

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE AUTOMAÇÃO E SISTEMAS
ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Matheus Kjellin

Modelagem e identificação de séries
temporais aplicados ao mercado
financeiro

Florianópolis
2020

Matheus Kjellin

Modelagem e identificação de séries temporais aplicados ao mercado financeiro

Relatório submetido à Universidade Federal de Santa Catarina como requisito para a aprovação na disciplina DAS 5511: Projeto de Fim de Curso do curso de Graduação em Engenharia de Controle e Automação.
Orientador: Prof. Eduardo Camponogara.

Florianópolis
2020

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Kjellin, Matheus
Modelagem e identificação de sériestemporais aplicados
ao mercadofinanceiro / Matheus Kjellin ; orientador,
Eduardo Camponogara, 2020.
60 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Engenharia de Controle e Automação,
Florianópolis, 2020.

Inclui referências.

1. Engenharia de Controle e Automação. 2. Modelagem. 3.
Séries temporais. 4. Mercado financeiro. I. Camponogara,
Eduardo. II. Universidade Federal de Santa Catarina.
Graduação em Engenharia de Controle e Automação. III. Título.

Matheus Kjellin

Modelagem e identificação de séries temporais aplicados ao mercado financeiro

Esta monografia foi julgada no contexto da disciplina DAS5511: Projeto de Fim de Curso e aprovada na sua forma final pelo Curso de Engenharia de Controle e Automação.

Florianópolis, 02 de março de 2020

Banca Examinadora:

Prof. Eduardo Camponogara

Orientador no Curso

Universidade Federal de Santa Catarina

Prof. Eric Aislan Antonello

Avaliador

Universidade Federal de Santa Catarina

José Augusto de Freitas Caetano

Debatedor

Universidade Federal de Santa Catarina

Thuany Karoline Stuart

Debatedor

Universidade Federal de Santa Catarina

Agradecimentos

Primeiramente a minha tia Carla, pelo amor, paciência, incentivo e todo o apoio que precisei durante a graduação. Aos grandes amigos que fiz nessa jornada da vida. Ao professor Eduardo Camponogara pela orientação e a todos os professores que direta ou indiretamente fizeram parte da minha graduação. E enfim a Universidade Federal de Santa Catarina pela oportunidade de realizar esse fantástico curso que proporcionou grande crescimento pessoal e profissional.

*“O investimento mais seguro, melhor e rentável
é o que você faz em você!”*

Resumo

Este relatório tem como tema central o emprego de conceitos de identificação de sistemas para gerar modelos para séries temporais e tem como objetivo geral a síntese de previsões em ativos do mercado financeiro. A metodologia utilizada vem da literatura técnica, sendo este um assunto relacionado com temas abordados na graduação de Eng. Controle e Automação e o mercado financeiro, trata-se de uma pesquisa explicativa. Os dados utilizados foram retirados diretamente do mercado financeiro, o que permite a reprodução dos resultados. Serão abordados modelos auto-regressivos, ARMA e ARIMA. Como resultado será apresentado a comparação de assertividade entre os modelos.

Palavras-chave: Modelagem. Séries Temporais. Mercado Financeiro.

Abstract

This report has as its central theme the use of systems identification concepts to generate models for time series and its general objective is the synthesis of forecasts in financial market assets. The methodology used comes from the technical literature, being a subject area connected with topics seen in the undergraduate degree of Control and Automation and the financial market. The data used were collected directly from the financial market, which allows the reproduction of results. Auto-regressive models, ARMA and ARIMA will be covered. As a result, the assertiveness comparison between the models will be presented.

Keywords: Modeling. Time Series. Financial Market.

Lista de ilustrações

Figura 1 – Ação PETR4 da empresa Petrobras nos últimos 10 anos. Fonte: Bolsa PT [1]	21
Figura 2 – Gráfico dos dados da Tabela 3. Fonte: Autor	25
Figura 3 – Componentes de um sistema de controle. Fonte: Coelho [2]	27
Figura 4 – Parametrização de modelos. Fonte: Dr. Oubbati, Kybernetik [3]	28
Figura 5 – Exemplo de séries temporais. Fonte: Resumo Wooldridge [4]	29
Figura 6 – Tendências de uma série temporal. Fonte: PUC-Rio [5]	30
Figura 7 – Sazonalidade de uma série temporal. Fonte: Portal Action [6]	30
Figura 8 – Pontos influentes de uma série temporal. Fonte: ReNom [7]	31
Figura 9 – Processo estocástico. Fonte: Morettin [8]	33
Figura 10 – Série temporal estacionária. Fonte: Abg Consultoria [9]	34
Figura 11 – Série não-estacionária homogênea. Fonte: Estatística UFMG [10]	35
Figura 12 – Série estacionária gerada da primeira diferença. Fonte: Estatística UFMG [10]	35
Figura 13 – Ruído branco. Fonte: Wikipédia [11]	36
Figura 14 – Diagrama de etapas iterativas na construção de modelos de previsão. Fonte: Autor	39
Figura 15 – Ilustração do <i>Telefone sem fio</i> . Fonte: StackExchange [12]	40
Figura 16 – Gráfico da PACF. Fonte: StackExchange [13]	41
Figura 17 – Gráfico da ACF. Fonte: Pennsylvania State University [14]	42
Figura 18 – Ativo financeiro EURUSD. Fonte: Autor	45
Figura 19 – Gráfico de atraso. Fonte: Autor	46
Figura 20 – Gráfico de autocorrelação parcial. Fonte: Autor	47
Figura 21 – Previsão do modelo AR(2). Fonte: Autor	48
Figura 22 – Previsão do modelo AR(7). Fonte: Autor	49
Figura 23 – Gráfico de autocorrelação. Fonte: Autor	50
Figura 24 – Previsão do modelo ARMA(7,2). Fonte: Autor	51
Figura 25 – Comparação ARMA(2,1) e AR(2). Fonte: Autor	52
Figura 26 – Série gerada da primeira diferença. Fonte: Autor	52
Figura 27 – Autocorrelação parcial da série diferenciada. Fonte: Autor	53
Figura 28 – Autocorrelação da série diferenciada. Fonte: Autor	53
Figura 29 – Previsão do modelo ARIMA(1,1,1). Fonte: Autor	54
Figura 30 – Previsão do modelo ARIMA(7,2,1). Fonte: Autor	55
Figura 31 – Previsão do modelo ARIMA(1,1,1) com identificação online. Fonte: Autor	56
Figura 32 – Previsão dos modelos. Fonte: Autor	56

