2015; BASHER; HAUG; SADORSKY, 2016) – apontam que, para países com forte dependência da exportação de *commodities*, o preço destas possui uma importância decisiva na determinação de sua taxa de câmbio em relação a moedas estrangeiras. Dado que as transações econômicas no meio internacional são em maior parte realizadas em torno de divisas consideradas “confiáveis” a nível mundial (ou pelo menos regional, no caso de transações dentro de blocos de integração regional), o preço das *commodities* irá impactar diretamente no volume de divisas estrangeiras auferidas com as exportações, o que por sua vez trará um reflexo para a taxa de câmbio. Assim, caso o preço de uma *commodity* se eleve, países importadores líquidos daquele recurso irão transferir mais de sua moeda para o exportador líquido, de modo que a moeda do exportador líquido tende a se valorizar. Sob o mesmo cenário e entre dois importadores líquidos da *commodity*, a moeda do país que importa em maiores quantidades tende a desvalorizar-se, pois maiores quantidades da moeda do maior comprador irá ser colocada em circulação no mercado mundial.

O artigo de Bodart, Candelon e Carpentier (2015), focando em economias emergentes, considerou *commodities* “específicas” para cada país analisado, a depender se suas pautas de exportação. Dado que o presente estudo também abarcou economias ditas “centrais”, as quais são importadoras líquidas de *commodities*, julgou-se que essa especificação não se fazia necessária, de modo que considerou-se duas *commodities* “globais”: petróleo e ouro – o primeiro por ser a principal matriz energética da economia mundial, e o segundo por ser considerado o ativo com maior reserva de valor. Para ambos, qualquer flutuação no seu nível de preços tende a impactar todos os agentes da economia global, especialmente o preço do petróleo, que historicamente promoveu impactos sistêmicos (com os choques de 1973 e 1979) e continua proporcionando “efeito contágio” em magnitude considerável (BRAHMASRENE; HUANG; SISSOKO, 2014; BOUOYOUR et al., 2015; FERRARO; ROGOFF; ROSSI, 2015; BASHER; HAUG; SADORSKY, 2016).

⁴Denomina-se “doença holandesa” (“*dutch disease*”) o fenômeno no qual países ricos em recursos naturais, possuindo um incentivo em depender prioritariamente da exportação dessas *commodities*, acabam apresentando uma economia pouco diversificada, o que, além de aumentar sua vulnerabilidade frente ao mercado internacional da referida *commodity*, solapa a força que esses países delegam para aumentar o nível de complexidade e sofisticação de suas economias. Para discussões mais aprofundadas, ver Corden (1984) e Krugman (1987).

3.2 Variáveis expurgadas

Além das 13 variáveis explicativas listadas na tabela 3.1, foram observadas na literatura científica outras variáveis propostas como preditores das taxas de câmbio, mas que foram expurgadas para o presente estudo, por razões elucidadas abaixo:

- (i) Produto Interno Bruto (PIB): Conforme discutido anteriormente, o PIB é uma variável macroeconômica que, apesar de ser apontada como relevante por diversos autores, é disponibilizada trimestralmente. Dado que a frequência mensal considerada no presente estudo já constitui em uma notável limitação – dada a natureza de transações de alta frequência no mercado FOREX –, a inclusão de uma variável trimestral forçaria a adoção dessa periodicidade em todos os outros preditores. Apesar de a frequência trimestral também ser amplamente analisada em estudos recentes, considerou-se que a escolha dessa frequência menor empobreceria o poder explicativo do modelo. Métodos de interpolação também foram descartados pois envolveriam outros indicadores também indisponíveis. Ademais, aliado ao fato de a literatura ter apontado uma *proxy* satisfatória, optou-se por desconsiderar o PIB como variável explicativa.
- (ii) Hiato do produto: O hiato do produto é definido como a diferença entre o PIB real e o PIB potencial e serve como uma medida da capacidade ociosa de uma economia. Sua definição conceitual e limitações para o escopo do presente estudo são as mesmas do PIB; por conseguinte, optou-se por excluir essa variável da análise pelos mesmos motivos.
- (iii) Taxa de câmbio a termo (*forward exchange rate*): O câmbio a termo é a taxa de câmbio, definida no presente, para uma operação a ser concretizada em data futura. Essa variável foi considerada por estudos como Clarida e Taylor (1997) e Clarida et al. (2003), porém diversos autores (CUMBY; OBSTFELD, 1984; FAMA, 1984; FRANKEL; POONAWALA, 2010) apontam que a taxa de câmbio a termo é um preditor ruim para o câmbio *spot*, oferecendo poder explicativo pífio, mesmo para a previsão da variação direcionada do câmbio. Adicionalmente, o câmbio a termo apresenta sensibilidade sensível à escolha do período (se a taxa a termo é de 1, 3, 6 ou 12 meses) e a períodos de alta volatilidade (SIMPSON; GROSSMANN, 2014). Portanto, optou-se por descartar o uso dessa variável como preditor.
- (iv) Conta de transações correntes do balanço de pagamentos: Conforme abordado em seção anterior, a incorporação de transações correntes como variável explicativa não apresentou significância preditiva sólida, além de ter sido utilizada em reduzido número de publicações (foi identificado apenas o estudo de Wright (2008)), justificando seu expurgo no presente estudo.

- (v) Derivativos: O presente estudo desconsiderou qualquer contrato de derivativos cambiais, dado que não foram encontrados estudos especialmente focados na previsão da taxa de câmbio que tenha utilizado essa variável como preditor. Em vez disso, contratos de derivativos cambiais têm sido utilizados pela literatura científica para analisar a exposição do risco cambial por firmas e o estudo de estratégias corporativas para o *hedge* desse risco (NGUYEN; FAFF; MARSHALL, 2007; YIP; NGUYEN, 2012; GATOPOULOS; LOUBERGÉ, 2013). Como o presente estudo considerou a taxa de câmbio como sendo exógena para indivíduos e firmas, o escopo do trabalho não reside em analisar como cada empresa irá reagir a mudanças no mercado cambial. Ademais, há uma grande variedade de derivativos cambiais (contratos a termo, futuros, opções, swaps), frequentemente com datas de maturação variados, o que tornou inviável a escolha de uma categoria uniforme para todos os 15 pares de câmbio analisados (por exemplo, não foram encontrados dados referentes a contratos de swap cambial entre libra esterlina e renminbi chinês); face a esse fato, grande parte dos estudos que abarcaram derivativos cambiais fizeram uso de uma variável *dummy* discriminando empresas que fizeram uso de derivativos cambiais (*forward*, futuro, opção ou swap) daquelas que não fizeram, uma abordagem que não seria aderente com as propostas do presente estudo. Dessa forma, optou-se por desconsiderar derivativos cambiais como variável explicativa.
- (vi) Commodities específicas: Conforme abordado anteriormente, o estudo de Bodart, Candelon e Carpentier (2015) realizou um estudo com dados em painel relativos a 33 países em desenvolvimento, e a *commodity* definida para cada país foi variante (*e.g.*: A *commodity* considerada para Argélia e Venezuela foi o petróleo, mas para o Chile foi o cobre, para o Paquistão foi o algodão, para o Níger foi o urânio, etc.). O presente estudo, tendo analisado também câmbios atrelados a economias ditas “centrais”, optou por descartar a definição específica para essa variável, se atendo a *commodities* que menos atrelados às peculiaridades de cada país analisado: petróleo e ouro.

Variável levantada	Principais referências
Diferencial da taxa de câmbio defasada em um período	Diebold, Gardeazabal e Yilmaz (1994), Engel (1994), Rapach e Wohar (2006), Chung e Hong (2007), Carriero, Kapetanios e Marcellino (2009), Donayre e Panovska (2016)
Diferencial da taxa de juros	Meese e Rogoff (1983), Chung e Hong (2007), Nootyaskool e Choengtong (2014), Fratzscher et al. (2015), Redl (2015), Beckmann e Schüssler (2016)
Diferencial da base monetária	Frenkel (1976), Meese e Rogoff (1983), Mark (1995), Balke, Ma e Wohar (2013), Djeutem e Kasa (2013), Dąbrowski, Papież e Śmiech (2014), Katusiime, Shamsuddin e Agbola (2015)
Diferencial da taxa de inflação	Meese e Rogoff (1983), Eng et al. (2008), Nootyaskool e Choengtong (2014), Fratzscher et al. (2015), Berg e Mark (2015), Ca'Zorzi, Kocięcki e Rubaszek (2015), Byrne, Korobilis e Ribeiro (2016), Donayre e Panovska (2016)
Diferencial do índice de produção industrial	Mitchell et al. (2005), Dick, MacDonald e Menkhoff (2015), Berg e Mark (2015), Donayre e Panovska (2016)
Diferencial da balança comercial	Kutan e Vukšić (2007), Eng et al. (2008), AbuHamad, Mohd e Salim (2013), Bown e Crowley (2013), Fratzscher et al. (2015), Chen e MacDonald (2015), Choudhri e Hakura (2015)
Diferencial da variação do índice da principal bolsa de valores	Hau e Rey (2006), Yuan (2013), Hassan, Mertens e Zhang (2016)
Diferencial do déficit público	Bouakez e Eyquem (2015), Bouoiyour et al. (2015), Chen e MacDonald (2015)
Diferencial da taxa de inflação esperada	Faust, Rogers e Wright (2003), Clarida e Waldman (2008), Conrad e Lamla (2010), Bacchetta e Wincoop (2013), Arghyrou e Pourpourides (2016)
Diferencial do movimento de capitais	Kutan e Vukšić (2007), Gelman et al. (2015), Katusiime, Shamsuddin e Agbola (2015)
Diferencial do volume de reservas internacionais	Klein e Ku (2006), Rogoff e Stavrakeva (2008), Nasirtoussi, Wah e Ling (2011), Corte, Sarno e Sestieri (2012), Katusiime, Shamsuddin e Agbola (2015)
Preço do petróleo	Brahmasrene, Huang e Sissoko (2014), Ferraro, Rogoff e Rossi (2015), Bouoiyour et al. (2015), Jammazi, Lahiani e Nguyen (2015), Basher, Haug e Sadorsky (2016), Zhang, Dufour e Galbraith (2016)
Preço do ouro	Bodart, Candelon e Carpentier (2015), Ferraro, Rogoff e Rossi (2015), Zhang, Dufour e Galbraith (2016)

Tabela 3.1: Variáveis independentes do modelo fundamentalista

4 METODOLOGIA

4.1 *Support Vector Machine*

4.1.1 SVM para classificação linear

O *Support Vector Machine* – SVM (BOSER; GUYON; VAPNIK, 1992; CORTES; VAPNIK, 1995) é um algoritmo de aprendizagem supervisionada que fornece uma função de decisão de classificação que discrimina os dados observados em duas classes complementares, $+1$ e -1 , de modo a maximizar a margem entre as classes. Basicamente, dadas n observações com p variáveis explicativas e a respectiva classe à qual a n -ésima observação pertence, o SVM é um método que permite computar a expressão de uma curva em \mathbb{R}^p que separa as duas classes, permitindo realizar inferências em relação à $\{n+1\}$ -ésima observação das mesmas p variáveis explicativas, predizendo a classe à qual essa nova observação irá pertencer.

Os insumos necessários para a formulação do problema são:

- (i) Um vetor de variáveis explicativas $\mathbf{x}_{(p \times 1)}$;
- (ii) Uma matriz de observações $\mathbf{A}_{(n \times p)}$ com n observações das p variáveis explicativas;
- (iii) Um vetor de variáveis dependentes $\mathbf{y}_{(n \times 1)}$, informando a classe à qual cada indivíduo observado pertence, tal que para todo $y_i \in \mathbf{y}, i = 1, 2, \dots, n, y_i \in \{-1, +1\}$.

Considerando *a priori* o caso mais simples, no qual os dados $\mathbf{x} \in \mathbb{R}^p$ são separáveis em relação às duas classes da variável dependente y por uma função linear em \mathbb{R}^p : $\sum_{i=1}^p w_i \cdot x_i - w_0 = 0$. O objetivo consiste em encontrar o hiperplano $\sum_{i=1}^p w_i \cdot x_i - w_0 = \mathbf{w} \cdot \mathbf{x} - w_0 = 0$ que maximiza a margem entre as classes, condicionado às observações tomadas. Assume-se, sem perda de generalidade, que as observações pertencentes à classe $+1$ estão acima do hiperplano $\mathbf{w} \cdot \mathbf{x} - w_0 = +1$, enquanto que as observações da classe -1 estão abaixo de

$w \cdot x - w_0 = -1$. Os pontos observados que tangenciam os hiperplanos de separação recebem a denominação de “*support vectors*” (“vetores de suporte”).

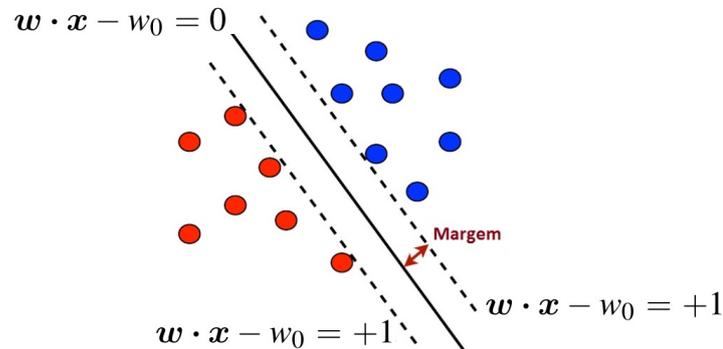


Figura 4.1: Representação do problema do SVM em dados linearmente separáveis.

Fonte: Adaptado de Mohri, Rostamizadeh e Talwalkar (2012, p. 65).

Conforme ilustra a figura 4.1, os pontos azuis e vermelhos representam as classes dicotômicas; deseja-se obter um hiperplano que separe as classes de modo a maximizar a menor distância geométrica possível entre pontos de diferentes classes, de modo a obter hiperplanos que “separem o máximo possível” as duas classes consideradas, representados pelas retas tracejadas na figura. A função de decisão para a predição da $\{n + 1\}$ -ésima observação será o hiperplano representado pela reta cheia na figura, paralelo a $w \cdot x - w_0 = +1$ e $w \cdot x - w_0 = -1$, tal que a sua distância para ambas as classes é igual¹. Assim, a função de decisão será expressa na forma:

$$f(x) = \text{sgn}(w^T x - w_0) = \begin{cases} +1, & w^T x - w_0 > 0, \\ -1, & w^T x - w_0 < 0. \end{cases}$$

com $w, x \in \mathbb{R}^p$

Dessa forma, a formulação do SVM se resume a um problema de otimização condicionada²; num caso genérico em que as n observações estão ordenadas pelas duas classes, de modo que primeiras $n_{\#} - 1$ observações pertencem à primeira classe e o restante (da $n_{\#}$ -ésima até a n -ésima observação) pertencem à classe complementar, a função objetivo e as restrições são

¹Tomando um segmento de reta ortogonal a $w \cdot x - w_0 = +1$ e a $w \cdot x - w_0 = -1$ que conecta os dois hiperplanos, o hiperplano de decisão deve interceptar o ponto médio deste segmento.

²Mais especificamente, o SVM é um problema de programação quadrática

definidas como a seguir:

$$\begin{aligned}
 \text{Maximizar : } & \frac{2}{\|\mathbf{w}\|^2} \\
 \text{Sujeito a : } & -w_0 + w_1 \cdot a_{1,1} + w_2 \cdot a_{1,2} + \dots + w_p \cdot a_{1,p} \leq -1 \\
 & \vdots \\
 & -w_0 + w_1 \cdot a_{n\#-1,1} + w_2 \cdot a_{n\#-1,2} + \dots + w_p \cdot a_{n\#-1,p} \leq -1 \\
 & -w_0 + w_1 \cdot a_{n\#,1} + w_2 \cdot a_{n\#,2} + \dots + w_p \cdot a_{n\#,p} \geq +1 \\
 & \vdots \\
 & -w_0 + w_1 \cdot a_{n,1} + w_2 \cdot a_{n,2} + \dots + w_p \cdot a_{n,p} \geq +1 \\
 \text{Tal que : } & w_0, w_1, w_2, \dots, w_p \in \mathbb{R}
 \end{aligned}$$

Onde $\|\mathbf{w}\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$ é a norma euclidiana.

A expressão da função objetivo deriva da expressão genérica da distância geométrica entre dois hiperplanos paralelos: Dado que a distância entre $\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n = c_1$ e $\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n = c_2$ pode ser calculada como $\frac{|c_1 - c_2|}{\sqrt{\alpha_1^2 + \dots + \alpha_n^2}}$, é possível escrever a distância entre $\mathbf{w} \cdot \mathbf{x} - w_0 = +1$ e $\mathbf{w} \cdot \mathbf{x} - w_0 = -1$ como $\frac{2}{\|\mathbf{w}\|}$ mediante manipulações algébricas básicas.

Constata-se que maximizar $\frac{2}{\|\mathbf{w}\|^2}$ equivale a minimizar seu recíproco $\frac{\|\mathbf{w}\|}{2}$; ademais, como $\|\mathbf{w}\|$ é um valor não-negativo, minimizar $\frac{\|\mathbf{w}\|}{2}$ equivale a minimizar $\frac{\|\mathbf{w}\|^2}{2} = \frac{\mathbf{w}^T \mathbf{w}}{2}$. Por conveniência, a função objetivo pode ser alterada de acordo sem inibições. Para efeitos de simplificação, também é adequado expressar o problema formulado em notação matricial, apresentada a seguir:

$$\begin{aligned}
 \text{Minimizar : } & \frac{1}{2} \mathbf{w}^T \mathbf{w} \\
 \text{Sujeito a : } & \mathbf{D}(\mathbf{A}\mathbf{w} - w_0 \mathbf{1}) \geq \mathbf{1} \\
 \text{Tal que : } & w_0 \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^p
 \end{aligned}$$

onde \mathbf{w} é o vetor coluna $p \times 1$ referente aos coeficientes angulares do hiperplano de decisão; w_0 é um escalar real relativo ao intercepto (coeficiente linear) do hiperplano de decisão; \mathbf{D} é a matriz diagonal $n \times n$ com os valores de \mathbf{y} ; e $\mathbf{1}$ é o vetor coluna $n \times 1$ com todos os valores iguais a 1.

4.1.2 SVM para classificação não linear

Para o caso em que as observações em \mathbb{R}^p não podem ser separadas por uma função linear, realiza-se uma transformação $\varphi(x) \in \mathbb{R}^q$ nas variáveis explicativas para uma dimensão superior³ na qual os dados se tornam linearmente separáveis. O espaço original \mathbb{R}^p é denominado “espaço de entrada” (*input space*) e o espaço induzido por φ , \mathbb{R}^q , é denominado “espaço de característica” (*feature space*). Assim, a matriz de observações $\mathbf{A}_{(n \times p)}$ será mapeada em uma dimensão superior como $\Phi_{(n \times q)}$. No espaço de característica \mathbb{R}^q , procede-se da mesma forma para o SVM do caso com dados linearmente separáveis: constroem-se dois hiperplanos que separem as duas classes, dos quais se deriva o hiperplano de decisão para realizar previsões para observações novas. Ao retornar para o espaço de entrada, a curva que separa as classes poderá ser uma função não linear, dado que as curvas de nível da secção transversal do hiperplano em \mathbb{R}^q com a transformação φ , projetadas na dimensão inferior \mathbb{R}^p , podem assumir formas não lineares. A figura 4.2 sintetiza essa ideia, ilustrando a redução do problema de classificação não-linear para um problema de classificação linear mediante a aplicação da transformação φ .

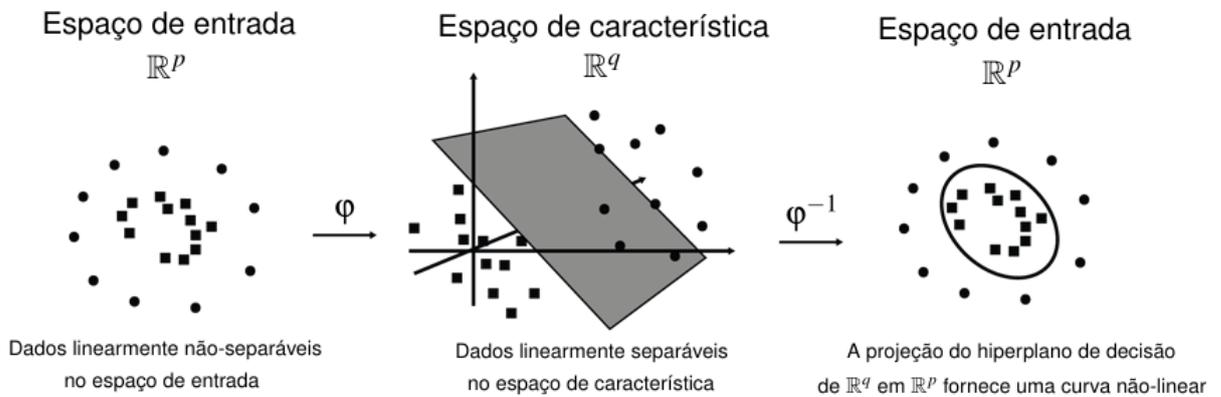


Figura 4.2: SVM para classificação não-linear.

Fonte: Modificado a partir de Soman, Loganathan e Ajay (2011, p. 147).

Dessa forma, realizando-se a transformação das observações originais:

$$\begin{aligned} \mathbf{x} &\rightarrow \varphi(\mathbf{x}) \\ \mathbb{R}^p &\rightarrow \mathbb{R}^q \end{aligned}$$

³Em geral, q é um número muito superior a p .

o SVM para separação não linear é dado por:

$$\begin{aligned} \text{Minimizar :} & \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} \\ \text{Sujeito a :} & \quad \mathbf{D}(\Phi \mathbf{w} - w_0 \mathbf{1}) \geq \mathbf{1} \\ \text{Tal que :} & \quad w_0 \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^q \end{aligned}$$

A forma da função de decisão será uma função linear no espaço de característica \mathbb{R}^q e possui estrutura análoga ao caso linearmente separável, a saber:

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) - w_0) = \begin{cases} +1, & \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) - w_0 > 0, \\ -1, & \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) - w_0 < 0. \end{cases}$$

com $\mathbf{w}, \boldsymbol{\varphi}(\mathbf{x}) \in \mathbb{R}^q$

4.1.3 Forma dual do SVM

Dado que o SVM não linear é um problema de programação quadrática, sua forma dual pode ser obtida facilmente resolvendo o dual de Wolfe (WOLFE, 1961), dado por

$$\max_{\boldsymbol{\lambda} \geq 0} \left[\min_{\mathbf{w}, w_0} \mathcal{L}(\mathbf{w}, w_0, \boldsymbol{\lambda}) \right]$$

onde

$$\mathcal{L}(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \boldsymbol{\lambda}^T [\mathbf{D}(\Phi \mathbf{w} - \gamma \mathbf{1}) - \mathbf{1}]$$

é a função Lagrangeana do SVM não linear, donde $\boldsymbol{\lambda}$ é o vetor coluna $n \times 1$ dos multiplicadores de Lagrange.

Resolvendo o Lagrangeano para \mathbf{w} e w_0 para reescrevê-lo em função apenas de $\boldsymbol{\lambda}$, obtêm-se as seguintes condições de primeira ordem:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}} \mathcal{L}(\mathbf{w}, w_0) &= \mathbf{w}^T - \boldsymbol{\lambda}^T \mathbf{D}\Phi = \mathbf{0} \Rightarrow \mathbf{w}^T = \boldsymbol{\lambda}^T \mathbf{D}\Phi \\ \frac{\partial}{\partial w_0} \mathcal{L}(\mathbf{w}, w_0) &= \boldsymbol{\lambda}^T \mathbf{D}\mathbf{1} = 0 \end{aligned}$$

Dado que $\boldsymbol{\lambda}^T \mathbf{D}\mathbf{1}$ resulta num escalar, a transposta desse termo será ele próprio ($\boldsymbol{\lambda}^T \mathbf{D}\mathbf{1} =$

$\lambda^T D \mathbf{1}$). Após manipulações básicas, o dual de Wolfe do SVM não linear é dado por:

$$\begin{aligned} \text{Maximizar :} & \quad -\frac{1}{2} \lambda^T D \cdot \Phi \Phi^T \cdot D^T \lambda + \lambda^T \mathbf{1} \\ \text{Sujeito a :} & \quad \mathbf{1}^T D \lambda = 0 \\ \text{Tal que :} & \quad \lambda \geq 0 \end{aligned}$$

O uso de φ garante que os dados serão linearmente separáveis em uma dimensão superior, porém a manipulação de $\varphi(\mathbf{x})$ em \mathbb{R}^q pode ser bastante onerosa, ou mesmo inviável, dado que $\varphi(\mathbf{x})$ pode ser um vetor em uma dimensão infinita. Em vez disso, recorre-se ao uso da **função Kernel**, a qual permitirá contornar o problema da alta dimensionalidade induzida por $\varphi(\mathbf{x})$, sem a necessidade de se computar explicitamente sua forma funcional. Observa-se que o termo $\Phi_{(n \times q)} \Phi_{(q \times n)}^T$ presente na forma dual do SVM resulta numa matriz $n \times n$ cujos elementos são os produtos internos das transformações φ , cada um dos quais fornecerá um escalar real. Dessa forma, a partir da transformação $\varphi(\mathbf{x})$, define-se a função *Kernel*⁴

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \varphi^T(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) \in \mathbb{R}, \quad i, j = 1, 2, 3, \dots, n$$

e a matriz gramiana \mathbf{K} :

$$\mathbf{K} = \Phi \Phi^T = \begin{pmatrix} \varphi(x_1)^T \varphi(x_1) & \varphi(x_1)^T \varphi(x_2) & \cdots & \varphi(x_1)^T \varphi(x_n) \\ \varphi(x_2)^T \varphi(x_1) & \varphi(x_2)^T \varphi(x_2) & \cdots & \varphi(x_2)^T \varphi(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi(x_n)^T \varphi(x_1) & \varphi(x_n)^T \varphi(x_2) & \cdots & \varphi(x_n)^T \varphi(x_n) \end{pmatrix}_{(n \times n)}$$

Ao agrupar o produto interno $\varphi^T(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$, nota-se que a ordem da matriz \mathbf{K} depende apenas das observações tomadas, e não mais da dimensionalidade do espaço de característica \mathbb{R}^q , a qual pode ser infinita, fato que reduz consideravelmente o custo computacional para obter a função de decisão ótima, a qual também pode ser escrita em termos da função *Kernel* em vez

⁴Dado que o produto interno é uma medida de similaridade em espaços de Hilbert, a função *Kernel* pode ser vista como uma forma de mensurar a “margem” entre as classes em espaços de dimensão elevada, inclusive infinita.

de φ mediante a condição $\mathbf{w}^T = \boldsymbol{\lambda}^T \mathbf{D}\Phi$ obtida anteriormente, donde obtém-se que⁵:

$$\begin{aligned} f(\mathbf{x}_i) &= \text{sgn}(\mathbf{w}^T \varphi(\mathbf{x}_i) - w_0) \\ &= \text{sgn}((\boldsymbol{\lambda}^T \mathbf{D}\Phi)\varphi(\mathbf{x}_i) - w_0) \\ &= \text{sgn}(\varphi(\mathbf{x}_i)^T (\Phi^T \mathbf{D}\boldsymbol{\lambda}) - w_0) \\ &= \text{sgn}\left(\sum_{j=1}^n \kappa(\mathbf{x}_i, \mathbf{x}_j) y_j \lambda_j - w_0\right) \quad \blacksquare \end{aligned}$$

Outra vantagem que a função *Kernel* propicia é a praticidade de se trocar a transformação $\varphi(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}^q$ por uma outra transformação $\varphi'(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}^{q'}$, basta trocar a matriz *Kernel* anterior $K = (\text{induzida por } \varphi(\mathbf{x}))$ pela matriz *Kernel* $K' = \Phi'(\Phi')^T$ induzida por $\varphi'(\mathbf{x})$, de modo que os elementos de K' representam os produtos internos $\varphi'^T(\mathbf{x}_i) \cdot \varphi'(\mathbf{x}_j)$. É fácil verificar que o restante da formulação do problema permanece constante. A possibilidade de se modificar o algoritmo de aprendizado apenas substituindo a matriz de produtos internos $K = \Phi\Phi^T$ é denominada **Kernel Trick**. Dessa forma, diferentes estruturas de não linearidade podem ser testadas apenas modificando a forma funcional da função *Kernel*; na verdade, qualquer função κ que satisfaz a condição de admissibilidade de Mercer (1909) – a ser abordada posteriormente – pode ser aplicada no algoritmo de aprendizagem.

4.2 Support Vector Regression

O *Support Vector Regression*–SVR (VAPNIK, 1995; DRUCKER et al., 1997) é a extensão do SVM cuja função de decisão, em vez de fornecer a classe à qual uma nova observação pertence, irá retornar uma estimativa numérica para a variável dependente. Sua construção matemática é análoga a um SVM para infinitas classes, em vez de apenas duas; dessa forma, a ideia básica é encontrar a função que melhor aproxima um conjunto de observações e que propicie a inferência acerca de uma nova observação tomada. Assim, o valor y não mais se limitará ao conjunto $\{-1, +1\}$, e poderá assumir qualquer valor real.

Em modelos de regressão, busca-se uma função de decisão “regularizada” – isto é, que tanto se adeque bem aos dados observados quanto possua boa capacidade de generalização. Associadas a essas duas características desejáveis estão associados dois problemas correspondentes, os quais constituem o chamado **dilema viés-variância**:

- (i) *Overfitting*: Caso se obrigue que a função de decisão passe por todas as observações

⁵Lembrando que $\mathbf{w}^T \varphi(\mathbf{x})$ é um escalar e \mathbf{D} é uma matriz simétrica

tomadas, o comportamento dessa função pode ser demasiadamente volátil, de modo que a capacidade de generalização seja prejudicada, o que afeta negativamente na predição de uma nova observação;

- (ii) *Underfitting*: Por outro lado, caso se opte por uma estrutura funcional mais simples (por exemplo, uma regressão linear), apesar de se simplificar o comportamento da função de decisão, os desvios dessa função em relação às observações tomadas tende a ser maior, prejudicando a aderência do modelo à amostra e conseqüentemente a validade das predições realizadas.

