

UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS  
DEPARTAMENTO DE MATEMÁTICA

INSTITUTO SUPERIOR DE CIÊNCIAS  
DO TRABALHO E DA EMPRESA  
DEPARTAMENTO DE FINANÇAS



**Ciências**  
**ULisboa**

**ISCTE**  **IUL**  
**Instituto Universitário de Lisboa**

Previsão de Séries Temporais Financeiras: Uma Abordagem com  
*Long Short-term Memory Deep Neural Networks*

Mestrado Matemática Financeira

Cristiana Costa de Almeida

Dissertação orientada por: Professora Doutora Diana Mendes

2019



## Agradecimentos

Antes demais, quero agradecer imenso aos meus familiares, principalmente à minha mãe, ao meu irmão e à minha avó, pelo apoio incondicional que me deram, por nunca me deixarem desistir dos meus sonhos e por proporcionarem a realização dos mesmos.

Também um grande obrigado a todos os meus amigos de infância, que não me acompanharam na faculdade, mas que mesmo assim sempre estiveram do meu lado e que ajudaram no que foi possível, sem exceções. Mas saliento, a Andrea Ceia, a Renata Assis, a Susana Soares, a Inês Salvador e o Mauro Damásio, pela paciência que tiveram em passar horas a ajudar-me e a ouvir-me.

Agradeço também aos meus parceiros de faculdade, tanto de licenciatura como de mestrado, pela ajuda que me proporcionaram tanto na faculdade como fora dela, pois foram um dos grandes alicerces para que isto fosse possível, salientando assim a Rita Ramalho, a Diana Salpico, a Angelina Boyarko, a Jossline Correia, o Tiago Jacinto e o Gustavo Mendes.

Ainda quero deixar um profundo agradecimento às minhas colegas de residência, principalmente à Patrícia Carvalho, à Sara Lopes e à Daniela Richicinski, pelo apoio que me deram desde o início, que foi essencial para a minha chegada até aqui. Sem esquecer de agradecer à governanta Dona Palmira Santos pelas longas conversas de apoio e motivação.

Um sincero obrigado também ao meu companheiro e amigo Fábio Chaves, pelo amor, pela paciência que teve nos muitos momentos de stress e por sempre conseguir fazer ver as coisas de outra forma, nunca deixando baixar os braços.

Quero agradecer à minha orientadora Professora Doutora Diana Mendes pelo apoio, pela paciência que teve em responder a todas as minhas dúvidas e pela motivação que me proporcionou na realização desta dissertação.

Por último, quero deixar um profundo obrigado a todas as pessoas que por mim passaram e deixaram um marco na minha vida, que também tornaram possível a minha chegada até aqui.



## Resumo

O intuito desta dissertação é identificar o método que proporciona resultados mais fidedignos na previsão de séries temporais financeiras, de entre os modelos tradicionais e as novas metodologias de *Machine Learning*.

A fim de alcançar esse objetivo, foi considerada uma base de dados com 4842 valores diários do fecho do índice bolsista *Standard & Poor's 500* (SP500), no intervalo temporal compreendido entre 3 de janeiro de 2000 e 1 de abril de 2019, excluindo os fins-de-semana e feriados.

Primeiramente, os dados foram trabalhados no programa *Eviews*, de forma a obter o modelo econométrico ARIMA mais adequado e através da metodologia de Box-Jenkins procedeu-se para a previsão da série financeira em estudo. Contudo, para obtenção de melhores resultados, a série foi reduzida, iniciando no dia 2 de março de 2009, devido à crise financeira de 2007-2008.

Para comparação de modelos, foi realizada a previsão dos dados da bolsa SP500, através de um modelo *Deep Neural Network – Long Short-Term Memory* (código programado em *Python*). Conclui-se que os melhores resultados de previsão foram obtidos com os modelos de redes neuronais, tanto para curto como para longo-prazo.

**Palavras-chave:** Série temporal financeira; Mercados Financeiros; *Standard & Poor's 500*; *Machine Learning*; *Deep Learning*; *Long Short-Term Memory*; Modelos Autorregressivos de Médias Móveis; Previsão.



## **Abstract**

The aim of this dissertation is to identify the method that provides the most reliable results in forecasting financial time series, between the traditional models and the new Machine Learning methodologies.

In order to achieve this objective, were considered 4842 daily closing values of the Standard & Poor's 500 (SP500) stock index, since January 3, 2000 until April 1, 2019, excluding weekends and holidays.

First, we use Eviews software, in order to obtain the right ARIMA econometric model, and by using the Box-Jenkins methodology, we forecast the SP500 financial time series. However, for best results, the series was split, starting on March 2, 2009, due to the 2007-2008 financial crisis.

For model comparison, was realized the prediction of the SP500 stock index data, using a Deep Neural Network - Long Short-Term Memory model (code programmed in Python). It is concluded that the best results were obtained with neural network models, for both short and long term forecast.

**Keywords:** Financial Time Series; Financial Market; Standard & Poor's 500; Machine Learning; Deep Learning; Long Short-Term Memory; Auto-Regressive Moving Average Models; Forecast.



## Abreviatura

SP500            *Standard & Poor's 500*

### **Econometria**

ACF	Função de Autocorrelação
ADF	Dickey – Fuller Aumentado
AIC	Critério de Akaike
AR	Autorregressivos
ARIMA	Autorregressivos Integrados de Médias Móveis
ARMA	Autorregressivos de Médias Móveis
B-J	Box – Jenkins
DF	Dickey – Fuller
DSP	<i>Difference Stationary Process</i>
HQC	Critério de Hannan – Quinn
JB	Jarque – Bera
KPSS	Kwiatkowski – Phillips – Schmidt – Shin
MA	Média Móvel
MAE	Erro Absoluto Médio
MAPE	Erro Absoluto Percentual Médio
ME	Erro Médio
MPE	Erro Percentual Médio
MSE	Erro Quadrático Médio
PACF	Função de Autocorrelação Parcial
PP	Phillips – Perron
RMSE	Raiz do Erro Quadrático Médio
SC	Critério de Schwarz Bayesiana
TSP	<i>Trend Stationary Process</i>

### ***Machine Learning***

AE	Codificadores Automáticos
BPNN	<i>Backpropagation Neural Network</i>
CPU	Unidade Central de Processamento

DAE	Denoising Autoencoder
DBF	<i>Deep Belief Network</i>
DL	<i>Deep Learning</i>
DNN	<i>Deep Neural Network</i>
FNN	<i>Feed-forward Neural Network</i>
GA	Algoritmo Genético
GPU	Unidade Processamento Gráfico
GRU	<i>Gated Recurrent Unit</i>
LSSVM	<i>Least Squares Support Vector Machines</i>
LSTM	<i>Long Short-Term Memory</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
ReLU	Unidades Lineares Retificadas
RNA	Rede Neural Artificial
RNC	Rede Neural de Convolução
RNR	Rede Neural Recorrente
SGD	Gradiente Descendente Estocástico
SVM	<i>Support Vector Machine</i>



# Índice

Agradecimentos .....	I
Resumo .....	III
Abstract.....	V
Abreviatura .....	VII
Índice .....	III
Índice de Figuras.....	VI
Índice de Tabelas.....	X
Introdução.....	1
1. Revisão de Literatura .....	4
1.1. Breve introdução a <i>Machine Learning</i> .....	4
1.2. Noções de Mercados Financeiros .....	5
2. Conceitos Econométricos.....	8
2.1. Séries Temporais .....	8
2.2. Processos Estocásticos.....	10
2.3. Processos para obter a estacionariedade.....	12
2.4. Modelos de Séries Temporais Univariadas.....	14
a. Modelos Autorregressivos – $AR(p)$ .....	14
b. Modelos Média Móvel - $MA(q)$ .....	15
c. Modelos Autorregressivos de Médias Móveis – $ARMA(p, q)$ .....	16
d. Modelos Autorregressivos Integrados de Médias Móveis – $ARIMA(p, d, q)$ .....	17
2.5. Critérios de Comparação de Modelos .....	17
2.6. Validação dos Modelos .....	18
2.7. Previsão.....	19
3. <i>Machine Learning</i> .....	22
3.1. Redes Neurais Artificiais (RNAs) .....	22
3.2. Support Vector Machines (SVMs).....	24
3.3. Deep Learning (DL).....	25
a. Redes Neurais Recorrentes.....	27
4. Estudo Empírico.....	32
4.1. Análise da Série Standard & Poor’s 500 .....	32
4.2. Previsão com Modelos Econométricos – Série Global .....	33
a. Estacionariedade .....	34
b. Análise Gráfica e Estatística dos retornos .....	34

c.	Estacionariedade da Série dos retornos .....	35
d.	Modelização da Série dos retornos .....	36
e.	Previsão da série .....	38
4.3.	Previsão com Modelos Econométricos – Série Reduzida.....	40
a.	Estacionariedade .....	41
b.	Análise Gráfica e Estatística dos Retornos .....	42
c.	Estacionariedade .....	43
d.	Modelização da Série dos retornos .....	43
e.	Previsão da série.....	45
4.4.	Síntese – Previsão com Modelos Econométricos .....	46
4.5.	Previsão com <i>Deep Neural Networks – Long Short Term Memory</i> .....	48
a.	Metodologia .....	48
b.	Modelo experimental.....	51
c.	Resultados obtidos.....	52
4.6.	Previsão com <i>Deep Neural Networks – Long Short Term Memory (Reduzida)</i> .....	54
a.	Modelo experimental .....	54
b.	Resultados obtidos.....	55
4.7.	Síntese – Previsão com <i>Deep Neural Networks</i> .....	56
4.8.	Comparação entre Modelos .....	56
Conclusão.....		59
Referências Bibliográficas .....		61
Anexos .....		63
Anexo A – Testes de Estacionariedade da Série Global.....		63
Anexo B – Testes de Estacionariedade da Série Global dos Retornos.....		64
Anexo C – Modelização da Série Global dos Retornos .....		65
Anexo D – Testes de Correlação e Heterocedasticidade da Série Global dos Retornos .		67
Anexo E – Previsão da Série Global dos Retornos .....		68
Anexo F – Testes de Estacionariedade da Série Reduzida .....		71
Anexo G – Testes de Estacionariedade da Série Reduzida dos Retornos .....		72
Anexo H – Modelização da Série Reduzida dos Retornos .....		73
Anexo I – Testes de Correlação e Heterocedasticidade da Série Reduzida dos Retornos		74
.....		
Anexo J – Previsão da Série Reduzida dos Retornos .....		75
Anexo K – Código e Experiências.....		76
Anexo L– Previsão da Série com DL - LSTM.....		82



# Índice de Figuras

Figura 1 – Ilustração da relação entre Inteligência Artificial, Machine Learning e Deep Learning.....	1
Figura 2 - Exemplo de rede LSTM.....	31
Figura 3 - Série global do índice Standard & Poor's 500 .....	32
Figura 4 - Gráfico de barras e estatísticas descritivas da série global do índice Standard & Poor's 500.....	33
Figura 5 - Série global dos retornos do índice Standard & Poor's 500 .....	35
Figura 6 - Gráfico de barras e estatísticas descritivas da série global dos retornos do índice Standard & Poor's 500 .....	35
Figura 7 - Correlograma da série global dos retornos do índice Standard & Poor's 500 ..	37
Figura 8 - Previsão out-of-sample do dia 2/4/2019 da série global dos retornos da SP500, com o modelo AR(1).....	39
Figura 9 - Série reduzida do índice Standard & Poor's 500.....	40
Figura 10 - Gráfico de barras e estatísticas descritivas da série reduzida do índice Standard & Poor's 500 .....	41
Figura 11 - Série reduzida dos retornos do índice Standard & Poor's.....	42
Figura 12 - Gráfico de barras e estatísticas descritivas da série reduzida dos retornos do índice Standard & Poor's 500 .....	42
Figura 13 - Correlograma da série reduzida dos retornos do índice Standard & Poor's 500 .....	44
Figura 14 - Previsão out-of-sample da série reduzida dos retornos da SP500, com o modelo MA(1) .....	46
Figura 15 - Ilustração de Redes Neurais Recorrentes possuem loops .....	48
Figura 16 - Ilustração de uma rede neural recorrente desenrolada .....	49
Figura 17 - Módulo de repetição em um RNN padrão contém uma única camada .....	49
Figura 18 - Módulo de repetição em um LSTM contém quatro camadas de interação.....	49
Figura 19 - Estado da célula de um LSTM .....	50
Figura 20 - Camada de rede neural sigmoide.....	50
Figura 21 - Previsão do índice bolsista SP500 com o modelo DNN.....	53
Figura 22 - Previsão da série reduzida do índice bolsista SP500 com o modelo DNN .....	55
Figura 23 - Gráfico de dispersão dos valores previstos com o modelo MA(1) e os valores reais .....	57
Figura 24 - Gráfico de dispersão dos valores previstos (-1) com o modelo MA(1) e os valores reais .....	57
Figura 25 - Previsão out-of-sample de 507 valores com o modelo ARMA .....	58
Figura 26 - Previsão out-of-sample com intervalo de confiança de 95% .....	58
Figura 27 - Output do teste de estacionariedade ADF da série global do índice SP500 .....	63
Figura 28 - Output do teste de estacionariedade PP da série global do índice SP500.....	63
Figura 29 - Output do teste de estacionariedade KPSS da série global do índice SP500....	63
Figura 30 - Output do teste de estacionariedade ADF da série global dos retornos do índice SP500 .....	64

Figura 31 - Output do teste de estacionariedade PP da série global dos retornos do índice SP500 .....	64
Figura 32 - Output do teste de estacionariedade KPSS da série global dos retornos do índice SP500.....	64
Figura 33 - Correlograma da série global dos retornos do índice SP500.....	65
Figura 34 - Output do modelo AR(1) da série global dos retornos do índice SP500 .....	65
Figura 35 - Output do modelo AR(2) da série global dos retornos do índice SP500 .....	65
Figura 36 - Output do modelo MA(1) da série global dos retornos do índice SP500 .....	66
Figura 37 - Output do modelo MA(2) da série global dos retornos do índice SP500 .....	66
Figura 38 - Output do modelo ARMA(1,1) da série global dos retornos do índice SP500 .....	66
Figura 39 - Output do modelo ARMA(1,2) da série global dos retornos do índice SP500 .....	66
Figura 40 - Output do modelo ARMA(2,1) da série global dos retornos do índice SP500 .....	67
Figura 41 - Output do modelo ARMA(2,2) da série global dos retornos do índice SP500 .....	67
Figura 42 - Output do teste de Heterocedasticidade dos resíduos da série global dos retornos do índice SP500 .....	67
Figura 43 - Output do teste de correlação dos resíduos da série global dos retornos do índice SP500.....	67
Figura 44 - Previsão in-sample com o modelo AR(1) da série global dos retornos do índice SP500 .....	68
Figura 45 - Previsão in-sample com o modelo AR(2) da série global dos retornos do índice SP500 .....	68
Figura 46 - Previsão in-sample com o modelo MA(1) da série global dos retornos do índice SP500 .....	68
Figura 47 - Previsão in-sample com o modelo MA(2) da série global dos retornos do índice SP500 .....	69
Figura 48 - Previsão in-sample com o modelo ARMA(1,1) da série global dos retornos do índice SP500.....	69
Figura 49 - Previsão in-sample com o modelo ARMA(1,2) da série global dos retornos do índice SP500.....	69
Figura 50 - Previsão in-sample com o modelo ARMA(2,1) da série global dos retornos do índice SP500.....	70
Figura 51 - Previsão in-sample com o modelo ARMA(2,2) da série global dos retornos do índice SP500.....	70
Figura 52 - Output do teste de estacionariedade ADF da série reduzida do índice SP500.....	71
Figura 53 - Output do teste de estacionariedade PP da série reduzida do índice SP500 .....	71
Figura 54 - Output do teste de estacionariedade KPSS da série reduzida do índice SP500.....	71
Figura 55 - Output do teste de estacionariedade ADF da série reduzida dos retornos do índice SP500.....	72
Figura 56 - Output do teste de estacionariedade PP da série reduzida dos retornos do índice SP500.....	72
Figura 57 - Output do teste de estacionariedade KPSS da série reduzida dos retornos do índice SP500.....	72
Figura 58 - Correlograma da série reduzida dos retornos do índice SP500 .....	73
Figura 59 - Output do modelo AR(1) da série reduzida dos retornos do índice SP500 .....	73
Figura 60 - Output do modelo MA(1) da série reduzida dos retornos do índice SP500 .....	73

<b>Figura 61 - Output do modelo ARMA(1,1) da série reduzida dos retornos do índice SP500</b>	<b>74</b>
<b>Figura 62 - Output do teste de Heterocedasticidade dos resíduos da série reduzida dos retornos do índice SP500</b>	<b>74</b>
<b>Figura 63 - Output do teste de correlação dos resíduos da série reduzida dos retornos do índice SP500</b>	<b>74</b>
<b>Figura 64 - Previsão in-sample com o modelo AR(1) da série reduzida dos retornos do índice SP500</b>	<b>75</b>
<b>Figura 65 - Previsão in-sample com o modelo MA(1) da série reduzida dos retornos do índice SP500</b>	<b>75</b>
<b>Figura 66 - Previsão in-sample com o modelo ARMA(1,1) da série reduzida dos retornos do índice SP500</b>	<b>75</b>



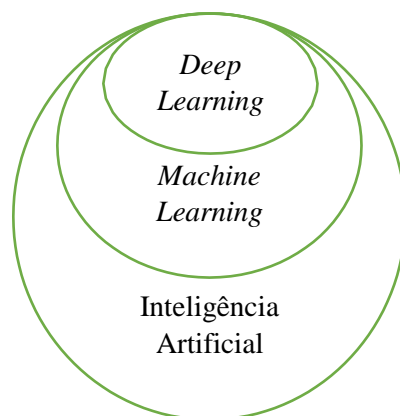
## Índice de Tabelas

Tabela 1 - Resultados dos testes de estacionariedade da série global do índice Standard & Poor's 500.....	34
Tabela 2 - Resultados dos testes de estacionariedade da série global dos retornos do índice Standard & Poor's 500 .....	36
Tabela 3 - Valores dos critérios de informação de todos os modelos potenciais da série global dos retornos do índice Standard & Poor's 500.....	37
Tabela 4 - Resultados dos testes de correlação e de heterocedasticidade da série global dos retornos do índice Standard & Poor's 500.....	38
Tabela 5 - Valores RMSE, MAE e MAPE de todos os modelos potenciais da série global dos retornos do índice Standard & Poor's 500.....	39
Tabela 6 - Valores Bias Proportion e Covariance Proportion dos modelos AR(1) e ARMA(1,2) da série global dos retornos do índice Standard & Poor's 500 .....	39
Tabela 7 - Resultados dos testes de estacionariedade da série reduzida do índice Standard & Poor's 500 .....	41
Tabela 8 - Resultados dos testes de estacionariedade da série reduzida dos retornos do índice Standard & Poor's 500 .....	43
Tabela 9 - Valores dos critérios de informação de todos os modelos potenciais da série reduzida dos retornos do índice Standard & Poor's 500 .....	44
Tabela 10 - Resultados dos testes de correlação e de heterocedasticidade da série reduzida dos retornos do índice Standard & Poor's 500.....	45
Tabela 11 - Valores RMSE, MAE e MAPE de todos os modelos potenciais da série reduzida dos retornos do índice Standard & Poor's 500 .....	45
Tabela 12 - Valores RMSE, MAE e MAPE dos modelos do índice Standard & Poor's 500 .....	47
Tabela 13 - O Coeficiente de Theil, Bias Proportion, Variance Proportion e Covariance Proportion dos respectivos modelos da série global e da série reduzida.....	47
Tabela 14 - Tabela síntese dos resultados de previsão out-of-sample no dia 2/3/2019 do índice SP500 com os modelos econométricos .....	47
Tabela 15 - Valores obtidos na previsão do índice bolsista SP500 como o modelo DNN....	53
Tabela 16 - Valores obtidos na previsão da série reduzida do índice bolsista SP500 como o modelo DNN.....	55
Tabela 17 - Tabela síntese dos resultados de previsão do índice SP500 com o modelo DNN .....	56
Tabela 18 - Valores previstos, erros de previsão e os valores de MAPE obtidos na previsão de todos os modelos .....	56
Tabela 19 - Resultados obtidos na previsão de 507 valores previsto tanto em MA(1) como em DNN-LSTM .....	58
Tabela 20 - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 7 e dimensionalidade de 256 de espaço de saída.....	76
Tabela 21 - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 6 e dimensionalidade de 256 de espaço de saída.....	76
Tabela 22 - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 6 e dimensionalidade de 256 de espaço de saída.....	77

<b>Tabela 23 - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 9 e dimensionalidade de 256 de espaço de saída.....</b>	<b>77</b>
<b>Tabela 24 - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 8 e dimensionalidade de 255 de espaço de saída.....</b>	<b>78</b>
<b>Tabela 25 - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 8 e dimensionalidade de 254 de espaço de saída.....</b>	<b>78</b>
<b>Tabela 26 - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 8 e dimensionalidade de 257 de espaço de saída.....</b>	<b>79</b>
<b>Tabela 27 - Resultados obtidos na previsão da série com o modelo DNN-LSTM .....</b>	<b>82</b>



## Introdução



**Figura 1** – Ilustração da relação entre Inteligência Artificial, *Machine Learning* e *Deep Learning*

Nestes últimos anos, *Machine Learning*, uma das grandes áreas da Inteligência Artificial, tem sido bastante desenvolvida, principalmente em relação à previsão de séries temporais financeiras, sendo este um dos grandes fatores para tomadas de decisões nas empresas, bancos e outras instituições.

Contudo, devido ao facto da existência de um vasto leque de projetos desenvolvidos nesta área, a comparação entre os métodos tradicionais econométricos com os modelos de *Deep Learning* (sub-área de aprendizagem profunda de *Machine Learning*), para se concluir qual destes retorna resultados fidedignos, é um dos temas bastante debatidos, tanto a nível internacional como a nível nacional.

Esta tese terá um carácter bastante inovador, principalmente a nível nacional, pois tem um tema pioneiro ao nível de mestrado, sendo o seu objetivo principal comparar os métodos tradicionais de previsão com novas metodologias de tipo LSTM (*Long Short Term Memory*) para *Deep Learning*, para dados financeiros, e onde toda a parte computacional foi desenvolvida na linguagem *Python*. A série temporal em estudo é proveniente da bolsa americana, nomeadamente o índice *Standard & Poor's 500*.

No início dos anos 90, Tang, de Almeida, e Fishwick defenderam que para séries temporais com memória longa, o modelo Box-Jenkins e as redes neurais apresentam bom desempenho. No entanto, o modelo Box-Jenkins mostra ser um pouco melhor na previsão de curto prazo e a rede neural na de longo prazo. Mas relativamente às séries de memória curta, são as redes neurais que apresentam melhores resultados.

De seguida, em 1996, através do estudo de Kohzadi, Boyd, Kermanshahi, e Kaastra, comprovou-se que as previsões com redes neurais são consideravelmente mais precisas do que as dos modelos tradicionais ARIMA (*Auto-Regressive Integrated Moving Average*). Além disso, chegaram à conclusão de que a razão por detrás disso, é porque os

dados utilizados no projeto contêm um comportamento não-linear, que não pode ser totalmente capturado pelo modelo ARIMA linear.

Para além disso, Mantovani, (2004) comparou os modelos ARIMA e as redes neurais artificiais, que mostrou que os modelos ARIMA, em grande parte dos casos, reproduzem razoavelmente as características das séries. Contudo, a classe das redes neurais artificiais reproduz de forma mais fidedigna estas características, principalmente em séries não-lineares.

Mais recentemente, Qian e Gao, (2017), chegaram à conclusão de que *Support Vector Machine* (SVM), um dos modelos mais utilizados em *Machine Learning*, é excelente na previsão de movimento de preços, comparativamente aos modelos ARIMA. Além disso, mencionam como sugestão o modelo LSTM, que pode obter um excelente desempenho.

De forma a seguir a sugestão de Qian e Gao, em 2018, Siami-Namini e Namin realizaram um estudo, onde comparam a precisão dos modelos ARIMA e LSTM, ao que indicaram que o modelo LSTM é superior ao modelo ARIMA. Além disso, nesse estudo defendem os benefícios da aplicação de algoritmos e técnicas de *Deep Learning* aos dados económicos e financeiros.

Por conseguinte, de forma a seguir os passos de Saimi-Namini e Namin (2018), o presente trabalho tem como objetivo concluir quais dos modelos ARIMA e LSTM, obtém melhores resultados na previsão de séries temporais financeiras, sendo que neste caso, será utilizado o índice bolsista *Standard & Poor's 500*.

Este projeto inicia-se com uma breve contextualização sobre *Machine Learning*, Econometria e Mercados Financeiros, de modo a facilitar o leitor a compreender um conjunto de conceitos importantes a serem aplicados. Essa contextualização tem como base alguns artigos, mas principalmente o livro “*Deep Learning with Python*” de Kethar (2018).

De seguida, é realizado um estudo empírico que se divide em duas grandes partes, sendo que a primeira é referente à previsão com modelo econométrico ARIMA e a segunda consiste na previsão da série temporal com o modelo *Deep Learning*, em particular com redes de tipo *Long Short-Term Memory* (LSTM).

Relativamente à primeira parte, inicialmente é realizado uma análise gráfica à série financeira e uma avaliação à estacionariedade da série, que ao ser obtida, leva ao procedimento da metodologia de Box-Jenkins. Após isso, é obtido o modelo que melhor se adequa à série e, posteriormente, é feita a respetiva previsão da série.

Referente à previsão do modelo *Deep Learning*, já se procede de forma diferente, pois a previsão da série é feita através de um código de programação na linguagem *Python*. Este código é corrido inúmeras vezes, com o intuito de obter o menor erro de previsão possível, alterando alguns parâmetros, como o número de épocas, a percentagem de dados em treino e em teste, entre outros.

Por fim, é realizada uma síntese, onde são comparados os valores de previsão obtidos, ao qual permite identificar o melhor método para previsão a série temporal financeira em consideração.

# 1. Revisão de Literatura

## 1.1. Breve introdução a *Machine Learning*

De acordo com Ketkar (2017), *Machine Learning* concentra-se no desenvolvimento de algoritmos que ficam melhores numa tarefa ao longo do tempo, ou seja, com a experiência.

Além disso, Henrique *et al.* (2019) referem também que as técnicas desta ciência buscam extrair padrões dos dados históricos – num processo conhecido como *training* ou *learning* para, posteriormente, fazer previsões sobre novos dados. Isto tem vindo a mostrar ao longo destes anos, um excelente desempenho na previsão de algoritmos, na maioria dos quais foram comprovados. Por esta razão, os seus modelos são alguns dos mais pesquisados, principalmente na área das Finanças.

No entanto, as suas técnicas podem ser agrupadas em quatro grupos: um dos grupos é a aprendizagem supervisionada, que transforma os dados em algoritmos para que estes sejam divididos em classes, que inclui as técnicas: *Decision Tree*, Regressão, *Support Vector Machines* e *Naive Bayes*; outro é a aprendizagem não supervisionada, em que a tarefa dos algoritmos é tirar conclusões dos dados sem serem classificados, que engloba as técnicas: *K-Means* e *Gaussian Mixture models*; existe também a aprendizagem de associação, onde se baseia no descobrir relações entre as variáveis de uma grande base de dados; e, por fim, a aprendizagem de reforço, que tem como tarefa encontrar a melhor ação a exercer, de forma a maximizar o retorno.

De entre muitos modelos, os mais utilizados para previsão de séries temporais financeiras envolvem as técnicas de *Support Vector Machines* (SVMs), redes neurais artificiais (RNAs), e em particular redes neurais recorrentes (RNR), que serão as mais focadas neste projeto.

Geralmente, nos modelos de *Machine Learning* o estudo dos dados é dividido em duas fases: a primeira consta na seleção de variáveis e modelos relevantes para a previsão, separando uma parte dos dados para o *training* e validação dos modelos, otimizando-os, e a segunda assenta na aplicação dos modelos aos dados destinados ao teste, medindo assim, o desempenho preditivo.

Em relação ao desempenho preditivo, é importante ter conhecimento de que se pode observar modelos com grau = 2 (melhores) em comparação com modelos de grau = 8 (piores), pois quando se considera que todos os conjuntos de dados contêm algum ruído, um modelo de maior capacidade, ou seja de maior grau, acaba por apenas ajustar o ruído, além do sinal nos dados.

## 1.2. Noções de Mercados Financeiros

A importância da análise dos mercados financeiros advém da sua influência nas tomadas de decisão relativamente aos gastos, tanto a nível pessoal como no mundo dos negócios.

Por sua vez, um dos aspetos mais importantes na análise destes mercados é perceber que tipo de ativos financeiros existem, como por exemplo, futuros, obrigações, opções, títulos de ações, entre outros. Sendo que, estes ativos são transacionados em índices de bolsa específicos.

No entanto, por índice de bolsa entende-se como o valor de determinada carteira de ações de vários setores, que é um indicador estatístico medido em pontos, que reflete o valor médio combinado de ativos, sendo que esse valor médio resulta da evolução de preços de um determinado mercado.

Relativamente às ações, estas constituem títulos de propriedade e representam uma fração do capital das sociedades, que são emitidas pelas empresas cotadas para obter recursos, fazendo face às necessidades de fundos.

Além disso, tendem a desvalorizarem-se ou a valorizarem-se em conjunto com outras ações, logo os índices bolsistas transmitem num só número, o comportamento global do mercado. Dado que, esse número constitui a média das subidas e descidas da cotação dessas ações.

Contudo, numa forma geral, pode-se dividir os índices bolsistas em dois tipos: o índice geral, em que são consideradas todas as ações cotadas numa bolsa, ou o índice de seleção, onde se restringe o índice às ações mais representativas da bolsa, de forma a refletir a sua importância no mercado.

É de notar que, certos fenómenos ocorrentes nos mercados financeiros são anunciados antecipadamente, principalmente a entrada ou saída de uma empresa na bolsa, o que permite aos investidores fazer certas previsões. No entanto que, quando uma empresa está prestes a entrar no índice bolsista, ocorre uma subida da cotação, devido a um aumento na procura da ação.

Contudo, o comportamento do índice ao longo do mercado depende da sua composição, do tipo e número de ações cotadas, da ponderação atribuída às ações e do tipo de média estabelecida.

Em relação à composição, os índices podem ser classificados por meio de dois critérios diferentes: um deles é o procedimento do cálculo da média, que pode ser aritmética ou geométrica; ou pelo sistema de ponderação. Este último critério leva a três tipos de índices: os índices de preço ponderado, em que o índice é proporcional à soma dos preços das ações das empresas que são incluídas na mesma (mais caras as ações da empresa, maior será a sua influência no índice); os índices de capitalização ponderada, no qual o índice é proporcional à capitalização do mercado das empresas (maior a capitalização de uma empresa, mais influência terá a mesma sobre o índice); e os índices

de igual ponderação, onde o índice é cotado com uma fórmula complexa, de modo a que a contribuição de todas as empresas seja igual.

Além disso, a variância de um índice bolsista depende dos diferentes pesos dos ativos, do valor do coeficiente de correlação médio e da volatilidade média de cada ação.

Relativamente à volatilidade do índice, esta consiste na variação dos preços em relação à sua média, durante um determinado intervalo de tempo, dependendo do comportamento das empresas e dos mercados.

Por sua vez, constatamos que o preço dos ativos financeiros oscila devido à lei da “procura e oferta”, ou seja, quando existem mais investidores a comprar ações do que a vender, o preço destas aumenta, enquanto que ocorre o inverso, se existirem mais investidores a vender, pois necessitam de baixar o preço das ações, no intuito de atrair os compradores. Contudo, essa lei sofre alterações devido às tendências, isto é, devido às notícias, à situação económica do país, aos dados económicos ou às previsões futuras de crescimento.

Para além disso, diversos outros fatores relacionados com a conjuntura económica de cada país podem levar a uma maior ou menor volatilidade dos mercados financeiros, como por exemplo, a inflação, as taxas de câmbio, as taxas de juro ou até mesmo de situações de instabilidade política.