Lista de tabelas

Tabela 1 – Dados do ativo EURUSD. Fonte: MetaTrader5 [15]	23
Tabela 2 – Dados utilizados do ativo EURUSD	24
Tabela 3 – Dados enumerados	24
Tabela 4 – Parâmetros AR(2)	47
Tabela 5 – MSE de modelos AR(p) na base de seleção	48
Tabela 6 – Parâmetros AR(7)	49
Tabela 7 – MSE de modelos ARMA(p,q) na base de seleção	50
Tabela 8 – Parâmetros ARMA(7,2)	51
Tabela 9 – Parâmetros ARIMA(1,1,1)	54
Tabela 10 – MSE na base de teste	57

Lista de abreviaturas e siglas

AR: modelos auto-regressivos

ARMA: modelos auto-regressivos de médias móveis

ARIMA: modelos auto-regressivos integrado de médias móveis

MSE: Erro médio quadrático

ACF: Função de autocorrelação

PACF: Função de autocorrelação parcial

Sumário

1	INTRODUÇÃO	21
1.1	Descrição geral do problema	21
1.2	Motivação e Argumentação	21
1.3	Objetivos	22
1.4	Estrutura do documento	22
2	DEFINIÇÃO DO PROBLEMA	23
2.1	Dados	23
2.2	Previsão	24
3	REVISÃO BIBLIOGRÁFICA	27
3.1	Sistemas e modelagem	27
3.2	Identificação de sistemas	28
3.2.1	Identificação <i>online</i>	28
3.2.2	Identificação <i>offline</i>	29
3.3	Séries temporais	29
3.4	Modelos para séries temporais	31
3.4.1	Operadores e Funções	31
3.4.2	Processo Estocástico	32
3.4.2.1	Estacionariedade	33
3.4.2.2	Ruído Branco	35
3.4.3	Classes de modelos	36
3.4.3.1	Modelos auto-regressivos (AR)	36
3.4.3.2	Modelos auto-regressivos de médias móveis (ARMA)	37
3.4.3.3	Modelos auto-regressivos integrado de médias móveis (ARIMA)	37
4	METODOLOGIA	39
4.1	Identificação de um modelo	39
4.1.1	PACF e termos autoregressivos	40
4.1.2	ACF e termos de médias móveis	41
4.2	Identificação de parâmetros	42
4.2.1	Ajuste de parâmetros para modelos AR	42
4.2.2	Ajuste de parâmetros para modelos ARMA	43
4.3	Verificação do modelo	44
5	RESULTADOS	45

5.1	Dados	45
5.2	Modelagem e identificação	46
5.2.1	Modelos AR	46
5.2.2	Modelos ARMA	49
5.2.3	Modelos ARIMA	52
5.3	Análise entre classes dos modelos	55
6	CONSIDERAÇÕES FINAIS	59
	REFERÊNCIAS	61

1 Introdução

1.1 Descrição geral do problema

O mercado financeiro é um ambiente de compra e venda de ativos, em que cada ativo tem seu preço variável no tempo. Um exemplo de ativo é a ação de uma empresa na bolsa de valores ou até commodities como o petróleo, como pode-se ver na Figura 1.



Figura 1 – Ação PETR4 da empresa Petrobras nos últimos 10 anos. Fonte: Bolsa PT [1]

Atualmente há dois tipos de análises em cima dos ativos para tentar prever um aumento ou queda do seu preço. A primeira é a análise fundamentalista, que busca entender os fatores econômicos que afetam o ativo. Um exemplo é ter acesso antecipado a informação de que foi descoberto uma grande fonte de petróleo e com o aumento da concorrência o seu preço logo irá descer. A segunda análise é a técnica, também conhecida como análise gráfica, que é baseada nos dados do gráfico do ativo e será a análise estudada no presente trabalho.

A análise técnica, em princípio, estuda o preço dos ativos no passado para encontrar padrões e indicadores que possam dar sinais sobre o futuro mais provável. Com o uso de conhecimentos de modelagem, identificação de sistemas e com os avanços no poder de processamento pode-se também pensar em utilizar os dados para treinar modelos matemáticos a fim de realizar previsões.

1.2 Motivação e Argumentação

Há duas principais motivações para resolver o problema de previsão do valor de ativos.

A primeira motivação é utilizar os conhecimentos obtidos durante o curso para entender se modelos matemáticos podem ser utilizados para prever sistemas aleatórios como o mercado de ativos.

Na segunda motivação, deve-se primeiramente entender a magnitude do mercado de ativos. De acordo com a reportagem da Infomoney [16], a IBOVESPA, índice de maior relevância no Brasil, chegou a movimentar R\$9,96 bilhões ao dia em setembro de 2018, valor este que ainda está em crescimento, provando que há uma alta liquidez e variação de valores. Tal movimentação acaba por permitir que se ganhe dinheiro operando no mercado para vender ativos por preços superiores ao de sua compra, mas para isso deve-se prever quando e quanto o preço vai subir.

1.3 Objetivos

O objetivo geral desse trabalho é analisar o desempenho das classes de modelos matemáticos AR, ARMA e ARIMA em dados do mercado financeiro.

Para isso, deve-se cumprir os seguintes objetivos específicos:

- Explicitar a relação de dados do ativo financeiro e séries temporais;
- Classificar e descrever os modelos matemáticos;
- Definir parâmetros para cada modelo;
- Analisar o desempenho dos modelos gerados em cada classe;
- Definir métricas para análise de previsão.

1.4 Estrutura do documento

No capítulo 2 será explicado o problema, os dados obtidos e o que será calculado. O capítulo 3 será composto por uma revisão bibliográfica onde serão abordados assuntos que vão ser usados como base para estruturar matematicamente o problema. O capítulo 4 descreve as metodologias utilizadas na solução. No capítulo 5 serão aplicadas as metodologias ao problema proposto e será realizada a descrição dos resultados. Por fim, no capítulo 6, serão apresentadas um resumo do que foi feito e sugestões para trabalhos futuros.

2 Definição do problema

Neste capítulo o problema será explicado e delimitado de forma que seja possível cumprir nossos objetivos. Serão apresentados os dados utilizados do mercado financeiro, o que será calculado através destes e os resultados esperados.

2.1 Dados

Com o objetivo geral sendo análise de modelos matemáticos e não de ativos, será utilizado apenas o ativo de mercado de câmbio *EURUSD*, um valor que indica a proporção da moeda dólar para cada moeda euro.

Após a definição do ativo de interesse, deve-se entender que os dados que são disponibilizados no mercado financeiro são valores discretos que dependem de um tempo de amostragem. Esse tempo é popularmente chamado de período ou *Timeframe*, para simplificar nosso problema, será utilizado apenas o período semanal, ou seja, uma amostra por semana.

Por fim, ao definir o período, se obtém dados relativos à data, preço e volume. O preço de um ativo é formado pelo valor financeiro das suas ordens de compra ou venda. Já o volume é o número de ordens de compra e venda negociadas no período especificado.