O dilema viés-variância preconiza que, em geral, há um *trade-off* entre a precisão das estimativas (“viés”) e a amplitude de seu intervalo de confiança (“variância”), de modo que é desejável uma função de decisão que contemple um meio-termo entre o *underfitting* e o *overfitting*. Para tal, a formulação do SVR considera dois parâmetros adicionais em relação ao SVM, relativos ao viés e à variância:

- (i) A fim de evitar *overfitting*, introduz-se uma banda de tolerância ε para o erro entre os valores predito e o observado, de modo que valores observados de y que estão a uma distância menor ou igual a ε da função de decisão (*i.e.*, os pontos observados $y_i, i = 1, 2, \dots, n$ tal que $|y_i - f(\mathbf{x}_i)| \leq \varepsilon$) são considerados “suficientemente próximos” para serem considerados como estatisticamente iguais a y ;
- (ii) Por outro lado, para minimizar o número de pontos fora do intervalo de confiança $[f(\mathbf{x}_i) - \varepsilon, f(\mathbf{x}_i) + \varepsilon]$ atribui-se uma penalização C na função objetivo para os pontos que ultrapassam esse intervalo de confiança em uma quantidade $\xi > 0$. Sem perda de generalidade, denota-se por ξ^* a magnitude de pontos enviesados positivamente ($y_i < f(\mathbf{x}_i) - \varepsilon$) e por ξ a magnitude de pontos enviesados negativamente ($y_i > f(\mathbf{x}_i) + \varepsilon$).

A partir da implicação dos parâmetros adicionais acima, define-se a função de perda ε -insensitiva (VAPNIK, 1995) $L_\varepsilon[y_i, f(\mathbf{x}_i)]$ ⁶:

$$L_\varepsilon[y_i, f(\mathbf{x}_i)] = \begin{cases} |y_i - f(\mathbf{x}_i)| - \varepsilon, & |y_i - f(\mathbf{x}_i)| > \varepsilon, \\ 0, & |y_i - f(\mathbf{x}_i)| \leq \varepsilon. \end{cases}$$

Vale salientar que a função de perda ε -insensitiva não é a única forma de determinar penalizações para o SVR. Parâmetros como a ponderação distinta dos custos atribuídos a pontos

⁶É fácil observar que a penalização na função objetivo do SVR ocorrerá se, e somente se, o valor de $L_\varepsilon[y_i, f(\mathbf{x}_i)]$ for diferente de zero.

enviesados positiva ou negativamente podem ser adicionados à formulação do problema, resultando em outras variações para o SVR. Especificamente, o SVR que considera a função perda $L_\varepsilon[y_i, f(x_i)]$, a ser utilizado no presente estudo, é conhecido como ε -SVR. Extensões que incluem estruturas de penalização distintas inclui o ν -SVR (CHANG; LIN, 2002), que introduz um parâmetro adicional de permissividade ao erro ν que permite o balanceamento mais preciso entre a taxa de observações fora da banda de confiança ε e o número de vetores de suporte. A formulação ε -SVR foi escolhida para o presente estudo por se tratar da forma mais utilizada na literatura de previsão do câmbio, além de exigir menor tempo computacional para se realizar a otimização; dado o grande número de casos considerados para o treinamento dos parâmetros (a ser detalhado posteriormente na seção 5.2), optou-se pelo ε -SVR. A figura 4.3 ilustra a ideia central do ε -SVR, que consiste em encontrar uma função de decisão que melhor aproxime um determinado conjunto de dados, considerando também o dilema viés-variância para preservar a capacidade de generalização.

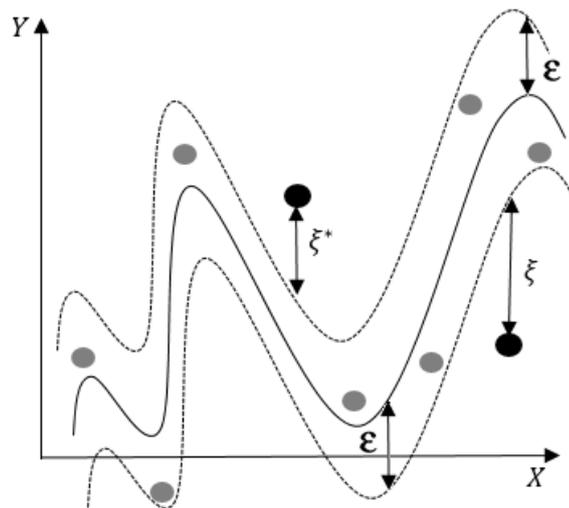


Figura 4.3: Parâmetros e função de decisão do ε -SVR.

Fonte: Extraído de Albuquerque (2014, p. 27).

Fazendo uso do mesmo artifício de dimensionalidade para dados não separáveis linearmente (mapeamento por φ), a função objetivo para o ε -SVR segue como a seguir:

$$\begin{aligned} \text{Minimizar :} & \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \boldsymbol{\xi}^T \mathbf{1} + C \boldsymbol{\xi}^{*T} \mathbf{1} \\ \text{Sujeito a :} & \quad \Phi \mathbf{w} + w_0 - \mathbf{y} \leq \varepsilon \mathbf{1} + \boldsymbol{\xi} \\ & \quad \mathbf{y} - \Phi \mathbf{w} - w_0 \leq \varepsilon \mathbf{1} + \boldsymbol{\xi}^* \\ \text{Tal que :} & \quad \boldsymbol{\xi}, \boldsymbol{\xi}^* \geq 0 \end{aligned}$$

onde $\mathbf{w}_{(q \times 1)}$ é o vetor dos coeficientes angulares do hiperplano de decisão em \mathbb{R}^q ; $w_0 \in \mathbb{R}$ é o

coeficiente linear (intercepto) do hiperplano de decisão em \mathbb{R}^q ; $\Phi_{(n \times q)}$ é a matriz de observações após serem aplicados a φ ; $\mathbf{y}_{(n \times 1)}$ é o vetor que fornece os valores da variável dependente dos pontos observados; $C \in \mathbb{R}$ é o custo do erro – sem perda de generalidade, assume o mesmo valor para pontos enviesados positiva ou negativamente; $\varepsilon > 0$ é a banda de tolerância que define o intervalo de confiança para o qual não há penalização; $\xi_{(n \times 1)}^*$ é o vetor relativo aos pontos acima da banda de tolerância; e $\xi_{(n \times 1)}$ é o vetor relativo aos pontos abaixo da banda de tolerância.

Colocando o problema na forma dual, tem-se que o Lagrangeano é dado por:

$$\begin{aligned} \mathcal{L}(\mathbf{w}, w_0, \xi, \xi^*, \lambda, \lambda^*, \mu, \mu^*) &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \xi^T \mathbf{1} + C \xi^{*T} \mathbf{1} - \lambda^T [-\Phi \mathbf{w} - w_0 + \mathbf{y} + \varepsilon \mathbf{1} + \xi] - \\ &\quad - \lambda^{*T} [\Phi \mathbf{w} + w_0 - \mathbf{y} + \varepsilon \mathbf{1} + \xi^*] - \mu^T \xi - \mu^{*T} \xi^* \end{aligned}$$

onde $\lambda_{(n \times 1)}$, $\lambda_{(n \times 1)}^*$, $\mu_{(n \times 1)}$ e $\mu_{(n \times 1)}^*$ são os multiplicadores de Lagrange associados aos quatro conjuntos de restrições.

As condições de primeira ordem para as variáveis de decisão são:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}} \mathcal{L}(\mathbf{w}, w_0, \xi, \xi^*) &= \mathbf{w}^T - \lambda^T \Phi - \lambda^{*T} \Phi = \mathbf{0} \Rightarrow \mathbf{w}^T = (\lambda^{*T} - \lambda^T) \Phi \\ \frac{\partial}{\partial w_0} \mathcal{L}(\mathbf{w}, w_0, \xi, \xi^*) &= \lambda^T \mathbf{1} - \lambda^{*T} \mathbf{1} = 0 \\ \frac{\partial}{\partial \xi} \mathcal{L}(\mathbf{w}, w_0, \xi, \xi^*) &= C \mathbf{1}^T - \lambda^T - \mu^T = 0 \\ \frac{\partial}{\partial \xi^*} \mathcal{L}(\mathbf{w}, w_0, \xi, \xi^*) &= C \mathbf{1}^T - \lambda^{*T} - \mu^{*T} = 0 \end{aligned}$$

Substituindo na expressão do Lagrangeano e realizando-se algumas manipulações algébricas, chega-se na forma dual a seguir:

$$\begin{aligned} \text{Maximizar :} &\quad -\frac{1}{2} (\lambda^{*T} - \lambda^T) \cdot \Phi \Phi^T \cdot (\lambda^* - \lambda) + \lambda^T \mathbf{1} - \varepsilon \lambda^T \mathbf{1} - \lambda^T \mathbf{y} - \varepsilon \lambda^{*T} \mathbf{1} + \lambda^{*T} \mathbf{y} \\ \text{Sujeito a :} &\quad \mathbf{0} \leq \lambda \leq C \mathbf{1} \\ &\quad \mathbf{0} \leq \lambda^* \leq C \mathbf{1} \end{aligned}$$

onde $\Phi \Phi^T$ é a matriz $K_{n \times n}$ cujos elementos k_{ij} são dados por $k_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j) = \varphi^T(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$. A expressão da função de decisão do SVR não linear é análoga ao caso do SVM não linear, com a distinção de retornar uma estimativa numérica, em vez de um sinal que indica uma das duas classes existentes. Dessa forma, a função de decisão será da forma $\mathbf{w}^T \varphi(\mathbf{x}) - w_0$; como a computação de φ é onerosa, é conveniente substituir a condição $\mathbf{w}^T = (\lambda^{*T} - \lambda^T) \Phi$ obtida

anteriormente para que a função de decisão seja escrita em termos da função *Kernel*

$$\begin{aligned}
 f(\mathbf{x}_i) &= \text{sgn}(\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - w_0) \\
 &= \text{sgn}((\boldsymbol{\lambda}^T \mathbf{D} \boldsymbol{\Phi}) \boldsymbol{\varphi}(\mathbf{x}_i) - w_0) \\
 &= \text{sgn}(\boldsymbol{\varphi}(\mathbf{x}_i)^T (\boldsymbol{\Phi}^T \mathbf{D} \boldsymbol{\lambda}) - w_0) \\
 &= \text{sgn} \left(\sum_{j=1}^n \kappa(\mathbf{x}_i, \mathbf{x}_j) y_j \lambda_j - w_0 \right) \quad \blacksquare
 \end{aligned}$$

4.3 Funções *Kernel*

4.3.1 Condição de Mercer

Conforme discutido anteriormente, o *Kernel Trick* provê grande versatilidade ao método do SVM, ao permitir testar qualquer função *Kernel* resultante de um produto interno entre as transformações φ . Porém, há de se notar que nem toda função κ sintetiza um produto interno em algum espaço de característica de dimensionalidade arbitrária. A condição de admissibilidade de Mercer (MERCER, 1909) delimita uma classe de funções κ que podem ser utilizadas para generalizar um produto interno – e que conseqüentemente poderão ser aplicadas ao algoritmo de aprendizagem. A condição é enunciada a seguir:

(Teorema de Mercer): Seja \mathcal{X} o espaço de entrada (*input space*). Uma função simétrica $\kappa(\mathbf{x}_i, \mathbf{x}_j) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ pode ser escrita como o produto interno $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i) \cdot \boldsymbol{\varphi}(\mathbf{x}_j)$ para algum mapeamento $\boldsymbol{\varphi}(\mathbf{x}) : \mathcal{X} \rightarrow \mathcal{H}$ que leva a algum espaço de característica (*feature space*) \mathcal{H} se, para todo subconjunto finito $\{\mathbf{x}_1, \dots, \mathbf{x}_p\} \in \mathcal{X}$, a matriz gramiana

$$\mathbf{K} = \begin{pmatrix} \kappa(\mathbf{x}_1, \mathbf{x}_1) & \kappa(\mathbf{x}_1, \mathbf{x}_2) & \cdots & \kappa(\mathbf{x}_1, \mathbf{x}_n) \\ \kappa(\mathbf{x}_2, \mathbf{x}_1) & \kappa(\mathbf{x}_2, \mathbf{x}_2) & \cdots & \kappa(\mathbf{x}_2, \mathbf{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa(\mathbf{x}_n, \mathbf{x}_1) & \kappa(\mathbf{x}_n, \mathbf{x}_2) & \cdots & \kappa(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix}_{(n \times n)}$$

é positiva semi-definida (possui todos os autovalores não-negativos).

Uma função κ que satisfaz a condição de Mercer é dita ser **admissível**, de modo que $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i) \cdot \boldsymbol{\varphi}(\mathbf{x}_j)$ pode ser especificado diretamente sem a necessidade de explicitar a forma funcional de $\boldsymbol{\varphi}$, o qual será um mapeamento implícito para algum espaço de característica \mathcal{H} . O teorema de Mercer garante que, para qualquer *Kernel* admissível κ é possível construir o espaço de Hilbert \mathcal{H} (e conseqüentemente $\boldsymbol{\varphi}$) no qual κ computa um produto interno em \mathcal{H} , o

qual é denominado *Reproducing Kernel Hilbert Space* – RKHS⁷.

Há de se notar que algoritmos de aprendizagem de máquinas podem ser aplicados com funções *Kernel* que não satisfazem a condição de Mercer; essa classe mais geral de funções diz respeito às funções **condicionalmente positiva-definida**, os quais podem ser aplicados – sobre algumas circunstâncias específicas – a algoritmos que utilizam de métodos *Kernel* (SCHÖLKOPF; SMOLA, 2002, p. 48–49). Desse modo, há funções *Kernel* condicionalmente positiva-definidas que generalizam métricas de similaridade em espaços vetoriais normados que não necessariamente são um produto interno⁸.

A literatura científica de aprendizado de máquinas mostra que, a depender da função κ utilizada, o desempenho preditivo do algoritmo sofre alterações significativas, de modo que a escolha de uma função *Kernel* adequada se torna um ponto crucial para a acurácia do modelo. Porém, a escolha ideal para a função *Kernel* permanece como uma questão em aberto nas produções científicas em aprendizado de máquinas; dada a ausência de um critério rigoroso para a seleção (ou a exclusão) de determinada função *Kernel*, a grande maioria dos estudos correlatos tem utilizado poucas funções, predominantemente os *Kernels* linear, polinomial e Gaussiano.

Porém, qualquer função κ que satisfaz a condição de admissibilidade de Mercer pode ser utilizada para derivar uma matriz K que, ao ser substituída no problema do SVR, fornecerá uma função de decisão diferente. Dado que a escolha da função *Kernel* possui um grande impacto na qualidade das previsões, o presente estudo irá testar as previsões realizadas por 9 funções *Kernel* distintas, comparando cada conjunto de previsões com o *benchmark* e verificar se as escolhas distintas para κ exercem influência significativa no poder explicativo do modelo. As funções *Kernel* consideradas para o presente estudo estão listadas na tabela 4.1.

As subseções a seguir farão uma breve discussão acerca da forma funcional das transformações ϕ que são generalizadas por cada função *Kernel* adotada; com isso, verificar-se-á que a introdução da função *Kernel* diminui consideravelmente a complexidade dos cálculos necessários para se modelar as interações em espaços de maiores dimensões. Para os *Kernels* polinomial e Gaussiano, serão computadas as formas funcionais explícitas para suas respectivas transformações ϕ cujo produto interno resulta na expressão sintética da função κ . Para os demais *Kernels*, essa demonstração não será desenvolvida devido à extensão e à complexidade dos cálculos; em vez disso, serão indicadas as referências de artigos que fazem essa demonstração.

⁷Para os fundamentos da teoria de RKHS, ver Aronszajn (1950)

⁸Schölkopf e Smola (2002, p. 49–51) discute a representação do mapeamento no espaço de características para funções condicionalmente positiva-definidas, bem como suas conexões com a classe de funções positiva-definidas.

Denominação/parâmetros	Forma funcional ($\kappa(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$)	Principais referências
<i>Kernel</i> polinomial(d, q)	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = [(\mathbf{x}_i \cdot \mathbf{x}_j) + d]^q, d \in \mathbb{R}, q \in \mathbb{N}$	Kamruzzaman, Sarker e Ahmad (2003), Cao, Pang e Bai (2005), Beltrami, Loch e Silva (2011), Sun e Li (2012)
<i>Kernel</i> Gaussiano(σ)	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{2\sigma^2}\right), \sigma \neq 0$	Ince e Trafalis (2006), Liu (2010), Li e Suohai (2013), Were et al. (2015), Gong et al. (2016)
<i>Kernel</i> Log(d)	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = -\log(\ \mathbf{x}_i - \mathbf{x}_j\ ^d + 1), d \geq 0$	Boughorbel, Tarel e Boujemaa (2005), Gaspar, Carbonell e Oliveira (2012), Eckhard et al. (2014)
<i>Kernel</i> multiquádrico inverso(c)	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\sqrt{\ \mathbf{x}_i - \mathbf{x}_j\ ^2 + c^2}}$	Micchelli (1986), Mercier e Lennon (2003)
<i>Kernel</i> Cauchy(σ)	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{1 + \frac{\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{\sigma^2}}, \sigma \neq 0$	Basak (2008), Gaspar, Carbonell e Oliveira (2012), Tian, Wang e Li (2014)
<i>Kernel</i> Spline estável(c, λ)	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} c \frac{\lambda^{2\mathbf{x}_i}}{2} \left(\lambda^{\mathbf{x}_j} - \frac{\lambda^{\mathbf{x}_i}}{3} \right), & \ \mathbf{x}_i\ \geq \ \mathbf{x}_j\ , \\ c \frac{\lambda^{2\mathbf{x}_j}}{2} \left(\lambda^{\mathbf{x}_i} - \frac{\lambda^{\mathbf{x}_j}}{3} \right), & \ \mathbf{x}_i\ < \ \mathbf{x}_j\ . \end{cases}, c \geq 0, 0 \leq \lambda \leq 1$	Pillonetto e Nicolao (2010), Chen, Ohlsson e Ljung (2012), Pillonetto et al. (2014)

Tabela 4.1: Funções *Kernel* utilizadas no estudo

4.3.2 *Kernel* polinomial

Seja o *Kernel* polinomial de ordem q dado por⁹ $\kappa(\mathbf{x}_i, \mathbf{x}_j) = [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^q$, proceder-se-á para a obtenção explícita da transformação φ que é generalizada por $\kappa(\mathbf{x}_i, \mathbf{x}_j)$. Para o caso em que $q = 1$ (“*Kernel* linear”), a verificação é imediata, bastando tomar a seguinte transformação:

$$\varphi(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}^{p+1}$$

$$\mathbf{x} \rightarrow \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \\ 1 \end{pmatrix}$$

⁹Por conveniência, assumiu-se o parâmetro $d = 1$; a derivação é análoga para qualquer $d \in \mathbb{R}$

Tomando o produto interno euclidiano (somatório dos produtos cruzados), obtém-se:

$$\begin{aligned}
 \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) &= \begin{pmatrix} x_{i1} & x_{i2} & \cdots & x_{ip} & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jp} \\ 1 \end{pmatrix} \\
 &= x_{i1}x_{j1} + x_{i2}x_{j2} + \cdots + x_{in}x_{jp} + 1 \\
 &= (x_{i1}x_{j1} + x_{i2}x_{j2} + \cdots + x_{in}x_{jp}) + 1 \\
 &= [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^1 \quad \blacksquare
 \end{aligned}$$

Para a decomposição do *Kernel* polinomial com $q = 2$ (“*Kernel* quadrático”), considera-se inicialmente o caso mais simples em que há apenas uma variável explicativa (o vetor de variáveis $\mathbf{x} \in \mathbb{R}^1$), observa-se que a transformação φ cujo produto interno é sintetizado pela expressão da função *Kernel* $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \kappa(x_{i1}, x_{j1}) = [(x_{i1} \cdot x_{j1}) + 1]^2$ é dado por:

$$\begin{aligned}
 \varphi(\mathbf{x}) : \mathbb{R}^1 &\rightarrow \mathbb{R}^3 \\
 \begin{pmatrix} x_1 \end{pmatrix} &\rightarrow \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1 \\ 1 \end{pmatrix}
 \end{aligned}$$

Desenvolvendo o produto interno $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$, tem-se que:

$$\begin{aligned}
 \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) &= \begin{pmatrix} x_{i1}^2 & \sqrt{2}x_{i1} & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{j1}^2 \\ \sqrt{2}x_{j1} \\ 1 \end{pmatrix} \\
 &= x_{i1}^2 x_{j1}^2 + 2x_{i1}x_{j1} + 1 \\
 &= (x_{i1}x_{j1} + 1)^2 \\
 &= [(x_{i1} \cdot x_{j1}) + 1]^2 \\
 &= [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^2 \quad \blacksquare
 \end{aligned}$$

Para o caso em que $\mathbf{x} \in \mathbb{R}^2$, a transformação $\varphi(\mathbf{x})$ adequada para que o produto interno

$\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ continue sendo expressado pelo *Kernel* quadrático $[(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^2$ é:

$$\varphi(\mathbf{x}) : \mathbb{R}^2 \rightarrow \mathbb{R}^6$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \sqrt{2}x_1x_2 \\ 1 \end{pmatrix}$$

Novamente, desenvolvendo o produto interno $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$, observa-se que a nova transformação φ induzida para o caso em que há duas variáveis explicativas resulta na mesma expressão definida pela função *Kernel* quadrática, conforme elucidado a seguir:

$$\begin{aligned} \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) &= \begin{pmatrix} x_{i1}^2 & x_{i2}^2 & \sqrt{2}x_{i1} & \sqrt{2}x_{i2} & \sqrt{2}x_{i1}x_{i2} & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{j1}^2 \\ x_{j2}^2 \\ \sqrt{2}x_{j1} \\ \sqrt{2}x_{j2} \\ \sqrt{2}x_{j1}x_{j2} \\ 1 \end{pmatrix} \\ &= x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + 2x_{i1}x_{j1}x_{i2}x_{j2} + 1 \\ &= (x_{i1}x_{j1})^2 + (x_{i2}x_{j2})^2 + 2(x_{i1}x_{j1}x_{i2}x_{j2}) + 2(x_{i1}x_{j1} + x_{i2}x_{j2}) + 1 \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2 + 2(x_{i1}x_{j1} + x_{i2}x_{j2}) + 1 \\ &= [(x_{i1}x_{j1} + x_{i2}x_{j2}) + 1]^2 \\ &= [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^2 \quad \blacksquare \end{aligned}$$

Para o caso em que $\mathbf{x} \in \mathbb{R}^3$, a transformação $\varphi(\mathbf{x})$ deve ser novamente ajustada para que o produto interno $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ ainda possa ser expresso pela expressão do *Kernel* quadrático

$[(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^2$, de modo que:

$$\varphi(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}^{10}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \sqrt{2}x_3 \\ \sqrt{2}x_1x_2 \\ \sqrt{2}x_1x_3 \\ \sqrt{2}x_2x_3 \\ 1 \end{pmatrix}$$

O produto interno $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ redefinido para $\mathbf{x} \in \mathbb{R}^3$ preserva a mesma expressão definida pela função *Kernel* quadrática, pois:

$$\begin{aligned} \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) &= \begin{pmatrix} x_{i1}^2 & x_{i2}^2 & x_{i3}^2 & \sqrt{2}x_{i1} & \sqrt{2}x_{i2} & \sqrt{2}x_{i3} & \sqrt{2}x_{i1}x_{i2} & \sqrt{2}x_{i1}x_{i3} & \sqrt{2}x_{i2}x_{i3} & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{j1}^2 \\ x_{j2}^2 \\ x_{j3}^2 \\ \sqrt{2}x_{j1} \\ \sqrt{2}x_{j2} \\ \sqrt{2}x_{j3} \\ \sqrt{2}x_{j1}x_{j2} \\ \sqrt{2}x_{j1}x_{j3} \\ \sqrt{2}x_{j2}x_{j3} \\ 1 \end{pmatrix} \\ &= x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + x_{i3}^2x_{j3}^2 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + 2x_{i3}x_{j3} + \\ &\quad + 2x_{i1}x_{j1}x_{i2}x_{j2} + 2x_{i1}x_{j1}x_{i3}x_{j3} + 2x_{i2}x_{j2}x_{i3}x_{j3} + 1 \\ &= (x_{i1}x_{j1})^2 + (x_{i2}x_{j2})^2 + (x_{i3}x_{j3})^2 + 2(x_{i1}x_{j1}x_{i2}x_{j2}) + 2(x_{i1}x_{j1}x_{i3}x_{j3}) + 2(x_{i2}x_{j2}x_{i3}x_{j3}) + \\ &\quad + 2(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3}) + 1 \\ &= \left\{ (x_{i1}x_{j1})^2 + 2(x_{i1}x_{j1})(x_{i2}x_{j2} + x_{i3}x_{j3}) + \left[(x_{i2}x_{j2})^2 + (x_{i3}x_{j3})^2 + 2(x_{i2}x_{j2}x_{i3}x_{j3}) \right] \right\} + \\ &\quad + 2(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3}) + 1 \\ &= \left[(x_{i1}x_{j1})^2 + 2(x_{i1}x_{j1})(x_{i2}x_{j2} + x_{i3}x_{j3}) + (x_{i2}x_{j2} + x_{i3}x_{j3})^2 \right] + \\ &\quad + 2(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3}) + 1 \\ &= [x_{i1}x_{j1} + (x_{i2}x_{j2} + x_{i3}x_{j3})]^2 + 2(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3}) + 1 \\ &= [(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3}) + 1]^2 \\ &= [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^2 \quad \blacksquare \end{aligned}$$

Procedendo por indução finita em p para o caso com vetores de variáveis explicativas $\mathbf{x} \in \mathbb{R}^p$, obtém-se a expressão geral para a transformação $\varphi(\mathbf{x})$ cujo produto interno $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ pode ser computado sucintamente pela função *Kernel* quadrática $[(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^2$, tal que:

$$\varphi(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}^{2p + \binom{p}{2} + 1}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_p^2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ \vdots \\ \sqrt{2}x_p \\ \sqrt{2}x_1x_2 \\ \sqrt{2}x_1x_3 \\ \vdots \\ \sqrt{2}x_1x_p \\ \sqrt{2}x_2x_3 \\ \sqrt{2}x_2x_4 \\ \vdots \\ \sqrt{2}x_2x_p \\ \vdots \\ \sqrt{2}x_{p-1}x_p \\ 1 \end{pmatrix}$$

A seguir, obter-se-á a expressão de φ para o *Kernel* polinomial de ordem 3, dado por $\kappa(\mathbf{x}_i, \mathbf{x}_j) = [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^3$. Novamente, num cenário inicial com apenas uma variável explicativa, a transformação adequada é dada por:

$$\varphi(\mathbf{x}) : \mathbb{R}^1 \rightarrow \mathbb{R}^4$$

$$\begin{pmatrix} x_1 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^3 \\ \sqrt{3}x_1^2 \\ \sqrt{3}x_1 \\ 1 \end{pmatrix}$$

Tomando o produto interno, tem-se que:

$$\begin{aligned}
 \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) &= \begin{pmatrix} x_{i1}^3 & \sqrt{3}x_{i1}^2 & \sqrt{3}x_{i1} & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{j1}^3 \\ \sqrt{3}x_{j1}^2 \\ \sqrt{3}x_{j1} \\ 1 \end{pmatrix} \\
 &= x_{i1}^3 x_{j1}^3 + 3x_{i1}^2 x_{j1}^2 + 3x_{i1} x_{j1} + 1 \\
 &= (x_{i1} x_{j1} + 1)^3 \\
 &= [(x_{i1} \cdot x_{j1}) + 1]^3 \\
 &= [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^3 \quad \blacksquare
 \end{aligned}$$

Quando $\mathbf{x} \in \mathbb{R}^2$, a transformação $\varphi(\mathbf{x})$ a ser utilizada $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ resulte no *Kernel* polinomial de ordem 3 é dada por:

$$\varphi(\mathbf{x}) : \mathbb{R}^2 \rightarrow \mathbb{R}^{10}$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^3 \\ x_2^3 \\ \sqrt{3}x_1^2 \\ \sqrt{3}x_2^2 \\ \sqrt{3}x_1 \\ \sqrt{3}x_2 \\ \sqrt{3}x_1^2 x_2 \\ \sqrt{3}x_1 x_2^2 \\ \sqrt{3}\sqrt{2}x_1 x_2 \\ 1 \end{pmatrix}$$