Em 1976, Black introduziu o efeito alavanca, que indica que o preço das ações tende a ser negativamente correlacionado com as variações na volatilidade, em que a volatilidade é mais elevada após impactos negativos do que após impactos positivos, com a mesma intensidade. Esta teoria defende que a alavancagem do mercado aumenta, à medida que o valor do índice de ações desce, tornando assim, o capital mais arriscado.

No entanto, Pindyck (1984), French, Schwert e Stambaugh (1987) e Campbell e Hentschel (1992) defendem que o aumento da volatilidade amplia o retorno sobre as ações, causando assim uma queda do preço. E contrariamente ao estudo de Black, esta teoria estabelece que as alterações na volatilidade provocam variações no preço.

Ao longo dos anos, a previsão do preço das ações sempre foi um tema muito debatido, pois devido à sua complexidade sempre estimulou o interesse dos investigadores com o intuito de desenvolver os melhores modelos de previsão.

Contudo, um dos fatores que torna a previsão bastante complexa é o facto de os acontecimentos futuros ser tão incertos, o que leva a que os investidores não tenham plena certeza das suas previsões.

Em 1970, foi desenvolvida a Teoria da Eficiência dos Mercados por Fama, que considera que o mercado é eficiente, quando toda a informação disponível é refletida nos preços. Também nessa teoria é possível serem identificadas três formas de eficiência: a eficiência na forma fraca, que defende que o preço atual representa apenas a informação relativa a todos os preços passados; a eficiência na forma semi-forte, que indica que o preço corrente, para além de refletir a informação referente a todos os preços do passado,

também retrata toda a informação pública; e a eficiência na forma forte, que refere que o preço retrata todas as informações, tanto públicas como privadas, e que nenhum investidor tem acesso a qualquer tipo de informação privilegiada.

Ainda referente à análise dos mercados financeiros, Santos (2015) destaca três grupos: os defensores da análise fundamental, os defensores da análise técnica e os analistas quantitativos, que se destacam dos dois anteriores pela relevância que tem na modelação quantitativa, no âmbito da moderna análise financeira dos mercados e derivados.

Além disso, destacam-se ainda dois grupos: os investigadores das finanças comportamentais, que defendem que existem evidências de anomalias nos mercados, e os investigadores das finanças tradicionais, que rejeitam a hipótese das anomalias, criticando as metodologias seguidas, e advogam que os mercados são eficientes.

## 2. Conceitos Econométricos

Nos últimos anos, vários modelos foram desenvolvidos para prever séries temporais financeiras, tais como os modelos auto-regressivos integrados de médias móveis (ARIMA). Estes modelos foram introduzidos por Box e Jenkins em 1970, (também conhecidos como a metodologia Box-Jenkins (B-J)), que têm como objetivo identificar, estimar e diagnosticar modelos com base em dados de séries temporais.

Além disso, mostram capacidade para gerar previsões eficientes de curto prazo, superando constantemente outros modelos complexos. Sendo que, este método capta as correlações históricas entre os dados e estende-as para períodos futuros. Isto é, quanto mais fortes, homogêneas e estáveis forem as correlações históricas entre os dados melhor será a adequabilidade dos modelos. Nos casos em que os dados são irregulares ou as correlações mudam com o tempo, o método pode extrapolar correlações inadequadas para períodos futuros (Box, Jenkins e Reinsel, (1994)).

Em síntese, os modelos ARIMA modelam a função de autocorrelação de uma série temporal com poucos parâmetros, utilizando uma combinação de termos de auto-regressão (AR), integração (I) e média móvel (MA).

### 2.1. Séries Temporais

Uma série temporal é conhecida como sendo um conjunto de observações efetuadas em instantes sucessivos de tempo, durante um determinado intervalo de tempo, que pode ser caracterizado da seguinte maneira:

$$Y_t(\textit{discreta}), t = 1, 2, \dots, n \quad \textit{ou} \quad Y(t) (\textit{contínua}), t \in I \subseteq \mathbb{R}$$

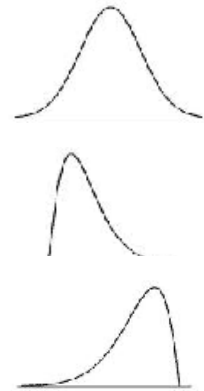
Estas séries podem ser divididas em dois tipos: séries temporais univariadas, que são constituídas por uma única observação em cada instante de tempo, e séries temporais multivariadas, que contêm várias observações em cada instante.

Além disso, podem reger-se de acordo com vários movimentos, que podem ser classificados por: movimentos de tendência, que descrevem uma direção global, segundo a qual a série temporal tende a desenvolver-se num intervalo de tempo longo; movimentos cíclicos, que traçam oscilações a longo prazo em torno da reta ou curva de tendência; movimentos por estações (sazonalidade), que delineiam padrões praticamente idênticos, a que uma série temporal parece obedecer durante os mesmos períodos de tempo sucessivos; movimentos aleatórios, que descrevem deslocamentos esporádicos das sucessões cronológicas, incitadas por acontecimentos aleatórios (eleições, cheias graves, etc.).

## Assimetria

A assimetria indica o grau de afastamento que uma distribuição apresenta do seu eixo de simetria, nomeadamente:

- Distribuição simétrica: Média = Mediana = Moda
- Distribuição assimétrica positiva (à direita): Média  $\geq$  Mediana  $\geq$  Moda
- Distribuição assimétrica negativa (à esquerda): Média  $\leq$  Mediana  $\leq$  Moda



Além disso, pode ser quantificada através do seu coeficiente do momento de assimetria, isto é:

$$s_k = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^3}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}^3$$

onde  $n$  = número de observações da série

$y_i$  = valor registado por esta série no instante  $i$

Após calculado o coeficiente do momento de assimetria, a distribuição é determinada da seguinte maneira, tendo a distribuição normal como termo comparativo:

- $s_k = 0$ , distribuição simétrica
- $s_k > 0$ , distribuição assimétrica positiva
- $s_k < 0$ , distribuição assimétrica negativa

## Curtose

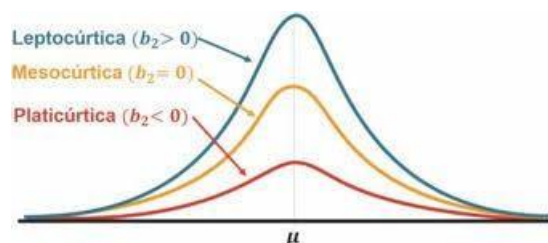
Relativamente à curtose, este termo corresponde ao grau de achatamento da distribuição, ou seja, relaciona o achatamento da curva de frequência com o da curva normal de referência.

Contudo, para avaliar a curtose associada a uma série, calcula-se o coeficiente do momento de curtose, isto é:

$$k = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^4}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}^2$$

De acordo com esse valor associado à série, assume-se que a mesma se rege de acordo com uma das seguintes distribuições:

- $k = 3$ , distribuição mesocúrtica
- $k > 3$ , distribuição leptocúrtica
- $k < 3$ , distribuição platicúrtica



De uma forma global, os coeficientes de assimetria e curtose, são utilizados para tirar conclusões sobre a normalidade da variável em estudo. No entanto, em forma de complemento, é também usual recorrer-se ao teste Jarque-Bera (JB).

Esse teste é um dos testes mais comuns para testar a normalidade de uma variável (Brooks, 2008), através da média, variância, simetria e curtose da mesma. Ou seja, assume-se que uma variável tem distribuição normal se apresentar uma distribuição simétrica e mesocúrtica (isto é, com assimetria nula e coeficiente de curtose igual a 3).

A estatística de teste do JB tem uma distribuição  $\chi^2$  com 2 graus de liberdade sobre a hipótese nula da normalidade,  $H_0$ . Assim, se a série tem uma distribuição normal, a estatística do teste de Jarque-Bera não será significativa, ou seja, o  $p$ -value associado ao teste será superior ao nível de significância 5%, levando à não rejeição da hipótese nula da normalidade.

## 2.2. Processos Estocásticos

Um processo estocástico pode ser definido como qualquer família de variáveis aleatórias  $\{Y(t), t \in I\}$ , em que  $I$  corresponde a um intervalo de tempo predefinido. De uma forma geral, o propósito da análise de séries temporais recai sobre registrar ocorrências dum processo estocástico desconhecido, logo as características de uma série temporal são as mesmas que as do processo estocástico associado.

Seja  $\{Y(t), t \in I\}$  um processo estocástico com variância finita, definem-se como suas estatísticas:

- Valor médio:  $\mu(t) = E[Y(t)]$
- Variância:  $\sigma^2(t) = Var(Y(t)) = E[(Y(t) - \mu(t))^2]$
- Covariância:
 
$$\begin{aligned} \gamma(t_1, t_2) &= cov[Y(t_1), Y(t_2)] = E\{[Y(t_1) - \mu(t_1)][Y(t_2) - \mu(t_2)]\} \\ &= E[Y(t_1)Y(t_2)] - \mu(t_1)\mu(t_2) \end{aligned}$$
- Coeficiente de correlação de Pearson:

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sigma(t_1)\sigma(t_2)} = \frac{\text{cov}[Y(t_1), Y(t_2)]}{\sqrt{\text{Var}(Y(t_1))\text{Var}(Y(t_2))}}$$

Uma das propriedades essenciais das séries temporais é a sua estacionariedade, uma vez que a maioria dos modelos econométricos só é válido para séries estacionárias, o que indica que a sua distribuição de probabilidade é estável ao longo do tempo.

No entanto, na parte econométrica, a utilização dada a este conceito é quase nula, dado que é muito difícil obter séries reais com este tipo de características (Ferreira, 2016). Por assim ser, assumir uma forma de estacionariedade mais fraca é por vezes suficiente.

Assim sendo, um processo estocástico  $\{Y(t), t \in I\}$  diz-se fracamente estacionário (estacionariedade fraca ou em covariância), se  $\forall t \in I$ , se tem:

- Média constante ao longo do tempo:  $E[Y(t)] = \mu$
- Variância constante e finita ao longo do tempo:  $\text{Var}(Y(t)) = \sigma^2 < \infty$
- Covariância dependente do desfasamento  $p$  e constante em relação ao tempo  $t$

$$\text{cov}[Y(t), Y(t-h)] = \gamma(t, t-h) = \gamma_h$$

Em séries temporais existem dois tipos de tendências: tendência determinística, em que os desvios em relação à linha de tendência são puramente aleatórios e não têm impacto sobre o comportamento ao longo do prazo, e tendência estocástica, em que a componente aleatória  $\varepsilon_t$  afeta a trajetória a longo prazo da série.

Esta distinção é de grande relevância uma vez que os dois tipos de tendência requerem diferentes tratamentos para induzir a estacionariedade na série. No primeiro caso, a estacionariedade é obtida através da remoção da tendência segundo a qual a série se desenvolve. Enquanto que, no segundo caso, a estacionariedade é induzida quando se efetua, geralmente, a primeira diferença da série.

Estas tendências podem ser representadas, respetivamente:

$$Y(t) = a + bt + \varepsilon_t \qquad Y(t) = \delta + Y_{t-1} + \varepsilon_t, \quad \text{com drift } \delta$$

$$Y(t) = Y_{t-1} + \varepsilon_t, \quad \text{sem drift}$$

Com base nas tendências, foram definidos os dois modelos mais populares para séries temporais não-estacionárias: o processo estacionário em tendência puro (TSP – *Trend Stationary Process*) e o processo estacionário em diferenças puro (DSP – *Difference Stationary Process*).

Em suma, para o seguinte modelo de regressão simples,

$$Y_t = \rho Y_{t-1} + \beta_0 + \beta_1 t + \varepsilon_t$$

Consideremos os seguintes cenários:

- Se  $\rho = 0$ , então existe uma tendência determinística e temos um processo estacionário em tendência puro, em que um processo que é estacionário em torno de uma tendência linear – série não estacionária;
- Se  $\rho = 1$  e  $\beta_1 = 0$ , então existe uma tendência estocástica nos dados, em que temos um processo estacionário em diferenças puro, designada por passeio aleatório com deriva (*drift*) – série não estacionária;
- Se  $\rho = 1$  e  $\beta_0, \beta_1 \neq 0$ , estamos perante um passeio aleatório com *drift* e tendência determinística – série não estacionária;
- Se  $0 < \rho < 1$  e  $\beta_1 = 0$ , existe correlação serial em  $Y$ , mas não existe tendência estocástica – série estacionária.

### 2.3. Processos para obter a estacionariedade

Como anteriormente dito, quando verificada a não-estacionariedade é necessário proceder a um dos seguintes processos para estabilizar as séries:

- Diferenças sucessivas:  $\Delta Y_t = Y_t - Y_{t-1}$   
 $\Delta^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$
- Diferenças dos logaritmos (retornos ou rendibilidades):

$$\Delta \log(Y_t) = \log(Y_t) - \log(Y_{t-1})$$

Para além disso, se uma série não-estacionária necessitar de ser diferenciada  $d$  vezes, para ficar estacionária, então a série diz-se integrada de ordem  $d$ ,  $Y_t \sim I(d)$ . Por sua vez, se uma série é estacionária sem necessitar de ser diferenciada, diz-se que é integrada de ordem zero,  $I(0)$ .

Uma sequência  $\{\varepsilon_t\}$  define o ruído branco, representada por  $\varepsilon_t \sim WN(0, \sigma^2)$  ou  $Y(t) \sim WN(0, \sigma^2)$ , se cada valor da série tiver média zero, variância constante e não apresentar correlação serial, isto é:

- $E(\varepsilon_t) = 0$ ;
- $Var(\varepsilon_t) = \sigma^2 < \infty$ ;
- $E(\varepsilon_t, \varepsilon_{t-1}) = cov(\varepsilon_t, \varepsilon_{t-s}) = 0$ , para  $s \neq 0$

É necessário ter em atenção que os erros (resíduos) dos modelos direcionados para previsão devem ser ruído branco, caso contrário, significa que os mesmos são correlacionados e, portanto, a previsão pode não ser fidedigna.

### Testes de Raiz Unitária

Se as técnicas de regressão standard forem aplicadas às séries não estacionárias, pode acontecer que o resultado de ser uma regressão de qualidade no que diz respeito às suas medidas standard (coeficientes dos estimadores significativos e um elevado  $R^2$ ),

mas, no entanto, tem muito pouco valor. Tal modelo tem a denominação de Regressão Espúria.

Com objetivo de evitar situações indesejáveis como apresentado anteriormente, foram criados os testes de raiz unitária, cujo propósito consiste em avaliar se a série é estacionária em termos estocásticos. Para isso, de uma forma geral é testada a hipótese nula de que a série é não estacionária ( $H_0: \rho = 1$ , possui uma raiz unitária) versus a hipótese de que a série é estacionária ( $H_1: \rho \neq 1$ , não possui raiz unitária).

Atendendo aos vários testes propostos destacam-se os testes desenvolvidos por Dickey e Fuller (1979), o teste KPSS desenvolvido por Kwiatkowski *et al.* (1992) e o teste PP de Phillips e Perron (1988).

No teste de Dickey – Fuller assume-se que os erros são independentes e identicamente distribuídos (i.i.d), ou seja, não apresenta qualquer autocorrelação.

Considerando o modelo de regressão inicial  $Y_t = \rho Y_{t-1} + \beta_0 + \beta_1 t + \varepsilon_t$ , este altera-se (ligeiramente) ao subtrair  $Y_{t-1}$  em ambos os lados da equação, obtendo-se:

$$Y_t - Y_{t-1} = (\rho - 1)Y_{t-1} + \beta_0 + \beta_1 t + \varepsilon_t$$

Assim, para este caso, as hipóteses a serem testadas são as que se seguem:

$$H_0: \alpha = \rho - 1 = 0 \quad \text{versus} \quad H_1: \alpha \neq 0$$

O teste em foco é utilizado quando os erros  $\varepsilon_t$  são ruído branco, ou seja não tem correlação serial. Porém, tal condição nem sempre é verificada, por isso deverá ser executado o teste com correção paramétrica, conhecido por teste de Dickey – Fuller Aumentado (ADF – *Augmented Dickey – Fuller*).

Este teste sugere o aumento da regressão, através da adição de um número de variáveis desfasadas da primeira diferença de  $Y_t$ , até que os erros  $\varepsilon_t$  obtidos não sejam autocorrelacionados, isto é:

$$\Delta Y_t = (\rho - 1)Y_{t-1} + \sum_{i=1}^p \beta_i \Delta Y_{t-i} + \varepsilon_t$$

Os testes Phillips - Perron (PP) são muito similares aos testes ADF, porém incorporam uma correção automática ao teste DF que levam a que sejam permitidos erros autocorrelacionados. Além disso, este teste em foco geralmente conduz às mesmas conclusões e possui as limitações importantes do teste ADF.

Por último, o teste Kwiatkowski – Phillips – Schmidt - Shin (KPSS) foi criado com o objetivo de conceber um contraponto aos tradicionais testes de raiz unitárias, com as seguintes hipóteses a serem consideradas:

$$H_0: \rho < 1 \quad \text{versus} \quad H_1: \rho = 1$$

Além disso, a estatística de teste é dada por:

$$LM = \sum_{l=1}^t \frac{S_t}{n^2 \hat{\theta}}$$

Onde  $s_t = \sum_{i=1}^t \varepsilon_i$  e  $\hat{\theta}$  é um estimador para a variância dos erros.

As hipóteses são não-rejeitadas ou rejeitadas de acordo com os valores críticos associados aos níveis de significância usuais, 1%, 5% e 10%. Sendo que se rejeita a hipótese nula  $H_0$  se:

$$LM_{KPSS} > \text{valores críticos}$$

## 2.4. Modelos de Séries Temporais Univariadas

Os modelos de séries temporais univariadas recorrem apenas a informações contidas nos valores passados registados pelas variáveis e, possivelmente, aos valores atuais e passados de um termo de erro, com o intuito de prever valores futuros.

Uma classe importante desses tipos de modelos são a família dos modelos ARMA. Estes modelos foram criados para séries estacionárias, uma vez que assumem que o processo está em equilíbrio. No entanto, caso a série não apresente estacionariedade, aplica-se o modelo ARIMA, sendo I correspondente ao processo de integração da série não-estacionária.

Estes tipos de modelos, ARMA e ARIMA, estão associados à metodologia de Box e Jenkins (1976), que visa captar o comportamento da correlação serial, ou autocorrelação, entre os valores da série, de forma a realizar futuras previsões.

### a. Modelos Autorregressivos – $AR(p)$

Os modelos autorregressivos são modelos onde o valor atual da variável em estudo,  $Y$ , depende apenas dos valores tomados por esta nos períodos passados e de um termo de erro. Por conseguinte, um modelo autorregressivo de ordem  $p$ , denotado por  $AR(p)$  é expresso da seguinte forma:

$$Y_t = a_0 + \sum_{i=1}^p a_i Y_{t-i} + \varepsilon_t \quad \text{com } \varepsilon_t \sim WN(0, \sigma^2)$$

Onde  $p$  designa-se por ordem de desfasamento (*lag order*) e  $\varepsilon_t$  é o processo de ruído branco que assinala a nova informação recebida no instante  $t$ .

Tomando a notação de  $L$  para o operador desfasamento (*operador lag*), representado por  $L^i(Y_t) = Y_{t-i}$ , a equação anterior pode ser escrita de forma equivalente:

$$\Phi(L)Y_t = a_0 + \varepsilon_t$$

Onde

$$\Phi(L) = 1 - \sum_{i=1}^p a_i L^i$$

Um processo  $AR(p)$  é estacionário, se e só se todas as raízes  $z = \frac{1}{\alpha_i}$  do seguinte polinómio são situadas no exterior do círculo unitário (ou seja,  $|\alpha_i| < 1$ ):

$$\Phi(z) = 1 - \sum_{i=1}^p a_i z^i = \prod_{i=1}^p (1 - \alpha_i z)$$

- Função Autocorrelação (ACF): Com média  $E(Y_t) = 0$  e covariância  $\gamma_h = E(Y_t, Y_{t+h})$ , define-se por:

$$\rho_k = \sum_{i=1}^p a_i \rho_{k-i}$$

Onde  $\rho = \frac{\gamma_h}{\gamma_0}$  e  $\gamma = Var(Y_t)$

- Variância:  $Var(Y_t) = \frac{\sigma^2}{\sum_{i=1}^p a_i \rho_i}$

- Função Autocorrelação Parcial (PACF):

$$R(p) = \begin{cases} a_h \neq 0, & \text{para } h < p \text{ (existe)} \\ \text{PACF } 0, & \text{para } h > p \text{ (não existe)} \\ \text{PACF} & \end{cases}$$

Esta função denota a convergência brusca para zero, ao fim de  $p$  passos, que por sua vez, apresenta uma convergência rápida para zero em comparação à função ACF. Além disso, é de notar que PACF diferencia um processo AR de um processo ARMA.

## b. Modelos Média Móvel - $MA(q)$

Sendo  $\varepsilon_t$  um processo de ruído branco, com  $E(Y_t) = 0$  e  $Var(\varepsilon_t) = \sigma^2$ , o modelo  $MA(q)$  representa-se por:

$$Y_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} = \varepsilon_t + \sum_{i=1}^q \theta_i L^i \varepsilon_t$$

Ou por,

$$Y_t = \theta_q(L) \varepsilon_t$$

Onde  $\theta_q(L) = 1 + \sum_{i=1}^q \theta_i L^i$  e  $\theta_1, \theta_2, \dots, \theta_q$  são constantes reais.

Ao considerar as equações anteriores, conclui-se que o processo  $MA(q)$  define-se como a média ponderada das últimas  $q + 1$  observações de um processo de ruído branco, em cada instante  $t$ . Além disso, independentemente dos valores das constantes reais  $\theta_1, \theta_2, \dots, \theta_q$ ,  $MA(q)$  é sempre um processo fracamente estacionário.

- Valor médio: Desde que  $E(\varepsilon_t) = 0$  também se constata que  $E(Y_t) = 0$ .

- Variância: Sendo  $E(Y_t) = 0$  e os erros  $\varepsilon_t$  não correlacionados, temos:

$$Var(Y_t) = (1 + \sum_{i=1}^q \theta_i^2) \sigma^2$$

- Função Autocovariância:

$$\gamma_s = \begin{cases} \sigma^2 (\theta_h + \sum_{i=s+1}^q \theta_i \theta_{i-s}), & \text{para } 0 < s \leq q \\ 0, & \text{para } s > q \end{cases}$$

- Função Autocorrelação:

$$\rho_k = \begin{cases} \frac{\theta_k + \sum_{i=k+1}^q \theta_i \theta_{i-k}}{1 + \sum_{i=1}^q \theta_i^2}, & \text{para } 0 < k \leq q \\ 0, & \text{para } k > q \end{cases}$$

Segundo a fórmula anterior, pode-se dizer que a função ACF de um processo  $MA(q)$  se anula no lag  $(q + 1)$ , o que por sua vez, indica que a função PACF converge suavemente para 0.

### c. Modelos Autorregressivos de Médias Móveis – $ARMA(p, q)$

Ao combinar os modelos  $AR(p)$  e  $MA(q)$ , surge o modelo  $ARMA(p, q)$ , que estabelece que o valor de uma série  $Y$  depende linearmente dos valores registados em momentos anteriores e de uma combinação de valores atuais e passados de um processo de ruído branco.

Por conseguinte, um modelo  $ARMA(p, q)$  é dado por:

$$Y_t = a_0 + \sum_{i=1}^p a_i Y_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

Com  $E(\varepsilon_t) = 0$ ;  $E(\varepsilon_t^2) = \sigma^2$  e  $E(\varepsilon_t, \varepsilon_s) = 0$  para  $t \neq s$ .

- Valor médio:  $E(Y_t) = \frac{a_0}{1 - \sum_{i=1}^p a_i}$

- Função de Autocorrelação (ACF): Esta função apresenta combinações de comportamentos derivados dos modelos AR e MA. Porém, para  $p < q$ , a função ACF vai ser idêntica à de um modelo  $AR(p)$ .

Além disso, o modelo  $ARMA(p, q)$  permite que sejam realizadas previsões dos dados já observados, funcionando muito bem como uma primeira aproximação. No entanto, salientam-se algumas limitações quando as séries temporais apresentam comportamentos divergentes, intervalos de tempo irregulares, altos e baixos na volatilidade e existência de quebras de estrutura.

### **d. Modelos Autorregressivos Integrados de Médias Móveis – $ARIMA(p, d, q)$**

Como foi dito anteriormente, este modelo é utilizado quando as séries não apresentam estacionariedade, o que se verifica quando não se observa

$$|1 - \sum_{i=1}^p a_i| < 1.$$

Neste tipo de modelos  $p$  retrata a ordem da componente AR,  $d$  apresenta o número de vezes que a série teve que ser diferenciada para se tornar estacionária e  $q$  representa a ordem da componente MA.

Matematicamente, sobre uma série  $Y_t$  não-estacionária devemos aplicar um modelo  $ARIMA(p, d, q)$ . No entanto, sobre a série  $Z_t = \Delta^d Y_t = (1 - L)^d Y_t$ , estacionária e resultante da diferenciação da série  $Y_t$ ,  $d$  vezes, terá de ser aplicado o modelo  $ARMA(p, q)$ .

Além disso, é de notar que todos os modelos anteriormente mencionados podem ser definidos em modelos  $ARIMA(p, d, q)$ , da seguinte forma:

- $ARMA(p, q) = ARIMA(p, 0, q)$
- $AR(p) = ARIMA(p, 0, 0)$
- $MA(q) = ARIMA(0, 0, q)$

## **2.5. Critérios de Comparação de Modelos**

Em relação à escolha de um modelo que represente a série temporal, pode-se prestar atenção aos valores das funções ACF e PACF, através da análise gráfica, para identificar quais as ordens a selecionar. Porém, por vezes pode não ser possível escolher as ordens, porque pode exibir padrões não muito esclarecedores.

De forma a evitar o acontecimento anteriormente mencionado, pode-se recorrer à apreciação dos critérios de informação. Estes critérios incorporam dois fatores: um termo que é uma função de soma dos resíduos ao quadrado e uma penalidade pela perda de graus de liberdade ao adicionarmos parâmetros extra. Por conseguinte, adicionar uma nova

variável ou um *lag* adicional ao modelo leva a que a soma dos resíduos quadrados diminua, enquanto que o valor do termo de penalidade aumenta.

Além disso, o intuito é escolher o número de parâmetros que diminui o valor do critério de informação, portanto adicionar um termo extra só irá reduzir esse valor, se a soma dos resíduos ao quadrado diminuir o suficiente, de modo a compensar o aumento no valor do termo da penalidade.

De acordo com Brooks (2008), os 3 critérios de informação mais importantes são:

- Critério de Akaike:  $AIC = \ln(\hat{\theta}) + \frac{2k}{T}$
- Critério de Schwarz Bayesiana:  $SC = \ln(\hat{\theta}) + \frac{k}{T} \ln(T)$
- Critério de Hannan – Quinn:  $HQC = \ln(\hat{\theta}) + \frac{2k}{T} \ln(\ln(T))$

Onde  $\hat{\theta}$  representa a variância residual, que se obtêm através do quociente entre a soma dos resíduos ao quadrado e o número de observações,  $k = p + q + 1$ , que diz respeito ao número total de parâmetros estimados e  $T$  que retrata o tamanho da amostra.

Contudo, é de notar que o SC integra um termo de penalidade mais rígido do que o AIC, enquanto que HQC se encontra entre ambos.

## 2.6. Validação dos Modelos

Para que o modelo escolhido seja validado, os seus resíduos devem verificar os pressupostos seguintes:

1. Os erros têm média nula ( $E(\varepsilon_t) = 0$ ).
2. A variância dos erros é constante e finita ( $\text{Var}(\varepsilon_t) = \sigma^2 < \infty$ ). Quando a variância dos erros é constante, está-se perante um caso de homocedasticidade. Caso contrário, denomina-se por heterocedasticidade.
3. Os erros são linearmente independentes ( $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ ),  $i \neq j$ .
4. Os erros são normalmente distribuídos ( $\varepsilon_t \sim N(0, \sigma^2)$ ).

Relativamente ao terceiro pressuposto mencionado anteriormente, um modelo adequa-se corretamente aos dados, quando os resíduos são linearmente independentes, para isso os resíduos não deverão rejeitar a hipótese nula: “Não existe correlação serial entre os resíduos”, do teste Breusch – Godfrey *Serial Correlation*.

Do mesmo modo, emprega-se o teste ARCH associado à hipótese nula: “Não existem efeitos ARCH nos resíduos (ou seja, a variância é constante)”, para verificar a existência de homocedasticidade nos resíduos.

## 2.7. Previsão

O processo de previsão realiza-se por duas partes. Primeiramente faz-se uma previsão *in-sample*, isto é, uma previsão dentro da janela temporal correspondente a cada amostra. Isto tem como intuito a comparação e seleção dos modelos não rejeitados e identificados, em que para isso analisa-se o erro absoluto médio (MAE) e a raiz do erro quadrático médio (RMSE). Será escolhido o modelo com menor valor de ambos.

Posteriormente, prossegue-se para uma previsão *out-of-sample* que consiste na previsão de valores fora da janela temporal de cada amostra. Neste caso serão previstos os valores fora do intervalo de tempo observado e a qualidade da previsão será avaliada usando as mesmas métricas.

No entanto, determinar a precisão de uma previsão feita por determinado modelo é uma maneira importante de testar a adequabilidade do mesmo aos dados em estudo. Contudo, de acordo com alguns economistas a adequabilidade estatística de um modelo, quer em termos dos cinco pressupostos, quer em termos da significância dos seus parâmetros, é bastante irrelevante desde que esse modelo produza previsões precisas (Brooks, 2008).

### Erro de Previsão

O erro de previsão  $e_i$  entre uma observação, ( $Y_i$ ), e o valor previsto associado ( $\hat{Y}$ ), é representado na seguinte forma:

$$e_i = Y_i - \hat{Y}$$

No intuito de avaliar o desempenho de determinado modelo, é importante utilizar um indicador que tenha a capacidade de classificar numericamente a previsão efetuada pelo modelo. Temos os seguintes indicadores:

- Erro Médio:  $ME = \frac{\sum_{i=1}^n e_i}{n}$
- Erro Absoluto Médio:  $MAE = \frac{\sum_{i=1}^n |e_i|}{n}$
- Erro Quadrático Médio:  $MSE = \frac{\sum_{i=1}^n e_i^2}{n}$
- Raiz do Erro Quadrático Médio:  $RMSE = \sqrt{MSE}$
- Erro Percentual Absoluto Médio:  $MAPE = \frac{100\%}{n} \times \sum_{i=1}^n \frac{|e_i|}{Y_i}$

### Previsão com Modelos ARIMA

Uma vez escolhido e estimado um modelo ARIMA, pode-se prever os valores futuros da série, tendo como base o modelo determinado. Contudo, é importante ter em atenção que quanto menor for o erro e a janela de previsão, melhores serão as previsões efetuadas pelo modelo.

Em termos de notação, a equação de previsão é dada por:

$$\hat{Y}(s) = \sum_{i=1}^{p+q} a_i \hat{Y}(s-1), \quad s > q$$

Onde  $\hat{Y}(s)$  é a previsão efetuada da série  $Y$ , no instante  $t$ , para  $s$  passos após esse mesmo instante. Ou seja, o intuito da equação anterior é prever um valor  $Y_{t+s}$ , com  $s \geq 1$ , considerando que todas as observações passadas são conhecidas.

## Metodologia de Box-Jenkins

Fase 1 – Identificação:

- Preparação dos dados e diferenciação dos dados para obter uma série estacionária
- Seleção do modelo: Examinar os dados, ACF e PACF para identificar os modelos potenciais

Fase 2 – Estimação e teste:

- Estimativas: Estimar os parâmetros dos modelos potenciais e selecionar o melhor modelo usando critérios de informação
- Diagnósticos: Analisar a ACF e PACF dos resíduos e verificar se os resíduos são ruído branco

Fase 3 – Aplicação:

- Previsão: Utilizar o modelo para a previsão de dados futuros.