O preço pode ser decomposto em quatro valores:

- Abertura: O preço ao início do período
- Máximo: O preço máximo durante o período
- Mínimo: O preço mínimo durante o período
- Fechamento: O preço ao fim do período

Na Tabela 1, pode-se visualizar todos os dados obtidos de um ativo.

Tabela 1 – Dados do ativo EURUSD. Fonte: MetaTrader5 [15]

Data	Abertura	Máximo	Mínimo	Fechamento	Volume
2014.07.01	1,36055	1,36401	1,33665	1,33877	616288
2014.08.01	1,33872	1,34446	1,31323	1,31323	844191
2014.09.01	1,31292	1,31600	1,25712	1,26307	1157836
2014.10.01	1,26306	1,28868	1,24851	1,25231	1746135

Para simplificar o problema de previsão, serão utilizados somente os dados de fechamento e data, como pode ser visto na Tabela 2. Presume-se então o problema a estas duas dimensões.

Tabela 2 – Dados utilizados do ativo EURUSD

Data	Fechamento
2014.07.01	1,33877
2014.08.01	1,31323
2014.09.01	1,26307
2014.10.01	1,25231

2.2 Previsão

Os valores de data são únicos e a cada nova amostra estes avançam no tempo conforme nosso período, pode-se até enumerar as amostras em ordem crescente conforme a Tabela 3.

Tabela 3 – Dados enumerados

Amostra	Data	Fechamento
1	2014.07.01	1,33877
2	2014.08.01	1,31323
3	2014.09.01	1,26307
4	2014.10.01	1,25231

Nota-se que o valor de fechamento pode ser uma variável dependente da amostra, declara-se então o índice t e os valores de fechamento como Y , chegando na Equação 2.1.

$$f(t) = X_t = Y \quad (2.1)$$

Utilizando os valores da Tabela 3, se tem o exemplo $f(1) = X_1 = 1,33877$.

Foi atribuída então uma ordem aos dados, o que leva ao questionamento de qual será o valor do fechamento X_5 . Ao visualizar os valores na Figura 2, é possível deduzir que o próximo valor seguiria em queda, mas não se tem conhecimento do valor exato.

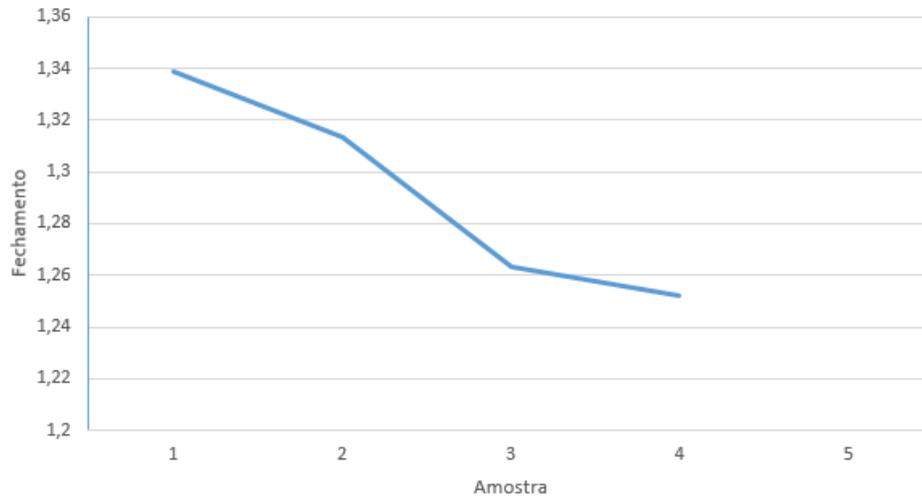


Figura 2 – Gráfico dos dados da Tabela 3. Fonte: Autor

Considerando que para qualquer conjunto de dados do mercado financeiro, o valor de fechamento mais recente obtido é X_t . Busca-se formas de prever X_{t+k} , onde k é o número de amostras à frente, através de diferentes modelos matemáticos.

3 Revisão Bibliográfica

Nesse capítulo será feita uma revisão teórica com o objetivo de embasar todo o projeto, em que será explicado o que é sistema e modelagem, séries temporais e suas características, modelagem de séries temporais, notações matemáticas e outros.

3.1 Sistemas e modelagem

Um sistema é a combinação de componentes que agem em conjunto para atingir determinado objetivo. A ideia de sistema não fica restrita apenas a algo físico, o conceito pode ser aplicado a fenômenos abstratos e dinâmicos, como aqueles encontrados na economia.

Em modelagem de sistemas, pode-se ver um sistema como um conjunto de elementos que interagem com o seu meio através de entradas e saídas, visando construir modelos que expliquem as características ou o comportamento de um sistema utilizando ferramentas matemáticas.

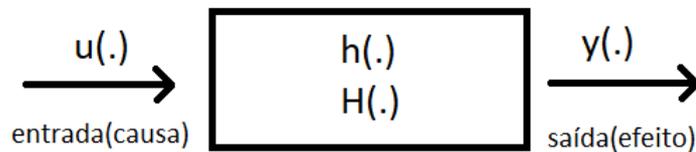


Figura 3 – Componentes de um sistema de controle. Fonte: Coelho [2]

Em que, segundo Coelho [2], os problemas associados à estrutura da Figura 3 são:

- Análise: é conhecido a entrada $u(\cdot)$, o sistema $h(\cdot)$ e deve-se obter a saída $y(\cdot)$;
- Projeto: é conhecido o sistema, $h(\cdot)$, a saída desejada, $y(\cdot)$ e deve-se obter a entrada $u(\cdot)$ para proporcionar tal saída;
- Identificação: é conhecido a entrada $u(\cdot)$, a saída $y(\cdot)$ e deve-se obter o sistema $h(\cdot)$, de modo que a saída gerada $y'(\cdot)$ seja aproximada da saída real $y(\cdot)$.

Ao se fazer a modelagem de um sistema, não se procura modelos com erro nulo, mas sim um modelo que mesmo com erros solucione o problema.

3.2 Identificação de sistemas

A identificação do sistema se concentra na criação de modelos de sistemas a partir de dados experimentais, com o principal objetivo de otimizar parâmetros de modelos candidatos a representar um processo real.

Podemos visualizar na Figura 4 um problema clássico de identificação, onde há o objetivo de localizar a_i, b_i de forma que $\hat{y} \approx y$.