Desenvolvendo $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$:

$$\begin{aligned}
 \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) &= \left(x_{i1}^3 \quad x_{i2}^3 \quad \sqrt{3}x_{i1}^2 \quad \sqrt{3}x_{i2}^2 \quad \sqrt{3}x_{i1} \quad \sqrt{3}x_{i2} \quad \sqrt{3}x_{i1}^2x_{i2} \quad \sqrt{3}x_{i1}x_{i2}^2 \quad \sqrt{3}\sqrt{2}x_{i1}x_{i2} \quad 1 \right) \cdot \begin{pmatrix} x_{j1}^3 \\ x_{j2}^3 \\ \sqrt{3}x_{j1}^2 \\ \sqrt{3}x_{j2}^2 \\ \sqrt{3}x_{j1} \\ \sqrt{3}x_{j2} \\ \sqrt{3}x_{j1}^2x_{j2} \\ \sqrt{3}x_{j1}x_{j2}^2 \\ \sqrt{3}\sqrt{2}x_{j1}x_{j2} \\ 1 \end{pmatrix} \\
 &= x_{i1}^3x_{j1}^3 + x_{i2}^3x_{j2}^3 + 3x_{i1}^2x_{j1}^2 + 3x_{i2}^2x_{j2}^2 + 3x_{i1}^2x_{j1}^2x_{i2}x_{j2} + 3x_{i1}x_{j1}x_{i2}^2x_{j2}^2 + \\
 &\quad + 3x_{i1}x_{j1} + 3x_{i2}x_{j2} + 6x_{i1}x_{j1}x_{i2}x_{j2} + 1 \\
 &= [x_{i1}^3x_{j1}^3 + x_{i2}^3x_{j2}^3 + 3x_{i1}^2x_{j1}^2x_{i2}x_{j2} + 3x_{i1}x_{j1}x_{i2}^2x_{j2}^2] + \\
 &\quad + 3[(x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{j1}x_{i2}x_{j2})] + 3(x_{i1}x_{j1} + x_{i2}x_{j2}) + 1 \\
 &= [(x_{i1}x_{j1} + x_{i2}x_{j2})^3] + 3[(x_{i1}x_{j1} + x_{i2}x_{j2})^2] + 3(x_{i1}x_{j1} + x_{i2}x_{j2}) + 1 \\
 &= [(x_{i1}x_{j1} + x_{i2}x_{j2}) + 1]^3 \\
 &= [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^3 \quad \blacksquare
 \end{aligned}$$

Com $\mathbf{x} \in \mathbb{R}^3$, a transformação $\varphi(\mathbf{x})$ cujo produto interno induz o *Kernel* polinomial de grau

3 segue como abaixo:

$$\varphi(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}^{20}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^3 \\ x_2^3 \\ x_3^3 \\ \sqrt{3}x_1^2 \\ \sqrt{3}x_2^2 \\ \sqrt{3}x_3^2 \\ \sqrt{3}x_1 \\ \sqrt{3}x_2 \\ \sqrt{3}x_3 \\ \sqrt{3}x_1^2x_2 \\ \sqrt{3}x_1x_2^2 \\ \sqrt{3}x_1^2x_3 \\ \sqrt{3}x_1x_3^2 \\ \sqrt{3}x_2^2x_3 \\ \sqrt{3}x_2x_3^2 \\ \sqrt{3}\sqrt{2}x_1x_2 \\ \sqrt{3}\sqrt{2}x_1x_3 \\ \sqrt{3}\sqrt{2}x_2x_3 \\ \sqrt{3}\sqrt{2}x_1x_2x_3 \\ 1 \end{pmatrix}$$

O produto interno $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ resulta na mesma expressão definida pelo *Kernel* polino-

mial, conforme verifica-se a seguir:

$$\begin{aligned}
\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) &= \begin{pmatrix} x_{i1}^3 \\ x_{i2}^3 \\ x_{i3}^3 \\ \sqrt{3}x_{i1}^2 \\ \sqrt{3}x_{i2}^2 \\ \sqrt{3}x_{i3}^2 \\ \sqrt{3}x_{i1} \\ \sqrt{3}x_{i2} \\ \sqrt{3}x_{i3} \\ \sqrt{3}x_{i1}^2x_{i2} \\ \sqrt{3}x_{i1}x_{i2}^2 \\ \sqrt{3}x_{i1}^2x_{i3} \\ \sqrt{3}x_{i1}x_{i3}^2 \\ \sqrt{3}x_{i2}^2x_{i3} \\ \sqrt{3}x_{i2}x_{i3}^2 \\ \sqrt{3}\sqrt{2}x_{i1}x_{i2} \\ \sqrt{3}\sqrt{2}x_{i1}x_{i3} \\ \sqrt{3}\sqrt{2}x_{i2}x_{i3} \\ \sqrt{3}\sqrt{2}x_{i1}x_{i2}x_{i3} \\ 1 \end{pmatrix}^T \cdot \begin{pmatrix} x_{j1}^3 \\ x_{j2}^3 \\ x_{j3}^3 \\ \sqrt{3}x_{j1}^2 \\ \sqrt{3}x_{j2}^2 \\ \sqrt{3}x_{j3}^2 \\ \sqrt{3}x_{j1} \\ \sqrt{3}x_{j2} \\ \sqrt{3}x_{j3} \\ \sqrt{3}x_{j1}^2x_{j2} \\ \sqrt{3}x_{j1}x_{j2}^2 \\ \sqrt{3}x_{j1}^2x_{j3} \\ \sqrt{3}x_{j1}x_{j3}^2 \\ \sqrt{3}x_{j2}^2x_{j3} \\ \sqrt{3}x_{j2}x_{j3}^2 \\ \sqrt{3}\sqrt{2}x_{j1}x_{j2} \\ \sqrt{3}\sqrt{2}x_{j1}x_{j3} \\ \sqrt{3}\sqrt{2}x_{j2}x_{j3} \\ \sqrt{3}\sqrt{2}x_{j1}x_{j2}x_{j3} \\ 1 \end{pmatrix} \\
&= x_{i1}^3x_{j1}^3 + x_{i2}^3x_{j2}^3 + x_{i3}^3x_{j3}^3 + 3x_{i1}^2x_{j1}^2 + 3x_{i2}^2x_{j2}^2 + 3x_{i3}^2x_{j3}^2 \\
&\quad + 3x_{i1}x_{j1} + 3x_{i2}x_{j2} + 3x_{i3}x_{j3} + 3x_{i1}^2x_{j1}^2x_{i2}x_{j2} + 3x_{i1}^2x_{j1}^2x_{i3}x_{j3} + 3x_{i2}^2x_{j2}^2x_{i3}x_{j3} + \\
&\quad + 3x_{i1}x_{j1}x_{i2}^2x_{j2}^2 + 3x_{i1}x_{j1}x_{i3}^2x_{j3}^2 + 3x_{i2}x_{j2}x_{i3}^2x_{j3}^2 + \\
&\quad + 6x_{i1}x_{j1}x_{i2}x_{j2} + 6x_{i1}x_{j1}x_{i3}x_{j3} + 6x_{i2}x_{j2}x_{i3}x_{j3} + 6x_{i1}x_{j1}x_{i2}x_{j2}x_{i3}x_{j3} + 1 \\
&= \{x_{i1}^3x_{j1}^3 + 3[x_{i1}^2x_{j1}^2(x_{i2}x_{j2} + x_{i3}x_{j3})] + 3[x_{i1}x_{j1}(x_{i2}^2x_{j2}^2 + x_{i3}^2x_{j3}^2 + 2x_{i2}x_{j2}x_{i3}x_{j3})] + \\
&\quad [x_{i2}^3x_{j2}^3 + x_{i3}^3x_{j3}^3 + 3(x_{i2}^2x_{j2}^2x_{i3}x_{j3}) + 3(x_{i2}x_{j2}x_{i3}^2x_{j3}^2)]\} + \\
&\quad + 3x_{i1}^2x_{j1}^2 + 3x_{i2}^2x_{j2}^2 + 3x_{i3}^2x_{j3}^2 + 3x_{i1}x_{j1} + 3x_{i2}x_{j2} + 3x_{i3}x_{j3} \\
&\quad + 6x_{i1}x_{j1}x_{i2}x_{j2} + 6x_{i1}x_{j1}x_{i3}x_{j3} + 6x_{i2}x_{j2}x_{i3}x_{j3} + 1 \\
&= \{(x_{i1}x_{j1})^3 + 3[x_{i1}^2x_{j1}^2(x_{i2}x_{j2} + x_{i3}x_{j3})] + 3[x_{i1}x_{j1}(x_{i2}x_{j2} + x_{i3}x_{j3})^2] + (x_{i2}x_{j2} + x_{i3}x_{j3})^3\} + \\
&\quad 3\{x_{i1}^2x_{j1}^2 + 2[x_{i1}x_{j1}(x_{i2}x_{j2} + x_{i3}x_{j3})] + [x_{i2}^2x_{j2}^2 + x_{i3}^2x_{j3}^2 + 2(x_{i2}x_{j2}x_{i3}x_{j3})]\} + \\
&\quad + 3(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3}) + 1 \\
&= [(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3})^3] + 3\{x_{i1}^2x_{j1}^2 + 2[x_{i1}x_{j1}(x_{i2}x_{j2} + x_{i3}x_{j3})] + [(x_{i2}x_{j2} + x_{i3}x_{j3})^2]\} + \\
&\quad + 3(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3}) + 1 \\
&= (x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3})^3 + 3(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3})^2 + 3(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3}) + 1 \\
&= [(x_{i1}x_{j1} + x_{i2}x_{j2} + x_{i3}x_{j3}) + 1]^3 \\
&= [(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^3 \quad \blacksquare
\end{aligned}$$

Realizando-se indução finita em n para o caso em que $\mathbf{x} \in \mathbb{R}^n$, a expressão geral para a $\varphi(\mathbf{x})$ cujo produto interno $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ resulta no *Kernel* polinomial de ordem 3 $[(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^3$, é dada por:

$$\varphi(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}^{3p+3\binom{p}{2}+1}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \rightarrow \begin{pmatrix} x_1^3 \\ \vdots \\ x_p^3 \\ \sqrt{3}x_1^2 \\ \vdots \\ \sqrt{3}x_p^2 \\ \sqrt{3}x_1 \\ \vdots \\ \sqrt{3}x_p \\ \sqrt{3}x_1^2x_2 \\ \vdots \\ \sqrt{3}x_1^2x_p \\ \sqrt{3}x_2^2x_3 \\ \vdots \\ \sqrt{3}x_2^2x_p \\ \vdots \\ \sqrt{3}x_{p-1}^2x_p \\ \sqrt{3}x_1x_2^2 \\ \vdots \\ \sqrt{3}x_1x_p^2 \\ \sqrt{3}x_2x_3^2 \\ \vdots \\ \sqrt{3}x_2x_p^2 \\ \vdots \\ \sqrt{3}x_{p-1}x_p^2 \\ \sqrt{3}\sqrt{2}x_1x_2 \\ \vdots \\ \sqrt{3}\sqrt{2}x_1x_p \\ \sqrt{3}\sqrt{2}x_2x_3 \\ \vdots \\ \sqrt{3}\sqrt{2}x_2x_p \\ \vdots \\ \sqrt{3}\sqrt{2}x_{p-1}x_p \\ \sqrt{3}\sqrt{2}x_1x_2x_3 \\ \vdots \\ \sqrt{3}\sqrt{2}x_1x_2x_p \\ \sqrt{3}\sqrt{2}x_1x_3x_4 \\ \vdots \\ \sqrt{3}\sqrt{2}x_1x_3x_p \\ \vdots \\ \sqrt{3}\sqrt{2}x_{p-2}x_{p-1}x_p \\ 1 \end{pmatrix}$$

Observa-se pelas expressões gerais do *Kernel* polinomial de graus 2 e 3 que o mapa φ contempla todas as interações possíveis entre as p variáveis explicativas que resultem num monômio de grau menor ou igual a 2 e 3, respectivamente. Dessa forma, é possível identificar a estrutura genérica para φ cujo produto interno $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ se resume à expressão $[(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^q$, de modo que, dado um vetor de p variáveis explicativas $\mathbf{x} \in \mathbb{R}^p$, a transformação φ irá conter:

(i) 1 termo de grau zero (escalar)

Para se chegar à expressão $[(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^q$, a constante deve ser igual a 1. Para a expressão mais geral do *Kernel* polinomial $[(\mathbf{x}_i \cdot \mathbf{x}_j) + d]^q$, $d \in \mathbb{R}$, basta trocar o último termo de φ de 1 para $\sqrt[q]{d}$, sem interferir no desenvolvimento dos outros termos. Dessa forma, o escalar presente em φ funciona como um “intercepto” no espaço de característica que essa transformação induz;

(ii) $p \cdot q$ termos correspondentes aos x_i elevados à j -ésima potência, com $i = 1, 2, \dots, p$; $j = 1, 2, \dots, q$;

Para o *Kernel* polinomial de grau q , a transformação φ deverá contar com os termos x_i^j , multiplicados por um escalar adequado para que os termos possam ser agrupados posteriormente num binômio perfeito. Conforme visto nos exemplos anteriores, quando $q = 2$, o “escalar adequado” para os monômios de grau 1 é $\sqrt{2}$, devido ao fato de o binômio $(a + b)^2$ ser fatorado como $a^2 + b^2 + 2ab$, o que faz com que, ao se tomar o produto interno $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$, o produto $\sqrt{2}x_i \cdot \sqrt{2}x_j$ resulte em $2x_ix_j$, termo que será posteriormente agrupado com outros fatores semelhantes e os termos quadráticos para se resumir ao quadrado perfeito. Analogamente, para o caso em que $q = 3$, o “escalar adequado” é $\sqrt{3}$ para os monômios de grau 2 ou 1 com apenas uma variável e para os monômios de grau 3 com duas variáveis, enquanto que é $\sqrt{3}\sqrt{2}$ para os monômios de grau 2 com 2 variáveis e o monômio de grau 3 com 3 variáveis. Isso se deve ao fato de a fatoração do cubo perfeito $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$ possuir os coeficientes 1, 3, 3, 1; no caso em que há p variáveis, o desenvolvimento de $(x_1 + x_2 + \dots + x_p)^3$ exigirá a forma explícita para $(x_2 + x_3 + \dots + x_p)^2$, um quadrado perfeito cujos coeficientes padrão são 1, 2, 1. Dessa forma, para um termo de interação que seja desenvolvido como um cubo perfeito, e posteriormente também como um quadrado perfeito, esse termo irá “herdar” os coeficientes cumulativamente, de modo que no mapa φ o escalar adequado seja $\sqrt{3}\sqrt{2}$, de modo que o produto $\sqrt{3}\sqrt{2}x_{i_1}x_{i_2}x_{i_3} \cdot \sqrt{3}\sqrt{2}x_{j_1}x_{j_2}x_{j_3}$ tornar-se-á $3 \cdot 2x_{i_1}x_{i_2}x_{i_3}x_{j_1}x_{j_2}x_{j_3}$, o qual, ao ser agrupado convenientemente com termos semelhantes, irá convergir para o binômio

perfeito $[(x_i \cdot x_j) + 1]^q$, a expressão exata do *Kernel* polinomial. Dada a relação com os binômios perfeitos, os escalares “adequados” para cada termo podem ser obtidos tomando-se a raiz quadrada dos números da pirâmide de Pascal, de modo que ao serem elevados ao quadrado possam ser posteriormente simplificados via binômio de Newton;

(iii) Para cada $k = 1, 2, \dots, p$, um subconjunto de x_1, x_2, \dots, x_p com k variáveis pode ser agrupado como um monômio de grau l de $\binom{l-1}{k-1}$ maneiras, $l = 1, 2, \dots, q$.¹⁰ Para poder ser sintetizado como um binômio perfeito, φ deve cobrir todas as interações possíveis que geram monômios de grau menor ou igual a q :

- Ao se tomar monômios com apenas uma variável (sem perda de generalidade, x_1), é imediato que se pode obter qualquer monômio de grau q , bastando para tal tomar a q -ésima potência. Para cobrir monômios de grau menor ou igual a um q natural qualquer, para cada $x_i, i = 1, 2, \dots, p$, há q maneiras distintas;
- Para monômios com duas variáveis (x_1 e x_2), não é possível alcançar grau 1, visto que, caso uma das variáveis recebesse expoente zero, o monômio reduzir-se-ia a um caso com apenas uma variável, já contemplados pelo item anterior. Dessa forma, um monômio com duas variáveis pode gerar termos de graus $2, 3, \dots, q$. Fixando um l que pertença a esse intervalo $2, 3, \dots, q$, é possível obter um monômio de grau exatamente igual a l de $\binom{l-1}{2-1}$ maneiras distintas. Assim, para obter o número de combinações que resultem em monômios de grau **menor ou igual** a q , basta realizar o somatório $\sum_{l=2}^q \binom{l-1}{2-1}$. Dado que há $\binom{p}{2}$ maneiras distintas de se tomar grupos de duas variáveis de um total de n variáveis, o número total de monômios a serem considerados para esse cenário é de $\binom{p}{2} \sum_{l=2}^q \binom{l-1}{2-1}$
- Para monômios com três variáveis, o grau mínimo é 3, dado que caso contrário haveria pelo menos uma variável com expoente zero, o que se reduziria a um dos itens anteriores. Novamente, fixando um l que pertença ao intervalo $3, 4, \dots, q$, obtém-se um monômio de grau exatamente igual a l de $\binom{l-1}{3-1}$ maneiras distintas. O somatório $\sum_{l=3}^q \binom{l-1}{3-1}$ fornece o número de combinações de três variáveis que resultem em monômios de grau **menor ou igual** a q . Sabendo que há $\binom{p}{3}$ maneiras distintas de se tomar grupos de três variáveis de um total de n variáveis, o número total de monômios a serem considerados para esse cenário é de $\binom{p}{3} \sum_{l=3}^q \binom{l-1}{3-1}$
- Generalizando a ideia para monômios com i variáveis, $i = 0, 1, 2, \dots, p$ o número de combinações que resultam em um monômio com grau menor ou igual a q é dado

¹⁰Esse resultado advém do problema clássico de combinatória “stars and bars”, o qual consiste em designar l objetos em k categorias, com cada categoria recebendo pelo menos 1 objeto

por $\binom{p}{i} \sum_{l=i}^q \binom{l-1}{i-1}$. De fato, para monômios constituídos por todas as p variáveis, há apenas uma combinação que gera um monômio de grau exatamente n , definido por $x_1 x_2 \cdots x_p$

- Por fim, o número total de termos de interação necessários em φ para cobrir todos os monômios de grau menor ou igual a q composto por qualquer número i de variáveis, $i = 0, 1, 2, \dots, p$, é dado pela expressão $\sum_{i=0}^p \left[\binom{p}{i} \sum_{l=i}^q \binom{l-1}{i-1} \right]$.

Dessa forma, considerando p variáveis explicativas, a transformação φ cujo produto interno $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ sintetiza o *Kernel* polinomial de ordem q $[(\mathbf{x}_i \cdot \mathbf{x}_j) + 1]^q$ será um mapeamento:

$$\varphi(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}^{\left\{ \sum_{i=0}^p \left[\binom{p}{i} \sum_{l=i}^q \binom{l-1}{i-1} \right] \right\}}$$

de modo que os elementos de φ correspondam à totalidade de monômios de grau menor ou igual a q , combinados entre todos os subconjuntos possíveis das p variáveis explicativas. Os coeficientes que multiplicam cada termo de interação dentro do vetor φ serão derivados da Pirâmide de Pascal.

Exemplificando com um caso simples com $p = 3$ (três variáveis x_1, x_2 e x_3) e $q = 6$: O número de termos necessários para se computar todos os monômios de grau menor ou igual a $q = 6$ é $\sum_{i=0}^3 \left[\binom{3}{i} \sum_{l=i}^6 \binom{l-1}{i-1} \right] = 1 + 18 + 45 + 20 = 84$ (ou seja, nesse caso $\varphi(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}^{84}$), enumerados a seguir:

- (i) Termos com 0 variáveis: Será um termo de grau zero – um escalar, conforme abordado anteriormente; para retornar a forma genérica do *Kernel* polinomial $[(\mathbf{x}_i \cdot \mathbf{x}_j) + d]^q$ (com $d \in \mathbb{R}$ o termo de intercepto), basta definir esse termo como sendo igual a $\sqrt[q]{d} \Rightarrow \binom{3}{0} \binom{0-1}{0-1} = 1$ termo
- (ii) Termos com 1 variável (monômios gerados têm grau maior ou igual a 1):
 - Monômios de grau 1: $x_1, x_2, x_3 \Rightarrow \binom{3}{1} \binom{1-1}{1-1} = 3$ termos
 - Monômios de grau 2: $x_1^2, x_2^2, x_3^2 \Rightarrow \binom{3}{1} \binom{2-1}{1-1} = 3$ termos
 - Monômios de grau 3: $x_1^3, x_2^3, x_3^3 \Rightarrow \binom{3}{1} \binom{3-1}{1-1} = 3$ termos
 - Monômios de grau 4: $x_1^4, x_2^4, x_3^4 \Rightarrow \binom{3}{1} \binom{4-1}{1-1} = 3$ termos
 - Monômios de grau 5: $x_1^5, x_2^5, x_3^5 \Rightarrow \binom{3}{1} \binom{5-1}{1-1} = 3$ termos
 - Monômios de grau 6: $x_1^6, x_2^6, x_3^6 \Rightarrow \binom{3}{1} \binom{6-1}{1-1} = 3$ termos

(iii) Termos com 2 variáveis (monômios gerados têm grau maior ou igual a 2):

- Monômios de grau 2: $x_1x_2, x_1x_3, x_2x_3 \Rightarrow \binom{3}{2} \binom{2-1}{2-1} = 3$ termos
- Monômios de grau 3: $x_1^2x_2, x_1^2x_3, x_2^2x_3, x_1x_2^2, x_1x_3^2, x_2x_3^2 \Rightarrow \binom{3}{2} \binom{3-1}{2-1} = 6$ termos
- Monômios de grau 4: $x_1^3x_2, x_1^3x_3, x_2^3x_3, x_1x_2^3, x_1x_3^3, x_2x_3^3, x_1^2x_2^2, x_1^2x_3^2, x_2^2x_3^2 \Rightarrow \binom{3}{2} \binom{4-1}{2-1} = 9$ termos
- Monômios de grau 5: $x_1^4x_2, x_1^4x_3, x_2^4x_3, x_1x_2^4, x_1x_3^4, x_2x_3^4, x_1^3x_2^2, x_1^3x_3^2, x_2^3x_3^2, x_1^2x_2^3, x_1^2x_3^3, x_2^2x_3^3 \Rightarrow \binom{3}{2} \binom{5-1}{2-1} = 12$ termos
- Monômios de grau 6: $x_1^5x_2, x_1^5x_3, x_2^5x_3, x_1x_2^5, x_1x_3^5, x_2x_3^5, x_1^4x_2^2, x_1^4x_3^2, x_2^4x_3^2, x_1^2x_2^4, x_1^2x_3^4, x_2^2x_3^4, x_1^3x_2^3, x_1^3x_3^3, x_2^3x_3^3 \Rightarrow \binom{3}{2} \binom{6-1}{2-1} = 15$ termos

(iv) Termos com 3 variáveis (monômios gerados têm grau maior ou igual a 3):

- Monômios de grau 3: $x_1x_2x_3 \Rightarrow \binom{3}{3} \binom{3-1}{3-1} = 1$ termo
- Monômios de grau 4: $x_1^2x_2x_3, x_1^2x_2^2x_3^2 \Rightarrow \binom{3}{3} \binom{4-1}{3-1} = 3$ termos
- Monômios de grau 5: $x_1^3x_2x_3, x_1x_2^3x_3, x_1x_2x_3^3, x_1^2x_2^2x_3, x_1^2x_2x_3^2, x_1x_2^2x_3^2 \Rightarrow \binom{3}{3} \binom{5-1}{3-1} = 6$ termos
- Monômios de grau 6: $x_1^4x_2x_3, x_1x_2^4x_3, x_1x_2x_3^4, x_1^3x_2^2x_3, x_1^3x_2x_3^2, x_1^2x_2^3x_3, x_1^2x_2x_3^3, x_1x_2^3x_3^2, x_1x_2^2x_3^3, x_1^2x_2^2x_3^2 \Rightarrow \binom{3}{3} \binom{6-1}{3-1} = 10$ termos

O presente estudo considerou valores de $q = 1, 2, 3, 4$. Valores maiores que 4 para q não foram considerados tendo em vista o fato de a grande parte dos estudos em aprendizado de máquina considerarem apenas os *Kernels* polinomiais de graus 1 e 2, de modo que para interações de grau superior em geral recorre-se ao *Kernel* Gaussiano, abordado do tópico seguinte.

4.3.3 *Kernel* Gaussiano

Uma das funções *Kernel* mais utilizadas na literatura de aprendizado de máquinas é o *Kernel* Gaussiano – cuja expressão se dá por $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$ – um caso específico da família de funções *Radial Basis Function (RBF)*. Esse *Kernel* é particularmente notável devido ao fato de computar diretamente um produto interno de mapas φ definidas em dimensões infinitas. Desenvolvendo a expressão e realizando a expansão por série de Taylor da função e^x , observa-se que a expressão do *Kernel* Gaussiano pode ser escrito como:

$$\begin{aligned}
\exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) &= \exp\left(-\frac{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}{2\sigma^2}\right) \\
&= \exp\left(-\frac{x_{i1}^2 + x_{j1}^2 + \dots + x_{ip}^2 + x_{jp}^2 - 2(x_{i1}x_{j1} + \dots + x_{ip}x_{jp})}{2\sigma^2}\right) \\
&= \exp\left(-\frac{\|\mathbf{x}_i\|^2 + \|\mathbf{x}_j\|^2 - 2(\mathbf{x}_i \cdot \mathbf{x}_j)}{2\sigma^2}\right) \\
&= \exp\left(-\frac{\|\mathbf{x}_i\|^2}{2\sigma^2} - \frac{\|\mathbf{x}_j\|^2}{2\sigma^2}\right) \cdot \exp\left(\frac{2(\mathbf{x}_i \cdot \mathbf{x}_j)}{2\sigma^2}\right) \\
&= \exp\left(-\frac{\|\mathbf{x}_i\|^2}{2\sigma^2} - \frac{\|\mathbf{x}_j\|^2}{2\sigma^2}\right) \cdot \left(\frac{e(0)}{0!} + \frac{e'(0) \cdot \frac{2(\mathbf{x}_i \cdot \mathbf{x}_j)}{2\sigma^2}}{1!} + \frac{e''(0) \cdot \left(\frac{2(\mathbf{x}_i \cdot \mathbf{x}_j)}{2\sigma^2}\right)^2}{2!} + \dots\right) \\
&= \exp\left(-\frac{\|\mathbf{x}_i\|^2}{2\sigma^2} - \frac{\|\mathbf{x}_j\|^2}{2\sigma^2}\right) \cdot \left(1 + \frac{2(\mathbf{x}_i \cdot \mathbf{x}_j)}{2\sigma^2} + \frac{\left(\frac{2(\mathbf{x}_i \cdot \mathbf{x}_j)}{2\sigma^2}\right)^2}{2!} + \dots\right)
\end{aligned}$$

Pela expressão acima fica evidente que o *Kernel* Gaussiano pode ser escrito como um produto interno de $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ em dimensão infinita¹¹, tal que:

¹¹Uma discussão detalhada acerca do RKHS induzido pelo *Kernel* Gaussiano pode ser encontrada em Steinwart, Hush e Scovel (2006).

$$\varphi(\mathbf{x}) : \mathbb{R}^p \rightarrow \mathbb{R}^\infty$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \rightarrow e^{-\frac{\|\mathbf{x}\|^2}{2\sigma^2}} \cdot \begin{pmatrix} 1 \\ x_1 \cdot \sqrt{\frac{2^0}{\sigma^2 \cdot 1!}} \\ \vdots \\ x_p \cdot \sqrt{\frac{2^0}{\sigma^2 \cdot 1!}} \\ x_1^2 \cdot \sqrt{\frac{2^1}{(\sigma^2)^2 \cdot 2!}} \\ \vdots \\ x_p^2 \cdot \sqrt{\frac{2^1}{(\sigma^2)^2 \cdot 2!}} \\ \sqrt{2} \cdot x_1 \cdot \sqrt{\frac{2^1}{(\sigma^2)^2 \cdot 2!}} \\ \vdots \\ \sqrt{2} \cdot x_n \cdot \sqrt{\frac{2^1}{(\sigma^2)^2 \cdot 2!}} \\ x_1^3 \cdot \sqrt{\frac{2^2}{(\sigma^2)^3 \cdot 3!}} \\ \vdots \\ x_p^3 \cdot \sqrt{\frac{2^2}{(\sigma^2)^3 \cdot 3!}} \\ \vdots \\ \sqrt{3} \cdot x_1^2 x_2 \cdot \sqrt{\frac{2^2}{(\sigma^2)^3 \cdot 3!}} \\ \vdots \\ \sqrt{3} \cdot x_1^2 x_p \cdot \sqrt{\frac{2^2}{(\sigma^2)^3 \cdot 3!}} \\ \vdots \\ \sqrt{3} \cdot x_p^2 x_1 \cdot \sqrt{\frac{2^2}{(\sigma^2)^3 \cdot 3!}} \\ \vdots \\ \sqrt{3} \cdot x_p^2 x_{p-1} \cdot \sqrt{\frac{2^2}{(\sigma^2)^3 \cdot 3!}} \\ \sqrt{3} \cdot x_1 x_2^2 \cdot \sqrt{\frac{2^2}{(\sigma^2)^3 \cdot 3!}} \\ \vdots \\ \sqrt{3} \cdot x_1 x_p^2 \cdot \sqrt{\frac{2^2}{(\sigma^2)^3 \cdot 3!}} \\ \vdots \\ \sqrt{3} \cdot x_p x_1^2 \cdot \sqrt{\frac{2^2}{(\sigma^2)^3 \cdot 3!}} \\ \vdots \\ \sqrt{3} \cdot x_p x_{p-1}^2 \cdot \sqrt{\frac{2^2}{(\sigma^2)^3 \cdot 3!}} \\ \sqrt{3} \sqrt{2} \cdot x_1 x_2 x_3 \\ \sqrt{3} \sqrt{2} \cdot x_1 x_3 x_4 \\ \vdots \\ \sqrt{3} \sqrt{2} \cdot x_1 x_{p-1} x_p \\ \sqrt{3} \sqrt{2} \cdot x_2 x_3 x_4 \\ \vdots \\ \sqrt{3} \sqrt{2} \cdot x_2 x_{p-1} x_p \\ \vdots \\ \sqrt{3} \sqrt{2} \cdot x_{p-2} x_{p-1} x_p \\ \vdots \end{pmatrix}$$

Nota-se que a estrutura da transformação φ do *Kernel* Gaussiano possui grande semelhança com o mapeamento do *Kernel* polinomial, englobando todos os monômios combinados pelos subconjuntos das variáveis explicativas x_1, x_2, \dots, x_p , mas agora incorporando os termos com grau até o infinito, em vez de limitar a um valor $q < \infty$. A convergência da série de Maclaurin para a função e^x garante¹² que o produto interno de dois vetores ∞ -dimensionais possa ser escrito na forma mais sintética como $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$.