A aplicação do método B-J depende de a série temporal em estudo ser estacionária, o que se verifica quando não se observa qualquer tendência e sazonalidade nos respectivos dados, ou seja, os dados apresentam uma média independente do tempo e a variância permanece constante. Contudo, para determinar se a série temporal é estacionária, analisa-se analiticamente e graficamente os coeficientes de autocorrelação (ACF) e de autocorrelação parcial (PACF) (Box, Jenkins e Reinsel (1994)).

No entanto, ao verificar a não-estacionariedade da série, retira-se a mesma através de transformações matemáticas simples: como, por exemplo, diferenciação e as diferenças de logaritmos dos elementos da série temporal (Makridakis, Wheelwright e Hyndman (1998)).

Uma vez verificada a estacionariedade da série, identificam-se as ordens apropriadas dos parâmetros de MA – coeficientes de autocorrelação (ACF) – e as ordens adequadas para os parâmetros de AR – coeficientes de autocorrelação parcial da série (PACF) (Box, Jenkins e Reinsel (1994)).

Em suma, sendo as séries não são estacionárias aplicam-se o modelo ARIMA e sendo estacionárias aplicam-se os modelos ARMA, que consiste nos modelos ARIMA, mas sem o fator de integração.

Por último, verificam-se se os resíduos (diferença entre os valores atuais dos valores estimados pelo modelo selecionado) são ruído branco, o que indica se o modelo é válido. Ao analisar a ACF e a PACF dos resíduos espera-se, no caso de erros aleatórios, que nenhum coeficiente de autocorrelação e autocorrelação parcial seja significativo (Makridakis e Hibon (1997); Makridakis, Wheelwright e Hyndman (1998)).

Definido o modelo, o método irá capturar as correlações históricas entre os dados e obter as previsões desejadas (Makridakis e Hibon (1997); Makridakis, Wheelwright e Hyndman (1998)).

No entanto, em relação à previsão de séries financeiras, esta pode ser feita *in-sample* (com objetivo de testar a precisão da previsão), em que obtêm valores para a janela temporal da amostra da série em estudo, ou *out-of-sample* (valores futuros), que é empregue para a obtenção de valores futuros fora do intervalo de tempo da amostra.

Contudo, o intuito por detrás da previsão de uma série temporal financeira é escolher entre os modelos o que melhor se adequa a prever a variável, por isso a seleção dos modelos deverá ser feita também com auxílio aos erros de previsão.

De entre as muitas medidas de adequabilidade existentes as mais utilizadas nos processos de previsão são: MAE (Erro absoluto médio); MPE (Erro percentual médio) e MAPE (Erros absolutos percentuais médios); MSE (Erro quadrático médio); RMSE (Raiz do erro quadrático médio).

### 3. *Machine Learning*

*Machine Learning* ou Aprendizagem Máquina é uma área vasta de Inteligência Artificial que vem a ser desenvolvida em muitas vertentes das áreas de saber. Os seus algoritmos tratam imagens, textos e dados. No entanto, nesta tese, vai-se abordar as suas aplicações (métodos, modelos, algoritmos) na área da previsão financeira, devido ao seu interesse teórico-prático e também à sua complexidade.

Neste capítulo está presente um pequeno aglomerado de matéria referente a este tema de ciência de dados, com o intuito de facilitar a compreensão do seu funcionamento. Os tópicos apresentados estão baseados em leitura de vários artigos, mas principalmente segue o livro de Ketkar, (2017).

#### 3.1. **Redes Neurais Artificiais (RNAs)**

Desde os últimos anos da década de 80, que as redes neurais artificiais (RNA) têm sido um tema popular na análise de dados, mas só no início da década de 90, é que começaram a ser usadas em Finanças.

De acordo com Arevalo (2016), a criação das RNAs é inspirada no cérebro humano e, por essa razão, as redes são compostas por muitos neurónios conectados entre si. Cada neurónio é uma unidade processadora que realiza uma agregação ponderada de múltiplos sinais de entrada. E, dependendo das suas entradas, é ativado e propaga um novo sinal de saída, por meio da aplicação de uma função não-linear.

A primeira RNA criada foi *Feed-forward Neural Network* (FNN), que contém vários neurónios conectados entre si, mas não existem ciclos nem *loops* ao longo da rede, e por esse motivo, as informações sempre avançam das entradas para os nós de saída.

Por essa razão, foi criado um dos modelos mais conhecidos nesta área, o modelo *Multilayer Perceptron* (MLP), que é um subtipo FNN. Este é composto por: uma camada de entrada, que aceita o vetor variável de entrada, várias camadas ocultas, que se encontram entre a camada de entrada e a de saída, e que têm como função aceitar vetor de saída por camada de entrada ou última camada oculta e uma camada de saída, que produz o resultado do processamento do vetor de entrada.

Cada camada tem um número finito de neurónios, no entanto podem conter quantidades distintas. Contudo, exceto nos neurónios na camada de entrada, todos os outros têm uma função de ativação não-linear, que normalmente assume a forma sigmoide, isto é

$$y(v_i) = \tanh(v_i) \quad \text{ou} \quad y(v_i) = \frac{1}{1 + e^{-v_i}}$$

onde  $v_i$  são as entradas lineares ponderadas para cada neurónio

Através da função de ativação, a rede MLP encontra relações não-lineares entre a entrada e a saída, o que o torna uma ferramenta bastante útil em muitos trabalhos, como classificação e regressão.

As RNAs, tradicionalmente, são treinadas com o algoritmo *Backpropagation*, que consiste em inicializar as matrizes de pesos do modelo com valores aleatórios, e de seguida, o erro entre a saída da rede e a saída desejada é avaliado.

No entanto, um problema sério desse algoritmo é que o erro é diluído exponencialmente à medida que passa pelas camadas ocultas, no caminho para o início da rede. Por esse motivo, apareceu o modelo *deep* MLP (um MLP com muitas camadas ocultas), em que apenas as últimas camadas são treinadas, enquanto as primeiras são pouco modificadas. Este modelo integra nos modelos de *Deep Learning*, que serão abordados posteriormente.

Resumidamente, em termos práticos, uma Rede Neural pode ser pensada como uma função  $f_{\theta}: x \rightarrow y$ , que recebe uma entrada  $x \in \mathbb{R}^n$  e produz uma saída  $y \in \mathbb{R}^m$ , e cujo comportamento é parametrizada por  $\theta \in \mathbb{R}^w$ , que é simplesmente uma coleção de todos os pesos  $w$  para todas as unidades na rede.

Isto é, numa forma geral, as redes neurais são construídas usando a unidade como base. Estas podem conter várias camadas (profundidade da rede), que englobam uma ou mais unidades, sendo isto referido como a largura da camada, que pode diferenciar de camada para camada.

Em relação à previsão gerada pela camada de saída, esta pode ser denotada por  $\hat{y}$  e avaliada através da função de perda, que mede o erro entre  $\hat{y}$  e  $y$  que denotamos por  $l$ .

Além disso, relativamente às funções de perda, pode-se referir como as mais utilizadas em *Machine Learning*, a entropia cruzada binária e a entropia cruzada, em que geralmente são derivadas, usando Máxima Verossimilhança. No entanto, enquanto que a entropia cruzada binária é utilizada no contexto de classificação binária, a entropia cruzada é usada no contexto de multiclassificação.

Ao supor que  $f(x, \theta)$  é o modelo que prevê a probabilidade de  $y$  dado  $x$ , onde  $\theta$  representa os parâmetros do modelo e que  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  são independentes, tendo uma distribuição de Bernoulli, pode-se representa a função entropia cruzada binária como:

$$P(D|\theta) = \prod_{i=1}^n f(x_i, \theta)^{y_i} \cdot (1 - f(x_i, \theta))^{(1-y_i)}$$

Com base na ideia de entropia cruzada binária, pode-se representar a entropia cruzada, por:

$$P(D|\theta) = \frac{n!}{n_1! \cdot n_2! \dots n_k!} \prod_{i=1}^n f(x_i, \theta)^{y_i}$$

onde  $y \in \{0, 1, \dots, k\}$  são classes,  $n_1, n_2, \dots, n_k$  são as contagens observadas de cada uma das  $k$  classes e  $\sum_{i=1}^k n_i = n$ . Além disso, de novo, assume-se que  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  são independentes, no entanto, neste caso seguem-se uma distribuição Multinomial.

Quando os algoritmos de *Machine Learning* são usados para regressão/previsão, a função de perda é quadrática e a estimação faz-se pelo método dos mínimos quadrados com gradiente decrescente estocástico.

### 3.2. Support Vector Machines (SVMs)

No campo das Finanças, os *Support Vector Machines* (SVMs) foram criados em 1963 por Vladimir N. Vapnik e Alexey Ya Chervonenkis, para classificarem problemas lineares numa forma mais eficiente que as redes neurais artificiais (RNAs).

No entanto, em 1992, Vapnik usou o método *kernel trick*, para que os SVMs fossem capazes de classificar modelos não-lineares. Este método tenta encontrar a melhor linha reta que pode classificar corretamente a maioria dos pontos de dados, que pertencem apenas a duas classes denotadas por vetores de recursos no espaço  $R^2$  (Qian e Gao, 2017).

Em 2003, Kim conseguiu melhores resultados usando o método *kernel Gaussian Radial Basis Function* como um classificador não-linear, no modelo SVM. Tendo sido comparado com as Redes Neurais usando o método *Backpropagation*, ao qual se chegou à conclusão de que SVM é melhor.

Pai e Lin, (2005) combinaram um SVM com média móvel integrada autorregressiva (ARIMA) num sistema híbrido capaz de erros menores do que aqueles obtidos usando os modelos separadamente.

H. Y. Yu et al., (2009) propuseram uma variação do SVM conhecido como *Least Squares* (LSSVM), que possui um custo computacional menor que o SVM original e boa capacidade de generalização, com o uso de Algoritmo Genético (GA) – um processo de seleção de valores otimizados para cada geração, que apresenta resultados superiores aos modelos SVM convencionais, ARIMA e redes neurais do tipo rede neural *Backpropagation* (BPNN).

Contudo, diferente de muitos modelos tradicionais de ML que tentam minimizar o risco empírico, ou seja, o erro de treino no conjunto de treino, SVM tentou minimizar o risco estrutural, uma combinação de risco empírico e complexidade do modelo, o que efetivamente evita o problema de *overfitting*, que significa que o modelo ajustado pode ter um bom desempenho no conjunto de treino, mas ruim no conjunto de teste. Para mais detalhes sobre as redes SVM vide Ketkar, (2017).

### 3.3. Deep Learning (DL)

*Deep learning* é uma forma de *Machine Learning* e como tal, também treina um modelo em dados para fazer previsões, mas diferencia-se pelas diferentes camadas de abstração pelos quais os dados passam, tendo uma aprendizagem mais profunda.

DL é hierárquico, no sentido de que em cada camada, o algoritmo extrai recursos em fatores, e os fatores de um nível mais profundo tornam-se os recursos do próximo nível. Ou seja, a própria palavra *deep* em *deep learning* indica a sua profundidade, em processar diretamente a partir de dados brutos.

Além disso, abrange vários derivados das RNAs, que incluem modelos como Rede Neural de Convolução (RNC), Rede Neural Recorrente (RNR), *Deep Belief Network* – DBF, entre outros.

De entre as várias possibilidades, os algoritmos mais procurados e com aplicações bem-sucedidas são os baseados nos modelos de redes recorrentes, em particular o LSTM e o GRU (*Gated Recurrent Unit*).

Como no modelo MLP, os neurónios nos algoritmos DL contêm funções de ativação sigmoidal ou de ativação tangente hiperbólica, isto é, respetivamente

$$\frac{1}{1 + e^{-x}} \quad \text{ou} \quad \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

de tipo *heaviside gate functions* (ex:  $\mathbb{I}(x > 0)$ ) ou unidades lineares retificadas (ReLU). A função ReLU (*Rectified Linear Unit*) retorna zero se recebe um input negativo e retorna o valor  $x$ , se  $x$  é positivo. As últimas funções de ativação mencionadas são das mais usadas recentemente, em particular nas unidades das camadas ocultas. Além disso, os resultados mostram que os neurónios com função de activação ReLU levam a gradientes grandes e consistentes, com o objetivo de uma rápida redução de dimensão dos dados.

Relativamente ao processamento de dados, é comum normalizar ou estandardizar os dados e depois dividir o conjunto de dados em três subconjuntos: o conjunto de treino que é usado para ajustar os pesos da rede, o conjunto de validação que é usado para minimizar a sobreposição e relaciona-se ao design da arquitetura (também chamado de seleção de modelo) e o conjunto de teste que é usado para confirmar o poder preditivo do modelo.

De acordo com Heaton *et al.*, (2016), em relação ao primeiro conjunto acima referido, existem 2 problemas-chave que podem ser abordados usando o desempenho preditivo de uma arquitetura:

- Um dos problemas é ter conhecimento de quanta regularização se deve adicionar à função de perda. Uma das soluções possíveis é usar validação cruzada, em que um conjunto independente de dados de *hold-out* (o conjunto de dados é dividido em conjunto de treino e conjunto de teste) é mantido em separado, para realizar uma medição fora da amostra do sucesso de treino numa segunda etapa, e ensinar o algoritmo a se calibrar para

um dado de treino. Isto leva a que à medida que variamos a quantidade de regularização, obtemos um caminho de regularização e escolhemos o nível de regularização para otimizar a perda preditiva fora da amostra. Além disso, pode-se também usar o estimador de risco não-enviesado de Stein (SURE), que estima de forma não-enviesada o erro de um estimador fixo qualquer. Assim, para qualquer limiar, o erro quadrático médio pode ser estimado para que o limiar com o menor erro seja selecionado. Para mais detalhes sobre SURE vide Ketkar, (2017).

- Um outro problema mais complicado é treinar o tamanho e a profundidade de cada camada da arquitetura, que é também conhecido como o problema de seleção de modelos. Para resolução desse problema, existe a técnica *dropout*, que é posteriormente mencionada.

Como referida anteriormente, a validação cruzada é uma técnica pelo qual dividimos nossos dados de treino num conjunto complementar para depois conduzir análises e validações em diferentes conjuntos, com o objetivo de reduzir o *overfitting* e aumentar o desempenho fora da amostra.

Em particular, quando treinamos séries temporais, podemos dividir os dados de treino em períodos de tempo separados, de duração idêntica, o que é particularmente desejável em aplicações financeiras em que é difícil obter indicadores confiáveis de tempo consistente e ter que ser treinados e testados extensivamente.

Além disso, a validação cruzada também fornece uma ferramenta para decidir quais níveis de regularização levam a uma boa generalização (ou seja, previsão), que é o *variance-bias trade-off* clássico. Isto é, se o modelo é demasiado simples e tem poucos parâmetros, então vai ter *bias* alto e variância baixa. Por outro lado, se o modelo tem demasiados parâmetros, então vai ter *bias* e variância alta. Portanto, terá de se encontrar o bom equilíbrio, de forma que a evitar tanto o *overfitting* como o *underfitting*. Uma das principais vantagens da mesma, em relação às métricas estatísticas tradicionais, é que permite avaliar o tamanho e a profundidade das camadas ocultas.

Relativamente à solução do segundo problema referido, *Dropout* (abandono) é projetado para evitar o *overfitting* excessivo no processo de treino e faz isso removendo as dimensões de entrada em  $X$  aleatoriamente com uma dada probabilidade  $p$ . E ainda, reduz a probabilidade de dependência excessiva de pequenos conjuntos de dados de entrada no treino.

Outra aplicação de regularização de *dropout* é a escolha do número de unidades ocultas numa camada, o que pode ser alcançado se eliminarmos unidades de camada oculta invés da camada de entrada e, em seguida, estabelecermos qual a probabilidade  $p$  dá os melhores resultados. Além disso, umas das propriedades de redução de dimensão de uma estrutura de rede é que, uma vez que uma variável de uma camada é descartada, todos os termos acima dela na rede também desaparecem.

De acordo com Heaton *et al.*, (2016) existem 2 projetos de arquitetura amplamente usados que se tornaram comuns em aplicações de *Machine Learning*: Redes Neurais Recorrentes (RNRs) e modelos LSTM.

Em relação às RNRs, apesar de estes terem-se mostrado bem-sucedidos em tarefas como reconhecimento de fala e geração de texto, têm dificuldade em aprender dinâmicas de longo prazo. Essa complicação advém em parte dos gradientes que desaparecem e explodem, que resultam da propagação dos gradientes, através das muitas camadas (correspondentes ao tempo) da rede neural recorrente. Contudo, estas redes são abordadas mais aprofundadamente, posteriormente.

Logo, apareceu os LSTMs, que têm a mesma estrutura de cadeia que RNR e aprendizagem profunda (DL), mas diferentes estruturas de módulos repetidos. Ao invés de uma camada de rede neural única da RNR, existem várias camadas interagindo numa forma especial. Além disso, são uma forma particular de rede recorrente que fornece uma solução, incorporando unidades de memória. Isso permite que a rede aprenda quando esquecer os estados ocultos anteriores e quando atualizar estados ocultos com novas informações. Todavia, mais à frente existe mais informações sobre estas redes.

Por fim, relativamente à previsão em termos gerais em DL tem várias vantagens comparando às previsões tradicionais: os dados de entrada podem ser expandidos para incluir todos os itens de possível relevância para o problema de previsão; as não-linearidades e interações complexas entre os dados de entrada são contabilizados, o que pode ajudar a aumentar o tamanho da amostra em relação aos modelos tradicionais; o *overfitting* é mais facilmente evitado.

## a. Redes Neurais Recorrentes

Vamos supor ao longo desta secção que a entrada consiste numa sequência de entidades  $x^{(1)}, x^{(2)}, \dots, x^{(\tau)}$  e que precisamos produzir uma sequência  $y^{(1)}, y^{(2)}, \dots, y^{(\tau)}$  ou apenas uma saída para a sequência de entrada inteira  $y$ . No entanto, para distinguir entre o que a RNR produz e o que é idealmente esperado que seja produzido, denotaremos por  $\hat{y}^{(1)}, \hat{y}^{(2)}, \dots, \hat{y}^{(\tau)}$  ou  $\hat{y}$  a saída que a RNR produz.

As RNRs produzem uma saída para cada entidade na sequência de entrada, logo pode ser descrito usando as seguintes equações:

$$h^{(t)} = \tanh(Ux^{(t)} + Wh^{(t-1)} + b)$$

$$\hat{y}^{(t)} = \text{softmax}(Vh^{(t)} + c)$$

onde  $h^{(t)}$  denota o cálculo do estado oculto de uma entidade na sequência,  $x^{(t)}$  a entrada na entidade,  $h^{(t-1)}$  o estado oculto anterior,  $\hat{y}^{(t)}$  a saída calculada usando o estado oculto  $h^{(t)}$ ,  $U$  e  $W$  os pesos associados à entrada e ao estado oculto anterior durante o cálculo do estado oculto atual, respetivamente,  $b$  o termo de enviesamento,  $V$  os pesos associados ao estado oculto durante o cálculo da saída e o seu termo de enviesamento  $c$ .

A função de ativação do *tanh* é usada no cálculo do estado oculto e a função de ativação *softmax* é usada no cálculo da saída.

Em relação à função *softmax*, esta é um tipo de função sigmoide, que é útil quando se tenta lidar com problemas de classificação. Além disso, transforma as saídas para cada classe em valores entre 0 e 1 e também divide pela soma das saídas, o que dá a probabilidade de a entrada estar numa determinada classe.

Contudo, conforme descrito pelas equações, a RNR pode processar uma sequência de entrada arbitrariamente grande. Além disso, os parâmetros da RNR nomeadamente,  $U$ ,  $W$ ,  $V$ ,  $b$ ,  $c$ , *etc* são compartilhadas através do cálculo da camada oculta e do valor de saída (para cada uma das entidades na sequência).

Atualmente, é essencial perceber como uma RNR é diferente de todas as redes neurais *feedforward* (incluindo redes de convolução). A principal diferença é o estado oculto, que representa um resumo das entidades vistas no passado (para a mesma sequência).

Considera-se agora uma variação na RNR em que, invés da recorrência usando o estado oculto, tem-se recorrência usando a saída produzida no estado anterior:

$$h^{(t)} = \tanh(Ux^{(t)} + W\hat{y}^{(t-1)} + b)$$

$$\hat{y}^{(t)} = \text{softmax}(Vh^{(t)} + c)$$

Comparativamente às equações descritas anteriormente, a única diferença está na equação do cálculo do estado oculto de uma entidade na sequência,  $h^{(t)}$ , pois usa a saída anterior  $\hat{y}^{(t-1)}$ . Ao invés do estado oculto anterior,  $h^{(t-1)}$ .

Também se pode considerar uma outra variação na RNR, onde apenas uma única saída é produzida para toda a sequência.

$$h^{(t)} = \tanh(Ux^{(t)} + W\hat{y}^{(t-1)} + b)$$

$$\hat{y} = \text{softmax}(Vh^{(\tau)} + c)$$

Em comparação com a variante da RNR anterior, haverá só uma única saída,  $\hat{y}$  e esta apenas usa o último estado oculto  $h^{(\tau)}$ .

Relativamente ao treino de uma RNR, este baseia-se simplesmente em desenrolar a RNR para um dado tamanho de entrada e treinar a RNR desenrolada via computação dos gradientes e usando o método de otimização do gradiente descendente estocástico (SGD).

Contudo, a ideia central do SGD é obter uma aproximação do gradiente ao analisar apenas alguns exemplos dos dados.

Em relação ao desenrolamento da relação de recorrência correspondente à RNR, este assenta na escrita das equações substituindo recursivamente o valor no qual a relação

de recorrência é definida, ou seja, o valor de  $h^{(t)}$  é definido em termos de  $h^{(t-1)}$ , que por sua vez é definido em termos de  $h^{(t-2)}$  e assim por diante até  $h^{(0)}$ .

Além disso, o processo de desenrolamento opera na suposição de que o comprimento da sequência da entrada,  $\tau$ , é conhecido de antemão e com base na recorrência é desenrolada. Contudo, uma vez desenrolados, temos essencialmente uma rede neural não recorrente, logo o processo de treino da RNR é idêntico ao treino de uma rede neural regular.

Dado que o conjunto de dados a ser treinado consiste em sequências de tamanhos variados, as sequências de entrada são agrupadas de modo que as sequências do mesmo tamanho caiam num grupo. Assim, é possível treinar a RNR com entradas de tamanhos variados.

Contudo, o treino das RNRs tem algumas desvantagens, pois sofre com os desafios de desaparecimento de gradientes, ou seja, ao calcular os gradientes nas RNRs desenrolados, o valor dos gradientes pode cair para um valor muito pequeno (próximo de zero), e com os desafios de explosão de gradientes, isto é, os gradientes podem aumentar para um valor muito alto.

De forma a lidar com gradientes explosivos pode-se redimensionar a norma do gradiente sempre que ultrapassar um limite definido pelo usuário. Especificamente, se o gradiente denotado por

$$\hat{g} = \frac{\partial L}{\partial W} \quad e \quad se \quad |g| > c$$

então, definimos

$$\hat{g} = \frac{c}{|g|} g$$

Computacionalmente eficiente, mas introduz um hiperparâmetro extra.

$$\frac{\partial L}{\partial W} = \sum_{1 \leq t \leq \tau} \frac{\partial L^{(t)}}{\partial h^{(t)}} \left[ \sum_{1 \leq k \leq t} \left[ \prod_{k \leq j \leq t-1} \frac{\partial h^{(j+1)}}{\partial h^{(j)}} \right] \frac{\partial h^{(k)}}{\partial W} \right]$$

onde  $\prod_{k \leq j \leq t-1} \frac{\partial h^{(j+1)}}{\partial h^{(j)}}$  envolve uma multiplicação de matriz repetida de  $w$ , que contribui para ambos os problemas de gradiente. Intuitivamente, isso é semelhante a multiplicar um número real valorizado repetidas vezes, o que pode levar o produto a encolher até zero ou explodir até o infinito.

### **Long Short Term Memory (LSTM)**

Os modelos LSTMs incluem uma porta de entrada, uma porta *Forget*, uma porta de modulação de entrada e uma célula de memória. A unidade de célula de memória combina a unidade de célula de memória anterior que é modulada pela porta de

modulação de entrada e de *Forget* juntamente com o estado oculto anterior, modulado pela porta de entrada. Essas células adicionais permitem que uma arquitetura LSTM aprenda dinâmicas temporais de longo prazo extremamente complexas.

Além disso, um modelo LSTM pode potencialmente melhorar os preditores utilizando dados do passado, memorizando padrões de volatilidade de períodos anteriores, o que permite automatizar a identificação das relações temporais nos dados, ao custo de conjuntos maiores de parâmetros a serem treinados.

A fim de compreender melhor o modelo LSTM, pode-se representá-lo da seguinte forma:

$$\begin{aligned}
 z^{(t)} &= g(W_z x^{(t)} + R_z \hat{y}^{(t-1)} + b_z) \\
 i^{(t)} &= \sigma(W_i x^{(t)} + R_i \hat{y}^{(t-1)} + p_i \odot c^{(t-1)} + b_i) \\
 f^{(t)} &= \sigma(W_f x^{(t)} + R_f \hat{y}^{(t-1)} + p_f \odot c^{(t-1)} + b_f) \\
 c^{(t)} &= i^{(t)} \odot z^{(t)} + f^{(t)} \odot c^{(t-1)} \\
 o^{(t)} &= \sigma(W_o x^{(t)} + R_o \hat{y}^{(t-1)} + p_o \odot c^{(t)} + b_o) \\
 \hat{y}^{(t)} &= o^{(t)} \odot h(c^{(t)})
 \end{aligned}$$

O elemento mais importante do LSTM é o estado da célula denotado por  $c^{(t)}$ . Este é atualizado com base na entrada do bloco  $z^{(t)}$ , que produz um valor baseado na entrada atual e na saída anterior, e no estado da célula anterior  $c^{(t-1)}$ . A porta de entrada  $i^{(t)}$  determina qual fração da entrada do bloco faz com que ela entre no estado da célula. A porta *Forget*  $f^{(t)}$  determina quanto do estado da célula anterior deve ser retido.

Além disso, a saída  $\hat{y}^{(t)}$  é determinada com base no estado da célula  $c^{(t)}$  e na porta de saída  $o^{(t)}$ , que determina o quanto o estado da célula afeta a saída. Todos os termos  $p$  são conexões *peephole*, que permitem que uma fração do estado da célula seja incluída no cálculo do termo em questão.

O cálculo do estado da célula  $c^{(t)}$  não encontra a questão do gradiente de desaparecimento. No entanto, os LSTMs são afetados por gradientes de explosão.

Por fim, note que o símbolo  $\odot$  observa multiplicação pontual de dois vetores (se  $a = [1,1,2]$  e  $b = [0.5,0.5,0.5]$  então  $a \odot b = [0.5,0.5,1]$ , as funções  $\sigma$ ,  $g$  e  $h$  não são funções de ativação linear, todas  $W$  e  $R$  são matrizes de peso e todos os termos  $b$  são termos de enviesamento).

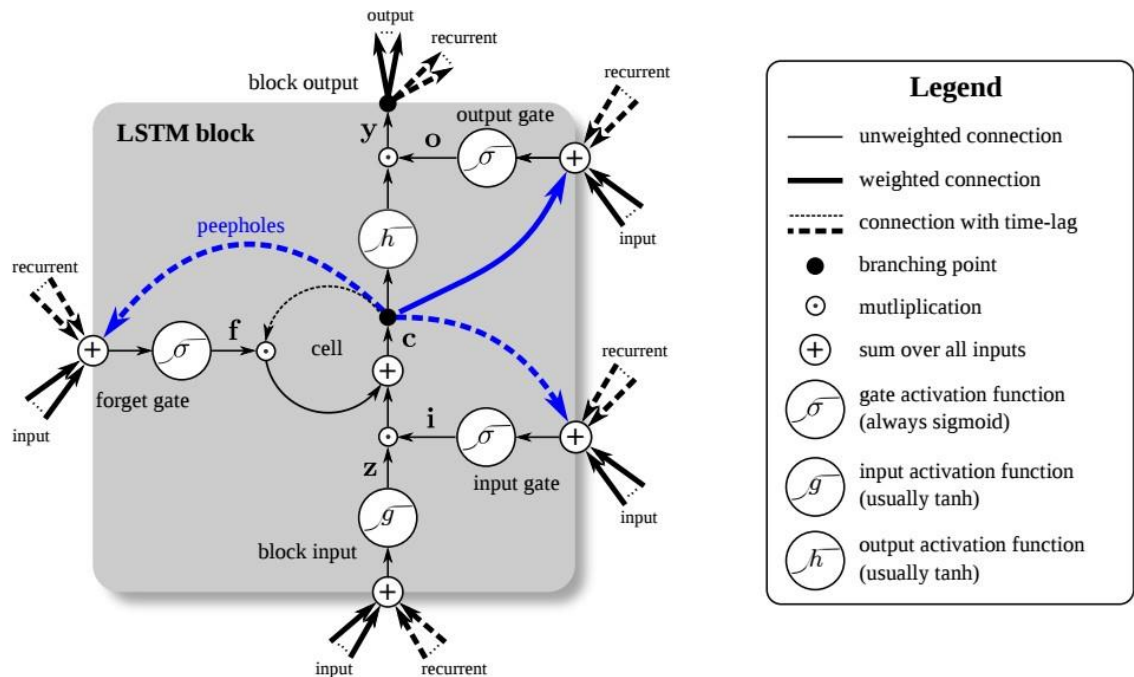


Figura 2 - Exemplo de rede LSTM

Resumindo, tendo em conta a Figura 2 os algoritmos LSTM tem controle sobre: a decisão de quando o input pode entrar no neurónio, de quando lembrar o que foi calculado no passo anterior e poder decidir também quando deixar o output passar para o próximo nível de tempo.

A complexidade e performance do LSTM traz um conjunto elevado de hiperparâmetros, cuja calibração é um processo bastante demorado e complicado.

## 4. Estudo Empírico

A fim de realizar o objetivo desta dissertação, vai se proceder à previsão da série Standard & Poor's 500, através de dois métodos distintos, com o intuito de posteriormente ser feita uma comparação entre os resultados obtidos.

Este capítulo é dividido por três seções: em que na primeira é feita uma breve análise da série, através de uma representação gráfica e dos seus valores estatísticos descritivos, na segunda é feita a previsão da série pelos métodos econométricos e por fim a previsão com a *deep neural network (LSTM)*.

Contudo, é importante salientar que a segunda seção deste capítulo é realizada no programa *Eviews*, tais como todos os gráficos presentes nesta dissertação são obtidos do mesmo *software*, no entanto a última seção, relativamente aos modelos de redes LSTM é implementada na linguagem de programação *Python*, usando o ambiente *Spyder*.

Os dados em análise encontram-se no intervalo temporal de 3 de janeiro de 2000 a 3 de abril de 2019, sendo estes diários, no entanto não englobam feriados e fins de semana. Os dados foram retirados de *Yahoo Finance*.

### 4.1. Análise da Série Standard & Poor's 500

A série em estudo, como já referido anteriormente, é constituída pelos valores da bolsa *Standard & Poor's 500 (SP500)*, devido à sua relevância relativamente ao estudo comportamental da economia americana.

Com intuito a um melhor entendimento do comportamento deste índice e a fim de retirar algumas conclusões, é apresentado o gráfico representativo da evolução da bolsa.