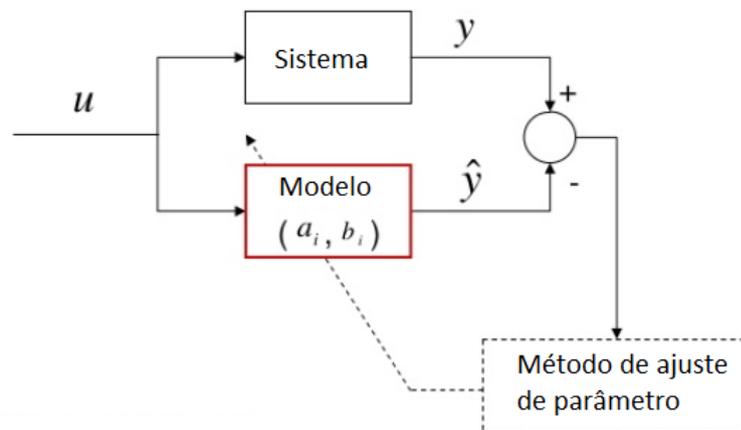


Figura 4 – Parametrização de modelos. Fonte: Dr. Oubbati, Kybernetik [3]

Coelho [2] cita que a área de identificação tem tido consideravelmente interesse nos últimos anos para fins de previsão, supervisão, diagnóstico e controle.

Existem diferentes procedimentos para identificar sistemas, sendo os principais:

- Identificação pelo teste de resposta ao degrau;
- Identificação pelo teste de resposta em frequência;
- Identificação *online*;
- Identificação *offline*.

Como não há a possibilidade de testar entradas no sistema do problema estudado, será revisado somente os tópicos de identificação *online* e identificação *offline*.

3.2.1 Identificação *online*

Os algoritmos de identificação *online* estimam os parâmetros de um modelo para cada novos dados disponíveis. A estimativa de parâmetros *online* pode ser realizada usando um algoritmo recursivo. Para estimar os valores dos parâmetros em um intervalo de tempo, algoritmos recursivos usam as medições atuais e as estimativas anteriores dos parâmetros.

Portanto, algoritmos recursivos são eficientes em termos de uso de memória. Além disso, algoritmos recursivos têm demandas computacionais menores.

3.2.2 Identificação *offline*

Na estimativa *offline*, você primeiro coleta todos os dados de entrada / saída e depois estima uma única vez os parâmetros do modelo. Os valores dos parâmetros estimados usando estimativa *online* podem variar com o tempo, mas os parâmetros estimados usando estimativa *offline* não.

3.3 Séries temporais

Uma série temporal é um conjunto de observações X_t feitas em sequência ao longo do tempo. Na maioria dos modelos, a ordem das observações é irrelevante para a análise. Por outro lado, em séries temporais a ordem dos dados é fundamental.

A principal característica de uma série temporal é que as observações vizinhas são relacionadas. Por exemplo, a quantidade de carros vendidos em dezembro pode estar relacionado à quantidade vendida em novembro, ou até mesmo relacionado com a de setembro.

É muito comum ver séries temporais relacionadas a temperatura ou preços, como pode ser visto na Figura 5.

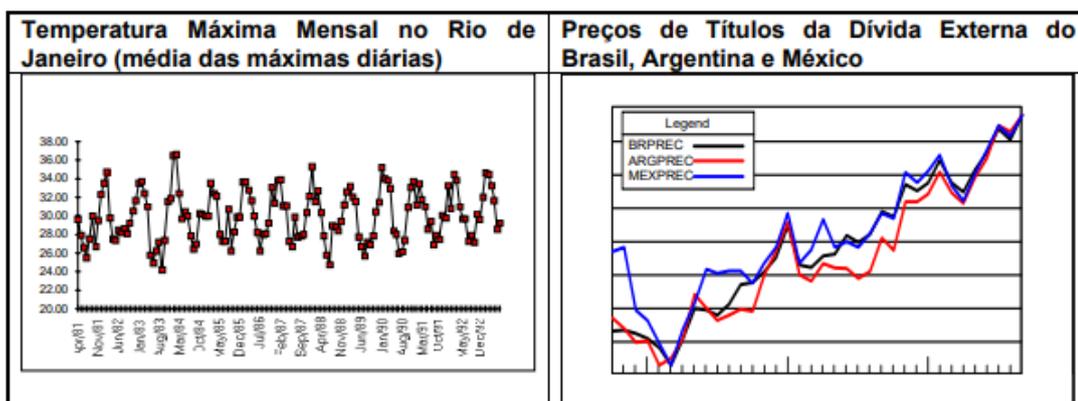


Figura 5 – Exemplo de séries temporais. Fonte: Resumo Wooldridge [4]

Pode-se separar séries temporais em dois grupos principais:

1. Univariado: A previsão depende apenas dos valores passados da própria variável. Exemplo: concentração mensal de CO_2 depende apenas de seus valores passados.

2. Multivariados: A previsão depende de valores passados de diversas variáveis, por exemplo: a precipitação depende da precipitação passada, temperatura, velocidade do vento, etc.

Ainda se têm séries temporais contínuas como a temperatura, ou discretas como a precipitação diária. Pode-se discretizar séries temporais contínuas através da amostragem em intervalo de tempos iguais.

Séries temporais possuem diversas propriedades, as principais estão listadas abaixo:

- Tendência: A média, a variância e a estrutura de autocorrelação são constantes, lineares ou quadráticas, como ilustrado na Figura 6.

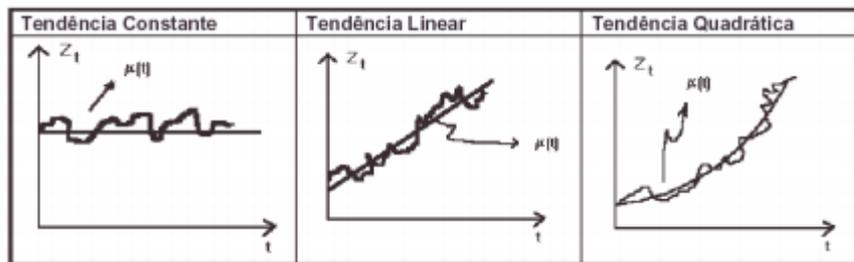


Figura 6 – Tendências de uma série temporal. Fonte: PUC-Rio [5]

- Sazonalidade: Flutuações periódicas, fenômenos que se repetem a cada período idêntico de tempo. Por exemplo, um crescimento de vendas no varejo em um dado período do ano, como pode ser visto na Figura 7.