4.3.4 *Kernel* multiquádrico inverso

O *Kernel* multiquádrico inverso

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\sqrt{\|\mathbf{x}_i - \mathbf{x}_j\|^2 + c^2}}$$

foi proposto por Micchelli (1986) para a interpolação de dados esparsos em superfícies multiquádricas¹³. Posteriormente, Mercier e Lennon (2003) utilizam essa função em um SVM para classificação de imagens, verificando que o *Kernel* multiquádrico inverso leva em consideração a similaridade espectral dos vetores de suporte ao delimitar a função de decisão, e que obteve resultados favoráveis no sentido de reduzir a ocorrência de classificações errôneas, além de também satisfazer a condição de Mercer. Micchelli (1986) demonstra que o mapeamento φ associado ao *Kernel* multiquádrico inverso, assim como o *Kernel* Gaussiano, também leva os dados originais a um espaço de característica com dimensão infinita. Uma discussão acerca do RKHS do *Kernel* multiquádrico inverso pode ser encontrado em Zhou (2003).

No âmbito de finanças, o *Kernel* multiquádrico inverso foi a função *Kernel* que obteve os melhores resultados em termos de risco-retorno no estudo de Henrique et al. (2016) acerca da seleção de portfólios a partir de empresas constituintes do índice S&P 100 (constituído pelas 100 empresas de maior porte listadas nas bolsas de valores dos EUA). Esse estudo em questão testou 17 funções *Kernel* distintas para classificar as 100 ações como “boas” ou “ruins”, e definir um portfólio como uma combinação dos ativos classificados como “bons” e compará-lo com o *benchmark* de mercado. O *Kernel* multiquádrico inverso obteve um retorno acumulado de 374.40% em 22 trimestres com um Value-At-Risk de -6.87%, sendo este o melhor desempenho entre todas as funções *Kernel* abordadas pelo estudo. Dessa forma, o presente estudo

¹²É fácil verificar que o raio de convergência da referida série é ∞ .

¹³As superfícies multiquádricas são a generalização das superfícies cônicas (elipses, hipérbolas e parábolas) para uma dimensão arbitrária n

incorporou essa função para a análise das previsões da taxa de câmbio, a fim de verificar se o desempenho satisfatório dessa função para a formação de portfólios persiste para uma aplicação alternativa.

4.3.5 Kernel Log

O *Kernel Log*

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = -\log(\|\mathbf{x}_i - \mathbf{x}_j\|^d + 1), d \geq 0$$

foi proposto por Boughorbel, Tarel e Boujemaa (2005) para um problema de reconhecimento de imagens e pertence à classe de *Kernels* condicionalmente positivo-definidos, os quais também podem ser aplicados ao algoritmo de aprendizagem do SVM, com a diferença que geram métricas de similaridade – não necessariamente o produto interno – em espaços de características normados. Por exemplo, o *Kernel Sigmóide* $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\alpha \mathbf{x}_i \cdot \mathbf{x}_j + r)$, amplamente utilizado na literatura de aprendizado de máquinas, é um *Kernel* condicionalmente positivo-definido¹⁴. Boughorbel, Tarel e Boujemaa (2005) deriva a prova de que o *Kernel Log* pode ser aplicado a algoritmos de aprendizagem por SVM, constatando que essa função obteve resultados superiores a funções *Kernel* “clássicas” Mercer-admissíveis. Eckhard et al. (2014) chegam a conclusão similar, aplicando o *Kernel Log* para a estimação de reflectância espectral, também no âmbito de reconhecimento de imagens. Para o âmbito de finanças, o *Kernel Log* também foi analisado por Henrique et al. (2016) para a seleção de portfólios. Especificamente para a previsão da taxa de câmbio, não foram encontrados estudos científicos que tenham feito uso dessa função em algoritmos de aprendizado de máquinas.

4.3.6 Kernel Cauchy

Proposto inicialmente por Basak (2008), o *Kernel Cauchy*

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{1 + \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}}, \sigma \neq 0$$

é baseado na distribuição de Cauchy e possui um comportamento de cauda pesada, o que o torna atrativo para modelar eventos sensíveis a “efeitos contágio” em espaços dimensões elevadas (CHANDRASEKHAR; REDDY, 2015), bem como elevando a precisão do SVM de classificação

¹⁴O *Kernel Sigmóide* não foi considerado para o presente estudo, visto que a literatura indica que o *Kernel Sigmóide* apresenta desempenho inferior ao *Kernel Gaussiano*. Para uma discussão aprofundada acerca desse ponto, ver Lin e Lin (2003).

em relação aos *Kernels* polinomial e Gaussiano (BASAK, 2008). Dentro do contexto de finanças, o *Kernel* Cauchy foi aplicado por Tian, Wang e Li (2014) para modelar o comportamento de valores extremos e de assimetria para medidas de risco, verificando que o *Kernel* Cauchy contribuiu positivamente para suavizar a distorção de medidas de risco causada pela subestimação dos eventos extremos, fenômeno mais propenso a ocorrer ao utilizar o *Kernel* Gaussiano. Dada a verificação do fato estilizado de presença de caudas pesadas na distribuição das taxas de câmbio no mercado FOREX, o presente estudo adotou o *Kernel* Cauchy para capturar esse comportamento de cauda e verificar se essa característica contribui para que as previsões geradas por este *Kernel* possam ser desejáveis.

Em relação às funções *Kernel* menos recorrentes na literatura de previsão por aprendizado de máquinas, Gaspar, Carbonell e Oliveira (2012) testam 6 funções *Kernel* – Linear, Polinomial, Gaussiano, Sigmóide, Log e Cauchy – e verificou suas acurácias de previsão em relação a 9 bancos de dados distintos, cujas naturezas variaram desde dados sobre qualidade de vinhos até estatísticas de ocorrência de hepatite e séries de retornos financeiros. O estudo concluiu que não foram encontradas evidências de algum *Kernel* que tenha provado ser “superior” aos demais ao lidar com conjuntos de dados heterogêneos; porém, as métricas de avaliação apontaram que os *Kernels* mais utilizados a literatura – linear, polinomial e Gaussiano – apresentaram desempenho ligeiramente superior. O presente estudo irá verificar se a incorporação de *Kernels* “não-mainstream” pode trazer melhorias no poder explicativo dos modelos de previsão SVR.

4.3.7 *Kernel* Spline Estável

Apresentado por Pillonetto e Nicolao (2010), o *Kernel* Spline Estável

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} c \frac{\lambda^{2\mathbf{x}_i}}{2} \left(\lambda^{\mathbf{x}_j} - \frac{\lambda^{\mathbf{x}_i}}{3} \right), & \|\mathbf{x}_i\| \geq \|\mathbf{x}_j\|, \\ c \frac{\lambda^{2\mathbf{x}_j}}{2} \left(\lambda^{\mathbf{x}_i} - \frac{\lambda^{\mathbf{x}_j}}{3} \right), & \|\mathbf{x}_i\| < \|\mathbf{x}_j\|. \end{cases}, c \geq 0, 0 \leq \lambda \leq 1$$

é uma extensão do *Kernel* Spline cúbico (GUNN, 1998)

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} c \frac{\mathbf{x}_i^2}{2} \left(\mathbf{x}_j - \frac{\mathbf{x}_i}{3} \right), & \|\mathbf{x}_i\| \geq \|\mathbf{x}_j\|, \\ c \frac{\mathbf{x}_j^2}{2} \left(\mathbf{x}_i - \frac{\mathbf{x}_j}{3} \right), & \|\mathbf{x}_i\| < \|\mathbf{x}_j\|. \end{cases}, c \geq 0, 0 \leq \lambda \leq 1$$

Pillonetto et al. (2014) apontam que as funções Spline possuem propriedades desejáveis para problemas de interpolação e aproximação de dados, de modo que funções suavizadas

por Splines não tendem a apresentar oscilações acentuadas, um problema que pode ocorrer quando polinômios de graus muito elevados são introduzidos¹⁵. Aplicações da função *Kernel Spline* cúbica em finanças incluem os artigos de Kamruzzaman e Sarker (2004) e de Liu (2010), relativos à predição do câmbio; e o trabalho de Coleman, Li e Wang (2013), que formula um algoritmo de otimização para a calibragem da função de volatilidade para a precificação de opções financeiras.

O estudo de Pilonetto e Nicolao (2010) formula o *Kernel Spline* Estável – o qual também induz um espaço de característica com dimensão infinita, além de ser capaz de preservar uma família de distribuições exponenciais estáveis (PILLONETTO; NICOLAO, 2010, p. 84). Testando seu desempenho em relação aos *Kernels Spline* cúbico e Gaussiano mediante simulações de Monte Carlo, o artigo mostra que o *Kernel Spline* Estável apresentou desempenho superior para a identificação de sistemas lineares dinâmicos. Não foram encontradas aplicações do *Kernel* para a previsão de taxas de câmbio, fato este que justifica seu uso no presente estudo.

¹⁵Esse fenômeno foi observado primeiramente por Runge (1901)

5 ANÁLISE EMPÍRICA

5.1 Definição do *benchmark*

O *benchmark* adotado para o presente estudo está embasada na hipótese do passeio aleatório (*Random Walk*), tal que abordado por Fama (1965), a qual preconiza que a trajetória entre o valor presente e o valor no período temporal seguinte possui máxima entropia (ou seja, um componente completamente estocástico). No âmbito de finanças, a implicação dessa hipótese é a de que os preços observados refletem toda a informação passada, de modo que nenhum modelo irá superar consistentemente o passeio aleatório para a previsão de valores futuros de séries financeiras, resultado conhecido como “Hipótese dos mercados eficientes” (FAMA, 1970).

Partindo-se da premissa de que a trajetória da taxa de câmbio observada no período $t - 1$ (s_{t-1}) até o período t (s_t) possui entropia máxima, é possível assumir, sem perda de generalidade, que a probabilidade de $s_{t-1} > s_t$ é igual à probabilidade de $s_{t-1} < s_t$. Dessa forma, sob o *Random Walk*, o incremento do câmbio predito em t (\hat{s}_t) em relação ao câmbio observado em $t - 1$ (s_{t-1}) é modelado por uma distribuição Normal centrada em s_{t-1} e com uma volatilidade condicional variante a cada período temporal¹.

Dado que a distribuição Normal é simétrica em relação à média, um movimento de subida é igualmente provável a um movimento de queda. Assim, **em média**, a variação $\hat{s}_t - s_{t-1}$ é igual a zero – Ou seja: dado o conjunto informacional disponível em $t - 1$ (composto pelas informações de todos os períodos em $t - 1$ e antes de $t - 1$), a hipótese do *Random Walk* diz que **o valor esperado** para t é simplesmente o valor observado em $t - 1$, tal que \hat{s}_t é igual a s_{t-1} , pois:

$$\mathbb{E}[s_t | s_{t-1}, s_{t-2}, \dots] = s_{t-1}$$

onde $\mathbb{E}(\cdot)$ é o valor esperado. Observa-se que a condição $\mathbb{E}[s_t | s_{t-1}, s_{t-2}, \dots] = s_{t-1}$ pode ser

¹Sob essa construção, a trajetória da série temporal é análoga a um movimento Browniano padrão clássico, a qual tem sido aplicado em diversos modelos em finanças, como o estudo clássico de Black e Scholes (1973).

reescrita como

$$\mathbb{E}[s_t - s_{t-1} | s_{t-1}, s_{t-2}, \dots] = 0$$

implicando que o ganho esperado entre $t - 1$ e t é zero. De fato, definindo o retorno entre $t - 1$ e t como $R_t = \log\left(\frac{\hat{s}_t}{s_{t-1}}\right)$, onde $\log(\cdot)$ é o logaritmo natural, é fácil verificar que $R_t = 0, \forall t$, pois $\hat{s}_t = s_{t-1}$ implica que $R_t = \log\left(\frac{s_{t-1}}{s_{t-1}}\right) = \log(1) = 0$. Especificamente para a previsão da taxa de câmbio, essa mesma construção do *benchmark* pode ser encontrada em Qi e Wu (2006).

5.2 Coleta e tratamento dos dados

As quatorze variáveis levantadas no modelo fundamentalista foram coletadas individualmente com auxílio da plataforma *Datastream*, acessada pelo sistema *Thomson Reuters Eikon* versão 4.0.33626. A periodicidade dos dados foi mensal, compreendendo o período entre janeiro de 2000 até dezembro de 2015, totalizando 192 observações de 14 variáveis cada. A seguir, os dados foram padronizados mediante a seguinte expressão:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sigma_j}; i = 1, 2, \dots, 192; j = 1, 2, \dots, 14$$

onde x_{ij} é o valor não padronizado da j -ésima variável na i -ésima observação, \bar{x}_j e σ_j são a média aritmética e o desvio padrão da j -ésima variável. A padronização foi realizada por conveniência computacional, visto que o processo de busca dos parâmetros ótimos do SVR demanda custo computacional menor para dados padronizados; ademais, a padronização é um procedimento que preserva a ordinalidade dos elementos amostrados, o que não interfere na otimização do problema de programação matemática que fornece a função de decisão.

Feito o tratamento da padronização, a nova base de dados composta pelos valores x_{ij}^* foi particionada em três frações mutuamente excludentes: conjuntos de treinamento, de validação e de teste. O propósito dessa segmentação é fazer com que o algoritmo de aprendizado de máquina possa identificar padrões em um conjunto de dados, mas testar seu desempenho em outro conjunto de dados dissociado da base à qual o algoritmo se submeteu, afim de melhor avaliar o real poder explicativo que a função de decisão encontrada possui face a dados “novos”. As proporções escolhidas para cada uma das partições foram:

- (i) Conjunto de treinamento: 96 períodos (50% do total de 192 períodos)
- (ii) Conjunto de validação: 38 períodos (20% do total de 192 períodos)

(iii) Conjunto de teste: 58 períodos (30% do total de 192 períodos)

A definição dos períodos temporais pertencentes a cada um dos subconjuntos de dados foi realizada aleatoriamente a fim de evitar o viés de tendência temporal oriundo dos ciclos macroeconômicos – ou seja, a capacidade preditiva poderia ser prejudicada ou melhorada em decorrência de uma tendência sistêmica, em vez da qualidade do modelo em si. Dessa forma, cada um dos 192 períodos forma sorteados sem reposição entre os três subconjuntos de dados com auxílio da função “*sample*” do *software R*.

Dessa forma, o procedimento de análise empírica pode ser sintetizado a seguir:

1. Conjunto de treinamento:

A esse conjunto de dados aplica-se a otimização do algoritmo SVR para cada uma das funções *Kernel* adotadas, performando-se o *grid search* para cada um dos parâmetros associados a cada caso. Em cada par de câmbio analisado e cada função *Kernel* distinta, uma função de decisão diferente será gerada. Basicamente, o algoritmo irá tentar “aprender” os padrões subjacentes ao conjunto de treinamento e fornecer os parâmetros que otimizem o problema de programação matemática do SVR. Os intervalos de variação dos parâmetros globais ϵ e C considerados para todos os casos (por serem integrantes da formulação do SVR) e os parâmetros específicos utilizados para cada função *Kernel* estão listadas na tabela 5.1.

Parâmetros globais	
$\epsilon = [0.05, 0.1, \dots, 0.95, 1]$	$C = [0.5, 1, 1.5, \dots, 9.5, 10]$
Função <i>Kernel</i>	Parâmetros específicos
<i>Kernel</i> polinomial (d, q)	$d = [0, 0.1, 0.2, \dots, 4.9, 5]; q = [1, 2, 3, 4]$
<i>Kernel</i> Gaussiano (σ)	$\sigma = [0.01, 0.02, \dots, 1.99, 2]$
<i>Kernel</i> Log (d)	$d = [0.1, 0.2, \dots, 4.9, 5]$
<i>Kernel</i> multiquádrico inverso (c)	$c = [0, 0.1, 0.2, \dots, 4.9, 5]$
<i>Kernel</i> Cauchy (σ)	$\sigma = [0.01, 0.02, \dots, 1.99, 2]$
<i>Kernel</i> Spline estável (c, λ)	$c = [0, 0.5, 1, \dots, 9.5, 10]; \lambda = [0, 0.1, \dots, 0.9, 1]$

Tabela 5.1: Intervalos de busca utilizados para os parâmetros de treinamento

2. Conjunto de validação:

Com base em cada combinação de parâmetros aplicados no conjunto de treinamento, a acurácia da respectiva função de decisão obtida com o treinamento é verificada para um conjunto de dados distinto, mediante alguma métrica de erros estabelecida. No presente estudo a

métrica adotada foi o RMSE (raiz quadrada do erro quadrático médio), definido como

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (s_t - \hat{s}_t)^2}{N}}$$

Conforme elucidado no referencial teórico, o RMSE constitui uma das principais métricas utilizadas na literatura de aprendizado de máquinas. Ademais, o RMSE considera uma “penalização adicional” para pontos preditos cujos desvios em relação ao valor observado são muito elevados, diferentemente da função de perda ϵ -insensitiva L_ϵ definida na formulação do SVR, a qual considera uma escala constante de penalização, independente do quão distante a previsão esteja do valor observado. Ou seja, a função L_ϵ considera que uma previsão que esteja a $3X$ unidades do observado é exatamente 3 vezes pior que uma previsão que esteja a X unidades do observado; por outro lado, o RMSE atribui peso maior a pontos muito distantes da média, de modo que uma previsão que esteja a $3X$ unidades do observado será considerada mais que 3 vezes pior que uma previsão que esteja a X unidades do observado².

Dessa forma, cada uma das funções de decisão obtidas no conjunto de treinamento será alimentada com os dados das variáveis independentes do conjunto de validação, de modo a computar a previsão da variável dependente para esses dados. Essa previsão será confrontada com os valores reais observados no conjunto de validação, e a métrica RMSE entre os valores predito e observado é calculada. O processo se repete para cada uma das combinações de parâmetros considerada, e a combinação ótima é aquela que minimiza o RMSE associado à sua previsão. Ao fim dessa etapa, serão fornecidas 135 combinações de parâmetros ótimos (uma para cada função *Kernel* em cada par de câmbio analisado), os quais serão aplicados no último conjunto de dados.

3. Conjunto de teste:

Por fim, após obter os parâmetros ótimos de cada um dos 135 modelos considerados, as 135 funções de decisão recebem os dados das variáveis independentes do conjunto de teste, gerando um vetor de valores preditos para a variável dependente (taxa de câmbio). Esse vetor será então comparado com o vetor de valores preditos do modelo *benchmark*, mediante as métricas de erro RMSE e MAE. Por fim, realizar-se-á o teste de significância *Reality Check* de White (2000) para verificar se os modelos SVR fornecem incrementos significativos de poder explicativo em relação ao *Random Walk*.

²É possível notar que a métrica de erro implícita na função de perda ϵ -insensitiva é o erro absoluto médio (MAE), a qual será utilizada para a avaliação das previsões no conjunto de teste.

A execução computacional se deu com o auxílio dos *softwares R*, versão 3.2.1 e *RStudio*, versão 0.99.467. Para a estimação dos modelos SVR, utilizou-se o pacote *Kernlab*, versão 0.9-24 (KARATZOGLOU et al., 2004).

6 RESULTADOS E DISCUSSÃO

6.1 Treinamento e validação

Para os intervalos de busca listados na tabela 5.1, os parâmetros ótimos obtidos na etapa de treinamento para os 135 modelos analisados e os seus respectivos RMSEs ótimos associados à previsão no conjunto de validação estão listadas nas tabelas 7.1 a 7.15.

Nota-se que, para as bases de dados e para as funções *Kernel* consideradas, os parâmetros ótimos que fornecem o menor RMSE para as previsões no conjunto de validação parecem não seguir algum padrão imediato, ou mesmo apresentar algum intervalo de prevalência; especialmente para os parâmetros globais ϵ e C , não parece haver uma concentração em algum subconjunto bem delimitado dos intervalos de busca adotados. De fato, o estado da arte da literatura científica em aprendizado de máquinas ainda encara a definição precisa do intervalo de busca dos parâmetros de treinamento como uma questão em aberto, de modo que ainda não foi estabelecido um consenso – ou mesmo heurísticas consistentes – para a escolha da melhor faixa de busca dos parâmetros para o *grid search*, assim como não há um consenso a respeito da melhor função *Kernel* a ser utilizada; por parcimônia, o presente estudo optou por testar várias funções *Kernel* e verificar a significância individual de cada modelo pelo teste de White.

6.2 Avaliação das previsões no conjunto de teste

Os parâmetros ótimos obtidos nas etapas de treinamento e validação foram aplicados para cada uma das 9 máquinas de cada um dos 15 pares de câmbio analisados, e a previsão foi comparada com o *benchmark*. Para essa etapa, foram consideradas as métricas de erro RMSE (raiz quadrada do erro quadrático médio) e MAE (erro absoluto médio) das previsões em relação aos valores observados para os respectivos períodos temporais.

Considerou-se ambas as métricas de erro RMSE e MAE porquanto são os indicadores de

erro mais recorrentes na literatura científica acerca de modelos de previsão por aprendizado de máquinas. Ademais, conforme abordado anteriormente, a construção das duas funções assumem escalas distintas de penalização em relação ao desvio entre os valores previsto e observado, de modo que se espera enriquecer a análise dos resultados incorporando ambas as medidas, cujas expressões são definidas como abaixo:

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (s_t - \hat{s}_t)^2}{N}}$$

$$MAE = \frac{\sum_{t=1}^N |s_t - \hat{s}_t|}{N}$$

Além disso, a taxa de acerto direcional em relação ao período imediatamente anterior também foi considerada – isto é, contabilizou-se como “acerto direcional” caso o sentido de variação do valor predito do período t em relação ao valor observado do período $t - 1$ tenha sido igual ao sentido de variação real do câmbio de $t - 1$ para t ; após avaliar o acerto direcional em todos os períodos do conjunto de teste, definiu-se como “taxa de acerto direcional” a proporção (por contagem simples) de acertos direcionais em relação ao total de períodos testados. Ao se considerar a estratégia de um agente especulador no mercado FOREX como “comprar barato, vender caro”, lograr êxito em antecipar o sentido da variação de um determinado par de câmbio confere a esse agente uma grande vantagem em relação ao restante do mercado, na medida em que pode ajustar suas operações de acordo com o sentido de oscilação favorável – por exemplo, caso se consiga prever que a moeda **A** irá se valorizar em relação à moeda **B**, é de se esperar que um agente especulador possa auferir ganhos comprando quantias da moeda **A** para vendê-las quando seu valor aumentar, mesmo que a magnitude da oscilação não seja capturada com grande precisão¹.

Os resultados para os conjuntos de teste dos 15 pares de câmbio analisados estão listados nas tabelas **7.16** a **7.30**.

Os resultados obtidos mostraram-se ser bastante favoráveis em favor dos modelos SVR, os quais em geral apresentaram taxa de acerto direcional superior e métricas de erro (tanto RMSE quanto MAE) inferiores ao *benchmark Random Walk*. Mais especificamente, é possível notar que os resultados foram significativamente divergentes entre as 9 funções *Kernel* consideradas; em especial, para uma mesma função *Kernel* também observou-se qualidade de previsão dis-

¹Considerando um cenário mais verossímil em que há custos de transação e restrições de operações – por exemplo, a impossibilidade de venda a descoberto – a magnitude do acerto teria grande relevância. Para efeitos de simplificação, este cenário mais sofisticado não foi incorporado na análise do presente estudo.

tinta para diferentes pares de câmbio analisados, reforçando a constatação da literatura de que a escolha da função *Kernel* interfere decisivamente no poder preditivo dos modelos.

Mais especificamente, observa-se que para alguns dos 135 modelos testados o *benchmark* não foi superado, tanto em relação ao acerto direcional quanto ao valor das métricas de erro – inclusive foram observados modelos SVR que apresentaram taxa de acerto melhores que o *benchmark*, porém métricas de erro piores; por outro lado, todos os modelos SVR que apresentaram métricas de erro inferiores ao *benchmark* também obtiveram taxa de acerto direcional superior. Em relação às métricas de erro RMSE e MAE, em todos os 135 modelos testados ambas as métricas estiveram na mesma direção relativa às respectivas métricas apresentadas pelo *benchmark* – isto é, para todos os 135 modelos, se o RMSE do SVR foi maior (menor) que o RMSE do *Random Walk*, o MAE do SVR também foi maior (menor) que o MAE do *Random Walk*.

Em relação às funções *Kernel* adotadas, observou-se que os *Kernels* Gaussiano, Log, Cauchy e Spline Estável apresentaram taxas de acerto direcional superiores ao *Random Walk* em todos os 15 pares de câmbio, sendo que os *Kernels* Log, Cauchy e Spline Estável também apresentaram RMSE e MAE inferiores aos do *Random Walk* em todos os câmbios estudados. Em contrapartida, os *Kernels* polinomiais de graus 1, 2, 3 e 4 apresentaram resultados menos favoráveis em relação ao *benchmark*, o que pode sugerir a existência de interações não lineares que tendem ao infinito, dado que o espaço de característica dos *Kernels* polinomiais possuem dimensão finita e não obtiveram resultados consistentemente melhores; mesmo o *Kernel* multiquádrico inverso, cujo ϕ implícito leva para um espaço vetorial de dimensão infinita, não apresentou resultados favoráveis em todas as ocorrências.

6.3 **Reality Check Test de White**

Apesar dos resultados aparentemente favoráveis verificados na seção anterior, autores como Moosa e Burns (2014) apontam que métricas de erro – como o RMSE e o MAE – falham em oferecer evidências estatísticas mais robustas para garantir que modelos de predição da taxa de câmbio ofereçam de fato poder explicativo superior em relação ao *Random Walk*; em especial, Moosa e Burns (2014) focam na verificação do poder preditivo em modelos que alegam ter superado o *Random Walk* e que utilizaram da métrica RMSE para a avaliação da acurácia, concluindo que os resultados desejáveis são em geral decorrentes da escolha das variáveis explicativas ou da incorporação de parâmetros variantes ao longo do tempo, e não da metodologia empregada para a análise, de modo que o simples fato de as estimativas for-

necerem RMSE menor que o *Random Walk* não garante que este último foi consistentemente superado.

Dessa forma, o presente estudo avalia os resultados das previsões utilizando o *Reality Check Test*, proposto por White (2000), um teste estatístico que verifica a existência do fenômeno denominado ***data snooping***, o qual diz respeito à constatação de resultados preditivos favoráveis por mero acaso, em vez de ser consequência da qualidade do modelo proposto. Dessa forma, o teste de White (2000) tem por objetivo verificar se o desempenho preditivo de um modelo é “suficientemente bom” para poder ser de fato ser avaliado como “melhor” em termos de significância estatística em relação a um *benchmark* estabelecido. Aplicações do teste de White para diagnosticar o *data snooping* no contexto de finanças incluem o trabalho de González-Rivera, Lee e Mishra (2004), o qual avalia previsões de volatilidade baseados em várias variantes do modelo GARCH e quatro métricas distintas de funções de perda; e o trabalho de Yen e Hsu (2010), que verifica o desempenho preditivo de cinco indicadores de análise técnica no mercado futuro de índices financeiros e *commodities*.

Mais especificamente para a previsão da taxa de câmbio, Qi e Wu (2006) testam 2127 indicadores de análise técnica para a previsão do câmbio de sete moedas (dólar canadense, marco alemão, libra esterlina, franco francês, lira italiana, iene japonês e franco suíço) em relação ao dólar norte-americano, baseado em observações diárias entre 1973 e 1998. O estudo constata a existência de modelos que apresentam métricas de erro desejáveis em relação ao *Random Walk*, mas que falham em apresentar significância estatística de que são de fato “suficientemente melhores” a ponto de se poder concluir consistentemente que esse modelos são “superiores”. Adicionando esse teste de hipótese à avaliação das previsões, o presente estudo pretende avaliar os modelos levantados com um nível de exigência mais rigoroso e identificar os modelos que agregam valor explicativo de forma consistente em relação ao *Random Walk*.

O *Reality Check Test* de White compara uma série de S estratégias preditivas em relação a um *benchmark*, comparando os modelos mediante uma métrica pré-estabelecida – nesse caso, a diferença entre os retornos em excesso. Ou seja, o teste de White irá verificar se há evidências estatísticas para afirmar se, dentre as S estratégias consideradas, há pelo menos uma cujo retorno supera consistentemente o retorno do modelo *benchmark*. Assim, a hipótese nula a ser testada afirma que o melhor dos S modelos considerados não é estatisticamente superior ao *benchmark*, de modo que:

$$H_0 : \max_{k=1,2,\dots,S} \{\mathbb{E}(RE_k)\}$$

onde $\mathbb{E}(\cdot)$ é o valor esperado computado pela média aritmética $\mathbb{E}(RE_k) = \frac{1}{N-1} \sum_{t=2}^N RE_{k,t}$, onde N é o número de observações do conjunto de teste² = 58.