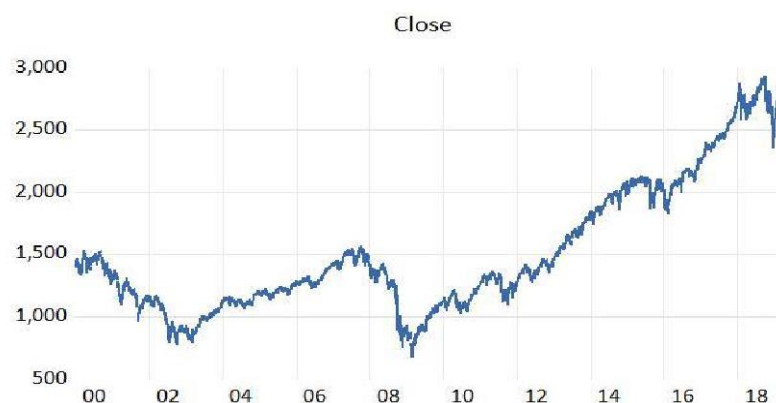


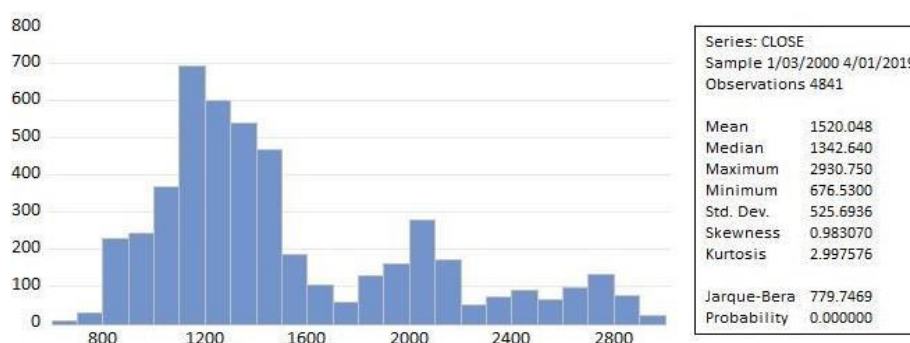
Figura 3 - Série global do índice *Standard & Poor's 500*

Ao observar a figura anterior (Figura 3), verifica-se a não existência de uma tendência geral definida, pois apresenta movimentos irregulares e sem padrões sazonais,

ou seja, uma oscilação constante. Estes aspetos dão indicação de uma variância e média não constantes, o que leva a concluir que a série não apresenta estacionariedade.

Além disso, é também evidente uma queda abrupta em 2007 a 2008, que poderá estar relacionada com a crise económica nesse período. Por essa razão, posteriormente será reduzida a janela temporal da série, iniciando só em 2 de março de 2009, com o objetivo de obter melhores resultados na previsão de valores futuros, em particular para os modelos lineares.

A fim de proceder a uma análise mais aprofundada, é agora apresentada a tabela com as estatísticas descritivas do índice SP500.



**Figura 4** - Gráfico de barras e estatísticas descritivas da série global do índice *Standard & Poor's 500*

De acordo com a Figura 4, a série apresenta um coeficiente de curtose ligeiramente inferior a 3, que independentemente de ser bastante próximo, esta apresenta uma distribuição platocúrtica. No entanto, ao observar o coeficiente de assimetria, este indica que a série revela uma distribuição assimétrica positiva, pois é superior a 0 e a média é superior à mediana. O índice SP500 atinge o seu valor máximo, 2930.750, nos meados do ano 2018 e o seu valor mínimo, 676.530, no início de 2009.

Relativamente ao teste de normalidade de Jarque-Bera, este apresenta um *p-value* inferior aos 3 níveis de significância, 1%, 5% e 10%, o que leva à rejeição da hipótese nula que corresponde à normalidade. Logo, a série não segue uma distribuição normal. Pela observação do gráfico apresentado na figura anterior, a distribuição parece ser multimodal.

## 4.2. Previsão com Modelos Econométricos – Série Global

Tendo em conta o que foi referido no capítulo 2, em que a maioria dos modelos econométricos só são válidos para séries estacionárias, nesta secção é analisada a sua estacionariedade e, posteriormente, são feitas as respetivas transformações, caso, inicialmente, seja comprovada a não-estacionariedade da série em níveis.

Por fim, realiza-se a previsão dos valores futuros do índice *Standard & Poor's 500*, usando a metodologia de Box-Jenkins aplicada aos modelos ARIMA.

## a. Estacionariedade

A análise da estacionariedade da série é feita através de três testes: teste Dickey – Fuller Aumentado (ADF), teste Phillips – Perron (PP) (testes de raiz unitária) e o teste Kwiatkowschi – Phillips – Schmidt – Shin (KPSS) (teste de estacionariedade).

Ao realizar os testes obteve-se os seguintes resultados:

Testes de estacionariedade		
ADF	PP	KPSS
0,9928 *	0,9959 *	6,116871 **

\*p-value  
\*\* estatística de teste

**Tabela 1** - Resultados dos testes de estacionariedade da série global do índice *Standard & Poor's 500*

Conforme apresentado na Tabela 1, pode-se concluir que a série não é estacionária, pois nos dois primeiros testes não se rejeita a hipótese nula, que representa a existência de uma raiz unitária, enquanto que no teste KPSS rejeita-se a estacionariedade da série<sup>1</sup>.

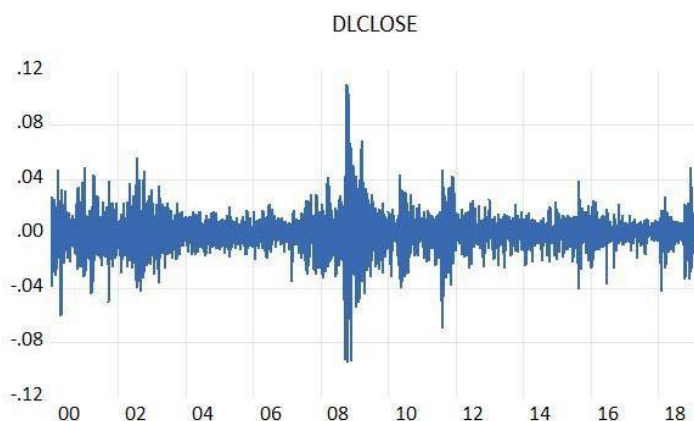
Estes resultados comprovam o que foi referido na análise da série, porém é importante saber que tipo de não-estacionariedade se encontra associada à *Standard & Poor's 500*, pois cada tipo requer tratamentos diferentes.

Existem dois tipos de não estacionariedade, a não estacionariedade em tendência determinística e a não estacionariedade em tendência estocástica. Neste caso, está-se perante a segunda tipologia de não-estacionariedade, em que só é necessário recorrer às diferenças dos logaritmos (retornos) para estabilizar a série.

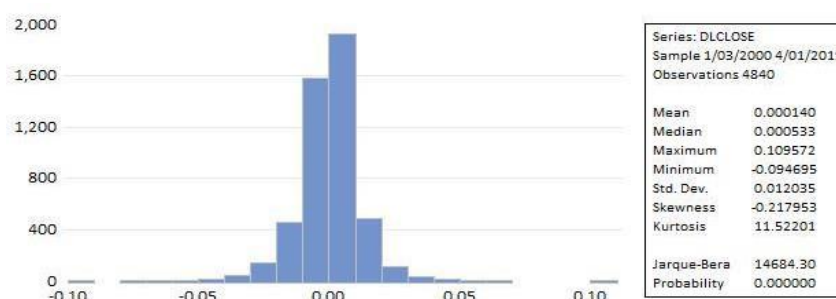
## b. Análise Gráfica e Estatística dos retornos

Após efetuada a diferença dos logaritmos sobre os dados do índice *Standard & Poor's 500*, é necessário analisar a série, novamente.

<sup>1</sup> Os outputs dos testes de estacionariedade, ADF, PP e KPSS da série global SP500 encontram-se no Anexo A – Testes de Estacionariedade da Série Global



**Figura 5** - Série global dos retornos do índice *Standard & Poor's 500*



**Figura 6** - Gráfico de barras e estatísticas descritivas da série global dos retornos do índice *Standard & Poor's 500*

Ao observar o gráfico anterior (Figura 5), pode-se constatar que a série dos retornos passou a apresentar um valor médio constante, contudo ainda demonstra clusters de volatilidade, isto é, sinais que a variância não é constante.

Ao analisar as estatísticas descritivas presentes na figura anterior (Figura 6), conclui-se que a série dos retornos de SP500 segue, agora, uma distribuição assimétrica negativa e leptocúrtica, pois o valor do coeficiente de *skewness* apresenta um valor negativo e a curtose, um valor muito superior a três. Observam-se também alguns valores muito extremos (*outliers*) fora das caudas da distribuição.

Em consequência, rejeita-se a normalidade dos retornos de *Standard & Poor's 500*, pois o *p-value* é inferior a qualquer nível de significância, o que leva, uma vez mais, à rejeição da hipótese nula.

### c. Estacionariedade da Série dos retornos

De novo, com intuito de estudar a estacionariedade do índice modificado, volta-se a testar nos três diferentes tipos de teste, Dickey – Fuller Aumentado (ADF), Phillips

– Perron (PP) e o Kwiatkowski – Phillips – Schmidt – Shin, obtendo assim os seguintes resultados:

Testes de estacionariedade		
ADF	PP	KPSS
0,0001 *	0,0001 *	0,334069 **

\*p-value  
\*\*estatística de teste

**Tabela 2** - Resultados dos testes de estacionariedade da série global dos retornos do índice *Standard & Poor's 500*

De acordo com a Tabela 2, pode-se afirmar que a série é estacionária, pois tanto no teste ADF como no teste Phillips-Perron, rejeita-se a hipótese nula, que representa a existência de uma raiz unitária.

Além disso, no teste KPSS não se rejeita a estacionariedade da série, pois o valor da estatística de teste é inferior a todos os outros valores críticos. Podendo assim, afirmar que a série dos retornos é estacionária<sup>2</sup>. Resumindo, a série SP500 é integrada de ordem 1, I(1), apresenta volatilidade elevada e valores extremos.

#### **d. Modelização da Série dos retornos**

Após obtido a estacionariedade da série, procede-se à escolha do modelo que melhor se adequa aos retornos do índice *Standard & Poor's 500*.

É importante saber que, por vezes, os modelos que melhor se ajustam aos dados da série nem sempre são esses que proporcionam melhores previsões. Por essa razão, nesta secção encontra-se o modelo que melhor se ajusta aos dados, no entanto, na previsão compara-se os resultados dos vários modelos possíveis.

Para alcançar o objetivo desta secção, recorre-se ao correlograma da série dos retornos para determinar quais os possíveis modelos representativos do índice, através das funções ACF e PACF.

---

<sup>2</sup>Os outputs dos testes de estacionariedade, ADF, PP e KPSS da série dos retornos de SP500 encontram-se no Anexo B – Testes de Estacionariedade da Série Global

Date: 09/26/19 Time: 09:56  
 Sample: 1/03/2000 4/01/2019  
 Included observations: 4840

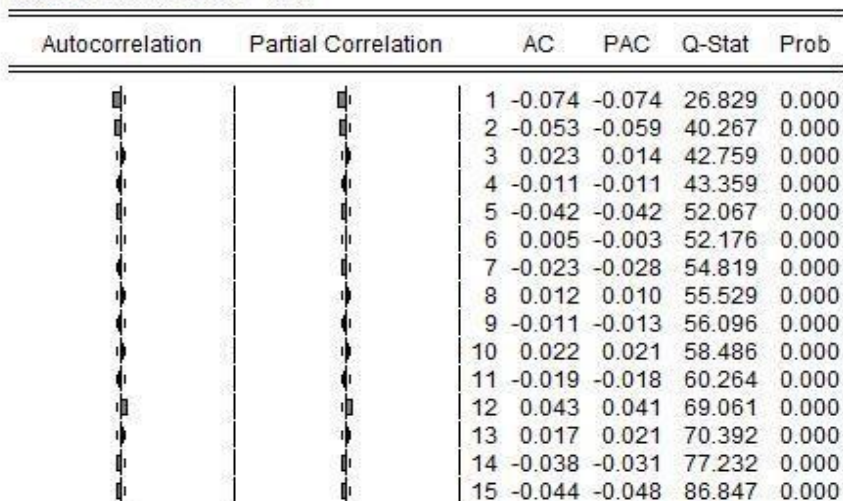


Figura 7 - Correlograma da série global dos retornos do índice *Standard & Poor's 500*

Ao analisar o correlograma<sup>3</sup> da Figura 7, é possível verificar que os dois primeiros lags demonstram ser significativos, tanto na função ACF como na PACF, o que leva a concluir que existe a possibilidade de estar perante um dos seguintes modelos AR(1), AR(2), MA(1), MA(2), ARMA(1,1), ARMA(1,2), ARMA(2,1) ou ARMA(2,2).

Tendo em consideração os modelos anteriormente referidos, segue-se a seleção do modelo que melhor se ajusta aos dados, através da comparação dos critérios de informação, Akaike, Schwarz Bayesiana e Hannan – Quinn, de todos os possíveis modelos.

Modelos	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)	ARMA(1,2)	ARMA(2,1)	ARMA(2,2)
AIC	-6,008911	-6,011735	-6,007611	-6,009791	-6,010609	-6,011593	-6,011528	-6,011131
SC	-6,006231	-6,007714	-6,004932	-6,005772	-6,006588	-6,006233	-6,006166	-6,004429
HQC	-6,007970	-6,010323	-6,006671	-6,008380	-6,009197	-6,009712	-6,009645	-6,008778

Tabela 3 - Valores dos critérios de informação de todos os modelos potenciais da série global dos retornos do índice *Standard & Poor's 500*

Ao comparar os critérios de informação<sup>4</sup> de cada modelo (Tabela 3), conclui-se que o modelo AR(2) é o modelo com AIC, SC e HQC mais baixos, logo será este, possivelmente, a ser utilizado na previsão dos valores futuros do índice.

Após esta conclusão, é agora necessário analisar os resíduos do modelo selecionado. Para isso recorre-se aos testes de correlação (Breusch-Pagan), em que um modelo que se adequa corretamente aos dados, têm de apresentar resíduos linearmente independentes, e ao teste de heterocedasticidade (ARCH), onde os resíduos possuem uma variância estável ao longo da janela temporal, se forem homocedásticos.

<sup>3</sup>A figura completa do correlograma da série dos retornos de SP500 está no Anexo C – Modelização da Série Global

<sup>4</sup>Os outputs de todos os modelos possíveis da série global dos retornos da SP500 estão no Anexo C – Modelização da Série Global .

Modelo	AR(2)
ARCH	0,0000
Breusch-Godfrey Serial Correlation	0,4906

**Tabela 4** - Resultados dos testes de correlação e de heterocedasticidade da série global dos retornos do índice *Standard & Poor's 500*

Tendo em consideração a tabela anterior (Tabela 4), pela comparação dos *p-value's* com os níveis de significância, pode-se afirmar que os resíduos da série os retornos de SP500 é linearmente independente, no entanto apresenta heterocedasticidade, ou seja, os resíduos têm uma variância instável<sup>5</sup>.

Por conseguinte, conclui-se que os resíduos não validam todos os pressupostos do ruído branco, o que era previsível, uma vez que o coeficiente de curtose é bastante elevado, existem valores extremos e possíveis quebras de estrutura. O facto de a série conter muitos dados e uma dinâmica complexa, dificulta a escolha de um modelo linear que se ajuste corretamente aos mesmos. Por essa razão, será realizada uma quebra na janela temporal, posteriormente, analisando a série num intervalo temporal com uma dinâmica mais suave.

### e. Previsão da série

A previsão é feita por todos os modelos possíveis encontrados durante o processo de seleção do modelo, comparando os erros de previsão, sendo o melhor, o que tiver menor valor.

Inicialmente, efetua-se uma previsão *in-sample*, em que representa uma previsão dentro do intervalo de tempo, e após a escolha do modelo com menor erro de previsão, faz-se uma previsão *out-of-sample* com esse mesmo modelo.

#### **Previsão *in-sample***

Nesta previsão prevê-se 4841 observações contidas no intervalo de 3 de janeiro de 2000 a 1 de abril de 2019 para todos os modelos possíveis.

Posteriormente, é feita uma comparação entre os valores do erro absoluto médio (MAE), do erro absoluto percentual médio (MAPE) e da raiz do erro quadrático médio (RMSE) de todos modelos, para que se escolha o melhor modelo para previsão *out-of-sample*.

<sup>5</sup>Os outputs dos testes de ARCH e de Breusch – Godfrey Serial Correlation encontram-se no Anexo D – Testes de Correlação e Heterocedasticidade da Série Global .

Modelos	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1,1)	ARMA(1,2)	ARMA(2,1)	ARMA(2,2)
RMSE	16,03021	16,01659	16,04750	16,03334	16,02252	16,01366	16,01509	16,01545
MAE	11,06845	11,08385	11,08190	11,09248	11,08568	11,08339	11,08226	11,08196
MAPE	0,798278	0,799274	0,799184	0,799956	0,799565	0,799255	0,799117	0,799067

**Tabela 5** - Valores RMSE, MAE e MAPE de todos os modelos potenciais da série global dos retornos do índice *Standard & Poor's 500*

De acordo com a tabela anterior<sup>6</sup>, constata-se que o modelo com menor raiz do erro quadrático médio (RMSE) é o ARMA(1,2), no entanto o modelo AR(1) é o que apresenta menor MAE, logo de forma a decidir qual destes deve ser o modelo escolhido, tem de se comparar os valores *Bias Proportion* e *Covariance Proportion*.

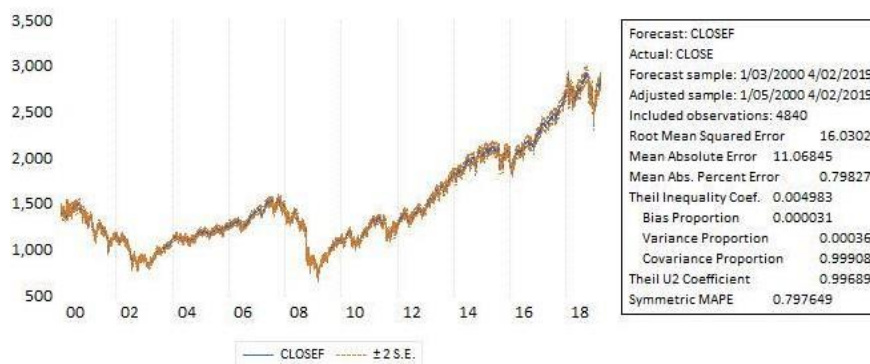
Modelos	AR(1)	ARMA(1,2)
<b>Bias Proportion</b>	0,000031	0,000039
<b>Covariance Proportion</b>	0,999083	0,999914

**Tabela 6** - Valores *Bias Proportion* e *Covariance Proportion* dos modelos AR(1) e ARMA(1,2) da série global dos retornos do índice *Standard & Poor's 500*

Ao analisar os dois indicadores presentes na Tabela 6, conclui-se que o melhor modelo é AR(1), pois este apresenta um *Bias Proportion* mais próximo de 0 e um *Covariance Proportion* aproximadamente igual a 1.

### Previsão *out-of-sample*

Uma vez escolhido o modelo, prevê-se o valor do dia seguinte após o intervalo de janela temporal. Uma vez que esse período termina a 1 de abril de 2019, a previsão terá sido efetuada para o dia 2 de abril de 2019.



**Figura 8** - Previsão *out-of-sample* do dia 2/4/2019 da série global dos retornos da SP500, com o modelo AR(1)

<sup>6</sup>Os outputs de todos os modelos possíveis encontrados encontram-se no Anexo E – Previsão da Série Global.

Através da Figura 8, observa-se que tanto o erro absoluto médio (MAE) como a raiz do erro quadrático médio (RMSE) apresentam valores elevados o que não sugere uma boa previsão. Porém, não indicam necessariamente que seja uma má previsão, uma vez que ambos os critérios dependem da escala utilizada.

Por isso, é preciso ter em atenção o coeficiente de Theil, a *Bias Proportion*, a *Variance Proportion* e a *Covariance Proportion*. Isto é, se o coeficiente de Theil, que varia entre 0 e 1, tem um valor próximo de 0, então os valores previstos aproximam-se bastante dos valores reais. O mesmo acontece, se a *Bias Proportion* e a *Variance Proportion* se aproximem relativamente de 0, e ainda se a *Covariance Proportion* registar um valor próximo de 1.

Tendo em consideração as afirmações feitas anteriormente, pode-se concluir que se está perante uma boa previsão, independentemente do MAE e RMSE serem elevados, pois o coeficiente de Theil, a *Bias Proportion* e a *Variance Proportion* são relativamente próximos de 0, e a *Covariance Proportion* próxima de 1.

Contudo, para obtenção de uma melhor comparação entre modelos, calcula-se agora o erro de previsão só para o dia 2 de abril de 2019, em que se obteve o valor previsto de 2865.190, sendo o valor verdadeiro 2867.23999:

$$\frac{|2865.190 - 2867.23999|}{2867.23999} = 0,0007149$$

Analisando o erro obtido, obtém-se que, a média percentual absoluta do erro (MAPE) é aproximadamente 0.07%, portanto, pode-se constatar que a previsão é excelente.

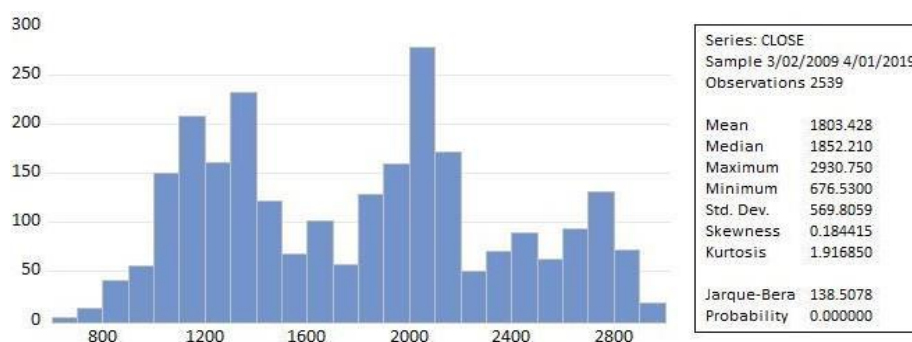
### 4.3. Previsão com Modelos Econométricos – Série Reduzida

Uma vez reduzida a janela temporal do índice *Standard & Poor's 500*, com início a 2 de março de 2009, volta-se a analisar a série.



Figura 9 - Série reduzida do índice *Standard & Poor's 500*

Comparativamente à série no seu global, é novamente visível que não existe uma tendência linear global, pois apresenta alguns padrões não-lineares. Isto indica que a variância e a média são instáveis, o que leva a concluir que a série não é estacionariedade.



**Figura 10** - Gráfico de barras e estatísticas descritivas da série reduzida do índice *Standard & Poor's 500*

Ao observar a Figura 10, verifica-se que a série, uma vez mais, segue uma distribuição platicúrtica e de assimetria positiva, pois apresenta um coeficiente de curtose inferior a 3 e um coeficiente de assimetria superior a 0, respetivamente.

Em relação ao teste de normalidade, este volta a apresentar um *p-value* inferior aos 3 níveis de significância, 1%, 5% e 10%, o que leva à rejeição da normalidade.

### a. Estacionariedade

Da mesma forma como na série global, o estudo da estacionariedade da série é feito através dos três típicos testes, em que se obteve os seguintes resultados:

Testes de estacionariedade		
ADF	PP	KPSS
0,8117 *	0,8386 *	6,125075 **

\*p-value  
\*\* estatística teste

**Tabela 7** - Resultados dos testes de estacionariedade da série reduzida do índice *Standard & Poor's 500*

De acordo com a Tabela 7, constata-se que a série não é estacionária, pois os dois primeiros testes não rejeitam a existência de uma raiz unitária, enquanto que o teste KPSS rejeita a estacionariedade da série<sup>7</sup>.

<sup>7</sup>Os outputs dos testes de estacionariedade, ADF, PP e KPSS da série reduzida da bolsa SP500 encontram-se no Anexo F – Testes de Estacionariedade da Série Reduzida.

Novamente, os testes corroboram as conclusões retiradas da análise da série e a série também apresenta uma não estacionariedade em tendência estocástica.

## b. Análise Gráfica e Estatística dos Retornos

Devido à tipologia de não-estacionariedade presente nesta série, efetua-se uma vez mais a diferença dos logaritmos sobre os dados.

Obtida a nova série, é agora necessário analisá-la, a fim de perceber se a série é estacionária ou não.

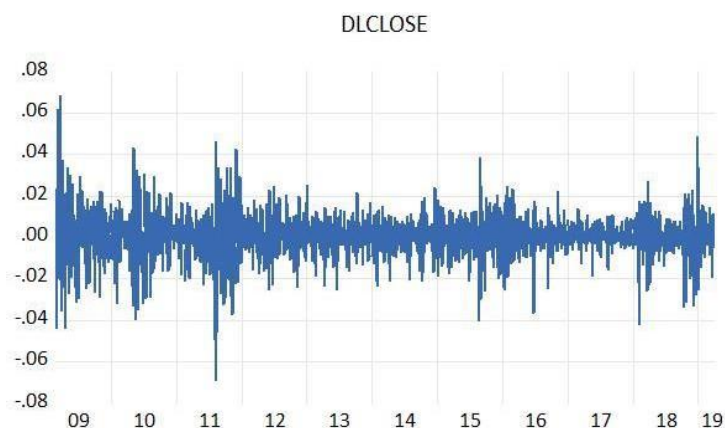


Figura 11 - Série reduzida dos retornos do índice *Standard & Poor's*

Tendo em conta o gráfico presente na Figura 11, verifica-se que a série dos retornos desenvolve-se ao longo de uma média constante nula e apresenta alguns clusters de volatilidade.

De forma a obter uma análise mais profunda, estuda-se as estatísticas descritivas da série dos retornos.

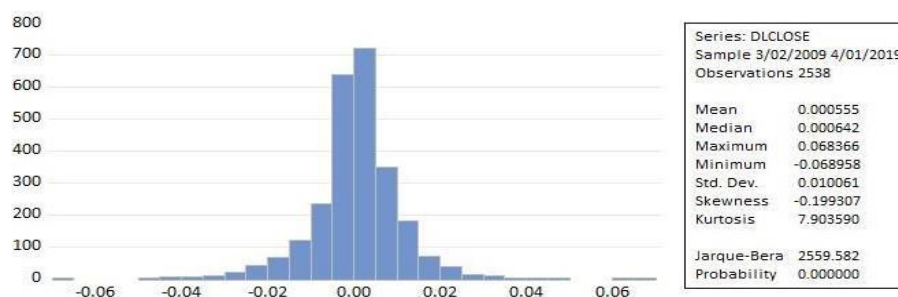


Figura 12 - Gráfico de barras e estatísticas descritivas da série reduzida dos retornos do índice *Standard & Poor's 500*

À semelhança da série global, a nova série dos retornos de SP500 apresenta o valor do coeficiente de *skewness* negativo e o valor da curtose muito superior a três, como se verifica na tabela anterior, o que indica que a série segue uma distribuição assimétrica

negativa e leptocúrtica, respetivamente. Também se pode concluir que o coeficiente de curtose da subamostra diminui em relação ao valor da curtose da amostra total.

Além disso, rejeita-se a hipótese de distribuição normal, pois o *p-value* do teste Jarque-Bera é inferior a qualquer nível de significância.

### c. Estacionariedade

Uma vez mais, faz-se o estudo à estacionariedade da nova série dos retornos da SP500, ao qual se obteve os resultados:

Testes de estacionariedade		
ADF	PP	KPSS
0,0001 *	0,0001 *	0,255026 **

\*p-value  
\*\* estatística teste

**Tabela 8** - Resultados dos testes de estacionariedade da série reduzida dos retornos do índice *Standard & Poor's 500*

Ao analisar a Tabela 8<sup>8</sup>, observa-se que tanto o teste ADF como o teste PP apresentam valores abaixo dos níveis de significância, o que leva à rejeição da hipótese nula, que patenteia a não estacionariedade da série.

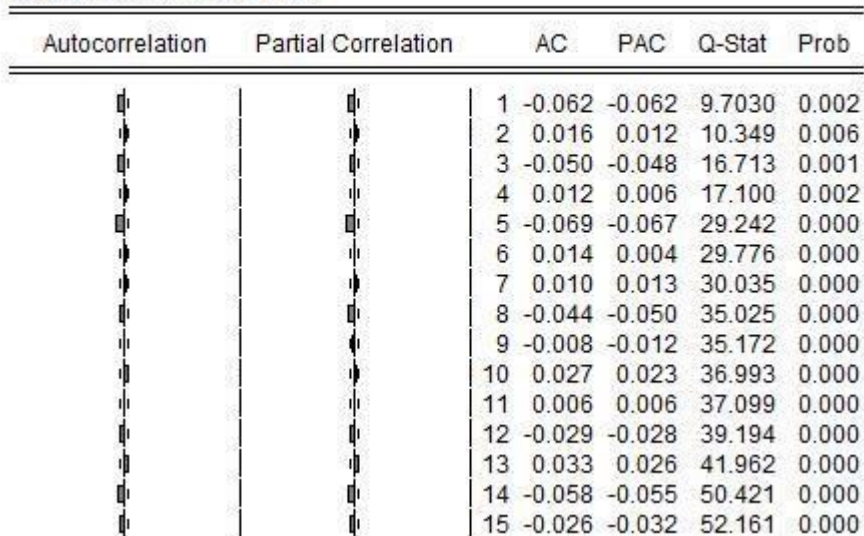
Além do mais, no teste KPSS não se rejeita a estacionariedade da série, o que leva a concluir, em conformidade com os três testes, que a série é estacionária.

### d. Modelização da Série dos retornos

De seguida, segue-se para a escolha do modelo que melhor se adequa aos dados, recorrendo, inicialmente, à determinação dos modelos potenciais. Para isso é necessário observar o correlograma da série.

<sup>8</sup>Os outputs dos testes de estacionariedade da série reduzida dos retornos da SP500 encontram-se no Anexo G – Testes de Estacionariedade da Série Reduzida .

Date: 09/26/19 Time: 10:19  
 Sample: 3/02/2009 4/01/2019  
 Included observations: 2538



**Figura 13** - Correlograma da série reduzida dos retornos do índice *Standard & Poor's 500*

Através do correlograma<sup>9</sup> presente na Figura 13, encontra-se apenas um lag, que apresenta ser significativo, tanto na função ACF como na PACF, o que leva aos modelos AR(1), MA(1) e ARMA(1,1) como modelos potenciais.

Para conseguir selecionar o modelo mais adequado aos dados, recorre-se à comparação dos critérios de informação, de todos os possíveis modelos.

Modelos	AR(1)	MA(1)	ARMA(1,1)
AIC	-6,362825	-6,363342	-6,372356
SC	-6,358222	-6,358742	-6,365452
HQC	-6,361155	-6,361673	-6,369852

**Tabela 9** - Valores dos critérios de informação de todos os modelos potenciais da série reduzida dos retornos do índice *Standard & Poor's 500*

Tendo em consideração aos valores dos critérios de informação<sup>10</sup> presentes na tabela anterior, conclui-se que o modelo ARMA(1,1) é o modelo que apresenta valores mais baixos. Logo, o modelo a ser utilizado na previsão dos valores futuros do índice, possivelmente, será este.

Dado esta conclusão, analisa-se agora os resíduos do modelo selecionado, para ter conhecimento do quão corretamente o modelo se adequa aos dados e de se o modelo possui uma variância estável ao longo do intervalo de tempo.

<sup>9</sup>O correlograma completo da série reduzida dos retornos da SP500 está presente no Anexo H – Modelização da Série Reduzida .

<sup>10</sup>Os outputs de todos os modelos possíveis encontrados da série reduzida dos retornos da SP500 encontram-se no Anexo H – Modelização da Série Reduzida .

Modelo	ARMA(1,1)
ARCH	0,0000
Breusch-Godfrey Serial Correlation	0,3961

**Tabela 10** - Resultados dos testes de correlação e de heterocedasticidade da série reduzida dos retornos do índice *Standard & Poor's 500*

De acordo com a Tabela 10, constata-se que a série dos resíduos é independente, ou seja, o modelo adequa-se aos dados, no entanto apresenta heterocedasticidade, ou seja, os resíduos têm variância instável<sup>11</sup>.

Como referido na análise da série global, independentemente, de os resíduos do modelo não validarem os pressupostos já era previsível, não quer dizer que não se obtenha uma boa previsão.

## e. Previsão da série

### Previsão *in-sample*

Nesta previsão prevê-se 2542 observações contidas no intervalo de 2 de março de 2009 a 1 de abril de 2019 para todos os modelos potenciais.