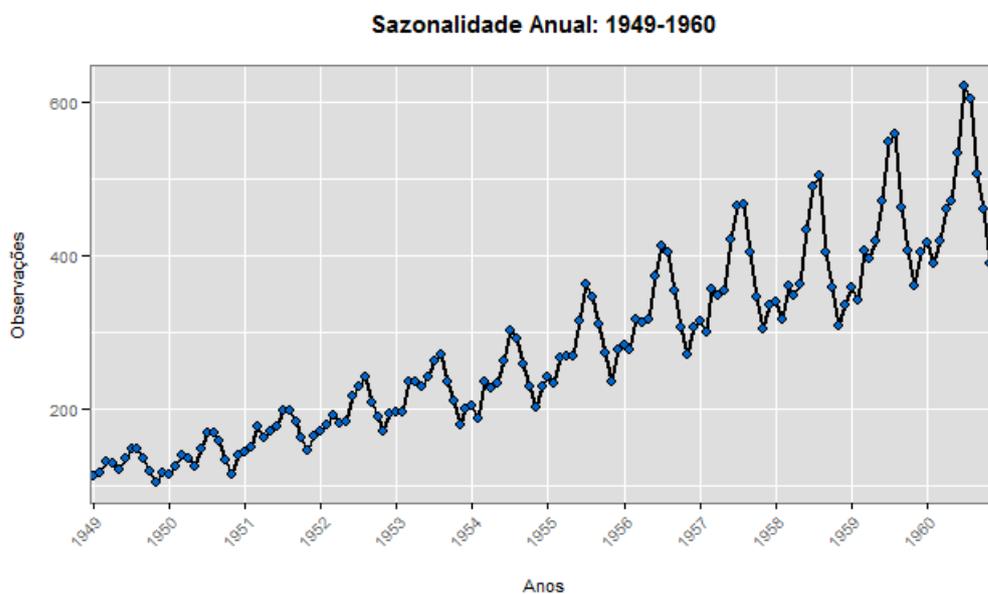


Figura 7 – Sazonalidade de uma série temporal. Fonte: Portal Action [6]

- **Pontos Influentes:** Observações geralmente com valores atípicos, conhecidas como *Outlier*, que podem influenciar nas previsões do modelo, um exemplo pode ser visto na Figura 8

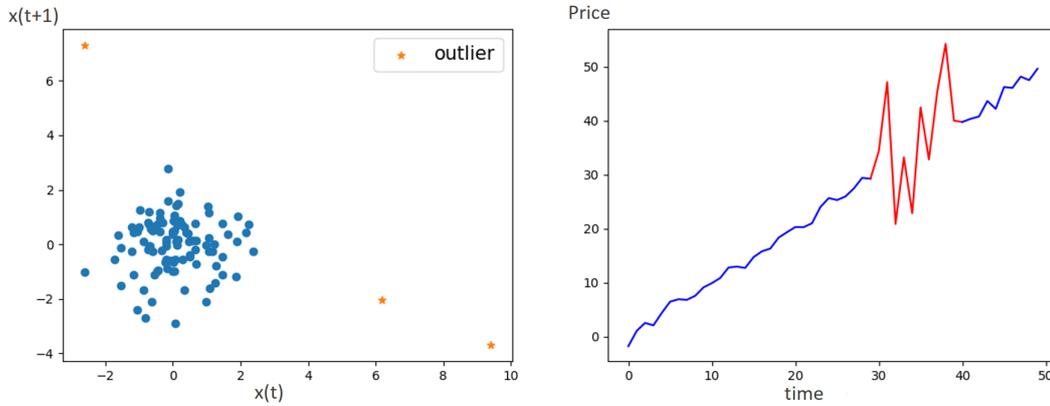


Figura 8 – Pontos influentes de uma série temporal. Fonte: ReNom [7]

3.4 Modelos para séries temporais

Para entender e formalizar os modelos deve-se primeiramente introduzir operadores e funções.

3.4.1 Operadores e Funções

Valor esperado: representa o valor médio μ de uma variável aleatória x com função densidade de probabilidade $f(x)$. É denotado por $E[x]$ e definido por:

$$E[x] = \mu = \int_{-\infty}^{\infty} x f(x) dx \quad (3.1)$$

Variância: representa o quão longe os valores de x se encontram do valor esperado. É denotado por $VAR[x]$ e definido por:

$$VAR[x] = E[(x - \mu)^2] \quad (3.2)$$

Desvio padrão: representa quão dispersos os dados estão da média. Quanto maior o desvio padrão, maior a dispersão nos dados. É denotado por $\sigma(x)$ e definido por:

$$\sigma(x) = \sqrt{VAR[x]} \quad (3.3)$$

Covariância: Representa como duas variáveis variam conjuntamente. É denotado por $COV(x, y)$ e definido por:

$$COV(x, y) = E[(x - \mu)(y - \mu)] \quad (3.4)$$

Isto equivale à seguinte fórmula:

$$COV(x, y) = E[xy] - E[x]E[y] \quad (3.5)$$

Função de autocorrelação: Representa qualquer relação estatísticas que envolva dependência entre duas variáveis. É denotado por $ACF(x_t, x_{t-k})$ e definido por:

$$ACF(x_t, x_{t-k}) = \frac{COV(x_t, x_{t-k})}{\sqrt{VAR(x_t)VAR(x_{t-k})}} \quad (3.6)$$

A autocorrelação tem propriedades importantes:

- (i) $1 \leq ACF(x_t, x_{t-k}) \leq -1$;
- (ii) $ACF(x_t, x_{t-k}) = 1$: relação linear perfeita e positiva;
- (iii) $ACF(x_t, x_{t-k}) = -1$: relação linear perfeita e negativa;
- (iv) $ACF(x_t, x_{t-k}) = 0$: inexistência de relação linear.

Função de autocorrelação parcial: Representa a correlação entre duas variáveis X_t e X_{t-k} eliminando a dependência dos termos intermediários $X_{t-1}, X_{t-2}, \dots, X_{t-(k+1)}$. É denotado por $PACF(x_t, x_{t-k})$ e definido por:

$$PACF(x_t, x_{t-k}) = ACF(x_t, x_{t-k} | x_{t-1}, \dots, x_{t-(k+1)}) \quad (3.7)$$

3.4.2 Processo Estocástico

Séries temporais não podem ser modeladas por modelos determinísticos. Há sempre erros que são controlados por estatística e leis probabilísticas, processos assim são nomeados de estocásticos.

Um processo estocástico é um fenômeno que varia em algum grau de forma imprevisível à medida que o tempo passa. Segundo Morettin [8] pode-se formalizar como:

$$Z = \{Z(t), t \in T\} \quad (3.8)$$

Onde T é um conjunto arbitrário, geralmente o conjunto dos reais \mathbb{R} , tal que $Z(t)$ é uma variável aleatória.

Sendo assim, Z é uma família de variáveis aleatórias, que supomos definidas num mesmo espaço de probabilidade (Ω, A, P) . Um espaço de probabilidade é uma tripla (Ω, A, P) formada por um conjunto Ω chamado de espaço amostral, elementos de A são chamados de eventos e a medida P é chamada de medida de probabilidade.

Para $t \in T$, $Z(t)$ é uma variável definida em Ω , sendo então que $Z(t)$ na realidade é uma função de dois argumentos $Z(t, \omega)$ dado que $\omega \in \Omega$, mas que em situações usuais ω é um valor fixo.