Define-se o retorno em excesso para cada estratégia preditiva $k = 1, 2, \dots, S$ pela seguinte expressão:

$$RE_{k,t} = R_{k,t} - R_{benchmark,t}, \quad k = 1, 2, \dots, S$$

Onde o retorno para cada estratégia preditiva $k = 1, 2, \dots, S$ é dado por:

$$R_{k,t} = \log\left(\frac{\hat{s}_t}{s_{t-1}}\right) \cdot I_{k,t-1}$$

onde \hat{s}_t é a taxa de câmbio predita para o período t , s_{t-1} é a taxa de câmbio no período $t-1$, $\log(\cdot)$ é o logaritmo natural e I_{t-1} é uma variável indicadora que indica o sentido da transação no instante t com base nas informações em $t-1$ – ou seja, caso a previsão no período $t-1$ seja que $\hat{s}_t > s_{t-1}$, o agente optaria por entrar em uma posição de compra em $t-1$ em relação a esse câmbio, pois poderia revendê-lo em t por um valor superior; analogamente, caso a previsão aponte que $\hat{s}_t < s_{t-1}$, o agente optaria por entrar em uma posição de venda em $t-1$. Dessa forma, I_{t-1} é uma variável *dummy* que pode assumir 2 valores distintos: sem perda de generalidade, $I_{t-1} = +1$ numa posição de compra (*long position*) e $I_{t-1} = -1$ numa posição de venda (*short position*).

Pela construção do modelo *benchmark*, em cada instante t , \hat{s}_t é igual a s_{t-1} , dado que a hipótese do *Random Walk* implica que o ganho esperado entre o período t e $t+1$ deve ser igual a zero. Dessa forma, o retorno esperado do *benchmark* é dado por:

$$\mathbb{E}(R_{benchmark,t}) = \log\left(\frac{\hat{s}_t}{s_{t-1}}\right) \cdot I_{t-1} = \log\left(\frac{s_{t-1}}{s_{t-1}}\right) \cdot I_{t-1} = \log(1) \cdot I_{t-1} = 0$$

Dessa forma, $\mathbb{E}(RE_k) = \mathbb{E}(R_{k,t}) - \mathbb{E}(R_{benchmark,t}) = \mathbb{E}(R_{k,t}) - 0 = \mathbb{E}(R_{k,t})$, $k = 1, 2, \dots, S$.

Para testar a hipótese nula H_0 , White (2000) indica o método do *bootstrap* estacionário, que consiste em tomar B reamostragens para os valores observados de $R_{k,t}$, a fim de diluir os efeitos de dependência temporal dos dados financeiros e evitar o “efeito contágio” de ciclos econômicos, motivação análoga ao sorteio aleatório dos períodos que compõem os conjuntos de treinamento, validação e teste. Os passos para o *bootstrap* estacionário podem ser sintetizados a seguir:

²Perde-se uma observação em decorrência da defasagem, visto que s_{t-1} não está definido para $t = 1$. Assim, RE_k será um vetor com $58 - 1 = 57$ elementos.

1. Calcular a estatística

$$\bar{V} = \max_{k=1,2,\dots,S} \{\sqrt{N-1} \cdot \mathbb{E}(RE_k)\}$$

com base nos vetores observados $R_{k,t}$;

2. Para cada $k = 1, 2, \dots, S$, com base na série de $R_{k,t}$, sortear aleatoriamente e com reposição elementos quaisquer pertencentes a $R_{k,t}$ até formar um vetor $R_{k,t}^*$ com dimensão igual a $R_{k,t}$ (nesse caso, um vetor com 57 elementos)³. O mesmo processo se repete B vezes;
3. Para cada uma das B reamostragens, calcular a média dos retornos do *bootstrap* $R_{k,t}^*$, dado por

$$\bar{R}_{k,t}^* = \frac{1}{N-1} \sum_{t=2}^N R_{k,t}^*;$$

4. Para cada reamostragem $i = 1, 2, \dots, B$, calcular as estatísticas

$$\bar{V}_i^* = \max_{k=1,2,\dots,S} \{\sqrt{N-1} \cdot [\bar{R}_{k,t}^* - \mathbb{E}(RE_k)]\};$$

5. Após reamostrar B vezes cada um dos S modelos propostos, compara-se \bar{V} com os percentis de \bar{V}_i^* de modo a obter o p-valor do teste: valores pequenos do p-valor implicam em rejeição da hipótese nula. O p-valor do teste pode ser obtido fazendo-se uma ordenação dos valores de \bar{V}_i^* e computar a proporção de valores superiores aos valores de \bar{V} , de modo que o p-valor é dado por

$$p = \frac{\text{número de valores em } \bar{V}_i^* \text{ maiores que } \bar{V}}{B}$$

No presente estudo, considerou-se três níveis de significância: 10%, 5% e 1%.

A rejeição da hipótese nula do teste de White implica que o melhor modelo dentre os S modelos considerados apresenta desempenho superior ao *benchmark*, pois equivale a dizer que há uma probabilidade pequena de se obter retornos em excesso tão “extremos” quanto os observados caso a hipótese nula de *Random Walk* seja válida. Porém, o teste não aponta qual dos S modelos é o melhor, tampouco apresenta conclusões em relação aos $S - 1$ modelos remanescentes. Desse modo, no presente estudo conduziu-se o teste para $S = 1$, testando individualmente cada um dos 9 modelos propostos (em vez de aninhar todos os 9) e comparando-os diretamente com o *benchmark* de cada par de câmbio, repetindo o processo para todos os 15

³O *bootstrap* estacionário exige a definição de um “parâmetro de suavização” Q que determina o tamanho dos “blocos” de observações consecutivas – isto é, frações da série original que preservam parcialmente a estrutura temporal na série reamostrada. Qi e Wu (2006) definem $Q = 0.5$ e apontam que a alteração desse parâmetro não acarreta alterações significativas nos resultados. Dessa forma, o presente estudo também adotará $Q = 0.5$.

pares, totalizando 135 testes de hipótese. A escolha de B seguiu o estudo de Qi e Wu (2006), os quais realizam aplicação similar para a avaliação de previsões de câmbio e apontam que $B = 500$ é suficiente para garantir ambas significância estatística e eficiência computacional. Dessa forma, adotou-se também $B = 500$. Os p-valores relativos aos 135 modelos considerados estão dispostos nas tabelas **7.31** a **7.33**.

Os p-valores do teste de White revelam que há uma quantidade considerável de modelos SVR que, mesmo obtendo taxa de acerto direcional superior e métricas de erros inferior ao *benchmark Random Walk*, não apresentaram evidências estatísticas de que são “suficientemente bons” a ponto de serem considerados de fato “melhores” do que o *Random Walk*, de modo que o desempenho favorável desses modelos SVR não pode ser creditado como mérito da metodologia em prover maior poder explicativo; mas, em vez disso, esse desempenho superior pode ser resultado do mero acaso.

É interessante notar que, para todos os 15 pares de câmbio analisados, pelo menos uma função *Kernel* apresentou significância estatística a 5% pelo teste de White. Ou seja: para os câmbios estudados, em 95% das vezes em que o teste for replicado, espera-se obter evidências de que pelo menos um dos *Kernels* adotados conseguiu de fato gerar retornos que superassem consistentemente o modelo *benchmark*. Também é notável a quantidade razoável de *Kernels* que conseguiram rejeitar a hipótese nula do teste de White a uma significância de 1%, o que fornece evidências robustas de que aqueles modelos de fato performaram melhor que o *Random Walk*. Por outro lado, é importante ressaltar a quantidade de modelos que falharam em rejeitar a hipótese nula do teste: dos 135 modelos testados, 64 não apresentaram significância mesmo a um nível de confiança mais brando de 90%; considerando a confiança usual de 95%, o número de modelos que falham em rejeitar a hipótese nula do teste de White sobe para 82 de 135, pouco mais de 60% dos modelos considerados.

Os resultados do teste de White também são consistentes com a literatura de aprendizado de máquinas no que se refere ao fato de não haver um “*Kernel* ótimo” que sempre seja superior em termos de acurácia preditiva. A função *Kernel* que obteve melhor significância foi o *Kernel Spline Estável*, cujo maior p-valor entre os 15 pares de câmbio foi 5.4% (para os outros 14 pares de câmbio, essa função apresentou significância de pelo menos 5%); porém, esse *Kernel* não foi o melhor em todos os 15 casos, sugerindo que, a depender da base de dados considerada, ou por sensibilidade dos parâmetros de *input* das previsões no conjunto de teste, o “melhor *Kernel*” tende a alterar-se. Os *Kernels* Log e Cauchy também apresentaram desempenho satisfatório, com várias ocorrências de significância a 1% para o teste de White, possivelmente em decorrência do fato de esses *Kernels* induzirem espaços de característica em dimensões

infinitas. No caso do *Kernel* Cauchy, o fato desta advir de uma distribuição de probabilidade com cauda pesada provavelmente tenha exercido uma influência positiva para a qualidade das previsões, dado que coaduna com um dos principais fatos estilizados do mercado FOREX.

O *Kernel* Gaussiano, o mais utilizado pela literatura de aprendizado de máquinas, apresentou um desempenho *overall* satisfatório, porém em 3 dos 15 pares de câmbio não obteve significância a 10% no teste de White, sugerindo que essa função, mesmo mapeando os dados em uma dimensão infinita, pode apresentar limitações em determinados conjuntos de dados; em comparação com o *Kernel* Cauchy, o Gaussiano obteve resultados ligeiramente inferiores, o que pode ser devido ao comportamento de cauda do *Kernel* Gaussiano (que se assemelha a uma distribuição Normal) convergir mais rapidamente a zero.

O *Kernel* multiquádrico inverso, que obteve melhor desempenho no estudo de Henrique et al. (2016), não obteve resultados de destaque no presente estudo; na verdade, chama a atenção o fato de que em 7 dos 15 pares de câmbio considerados, essa função – que também mapeia os dados em dimensão infinita – não apresentou significância mesmo a 10% para o teste de White, apresentando inclusive alguns p-valores bastante elevados, mais uma vez sugerindo que a “melhor escolha” para a função *Kernel* não parece ter um critério bem estabelecido, além de evidenciar que uma mesma função *Kernel* pode apresentar desempenho bastante distinto ao ser aplicada em contextos distintos, uma vez que o *Kernel* multiquádrico inverso, que obteve os melhores resultados para a formação de portfólios, não repetiu o bom resultado no contexto da predição do câmbio.

Por fim, os quatro *Kernels* polinomiais (linear, quadrático, cúbico e quártico) não apresentaram desempenho satisfatório nas previsões realizadas; em especial, os *Kernels* polinomiais de graus 1, 2 e 3 não obtiveram significância a 10% no teste de White para todos os pares de câmbio exceto USD-BRL, no qual todos os 9 *Kernels* obtiveram êxito em rejeitar a hipótese nula do teste. O *Kernel* polinomial de grau 4 obteve resultado ligeiramente superior, porém ainda bastante aquém do desempenho apresentado pelos demais *Kernels*. Esse resultado vai de encontro ao *mainstream* observado na literatura, uma vez que grande parte dos estudos em aprendizado de máquinas se concentra fortemente nos *Kernels* polinomial e Gaussiano; com os *Kernels* polinomiais demonstrando pobre poder explicativo para as previsões da taxa de câmbio, o presente estudo fornece um possível indício de que as interações subjacentes no mercado cambial extrapolam espaços de característica com dimensão finita, de modo que a incorporação de funções *Kernel* que abarcam extensões a *feature spaces* de dimensão infinita pode representar um incremento significativo na qualidade das previsões para essa área.

7 CONCLUSÃO E RECOMENDAÇÕES

O presente estudo avaliou o poder preditivo de modelos SVR para 15 taxas de câmbio utilizando variáveis macroeconômicas utilizadas na literatura científica; ademais, o estudo verificou a acurácia das previsões de 9 funções *Kernel*, totalizando 135 modelos. Constatou-se que a introdução de modelos SVR logrou êxito em fornecer previsões com maior precisão e menor margem de erro, devido à incorporação de interações não-lineares entre as variáveis explicativas. Nesse sentido, o presente estudo apresenta a contribuição de fornecer evidências adicionais de que a introdução de formas funcionais não-lineares para as previsões da taxa de câmbio é um fator de grande relevância para a elevação do poder explicativo, preenchendo uma lacuna ainda existente no estado da arte dessa agenda de pesquisa, dado que a maioria dos estudos recentes em predição do câmbio ainda considera modelos lineares.

Ademais, o estudo verificou que a escolha da função *Kernel* de fato possui grande importância, uma vez que, dado um conjunto de dados, funções *Kernel* distintos acarretam impactos diretos na qualidade das previsões geradas pelo algoritmo de aprendizado de máquinas. Por outro lado, uma mesma função *Kernel* pode apresentar desempenho notavelmente heterogêneo, a depender do conjunto de dados em que se aplica o algoritmo de aprendizagem e da natureza da temática abordada. Foi apresentada uma breve descrição e análise do espaço de características induzidos pelas funções *Kernel*, e de que forma a incorporação dessa função é capaz de generalizar interações não-lineares em dimensões elevadas, ou até mesmo infinitas. Identificou-se também que as funções *Kernel* mais utilizadas na literatura acadêmica correlata – *Kernels* polinômiais e Gaussiano – não obtiveram resultado de destaque nos testes realizados; em especial, as previsões geradas pelos *Kernels* polinômiais de graus 1, 2, 3 e 4 não obtiveram sucesso em superar o *Random Walk*. Essa observação pode contribuir para o desenvolvimento de trabalhos futuros na área, dado que no presente estudo os *Kernels* que forneceram as melhores previsões – em termos de elevado acerto direcional, reduzido valor nas métricas de erro e nos p-valores do teste de White – foram os *Kernels* Log, Cauchy e Spline Estável, todos de aplicação inédita especificamente para previsão da taxa de câmbio.

Outra contribuição do estudo reside no uso de um teste estatístico mais robusto – *Reality*

Check Test de White (2000) – para averiguar a qualidade das previsões, em vez de recorrer meramente a métodos “*soft*” de avaliação do erro – como o RMSE e o MAE – um fenômeno comum na literatura atual, dado que essas métricas podem carregar um componente do acaso e mascarar um possível *data snooping*, em vez de avaliar de fato a qualidade da metodologia proposta. Enquanto a aplicação desse teste não é inédita em trabalhos de previsão do câmbio, representa um aditivo em relação ao *mainstream* observado no estado da arte, e possui o potencial de ajudar a eliminar modelos defeituosos que possam ter obtido bons resultados por circunstâncias específicas de recorte temporal ou natureza da base de dados testada.

Enquanto o presente estudo observou uma quantidade razoável de modelos SVR que obtiveram boa significância no teste de White, é precipitado concluir que esses modelos são **de fato** superiores a um *Random Walk*, na medida em que os resultados obtidos estão condicionados à amostra selecionada, ao período temporal considerado e às variáveis explicativas selecionadas. Dessa forma, não é prudente supor que os resultados obtidos são suficientes para se refutar categoricamente o *Meese-Rogoff puzzle*.

Mais especificamente, o estudo realizou um breve mapeamento da produção científica recente na área de previsão de taxa de câmbio, com ênfase em estudos que fizeram aplicação de métodos de aprendizado de máquina e inteligência artificial, especialmente modelos SVM e suas extensões, como o SVR. Ademais, com base nos artigos analisados, foi apresentado um modelo fundamentalista composto por 13 variáveis explicativas embasadas na literatura para servirem como preditores da taxa de câmbio *spot*. Apesar de não possuir vocação de ser exaustiva, espera-se que esse modelo possa ser de interesse de acadêmicos e participantes do mercado cambial, de modo a subsidiar estudos futuros nessa temática, à luz dos resultados em geral satisfatórios que esse modelo atingiu na presente aplicação.

7.1 Limitações e desenvolvimentos futuros

O presente estudo apresenta diversas limitações, a serem analisadas e transpostas em estudos futuros. A mais proeminente reside na periodicidade mensal adotada para os dados, um cenário de baixa verossimilhança com as características do mercado FOREX, o qual possui grandes liquidez e volume de transações e se notabiliza pela alta frequência das operações executadas e elevado número de participantes. Ao tomar como base variáveis explicativas fundamentalistas, o estudo se depara com o entrave da periodicidade dos dados, na medida em que, por mais que haja disponibilidade de dados de alta frequência para variáveis financeiras como preços de ações, taxas de câmbio e índices de bolsas de valores, a divulgação

dos fundamentos macroeconômicos tende a ter periodicidade menor – algumas variáveis inclusive são atualizadas apenas trimestralmente, como o PIB, que acabou sendo desconsiderado no presente estudo face a essa limitação. Constatou-se que a literatura especializada em modelos fundamentalistas tem sistematicamente realizado previsões para os horizontes de baixa frequência mensal e trimestral, evidenciando que, muito embora retrate um cenário pouco aderente com a realidade, modelos que consideram previsões de baixa frequência possuem validade acadêmica e rigor científico, e podem contribuir para a construção do conhecimento nessa área temática.

A maioria dos trabalhos que consideram mais especificamente a análise técnica realizam previsões do câmbio em frequência diária, com alguns autores inclusive consideram frequências de horas ou mesmo minutos (DIEBOLD; HAHN; TAY, 1999; SANTOS; COSTA; COELHO, 2007; RIME DAGFINN; SOJLI, 2010). Em se tratando de modelos fundamentalistas, esse entrave é mais difícil de ser rompido, porém pode se recorrer ao uso de interpolações para a obtenção de dados em periodicidade maior, ou a definição de *proxies* com disponibilidade de dados em alta frequência para variáveis de baixa frequência.

A escolha das variáveis explicativas também possui diversas imperfeições em potencial, muito em medida da limitação da periodicidade supracitada. Adicionalmente, algumas variáveis propostas na literatura foram expurgadas, de modo que a verificação do desempenho de modelos futuros que incorporem algumas dessas variáveis expurgadas é desejável e pertinente. Outras recomendações consistem em incorporar também indicadores de análise técnica para figurarem como variáveis explicativas, ou mesmo incorporar todas as variáveis observadas na literatura, e posteriormente aplicar técnicas de redução de dados, como análise fatorial ou análise de componentes principais. No presente estudo, tais técnicas não se fizeram necessárias, dado que considerou-se que o número de preditores (13) implicava em uma dimensionalidade de manipulação conveniente. Em estudos que consideram o nível corporativo, por contrapartida, frequentemente são consideradas várias dezenas de indicadores, muitos dos quais podem possuir redundância conceitual, de modo que a redução dos dados tornar-se-ia bastante atrativa.

Custos de transação foram desconsiderados no presente trabalho, dado que não foram identificados na literatura estudos que incorporassem esse aspecto na previsão da taxa de câmbio. Sua introdução poderia ser relevante no sentido de identificar modelos e estratégias que, mesmo fornecendo retornos consistentemente maiores que zero, podem não ser “bons o bastante” para compensar a concretização de uma operação no mercado FOREX, caso se considerem custos de transação. Com essa restrição adicional, *benchmarks* alternativos poderiam

ser considerados para o teste da acurácia e do desempenho das previsões geradas pelo SVR.

A boa capacidade preditiva alcançada por algumas das funções *Kernel* corrobora o achado na literatura da existência de padrões identificáveis subjacentes às séries temporais financeiras, como observado em Caraiani e Haven (2015), que aponta evidências de multifractalidade. Com base nessa constatação, funções *Kernel* que sejam capazes de generalizar fractais possuem potencial de incrementar ainda mais o poder explicativo de modelos de previsão ao serem incorporados futuramente em algoritmos de aprendizado de máquinas.

Em relação à metodologia empregada, sugerem-se como desenvolvimentos futuros a incorporação de outros métodos de aprendizado de máquinas – em especial redes neurais artificiais, visto que essa classe de modelos ainda é amplamente utilizada na literatura de predição da taxa de câmbio. Apesar de numerosos estudos terem apontado a superioridade dos modelos SVM e suas extensões em relação às redes neurais, uma comparação direta entre as duas metodologias no contexto de taxas de câmbio pode ajudar a fornecer *insights* e abrir novas possibilidades de pesquisa nessa área. Outras metodologias de aprendizado de máquinas como algoritmos genéticos (BRITO; OLIVEIRA, 2012; FALAT; PANCIKOVA, 2015) e modelos *deep learning* (SHEN; CHAO; ZHAO, 2015) também possuem grande destaque na produção recente em aprendizado de máquinas e a incorporação desses métodos em pesquisas futuras tendem a enriquecer significativamente a análise apresentada pelo presente estudo.

Ademais, outras versões do SVR poderiam ser incorporadas: o presente estudo fez uso do ϵ -SVR, a qual adota a função de perda ϵ -insensitiva $L_\epsilon[y_i, f(x_i)]$. O SVR pode ser formulado com diferentes funções perda ou a introdução de parâmetros adicionais, culminando em versões distintas que, ao serem testadas, podem revelar estruturas interessantes específicas ao mercado cambial. Extensões como o ν -SVR (CHANG; LIN, 2002), que introduz um parâmetro adicional de permissividade ao erro, podem ser consideradas em estudos futuros e comparados aos modelos ϵ -SVR. Modelos híbridos que combinam duas ou mais metodologias distintas de aprendizado de máquinas também podem ser levadas em consideração (NI; YIN, 2009; BRITO; OLIVEIRA, 2012; SERMPINIS et al., 2015).

Com respeito à verificação da significância estatística obtida pelos modelos de aprendizado de máquinas em relação ao *benchmark*, a literatura aponta a possibilidade de utilizar diversos outros testes para avaliar a precisão das previsões, como o teste Diebold-Mariano (DIEBOLD; MARIANO, 1995). Especificamente para identificar o fenômeno do *data snooping*, possibilidades alternativas incluem o *Superior Predictive Ability Test* de (HANSEN, 2005) e o “*New Superior Predictive Ability Test*” de (CAI; JIANG; ZHANG, 2013). Mais especificamente, estudos futuros podem testar o efeito que a alteração dos intervalos de busca para os parâmetros ótimos nos

conjuntos de treinamento e validação exerce na qualidade das previsões, bem como a alteração do número de reamostragens consideradas no *bootstrap* estacionário para o teste de White.

Em aspectos mais gerais, uma replicação futura do presente estudo para um horizonte temporal distinto e para amostras maiores, ou para pares de câmbio não considerados no presente trabalho, é desejável, a fim de verificar a significância do ϵ -SVR e dos *Kernels* levantados em conjuntos de dados distintos, o que tende a fornecer evidências adicionais em relação ao poder explicativo que a introdução dessa classe de modelos agrega em relação a metodologias tradicionais, como a análise econométrica de séries temporais ou de dados em painel.

Por fim, dado que a introdução de funções *Kernel* até então inéditas no contexto específico de previsão de taxas de câmbio trouxe resultados desejáveis, uma possibilidade de desenvolvimento dessa linha de pesquisa é a criação de novas funções *Kernel* que satisfaçam a condição de Mercer – e que portanto poderão ser utilizadas no algoritmo do SVR via *Kernel trick*. A proposta de Sun e Li (2012) de ponderação de *Kernels* pode ser levada em consideração: dado que a combinação linear de *Kernels* Mercer-admissíveis também são Mercer-admissíveis, os *Kernels* utilizados no presente estudo – ou apenas os que apresentaram poder preditivo elevado – podem ser combinados para gerar “*Kernels* híbridos”, cujos pesos ótimos podem ser obtidos via *grid search*.

A questão de criação de *Kernels* é também abordada por Daoud e Turabieh (2013), que propõe a criação de novas funções *Kernels* com base no estudo de espaços de matrizes positiva semi-definidas. Seguindo proposta similar Simiński (2014) apresenta um *neuro-fuzzy system* capaz de gerar funções *Kernel* Mercer-admissíveis capaz de realizar uma aproximação adaptativa da “melhor” função *Kernel* para o conjunto de dados considerado. Aiolli e Donini (2015), por sua vez, propõe um algoritmo de *Multiple Kernel Learning*, o qual consiste em combinar funções *Kernel* Mercer-admissíveis e realizar combinações entre elas para gerar *Kernels* novos, almejando a melhoria da acurácia das previsões de uma máquina de aprendizado. Os autores afirmam que o algoritmo proposto é capaz de lidar com centenas de milhares de *Kernels* “fracos” de maneira computacionalmente eficiente para chegar em *Kernels* “fortes” que são combinações dos anteriores. Dado que a escolha ideal da função *Kernel* é uma das principais perguntas em aberto na literatura de aprendizado de máquinas, estudos futuros que focam nessa problemática são bastante pertinentes para o desenvolvimento dessa linha de pesquisa.

ANEXO A: TABELAS

A.1: Parâmetros ótimos para os conjuntos de treinamento e validação

Dólar norte-americano – Euro		
Função <i>Kernel</i>	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.65; C = 7.5; d = 1.4$	0.0063358
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.65; C = 5.5; d = 4.5$	0.0058146
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.30; C = 4.5; d = 4.2$	0.0050819
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.25; C = 6.0; d = 0.4$	0.0041082
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.65; C = 7.0; \sigma = 1.03$	0.0029817
<i>Kernel</i> Log (d)	$\varepsilon = 0.45; C = 3.0; d = 2.2$	0.0034852
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.80; C = 1.5; c = 4.4$	0.0039081
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.40; C = 9.0; \sigma = 0.14$	0.0020613
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.10; C = 8.0; c = 4.0; \lambda = 0.1$	0.0022854

Tabela 7.1: Parâmetros ótimos para o câmbio USD-EUR após treinamento e validação

Dólar norte-americano – Libra esterlina		
Função <i>Kernel</i>	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.75; C = 0.5; d = 0.6$	0.0067641
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.20; C = 10.0; d = 3.9$	0.0044674
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.70; C = 8.0; d = 4.8$	0.0049352
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.60; C = 7.5; d = 2.5$	0.0041198
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.75; C = 1.5; \sigma = 0.26$	0.0037862
<i>Kernel</i> Log (d)	$\varepsilon = 0.75; C = 10.0; d = 0.7$	0.0038905
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.75; C = 8.0; c = 1.8$	0.0024469
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.15; C = 9.5; \sigma = 0.02$	0.0020256
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.20; C = 9.0; c = 3.5; \lambda = 0$	0.0017943

Tabela 7.2: Parâmetros ótimos para o câmbio USD-GBP após treinamento e validação

Dólar norte-americano – Iene japonês		
Função Kernel	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.85; C = 9.5; d = 4.5$	0.3717248
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.65; C = 9.5; d = 2.7$	0.4142056
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.45; C = 3.0; d = 4.7$	0.2936487
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.10; C = 1.0; d = 0.9$	0.2450781
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.05; C = 10.0; \sigma = 1.97$	0.1206370
<i>Kernel</i> Log (d)	$\varepsilon = 0.05; C = 2.0; d = 0.3$	0.0956891
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.75; C = 1.0; c = 3.6$	0.2785024
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.05; C = 9.0; \sigma = 0.34$	0.1296306
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.20; C = 8.5; c = 1.5; \lambda = 0.4$	0.1834687

Tabela 7.3: Parâmetros ótimos para o câmbio USD-JPY após treinamento e validação

Dólar norte-americano – Renminbi chinês		
Função Kernel	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.45; C = 9.5; d = 0.6$	0.0048506
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.30; C = 5.0; d = 0.8$	0.0048497
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.30; C = 0.5; d = 3.5$	0.0041498
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.10; C = 8.0; d = 5.0$	0.0045237
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.47; C = 1; \sigma = 0.01$	0.0028659
<i>Kernel</i> Log (d)	$\varepsilon = 0.60; C = 10.0; d = 0.5$	0.0027073
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.30; C = 1.0; c = 3.2$	0.0036510
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.55; C = 1.5; \sigma = 0.02$	0.0028556
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.40; C = 4.0; c = 2.0; \lambda = 0.9$	0.0021507

Tabela 7.4: Parâmetros ótimos para o câmbio USD-CNY após treinamento e validação

Dólar norte-americano – Real brasileiro		
Função Kernel	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.85; C = 6.5; d = 3.0$	0.0056456
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.65; C = 6.5; d = 4.4$	0.0057056
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.20; C = 9.5; d = 4.7$	0.0055017
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.25; C = 9.5; d = 2.3$	0.0051473
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.80; C = 3.5; \sigma = 0.26$	0.0039018
<i>Kernel</i> Log (d)	$\varepsilon = 0.40; C = 7.5; d = 2.0$	0.0026804
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.80; C = 5.5; c = 3.7$	0.0040863
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.55; C = 7.0; \sigma = 0.04$	0.0029745
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.25; C = 2.5; c = 4.5; \lambda = 0.2$	0.0032183

Tabela 7.5: Parâmetros ótimos para o câmbio USD-BRL após treinamento e validação

Euro – Libra esterlina		
Função Kernel	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.60; C = 2.0; d = 4.6$	0.0022682
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.65; C = 3.5; d = 4.2$	0.0008684
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.10; C = 4.5; d = 3.5$	0.0011650
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.40; C = 9.0; d = 1.1$	0.0008261
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.70; C = 1.5; \sigma = 0.08$	0.0006765
<i>Kernel</i> Log (d)	$\varepsilon = 0.05; C = 9.0; d = 0.6$	0.0004751
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.75; C = 2.0; c = 2.5$	0.0009672
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.30; C = 8.0; \sigma = 0.05$	0.0003394
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.60; C = 9.5; c = 2.0; \lambda = 0.2$	0.0003052

Tabela 7.6: Parâmetros ótimos para o câmbio EUR-GBP após treinamento e validação

Euro – Iene japonês		
Função Kernel	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.35; C = 6.0; d = 1.4$	0.9257941
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.45; C = 8.5; d = 4.2$	0.5961265
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.80; C = 1.5; d = 4.9$	0.6142661
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.75; C = 7.5; d = 1.3$	0.5622811
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.60; C = 3.0; \sigma = 1.83$	0.3379625
<i>Kernel</i> Log (d)	$\varepsilon = 0.10; C = 6.0; d = 1.6$	0.4359203
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.55; C = 6.0; c = 0.8$	0.5267802
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.25; C = 7.5; \sigma = 0.34$	0.2638739
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.15; C = 7.5; c = 0.0; \lambda = 0.3$	0.2912408