Modelos	AR(1)	MA(1)	ARMA(1,1)
RMSE	16,64764	16,64328	16,65036
MAE	11,41143	11,40879	11,41035
MAPE	0,683101	0,683130	0,682049

**Tabela 11** - Valores RMSE, MAE e MAPE de todos os modelos potenciais da série reduzida dos retornos do índice *Standard & Poor's 500*

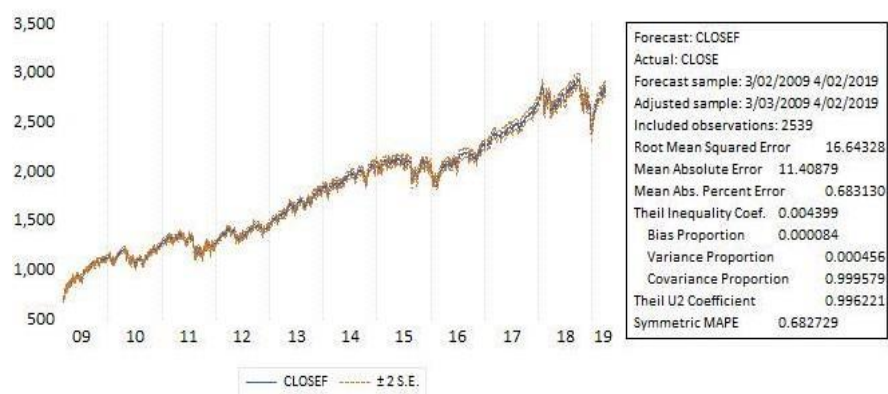
Ao verificar os valores do erro absoluto médio (MAE) e da raiz do erro quadrático médio (RMSE) de todos modelos<sup>12</sup>, conclui-se que o modelo que apresenta valores mais baixos é o modelo MA(1). Por isso, será este o melhor modelo para previsão *out-of-sample*, apesar de na comparação dos critérios de informação, ter-se optado pelo modelo ARMA(1,1).

<sup>11</sup> Os outputs dos testes de ARCH e de Breusch-Godfrey Serial Correlation estão presentes no Anexo I – Testes de Correlação e Heterocedasticidade da Série Reduzida .

<sup>12</sup> Os outputs das previsões de todos os vários modelos da série reduzida dos retornos da bolsa SP500 encontram-se no Anexo J – Previsão da Série Reduzida .

## Previsão *out-of-sample*

Em relação à previsão *out-of-sample*, esta prevê o valor do dia seguinte fora da amostra, que neste caso, se refere ao dia 2 de abril de 2019.



**Figura 14** - Previsão *out-of-sample* da série reduzida dos retornos da SP500, com o modelo MA(1)

Ao analisar a Figura 14, observa-se que o valor MAE e o valor RMSE são elevados, contudo, como na previsão de série global, isso não indica uma má previsão, pois ambos os critérios dependem da escala.

Por isso, ao verificar o coeficiente de *Theil*, a *Bias Proportion*, a *Variance Proportion* e a *Covariance Proportion*, conclui-se que os valores previstos são próximos dos valores reais, pois os três primeiros referidos se aproximam de 0 e o último de 1.

Uma vez mais, se calcula o erro de previsão para o dia 2 de abril de 2019, para obtenção de uma melhor comparação entre modelos. Neste caso, obteve-se o valor previsto de 2866.810, sendo o valor verdadeiro 2867.23999:

$$\frac{|2866.810 - 2867.23999|}{2867.23999} = 0,0001499$$

Contas feitas, o erro MAPE é aproximadamente 0.015%, e, mais uma vez, conclui-se que os modelos da família ARIMA são capazes de previsões de alta qualidade para um único valor no futuro.

## 4.4. Síntese – Previsão com Modelos Econométricos

De forma a compreender qual dos modelos obteve uma melhor previsão do índice bolsista SP500, faz-se agora um aglomerado de todas as conclusões chegadas na previsão dos modelos econométricos:

Modelos	Global	Reduzida
	AR(1)	MA(1)
RMSE	16,03021	16,64328
MAE	11,06845	11,40879
MAPE	0,798278	0,68313

**Tabela 12** - Valores RMSE, MAE e MAPE dos modelos do índice *Standard & Poor's 500*

De acordo com a Tabela 12, onde estão apresentados os valores RMSE, MAE e MAPE, o modelo obtido da série reduzida contém valores mais elevados, o que leva a concluir que o modelo AR(1) da série global, adequa-se melhor aos dados do índice bolsista SP500, do que o modelo da série reduzida. Posto isto, pode-se novamente afirmar que os modelos que melhor se adequam aos dados, podem não levar às melhores previsões.

Modelos	Global	Reduzida
	AR(1)	MA(1)
Coefficiente Theil	0,004983	0,004399
Bias Proportion	0,000031	0,000084
Variance Proportion	0,000368	0,000456
Covariance Proportion	0,999083	0,999579

**Tabela 13** - O Coeficiente de Theil, *Bias Proportion*, *Variance Proportion* e *Covariance Proportion* dos respetivos modelos da série global e da série reduzida

Porém, ao analisar a Tabela 13, não se consegue indicar qual das séries se adequa melhor aos dados, pois ambas as séries levaram a boas previsões. Esta conclusão advém do facto de o coeficiente de Theil nas duas séries ser próximo de 0, tal como o *Bias Proportion* e o *Variance Proportion*. E ainda, o resultado do *Covariance Proportion* corrobora com essa afirmação, pois é bastante próximo de 1, em ambas as séries.

Para além disso, o facto de a série global apresentar um *Bias Proportion* e um *Covariance Proportion* mais próximos de 0, e a série reduzida apresentar melhores valores no Coeficiente de Theil e no *Covariance Proportion*, leva a essa conclusão.

A fim de conseguir chegar à conclusão de qual modelo levou a uma previsão mais fidedigna, elaborou-se a seguinte tabela:

Séries	Modelos	Valores previstos	Valor Real	Erros de previsão
Global	AR(1)	2865,190	2867,240	0,0007149
Reduzida	MA(1)	2866,810		0,0001499

**Tabela 14** - Tabela síntese dos resultados de previsão *out-of-sample* no dia 2/3/2019 do índice SP500 com os modelos econométricos

Tendo em consideração a Tabela 14, pode-se, finalmente, concluir que a série reduzida obteve um melhor valor previsto em comparação com a série global. Com este resultado, pode-se afirmar que os modelos ARMA/ARIMA obtêm melhores resultados se processarem menos dados, isto porque não conseguem captar existência de vários

padrões ao longo do tempo. Porém, no caso de amostras menores, como no caso da série reduzida, os modelos mencionados já se adequam melhores para posteriores previsões. Sendo que, neste caso a diferença não foi muito significativa.

#### 4.5. Previsão com *Deep Neural Networks – Long Short Term Memory*

A previsão de séries temporais tem um grande número de aplicações no mundo atual, desde previsões de preços do mercado de ações, previsão do tempo, previsão de vendas, etc.

No entanto, neste projeto, o modelo de séries temporais utilizado é na área de previsão dos preços de mercado de ações, *Standard & Poor's 500*, em que prevê valores futuros baseados em valores observados. Isto acontece, através do estudo dos dados da série para extrair informações e padrões estatísticos úteis.

De forma a obter melhores resultados, recorre-se aos modelos de redes LSTM, que consistem na memorização de valores ao longo de um período de tempo arbitrário. Além disso, são constituídos por três portas que regulam o fluxo de informação dentro e fora da célula, tornando-o assim num modelo adequado na previsão de séries temporais financeiras.

##### a. Metodologia

##### Redes Neurais Recorrentes

As redes neurais tradicionais não permitem memorizar grande número de informações, o que leva a que não haja um raciocínio contínuo. Contudo, as RNRs conseguem resolver esse problema, pois são redes com *loops*, permitindo que as informações persistam, como está representado na figura seguinte.

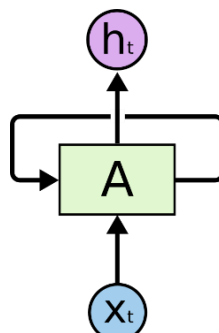
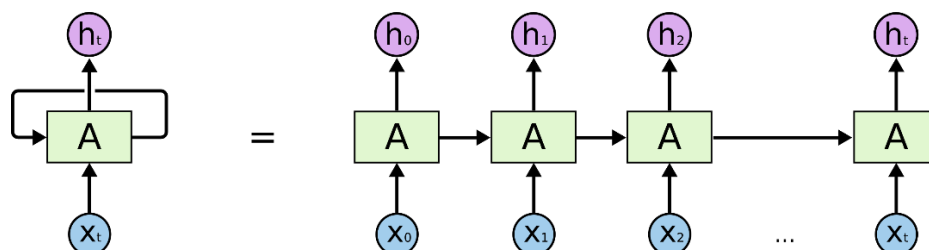


Figura 15 - Ilustração de Redes Neurais Recorrentes possuem *loops*

Onde  $A$  processa os inputs  $x_t$  e gera um valor  $h_t$ .

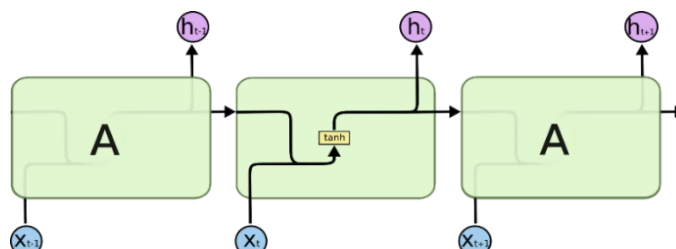
Este tipo de redes pode ser definido como sendo múltiplas cópias da mesma rede, em que há passagem de informações entre elas. Logo, pode-se afirmar que um *loop* permite que as informações passem de etapa para etapa, como é representado na próxima figura:



**Figura 16** - Ilustração de uma rede neural recorrente desenrolada

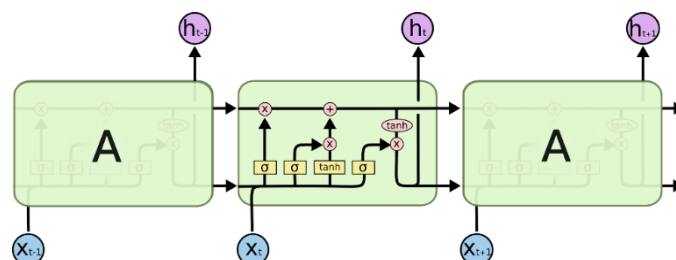
Ao longo dos últimos anos, as RNRs têm tido bastante sucesso em algumas aplicações, como o reconhecimento de fala, a modelagem de idiomas, tradução, legendagem de imagens, entre outros. Isso advém do uso dos modelos LSTMs, que proporcionam melhores resultados, em algumas tarefas, comparativamente com outros.

Esses modelos são um tipo de RNR, capazes de memorizar informações a longo prazo. Contudo, o que os diferencia é que todas as redes neurais recorrentes têm a forma de uma cadeia de módulos repetitivos da rede neural, como está representado na figura seguinte:



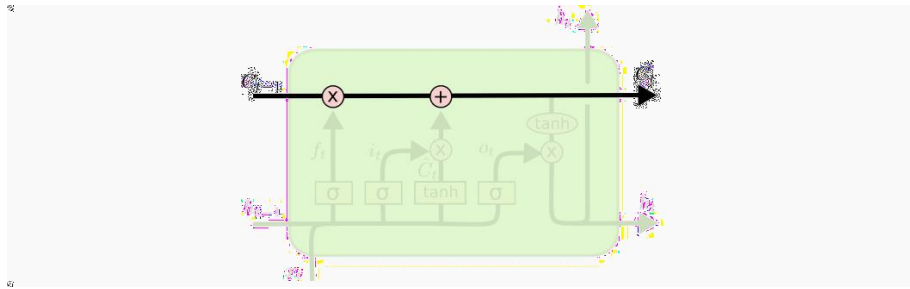
**Figura 17** - Módulo de repetição em um RNN padrão contém uma única camada

No entanto, os LSTMs ao invés de terem uma única camada de rede neural (estrutura mais simples), contêm quatro camadas, que interagem entre elas.



**Figura 18** - Módulo de repetição em um LSTM contém quatro camadas de interação

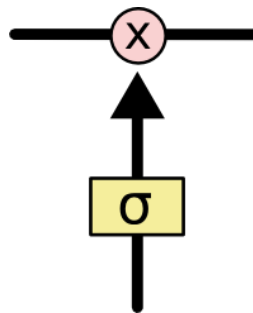
O elemento-chave destes modelos é o estado da célula, isto é, a linha horizontal que passa pela parte superior do diagrama, como representado na Figura 19. Esta percorre toda a cadeia, com apenas algumas interações lineares menores, para que as informações fluam sem alterações.



**Figura 19** - Estado da célula de um LSTM

Além disso, estas redes têm a capacidade de remover ou adicionar informações ao estado da célula (Figura 19), cuidadosamente reguladas por estruturas chamadas portas, em que um LSTM possui três, para proteger e controlar o estado da célula.

Contudo, estas portas são compostas por uma camada de rede neural sigmoide, como representado abaixo:



**Figura 20** - Camada de rede neural sigmoide

Essa camada sigmoide produz valores entre zero e um, que indica a proporção de informação que deve ser transmitida, sendo que zero não permite transpassar nenhuma informação e o valor um refere que toda a informação é importante.

Em suma, os LSTMs são adequados para problemas de dependência de longo prazo, pois têm a capacidade de reter informações passadas, enquanto tomam decisões. Por esse motivo, são bastante utilizados para previsões de séries temporais, o que levou à escolha deste modelo para prever os preços de fecho do índice SP500.

## Função de Custo e Otimização

A escolha do algoritmo de otimização a utilizar no modelo de *Deep Learning* é o um dos aspetos mais importante na previsão das séries temporais, pois é o que torna o modelo mais ou menos eficiente.

Neste projeto foi escolhido o algoritmo *Adam*, pois é um algoritmo bastante popular no campo da *Deep Learning*, que advém do facto de conseguir bons resultados rapidamente. Além disso, é uma extensão do gradiente descendente estocástico (SGD), que atualiza os pesos da rede de forma iterativa, com base nos dados de treino.

Este algoritmo foi introduzido por Diederik Kingma e Jimmy Ba, no artigo Kingma e Ba, (2014). Neste artigo referem alguns dos seus benefícios, como por exemplo, a simplicidade em implementar, a eficiência computacionalmente, os poucos requisitos de memória, entre outros.

Para além disso, os autores também descrevem o algoritmo como sendo a combinação das vantagens de outras duas extensões de gradiente estocástico. Como o algoritmo AdaGrad, este também mantém *learning rates* por-parâmetro, que melhora o desempenho em problemas com gradientes dispersos, por exemplo, na linguagem natural e em problemas de visão computacional, e como o algoritmo RMSProp, também mantém *learning rates* por-parâmetro, que são adaptadas com base na média das magnitudes recentes dos gradientes para o peso. Isso significa que o algoritmo funciona bem em problemas on-line e não estacionários.

Contudo, ao invés de adaptar as *learning rates* de parâmetros com base no primeiro momento médio (a média) como em RMSProp, Adam também faz uso da média dos segundos momentos dos gradientes (a variância não centralizada).

## **b. Modelo experimental**

Neste projeto, o modelo LSMT usa o pacote Keras, que é uma biblioteca que fornece blocos de construção altamente poderosos e abstratos para construir redes de *deep learning*. Além disso, esses blocos de construção que Keras fornece são construídos usando Theano, bem como TensorFlow, que é uma alternativa para Theano para construir gráficos computacionais, automaticamente derivando gradientes, etc.

Este suporta computação de unidade central de processamento (CPU) e unidade de processamento gráfico (GPU) e é uma ótima ferramenta para fazer protótipos de ideias rapidamente. Também implementa uma série de optimizadores, como Gradiente Descendente Estocástico (SGD), o RMSProp, o AdaGrad, entre outros, mas, neste caso, é usado o Adam.

Logo, conforme os requisitos, no modelo temos também Theano, que permite que o usuário defina expressões matemáticas que codifiquem funções de perda e, uma vez definidas, Theano permite que o usuário calcule os gradientes dessas expressões.

Ainda em relação ao modelo LSTM, este tem implementado o treino em lotes, ou seja, o modelo gera um único valor para cada lote de um determinado número de valores de entrada. Tem também uma dimensionalidade do espaço de saída que será definida e pode conter uma ou mais camadas densas totalmente conectadas no final.

Neste projeto, os dados foram divididos em duas partes, em que a primeira parte é usada para treinar o modelo e o restante para testar o mesmo. Além disso, foram usadas épocas para treinar o modelo. Foram processados os preços de fecho do índice *Standard & Poor's 500*, usando o conjunto de dados completo. Para poder usar a função de activação sigmoide, a série temporal é estandardizada para tomar valores entre 0 e 1.

De forma a obter o melhor resultado possível, foram efetuadas várias experiências<sup>13</sup>, que se basearam na execução repetitiva do código<sup>14</sup> em *Python*, à medida que se alterava os parâmetros do mesmo. Como o objetivo principal da tese é a obtenção de uma previsão fidedigna, o critério na escolha dos melhores hiperparâmetros do algoritmo LSTM foram sempre baseados no menor erro de previsão.

Inicialmente, testou-se o código para obter a melhor quantia de valores em cada lote, que variou entre 7 e 365, ao qual se chegou a uma quantia de 8 valores.

Relativamente à divisão dos dados, variou-se entre as percentagens 75%-25% e 95%-5%, em que o melhor resultado obtido foi na execução do código quando o conjunto de treino era de 90% dos dados e 10% de teste.

Também, para a determinação da quantidade de camadas densas a serem usadas no modelo, correu-se o código e obteve-se o valor de uma única camada, para obter melhores resultados. Para definição do número de neurónios na camada oculta, testou-se novamente o código várias vezes, minimizando o erro de previsão, em que se alcançou o valor de 256. A camada de saída foi definida para prever um valor de cada vez.

Por fim, em relação às épocas de treino dos dados alcançou-se um valor ótimo de 250 épocas, em que quanto maior é o número de época menor é o erro, pois é no treino que ocorrem os ajustes na rede. Entretanto, o erro diminui apenas no começo e após algumas épocas, começa a aumentar, devido ao *overfitting* do sistema.

### **c. Resultados obtidos**

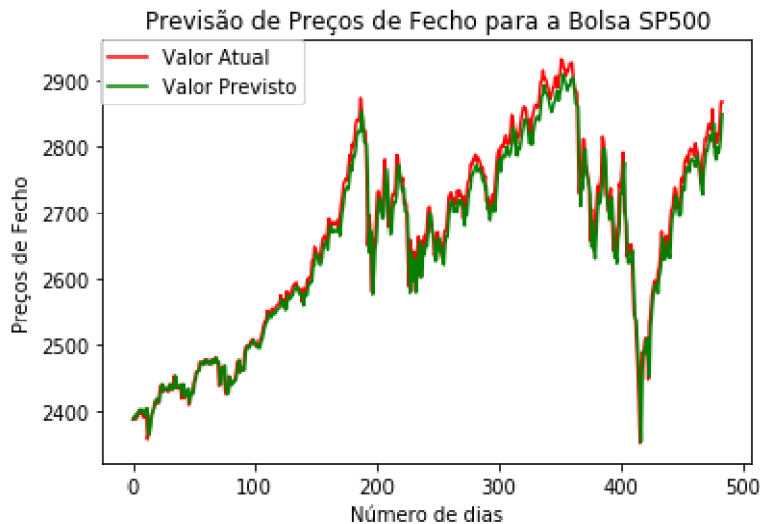
O algoritmo LSTM com a arquitetura final, determinou os seguintes resultados<sup>15</sup>: na Figura 21, temos a previsão sobre o conjunto de teste, isto é, os últimos 10% da amostra total. Observa-se um padrão bastante semelhante entre a linha dos valores preditos e valores verdadeiros da série. Seguidamente, a Tabela 15 representa os vários erros de previsão.

---

<sup>13</sup> Em Anexo K – Código e Experiências

<sup>14</sup> Em Anexo K – Código e Experiências

<sup>15</sup> Ver em Anexo L – Previsão da Série com DL - LSTM



**Figura 21** - Previsão do índice bolsista SP500 com o modelo DNN

Erros	Valores obtidos
RMSE	564,952041
RMSE em %	0,86
MAPE	0,00609598

**Tabela 15** - Valores obtidos na previsão do índice bolsista SP500 como o modelo DNN

De forma a facilitar a compreensão dos valores obtidos na execução do código, é feita uma breve explicação do significado de cada valor presente.

Com se tinha referido anteriormente, o programa é percorrido por 250 épocas e por cada época é obtido uma função de perda (*loss function*), que representa a soma dos resíduos ao quadrado, ou seja, a função que otimizamos no método dos mínimos quadrados.

Contudo, como se pode observar nos resultados<sup>16</sup>, está presente em cada época um valor *loss*, que representa o valor barreira, em que no critério de otimização, os resíduos devem ser inferiores a este valor.

Para além disso, está também presente um valor *loss\_value*, que indica o valor da função perda, ou seja, a soma dos quadrados dos resíduos.

De seguida, é obtido o valor *Mean squared error*, que é o erro quadrático médio, que mede o desempenho da previsão do código, ou seja, quanto menor for este valor, melhor será a previsão.

No entanto, ao analisar esse valor, pode-se verificar que é um valor elevado, contudo os erros próximos de 500 são considerados bons, pois na escala de 0 a 1, são equivalentes a 0.0001.

<sup>16</sup> Ver em Anexo L– Previsão da Série com DL - LSTM

Contudo, ao observar o último valor previsto, dia 2 de abril de 2019, pode-se calcular o erro, da seguinte forma:

$$\frac{|2852.0828 - 2867.23999|}{2867.23999} = 0.005286$$

Ao considerar o valor anterior, pode-se concluir que é uma boa previsão, pois o erro de previsão é relativamente bom, o que indica que o valor previsto está próximo do valor real do fecho do índice bolsista SP500 do dia 2 de abril de 2019. Nota-se que este valor, de facto, é predito numa amostra de teste desconhecida de cerca 500 observações, ou seja, antes deste valor são preditos todos os outros tendo como base só a aprendizagem retirada do conjunto de treino. Visto desta forma, a capacidade de previsão da rede é excelente.

## **4.6. Previsão com *Deep Neural Networks – Long Short Term Memory* (Reduzida)**

### **a. Modelo experimental**

Da mesma forma que ocorreu no método econométrico, a série é dividida com intuito de estudar o desempenho da rede numa amostra menor. Assim sendo, a quebra da série é feita no dia 2 de março de 2009, isto é, os dados da série reduzida têm início no dia 2 de março de 2009 até dia 2 de abril de 2019, não incluindo fins-de-semana.

Uma vez mais, foram efetuadas várias experiências, alterando os parâmetros do código<sup>17</sup>, para obter os melhores resultados.

Tendo procedido da mesma forma, como anteriormente foi descrito, chegou-se aos seguintes valores: lotes de 8, percentagem de 85% dos dados no conjunto de treino e 15% no teste, dimensionalidade de 256 neurónio na camada oculta, uma única camada densa e 200 épocas.

---

<sup>17</sup> Em Anexo K – Código e Experiências

## b. Resultados obtidos

Após várias tentativas, obteve-se os seguintes resultados:



Figura 22 - Previsão da série reduzida do índice bolsista SP500 com o modelo DNN

Erros	Valores obtidos
RMSE	652,104347
RMSE em %	0,93
MAPE	0,00663777

Tabela 16 - Valores obtidos na previsão da série reduzida do índice bolsista SP500 como o modelo DNN

Na Tabela 16 pode-se observar o valor RMSE, que como dito anteriormente, mede o desempenho da previsão do código, que, neste caso, apresenta um valor elevado. Porém, os erros próximos de 500 são considerados bons, pois na escala de 0 a 1, são equivalentes a 0.0001, no entanto, como se pode verificar este valor já se apresenta um valor afastado de 500, mas mesmo assim, é um erro relativamente bom.

Uma vez obtido o último valor previsto de 2848.6018, calcula-se agora o erro de previsão, da seguinte forma:

$$\frac{|2848.6018 - 2867.23999|}{2867.23999} = 0.0065$$

Ao analisar o erro obtido, considera-se como um bom erro de previsão, pois não se encontra muito afastado do valor real do fechamento do índice bolsista SP500 do dia 2 de abril de 2019.

## 4.7. Síntese – Previsão com Deep Neural Networks

A fim de concluir qual das séries (global e reduzida) levou a melhores resultados com o modelo *Deep Neural Network*, com *Long Short-Term Memory*, elaborou-se a seguinte tabela:

Séries	Valores previstos	Valor Real	Erros de previsão	RMSE	RMSE em %	MAPE
Global	2852,083	2867,240	0,0052863	564,952041	0,86	0,006096
Reduzida	2848,602		0,0065003	652,104347	0,93	0,006638

Tabela 17 - Tabela síntese dos resultados de previsão do índice SP500 com o modelo DNN

Tendo em consideração a Tabela 17, determina-se que a série global apresenta um valor mais próximo do valor real, uma vez que gerou um erro de previsão menor. Além disso, ao observar os valores RMSE, em absoluto e em percentagem, e MAPE, verifica-se que os valores menores estão presentes na série global, o que indica que é uma boa previsão e o que o modelo se ajusta adequadamente aos dados. Com isto, conclui-se que o modelo DNN apresenta uma melhor previsão, quando contém um grande número de dados.

Este modelo se adequa bem aos dados, devido à sua capacidade de fazer ajuste durante as épocas de treino de dados que quanto maior é o número de épocas menor é o RMSE, porém o erro diminui apenas no começo, pois após algumas épocas começa de novo a aumentar, devido ao *overfitting*.

## 4.8. Comparação entre Modelos

Ao aglomerar todos os resultados obtidos na previsão no dia 2 de abril de 2019, obtém-se o seguinte quadro:

Séries	Modelos	Valores previstos	Valor Real	Erros de previsão	MAPE
Global	AR(1)	2865,190	2867,240	0,0007150	0,798278
Reduzida	MA(1)	2866,810		0,0001500	0,683130
Global	DNN	2852,083		0,0052863	0,609598
Reduzida	DNN	2848,602		0,0065003	0,663777

Tabela 18 - Valores previstos, erros de previsão e os valores de MAPE obtidos na previsão de todos os modelos

Ao analisar a tabela anterior, a partir da coluna dos erros de previsão, afirma-se que o modelo que apresentou um valor previsto mais próximo do valor real é o modelo MA(1) da série reduzida.

Em relação ao MAPE, relembra-se que este valor é associado a amostras diferentes em termos de números de observações, nomeadamente, no caso dos modelos ARMA temos a amostra inteira, enquanto que para os modelos LSTM só temos o conjunto de teste. Em proporção (10% e 15% respetivamente), o MAPE dos modelos

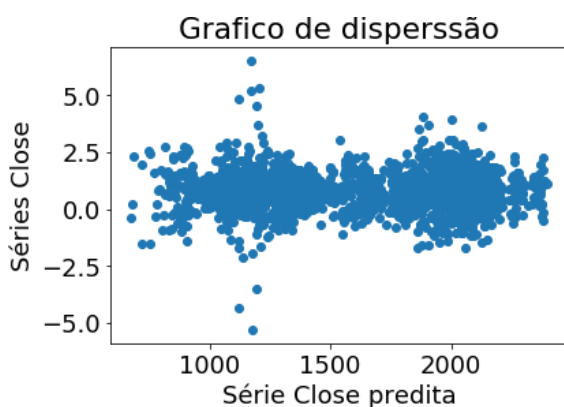
ARMA fica em 0.079 para AR(1) e 0.1024 para MA(1), e sendo assim, as redes recorrentes de tipo LSTM, dominam a qualidade de previsão.

Da mesma forma, como explicado anteriormente, o valor predito fora de amostra, para o dia 2 de abril de 2019, pela aplicação dos modelos LSTM, consta no valor final predito no conjunto de teste, sendo todos os outros valores do conjunto de teste preditos também no mesmo algoritmo. Fazendo um simples exercício, e considerando nos modelos LSTM o conjunto de treino como sendo em proporção de 0.9999% da amostra total, resulta que o conjunto de teste iria conter apenas um dia para prever, nomeadamente, o dia 2 de abril de 2019. No caso da amostra completa, obtém-se o seguinte erro para a previsão de um valor:

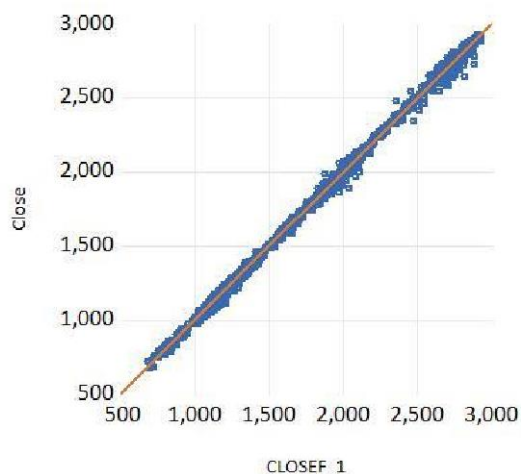
$$\frac{|2867.9255 - 2867.23999|}{2867.23999} = 0.000239$$

que já é comparável com os erros obtidos pela previsão com os modelos ARMA.

Visto de outra forma, a previsão a partir dos modelos ARMA, pode ser representada graficamente usando os gráficos de dispersão, onde no eixo dos xx temos a série predito pelo modelo e no eixo dos yy temos a série temporal original.



**Figura 23** - Gráfico de dispersão dos valores previstos com o modelo MA(1) e os valores reais



**Figura 24** - Gráfico de dispersão dos valores previstos (-1) com o modelo MA(1) e os valores reais

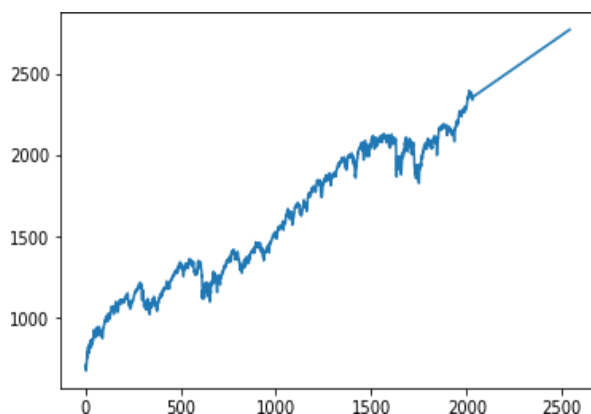
Ao analisar o gráfico de dispersão anterior, que é uma nuvem aleatória, conclui-se que para séries temporais longas e com dinâmicas complexas, os modelos ARMA/ARIMA levam a previsões *in-sample* e *out-of-sample* “bastante naïves”. Isto é, devido ao facto de os modelos econométricos ARMA/ARIMA só terem em consideração o valor real do dia anterior, ou seja, como no dia 1 de abril de 2019, o valor de fecho do índice bolsista SP500 foi de 2867,19, o modelo tradicional ARMA/ARIMA prevê que o valor do dia seguinte será próximo desse valor, e assim adiante, ou seja, o valor predito de amanhã é o valor observado hoje. Para comprovar isto, representa-se graficamente o gráfico de dispersão, onde no eixo dos xx temos a série predita desfasada com uma unidade e no eixo dos yy temos a série original. Observa-se a correlação quase perfeita

entre as duas séries, o que indica que os valores previstos têm em consideração só o valor do dia anterior, o que na verdade deveria considerar o comportamento da bolsa na sua totalidade.

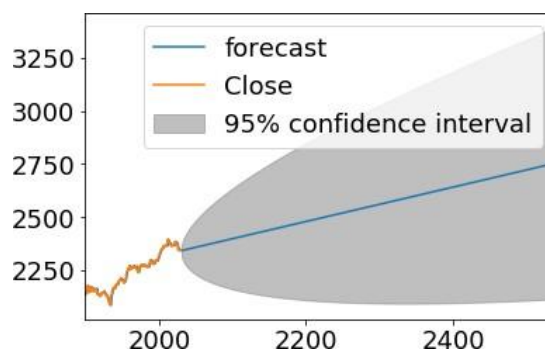
Pelos motivos ditos anteriormente, independentemente de os modelos ARMA/ARIMA indicarem relativamente melhores resultados para a previsão de um único valor no futuro, os modelos DNN-LSTM são mais adequados para previsões de séries temporais financeiras, pois estes têm em consideração todo o comportamento da série, descobrem padrões não-lineares e têm uma capacidade de aprendizagem e de generalização que os modelos econométricos ARMA/ARIMA não têm.