Na Figura 9, se tem um exemplo de processo estocástico, onde para cada t , se tem uma variável aleatória $Z(t, \omega)$ com uma distribuição de probabilidade $f_z(Z)$. Ao aplicar a função de valor esperado $\mu(t)$, obtemos então uma trajetória do processo, ou ainda, uma série temporal.

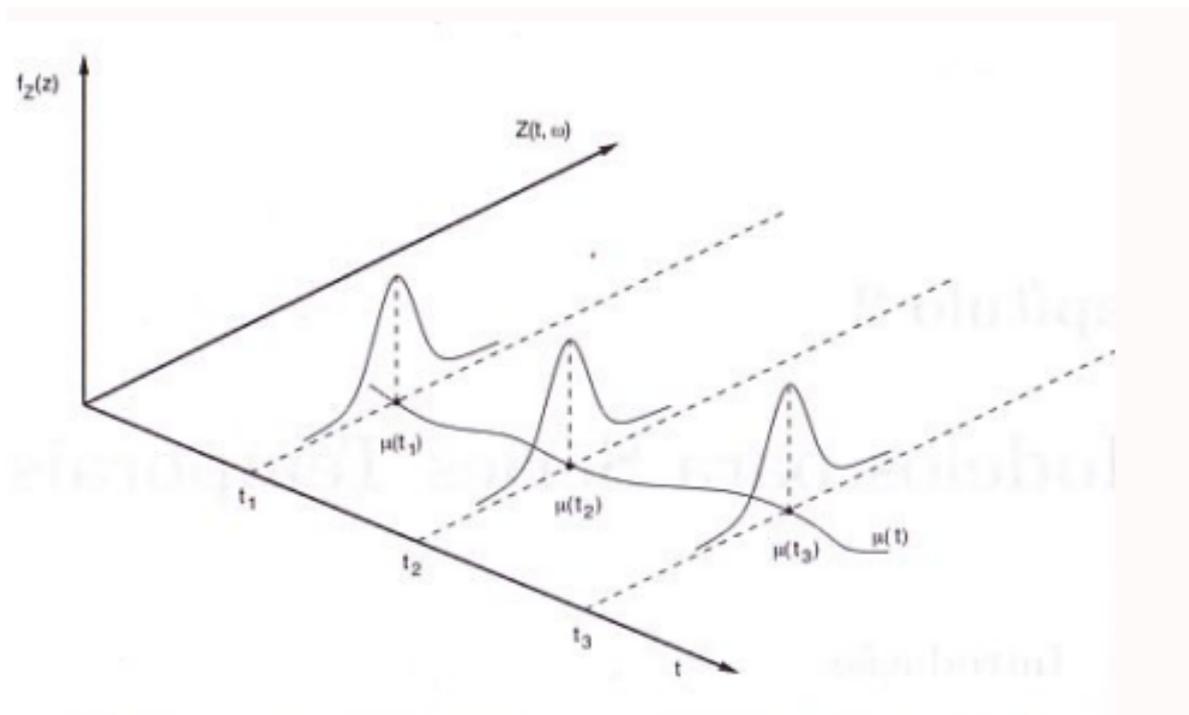


Figura 9 – Processo estocástico. Fonte: Morettin [8]

3.4.2.1 Estacionariedade

Se o processo estocástico que gerou a série de observações não é alterado no tempo, ou seja, o processo se desenvolve no tempo em torno da média, de modo que a escolha de uma origem dos tempos não é importante, se diz que o mesmo é estacionário. Pode-se visualizar essa propriedade na Figura 10.

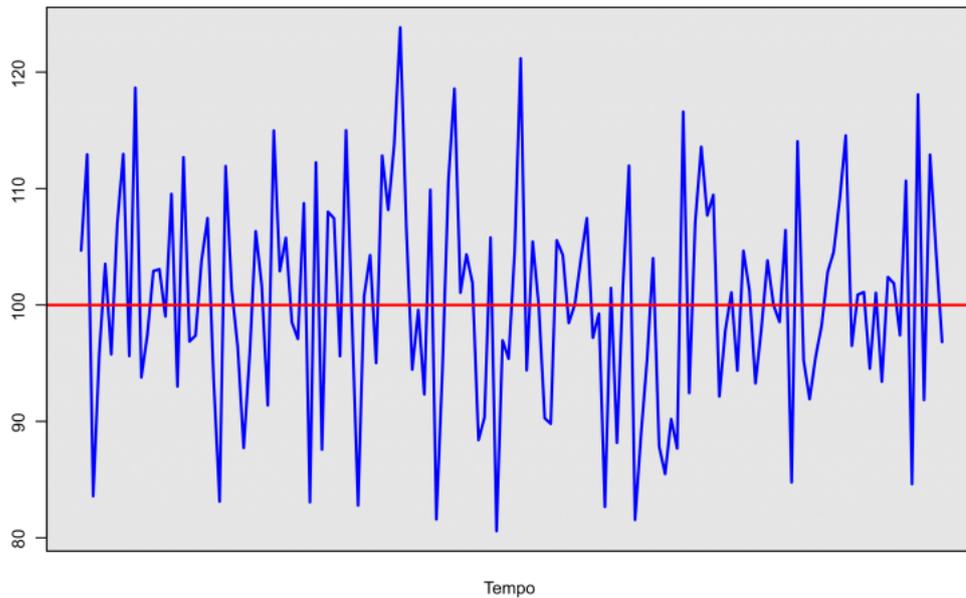


Figura 10 – Série temporal estacionária. Fonte: Abg Consultoria [9]

Morettin [8] classifica os processos estacionários como:

1. Estritamente estacionário: Quando todas as distribuições da série X_t são invariantes sob translações do tempo, logo a média $\mu(t)$ e variância $VAR(t)$ são constantes. Exemplo: Ruídos.
2. Fracamente estacionário: Quando o seu valor médio é constante, variância finita e sua função de covariância $COV(t_1, t_2)$ é uma função de $|t_1 - t_2|$.

Denominaremos por processo estacionário somente quando o processo for fracamente estacionário.

Há também uma classe de processos chamada **não estacionários homogêneos**. Tais processos são definidos assim pois ao tomarmos suas diferenças pela equação (3.9), encontramos um processo estacionário.

$$Dif(X_t) = X_t - X_{t-1} \quad (3.9)$$

Pode-se visualizar a transformação de uma série não estacionária em uma estacionária pelas Figuras 11 e 12, respectivamente.