Tabela 7.7: Parâmetros ótimos para o câmbio EUR-JPY após treinamento e validação

Euro – Renminbi chinês		
Função Kernel	Parâmetros ótimos	RMSE ótimo
Kernel polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.45; C = 9.0; d = 1.9$	0.0240376
Kernel polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.20; C = 3.5; d = 0.7$	0.0290395
Kernel polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.60; C = 1.5; d = 2.1$	0.0250341
Kernel polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.65; C = 0.5; d = 3.6$	0.0240726
Kernel Gaussiano (σ)	$\varepsilon = 0.60; C = 5.0; \sigma = 1.29$	0.0140421
Kernel Log (d)	$\varepsilon = 0.60; C = 9.5; d = 0.1$	0.0065808
Kernel multiquádrico inverso (c)	$\varepsilon = 0.15; C = 3.5; c = 1.3$	0.0190988
Kernel Cauchy (σ)	$\varepsilon = 0.55; C = 2.0; \sigma = 0.34$	0.0179911
Kernel Spline estável (c, λ)	$\varepsilon = 0.80; C = 4.0; c = 3.0; \lambda = 0.2$	0.0124689

Tabela 7.8: Parâmetros ótimos para o câmbio EUR-CNY após treinamento e validação

Euro – Real brasileiro		
Função Kernel	Parâmetros ótimos	RMSE ótimo
Kernel polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.60; C = 7.5; d = 3.2$	0.0180785
Kernel polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.35; C = 3.5; d = 2.7$	0.0190831
Kernel polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.70; C = 9.5; d = 2.0$	0.0189252
Kernel polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.40; C = 4.0; d = 3.0$	0.0166709
Kernel Gaussiano (σ)	$\varepsilon = 0.65; C = 5.0; \sigma = 1.69$	0.0098531
Kernel Log (d)	$\varepsilon = 0.20; C = 5.0; d = 1.4$	0.0086046
Kernel multiquádrico inverso (c)	$\varepsilon = 0.65; C = 8.5; c = 1.3$	0.0118463
Kernel Cauchy (σ)	$\varepsilon = 0.20; C = 2.0; \sigma = 0.57$	0.0051394
Kernel Spline estável (c, λ)	$\varepsilon = 0.85; C = 4.0; c = 1.0; \lambda = 0$	0.0068610

Tabela 7.9: Parâmetros ótimos para o câmbio EUR-BRL após treinamento e validação

Libra esterlina – Iene japonês		
Função Kernel	Parâmetros ótimos	RMSE ótimo
Kernel polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.50; C = 6.0; d = 1.7$	0.7403743
Kernel polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.65; C = 7.0; d = 3.3$	0.6749107
Kernel polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.40; C = 1.5; d = 0.8$	0.4921966
Kernel polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.10; C = 8.0; d = 2.3$	0.4307182
Kernel Gaussiano (σ)	$\varepsilon = 0.10; C = 8.0; \sigma = 1.72$	0.2285132
Kernel Log (d)	$\varepsilon = 0.05; C = 1.5; d = 0.3$	0.2651681
Kernel multiquádrico inverso (c)	$\varepsilon = 0.10; C = 9.5; c = 2.3$	0.3864509
Kernel Cauchy (σ)	$\varepsilon = 0.05; C = 2.0; \sigma = 0.39$	0.2144541
Kernel Spline estável (c, λ)	$\varepsilon = 0.30; C = 4.0; c = 4.0; \lambda = 0.4$	0.2701830

Tabela 7.10: Parâmetros ótimos para o câmbio GBP-JPY após treinamento e validação

Libra esterlina – Renminbi chinês		
Função Kernel	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 1.00; C = 8.0; d = 4.6$	0.0456466
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.30; C = 6.5; d = 2.0$	0.0392148
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.35; C = 4.5; d = 4.6$	0.0350913
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.30; C = 1.5; d = 0.9$	0.0361291
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.70; C = 6.5; \sigma = 0.18$	0.0247032
<i>Kernel</i> Log (d)	$\varepsilon = 0.05; C = 9.5; d = 0.5$	0.0175669
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.70; C = 4.0; c = 3.1$	0.0267838
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.70; C = 4.5; \sigma = 0.37$	0.0292138
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.45; C = 7.0; c = 2.5; \lambda = 0.8$	0.0146723

Tabela 7.11: Parâmetros ótimos para o câmbio GBP-CNY após treinamento e validação

Libra esterlina – Real brasileiro		
Função Kernel	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.90; C = 1.5; d = 1.3$	0.0203587
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.60; C = 6.0; d = 4.4$	0.0191714
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.20; C = 7.5; d = 2.1$	0.0210085
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.30; C = 2.0; d = 3.9$	0.0167923
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.55; C = 8.0; \sigma = 1.30$	0.0234786
<i>Kernel</i> Log (d)	$\varepsilon = 0.05; C = 10.0; d = 0.7$	0.0139745
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.65; C = 6.0; c = 0.0$	0.0237581
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.35; C = 1.5; \sigma = 0.04$	0.0118620
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.60; C = 1.5; c = 2.0; \lambda = 0.7$	0.0097238

Tabela 7.12: Parâmetros ótimos para o câmbio GBP-BRL após treinamento e validação

Renminbi chinês – lene japonês		
Função Kernel	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 1.00; C = 7.0; d = 1.7$	0.0432831
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.50; C = 7.0; d = 2.7$	0.0473678
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.20; C = 3.0; d = 2.1$	0.0387309
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.10; C = 4.0; d = 0.9$	0.0383520
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.15; C = 9.0; \sigma = 0.88$	0.0292701
<i>Kernel</i> Log (d)	$\varepsilon = 0.45; C = 6.5; d = 1.7$	0.0220003
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.15; C = 4.0; c = 3.0$	0.0318921
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.20; C = 4.0; \sigma = 1.53$	0.0177614
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.20; C = 9.5; c = 3.5; \lambda = 0.1$	0.0191558

Tabela 7.13: Parâmetros ótimos para o câmbio CNY-JPY após treinamento e validação

Real brasileiro – lene japonês		
Função Kernel	Parâmetros ótimos	RMSE ótimo
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.95; C = 8.0; d = 3.8$	0.3280816
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.65; C = 3.5; d = 3.9$	0.3031843
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.40; C = 5.0; d = 4.6$	0.2918964
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.25; C = 1.0; d = 2.1$	0.2211873
<i>Kernel</i> Gaussiano (σ)	$\varepsilon = 0.15; C = 8.0; \sigma = 1.47$	0.2647986
<i>Kernel</i> Log (d)	$\varepsilon = 0.25; C = 8.0; d = 0.2$	0.2074212
<i>Kernel</i> multiquádrico inverso (c)	$\varepsilon = 0.30; C = 9.5; c = 3.3$	0.2599215
<i>Kernel</i> Cauchy (σ)	$\varepsilon = 0.20; C = 5.5; \sigma = 1.04$	0.1792225
<i>Kernel</i> Spline estável (c, λ)	$\varepsilon = 0.05; C = 1.5; c = 1.0; \lambda = 0.3$	0.1866184

Tabela 7.14: Parâmetros ótimos para o câmbio BRL-JPY após treinamento e validação

Real brasileiro – Renminbi chinês		
Função Kernel	Parâmetros ótimos	RMSE ótimo
Kernel polinomial grau 1 (d), $q = 1$	$\varepsilon = 0.90; C = 3.5; d = 1.8$	0.0219874
Kernel polinomial grau 2 (d), $q = 2$	$\varepsilon = 0.65; C = 3.5; d = 2.6$	0.0167363
Kernel polinomial grau 3 (d), $q = 3$	$\varepsilon = 0.30; C = 5.0; d = 4.9$	0.0161124
Kernel polinomial grau 4 (d), $q = 4$	$\varepsilon = 0.20; C = 0.5; d = 0.8$	0.0135723
Kernel Gaussiano (σ)	$\varepsilon = 0.06; C = 61; \sigma = 1.83$	0.0129862
Kernel Log (d)	$\varepsilon = 0.20; C = 8.0; d = 1.6$	0.0086252
Kernel multiquádrico inverso (c)	$\varepsilon = 0.30; C = 6.5; c = 4.2$	0.0130947
Kernel Cauchy (σ)	$\varepsilon = 0.10; C = 5.5; \sigma = 0.61$	0.0106582
Kernel Spline estável (c, λ)	$\varepsilon = 0.70; C = 5.0; c = 3.0; \lambda = 0.2$	0.0078609

Tabela 7.15: Parâmetros ótimos para o câmbio BRL-CNY após treinamento e validação

A.2: Avaliação das previsões nos conjuntos de teste

Dólar norte-americano – Euro			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	54.38%	0.004279	0.002974
Kernel polinomial grau 1 (d), $q = 1$	47.36%	0.006812	0.004318
Kernel polinomial grau 2 (d), $q = 2$	56.14%	0.006102	0.004015
Kernel polinomial grau 3 (d), $q = 3$	63.15%	0.005596	0.003512
Kernel polinomial grau 4 (d), $q = 4$	68.42%	0.004364	0.003029
Kernel Gaussiano (σ)	78.94%	0.003365	0.002496
Kernel Log (d)	73.68%	0.003801	0.002701
Kernel multiquádrico inverso (c)	68.42%	0.004295	0.003001
Kernel Cauchy (σ)	80.70%	0.002591	0.001753
Kernel Spline estável (c, λ)	87.71%	0.002818	0.001937

Tabela 7.16: Avaliação das previsões no conjunto de teste para o câmbio USD-EUR

Dólar norte-americano – Libra esterlina			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	50.87%	0.005477	0.003816
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	63.15%	0.006915	0.004973
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	57.89%	0.005169	0.004067
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	68.42%	0.005528	0.004492
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	73.68%	0.004671	0.003858
<i>Kernel</i> Gaussiano (σ)	82.45%	0.004281	0.002816
<i>Kernel</i> Log (d)	87.71%	0.004916	0.003206
<i>Kernel</i> multiquádrico inverso (c)	68.42%	0.003383	0.002761
<i>Kernel</i> Cauchy (σ)	82.45%	0.002841	0.002183
<i>Kernel</i> Spline estável (c, λ)	91.22%	0.002392	0.001670

Tabela 7.17: Avaliação das previsões no conjunto de teste para o câmbio USD-GBP

Dólar norte-americano – Iene japonês			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	50.87%	0.323567	0.241017
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	54.38%	0.401527	0.291643
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	47.36%	0.448166	0.328711
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	56.14%	0.336172	0.256061
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	61.40%	0.288152	0.200627
<i>Kernel</i> Gaussiano (σ)	70.17%	0.138325	0.089196
<i>Kernel</i> Log (d)	78.94%	0.101482	0.067917
<i>Kernel</i> multiquádrico inverso (c)	71.92%	0.292761	0.203815
<i>Kernel</i> Cauchy (σ)	85.96%	0.168295	0.117674
<i>Kernel</i> Spline estável (c, λ)	82.45%	0.213527	0.149915

Tabela 7.18: Avaliação das previsões no conjunto de teste para o câmbio USD-JPY

Dólar norte-americano – Renminbi chinês			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	54.38%	0.003715	0.001787
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	50.87%	0.005012	0.002847
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	45.61%	0.005187	0.003081
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	59.64%	0.004129	0.002312
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	56.14%	0.004880	0.002716
<i>Kernel</i> Gaussiano (σ)	78.94%	0.003093	0.001409
<i>Kernel</i> Log (d)	77.19%	0.002951	0.001387
<i>Kernel</i> multiquádrico inverso (c)	57.89%	0.003903	0.001946
<i>Kernel</i> Cauchy (σ)	68.42%	0.003266	0.001593
<i>Kernel</i> Spline estável (c, λ)	80.70%	0.002317	0.001295

Tabela 7.19: Avaliação das previsões no conjunto de teste para o câmbio USD-CNY

Dólar norte-americano – Real brasileiro			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	45.61%	0.012856	0.009057
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	61.40%	0.006372	0.004520
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	66.67%	0.006738	0.004804
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	64.91%	0.005864	0.004233
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	71.92%	0.005302	0.003841
<i>Kernel</i> Gaussiano (σ)	71.92%	0.004226	0.003302
<i>Kernel</i> Log (d)	78.94%	0.002922	0.001963
<i>Kernel</i> multiquádrico inverso (c)	73.68%	0.004398	0.003358
<i>Kernel</i> Cauchy (σ)	91.22%	0.003392	0.002198
<i>Kernel</i> Spline estável (c, λ)	85.96%	0.003504	0.002366

Tabela 7.20: Avaliação das previsões no conjunto de teste para o câmbio USD-BRL

Euro – Libra esterlina			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	56.14%	0.001142	0.000868
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	49.12%	0.002545	0.001782
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	61.40%	0.001031	0.000796
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	52.63%	0.001478	0.000918
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	63.15%	0.000996	0.000785
<i>Kernel</i> Gaussiano (σ)	73.68%	0.000719	0.000536
<i>Kernel</i> Log (d)	80.70%	0.000491	0.000401
<i>Kernel</i> multiquádrico inverso (c)	68.42%	0.001090	0.000710
<i>Kernel</i> Cauchy (σ)	84.21%	0.000344	0.000237
<i>Kernel</i> Spline estável (c, λ)	84.21%	0.000317	0.000248

Tabela 7.21: Avaliação das previsões no conjunto de teste para o câmbio EUR-GBP

Euro – Iene japonês			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	54.38%	0.611865	0.333312
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	42.10%	0.957012	0.501266
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	61.40%	0.614935	0.321673
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	54.38%	0.640676	0.350148
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	61.40%	0.589177	0.318729
<i>Kernel</i> Gaussiano (σ)	78.94%	0.361019	0.196037
<i>Kernel</i> Log (d)	73.68%	0.460328	0.247128
<i>Kernel</i> multiquádrico inverso (c)	64.91%	0.558473	0.296759
<i>Kernel</i> Cauchy (σ)	89.47%	0.281645	0.14802
<i>Kernel</i> Spline estável (c, λ)	84.21%	0.303568	0.160629

Tabela 7.22: Avaliação das previsões no conjunto de teste para o câmbio EUR-JPY

Euro – Renminbi chinês			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	50.87%	0.030595	0.021964
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	54.38%	0.028004	0.020426
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	50.87%	0.031674	0.022738
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	54.38%	0.028279	0.019452
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	52.63%	0.027460	0.019034
<i>Kernel</i> Gaussiano (σ)	73.68%	0.017727	0.011682
<i>Kernel</i> Log (d)	77.19%	0.007359	0.005016
<i>Kernel</i> multiquádrico inverso (c)	64.91%	0.021168	0.014832
<i>Kernel</i> Cauchy (σ)	70.17%	0.018362	0.012803
<i>Kernel</i> Spline estável (c, λ)	82.45%	0.013429	0.010259

Tabela 7.23: Avaliação das previsões no conjunto de teste para o câmbio EUR-CNY

Euro – Real brasileiro			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	59.64%	0.016265	0.011835
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	43.85%	0.020399	0.014277
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	50.87%	0.021561	0.014985
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	52.63%	0.021844	0.015057
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	56.14%	0.017402	0.010420
<i>Kernel</i> Gaussiano (σ)	64.91%	0.011184	0.007354
<i>Kernel</i> Log (d)	68.42%	0.009257	0.006810
<i>Kernel</i> multiquádrico inverso (c)	61.40%	0.013061	0.007712
<i>Kernel</i> Cauchy (σ)	85.96%	0.007457	0.004397
<i>Kernel</i> Spline estável (c, λ)	73.68%	0.007905	0.004612

Tabela 7.24: Avaliação das previsões no conjunto de teste para o câmbio EUR-BRL

Libra esterlina – Iene japonês			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	59.64%	0.521000	0.379208
<i>Kernel polinomial grau 1 (d), q = 1</i>	36.84%	0.751874	0.481123
<i>Kernel polinomial grau 2 (d), q = 2</i>	52.63%	0.688157	0.439624
<i>Kernel polinomial grau 3 (d), q = 3</i>	56.14%	0.501271	0.355826
<i>Kernel polinomial grau 4 (d), q = 4</i>	63.15%	0.449371	0.290632
<i>Kernel Gaussiano (σ)</i>	80.70%	0.234276	0.153561
<i>Kernel Log (d)</i>	68.42%	0.273445	0.178425
<i>Kernel multiquádrico inverso (c)</i>	66.67%	0.39256	0.248157
<i>Kernel Cauchy (σ)</i>	89.47%	0.220226	0.161173
<i>Kernel Spline estável (c, λ)</i>	71.92%	0.277297	0.193926

Tabela 7.25: Avaliação das previsões no conjunto de teste para o câmbio GBP-JPY

Libra esterlina – Renminbi chinês			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	57.89%	0.040310	0.029138
<i>Kernel polinomial grau 1 (d), q = 1</i>	52.63%	0.047129	0.032818
<i>Kernel polinomial grau 2 (d), q = 2</i>	57.89%	0.041954	0.030412
<i>Kernel polinomial grau 3 (d), q = 3</i>	61.40%	0.038120	0.028527
<i>Kernel polinomial grau 4 (d), q = 4</i>	61.40%	0.038784	0.028610
<i>Kernel Gaussiano (σ)</i>	73.68%	0.027431	0.019926
<i>Kernel Log (d)</i>	77.19%	0.019304	0.015038
<i>Kernel multiquádrico inverso (c)</i>	70.17%	0.028854	0.021533
<i>Kernel Cauchy (σ)</i>	68.42%	0.031059	0.022573
<i>Kernel Spline estável (c, λ)</i>	82.45%	0.017009	0.013703

Tabela 7.26: Avaliação das previsões no conjunto de teste para o câmbio GBP-CNY

Libra esterlina – Real brasileiro			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	50.87%	0.020960	0.015091
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	50.87%	0.022158	0.016728
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	54.38%	0.020927	0.015258
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	47.36%	0.022718	0.016692
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	59.64%	0.018424	0.014248
<i>Kernel</i> Gaussiano (σ)	54.38%	0.024412	0.017142
<i>Kernel</i> Log (d)	64.91%	0.014578	0.099128
<i>Kernel</i> multiquádrico inverso (c)	43.85%	0.024910	0.017521
<i>Kernel</i> Cauchy (σ)	70.17%	0.013219	0.008731
<i>Kernel</i> Spline estável (c, λ)	71.92%	0.010523	0.074612

Tabela 7.27: Avaliação das previsões no conjunto de teste para o câmbio GBP-BRL

Renminbi chinês – Iene japonês			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	50.87%	0.048338	0.034934
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	50.87%	0.046121	0.032817
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	49.12%	0.049314	0.035781
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	63.15%	0.039850	0.030012
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	63.15%	0.040084	0.029970
<i>Kernel</i> Gaussiano (σ)	71.92%	0.031163	0.020818
<i>Kernel</i> Log (d)	75.43%	0.023871	0.016048
<i>Kernel</i> multiquádrico inverso (c)	66.67%	0.033061	0.024418
<i>Kernel</i> Cauchy (σ)	78.94%	0.019367	0.013809
<i>Kernel</i> Spline estável (c, λ)	75.43%	0.020659	0.014805

Tabela 7.28: Avaliação das previsões no conjunto de teste para o câmbio CNY-JPY

Real brasileiro – Iene japonês			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	45.61%	0.321020	0.199892
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	43.85%	0.339762	0.210578
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	50.87%	0.307182	0.195376
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	54.38%	0.296124	0.192001
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	63.15%	0.239909	0.140182
<i>Kernel</i> Gaussiano (σ)	56.14%	0.280834	0.188034
<i>Kernel</i> Log (d)	64.91%	0.211743	0.126593
<i>Kernel</i> multiquádrico inverso (c)	64.91%	0.264157	0.173354
<i>Kernel</i> Cauchy (σ)	71.92%	0.183072	0.138657
<i>Kernel</i> Spline estável (c, λ)	71.92%	0.190331	0.111673

Tabela 7.29: Avaliação das previsões no conjunto de teste para o câmbio BRL-JPY

Real brasileiro – Renminbi chinês			
Modelo	Taxa de acerto direcional	RMSE	MAE
<i>Random Walk</i>	42.10%	0.017291	0.011874
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	40.35%	0.022801	0.013872
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	49.12%	0.017520	0.011996
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	52.63%	0.016923	0.011428
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	64.91%	0.014079	0.009638
<i>Kernel</i> Gaussiano (σ)	70.17%	0.013351	0.009214
<i>Kernel</i> Log (d)	75.43%	0.009479	0.005893
<i>Kernel</i> multiquádrico inverso (c)	66.67%	0.013782	0.009070
<i>Kernel</i> Cauchy (σ)	57.89%	0.015334	0.010563
<i>Kernel</i> Spline estável (c, λ)	84.21%	0.009080	0.005695

Tabela 7.30: Avaliação das previsões no conjunto de teste para o câmbio BRL-CNY

A.3: P-valores do *Reality Check Test* de White

Modelo	USD-EUR	USD-GBP	USD-JPY	USD-CNY	USD-BRL
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	0.836	0.623	0.404	0.614	0.062*
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	0.732	0.164	0.562	0.628	0.078*
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	0.680	0.332	0.368	0.474	0.038**
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	0.528	0.078*	0.138	0.590	0.030**
<i>Kernel</i> Gaussiano (σ)	0.094*	0.046**	0.020**	0.092*	0.012**
<i>Kernel</i> Log (d)	0.268	0.124	0.002***	0.084*	0.002***
<i>Kernel</i> multiquádrico inverso (c)	0.496	0.028**	0.181	0.430	0.012**
<i>Kernel</i> Cauchy (σ)	0.028**	0.016**	0.036**	0.230	0.004***
<i>Kernel</i> Spline estável (c, λ)	0.048**	0.008***	0.054*	0.036**	0.004***

(*: Significante a 10% **: Significante a 5% ***: Significante a 1%)

Tabela 7.31: P-valores do teste de White – Parte 1

Modelo	EUR-GBP	EUR-JPY	EUR-CNY	EUR-BRL	GBP-JPY
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	0.652	0.894	0.404	0.656	0.863
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	0.422	0.370	0.492	0.712	0.632
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	0.490	0.438	0.418	0.770	0.378
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	0.328	0.336	0.422	0.482	0.204
<i>Kernel</i> Gaussiano (σ)	0.082*	0.038**	0.038**	0.120	0.002***
<i>Kernel</i> Log (d)	0.010***	0.098*	0.002***	0.046**	0.024**
<i>Kernel</i> multiquádrico inverso (c)	0.350	0.284	0.062*	0.232	0.085*
<i>Kernel</i> Cauchy (σ)	0.008***	0.012**	0.050**	0.010***	0.000***
<i>Kernel</i> Spline estável (c, λ)	0.004***	0.020**	0.010***	0.014**	0.028**

(*: Significante a 10% **: Significante a 5% ***: Significante a 1%)

Tabela 7.32: P-valores do teste de White – Parte 2

Modelo	GBP-CNY	GBP-BRL	CNY-JPY	BRL-JPY	BRL-CNY
<i>Kernel</i> polinomial grau 1 (d), $q = 1$	0.692	0.572	0.456	0.478	0.618
<i>Kernel</i> polinomial grau 2 (d), $q = 2$	0.470	0.418	0.582	0.292	0.422
<i>Kernel</i> polinomial grau 3 (d), $q = 3$	0.298	0.580	0.118	0.198	0.286
<i>Kernel</i> polinomial grau 4 (d), $q = 4$	0.308	0.200	0.120	0.068*	0.094*
<i>Kernel</i> Gaussiano (σ)	0.040**	0.588	0.048**	0.150	0.080**
<i>Kernel</i> Log (d)	0.004***	0.078*	0.020**	0.042**	0.020**
<i>Kernel</i> multiquádrico inverso (c)	0.042**	0.616	0.070*	0.080*	0.084*
<i>Kernel</i> Cauchy (σ)	0.056*	0.034**	0.010***	0.022**	0.154
<i>Kernel</i> Spline estável (c, λ)	0.002***	0.024**	0.012**	0.026**	0.014**

(*: Significante a 10% **: Significante a 5% ***: Significante a 1%)

Tabela 7.33: P-valores do teste de White – Parte 3

ANEXO B: CÓDIGO DE PROGRAMAÇÃO EM R

Segue abaixo o código de programação utilizado para a análise empírica descrita no capítulo 5, contemplando a separação da base de dados, a definição das funções *Kernel* e do *benchmark*, o *grid search* dos parâmetros ótimos, a estimação das previsões no conjunto de testes, o cálculo da taxa de acerto direcional e das métricas de erro e a execução do teste de White. Para efeitos de simplificação e sem perda de generalidade, foi anexado apenas o código referente aos modelos SVR com o *Kernel* polinomial de grau 1; a construção dos demais 8 *Kernels* é análoga, com alterações apenas nos parâmetros considerados, suas faixas de variação para a busca e a nomenclatura dos objetos específicos a cada *Kernel*.

```
##### FUNCOES KERNEL #####

Kernelpoli1<-function(x,y) #Kernel polinomial grau 1
{
  res<-t(x)%*%y+cc
  return(res)
}
class(Kernelpoli1) <- "kernel"

Kernelpoli2<-function(x,y) #Kernel polinomial grau 2
{
  res<-(t(x)%*%y+cc)^2
  return(res)
}
class(Kernelpoli2) <- "kernel"

Kernelpoli3<-function(x,y) #Kernel polinomial grau 3
{
  res<-(t(x)%*%y+cc)^3
  return(res)
}
class(Kernelpoli3) <- "kernel"

Kernelpoli4<-function(x,y) #Kernel polinomial grau 4
{
  res<-(t(x)%*%y+cc)^4
```

```

    return(res)
}
class(Kernelpoli4) <- "kernel"

Kernelgauss<-function(x,y) #Kernel Gaussiano
{
  res<-exp(-sigma*(sum((x-y)^2)))
  return(res)
}
class(Kernelgauss) <- "kernel"

Kernellog<-function(x,y) #Kernel Log
{
  res<-(-log((sqrt(sum((x-y)^2))^d)+1))
  return(res)
}
class(Kernellog) <- "kernel"

Kernelinv<-function(x,y) #Kernel inverso multiquadrico
{
  res<-1/(sqrt(sum((x-y)^2)+(cc^2)))
  return(res)
}
class(Kernelinv) <- "kernel"

Kernelcau<-function(x,y) #Kernel Cauchy
{
  den<-1+((sum((x-y)^2))/(sigma^2))
  res<-1/den
  return(res)
}
class(Kernelcau) <- "kernel"

Kernelspli<-function(x,y) #Kernel Spline estavel
{
  res<-(cc/2*(lambda^(2*max(x,y)))*((lambda^min(x,y))-((lambda^max(x,y))/3)))
  return(res)
}
class(Kernelspli) <- "kernel"

##### SEPARACAO DA BASE #####

BBaseTotal<-read.table(file.choose()) # Le base de dados

```

```

set.seed(18) # Fixa semente para garantir mesmas linhas de treinamento/
# validacao/teste para todos os modelos
linhasDEtreino<-sample(1:nrow(BBaseTotal), size=(96), replace = FALSE)
# sorteia linhas de treinamento
linhasDEtreino<-sort(linhasDEtreino) # ordem crescente
prov<-BBaseTotal[-linhasDEtreino,] # objeto provisório
linhasDEvalidacao<-sample(1:nrow(prov), size=(38), replace = FALSE)
# sorteia linhas de validacao
linhasDEvalidacao<-sort(linhasDEvalidacao) # ordem crescente

treinamento<-BBaseTotal[linhasDEtreino,] # as 96 linhas sorteadas viram o conjunto de treinamento
validacao<-prov[linhasDEvalidacao,] # as 38 linhas sorteadas viram o conjunto de validacao
teste<-prov[-linhasDEvalidacao,] # o que sobrou e' o conjunto de teste
linhasteste<-as.numeric(row.names(teste)) # linhas de teste
rm(prov) # remover objeto provisório
colnames(treinamento) # checar colunas

##### BENCHMARK #####

benchmark<-NA
for (i in 1:nrow(teste))
{
  benchmark<-c(benchmark, teste$cambio[i-1]) # valor predito em t e' o valor observado em t-1
}
print(benchmark)

##### GRID SEARCH: TREINAMENTO E VALIDACAO #####

# carregando pacotes
library(kernlab)
library(foreach)
library(doParallel)
library(dosNOW)

# para rodar em paralelo
ncl<-detectCores()
ncl
cl <- makeCluster(ncl)
registerDoParallel(cl)

# definindo intervalo de busca para grid search
epsilonpol<-seq(0.05,1,by=0.05)

```

```

Cpol<-seq(0.5,10,by=0.5)
ccpol<-seq(0,5,by=0.1)
parametropol<-as.data.frame(expand.grid(epsilon=epsilonpol,C=Cpol,cc=ccpol))
# tabela com todas as combinacoes
performacepol<-as.data.frame(rep(0.0,nrow(parametropol)))
# tabela que vai receber o RMSE dos parametros aplicados no conjunto de validacao

# testando para cada combinacao
acuracia <-foreach(i=(1:nrow(parametropol)),.packages="kernlab",.combine='rbind') %dopar%
{
  epsilon<-parametropol[i,1]
  C<-parametropol[i,2]
  cc<-parametropol[i,3]
  Kernelpoli1<-function(x,y)
  {
    res<-t(x)%*%y+cc
    return(res)
  }
  class(Kernelpoli1) <- "kernel"

  # 'cambio' sendo explicado pelas outras variaveis ('cambiodef', 'juros',...)
  svm <- ksvm(c(treinamento$cambiodef,treinamento$juros,treinamento$bolsa,
  treinamento$ofertmon,treinamento$infla,treinamento$prod,treinamento$trade,
  treinamento$gov,treinamento$soil,treinamento$gold,treinamento$movkap,
  treinamento$sdr,treinamento$inflaexp),treinamento$cambio,epsilon=epsilon,
  C=C,kernel=Kernelpoli1,scaled=T)

  # valores preditos para cada combinacao de parametros aplicados no conjunto de validacao
  ypred<-predict(svm,validacao$cambio)
  CXP<-as.data.frame(as.numeric(ypred))
  # calculo do RMSE da predicao no conjunto de validacao para cada combinacao de parametros
  CXP$observado<-validacao$cambio
  CXP$desvio<-CXP[,1]-CXP[,2]
  CXP$desvioquad<-(CXP[,1]-CXP[,2])^2

  # guarda RMSE de cada combinacao de parametros
  performacepol[i,1]<-sqrt(mean(CXP$desvioquad))
  print(performacepol)
}

##### PARAMETROS OTIMOS #####

which(performacepol==min(performacepol)) # linha do menor RMSE