De forma a aprofundar a comparação entre os métodos, foi realizada a previsão de vários valores, tanto para o modelo ARMA como para o modelo DNN-LSTM. Na amostra reduzida, prevemos 507 valores com ARMA e 507 valores com LSTM (o que equivale ao conjunto de teste de LSTM).

Ao observar a Figura 25, conclui-se que apesar de o ARMA ser o melhor na previsão de um valor *out-of-sample*, na previsão de vários valores no futuro, as redes neurais levam a previsões mais fidedignas, sendo que os modelos ARMA são incapazes de prever 50 ou 500 valores.



**Figura 25** - Previsão *out-of-sample* de 507 valores com o modelo ARMA



**Figura 26** - Previsão *out-of-sample* com intervalo de confiança de 95%

	ARMA	LSTM
<b>RMSE%</b>	12.36%	3%

**Tabela 19** - Resultados obtidos na previsão de 507 valores previsto tanto em MA(1) como em DNN-LSTM

Para além disso, ao verificar a Tabela 19, o modelo MA(1) levou a uma RMSE em percentagem muito mais elevado que no modelo DNN-LSTM, o que corrobora com a afirmação anterior, em que os modelos ARMA/ARIMA não obtêm previsões fidedignas em previsão de vários valores fora da amostra.

## Conclusão

O objetivo principal desta dissertação consistia na determinação do método que resulta em previsões de séries temporais financeiras mais precisas. Para isso, o estudo empírico teve como base os valores de fecho do índice bolsista SP500, sendo este um dos principais índices de bolsa de valores. Além disso, a fim de obter o melhor resultado possível foi feita uma quebra na série temporal financeira, tendo início a partir do após-crise financeira de 2007-2008.

Neste projeto, as previsões foram efetuadas sobre a série SP500 usando os modelos ARMA/ARIMA e os métodos de *Machine Learning*, mais especificamente os modelos de *Deep Neural Network*, com *Long Short-Term Memory*.

A utilização dos modelos ARMA/ARIMA, nesta tese, advém do facto do seu histórico nas previsões de séries temporais financeiras, porém geralmente levam a resultados mais fidedignos, quando são efetuados com base num intervalo de tempo pequeno. Isto é, quanto maior é a quantidade de dados, maior é a dificuldade deste tipo de modelos, enquanto modelos lineares, de captarem quebras de estrutura e não-linearidades presentes na série.

Por outro lado, a escolha do modelo DNN-LSTM tem por base o facto de o método ser dos mais recentes utilizados na previsão de séries temporais financeiras, de ter memória e capacidade de reproduzir padrões não-lineares e de ser um dos modelos mais bem-sucedidos de *Deep Learning*.

Ao analisar os erros de previsão obtidos pelos modelos, chegou-se à conclusão de que os modelos ARMA/ARIMA levaram a melhores previsões para o dia 2 de abril de 2019 do que o modelo DNN-LSTM.

Contudo, apesar de os modelos ARMA/ARIMA apresentarem valores mais próximos dos valores reais de fecho do índice bolsista SP500, não apresentam previsões “verdadeiras”, pois estes modelos só projetam os valores do dia anterior quando encontram graus elevados de variância e complexidade, o que ao contrário do modelo DNN-LSTM, que considera todos os padrões no comportamento da bolsa. Por isso, independentemente de ter apresentado valores relativamente mais afastados da realidade, nas terceiras casas decimais, o modelo DNN-LSTM continua a ser o mais adequado para previsão de séries temporais financeiras. A nossa conclusão fortalece-se quando se faz uma previsão para um intervalo de tempo maior e onde as redes de tipo LSTM mostram a sua superioridade na previsão.

Nesta dissertação, como em qualquer outro projeto, foram encontradas algumas limitações durante a sua realização, o que pode ter influenciado nos resultados. A principal limitação encontra-se relacionada com uma questão de material informático, pois para a execução do código é necessário ocorrer em computadores relativamente “bons”, ou seja, com alguma potência, pois a sua execução demorava mais tempo do que o suposto. Sendo assim, com mais tempo, podiam-se alcançar modelos com um desempenho ainda melhor. Contudo, não podemos deixar de referir, que toda a parte de

redes neuronais, *Machine Learning*, modelos LSTM foram alvo de autoestudo, não fazendo parte do programa do mestrado.

## Referências Bibliográficas

- Arevalo, A. (2016). Short-Term Forecasting of Financial Time Series with Deep Neural Networks, 42. Disponível em: <http://www.bdigital.unal.edu.co/54538/>. (Acedido a: 25-07-2019)
- Box, G.; Jenkins, G.; Reinsel, G. (1994). Time Series Analysis: Forecasting and Control. *New Jersey: Prentice Hall*.
- Brownlee, J. (2017). Gentle Introduction to the Adam Optimization Algorithm for Deep Learning. Disponível em: <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>. (Acedido a: 25-07-2019)
- Dickey, D. e Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*. 74, 427-431
- Ferreira, P. (2016). Princípio de Econometria. *Carcavelos, Rei dos Livros*.
- Heaton, J. B., Polson, N. G., e Witte, J. H. (2016). Deep Learning in Finance, (February), 1–20. <https://doi.org/10.1016/j.chest.2016.05.010>
- Henrique, B. M., Sobreiro, V. A., e Kimura, H. (2019). Literature Review: Machine Learning Techniques Applied to Financial Market Prediction. *Expert Systems with Applications*. <https://doi.org/10.1016/J.ESWA.2019.01.012>
- Keras: The Python Deep Learning library, “Keras Documentation. Disponível em: <https://keras.io/>. (Acedido a: 25-07-2019)
- Ketkar, N. (2017). *Deep Learning with Python*. <https://doi.org/10.1007/978-1-4842-2766-4>
- Kingma, D. P., e Ba, J. (2014). Adam: A Method for Stochastic Optimization, 1–15. Disponível em: <http://arxiv.org/abs/1412.6980>. (Acedido a: 08-08-2019)
- Kohzadi, N., Boyd, M. S., Kermanshahi, B., e Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, 10(2), 169–181. [https://doi.org/10.1016/0925-2312\(95\)00020-8](https://doi.org/10.1016/0925-2312(95)00020-8)
- Kwiatkowski, D.; Phillips, P.; Schmidt, P. e Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*. 54, 159-178.
- Long Short-Term Memory (LSTM). (2018). Disponível em: <https://developer.nvidia.com/discover/lstm>. (Acedido a: 20-09-2019)
- Makridakis, S.; Hibon, M.(1997). ARMA Models and the Box-Jenkins Methodology. *Journal of Forecasting*. (16)3, 147-163
- Makridakis, S.; Wheelwright, S. e Hyndman, R. (1998). Forecasting: Methods

Applications. 3ª Edição. *Nova Iorque: John Wiley e Sons.*

- Mantovani, G. (2004). *Previsão De Séries Temporais : Redes Neurais Artificiais Vs . Modelos Arima.* Universidade Federal Do Rio Grande Do Sul.
- Olah, C. (2015). Understanding LSTM Networks. Disponível em: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>. (Acedido a: 25-07-2019)
- Pai, P.-F., e Lin, C.-S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, 33(6), 497–505. <https://doi.org/10.1016/J.OMEGA.2004.07.024>
- Phillips, P. e Perron, P. (1988). Testing for a unit root in time series analysis. *Biometrika*. 75, 335-346.
- Qian, X.-Y., e Gao, S. (2017). Financial Series Prediction: Comparison Between Precision of Time Series Models and Machine Learning Methods, 2017(i), 1–9. <https://doi.org/10.1063/1.107544>
- RSNA Pneumonia Detection Challenge | Kaggle. Disponível em: <https://www.kaggle.com/amarpreetsingh/stock-prediction-lstm-using-keras>. (Acedido a: 25-07-2019)
- Siarni-Namini, S., e Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM, 1–19. Disponível em: <http://arxiv.org/abs/1803.06386>. (Acedido a 08-08-2019)
- Tang, Z., de Almeida, C., e Fishwick, P. A. (1991). Time series forecasting using neural networks vs. Box- Jenkins methodology. *SIMULATION*, 57(5), 303–310. <https://doi.org/10.1177/003754979105700508>
- Yu, H. Y., Niu, X. Y., Lin, H. J., Ying, Y. B., Li, B. B., e Pan, X. X. (2009). A feasibility study on on-line determination of rice wine composition by Vis–NIR spectroscopy and least-squares support vector machines. *Food Chemistry*, 113(1), 291–296. <https://doi.org/10.1016/J.FOODCHEM.2008.06.083>

# Anexos

## Anexo A – Testes de Estacionariedade da Série Global

Null Hypothesis: CLOSE has a unit root  
Exogenous: Constant  
Lag Length: 2 (Automatic - based on SIC, maxlag=31)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.726969	0.9928
Test critical values:		
1% level	-3.431520	
5% level	-2.861942	
10% level	-2.567027	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
Dependent Variable: D(CLOSE)  
Method: Least Squares  
Date: 09/26/19 Time: 09:52  
Sample (adjusted): 1/06/2000 4/01/2019  
Included observations: 4838 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CLOSE(-1)	0.000319	0.000439	0.726969	0.4673
D(CLOSE(-1))	-0.056789	0.014384	-3.948139	0.0001
D(CLOSE(-2))	-0.046448	0.014367	-3.233041	0.0012
C	-0.152105	0.705295	-0.215661	0.8293

R-squared	0.005128	Mean dependent var	0.302828
Adjusted R-squared	0.004511	S.D. dependent var	16.05307
S.E. of regression	16.01682	Akaike info criterion	8.385983
Sum squared resid	1240108.	Schwarz criterion	8.391344
Log likelihood	-20281.69	Hannan-Quinn criter.	8.387865
F-statistic	8.306317	Durbin-Watson stat	1.998048
Prob(F-statistic)	0.000017		

**Figura 27** - Output do teste de estacionariedade ADF da série global do índice SP500

Null Hypothesis: CLOSE has a unit root  
Exogenous: Constant  
Bandwidth: 31 (Newey-West automatic) using Bartlett kernel

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	0.928101	0.9959
Test critical values:		
1% level	-3.431519	
5% level	-2.861941	
10% level	-2.567026	

\*MacKinnon (1996) one-sided p-values.

Residual variance (no correction) 258.1806  
HAC corrected variance (Bartlett kernel) 184.0000

Phillips-Perron Test Equation  
Dependent Variable: D(CLOSE)  
Method: Least Squares  
Date: 09/26/19 Time: 09:52  
Sample (adjusted): 1/04/2000 4/01/2019  
Included observations: 4840 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CLOSE(-1)	0.000210	0.000440	0.477815	0.6328
C	-0.027592	0.707096	-0.039022	0.9689

R-squared	0.000047	Mean dependent var	0.291729
Adjusted R-squared	-0.000159	S.D. dependent var	16.07004
S.E. of regression	16.07132	Akaike info criterion	8.392363
Sum squared resid	1249594.	Schwarz criterion	8.395043
Log likelihood	-20307.52	Hannan-Quinn criter.	8.393304
F-statistic	0.228307	Durbin-Watson stat	2.104792
Prob(F-statistic)	0.632804		

**Figura 28** - Output do teste de estacionariedade PP da série global do índice SP500

Null Hypothesis: CLOSE is stationary  
Exogenous: Constant  
Bandwidth: 54 (Newey-West automatic) using Bartlett kernel

	LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic	6.116871
Asymptotic critical values*:	
1% level	0.739000
5% level	0.463000
10% level	0.347000

\*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction) 276296.7  
HAC corrected variance (Bartlett kernel) 14928203

KPSS Test Equation  
Dependent Variable: CLOSE  
Method: Least Squares  
Date: 09/26/19 Time: 09:53  
Sample: 1/03/2000 4/01/2019  
Included observations: 4841

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1520.048	7.555534	201.1834	0.0000

R-squared	0.000000	Mean dependent var	1520.048
Adjusted R-squared	0.000000	S.D. dependent var	525.6936
S.E. of regression	525.6936	Akaike info criterion	15.36752
Sum squared resid	1.34E+09	Schwarz criterion	15.36886
Log likelihood	-37196.08	Hannan-Quinn criter.	15.36799
Durbin-Watson stat	0.000935		

**Figura 29** - Output do teste de estacionariedade KPSS da série global do índice SP500

## Anexo B – Testes de Estacionariedade da Série Global dos Retornos

Null Hypothesis: DLCLOSE has a unit root  
Exogenous: Constant  
Lag Length: 1 (Automatic - based on SIC, maxlag=31)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-54.06545	0.0001
Test critical values:		
1% level	-3.431520	
5% level	-2.861942	
10% level	-2.567027	

\*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(DLCLOSE)

Method: Least Squares

Date: 09/26/19 Time: 09:55

Sample (adjusted): 1/06/2000 4/01/2019

Included observations: 4838 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLCLOSE(-1)	-1.137446	0.021038	-54.06545	0.0000
D(DLCLOSE(-1))	0.058563	0.014343	4.083094	0.0000
C	0.000167	0.000172	0.972115	0.3310
R-squared	0.538799	Mean dependent var	1.98E-06	
Adjusted R-squared	0.538608	S.D. dependent var	0.017626	
S.E. of regression	0.011973	Akaike info criterion	-6.011735	
Sum squared resid	0.693090	Schwarz criterion	-6.007714	
Log likelihood	14545.39	Hannan-Quinn criter.	-6.010323	
F-statistic	2824.251	Durbin-Watson stat	1.998116	
Prob(F-statistic)	0.000000			

Figura 30 - Output do teste de estacionariedade ADF da série global dos retornos do índice SP500

Null Hypothesis: DLCLOSE has a unit root  
Exogenous: Constant  
Bandwidth: 22 (Newey-West automatic) using Bartlett kernel

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-75.91357	0.0001
Test critical values:		
1% level	-3.431519	
5% level	-2.861942	
10% level	-2.567027	

\*Mackinnon (1996) one-sided p-values.

Residual variance (no correction)	0.000144
HAC corrected variance (Bartlett kernel)	0.000118

Phillips-Perron Test Equation

Dependent Variable: D(DLCLOSE)

Method: Least Squares

Date: 09/26/19 Time: 09:55

Sample (adjusted): 1/05/2000 4/01/2019

Included observations: 4839 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLCLOSE(-1)	-1.074443	0.014324	-75.00986	0.0000
C	0.000158	0.000172	0.919348	0.3580
R-squared	0.537725	Mean dependent var	1.05E-05	
Adjusted R-squared	0.537630	S.D. dependent var	0.017634	
S.E. of regression	0.011991	Akaike info criterion	-6.008911	
Sum squared resid	0.695481	Schwarz criterion	-6.006231	
Log likelihood	14540.56	Hannan-Quinn criter.	-6.007970	
F-statistic	5626.479	Durbin-Watson stat	2.008661	
Prob(F-statistic)	0.000000			

Figura 31 - Output do teste de estacionariedade PP da série global dos retornos do índice SP500

Null Hypothesis: DLCLOSE is stationary  
Exogenous: Constant  
Bandwidth: 25 (Newey-West automatic) using Bartlett kernel

	LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic	0.334069
Asymptotic critical values*:	
1% level	0.739000
5% level	0.463000
10% level	0.347000

\*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	0.000145
HAC corrected variance (Bartlett kernel)	0.000103

KPSS Test Equation

Dependent Variable: DLCLOSE

Method: Least Squares

Date: 09/26/19 Time: 09:56

Sample (adjusted): 1/04/2000 4/01/2019

Included observations: 4840 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000140	0.000173	0.809968	0.4180
R-squared	0.000000	Mean dependent var	0.000140	
Adjusted R-squared	0.000000	S.D. dependent var	0.012035	
S.E. of regression	0.012035	Akaike info criterion	-6.001763	
Sum squared resid	0.700905	Schwarz criterion	-6.000423	
Log likelihood	14525.27	Hannan-Quinn criter.	-6.001293	
Durbin-Watson stat	2.146478			

Figura 32 - Output do teste de estacionariedade KPSS da série global dos retornos do índice SP500

## Anexo C – Modelização da Série Global dos Retornos

Date: 09/26/19 Time: 09:56  
 Sample: 1/03/2000 4/01/2019  
 Included observations: 4840

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.074	-0.074	26.829	0.000		
2	-0.053	-0.059	40.267	0.000		
3	0.023	0.014	42.759	0.000		
4	-0.011	-0.011	43.359	0.000		
5	-0.042	-0.042	52.067	0.000		
6	0.005	-0.003	52.176	0.000		
7	-0.023	-0.028	54.819	0.000		
8	0.012	0.010	55.529	0.000		
9	-0.011	-0.013	56.096	0.000		
10	0.022	0.021	58.486	0.000		
11	-0.019	-0.018	60.264	0.000		
12	0.043	0.041	69.061	0.000		
13	0.017	0.021	70.392	0.000		
14	-0.038	-0.031	77.232	0.000		
15	-0.044	-0.048	86.847	0.000		
16	0.064	0.052	106.85	0.000		
17	0.010	0.021	107.35	0.000		
18	-0.052	-0.044	120.65	0.000		
19	-0.003	-0.012	120.69	0.000		
20	0.019	0.010	122.53	0.000		
21	-0.037	-0.030	129.14	0.000		
22	0.026	0.019	132.31	0.000		
23	0.004	0.003	132.39	0.000		
24	-0.004	-0.002	132.46	0.000		
25	-0.003	-0.006	132.52	0.000		
26	0.001	-0.002	132.53	0.000		
27	0.023	0.031	135.07	0.000		
28	-0.008	-0.009	135.40	0.000		
29	0.016	0.013	136.72	0.000		
30	0.010	0.014	137.21	0.000		
31	-0.008	0.007	137.51	0.000		
32	0.011	0.005	138.08	0.000		
33	-0.010	-0.014	138.62	0.000		
34	-0.068	-0.063	160.94	0.000		
35	0.014	0.003	161.88	0.000		
36	0.022	0.014	164.15	0.000		

**Figura 33** - Correlograma da série global dos retornos do índice SP500

Dependent Variable: D(LOG(CLOSE))  
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)  
 Date: 09/26/19 Time: 09:57  
 Sample (adjusted): 1/05/2000 4/01/2019  
 Included observations: 4839 after adjustments  
 Convergence achieved after 2 iterations  
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000148	0.000160	0.919408	0.3579
AR(1)	-0.074443	0.014324	-5.197089	0.0000
R-squared	0.005553	Mean dependent var	0.000148	
Adjusted R-squared	0.005347	S.D. dependent var	0.012023	
S.E. of regression	0.011991	Akaike info criterion	-6.008911	
Sum squared resid	0.695481	Schwarz criterion	-6.006231	
Log likelihood	14540.56	Hannan-Quinn criter.	-6.007970	
F-statistic	27.00973	Durbin-Watson stat	2.008661	
Prob(F-statistic)	0.000000			
Inverted AR Roots	-0.7			

Dependent Variable: D(LOG(CLOSE))  
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)  
 Date: 09/26/19 Time: 09:58  
 Sample (adjusted): 1/06/2000 4/01/2019  
 Included observations: 4838 after adjustments  
 Convergence achieved after 2 iterations  
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000147	0.000151	0.972259	0.3310
AR(1)	-0.078883	0.014358	-5.493976	0.0000
AR(2)	-0.058563	0.014343	-4.083094	0.0000
R-squared	0.008968	Mean dependent var	0.000148	
Adjusted R-squared	0.008558	S.D. dependent var	0.012024	
S.E. of regression	0.011973	Akaike info criterion	-6.011735	
Sum squared resid	0.693090	Schwarz criterion	-6.007714	
Log likelihood	14545.39	Hannan-Quinn criter.	-6.010323	
F-statistic	21.87534	Durbin-Watson stat	1.998116	
Prob(F-statistic)	0.000000			
Inverted AR Roots	-0.4-.24i	-0.4+.24i		

**Figura 34** - Output do modelo AR(1) da série global dos retornos do índice SP500 **Figura 35** - Output do modelo AR(2) da série global dos retornos do índice SP500

Dependent Variable: D(LOG(CLOSE))  
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)  
 Date: 09/26/19 Time: 09:58  
 Sample (adjusted): 1/04/2000 4/01/2019  
 Included observations: 4840 after adjustments  
 Failure to improve likelihood (non-zero gradients) after 5 iterations  
 Coefficient covariance computed using outer product of gradients  
 MA Backcast: 1/03/2000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000141	0.000158	0.889921	0.3736
MA(1)	-0.083706	0.014328	-5.842292	0.0000
R-squared	0.006242	Mean dependent var	0.000140	
Adjusted R-squared	0.006037	S.D. dependent var	0.012035	
S.E. of regression	0.011999	Akaike info criterion	-6.007611	
Sum squared resid	0.696530	Schwarz criterion	-6.004932	
Log likelihood	14540.42	Hannan-Quinn criter.	-6.006671	
F-statistic	30.38949	Durbin-Watson stat	1.988600	
Prob(F-statistic)	0.000000			
Inverted MA Roots	.08			

**Figura 36** - Output do modelo MA(1) da série global dos retornos do índice SP500

Dependent Variable: D(LOG(CLOSE))  
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)  
 Date: 09/26/19 Time: 09:59  
 Sample (adjusted): 1/04/2000 4/01/2019  
 Included observations: 4840 after adjustments  
 Failure to improve likelihood (non-zero gradients) after 4 iterations  
 Coefficient covariance computed using outer product of gradients  
 MA Backcast: 12/31/1999 1/03/2000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000141	0.000150	0.939546	0.3475
MA(1)	-0.077694	0.014362	-5.409811	0.0000
MA(2)	-0.052285	0.014361	-3.640906	0.0003
R-squared	0.008816	Mean dependent var	0.000140	
Adjusted R-squared	0.008406	S.D. dependent var	0.012035	
S.E. of regression	0.011984	Akaike info criterion	-6.009791	
Sum squared resid	0.694726	Schwarz criterion	-6.005772	
Log likelihood	14546.69	Hannan-Quinn criter.	-6.008380	
F-statistic	21.51050	Durbin-Watson stat	1.998621	
Prob(F-statistic)	0.000000			
Inverted MA Roots	.27	-.19		

**Figura 37** - Output do modelo MA(2) da série global dos retornos do índice SP500

Dependent Variable: D(LOG(CLOSE))  
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)  
 Date: 09/26/19 Time: 09:59  
 Sample (adjusted): 1/05/2000 4/01/2019  
 Included observations: 4839 after adjustments  
 Failure to improve likelihood (non-zero gradients) after 14 iterations  
 Coefficient covariance computed using outer product of gradients  
 MA Backcast: 1/04/2000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000148	0.000148	1.001255	0.3168
AR(1)	0.398582	0.124283	3.207042	0.0013
MA(1)	-0.482372	0.118727	-4.062868	0.0000
R-squared	0.007650	Mean dependent var	0.000148	
Adjusted R-squared	0.007240	S.D. dependent var	0.012023	
S.E. of regression	0.011980	Akaike info criterion	-6.010609	
Sum squared resid	0.694015	Schwarz criterion	-6.006588	
Log likelihood	14545.67	Hannan-Quinn criter.	-6.009197	
F-statistic	18.64022	Durbin-Watson stat	1.989600	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.40			
Inverted MA Roots	.48			

**Figura 38** - Output do modelo ARMA(1,1) da série global dos retornos do índice SP500

Dependent Variable: D(LOG(CLOSE))  
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)  
 Date: 09/26/19 Time: 10:00  
 Sample (adjusted): 1/05/2000 4/01/2019  
 Included observations: 4839 after adjustments  
 Failure to improve likelihood (non-zero gradients) after 9 iterations  
 Coefficient covariance computed using outer product of gradients  
 MA Backcast: 1/03/2000 1/04/2000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000146	0.000151	0.967649	0.3333
AR(1)	-0.221758	0.194399	-1.140738	0.2540
MA(1)	0.144115	0.194183	0.742161	0.4580
MA(2)	-0.070394	0.019617	-3.588403	0.0003
R-squared	0.009036	Mean dependent var	0.000148	
Adjusted R-squared	0.008422	S.D. dependent var	0.012023	
S.E. of regression	0.011972	Akaike info criterion	-6.011593	
Sum squared resid	0.693045	Schwarz criterion	-6.006233	
Log likelihood	14549.05	Hannan-Quinn criter.	-6.009712	
F-statistic	14.69659	Durbin-Watson stat	2.000583	
Prob(F-statistic)	0.000000			
Inverted AR Roots	-.22			
Inverted MA Roots	.20	-.35		

**Figura 39** - Output do modelo ARMA(1,2) da série global dos retornos do índice SP500

Dependent Variable: D(LOG(CLOSE))  
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)  
 Date: 09/26/19 Time: 10:00  
 Sample (adjusted): 1/06/2000 4/01/2019  
 Included observations: 4838 after adjustments  
 Failure to improve likelihood (non-zero gradients) after 16 iterations  
 Coefficient covariance computed using outer product of gradients  
 MA Backcast: 1/05/2000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000147	0.000153	0.962018	0.3361
AR(1)	-0.288980	0.226898	-1.273616	0.2029
AR(2)	-0.074590	0.020000	-3.729553	0.0002
MA(1)	0.210911	0.227487	0.927136	0.3539

R-squared	0.009172	Mean dependent var	0.000148
Adjusted R-squared	0.008557	S.D. dependent var	0.012024
S.E. of regression	0.011973	Akaike info criterion	-6.011528
Sum squared resid	0.692947	Schwarz criterion	-6.006166
Log likelihood	14545.89	Hannan-Quinn criter.	-6.009645
F-statistic	14.91578	Durbin-Watson stat	1.999670
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.14+.23i	-.14-.23i
Inverted MA Roots	-.21	

Figura 40 - Output do modelo ARMA(2,1) da série global dos retornos do índice SP500

Dependent Variable: D(LOG(CLOSE))  
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)  
 Date: 09/26/19 Time: 10:01  
 Sample (adjusted): 1/06/2000 4/01/2019  
 Included observations: 4838 after adjustments  
 Failure to improve likelihood (non-zero gradients) after 18 iterations  
 Coefficient covariance computed using outer product of gradients  
 MA Backcast: 1/04/2000 1/05/2000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000147	0.000154	0.955538	0.3394
AR(1)	-0.291877	0.243962	-1.196407	0.2316
AR(2)	-0.121569	0.145370	-0.836275	0.4030
MA(1)	0.213960	0.244878	0.873743	0.3823
MA(2)	0.047646	0.154501	0.308390	0.7578

R-squared	0.009188	Mean dependent var	0.000148
Adjusted R-squared	0.008368	S.D. dependent var	0.012024
S.E. of regression	0.011974	Akaike info criterion	-6.011131
Sum squared resid	0.692936	Schwarz criterion	-6.004429
Log likelihood	14545.92	Hannan-Quinn criter.	-6.008778
F-statistic	11.20449	Durbin-Watson stat	1.999858
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.15-.32i	-.15+.32i
Inverted MA Roots	-.11+.19i	-.11-.19i

Figura 41 - Output do modelo ARMA(2,2) da série global dos retornos do índice SP500

## Anexo D – Testes de Correlação e Heterocedasticidade da Série Global dos Retornos

Heteroskedasticity Test: ARCH

F-statistic	513.0644	Prob. F(2,4833)	0.0000
Obs*R-squared	846.9448	Prob. Chi-Square(2)	0.0000

Test Equation:

Dependent Variable: RESID^2  
 Method: Least Squares  
 Date: 09/26/19 Time: 10:03  
 Sample (adjusted): 1/10/2000 4/01/2019  
 Included observations: 4836 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.14E-05	6.39E-06	11.16549	0.0000
RESID^2(-1)	0.127658	0.013341	9.568531	0.0000
RESID^2(-2)	0.373327	0.013341	27.98265	0.0000

R-squared	0.175133	Mean dependent var	0.000143
Adjusted R-squared	0.174792	S.D. dependent var	0.000453
S.E. of regression	0.000412	Akaike info criterion	-12.75167
Sum squared resid	0.000819	Schwarz criterion	-12.74765
Log likelihood	30836.54	Hannan-Quinn criter.	-12.75026
F-statistic	513.0644	Durbin-Watson stat	2.074971
Prob(F-statistic)	0.000000		

Figura 42 - Output do teste de Heterocedasticidade dos resíduos da série global dos retornos do índice SP500

Breusch-Godfrey Serial Correlation LM Test:

Null hypothesis: No serial correlation at up to 2 lags

F-statistic	0.712197	Prob. F(2,4833)	0.4906
Obs*R-squared	1.425447	Prob. Chi-Square(2)	0.4903

Test Equation:

Dependent Variable: RESID  
 Method: Least Squares  
 Date: 09/26/19 Time: 10:02  
 Sample: 1/06/2000 4/01/2019  
 Included observations: 4838  
 Coefficient covariance computed using outer product of gradients  
 Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.31E-07	0.000151	0.006149	0.9951
AR(1)	-0.408664	0.350094	-1.167298	0.2431
AR(2)	0.097889	0.175494	0.556650	0.5778
RESID(-1)	0.409610	0.350507	1.168621	0.2426
RESID(-2)	-0.129556	0.194758	-0.665216	0.5059

R-squared	0.000295	Mean dependent var	2.47E-19
Adjusted R-squared	-0.000533	S.D. dependent var	0.011970
S.E. of regression	0.011974	Akaike info criterion	-6.011203
Sum squared resid	0.692886	Schwarz criterion	-6.004501
Log likelihood	14546.10	Hannan-Quinn criter.	-6.008850
F-statistic	0.356098	Durbin-Watson stat	1.999940
Prob(F-statistic)	0.839930		

Figura 43 - Output do teste de correlação dos resíduos da série global dos retornos do índice SP500

## Anexo E – Previsão da Série Global dos Retornos

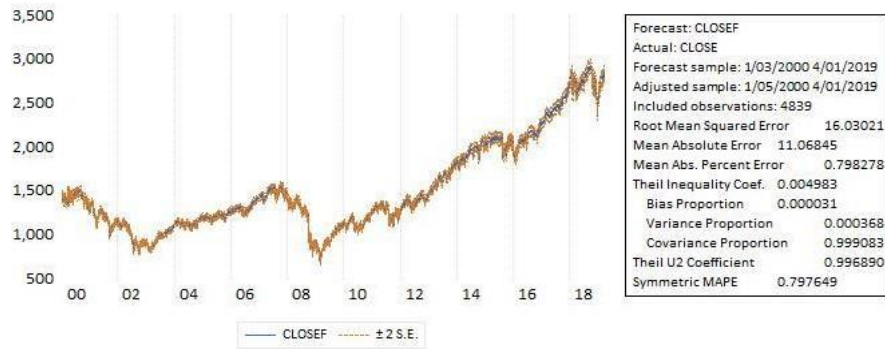


Figura 44 - Previsão *in-sample* com o modelo AR(1) da série global dos retornos do índice SP500

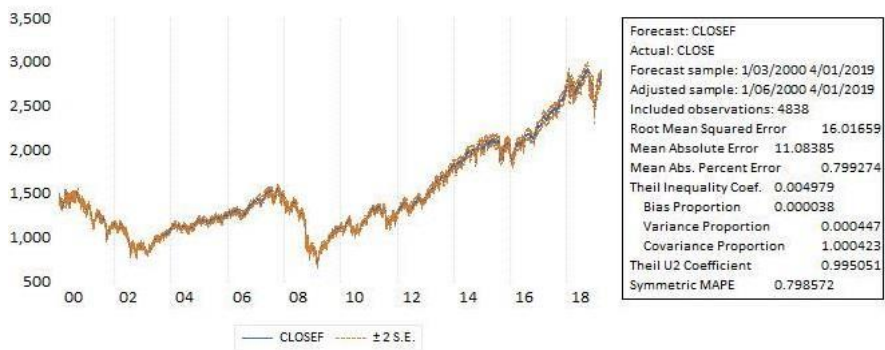


Figura 45 - Previsão *in-sample* com o modelo AR(2) da série global dos retornos do índice SP500