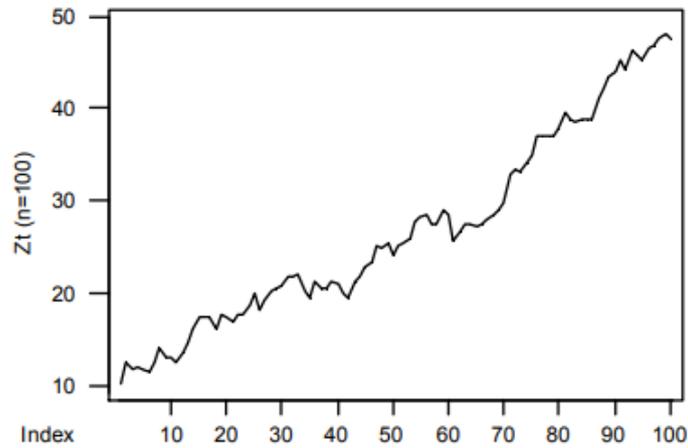


Figura 11 – Série não-estacionária homogênea. Fonte: Estatística UFMG [10]

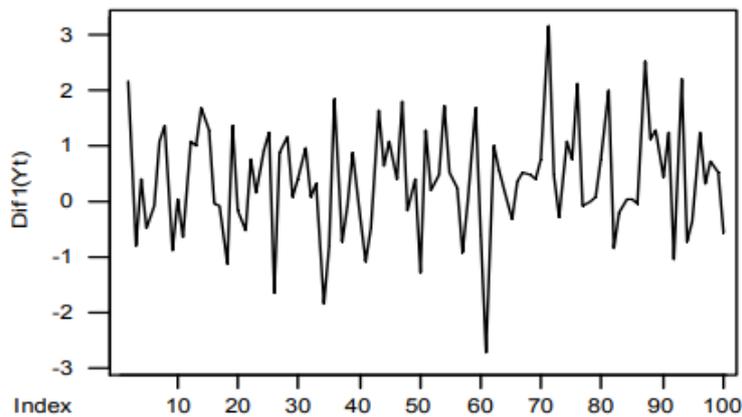


Figura 12 – Série estacionária gerada da primeira diferença. Fonte: Estatística UFMG [10]

3.4.2.2 Ruído Branco

Ruído branco é uma sequência de variáveis aleatórias, não autocorrelacionadas com valor médio zero e variância finita, denotado por ε_t .

$$E[\varepsilon_t] = 0 \quad (3.10)$$

$$Var[\varepsilon_t] = \sigma^2 < \infty \quad (3.11)$$

$$Corr(\varepsilon_t, \varepsilon_{t-1}) = 0 \quad (3.12)$$

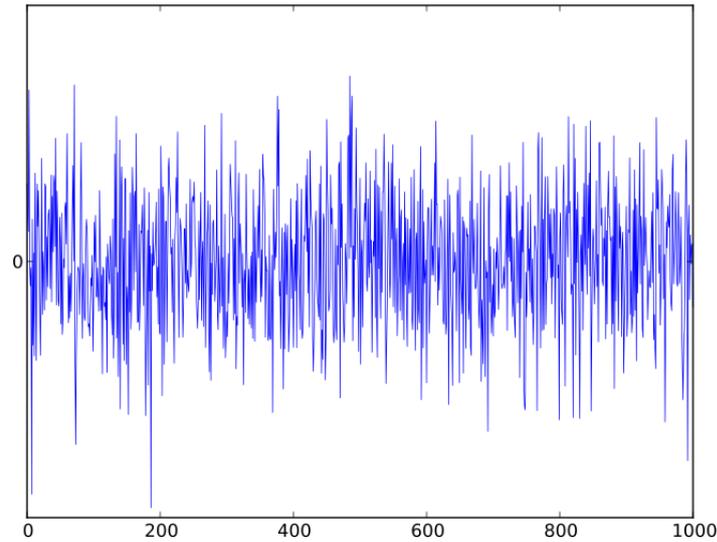


Figura 13 – Ruído branco. Fonte: Wikipédia [11]

3.4.3 Classes de modelos

Os modelos para séries temporais podem ser classificados em modelos paramétricos, nos quais o número de parâmetros é finito e, modelos não-paramétricos, os quais envolvem um número infinito de parâmetros. Na classe de modelos paramétricos, a análise é feita no domínio do tempo. Dentre estes modelos, os mais frequentemente utilizados são os modelos auto-regressivo (AR), o auto-regressivo e de média móvel (ARMA), o modelo auto-regressivo integrado e de média móvel (ARIMA), modelos de memória longa (ARFIMA), modelos estruturais e modelos não-lineares. Os modelos não-paramétricos mais usados são a função de auto-covariância e sua transformada de Fourier. Morettin [8].

3.4.3.1 Modelos auto-regressivos (AR)

Dada uma série temporal X_t , pode-se modelar esta pelo modelo $AR(p)$, onde p é a ordem do modelo.

O modelo $AR(p)$ tem como entrada os valores passados $X_{t-1}, X_{t-2}, \dots, X_{t-p}$. Estas entradas são multiplicadas por parâmetros otimizados ϕ para prever o próximo valor X_t .

Uma forma comum de representar $AR(p)$ é pela equação (3.13).

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t \quad (3.13)$$

onde c é uma constante, ϕ_i são os parâmetros e ε_t é o erro aleatório, chamado de ruído branco. Pode-se dar como exemplo o modelo de primeira ordem $AR(1)$, sendo este definido

de acordo com a equação (3.14).

$$X_t = c + \phi_1 X_{t-1} + \varepsilon_t \quad (3.14)$$

3.4.3.2 Modelos auto-regressivos de médias móveis (ARMA)

Dada uma série temporal X_t , pode-se modelar esta pelo modelo $ARMA(p, q)$, onde p, q são a ordem do modelo.

Este modelo utiliza tanto um modelo auto-regressivo como o modelo de média móvel. O modelo de média móvel utiliza os erros anteriores para melhorar a previsão, trazendo então para equação os valores passados do ruído branco $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$, os quais são também multiplicadas por parâmetros otimizados θ que visam melhorar a previsão do próximo valor X_t .

Pode-se representar $ARMA(p, q)$ pela equação (3.15).

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (3.15)$$

onde c é uma constante, ϕ_i e θ_i são parâmetros e ε_t é o erro aleatório. Pode-se dar como exemplo o modelo de primeira ordem $ARMA(1, 1)$, definido de acordo com a equação (3.16).

$$X_t = c + \varepsilon_t + \phi_1 X_{t-1} + \theta_1 \varepsilon_{t-1} \quad (3.16)$$

3.4.3.3 Modelos auto-regressivos integrado de médias móveis (ARIMA)

Dada uma série temporal X_t não estacionária homogênea, utilizamos a equação da diferença (3.9) para criar uma nova série temporal \dot{X}_t , definida na Equação (3.17).

$$\dot{X}_t = \Delta X_t = X_t - X_{t-1} \quad (3.17)$$

\dot{X}_t é a série diferenciada uma vez de X_t , mas esse método é aplicado até se encontrar uma série estacionária. Denotamos então por $\dot{X}_t^d = \Delta^d X_t$ a série temporal diferenciada d vezes de X_t .