```

```

performacepol[which(performacepol==min(performacepol)),1] # valor do menor RMSE
parametropol[which(performacepol==min(performacepol)),] # parametros que forneceram menor RMSE

##### TESTE #####

rmse<-function(a){sqrt(mean(a^2))} # RMSE
mae<-function(a){mean(abs(a))} # MAE

# aplicando parametros otimos do treinamento/validacao no conjunto de teste
svmtestepol<-ksvm(c(teste$cambiodef, teste$juros, teste$bolsa, teste$ofertmon,
teste$infla, teste$prod, teste$trade, teste$gov, teste$oil, teste$gold,
teste$movkap, teste$sdr, teste$inflaexp), teste$cambio, epsilon=0.63, C=74,
kernel=Kernelpolil, scaled=T)
preditopol<-predict(svmtestepol, teste$cambio) # vetor de valores preditos

rmse(preditopol-teste$cambio)
rmse(benchmark[-1]-teste$cambio[-1]) # valor predito do benchmark para o primeiro periodo e' NA
mae(preditopol-teste$cambio)
mae(benchmark[-1]-teste$cambio[-1]) # valor predito do benchmark para o primeiro periodo e' NA

##### TAXA DE ACERTO DIRECIONAL #####

# para cada periodo, verificar se cambio observado subiu ou desceu
osci<-c()
for (i in 2:58)
{
  if (teste$cambio[i]<teste$cambio[i-1])
  {
    res<--1
  }
  else
  {
    res<-1
  }
  osci<-c(osci, res)
}
print(osci)

# para cada periodo, verificar se cambio predito pelo benchmark subiu ou desceu;
#NA para primeiro periodo
osciben<-c()
for (i in 3:58)
{

```

```

if (benchmark[i]<benchmark[i-1])
{
  res<--1
}
else
{
  res<-1
}
osciben<-c(osciben,res)
}
print(osciben)

# para cada periodo, verificar se cambio predito pelo SVR subiu ou desceu
osci2<-c()
for (i in 2:58)
{
  if (preditopol[i]<preditopol[i-1])
  {
    res<--1
  }
  else
  {
    res<-1
  }
  osci2<-c(osci2,res)
}
print(osci2)

which(osci[-1]!=osciben) # em quantos periodos benchmark errou direcao em relacao ao verdadeiro
(1-(length(which(osci[-1]!=osciben))/57))*100 # porcentagem de acerto

which(osci!=osci2) # em quantos periodos SVR errou direcao em relacao ao verdadeiro
(1-(length(which(osci!=osci2))/57))*100 # porcentagem de acerto

##### TESTE DE WHITE #####

# retorno esperado em excesso = retorno esperado do SVR
# ja que retorno esperado do benchmark e' zero por construcao
excessreturn<-matrix(nrow=1,ncol=57)
for (i in 2:58)
{
  excessreturn[i-1]<-log(preditopol[i]/preditopol[i-1])*osci2[i-1]
# log-retorno do SVR, corrigido pela direcao da operacao (compra ou venda)

```

```

}
excessreturn

nboot<-500 # numero de reamostragens
nobsteste<-57 # numero de retornos considerados
# perde-se a primeira das 58 linhas do conjunto de testes

# bootstrap estacionario
thetaboot<-matrix(nrow=nboot, ncol=nobsteste)
for(i in 1:nboot)
{
  thetaboot[i,]<-sample(nobsteste,nobsteste,replace=TRUE)
}
thetaboot

# media dos retornos do bootstrap, para cada reamostragem
mediaboot<-matrix(nrow=1, ncol=nboot)
for(i in 1:nboot)
{
  mediaboot[1,i]=mean(excessreturn[1,thetaboot[i,]])
}
mediaboot

# media do retorno em excesso observado
mediaretorno<-mean(excessreturn[1,])
mediaretorno

# estatistica V-barra; a conta e' direta porque S=1
# (comparou-se apenas uma estrategia por vez com o benchmark)
Vmedia<-nobsteste^(1/2)*mediaretorno
Vmedia

# estatisticas V-barra-estrela
Vboot<-numeric(nboot)
for(i in 1:nboot)
{
  Vboot[i]<-nobsteste^(1/2)*(mediaboot[1,i]-mediaretorno)
}
Vboot

# comparacao: contagem de valores de V-barra-estrela maiores que V-barra
pvalor<-length(which(Vboot>Vmedia))/nboot
pvalor # p-valor do teste; rejeitar H0 de random walk se for menor que 10/5/1%

```

REFERÊNCIAS BIBLIOGRÁFICAS

ABIDIN, S. N. Z.; JAAFAR, M. M. Surveying the best volatility measurements in stock market forecasting techniques involving small size companies in bursa malaysia. **IEEE Symposium on Humanities, Science and Engineering Research (SHUSER)**. [S.l.], 2012. p. 975–979.

ABOUNOORI, E.; ELMI, Z. M.; NADEMI, Y. Forecasting tehran stock exchange volatility; Markov switching garch approach. *Physica A: Statistical Mechanics and its Applications*, Elsevier BV, v. 445, p. 264–282, Mar 2016. ISSN 0378-4371.

ABRAHAM, A. Analysis of hybrid soft and hard computing techniques for forex monitoring systems. **Proceedings of the 2002 IEEE International Conference on Fuzzy Systems**. [S.l.], 2002. v. 2, p. 1616–1622.

ABUHAMAD, M.; MOHD, M.; SALIM, J. Event-driven business intelligence approach for real-time integration of technical and fundamental analysis in forex market. *Journal of Computer Science*, Science Publications, v. 9, n. 4, p. 488–499, Apr 2013. ISSN 1549-3636.

ADDAM, O.; CHEN, A.; HOANG, W.; ROKNE, J.; ALHAJJ, R. Foreign exchange data crawling and analysis for knowledge discovery leading to informative decision making. *Knowledge-Based Systems*, Elsevier BV, Mar 2016. ISSN 0950-7051.

AHMED, S.; LIU, X.; VALENTE, G. Can currency-based risk factors help forecast exchange rates? *International Journal of Forecasting*, Elsevier BV, v. 32, n. 1, p. 75–97, Jan 2016. ISSN 0169-2070.

AIOELLI, F.; DONINI, M. EasyMKL: a scalable multiple kernel learning algorithm. *Neurocomputing*, Elsevier BV, v. 169, p. 215–224, Dec 2015. ISSN 0925-2312.

ALBUQUERQUE, P. H. M. *Previsão de Séries Temporais Financeiras por meio de Máquinas de Suporte Vetorial e Ondaletas*. [S.l.], 2014.

ANDERSEN, T. G.; BOLLERSLEV, T.; DIEBOLD, F. X.; EBENS, H. The distribution of realized stock return volatility. *Journal of financial economics*, Elsevier, v. 61, n. 1, p. 43–76, 2001.

ARGHYROU, M. G.; POURPOURIDES, P. Inflation announcements and asymmetric exchange rate responses. *Journal of International Financial Markets, Institutions and Money*, Elsevier BV, v. 40, p. 80–84, Jan 2016. ISSN 1042-4431.

ARONSZAJN, N. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, American Mathematical Society (AMS), v. 68, n. 3, p. 337–337, Mar 1950. ISSN 0002-9947.

ATSALAKIS, G. S.; VALAVANIS, K. P. Surveying stock market forecasting techniques—part ii: Soft computing methods. *Expert Systems with Applications*, Elsevier, v. 36, n. 3, p. 5932–5941, 2009.

- AUER, P.; BURGSTEINER, H.; MAASS, W. A learning rule for very simple universal approximators consisting of a single layer of perceptrons. *Neural Networks*, Elsevier, v. 21, n. 5, p. 786–795, 2008.
- BAASHER, A. A.; FAKHR, M. W. Forex trend classification using machine learning techniques. *Recent Researches in Applied Informatics and Remote Sensing*, v. 8, p. 41–47, 2012.
- BACCHETTA, P.; WINCOOP, E. van. On the unstable relationship between exchange rates and macroeconomic fundamentals. *Journal of International Economics*, Elsevier BV, v. 91, n. 1, p. 18–26, Sep 2013. ISSN 0022-1996.
- BACHELIER, L. *Théorie de la spéculation*. 21–86 p. Tese (Doutorado) — Gauthier-Villars, 1900.
- BADEA, L. M. Supporting management decisions by using artificial neural networks for exchange rate prediction. *Accounting and Management Information Systems*, Bucharest Academy of Economic Studies, v. 12, n. 4, p. 578–594, 2013.
- BALDWIN, R. Global manufacturing value chains and trade rules. *The shifting Geography of Global Value Chains: implications for developing countries and trade policy*, 2012.
- BALDWIN, R. The world trade organization and the future of multilateralism. *Journal of Economic Perspectives*, American Economic Association, v. 30, n. 1, p. 95–116, Feb 2016. ISSN 0895-3309.
- BALKE, N. S.; MA, J.; WOHAR, M. E. The contribution of economic fundamentals to movements in exchange rates. *Journal of International Economics*, Elsevier BV, v. 90, n. 1, p. 1–16, May 2013. ISSN 0022-1996.
- BANSAL, R.; SHALIASTOVICH, I. A long-run risks explanation of predictability puzzles in bond and currency markets. *Review of Financial Studies*, Soc Financial Studies, v. 26, n. 1, p. 1–33, 2013.
- BASAK, J. A least square kernel machine with box constraints. **19th International Conference on Pattern Recognition**. [S.l.]: Institute of Electrical and Electronics Engineers (IEEE), 2008. ISBN <http://id.crossref.org/isbn/978-1-4244-2174-9>. ISSN 1051-4651.
- BASHER, S. A.; HAUG, A. A.; SADORSKY, P. The impact of oil shocks on exchange rates: A markov-switching approach. *Energy Economics*, Elsevier, v. 54, p. 11–23, 2016.
- BAYOUMI, T.; EICHENGREEN, B. Exchange rate volatility and intervention: implications of the theory of optimum currency areas. *Journal of International Economics*, Elsevier, v. 45, n. 2, p. 191–209, 1998.
- BECKMANN, J.; SCHÜSSLER, R. Forecasting exchange rates under parameter and model uncertainty. *Journal of International Money and Finance*, Elsevier BV, v. 60, p. 267–288, Feb 2016. ISSN 0261-5606.
- BEKIROU, S. D. Heuristic learning in intraday trading under uncertainty. *Journal of Empirical Finance*, Elsevier BV, v. 30, p. 34–49, Jan 2015. ISSN 0927-5398.
- BELTRAMI, M.; LOCH, G. V.; SILVA, A. Comparação das técnicas de support vector regression e redes neurais na precificação de opções. *XLII Simpósio Brasileiro de Pesquisa Operacional*, p. 572–583, 2011.

BERG, K. A.; MARK, N. C. Third-country effects on the exchange rate. *Journal of International Economics*, Elsevier BV, v. 96, n. 2, p. 227–243, Jul 2015. ISSN 0022-1996.

BISSOONDEEAL, R. K.; BINNER, J. M.; BHURUTH, M.; GAZELY, A.; MOOTANAH, V. P. Forecasting exchange rates with linear and nonlinear models. *Global Business and Economics Review*, Inderscience Publishers, v. 10, n. 4, p. 414–429, 2008.

BISSOONDEEAL, R. K.; KAROGLU, M.; GAZELY, A. M. Forecasting the UK/US exchange rate with divisia monetary models and neural networks. *Scottish journal of political economy*, Wiley Online Library, v. 58, n. 1, p. 127–152, 2011.

BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. *The journal of political economy*, JSTOR, p. 637–654, 1973.

BODART, V.; CANDELON, B.; CARPANTIER, J.-F. Real exchanges rates, commodity prices and structural factors in developing countries. *Journal of International Money and Finance*, Elsevier BV, v. 51, p. 264–284, Mar 2015. ISSN 0261-5606.

BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, Elsevier, v. 31, n. 3, p. 307–327, 1986.

BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. **Proceedings of the fifth annual workshop on Computational learning theory**. [S.l.]: Association for Computing Machinery (ACM), 1992. ISBN <http://id.crossref.org/isbn/089791497X>.

BOUAKEZ, H.; EYQUEM, A. Government spending, monetary policy, and the real exchange rate. *Journal of International Money and Finance*, Elsevier BV, v. 56, p. 178–201, Sep 2015. ISSN 0261-5606.

BOUGHORBEL, S.; TAREL, J.-P.; BOUJEMAA, N. Conditionally positive definite kernels for SVM based image recognition. **IEEE International Conference on Multimedia and Expo**. [S.l.]: Institute of Electrical and Electronics Engineers (IEEE), 2005. ISBN <http://id.crossref.org/isbn/0-7803-9331-7>.

BOUOUIYOUR, J.; SELMI, R.; TIWARI, A. K.; SHAHBAZ, M. The nexus between oil price and Russia's real exchange rate: Better paths via unconditional vs conditional analysis. *Energy Economics*, Elsevier BV, v. 51, p. 54–66, Sep 2015. ISSN 0140-9883.

BOWN, C. P.; CROWLEY, M. A. Import protection, business cycles, and exchange rates: Evidence from the great recession. *Journal of International Economics*, Elsevier BV, v. 90, n. 1, p. 50–64, May 2013. ISSN 0022-1996.

BOŽIĆ, J.; VUKOTIĆ, S.; BABIĆ, Đ. Prediction of the RSD exchange rate by using wavelets and neural networks. **19th Telecommunications Forum (TELFOR)**. [S.l.], 2011. p. 703–706.

BRAHMASRENE, T.; HUANG, J.-C.; SISSOKO, Y. Crude oil prices and exchange rates: Causality, variance decomposition and impulse response. *Energy Economics*, Elsevier BV, v. 44, p. 407–412, Jul 2014. ISSN 0140-9883.

BRITO, R. F. B. de; OLIVEIRA, A. L. I. Comparative Study of FOREX Trading Systems Built with SVR+GHSOM and Genetic Algorithms Optimization of Technical Indicators. *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, Institute of Electrical & Electronics Engineers (IEEE), p. 351–358, Nov 2012.

- BUNCIC, D.; PIRAS, G. D. Heterogeneous agents, the financial crisis and exchange rate predictability. *Journal of International Money and Finance*, Elsevier BV, v. 60, p. 313–359, Feb 2016. ISSN 0261-5606.
- BURNS, K.; MOOSA, I. A. Enhancing the forecasting power of exchange rate models by introducing nonlinearity: Does it work? *Economic Modelling*, Elsevier BV, v. 50, p. 27–39, Nov 2015. ISSN 0264-9993.
- BYRNE, J. P.; KOROBILIS, D.; RIBEIRO, P. J. Exchange rate predictability in a changing world. *Journal of International Money and Finance*, Elsevier BV, v. 62, p. 1–24, Apr 2016. ISSN 0261-5606.
- CAI, Z.; JIANG, J.; ZHANG, J. *A New Test for Superior Predictive Ability*. [S.l.], 2013.
- CAO, D.-Z.; PANG, S.-L.; BAI, Y.-H. Forecasting exchange rate using support vector machines. **Proceedings of 2005 International Conference on Machine Learning and Cybernetics**. [S.l.], 2005. v. 6, p. 3448–3452.
- CAO, L. Support vector machines experts for time series forecasting. *Neurocomputing*, Elsevier, v. 51, p. 321–339, 2003.
- CARAIANI, P.; HAVEN, E. Evidence of multifractality from CEE exchange rates against Euro. *Physica A: Statistical Mechanics and its Applications*, Elsevier BV, v. 419, p. 395–407, Feb 2015. ISSN 0378-4371.
- CARRIERO, A.; KAPETANIOS, G.; MARCELLINO, M. Forecasting Exchange Rates with a Large Bayesian VAR. *International Journal of Forecasting*, v. 25, n. 2, p. 400–417, 2009.
- CA'ZORZI, M.; KOCIEŃKI, A.; RUBASZEK, M. Bayesian forecasting of real exchange rates with a dornbusch prior. *Economic Modelling*, Elsevier BV, v. 46, p. 53–60, Apr 2015. ISSN 0264-9993.
- CHABOUD, A. P.; CHIQUOINE, B.; HJALMARSSON, E.; VEGA, C. Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, Wiley Online Library, v. 69, n. 5, p. 2045–2084, 2014.
- CHANDRASEKHAR, P.; REDDY, B. M. Training a support vector classifier using a cauchy-laplace product kernel. *International Journal of Computer Applications*, v. 116, n. 21, p. 48–, Apr 2015.
- CHANG, C.-C.; LIN, C.-J. Training ν -support vector regression: Theory and algorithms. *Neural Computation*, v. 14, n. 8, p. 1959–1977, 2002.
- CHAO, J.; SHEN, F.; ZHAO, J. Forecasting exchange rate with deep belief networks. **International Joint Conference on Neural Networks (IJCNN)**. [S.l.], 2011. p. 1259–1266.
- CHEN, A.-S.; LEUNG, M. T. Regression neural network for error correction in foreign exchange forecasting and trading. *Computers & Operations Research*, Elsevier, v. 31, n. 7, p. 1049–1068, 2004.
- CHEN, T.; OHLSSON, H.; LJUNG, L. On the estimation of transfer functions, regularizations and Gaussian processes—revisited. *Automatica*, Elsevier BV, v. 48, n. 8, p. 1525–1535, Aug 2012. ISSN 0005-1098.

- CHEN, X.; MACDONALD, R. Measuring the dollar-euro permanent equilibrium exchange rate using the unobserved components model. *Journal of International Money and Finance*, Elsevier BV, v. 53, p. 20–35, May 2015. ISSN 0261-5606.
- CHEUNG, Y.-W.; CHINN, M. D.; PASCUAL, A. G. Empirical exchange rate models of the nineties: Are any fit to survive? *Journal of international money and finance*, Elsevier, v. 24, n. 7, p. 1150–1175, 2005.
- CHEUNG, Y.-W.; RIME, D. The offshore renminbi exchange rate: Microstructure and links to the onshore market. *Journal of International Money and Finance*, Elsevier BV, v. 49, p. 170–189, Dec 2014. ISSN 0261-5606.
- CHIȚU, L.; EICHENGREEN, B.; MEHL, A. History, gravity and international finance. *Journal of International Money and Finance*, Elsevier BV, v. 46, p. 104–129, Sep 2014. ISSN 0261-5606.
- CHOUDHRI, E. U.; HAKURA, D. S. The exchange rate pass-through to import and export prices: The role of nominal rigidities and currency choice. *Journal of International Money and Finance*, Elsevier BV, v. 51, p. 1–25, Mar 2015. ISSN 0261-5606.
- CHOWDHURY, I.; SARNO, L. Time-varying volatility in the foreign exchange market: New evidence on its persistence and on currency spillovers. *Journal of Business Finance & Accounting*, Wiley Online Library, v. 31, n. 5-6, p. 759–793, 2004.
- CHUNG, J.; HONG, Y. Model-free evaluation of directional predictability in foreign exchange markets. *Journal of Applied Econometrics*, Wiley Online Library, v. 22, n. 5, p. 855–889, 2007.
- CLARIDA, R. H.; SARNO, L.; TAYLOR, M. P.; VALENTE, G. The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond. *Journal of International Economics*, v. 60, n. 1, p. 61–83, 2003.
- CLARIDA, R. H.; TAYLOR, M. P. The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors. *Review of Economics and Statistics*, v. 79, n. 3, p. 353–361, 1997.
- CLARIDA, R. H.; WALDMAN, D. Is bad news about inflation good news for the exchange rate? and, if so, can that tell us anything about the conduct of monetary policy? In: *Asset Prices and Monetary Policy*. [S.I.]: University of Chicago Press, 2008. p. 371–396.
- CLARK, T. E.; WEST, K. D. Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, Elsevier, v. 138, n. 1, p. 291–311, 2007.
- COLEMAN, T. F.; LI, Y.; WANG, C. Stable local volatility function calibration using spline kernel. *Comput Optim Appl*, Springer Science + Business Media, v. 55, n. 3, p. 675–702, Feb 2013. ISSN 1573-2894.
- CONRAD, C.; LAMLA, M. J. The high-frequency response of the EUR-USD exchange rate to ECB communication. *Journal of Money, Credit and Banking*, Wiley Online Library, v. 42, n. 7, p. 1391–1417, 2010.
- COOPER, R. N. Trade policy is foreign policy. *Foreign Policy*, JSTOR, n. 9, p. 18–36, 1972.
- CORDEN, W. M. Booming sector and Dutch disease economics: survey and consolidation. *Oxford Economic Papers*, v. 36, n. 3, p. 359–380, 1984.

- CORRADO, C. J. Event studies: A methodology review. *Accounting & Finance*, Wiley Online Library, v. 51, n. 1, p. 207–234, 2011.
- CORTE, P. D.; SARNO, L.; SESTIERI, G. The Predictive Information Content of External Imbalances for Exchange Rate Returns: How Much Is It Worth? *Review of Economics and Statistics*, v. 94, n. 1, p. 100–115, 2012.
- CORTE, P. D.; SARNO, L.; TSIAKAS, I. Spot and forward volatility in foreign exchange. *Journal of Financial Economics*, Elsevier, v. 100, n. 3, p. 496–513, 2011.
- CORTES, C.; VAPNIK, V. N. Support-vector networks. *Machine Learning*, v. 20, p. 273–297, 1995.
- CUMBY, R. E.; OBSTFELD, M. *International interest-rate and price-level linkages under flexible exchange rates: A review of recent evidence*. [S.l.]: National Bureau of Economic Research Cambridge, Mass., USA, 1982.
- CUMBY, R. F.; OBSTFELD, M. *International interest-rate linkages under flexible exchange rates: A review of recent evidence*. [S.l.]: University of Chicago Press, 1984.
- CYBENKO, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, Springer, v. 5, n. 4, p. 455–455, 1992.
- DĄBROWSKI, M. A.; PAPIEŻ, M.; ŚMIECH, S. Exchange rates and monetary fundamentals in CEE countries: Evidence from a panel approach. *Journal of Macroeconomics*, Elsevier BV, v. 41, p. 148–159, Sep 2014. ISSN 0164-0704.
- DAOUD, E. A.; TURABIEH, H. New empirical nonparametric kernels for support vector machine classification. *Applied Soft Computing*, Elsevier BV, v. 13, n. 4, p. 1759–1765, Apr 2013. ISSN 1568-4946.
- DENG, S.; YOSHIYAMA, K.; MITSUBUCHI, T.; SAKURAI, A. Hybrid method of multiple kernel learning and genetic algorithm for forecasting short-term foreign exchange rates. *Computational Economics*, Springer, v. 45, n. 1, p. 49–89, 2015.
- DICK, C. D.; MACDONALD, R.; MENKHOFF, L. Exchange rate forecasts and expected fundamentals. *Journal of International Money and Finance*, Elsevier BV, v. 53, p. 235–256, May 2015. ISSN 0261-5606.
- DIEBOLD, F. X.; GARDEAZABAL, J.; YILMAZ, K. On cointegration and exchange rate dynamics. *Journal of Finance*, v. 49, n. 2, p. 727–735, 1994.
- DIEBOLD, F. X.; HAHN, J.; TAY, A. S. Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange. *Review of Economics and Statistics*, v. 81, n. 4, p. 661–673, 1999.
- DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, Informa UK Limited, v. 13, n. 3, p. 253–263, Jul 1995. ISSN 1537-2707.
- DJEUTEM, E.; KASA, K. Robustness and exchange rate volatility. *Journal of International Economics*, Elsevier BV, v. 91, n. 1, p. 27–39, Sep 2013. ISSN 0022-1996.

- DONAYRE, L.; PANOVSKA, I. State-dependent exchange rate pass-through behavior. *Journal of International Money and Finance*, Elsevier BV, v. 64, p. 170–195, Jun 2016. ISSN 0261-5606.
- DRUCKER, H.; BURGESS, C. J.; KAUFMAN, L.; SMOLA, A.; VAPNIK, V. N. Support vector regression machines. **Advances in neural information processing systems**. [S.l.]: Morgan Kaufmann Publishers, 1997. p. 155–161.
- DUEKER, M.; NEELY, C. J. Can markov switching models predict excess foreign exchange returns? *Journal of Banking & Finance*, Elsevier, v. 31, n. 2, p. 279–296, 2007.
- DUNIS, C. L.; HUANG, X. Forecasting and trading currency volatility: An application of recurrent neural regression and model combination. *Journal of forecasting*, Wiley Online Library, v. 21, n. 5, p. 317–354, 2002.
- DYMOVA, L.; SEVASTJANOV, P.; KACZMAREK, K. A forex trading expert system based on a new approach to the rule-based evidential reasoning. *Expert Systems with Applications*, Elsevier BV, v. 51, p. 1–13, Jun 2016. ISSN 0957-4174.
- ECKHARD, T.; VALERO, E. M.; HERNÁNDEZ-ANDRÉS, J.; HEIKKINEN, V. Evaluating logarithmic kernel for spectral reflectance estimation—effects on model parametrization, training set size, and number of sensor spectral channels. *Journal of the Optical Society of America A, Optics and Image Science*, The Optical Society, v. 31, n. 3, p. 541–549, Feb 2014. ISSN 1520-8532.
- EHRMANN, M.; FRATZSCHER, M. Exchange rates and fundamentals: new evidence from real-time data. *Journal of International Money and Finance*, Elsevier, v. 24, n. 2, p. 317–341, 2005.
- EICHENGREEN, B. The renminbi as an international currency. *Journal of Policy Modeling*, Elsevier BV, v. 33, n. 5, p. 723–730, Sep 2011. ISSN 0161-8938.
- ENG, M. H.; LI, Y.; WANG, Q.-G.; LEE, T. H. Forecast Forex with ANN Using Fundamental Data. *2008 International Conference on Information Management, Innovation Management and Industrial Engineering*, Institute of Electrical & Electronics Engineers (IEEE), p. 279–282, Dec 2008.
- ENGEL, C. Can the markov switching model forecast exchange rates? *Journal of International Economics*, v. 36, n. 1–2, p. 151–165, 1994.
- ENGEL, C.; MARK, N. C.; WEST, K. D. Factor model forecasts of exchange rates. *Econometric Reviews*, Taylor & Francis, v. 34, n. 1-2, p. 32–55, 2015.
- ENGEL, C.; WEST, K. Exchange rates and fundamentals. *Journal of Political Economy*, University of Chicago Press, v. 113, n. 3, p. 485–517, Jun 2005. ISSN 1537-534X.
- EVANS, C.; PAPPAS, K.; XHAFSA, F. Utilizing artificial neural networks and genetic algorithms to build an algo-trading model for intra-day foreign exchange speculation. *Mathematical and Computer Modelling*, Elsevier BV, v. 58, n. 5-6, p. 1249–1266, Sep 2013. ISSN 0895-7177.
- EVANS, M. D.; LYONS, R. K. Do currency markets absorb news quickly? *Journal of International Money and Finance*, Elsevier, v. 24, n. 2, p. 197–217, 2005.

- FALAT, L.; PANCIKOVA, L. Quantitative modelling in economics with advanced artificial neural networks. *Procedia Economics and Finance*, Elsevier BV, v. 34, p. 194–201, 2015. ISSN 2212-5671.
- FAMA, E. F. Random walks in stock market prices. *Financial Analysts Journal*, CFA Institute, v. 21, n. 5, p. 55–59, Sep 1965. ISSN 0015-198X.
- FAMA, E. F. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, JSTOR, v. 25, n. 2, p. 383, May 1970. ISSN 0022-1082.
- FAMA, E. F. Forward and spot exchange rates. *Journal of monetary economics*, Elsevier, v. 14, n. 3, p. 319–338, 1984.
- FATUM, R.; HUTCHISON, M.; WU, T. Asymmetries and state dependence: The impact of macro surprises on intraday exchange rates. *Journal of the Japanese and International Economies*, Elsevier, v. 26, n. 4, p. 542–560, 2012.
- FAUST, J.; ROGERS, J. H.; WRIGHT, J. H. Exchange rate forecasting: the errors we've really made. *Journal of International Economics*, Elsevier BV, v. 60, n. 1, p. 35–59, May 2003. ISSN 0022-1996.
- FERRARO, D.; ROGOFF, K.; ROSSI, B. Can oil prices forecast exchange rates? an empirical analysis of the relationship between commodity prices and exchange rates. *Journal of International Money and Finance*, Elsevier BV, v. 54, p. 116–141, Jun 2015. ISSN 0261-5606.
- FIORENTINO, R. V.; CRAWFORD, J.-A.; TOQUEBOEUF, C. The landscape of regional trade agreements and WTO surveillance. *Multilateralizing regionalism: Challenges for the global trading system*, Cambridge University Press, p. 28, 2009.
- FRANKEL, J.; POONAWALA, J. The forward market in emerging currencies: Less biased than in major currencies. *Journal of International Money and Finance*, Elsevier BV, v. 29, n. 3, p. 585–598, Apr 2010. ISSN 0261-5606.
- FRATZSCHER, M.; RIME, D.; SARNO, L.; ZINNA, G. The scapegoat theory of exchange rates: the first tests. *Journal of Monetary Economics*, Elsevier BV, v. 70, p. 1–21, Mar 2015. ISSN 0304-3932.
- FRENKEL, J. A. A monetary approach to the exchange rate: Doctrinal aspects and empirical evidence. *Scandinavian Journal of Economics*, v. 78, n. 2, p. 200–224, 1976.
- GALESHCHUK, S. Neural networks performance in exchange rate prediction. *Neurocomputing*, Elsevier BV, v. 172, p. 446–452, Jan 2016. ISSN 0925-2312.
- GARDNER, R. N. *Sterling-dollar diplomacy in current perspective: the origins and the prospects of our international economic order*. [S.l.]: Columbia University Press New York, 1980.
- GASPAR, P.; CARBONELL, J.; OLIVEIRA, J. L. On the parameter optimization of support vector machines for binary classification. *Journal of Integrative Bioinformatics*, v. 9, n. 3, p. 1–11, 2012.
- GATOPOULOS, G.; LOUBERGÉ, H. Combined use of foreign debt and currency derivatives under the threat of currency crises: The case of latin American firms. *Journal of International Money and Finance*, Elsevier BV, v. 35, p. 54–75, Jun 2013. ISSN 0261-5606.