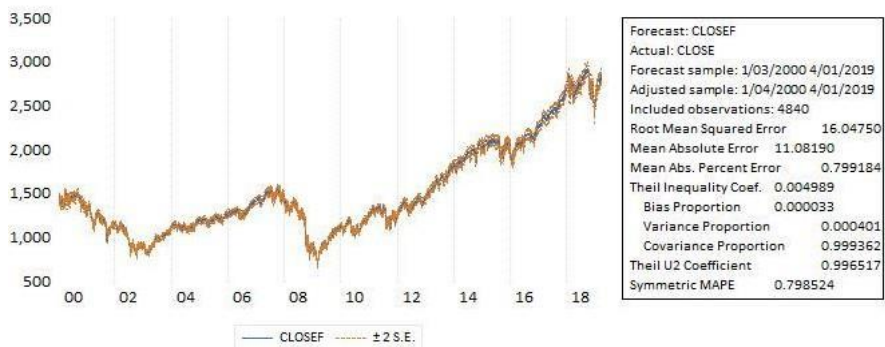
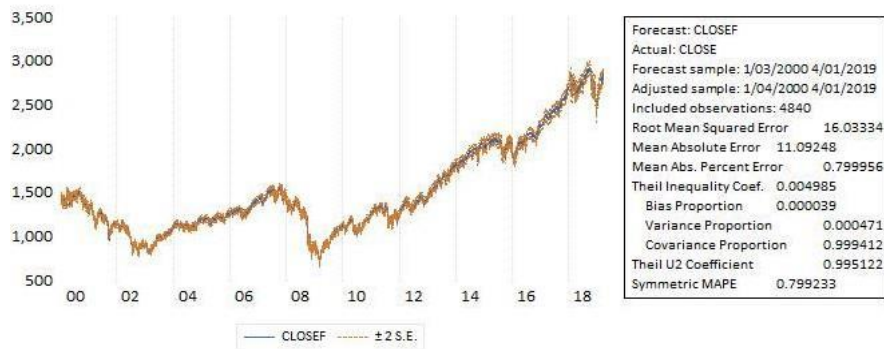
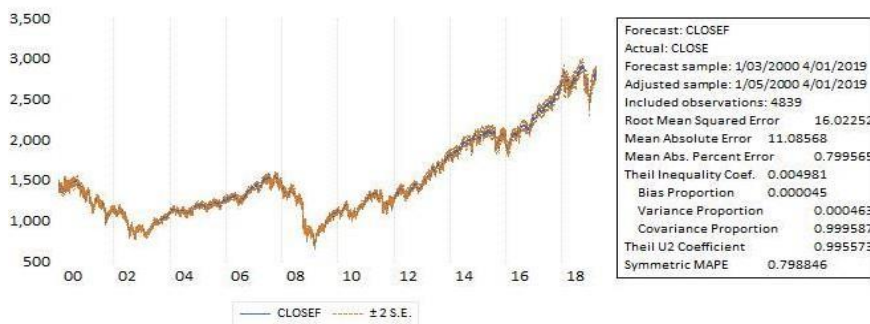


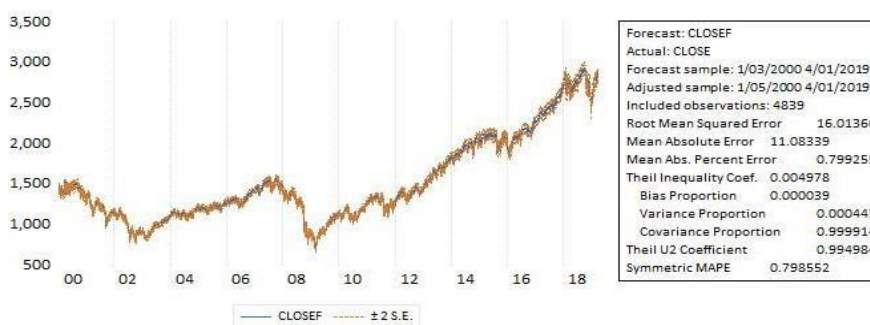
Figura 46 - Previsão *in-sample* com o modelo MA(1) da série global dos retornos do índice SP500



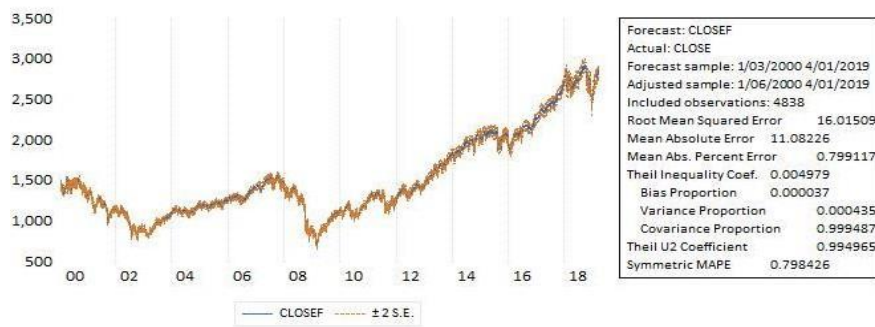
**Figura 47** - Previsão *in-sample* com o modelo MA(2) da série global dos retornos do índice SP500



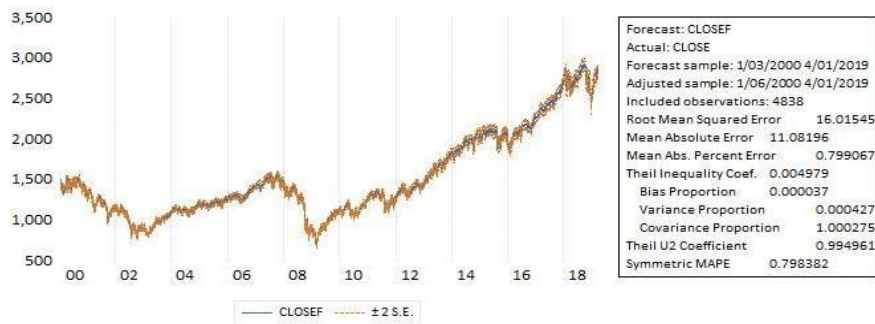
**Figura 48** - Previsão *in-sample* com o modelo ARMA(1,1) da série global dos retornos do índice SP500



**Figura 49** - Previsão *in-sample* com o modelo ARMA(1,2) da série global dos retornos do índice SP500



**Figura 50** - Previsão *in-sample* com o modelo ARMA(2,1) da série global dos retornos do índice SP500



**Figura 51** - Previsão *in-sample* com o modelo ARMA(2,2) da série global dos retornos do índice SP500

## Anexo F – Testes de Estacionariedade da Série Reduzida

Null Hypothesis: CLOSE has a unit root  
Exogenous: Constant  
Lag Length: 0 (Automatic - based on SIC, maxlag=26)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.823968	0.8117
Test critical values:		
1% level	-3.432733	
5% level	-2.862479	
10% level	-2.567315	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
Dependent Variable: D(CLOSE)  
Method: Least Squares  
Date: 09/26/19 Time: 10:15  
Sample (adjusted): 3/03/2009 4/01/2019  
Included observations: 2538 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CLOSE(-1)	-0.000478	0.000580	-0.823968	0.4100
C	1.715416	1.096887	1.563894	0.1180
R-squared	0.000268	Mean dependent var	0.853574	
Adjusted R-squared	-0.000127	S.D. dependent var	16.64050	
S.E. of regression	16.64155	Akaike info criterion	8.462470	
Sum squared resid	702322.9	Schwarz criterion	8.467071	
Log likelihood	-10736.87	Hannan-Quinn criter.	8.464139	
F-statistic	0.678923	Durbin-Watson stat	2.064194	
Prob(F-statistic)	0.410035			

Figura 52 - Output do teste de estacionariedade ADF da série reduzida do índice SP500

Null Hypothesis: CLOSE has a unit root  
Exogenous: Constant  
Bandwidth: 24 (Newey-West automatic) using Bartlett kernel

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-0.724800	0.8386
Test critical values:		
1% level	-3.432733	
5% level	-2.862479	
10% level	-2.567315	

\*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	276.7230
HAC corrected variance (Bartlett kernel)	196.6236

Phillips-Perron Test Equation  
Dependent Variable: D(CLOSE)  
Method: Least Squares  
Date: 09/26/19 Time: 10:15  
Sample (adjusted): 3/03/2009 4/01/2019  
Included observations: 2538 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CLOSE(-1)	-0.000478	0.000580	-0.823968	0.4100
C	1.715416	1.096887	1.563894	0.1180
R-squared	0.000268	Mean dependent var	0.853574	
Adjusted R-squared	-0.000127	S.D. dependent var	16.64050	
S.E. of regression	16.64155	Akaike info criterion	8.462470	
Sum squared resid	702322.9	Schwarz criterion	8.467071	
Log likelihood	-10736.87	Hannan-Quinn criter.	8.464139	
F-statistic	0.678923	Durbin-Watson stat	2.064194	
Prob(F-statistic)	0.410035			

Figura 53 - Output do teste de estacionariedade PP da série reduzida do índice SP500

Null Hypothesis: CLOSE is stationary  
Exogenous: Constant  
Bandwidth: 40 (Newey-West automatic) using Bartlett kernel

	LM-Stat
Kwiatkowski-Phillips-Schmidt-Shin test statistic	6.125075
Asymptotic critical values*:	
1% level	0.739000
5% level	0.463000
10% level	0.347000

\*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	324550.9
HAC corrected variance (Bartlett kernel)	13010693

KPSS Test Equation  
Dependent Variable: CLOSE  
Method: Least Squares  
Date: 09/26/19 Time: 10:16  
Sample: 3/02/2009 4/01/2019  
Included observations: 2539

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1803.428	11.30826	159.4789	0.0000
R-squared	0.000000	Mean dependent var	1803.428	
Adjusted R-squared	0.000000	S.D. dependent var	569.8059	
S.E. of regression	569.8059	Akaike info criterion	15.52886	
Sum squared resid	8.24E+08	Schwarz criterion	15.53116	
Log likelihood	-19712.89	Hannan-Quinn criter.	15.52970	
Durbin-Watson stat	0.000855			

Figura 54 - Output do teste de estacionariedade KPSS da série reduzida do índice SP500

## Anexo G – Testes de Estacionariedade da Série Reduzida dos Retornos

Null Hypothesis: DLCLOSE has a unit root  
Exogenous: Constant  
Lag Length: 0 (Automatic - based on SIC, maxlag=26)

	t-Statistic	Prob.*
<b>Augmented Dickey-Fuller test statistic</b>	<b>-53.55653</b>	<b>0.0001</b>
Test critical values:		
1% level	-3.432734	
5% level	-2.862479	
10% level	-2.567315	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
Dependent Variable: D(DLCLOSE)  
Method: Least Squares  
Date: 09/26/19 Time: 10:18  
Sample (adjusted): 3/04/2009 4/01/2019  
Included observations: 2537 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLCLOSE(-1)	-1.061823	0.019826	-53.55653	0.0000
C	0.000592	0.000200	2.963720	0.0031
R-squared	0.530842	Mean dependent var	7.07E-06	
Adjusted R-squared	0.530657	S.D. dependent var	0.014661	
S.E. of regression	0.010044	Akaike info criterion	-6.362825	
Sum squared resid	0.255753	Schwarz criterion	-6.358222	
Log likelihood	8073.244	Hannan-Quinn criter.	-6.361155	
F-statistic	2868.302	Durbin-Watson stat	1.994790	
Prob(F-statistic)	0.000000			

Figura 55 - Output do teste de estacionariedade ADF da série reduzida dos retornos do índice SP500

Null Hypothesis: DLCLOSE has a unit root  
Exogenous: Constant  
Bandwidth: 22 (Newey-West automatic) using Bartlett kernel

	Adj. t-Stat	Prob.*
<b>Phillips-Perron test statistic</b>	<b>-54.22944</b>	<b>0.0001</b>
Test critical values:		
1% level	-3.432734	
5% level	-2.862479	
10% level	-2.567315	

\*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.000101
HAC corrected variance (Bartlett kernel)	8.11E-05

Phillips-Perron Test Equation  
Dependent Variable: D(DLCLOSE)  
Method: Least Squares  
Date: 09/26/19 Time: 10:18  
Sample (adjusted): 3/04/2009 4/01/2019  
Included observations: 2537 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLCLOSE(-1)	-1.061823	0.019826	-53.55653	0.0000
C	0.000592	0.000200	2.963720	0.0031
R-squared	0.530842	Mean dependent var	7.07E-06	
Adjusted R-squared	0.530657	S.D. dependent var	0.014661	
S.E. of regression	0.010044	Akaike info criterion	-6.362825	
Sum squared resid	0.255753	Schwarz criterion	-6.358222	
Log likelihood	8073.244	Hannan-Quinn criter.	-6.361155	
F-statistic	2868.302	Durbin-Watson stat	1.994790	
Prob(F-statistic)	0.000000			

Figura 56 - Output do teste de estacionariedade PP da série reduzida dos retornos do índice SP500

Null Hypothesis: DLCLOSE is stationary  
Exogenous: Constant  
Bandwidth: 25 (Newey-West automatic) using Bartlett kernel

	LM-Stat.
<b>Kwiatkowski-Phillips-Schmidt-Shin test statistic</b>	<b>0.255026</b>
Asymptotic critical values*:	
1% level	0.739000
5% level	0.463000
10% level	0.347000

\*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	0.000101
HAC corrected variance (Bartlett kernel)	7.14E-05

KPSS Test Equation  
Dependent Variable: DLCLOSE  
Method: Least Squares  
Date: 09/26/19 Time: 10:19  
Sample (adjusted): 3/03/2009 4/01/2019  
Included observations: 2538 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000555	0.000200	2.779660	0.0055
R-squared	0.000000	Mean dependent var	0.000555	
Adjusted R-squared	0.000000	S.D. dependent var	0.010061	
S.E. of regression	0.010061	Akaike info criterion	-6.359990	
Sum squared resid	0.256783	Schwarz criterion	-6.357689	
Log likelihood	8071.827	Hannan-Quinn criter.	-6.359155	
Durbin-Watson stat	2.122933			

Figura 57 - Output do teste de estacionariedade KPSS da série reduzida dos retornos do índice SP500

## Anexo H – Modelização da Série Reduzida dos Retornos

Date: 09/26/19 Time: 10:19  
 Sample: 3/02/2009 4/01/2019  
 Included observations: 2538

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1			-0.062	-0.062	9.7030	0.002
2			0.016	0.012	10.349	0.006
3			-0.050	-0.048	16.713	0.001
4			0.012	0.006	17.100	0.002
5			-0.069	-0.067	29.242	0.000
6			0.014	0.004	29.776	0.000
7			0.010	0.013	30.035	0.000
8			-0.044	-0.050	35.025	0.000
9			-0.008	-0.012	35.172	0.000
10			0.027	0.023	36.993	0.000
11			0.006	0.006	37.099	0.000
12			-0.029	-0.028	39.194	0.000
13			0.033	0.026	41.962	0.000
14			-0.058	-0.055	50.421	0.000
15			-0.026	-0.032	52.161	0.000
16			0.029	0.029	54.299	0.000
17			0.061	0.055	63.819	0.000
18			-0.022	-0.012	65.089	0.000
19			-0.041	-0.048	69.356	0.000
20			0.023	0.017	70.675	0.000
21			0.013	0.022	71.085	0.000
22			-0.022	-0.021	72.326	0.000
23			0.037	0.028	75.820	0.000
24			0.003	0.007	75.849	0.000
25			-0.072	-0.064	89.187	0.000
26			0.006	-0.001	89.286	0.000
27			0.013	0.007	89.713	0.000
28			-0.024	-0.025	91.193	0.000
29			0.019	0.020	92.112	0.000
30			-0.001	-0.010	92.114	0.000
31			0.010	0.017	92.391	0.000
32			-0.043	-0.032	97.113	0.000
33			0.022	-0.004	98.326	0.000
34			-0.027	-0.027	100.18	0.000
35			0.004	0.010	100.22	0.000
36			0.009	0.013	100.44	0.000

Figura 58 - Correlograma da série reduzida dos retornos do índice SP500

Dependent Variable: D(LOG(CLOSE))  
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)

Date: 09/26/19 Time: 10:20  
 Sample (adjusted): 3/04/2009 4/01/2019  
 Included observations: 2537 after adjustments  
 Convergence achieved after 2 iterations  
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000557	0.000188	2.968159	0.0030
AR(1)	-0.061823	0.019826	-3.118267	0.0018
R-squared	0.003821	Mean dependent var		0.000558
Adjusted R-squared	0.003428	S.D. dependent var		0.010062
S.E. of regression	0.010044	Akaike info criterion		-6.362825
Sum squared resid	0.255753	Schwarz criterion		-6.358222
Log likelihood	8073.244	Hannan-Quinn criter.		-6.361155
F-statistic	9.723592	Durbin-Watson stat		1.994790
Prob(F-statistic)	0.001840			
Inverted AR Roots	-.06			

Figura 59 - Output do modelo AR(1) da série reduzida dos retornos do índice SP500

Dependent Variable: D(LOG(CLOSE))  
 Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)

Date: 09/26/19 Time: 10:20  
 Sample (adjusted): 3/03/2009 4/01/2019  
 Included observations: 2538 after adjustments  
 Failure to improve likelihood (non-zero gradients) after 3 iterations  
 Coefficient covariance computed using outer product of gradients  
 MA Backcast: 3/02/2009

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000555	0.000187	2.963431	0.0031
MA(1)	-0.060677	0.019826	-3.060427	0.0022
R-squared	0.003739	Mean dependent var		0.000555
Adjusted R-squared	0.003346	S.D. dependent var		0.010061
S.E. of regression	0.010044	Akaike info criterion		-6.362947
Sum squared resid	0.255823	Schwarz criterion		-6.358346
Log likelihood	8076.580	Hannan-Quinn criter.		-6.361278
F-statistic	9.516665	Durbin-Watson stat		2.000489
Prob(F-statistic)	0.002058			
Inverted MA Roots	.06			

Figura 60 - Output do modelo MA(1) da série reduzida dos retornos do índice SP500

Dependent Variable: D(LOG(CLOSE))  
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)  
Date: 09/26/19 Time: 10:21  
Sample (adjusted): 3/04/2009 4/01/2019  
Included observations: 2537 after adjustments  
Failure to improve likelihood (non-zero gradients) after 9 iterations  
Coefficient covariance computed using outer product of gradients  
MA Backcast: 3/03/2009

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000555	0.000194	2.857810	0.0043
AR(1)	-0.797954	0.069261	-11.52101	0.0000
MA(1)	0.759879	0.075389	10.07948	0.0000

R-squared	0.014048	Mean dependent var	0.000558
Adjusted R-squared	0.013270	S.D. dependent var	0.010062
S.E. of regression	0.009995	Akaike info criterion	-6.372356
Sum squared resid	0.253128	Schwarz criterion	-6.365452
Log likelihood	8086.334	Hannan-Quinn criter.	-6.369852
F-statistic	18.05289	Durbin-Watson stat	2.023820
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.80
Inverted MA Roots	-.76

**Figura 61** - Output do modelo ARMA(1,1) da série reduzida dos retornos do índice SP500

## Anexo I – Testes de Correlação e Heterocedasticidade da Série Reduzida dos Retornos

Heteroskedasticity Test: ARCH

F-statistic	89.88740	Prob. F(1,2534)	0.0000
Obs*R-squared	86.87661	Prob. Chi-Square(1)	0.0000

Test Equation:  
Dependent Variable: RESID^2  
Method: Least Squares  
Date: 09/26/19 Time: 10:22  
Sample (adjusted): 3/05/2009 4/01/2019  
Included observations: 2536 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.13E-05	5.32E-06	15.28396	0.0000
RESID^2(-1)	0.185082	0.019522	9.480896	0.0000

R-squared	0.034257	Mean dependent var	9.98E-05
Adjusted R-squared	0.033876	S.D. dependent var	0.000254
S.E. of regression	0.000249	Akaike info criterion	-13.75389
Sum squared resid	0.000158	Schwarz criterion	-13.74929
Log likelihood	17441.94	Hannan-Quinn criter.	-13.75222
F-statistic	89.88740	Durbin-Watson stat	2.092460
Prob(F-statistic)	0.000000		

**Figura 62** - Output do teste de Heterocedasticidade dos resíduos da série reduzida dos retornos do índice SP500

Breusch-Godfrey Serial Correlation LM Test:  
Null hypothesis: No serial correlation at up to 1 lag

F-statistic	0.720453	Prob. F(1,2533)	0.3961
Obs*R-squared	0.721386	Prob. Chi-Square(1)	0.3957

Test Equation:  
Dependent Variable: RESID  
Method: Least Squares  
Date: 09/26/19 Time: 10:21  
Sample: 3/04/2009 4/01/2019  
Included observations: 2537  
Coefficient covariance computed using outer product of gradients  
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.91E-07	0.000194	-0.002011	0.9984
AR(1)	-0.019879	0.073118	-0.271872	0.7857
MA(1)	0.031144	0.083849	0.371424	0.7104
RESID(-1)	-0.023383	0.027549	-0.848795	0.3961

R-squared	0.000284	Mean dependent var	-1.97E-06
Adjusted R-squared	-0.000900	S.D. dependent var	0.009991
S.E. of regression	0.009995	Akaike info criterion	-6.371852
Sum squared resid	0.253056	Schwarz criterion	-6.362647
Log likelihood	8086.695	Hannan-Quinn criter.	-6.368513
F-statistic	0.240118	Durbin-Watson stat	2.000440
Prob(F-statistic)	0.868397		

**Figura 63** - Output do teste de correlação dos resíduos da série reduzida dos retornos do índice SP500

## Anexo J – Previsão da Série Reduzida dos Retornos



Figura 64 - Previsão *in-sample* com o modelo AR(1) da série reduzida dos retornos do índice SP500



Figura 65 - Previsão *in-sample* com o modelo MA(1) da série reduzida dos retornos do índice SP500

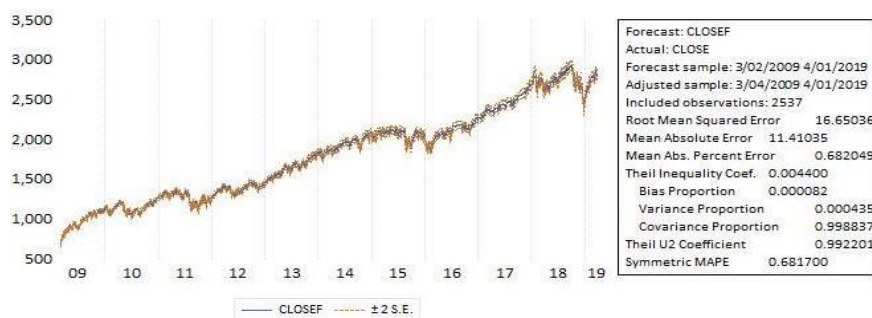


Figura 66 - Previsão *in-sample* com o modelo ARMA(1,1) da série reduzida dos retornos do índice SP500

## Anexo K – Código e Experiências

	Lote = 7, N = 256							
	Teste 80%		Teste 90%		Teste 95%		Teste 99%	
Epoch 1/20	0,0047	0,0014	0,003	0,003	0,0035	0,001	0,0036	0,0029
Epoch 2/20	0,0015	0,002	0,0014	0,0034	0,0014	7,40E-04	0,0015	5,27E-04
Epoch 3/20	0,0016	0,0029	0,0016	0,0049	0,0017	8,61E-04	0,0017	0,003
Epoch 4/20	0,0016	0,0038	0,0016	0,0041	0,002	0,0012	0,002	0,0037
Epoch 5/20	0,0018	0,0056	0,0019	0,005	0,0023	9,65E-04	0,0022	0,0018
Epoch 6/20	0,0018	0,0061	0,0019	0,0065	0,0029	0,0015	0,0029	0,0015
Epoch 7/20	0,002	0,0085	0,0021	0,0074	0,0045	0,0028	0,0039	0,0026
Epoch 8/20	0,002	0,0097	0,0023	0,0071	0,0061	0,0012	0,005	0,0013
Epoch 9/20	0,0024	0,0092	0,0028	0,0034	0,0065	0,0011	0,0067	7,67E-04
Epoch 10/20	0,0027	0,0086	0,0035	0,0023	0,008	0,0045	0,0087	5,09E-04
Epoch 11/20	0,003	0,0116	0,0037	0,0061	0,0087	0,0014	0,0095	0,0023
Epoch 12/20	0,0029	0,0131	0,0036	0,0049	0,008	0,0015	0,0099	0,0022
Epoch 13/20	0,0027	0,0138	0,0038	0,0102	0,0059	0,0016	0,0091	0,0016
Epoch 14/20	0,0025	0,0119	0,0035	0,0128	0,0038	9,52E-04	0,0082	0,0017
Epoch 15/20	0,0023	0,0118	0,0028	0,0108	0,0021	4,19E-04	0,0069	0,0017
Epoch 16/20	0,0021	0,0099	0,0018	0,0069	8,41E-04	2,20E-04	0,0051	0,0021
Epoch 17/20	0,002	0,0097	0,001	0,004	3,12E-04	2,18E-04	0,0037	0,0016
Epoch 18/20	0,0018	0,008	5,29E-04	0,0027	1,54E-04	2,07E-04	0,0025	0,0014
Epoch 19/20	0,0017	0,0088	2,81E-04	0,0019	1,03E-04	1,92E-04	0,0016	9,93E-04
Epoch 20/20	0,0015	0,0078	1,82E-04	0,0013	8,24E-05	1,81E-04	0,0011	7,54E-04
MSEAP		39777,85707		6666,33576		918,74398		3829,03287

**Tabela 20** - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 7 e dimensionalidade de 256 de espaço de saída

	Lote = 6, N = 256							
	Teste 80%		Teste 90%		Teste 95%		Teste 99%	
	0,0054	0,002	0,0037	0,0025	0,0045	8,32E-04	0,0039	0,0018
	0,0013	0,0017	0,0014	0,0023	0,0015	4,45E-04	0,0013	4,37E-04
	0,0014	0,0027	0,0016	0,0036	0,0018	8,42E-04	0,0016	0,0012
	0,0014	0,0023	0,0016	0,0057	0,002	8,74E-04	0,0017	5,72E-04
	0,0014	0,0035	0,0017	0,0065	0,0026	8,96E-04	0,0021	0,0017
	0,0014	0,0036	0,0019	0,0075	0,0036	0,0025	0,0026	4,59E-04
	0,0015	0,0054	0,0021	0,0073	0,0049	0,0029	0,0038	0,0031
	0,0014	0,0064	0,0025	0,0065	0,0064	0,0012	0,0048	0,0011
	0,0015	0,0072	0,0029	0,0049	0,0077	0,0061	0,0057	3,96E-04
	0,0016	0,008	0,0035	0,0036	0,0093	9,96E-04	0,0075	3,42E-04
	0,0016	0,0085	0,004	0,0061	0,0089	0,0019	0,0084	0,0034
	0,0017	0,0087	0,0042	0,009747	0,0074	0,0016	0,0088	0,0032
	0,0019	0,0097	0,0043	0,0197	0,0052	0,0015	0,0083	0,0026
	0,0022	0,0132	0,0036	0,0167	0,0032	7,69E-04	0,0073	0,0023
	0,0025	0,015926	0,0025	0,0125	0,0015	2,95E-04	0,0059	0,002
	0,0024	0,0165	0,0016	0,0085	5,67E-04	2,45E-04	0,0046	0,0018
	0,0023	0,015	8,72E-04	0,005804	2,29E-04	2,34E-04	0,0032	0,0015
	0,0022	0,0146	4,72E-04	0,0041	1,31E-04	2,19E-04	0,002	0,0014
	0,0021	0,0131	2,92E-04	0,003	9,69E-05	2,06E-04	0,0013	9,50E-04
	0,0022	0,0111	2,12E-04	0,0021	8,28E-05	1,96E-04	8,33E-04	7,62E-04
		56535,2787		10788,3434		996,83647		3871,7429

**Tabela 21** - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 6 e dimensionalidade de 256 de espaço de saída

Lote = 8, N = 256							
Teste 80%		Teste 90%		Teste 95%		Teste 99%	
0,0036	0,0013	0,0039	0,0032	0,0028	7,58E-04	0,0031	0,0035
0,0015	0,0019	0,0015	0,0033	0,0014	7,42E-04	0,0015	0,0015
0,0015	0,0033	0,0016	0,0034	0,0016	5,81E-04	0,0018	0,0043
0,0016	0,0045	0,0017	0,0055	0,0019	6,23E-04	0,002	0,003
0,0016	0,0054	0,0018	0,0059	0,0021	0,001104	0,0024	0,005
0,0019	0,0063	0,002	0,0063	0,0025	0,0043	0,003	0,0021
0,0019	0,0074	0,0023	0,0072	0,0032	0,0011	0,0041	7,74E-04
0,0021	0,0085	0,0028	0,0039	0,0046	0,0019	0,0055	0,0023
0,0022	0,0088	0,0034	0,0035	0,0057	9,69E-04	0,0056	9,57E-04
0,0023	0,0073	0,0036	0,0019	0,0063	0,0029	0,0071	0,0014
0,0023	0,0093	0,0038	0,0057	0,0076	7,99E-04	0,0085	0,0032
0,0024	0,0111	0,0037	0,0066	0,0075	0,0018	0,0089	0,0031
0,0023	0,0119	0,0038	0,0123	0,0059	0,0013	0,0086	0,0019
0,0021	0,0107	0,003	0,0103	0,0039	8,42E-04	0,0074	0,0024
0,002	0,0111	0,002	0,0068	0,0023	3,37E-04	0,0062	0,0016
0,0017	0,0099	0,0012	0,0035	9,99E-04	2,09E-04	0,0051	0,0014
0,0016	0,0085	6,73E-04	0,0023	3,75E-04	2,15E-04	0,0035	0,0012
0,0014	0,0091	3,53E-04	0,0014	1,72E-04	2,01E-04	0,0021	0,0012
0,0013	0,00793	2,09E-04	9,37E-04	1,10E-04	1,85E-04	0,0012	9,61E-04
0,0012	0,0077	1,49E-04	6,31E-04	8,75E-05	1,73E-04	7,70E-04	6,85E-04
	39184,284		3208,9279		<b>877,9914</b>		3480,3131

**Tabela 22** - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 6 e dimensionalidade de 256 de espaço de saída

Lote = 9, N = 256							
Teste 80%		Teste 90%		Teste 95%		Teste 99%	
0,0035	0,0012	0,003	0,0013	0,003	4,52E-04	0,0029	0,0036
0,0014	0,0018	0,0014	0,004	0,0015	9,32E-04	0,0016	0,0017
0,0015	0,0035	0,0015	0,0038	0,0017	6,22E-04	0,0019	0,0059
0,0016	0,0042	0,0016	0,0029	0,0021	0,0013	0,0022	0,0062
0,0016	0,005	0,0018	0,0037	0,0024	8,21E-04	0,0025	0,0039
0,0018	0,0061	0,0019	0,0063	0,0031	9,99E-04	0,0034	0,0057
0,002	0,0075	0,002	0,0064	0,0039	0,0017	0,0041	0,0022
0,0021	0,0077	0,0024	0,0073	0,0048	0,0015	0,005	0,003
0,0022	0,0079	0,0026	0,0075	0,0059	9,46E-04	0,0055	4,56E-04
0,0023	0,0082	0,0028	0,0068	0,0063	0,0041	0,0073	8,79E-04
0,0023	0,007	0,0029	0,0017	0,0075	0,0011	0,0079	0,0032
0,0023	0,0092	0,003	0,0041	0,007	0,0015	0,0078	0,0024
0,0021	0,0098	0,0029	0,0036	0,0053	9,51E-04	0,0069	0,0025
0,002	0,0091	0,0029	0,0042	0,0034	5,63E-04	0,006	0,0016
0,0017	0,0082	0,0027	0,0069	0,0019	2,35E-04	0,0048	0,0015
0,0015	0,0073	0,0019	0,0035	7,62E-04	2,54E-04	0,0033	0,0011
0,0013	0,0069	0,0013	0,0028	2,93E-04	2,34E-04	0,002	9,95E-04
0,0012	0,0063	8,01E-04	0,0016	1,49E-04	2,07E-04	0,0011	7,01E-04
0,0011	0,0056	4,52E-04	0,0011	1,05E-04	1,89E-04	5,52E-04	4,73E-04
9,31E-04	0,0059	2,48E-04	7,34E-04	8,83E-05	1,76E-04	3,18E-04	3,37E-04
	30117,956		3729,2525		894,82323		1710,6864