- GELMAN, M.; JOCHEM, A.; REITZ, S.; TAYLOR, M. P. Real financial market exchange rates and capital flows. *Journal of International Money and Finance*, Elsevier BV, v. 54, p. 50–69, Jun 2015. ISSN 0261-5606.
- GENÇAY, R.; GRADOJEVIC, N. Crash of 87Was it expected?: Aggregate market fears and long-range dependence. *Journal of Empirical Finance*, Elsevier, v. 17, n. 2, p. 270–282, 2010.
- GERLEIN, E. A.; MCGINNITY, M.; BELATRECHE, A.; COLEMAN, S. Evaluating machine learning classification for financial trading: An empirical approach. *Expert Systems with Applications*, Elsevier BV, v. 54, p. 193–207, Jul 2016. ISSN 0957-4174.
- GIACOMINI, R.; ROSSI, B. Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, Wiley-Blackwell, v. 25, n. 4, p. 595–620, Apr 2010. ISSN 0883-7252.
- GOLLER, C.; KUCHLER, A. Learning task-dependent distributed representations by backpropagation through structure. **IEEE International Conference on Neural Networks**. [S.l.], 1996. v. 1, p. 347–352.
- GONG, X.; SI, Y.-W.; FONG, S.; BIUK-AGHAI, R. P. Financial time series pattern matching with extended UCR Suite and Support Vector Machine. *Expert Systems with Applications*, Elsevier BV, v. 55, p. 284–296, Aug 2016. ISSN 0957-4174.
- GONZÁLEZ-RIVERA, G.; LEE, T.-H.; MISHRA, S. Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting*, v. 20, p. 629–645, 2004.
- GOODMAN, S. H. Foreign exchange rate forecasting techniques: implications for business and policy. *The Journal of Finance*, Wiley Online Library, v. 34, n. 2, p. 415–427, 1979.
- GRADOJEVIC, N. *Non-linear exchange rate forecasting: The role of market microstructure variables*. Tese (Doutorado) — University of British Columbia, 2002.
- GREENAWAY, R.; MARK, N. C.; SUL, D.; WU, J.-L. Exchange rates as exchange rate common factors. *Working paper, Department of Economics, University of Notre Dame.*, HKIMR Working Paper, 2014.
- GRIFFITH-JONES, S.; KIMMIS, J. *The Role of the SDR in the International Financial System*. [S.l.], 2001.
- GUNN, S. *Support Vector Machines for Classification and Regression*. [S.l.], May 1998.
- HANN, T. H.; STEURER, E. Much ado about nothing? exchange rate forecasting: Neural networks vs. linear models using monthly and weekly data. *Neurocomputing*, Elsevier, v. 10, n. 4, p. 323–339, 1996.
- HANSEN, P. R. A test for superior predictive ability. *Journal of Business and Economic Statistics*, v. 23, p. 365–380, 2005.
- HASSAN, T. A.; MERTENS, T. M.; ZHANG, T. Not so disconnected: Exchange rates and the capital stock. *Journal of International Economics*, Elsevier BV, v. 99, p. S43–S57, Mar 2016. ISSN 0022-1996.

- HAU, H.; REY, H. Exchange rates, equity prices, and capital flows. *Review of financial studies*, Soc Financial Studies, v. 19, n. 1, p. 273–317, 2006.
- HE, H.; SHEN, X. Bootstrap methods for foreign currency exchange rates prediction. **International Joint Conference on Neural Networks, 2007. IJCNN**. [S.l.], 2007. p. 1272–1277.
- HENRIQUE, P. A.; MARCELINO, S. S. F.; LIMA, M.; FERREIRA, F. N.; ALBUQUERQUE, P. H. M. Portfolio selection with support vector regression. **R Finance Chicago**. [S.l.: s.n.], 2016.
- HINTON, G. E.; OSINDERO, S.; TEH, Y.-W. A fast learning algorithm for deep belief nets. *Neural computation*, MIT Press, v. 18, n. 7, p. 1527–1554, 2006.
- HOLSTI, K. J. Politics in command: foreign trade as national security policy. *International Organization*, JSTOR, p. 643–671, 1986.
- HONG, W.-C. A hybrid support vector machine regression for exchange rate prediction. *International Journal of Information and Management Sciences*, v. 17, n. 2, p. 19–32, 2006.
- HOOY, C.-W.; SIONG-HOOK, L.; TZE-HAW, C. The impact of the Renminbi real exchange rate on ASEAN disaggregated exports to China. *Economic Modelling*, Elsevier BV, v. 47, p. 253–259, Jun 2015. ISSN 0264-9993.
- HSIEH, J.-G.; LIN, Y.-L.; JENG, J.-H. Preliminary study on wilcoxon learning machines. *IEEE Transactions on Neural Networks*, IEEE, v. 19, n. 2, p. 201–211, 2008.
- HUA, X.; ZHANG, D.; LEUNG, S. C. H. Exchange rate prediction through ann based on kernel regression. *2010 Third International Conference on Business Intelligence and Financial Engineering*, Institute of Electrical & Electronics Engineers (IEEE), p. 39–43, Aug 2010.
- HUANG, S.-C.; CHUANG, P.-J.; WU, C.-F.; LAI, H.-J. Chaos-based support vector regressions for exchange rate forecasting. *Expert Systems with Applications*, Elsevier BV, v. 37, n. 12, p. 8590–8598, Dec 2010. ISSN 0957-4174.
- HUANG, Y.-L.; HUANG, C.-H. The persistence of taiwan's output fluctuations: an empirical study using innovation regime-switching model. *Applied Economics*, Taylor & Francis, v. 39, n. 20, p. 2673–2679, 2007.
- HURRELL, A. Power transitions, emerging powers, and the shifting terrain of the middle ground. In: *Ethical Reasoning in International Affairs*. [S.l.]: Springer, 2013. p. 222–245.
- INCE, H.; TRAFALIS, T. B. A hybrid model for exchange rate prediction. *Decision Support Systems*, Elsevier BV, v. 42, n. 2, p. 1054–1062, Nov 2006. ISSN 0167-9236.
- ISMAIL, M. T.; ISA, Z. Modelling exchange rates using regime switching models. *Sains Malaysiana*, Universiti Kebangsaan Malaysia, v. 35, n. 2, p. 55–62, 2006.
- JAMMAZI, R.; LAHIANI, A.; NGUYEN, D. K. A wavelet-based nonlinear ARDL model for assessing the exchange rate pass-through to crude oil prices. *Journal of International Financial Markets, Institutions and Money*, Elsevier BV, v. 34, p. 173–187, Jan 2015. ISSN 1042-4431.
- JUNTTILA, J.; KORHONEN, M. Nonlinearity and time-variation in the monetary model of exchange rates. *Journal of Macroeconomics*, Elsevier, v. 33, n. 2, p. 288–302, 2011.

KAMAL, Y. Modeling the exchange rate volatility, using generalized autoregressive conditionally heteroscedastic (GARCH) type models: Evidence from Pakistan. *African Journal of Business Management*, Academic Journals, v. 6, n. 8, p. 2830–2838, Feb 2012. ISSN 1993-8233.

KAMRUZZAMAN, J.; SARKER, R. A. Application of support vector machine to forex monitoring. *Journal of The Association of Eletronics (Japanese)*, The Institute of Electrical Engineers of Japan, v. 124, n. 10, p. 1944–1951, 2004.

KAMRUZZAMAN, J.; SARKER, R. A.; AHMAD, I. SVM based models for predicting foreign currency exchange rates. **Third IEEE International Conference on Data Mining, ICDM**. [S.I.], 2003. p. 557–560.

KARATZOGLOU, A.; SMOLA, A.; HORNIK, K.; ZEILEIS, A. Kernlab: An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, v. 11, n. 9, p. 1–20, 2004.

KATUSIIME, L.; SHAMSUDDIN, A.; AGBOLA, F. W. Macroeconomic and market microstructure modelling of Ugandan exchange rate. *Economic Modelling*, Elsevier BV, v. 45, p. 175–186, Feb 2015. ISSN 0264-9993.

KHAN, A. K.; ANWER, M.; BANIK, S. Exchange rate prediction using fuzzy system neural network approach. **IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)**. [S.I.], 2013. p. 1321–1326.

KILIAN, L.; TAYLOR, M. P. Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, Elsevier BV, v. 60, n. 1, p. 85–107, May 2003. ISSN 0022-1996.

KIM, K.-J. Financial time series forecasting using support vector machines. *Neurocomputing*, Elsevier BV, v. 55, n. 1-2, p. 307–319, Sep 2003. ISSN 0925-2312.

KLEIN, C.; KU, L. A value investing strategy for forecasting exchange rates. **Proceedings of the IIE Annual Conference**. [S.I.], 2006. p. 1–7.

KODOGIANNIS, V.; LOLIS, A. Forecasting financial time series using neural network and fuzzy system-based techniques. *Neural computing & applications*, Springer, v. 11, n. 2, p. 90–102, 2002.

KOHLSCHEEN, E. The impact of monetary policy on the exchange rate: A high frequency exchange rate puzzle in emerging economies. *Journal of International Money and Finance*, Elsevier BV, v. 44, p. 69–96, Jun 2014. ISSN 0261-5606.

KOROL, T. A fuzzy logic model for forecasting exchange rates. *Knowledge-Based Systems*, Elsevier BV, v. 67, p. 49–60, Sep 2014. ISSN 0950-7051.

KOSE, M. A. *Emerging markets: Resilience and growth amid global turmoil*. [S.I.]: Brookings Institution Press, 2011.

KROLLNER, B.; VANSTONE, B.; FINNIE, G. Financial time series forecasting with machine learning techniques: A survey. *European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning*, 2010.

- KRUGMAN, P. The narrow moving band, the dutch disease, and the competitive consequences of mrs. thatcher: Notes on trade in the presence of dynamic scale economies. *Journal of development Economics*, v. 27, n. 1–2, p. 41–55, 1987.
- KUAN, C.-M.; LIU, T. Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*, Wiley Online Library, v. 10, n. 4, p. 347–364, 1995.
- KUTAN, A. M.; VUKŠIĆ, G. Foreign direct investment and export performance: empirical evidence. *Comparative Economic Studies*, v. 49, n. 3, p. 430–445, 2007.
- LAWS, J.; THOMPSON, J. The efficiency of financial futures markets: Tests of prediction accuracy. *European Journal of Operational Research*, Elsevier, v. 155, n. 2, p. 284–298, 2004.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.
- LEE, Y.; OW, L. T. C.; LING, D. N. C. Hidden markov models for forex trends prediction. **International Conference on Information Science and Applications (ICISA)**. [S.I.], 2014. p. 1–4.
- LEI, Z. Nonlinear prediction of exchange rate: A new approach to multiple time series analysis. **International Conference on Management Science and Engineering (ICMSE)**. [S.I.], 2013. p. 1798–1803.
- LENTO, C. A combined signal approach to technical analysis on the S&P 500. *SSRN 1113622*, 2008.
- LI, M.; SUOHAI, F. Forex Prediction Based on SVR Optimized by Artificial Fish Swarm Algorithm. *Fourth Global Congress on Intelligent Systems*, Institute of Electrical & Electronics Engineers (IEEE), p. 47–52, Dec 2013.
- LI, W.; WONG, M. C.; CENEV, J. High frequency analysis of macro news releases on the foreign exchange market: A survey of literature. *Big Data Research*, Elsevier BV, v. 2, n. 1, p. 33–48, Mar 2015. ISSN 2214-5796.
- LIN, H.-T.; LIN, C.-J. *A Study on Sigmoid Kernels for SVM and the Training of non-PSD Kernels by SMO-type Methods*. [S.I.], 2003.
- LIN, W.-Y.; HU, Y.-H.; TSAI, C.-F. Machine learning in financial crisis prediction: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, IEEE, v. 42, n. 4, p. 421–436, 2012.
- LISI, F.; SCHIAVO, R. A. A comparison between neural networks and chaotic models for exchange rate prediction. *Computational Statistics & Data Analysis*, Elsevier, v. 30, n. 1, p. 87–102, 1999.
- LIU, F.-Y. The hybrid prediction model of CNY/USD exchange rate based on wavelet and support vector regression. **International Conference on Advanced Computer Control (ICACC)**. [S.I.], 2010. v. 4, p. 561–565.
- LÓPEZ-SUÁREZ, C. F.; RODRÍGUEZ-LÓPEZ, J. A. Nonlinear exchange rate predictability. *Journal of International Money and Finance*, Elsevier, v. 30, n. 5, p. 877–895, 2011.

- LUSTIG, H.; VERDELHAN, A. The cross section of foreign currency risk premia and consumption growth risk. *The American Economic Review*, American Economic Association, v. 97, n. 1, p. 89–117, 2007.
- LUX, T.; MARCHESI, M. Volatility clustering in financial markets: a microsimulation of interacting agents. *International journal of theoretical and applied finance*, World Scientific, v. 3, n. 04, p. 675–702, 2000.
- MAJHI, B.; ROUT, M.; MAJHI, R.; PANDA, G.; FLEMING, P. J. New robust forecasting models for exchange rates prediction. *Expert Systems with Applications*, Elsevier BV, v. 39, n. 16, p. 12658–12670, Nov 2012. ISSN 0957-4174.
- MAJHI, R.; PANDA, G.; SAHOO, G. Efficient prediction of exchange rates with low complexity artificial neural network models. *Expert Systems with Applications*, Elsevier BV, v. 36, n. 1, p. 181–189, Jan 2009. ISSN 0957-4174.
- MARK, N. C. Exchange rates and fundamentals: Evidence on long-horizon predictability. *The American Economic Review*, JSTOR, p. 201–218, 1995.
- MEESE, R. A.; ROGOFF, K. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, Elsevier, v. 14, n. 1, p. 3–24, 1983.
- MEESE, R. A.; ROSE, A. K. An empirical assessment of non-linearities in models of exchange rate determination. *Review of Economic Studies*, v. 58, n. 3, p. 603–619, 1991.
- MENKHOFF, L. High-frequency analysis of foreign exchange interventions: What do we learn? *Journal of Economic Surveys*, Wiley Online Library, v. 24, n. 1, p. 85–112, 2010.
- MERCER, J. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London*, v. 209, n. A, p. 415–446, 1909.
- MERCIER, G.; LENNON, M. Support vector machines for hyperspectral image classification with spectral-based kernels. **IEEE International Geoscience and Remote Sensing Symposium. Proceedings**. [S.l.]: Institute of Electrical & Electronics Engineers (IEEE), 2003. ISBN <http://id.crossref.org/isbn/0-7803-7929-2>.
- MICCHELLI, C. A. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constructive Approximation*, Springer Science + Business Media, v. 2, n. 1, p. 11–22, Dec 1986. ISSN 1432-0940.
- MITCHELL, J.; SMITH, R. J.; WEALE, M. R.; WRIGHT, S.; SALAZAR, E. L. An indicator of monthly gdp and an early estimate of quarterly gdp growth. *The Economic Journal*, v. 115, p. F108–F129, Feb 2005.
- MOHRI, M. M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of Machine Learning*. [S.l.]: MIT Press, 2012.
- MONTAVON, G.; BRAUN, M. L.; MÜLLER, K.-R. Kernel analysis of deep networks. *The Journal of Machine Learning Research*, JMLR. org, v. 12, p. 2563–2581, 2011.
- MOOSA, I.; BURNS, K. The unbeatable random walk in exchange rate forecasting: Reality or myth? *Journal of Macroeconomics*, v. 40, p. 69–81, 2014.

- MURIEL, A. Short-term predictions in forex trading. *Physica A: Statistical Mechanics and its Applications*, Elsevier BV, v. 344, n. 1-2, p. 190–193, Dec 2004. ISSN 0378-4371.
- NASSIRTOUSSI, A. K.; AGHABOZORGI, S.; WAH, T. Y.; NGO, D. C. L. Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. *Expert Systems with Applications*, Elsevier BV, v. 42, n. 1, p. 306–324, Jan 2015. ISSN 0957-4174.
- NASSIRTOUSSI, A. K.; WAH, T. Y.; LING, D. N. C. A novel FOREX prediction methodology based on fundamental data. *African Journal of Business Management*, Academic Journals, v. 5, n. 20, p. 8322–8330, 2011.
- NEELY, C. A survey of exchange rate announcements on foreign exchange returns. *Fed. Reserve Bank St. Louis Rev.*, v. 92, p. 417–463, 2010.
- NGUYEN, H.; FAFF, R.; MARSHALL, A. Exchange rate exposure, foreign currency derivatives and the introduction of the euro: French evidence. *International Review of Economics & Finance*, Elsevier BV, v. 16, n. 4, p. 563–577, Jan 2007. ISSN 1059-0560.
- NI, H.; YIN, H. Exchange rate prediction using hybrid neural networks and trading indicators. *Neurocomputing*, Elsevier BV, v. 72, n. 13-15, p. 2815–2823, Aug 2009. ISSN 0925-2312.
- NOOTYASKOOL, S.; CHOENGTONG, W. Hidden markov models predict foreign exchange rate. **International Symposium on Communications and Information Technologies (ISCIT)**. [S.I.], 2014. p. 99–101.
- OZTURK, M.; TOROSLU, I. H.; FIDAN, G. Heuristic based trading system on forex data using technical indicator rules. *Applied Soft Computing*, Elsevier BV, v. 43, p. 170–186, Jun 2016. ISSN 1568-4946.
- PAVLIDIS, E. G.; PAYA, I.; PEEL, D. A. Testing for linear and nonlinear granger causality in the real exchange rate-consumption relation. *Economics Letters*, Elsevier BV, v. 132, p. 13–17, Jul 2015. ISSN 0165-1765.
- PHILIP, A. A.; TAOFIKI, A. A.; BIDEMI, A. A. Artificial neural network model for forecasting foreign exchange rate. *World of computer science and information technology Journal*, v. 1, n. 3, p. 110–118, 2011.
- PIERDZIOCH, C.; RÜLKE, J.-C. On the directional accuracy of forecasts of emerging market exchange rates. *International Review of Economics & Finance*, Elsevier BV, v. 38, p. 369–376, Jul 2015. ISSN 1059-0560.
- PILLONETTO, G.; DINUZZO, F.; CHEN, T.; NICOLAO, G. D.; LJUNG, L. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, Elsevier, v. 50, n. 3, p. 657–682, 2014.
- PILLONETTO, G.; NICOLAO, G. D. A new kernel-based approach for linear system identification. *Automatica*, Elsevier BV, v. 46, n. 1, p. 81–93, Jan 2010. ISSN 0005-1098.
- PINCAK, R. The string prediction models as invariants of time series in the forex market. *Physica A: Statistical Mechanics and its Applications*, Elsevier BV, v. 392, n. 24, p. 6414–6426, Dec 2013. ISSN 0378-4371.

- PREMANODE, B.; TOUMAZOU, C. Improving prediction of exchange rates using Differential EMD. *Expert Systems with Applications*, Elsevier BV, v. 40, n. 1, p. 377–384, Jan 2013. ISSN 0957-4174.
- QI, M.; WU, Y. Technical trading-rule profitability, data snooping, and reality check: Evidence from the foreign exchange market. *Journal of Money, Credit and Banking*, v. 38, n. 8, p. 2135–2158, Dec 2006.
- RAPACH, D. E.; WOCHAR, M. E. The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior. *International Journal of Forecasting*, v. 22, n. 2, p. 341–361, 2006.
- RAVISANKAR, P.; RAVI, V.; RAO, G. R.; BOSE, I. Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, Elsevier BV, v. 50, n. 2, p. 491–500, Jan 2011. ISSN 0167-9236.
- REDL, C. Noisy news and exchange rates: A SVAR approach. *Journal of International Money and Finance*, Elsevier BV, v. 58, p. 150–171, Nov 2015. ISSN 0261-5606.
- REGNAULT, J. *Calcul des chances et philosophie de la bourse*. [S.l.]: Mallet-Bachelier, 1863.
- REHMAN, M.; KHAN, G. M.; MAHMUD, S. A. Foreign currency exchange rates prediction using cgp and recurrent neural network. *IERI Procedia*, Elsevier BV, v. 10, p. 239–244, 2014. ISSN 2212-6678.
- RIME DAGFINN, S. L.; SOJLI, E. Exchange rate forecasting, order flow and macroeconomic information. *Journal of International Economics*, v. 80, n. 1, p. 72–88, 2010.
- RIVERA-PIZA, H. *Towards the Forex Advisor: a neural network model for forecasting foreign exchange rates*. [S.l.: s.n.], 1996.
- RODRIK, D. The past, present, and future of economic growth. *Challenge*, Informa UK Limited, v. 57, n. 3, p. 5–39, Apr 2014. ISSN 0577-5132.
- ROGOFF, K. S.; STAVRAKEVA, V. *The Continuing Puzzle of Short Horizon Exchange Rate Forecasting*. [S.l.], 2008.
- ROSSI, B. Exchange rate predictability. *Journal of Economic Literature*, American Economic Association, v. 51, n. 4, p. 1063–1119, Dec 2013. ISSN 0022-0515.
- ROUT, M.; MAJHI, B.; MOHAPATRA, U. M. Efficient long range prediction of exchange rates using radial basis function neural network models. **2012 International Conference on Advances in Engineering, Science and Management (ICAESM)**. [S.l.], 2012. p. 530–535.
- RUNGE, C. Über empirische funktionen und die interpolation zwischen äquidistanten ordinaten. *Zeitschrift für Mathematik und Physik*, v. 46, p. 224–243, 1901.
- SAGER, M. J.; TAYLOR, M. P. Under the microscope: the structure of the foreign exchange market. *International Journal of Finance & Economics*, Wiley-Blackwell, v. 11, n. 1, p. 81–95, Jan 2006. ISSN 1099-1158.
- SANTOS, A. A. P.; COSTA, N. C. A. da; COELHO, L. dos S. Computational intelligence approaches and linear models in case studies of forecasting exchange rates. *Expert Systems with Applications*, Elsevier, v. 33, n. 4, p. 816–823, 2007.

- SCHINASI, G. J.; SWAMY, P. A. V. B. The out-of-sample forecasting performance of exchange rate models when coefficients are allowed to change. *Journal of International Money and Finance*, v. 8, n. 3, p. 375–390, 1989.
- SCHÖLKOPF, B.; SMOLA, A. J. *Learning with kernels: Support Vector Machines, Regularization, Optimization and Beyond*. [S.l.]: The MIT Press, 2002.
- SERMPINIS, G.; DUNIS, C.; LAWS, J.; STASINAKIS, C. Forecasting and trading the EUR/USD exchange rate with stochastic Neural Network combination and time-varying leverage. *Decision Support Systems*, Elsevier BV, v. 54, n. 1, p. 316–329, Dec 2012. ISSN 0167-9236.
- SERMPINIS, G.; STASINAKIS, C.; THEOFILATOS, K.; KARATHANASOPOULOS, A. Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms–Support vector regression forecast combinations. *European Journal of Operational Research*, Elsevier BV, v. 247, n. 3, p. 831–846, Dec 2015. ISSN 0377-2217.
- SEWELL, M. Characterization of financial time series. *RN*, v. 11, n. 01, p. 01, 2011.
- SHEN, F.; CHAO, J.; ZHAO, J. Forecasting exchange rate using deep belief networks and conjugate gradient method. *Neurocomputing*, Elsevier BV, v. 167, p. 243–253, Nov 2015. ISSN 0925-2312.
- SHI, Y.; HO, K.-Y. Long memory and regime switching: A simulation study on the markov regime-switching arfima model. *Journal of Banking & Finance*, Elsevier, v. 61, p. S189–S204, 2015.
- SIMIŃSKI, K. Neuro-fuzzy system based kernel for classification with support vector machines. *Man-Machine Interactions 3*, Springer Science + Business Media, p. 415–422, 2014. ISSN 2194-5365.
- SIMPSON, M. W.; GROSSMANN, A. An examination of the forward prediction error of U.S. dollar exchange rates and how they are related to bid-ask spreads, purchasing power parity disequilibria, and forward premium asymmetry. *The North American Journal of Economics and Finance*, Elsevier BV, v. 28, p. 221–238, Apr 2014. ISSN 1062-9408.
- SLAVTCHEVA, D. Financial development, exchange rate regimes and productivity growth: Theory and evidence. *Journal of Macroeconomics*, Elsevier BV, v. 44, p. 109–123, Jun 2015. ISSN 0164-0704.
- SOMAN, K. P.; LOGANATHAN, R.; AJAY, V. *Machine Learning with SVM and Other Kernel Methods*. [S.l.]: PHI Learning Private Limited, 2011.
- STEINWART, I.; HUSH, D.; SCOVEL, C. An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *IEEE Transactions on Information Theory*, Institute of Electrical and Electronics Engineers (IEEE), v. 52, n. 10, p. 4635–4643, Oct 2006. ISSN 0018-9448.
- SUN, J.; LI, H. Financial distress prediction using support vector machines: Ensemble vs. individual. *Applied Soft Computing*, Elsevier BV, v. 12, n. 8, p. 2254–2265, Aug 2012. ISSN 1568-4946.

- TALEBI, H.; HOANG, W.; GAVRILOVA, M. L. Multi-scale foreign exchange rates ensemble for classification of trends in forex market. *Procedia Computer Science*, Elsevier BV, v. 29, p. 2065–2075, 2014. ISSN 1877-0509.
- TAY, F. E.; CAO, L. Application of support vector machines in financial time series forecasting. *Omega*, Elsevier BV, v. 29, n. 4, p. 309–317, Aug 2001. ISSN 0305-0483.
- TIAN, W.; WANG, T.; LI, B. Risk measures with wang transforms under flexible skew-generalized settings. *International Journal of Intelligent Technologies and Applied Statistics*, v. 7, n. 3, p. 185–205, Sep 2014.
- VAPNIK, V. N. *The nature of statical learning theory*. [S.I.]: Springer Verlag, 1995.
- WERE, K.; BUI, D. T.; DICK, O. B.; SINGH, B. R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an afromontane landscape. *Ecological Indicators*, Elsevier BV, v. 52, p. 394–403, May 2015. ISSN 1470-160X.
- WHITE, H. A reality check for data snooping. *Econometrica*, v. 68, n. 5, p. 1097–1126, Sep 2000.
- WIDROW, B.; MCCOOL, J.; LARIMORE, M. G.; JR, C. R. J. Stationary and nonstationary learning characteristics of the lms adaptive filter. In: *Aspects of Signal Processing*. [S.I.]: Springer, 1977. p. 355–393.
- WIELAND, V.; WOLTERS, M. H. et al. Forecasting and policy making. *Handbook of economic forecasting*, North-Holland Amsterdam, v. 2, p. 239–325, 2013.
- WOLFE, P. A duality theorem for non-linear programming. *Quarterly of Applied Mathematics*, v. 19, n. 3, p. 239–244, 1961.
- WOLFF, C. C. P. Time-varying parameters and the out-of-sample forecasting performance of structural exchange rate models. *Journal of Business and Economic Statistics*, v. 5, n. 1, p. 87–97, 1987.
- WRIGHT, J. H. Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics*, v. 146, n. 2, p. 329–341, 2008.
- XIA, M.; XU, Z.; ZHU, B. Some issues on intuitionistic fuzzy aggregation operators based on archimedean t-conorm and t-norm. *Knowledge-Based Systems*, Elsevier, v. 31, p. 78–88, 2012.
- YAO, C.-Z.; LIN, J.-N.; ZHENG, X.-Z.; LIU, X.-F. The study of RMB exchange rate complex networks based on fluctuation mode. *Physica A: Statistical Mechanics and its Applications*, Elsevier BV, v. 436, p. 359–376, Oct 2015. ISSN 0378-4371.
- YEN, S. M.-F.; HSU, Y.-L. Profitability of technical analysis in financial and commodity futures markets A reality check. *Decision Support Systems*, v. 50, p. 128–139, 2010.
- YIN, W.; LI, J. Macroeconomic fundamentals and the exchange rate dynamics: A no-arbitrage macro-finance approach. *Journal of International Money and Finance*, Elsevier BV, v. 41, p. 46–64, Mar 2014. ISSN 0261-5606.

- YIP, W. H.; NGUYEN, H. Exchange rate exposure and the use of foreign currency derivatives in the Australian resources sector. *Journal of Multinational Financial Management*, Elsevier BV, v. 22, n. 4, p. 151–167, Oct 2012. ISSN 1042-444X.
- YUAN, Y. Forecasting the movement direction of exchange rate with polynomial smooth support vector machine. *Mathematical and Computer Modelling*, Elsevier BV, v. 57, n. 3-4, p. 932–944, Feb 2013. ISSN 0895-7177.
- ZHANG, H. J.; DUFOUR, J.-M.; GALBRAITH, J. W. Exchange rates and commodity prices: Measuring causality at multiple horizons. *Journal of Empirical Finance*, Elsevier BV, v. 36, p. 100–120, Mar 2016. ISSN 0927-5398.
- ZHANG, Z.; ZHAO, Q. The application of SVMs method on exchange rates fluctuation. *Discrete Dynamics in Nature and Society*, Hindawi Publishing Corporation, v. 2009, p. 1–8, 2009. ISSN 1607-887X.
- ZHOU, D.-X. Capacity of reproducing kernel spaces in learning theory. *IEEE Transactions on Information Theory*, Institute of Electrical and Electronics Engineers (IEEE), v. 49, n. 7, p. 1743–1752, Jul 2003. ISSN 0018-9448.