**Tabela 23** - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 9 e dimensionalidade de 256 de espaço de saída

Lote = 8, N = 255							
Teste 80%		Teste 90%		Teste 95%		Teste 99%	
0,0038	0,0013	0,003	0,0025	0,0025	4,43E-04	0,0029	0,0033
0,0015	0,0022	0,0013	0,0035	0,0014	8,95E-04	0,0015	0,0017
0,0016	0,0037	0,0015	0,0044	0,0016	7,24E-04	0,0018	0,0045
0,0017	0,0042	0,0016	0,0037	0,002	0,0011	0,002	0,0031
0,0018	0,0055	0,0018	0,0031	0,0022	7,69E-04	0,0025	0,0059
0,0019	0,0073	0,0019	0,006	0,0028	0,0014	0,0031	0,0048
0,002	0,0079	0,0021	0,0066	0,0036	0,0012	0,0041	0,0014
0,0022	0,0104	0,0025	0,0059	0,0047	0,0033	0,0051	0,0039
0,0023	0,0098	0,0028	0,0076	0,0062	0,0012	0,006	5,65E-04
0,0025	0,0078	0,0031	0,0021	0,0068	0,0035	0,0079	6,55E-04
0,0027	0,0101	0,0032	0,0039	0,0081	0,0011	0,0084	0,0034
0,0026	0,0131	0,0032	0,0039	0,0077	0,0018	0,0085	0,0025
0,0024	0,012	0,0032	0,0048	0,0063	0,0014	0,0076	0,0025
0,0022	0,0111	0,003	0,0054	0,0042	9,69E-04	0,0065	0,0017
0,0019	0,0104	0,0026	0,0075	0,0025	3,96E-04	0,0053	0,0016
0,0017	0,009	0,0018	0,0046	0,0011	2,13E-04	0,0038	0,0014
0,0016	0,0087	0,0011	0,0027	4,01E-04	2,19E-04	0,0024	0,0012
0,0015	0,0074	6,28E-04	0,0016	1,81E-04	2,06E-04	0,0014	8,96E-04
0,0014	0,0095	3,26E-04	0,001	1,15E-04	1,89E-04	8,32E-04	6,95E-04
0,0012	0,0073	1,87E-04	7,03E-04	8,94E-05	1,74E-04	5,08E-04	5,11E-04
	37280,895		3571,9396		883,66801		2595,4787

**Tabela 24** - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 8 e dimensionalidade de 255 de espaço de saída

Lote = 8, N = 254							
Teste 80%		Teste 90%		Teste 95%		Teste 99%	
0,0034	0,0013	0,0029	0,0023	0,003	4,43E-04	0,0035	0,0037
0,0016	0,0021	0,0013	0,0036	0,0016	8,77E-04	0,0014	0,0011
0,0017	0,0041	0,0015	0,0048	0,0018	8,96E-04	0,0017	0,0044
0,0018	0,0041	0,0016	0,0043	0,0021	7,37E-04	0,0019	0,0051
0,0018	0,0062	0,0018	0,0036	0,0025	9,44E-04	0,0023	0,0023
0,002	0,0069	0,0019	0,0038	0,003	0,0011	0,0028	0,0016
0,0022	0,0095	0,0021	0,0061	0,0042	0,0027	0,0036	0,0059
0,0023	0,0111	0,0023	0,0073	0,0062	0,0012	0,0045	0,0027
0,0025	0,0105	0,0027	0,007	0,0062	0,001	0,0053	0,0012
0,0026	0,0079	0,003	0,003	0,007	0,0045	0,0059	4,78E-04
0,0028	0,0101	0,0033	0,0027	0,0082	0,0016	0,0076	5,21E-04
0,0027	0,013	0,0033	0,0042	0,0078	0,0017	0,0078	0,003
0,0024	0,0116	0,0032	0,0051	0,0062	0,0015	0,0081	0,0029
0,0022	0,0099	0,0032	0,008	0,0041	0,0011	0,0069	0,0023
0,002	0,0086	0,0026	0,0074	0,0024	4,67E-04	0,0057	0,0017
0,0019	0,0074	0,0018	0,0044	0,0011	2,21E-04	0,0045	0,0015
0,0017	0,0077	0,0011	0,0027	3,88E-04	2,19E-04	0,003	0,0014
0,0015	0,0071	6,24E-04	0,0015	1,79E-04	2,08E-04	0,0018	0,0011
0,0014	0,0068	3,17E-04	0,0011	1,13E-04	1,92E-04	0,0011	8,10E-04
0,0012	0,0073	1,82E-04	7,38E-04	8,81E-05	1,78E-04	6,50E-04	5,77E-04
	37073,593		3749,3458		903,44542		2932,7926

**Tabela 25** - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 8 e dimensionalidade de 254 de espaço de saída

Lote = 8, N = 257							
Teste 80%		Teste 90%		Teste 95%		Teste 99%	
0,0038	0,0014	0,0039	0,0033	0,0031	5,98E-04	0,0032	0,0038
0,0016	0,0024	0,0014	0,0034	0,0015	9,16E-04	0,0015	0,0012
0,0016	0,0035	0,0016	0,0053	0,0018	6,51E-04	0,0018	0,0053
0,0017	0,0042	0,0017	0,0061	0,0022	6,95E-04	0,0021	0,0046
0,0018	0,0051	0,0018	0,0058	0,0024	8,33E-04	0,0025	0,0061
0,0019	0,007	0,002	0,0064	0,0032	0,0019	0,003	0,0044
0,0021	0,0091	0,0021	0,0074	0,0048	0,0027	0,0042	0,0043
0,0023	0,01	0,0024	0,0073	0,0065	0,0012	0,0052	0,0012
0,0024	0,0102	0,0029	0,0052	0,0068	9,93E-04	0,0056	0,001
0,0025	0,0079	0,0033	0,001918	0,0078	0,0051	0,0064	0,0011
0,0027	0,0092	0,0036	0,0053	0,0088	0,0013	0,008	0,0027
0,0026	0,0108	0,0035	0,0043	0,0077	0,0014	0,0083	0,0032
0,0024	0,0122	0,0036	0,0069	0,0057	0,0015	0,0082	0,0027
0,0021	0,0103	0,0033	0,0104	0,0038	8,54E-04	0,007	0,0022
0,002	0,0097	0,0025	0,0085	0,002	3,36E-04	0,0059	0,0017
0,0018	0,0096	0,0015	0,005	8,10E-04	2,16E-04	0,0044	0,0014
0,0016	0,0084	9,27E-04	0,003	3,04E-04	2,19E-04	0,003	0,0012
0,0015	0,008	5,03E-04	0,002	1,51E-04	2,02E-04	0,0018	0,001
0,0013	0,0074	2,79E-04	0,0013	1,02E-04	1,86E-04	0,001	7,53E-04
0,0013	0,0071	1,80E-04	8,73E-04	8,33E-05	1,74E-04	6,18E-04	5,52E-04
	36302,527		4438,5574		881,94155		2804,7882

**Tabela 26** - Resultados obtidos com o modelo DNN - LSTM, com 20 épocas, lote de 8 e dimensionalidade de 257 de espaço de saída

O código utilizado para a previsão do modelo DNN-LSTM:

```
import numpy as np
import pandas as pd
from keras.models import Sequential
from keras.layers import LSTM,Dense
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error
from math import sqrt

#Function for processing training and testing data
def getdata(data,ln):
    X,Y = [],[]
    for i in range( len(data)-ln-1 ):
        X.append(data[ i:(i+ln),0] )
```

```

        Y.append(data[ (i+ln),0] )

return np.array(X),np.array(Y)

#Function for loading input data

data = pd.read_csv('C:/Users/crist/OneDrive/Área de Trabalho/Tese/Documentos/
SeP500.csv', engine = 'python')

dt = data['Close']

dt.dropna(inplace=True)

#Function for scaling input data

scale = MinMaxScaler()

dt = dt.values.reshape(dt.shape[0],1)

dt = scale.fit_transform(dt)

dt

#Splitting input data into training and testing data

X,y = getdata(dt,8)

X_train,X_test = X[:int(X.shape[0]*0.90)],X[int(X.shape[0]*0.90):]

y_train,y_test = y[:int(y.shape[0]*0.90)],y[int(y.shape[0]*0.90):]

#Parameters of the model

model = Sequential()

model.add(LSTM(256,input_shape=(8,1)))

model.add(Dense(1))

model.compile(optimizer='adam',loss='mse')

#Training the model

X_train = X_train.reshape((X_train.shape[0],X_train.shape[1],1))

X_test = X_test.reshape((X_test.shape[0],X_test.shape[1],1))

history = model.fit(X_train,y_train,epochs = 250,validation_data=(X_test,y_test),
shuffle=False)

```

```

#Testing the model using test data
Xt = model.predict(X_test)

#Plotting actual data vs predicted data
plt.rcParams.update({'font.size': 10})
plt.plot(scale.inverse_transform(y_test.reshape(-1,1)), color='red', label='Valor Atual')
plt.plot(scale.inverse_transform(Xt), color='green', label='Valor Previsto')
plt.legend(bbox_to_anchor=(0.31, 1), loc=1, borderaxespad=0.)
plt.xlabel('Número de dias')
plt.ylabel('Preços de Fecho')
plt.title('Previsão de Preços de Fecho para a Bolsa SP500')
plt.show()

#Calculating the mean squared error
xin = scale.inverse_transform(Xt)
yin = scale.inverse_transform(y_test.reshape(-1,1))
final_mse = mean_squared_error(xin, yin)
print ("Mean squared error after predictions: %f"%(final_mse))
print('RMSE%',str(round((sqrt(mean_squared_error(yin, xin))/yin.mean()*100,2))+'%')

mape1 = np.mean(np.abs(xin - yin)/np.abs(yin))*100
print('MAPE1', mape1)

act = []
pred = []
i=483
Xt = model.predict(X_test[i].reshape(1,8,1))
print('predicted: {0}, actual: {1}'.format(scale.inverse_transform(Xt),
scale.inverse_transform(y_test[i].reshape(-1,1))))
pred.append(scale.inverse_transform(Xt))

```

```
act.append(scale.inverse_transform(y_test[i].reshape(-1,1)))
```

### Anexo L– Previsão da Série com DL - LSTM

Epoch 246/250 4350/4350 [=====] - 14s 3ms/step - loss: 1.2800e-04 - val_loss: 1.1103e-04
Epoch 247/250 4350/4350 [=====] - 12s 3ms/step - loss: 1.2233e-04 - val_loss: 1.1245e-04
Epoch 248/250 4350/4350 [=====] - 12s 3ms/step - loss: 1.1736e-04 - val_loss: 1.1340e-04
Epoch 249/250 4350/4350 [=====] - 13s 3ms/step - loss: 1.1267e-04 - val_loss: 1.1171e-04=>. .... ] - ETA: 5s - loss: 1.4214e-04
Epoch 250/250 4350/4350 [=====] - 14s 3ms/step - loss: 1.0753e-04 - val_loss: 1.1118e-04

Tabela 27 - Resultados obtidos na previsão da série com o modelo DNN-LSTM

Valores previstos com o modelo DNN-LSTM para a série no seu global:

Atual	Predicted	Atual	Predicted	Atual	Predicted
<b>2388.3301</b>	2389.8450	<b>2681.4700</b>	2684.4248	<b>2821.9299</b>	2826.0276
<b>2391.1699</b>	2393.1003	<b>2679.2500</b>	2678.5977	<b>2839.9600</b>	2814.1467
<b>2388.1299</b>	2395.9878	<b>2684.5701</b>	2675.8430	<b>2818.3701</b>	2827.8752
<b>2389.5200</b>	2393.5471	<b>2683.3401</b>	2680.0725	<b>2840.6899</b>	2811.6240
<b>2399.2900</b>	2394.5171	<b>2680.5000</b>	2679.5359	<b>2850.1299</b>	2828.2212
<b>2399.3799</b>	2403.3186	<b>2682.6201</b>	2677.0210	<b>2857.0500</b>	2838.4377
<b>2396.9199</b>	2404.2476	<b>2687.5400</b>	2678.5332	<b>2862.9600</b>	2845.5046
<b>2399.6299</b>	2402.0811	<b>2673.6101</b>	2682.9490	<b>2861.8201</b>	2851.3201
<b>2394.4399</b>	2404.2478	<b>2695.8101</b>	2671.5105	<b>2856.9800</b>	2851.0742
<b>2390.8999</b>	2399.8452	<b>2713.0601</b>	2689.1633	<b>2874.6899</b>	2846.9006
<b>2402.3201</b>	2396.2505	<b>2723.9900</b>	2706.1443	<b>2896.7400</b>	2861.1794
<b>2400.6699</b>	2406.1001	<b>2743.1499</b>	2717.4443	<b>2897.5200</b>	2881.5269
<b>2357.0300</b>	2405.6318	<b>2747.7100</b>	2735.0332	<b>2914.0400</b>	2884.8394
<b>2365.7200</b>	2366.6218	<b>2751.2900</b>	2740.9600	<b>2901.1299</b>	2898.9070

<b>2381.7300</b>	2370.5664	<b>2748.2300</b>	2744.5691	<b>2901.5200</b>	2890.0017
<b>2394.0200</b>	2385.5583	<b>2767.5601</b>	2742.3328	<b>2896.7200</b>	2888.9941
<b>2398.4199</b>	2398.0176	<b>2786.2400</b>	2758.4292	<b>2888.6001</b>	2884.8823
<b>2404.3899</b>	2403.0522	<b>2776.4199</b>	2776.2986	<b>2878.0500</b>	2877.5498
<b>2415.0701</b>	2408.7078	<b>2802.5601</b>	2770.1284	<b>2871.6799</b>	2867.7573
<b>2415.8201</b>	2418.6382	<b>2798.0300</b>	2791.3181	<b>2877.1299</b>	2861.1616
<b>2412.9099</b>	2420.2195	<b>2810.3000</b>	2790.1899	<b>2887.8899</b>	2864.8970
<b>2411.8000</b>	2417.6958	<b>2832.9700</b>	2800.2590	<b>2888.9199</b>	2874.4709
<b>2430.0601</b>	2416.4163	<b>2839.1299</b>	2820.6501	<b>2904.1799</b>	2876.6243
<b>2439.0701</b>	2432.4988	<b>2837.5400</b>	2828.4011	<b>2904.9800</b>	2889.5874
<b>2436.1001</b>	2442.0962	<b>2839.2500</b>	2827.9128	<b>2888.8000</b>	2891.9863
<b>2429.3301</b>	2440.3110	<b>2872.8701</b>	2829.2014	<b>2904.3101</b>	2878.7283
<b>2433.1399</b>	2434.0134	<b>2853.5300</b>	2857.5300	<b>2907.9500</b>	2889.8159
<b>2433.7900</b>	2436.7241	<b>2822.4299</b>	2845.0657	<b>2930.7500</b>	2894.4260
<b>2431.7700</b>	2437.6167	<b>2823.8101</b>	2817.0652	<b>2929.6699</b>	2913.9580
<b>2429.3899</b>	2435.9216	<b>2821.9800</b>	2814.5310	<b>2919.3701</b>	2915.7190
<b>2440.3501</b>	2433.6348	<b>2762.1299</b>	2812.8416	<b>2915.5601</b>	2907.2703
<b>2437.9199</b>	2443.1194	<b>2648.9399</b>	2762.0610	<b>2905.9700</b>	2902.8772
<b>2432.4600</b>	2441.9424	<b>2695.1399</b>	2658.7266	<b>2914.0000</b>	2894.3296
<b>2433.1499</b>	2436.9224	<b>2681.6599</b>	2685.7766	<b>2913.9800</b>	2899.8618
<b>2453.4600</b>	2437.0127	<b>2581.0000</b>	2678.5383	<b>2924.5901</b>	2900.6807
<b>2437.0300</b>	2455.0295	<b>2619.5500</b>	2590.8105	<b>2923.4299</b>	2909.5896
<b>2435.6101</b>	2442.3291	<b>2656.0000</b>	2613.9045	<b>2925.5100</b>	2909.8486
<b>2434.5000</b>	2439.6470	<b>2662.9399</b>	2648.8181	<b>2901.6101</b>	2911.5593
<b>2438.3000</b>	2438.4446	<b>2698.6299</b>	2658.8938	<b>2885.5701</b>	2891.8906
<b>2439.0701</b>	2441.7227	<b>2731.2000</b>	2690.6057	<b>2884.4299</b>	2875.7842
<b>2419.3799</b>	2442.7664	<b>2732.2200</b>	2722.0791	<b>2880.3401</b>	2872.7380
<b>2440.6899</b>	2425.4128	<b>2716.2600</b>	2726.4529	<b>2785.6799</b>	2869.0476
<b>2419.7000</b>	2442.5203	<b>2701.3301</b>	2713.1035	<b>2728.3701</b>	2788.9890
<b>2423.4099</b>	2425.7791	<b>2703.9600</b>	2698.5325	<b>2767.1299</b>	2729.6599
<b>2429.0100</b>	2427.2886	<b>2747.3000</b>	2699.0320	<b>2750.7900</b>	2755.5681
<b>2432.5400</b>	2432.5193	<b>2779.6001</b>	2736.2515	<b>2809.9199</b>	2745.6416
<b>2409.7500</b>	2436.1885	<b>2744.2800</b>	2768.3349	<b>2809.2100</b>	2794.5234
<b>2425.1799</b>	2416.3003	<b>2713.8301</b>	2742.0654	<b>2768.7800</b>	2800.2585
<b>2427.4299</b>	2427.9734	<b>2677.6699</b>	2712.4629	<b>2767.7800</b>	2766.4084
<b>2425.5300</b>	2431.2649	<b>2691.2500</b>	2677.8604	<b>2755.8799</b>	2761.0933
<b>2443.2500</b>	2429.8711	<b>2720.9399</b>	2685.5117	<b>2740.6899</b>	2750.5220
<b>2447.8301</b>	2445.3958	<b>2728.1201</b>	2712.2422	<b>2656.1001</b>	2736.3154
<b>2459.2700</b>	2451.0010	<b>2726.8000</b>	2721.6917	<b>2705.5701</b>	2662.2278
<b>2459.1399</b>	2461.5459	<b>2738.9700</b>	2721.5347	<b>2658.6899</b>	2695.6309
<b>2460.6101</b>	2462.4343	<b>2786.5701</b>	2731.7820	<b>2641.2500</b>	2660.0935
<b>2473.8301</b>	2463.7065	<b>2783.0200</b>	2773.4678	<b>2682.6299</b>	2640.6892

<b>2473.4500</b>	2475.4592	<b>2765.3101</b>	2775.6187	<b>2711.7400</b>	2674.2944
<b>2472.5400</b>	2476.3208	<b>2749.4800</b>	2760.5588	<b>2740.3701</b>	2703.4846
<b>2469.9099</b>	2475.5073	<b>2747.3301</b>	2745.0574	<b>2723.0601</b>	2731.2676
<b>2477.1299</b>	2473.0686	<b>2752.0100</b>	2741.3201	<b>2738.3101</b>	2719.6941
<b>2477.8301</b>	2479.1694	<b>2712.9199</b>	2744.9714	<b>2755.4500</b>	2730.9548
<b>2475.4199</b>	2480.4407	<b>2716.9399</b>	2712.1697	<b>2813.8899</b>	2746.9519
<b>2472.1001</b>	2478.4131	<b>2711.9299</b>	2711.4832	<b>2806.8301</b>	2798.4136
<b>2470.3000</b>	2475.2478	<b>2643.6899</b>	2707.3652	<b>2781.0100</b>	2798.8323
<b>2476.3501</b>	2473.3374	<b>2588.2600</b>	2648.3167	<b>2726.2200</b>	2776.6538
<b>2477.5701</b>	2478.5002	<b>2658.5500</b>	2593.2285	<b>2722.1799</b>	2727.0444
<b>2472.1599</b>	2480.1421	<b>2612.6201</b>	2648.0374	<b>2701.5801</b>	2717.4016
<b>2476.8301</b>	2475.5100	<b>2605.0000</b>	2615.3518	<b>2730.2000</b>	2698.9958
<b>2480.9099</b>	2479.1191	<b>2640.8701</b>	2604.7073	<b>2736.2700</b>	2721.4058
<b>2474.9199</b>	2483.1167	<b>2581.8799</b>	2634.7202	<b>2690.7300</b>	2729.5479
<b>2474.0200</b>	2478.2363	<b>2614.4500</b>	2587.1685	<b>2641.8899</b>	2691.5000
<b>2438.2100</b>	2476.8999	<b>2644.6899</b>	2609.7634	<b>2649.9299</b>	2644.6887
<b>2441.3201</b>	2445.1577	<b>2662.8401</b>	2638.8799	<b>2632.5601</b>	2646.2368
<b>2465.8401</b>	2444.6377	<b>2604.4700</b>	2657.7712	<b>2673.4500</b>	2631.8862
<b>2464.6101</b>	2466.5369	<b>2613.1599</b>	2609.3257	<b>2682.1699</b>	2665.5999
<b>2468.1101</b>	2467.7808	<b>2656.8701</b>	2610.9797	<b>2743.7900</b>	2677.2458
<b>2430.0100</b>	2470.8564	<b>2642.1899</b>	2649.2124	<b>2737.8000</b>	2731.1453
<b>2425.5500</b>	2437.4333	<b>2663.9900</b>	2641.1108	<b>2760.1699</b>	2732.4346
<b>2428.3701</b>	2429.9722	<b>2656.3000</b>	2658.7151	<b>2790.3701</b>	2751.2488
<b>2452.5100</b>	2431.9978	<b>2677.8401</b>	2654.1453	<b>2700.0601</b>	2779.0730
<b>2444.0400</b>	2453.7324	<b>2706.3899</b>	2671.9631	<b>2695.9500</b>	2705.4729
<b>2438.9700</b>	2448.4697	<b>2708.6399</b>	2698.6072	<b>2633.0801</b>	2692.4233
<b>2443.0500</b>	2443.2837	<b>2693.1299</b>	2703.6101	<b>2637.7200</b>	2637.2419
<b>2444.2400</b>	2446.3323	<b>2670.1399</b>	2690.6702	<b>2636.7800</b>	2634.7705
<b>2446.3000</b>	2447.7343	<b>2670.2900</b>	2669.2468	<b>2651.0701</b>	2634.2583
<b>2457.5901</b>	2449.6926	<b>2634.5601</b>	2666.8242	<b>2650.5400</b>	2646.7197
<b>2471.6499</b>	2459.8711	<b>2639.3999</b>	2635.9309	<b>2599.9500</b>	2647.7603
<b>2476.5500</b>	2473.2896	<b>2666.9399</b>	2636.5095	<b>2545.9399</b>	2604.0779
<b>2457.8501</b>	2478.9026	<b>2669.9099</b>	2660.5999	<b>2546.1599</b>	2551.9045
<b>2465.5400</b>	2462.8535	<b>2648.0500</b>	2666.0750	<b>2506.9600</b>	2546.5076
<b>2465.1001</b>	2467.8867	<b>2654.8000</b>	2647.7361	<b>2467.4199</b>	2512.1460
<b>2461.4299</b>	2468.1182	<b>2635.6699</b>	2651.2937	<b>2416.6201</b>	2473.8335
<b>2488.1101</b>	2464.9009	<b>2629.7300</b>	2635.2952	<b>2351.1001</b>	2425.1953
<b>2496.4800</b>	2488.0969	<b>2663.4199</b>	2628.2935	<b>2467.7000</b>	2362.2634
<b>2498.3701</b>	2497.8872	<b>2672.6299</b>	2656.6897	<b>2488.8301</b>	2460.1348
<b>2495.6201</b>	2500.3945	<b>2671.9199</b>	2668.0430	<b>2485.7400</b>	2489.1545
<b>2500.2300</b>	2498.1177	<b>2697.7900</b>	2668.5886	<b>2506.8501</b>	2488.8760
<b>2503.8701</b>	2501.8572	<b>2723.0701</b>	2690.7473	<b>2510.0300</b>	2507.0203

<b>2506.6499</b>	2505.4475	<b>2727.7200</b>	2714.9829	<b>2447.8899</b>	2511.5374
<b>2508.2400</b>	2508.2551	<b>2730.1299</b>	2721.7275	<b>2531.9399</b>	2457.0842
<b>2500.6001</b>	2509.9109	<b>2711.4500</b>	2724.4043	<b>2549.6899</b>	2525.2832
<b>2502.2200</b>	2503.3413	<b>2722.4600</b>	2708.6316	<b>2574.4099</b>	2548.3564
<b>2496.6599</b>	2504.0371	<b>2720.1299</b>	2716.0303	<b>2584.9600</b>	2572.0425
<b>2496.8401</b>	2499.2720	<b>2712.9700</b>	2715.0593	<b>2596.6399</b>	2583.5637
<b>2507.0400</b>	2498.9333	<b>2733.0100</b>	2708.7900	<b>2596.2600</b>	2594.7061
<b>2510.0601</b>	2507.9036	<b>2724.4399</b>	2725.1379	<b>2582.6101</b>	2595.4238
<b>2519.3601</b>	2511.5229	<b>2733.2900</b>	2719.8840	<b>2610.3000</b>	2583.5547
<b>2529.1201</b>	2520.0000	<b>2727.7600</b>	2726.6753	<b>2616.1001</b>	2606.1877
<b>2534.5801</b>	2529.4143	<b>2721.3301</b>	2722.8184	<b>2635.9600</b>	2613.8660
<b>2537.7400</b>	2535.1208	<b>2689.8601</b>	2716.7869	<b>2670.7100</b>	2631.7832
<b>2552.0701</b>	2538.4072	<b>2724.0100</b>	2689.1091	<b>2632.8999</b>	2663.7319
<b>2549.3301</b>	2551.2190	<b>2705.2700</b>	2714.9958	<b>2638.7000</b>	2634.6833
<b>2544.7300</b>	2550.1726	<b>2734.6201</b>	2702.3574	<b>2642.3301</b>	2635.9731
<b>2550.6399</b>	2545.9260	<b>2746.8701</b>	2725.7754	<b>2664.7600</b>	2639.4072
<b>2555.2400</b>	2550.5955	<b>2748.8000</b>	2739.1660	<b>2643.8501</b>	2659.1655
<b>2550.9299</b>	2555.1467	<b>2772.3501</b>	2742.3135	<b>2640.0000</b>	2643.4661
<b>2553.1699</b>	2551.8579	<b>2770.3701</b>	2762.5854	<b>2681.0500</b>	2638.1348
<b>2557.6399</b>	2553.4128	<b>2779.0300</b>	2763.4072	<b>2704.1001</b>	2672.9207
<b>2559.3601</b>	2557.4917	<b>2782.0000</b>	2770.6707	<b>2706.5300</b>	2696.8950
<b>2561.2600</b>	2559.4329	<b>2786.8501</b>	2774.0774	<b>2724.8701</b>	2701.6211
<b>2562.1001</b>	2561.2764	<b>2775.6299</b>	2778.5376	<b>2737.7000</b>	2717.6531
<b>2575.2100</b>	2562.1870	<b>2782.4900</b>	2769.5469	<b>2731.6101</b>	2730.4275
<b>2564.9800</b>	2573.6865	<b>2779.6599</b>	2774.1750	<b>2706.0500</b>	2726.6567
<b>2569.1299</b>	2566.0415	<b>2773.7500</b>	2772.4053	<b>2707.8799</b>	2704.2153
<b>2557.1499</b>	2568.7122	<b>2762.5901</b>	2767.1516	<b>2709.8000</b>	2702.9802
<b>2560.3999</b>	2558.5930	<b>2767.3201</b>	2757.0120	<b>2744.7300</b>	2704.6257
<b>2581.0701</b>	2560.2861	<b>2749.7600</b>	2759.7788	<b>2753.0300</b>	2734.7236
<b>2572.8301</b>	2578.5630	<b>2754.8799</b>	2745.2798	<b>2745.7300</b>	2745.5193
<b>2575.2600</b>	2573.4429	<b>2717.0701</b>	2747.8154	<b>2775.6001</b>	2740.4314
<b>2579.3601</b>	2574.8423	<b>2723.0601</b>	2716.0037	<b>2779.7600</b>	2765.0452
<b>2579.8501</b>	2578.5696	<b>2699.6299</b>	2717.1448	<b>2784.7000</b>	2771.6650
<b>2587.8401</b>	2579.3892	<b>2716.3101</b>	2697.4963	<b>2774.8799</b>	2776.5142
<b>2591.1299</b>	2586.4065	<b>2718.3701</b>	2709.4160	<b>2792.6699</b>	2768.7214
<b>2590.6399</b>	2590.0481	<b>2726.7100</b>	2712.8162	<b>2796.1101</b>	2782.7183
<b>2594.3799</b>	2589.9631	<b>2713.2200</b>	2720.3330	<b>2793.8999</b>	2787.4470
<b>2584.6201</b>	2593.1611	<b>2736.6101</b>	2709.6819	<b>2792.3799</b>	2786.1108
<b>2582.3000</b>	2585.0259	<b>2759.8201</b>	2728.2969	<b>2784.4900</b>	2784.5852
<b>2584.8401</b>	2582.0537	<b>2784.1699</b>	2750.4089	<b>2803.6899</b>	2777.6953
<b>2578.8701</b>	2583.9854	<b>2793.8401</b>	2773.7400	<b>2792.8101</b>	2793.0554
<b>2564.6201</b>	2579.0481	<b>2774.0200</b>	2784.6865	<b>2789.6499</b>	2785.8938

<b>2585.6399</b>	2566.0789	<b>2798.2900</b>	2769.0637	<b>2771.4500</b>	2782.1968
<b>2578.8501</b>	2582.9644	<b>2801.3101</b>	2787.4707	<b>2748.9299</b>	2766.3169
<b>2582.1399</b>	2579.0596	<b>2798.4299</b>	2792.4521	<b>2743.0701</b>	2745.1670
<b>2599.0300</b>	2581.3911	<b>2809.5500</b>	2790.5427	<b>2783.3000</b>	2737.6125
<b>2597.0801</b>	2596.3396	<b>2815.6201</b>	2799.6350	<b>2791.5200</b>	2771.0493
<b>2602.4199</b>	2596.3086	<b>2804.4900</b>	2805.9172	<b>2810.9199</b>	2782.3472
<b>2601.4199</b>	2600.8188	<b>2801.8301</b>	2797.2563	<b>2808.4800</b>	2800.0093
<b>2627.0400</b>	2600.4409	<b>2806.9800</b>	2793.8403	<b>2822.4800</b>	2800.0835
<b>2626.0701</b>	2622.5383	<b>2820.3999</b>	2797.7754	<b>2832.9399</b>	2811.7361
<b>2647.5801</b>	2624.2273	<b>2846.0701</b>	2809.6431	<b>2832.5701</b>	2822.0144
<b>2642.2200</b>	2642.8486	<b>2837.4399</b>	2832.7876	<b>2824.2300</b>	2822.9634
<b>2639.4399</b>	2640.3188	<b>2818.8201</b>	2828.4795	<b>2854.8799</b>	2815.9517
<b>2629.5701</b>	2637.4573	<b>2802.6001</b>	2812.0466	<b>2800.7100</b>	2840.7712
<b>2629.2700</b>	2628.5581	<b>2816.2900</b>	2796.1675	<b>2798.3601</b>	2798.4363
<b>2636.9800</b>	2627.2761	<b>2813.3601</b>	2805.7466	<b>2818.4600</b>	2790.7292
<b>2651.5000</b>	2633.8545	<b>2827.2200</b>	2804.6677	<b>2805.3701</b>	2806.9502
<b>2659.9900</b>	2647.2014	<b>2840.3501</b>	2816.1990	<b>2815.4399</b>	2798.1255
<b>2664.1101</b>	2656.0413	<b>2850.3999</b>	2828.7515	<b>2834.3999</b>	2805.4163
<b>2662.8501</b>	2660.5220	<b>2858.4500</b>	2838.7656	<b>2867.1899</b>	2822.4050
<b>2652.0100</b>	2659.8596	<b>2857.7000</b>	2846.7427	<b>2867.2400</b>	2852.0828
<b>2675.8101</b>	2650.3713	<b>2853.5801</b>	2847.0845		
<b>2690.1599</b>	2669.7239	<b>2833.2800</b>	2843.5789		