Gerada a série estacionária \dot{X}_t , pode-se representá-la por um modelo ARMA. Como \dot{X}_t é uma diferença de X_t , então X_t é uma integral de \dot{X}_t , assim dizemos que X_t segue um modelo autorregressivo integrado e de médias móveis, ou seja, segue um modelo $ARIMA(p, d, q)$, onde p, q são a ordem do modelo $ARMA(p, q)$ e d a ordem da diferenciação.

Portanto, o modelo ARIMA supõe que a diferença da série X_t pode ser representada por um modelo ARMA, estacionário. Na maioria dos casos usuais, $d = 1$, representado na Equação (3.18)

$$\dot{X}_t = c + \varepsilon_t + \sum_{i=1}^p \phi_i \dot{X}_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (3.18)$$

4 Metodologia

Uma abordagem amplamente utilizada para a construção de modelos para séries temporais é conhecida como metodologia Box e Jenkins. Morettin [8]

O método utilizado para elaborar modelos para séries temporais será baseado em um ciclo iterativo, ilustrado na Figura 14, no qual a escolha da estrutura do modelo é baseada nos próprios dados. Os estágios do ciclo iterativo são:

1. Identificar um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
2. A seguir vem a fase de estimação, na qual os parâmetros do modelo identificado são estimados;
3. Finalmente, realizar a verificação ou diagnóstico do modelo ajustado, através de uma análise de erros, para se saber se este é adequado para os fins desejados (previsão, por exemplo)

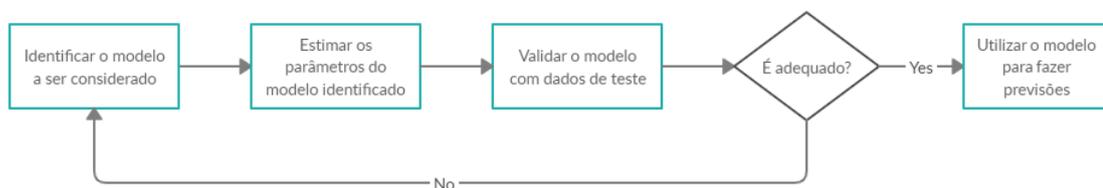


Figura 14 – Diagrama de etapas iterativas na construção de modelos de previsão. Fonte: Autor

Nesse capítulo será analisada cada etapa desse estágio e serão realizados ajustes em relação aos modelos escolhidos para a aplicação em nosso problema.

4.1 Identificação de um modelo

Nessa etapa precisa-se estimar a ordem p do modelo auto-regressivo, o grau de diferenciação d e a ordem q do modelo de média móvel.

A identificação da ordem d é feita através da análise da estacionariedade da série temporal. A análise pode ser feita graficamente, onde tenta-se verificar se o valor médio é constante, em caso negativo, aplica-se a diferença pela Equação (3.17) recursivamente até se atender o requisito de valor médio constante. O valor da ordem d é definido pela

quantidade de vezes em que se é aplicado a diferença na série temporal. Nas Figuras 11 e 12 consegue-se ver um exemplo de uma série com ordem $d = 1$.

Nas ordens p e q , é comum a identificação ao selecionar modelos de menor ordem, parametrizá-los e então aplicar algum critério de verificação aos erros gerados para selecioná-los. Mas segundo Tsay [17], pode-se estimar as ordens p e q através dos gráficos gerados das Equações ACF (3.6) e PACF (3.7), utilizando dados atrasados da própria série temporal. O gráfico da ACF nos mostra os coeficientes de correlação de uma série temporal a cada atraso. Já o gráfico PACF é um gráfico dos coeficientes de correlação parciais entre as séries e os atrasos em si.

4.1.1 PACF e termos autoregressivos

A PACF é a quantidade de correlação entre duas variáveis, em que não se é considerado as correlações intermediárias. Para se entender melhor, considere a ACF de uma série temporal X_t com seu atraso 1, ou seja $ACF(X_t, X_{t-1})$, que presumivelmente é também a correlação entre $ACF(X_{t-1}, X_{t-2})$. Mas se X_t estiver correlacionado com X_{t-1} e X_{t-1} estiver igualmente correlacionado com X_{t-2} , também devemos esperar encontrar correlação entre X_t e X_{t-2} . De fato, a ACF que devemos esperar no atraso 2 é precisamente o quadrado da ACF do atraso 1. Assim, a correlação no atraso 1 se propaga para o atraso 2 e até para defasagem de ordens superiores. Já a PACF no atraso 2, verifica a ACF entre X_t e X_{t-2} desconsiderando as ACF feitas com X_{t-1} , isto é, a diferença entre a correlação real no atraso 2 e a correlação esperada devido à propagação da correlação no atraso 1.

Uma forma simples de entender a PACF é pela brincadeira *Telefone sem fio*.

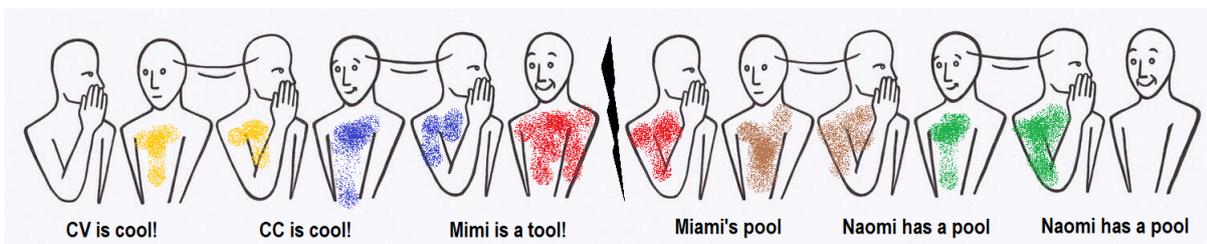


Figura 15 – Ilustração do *Telefone sem fio*. Fonte: StackExchange [12]

A mensagem passada na Figura 15 fica distorcida à medida que é sussurrada de pessoa para pessoa, se for comparada a mensagem do primeiro participante ao segundo, há poucas diferenças, da mesma forma que se comparado a mensagem do segundo participante com a mensagem do terceiro. Entretanto, ao se comparar a mensagem do primeiro participante diretamente com a do terceiro nota-se que o erro aumenta consideravelmente, e todo o sentido da frase é perdido após o quarto participante.

Se fosse possível fazer um gráfico da PACF desta ilustração, este seria parecido com o gráfico representado na Figura 16, nos mostrando que à partir do terceiro participante, ou terceiro atraso, há pouca contribuição de informação. É muito comum em gráficos de

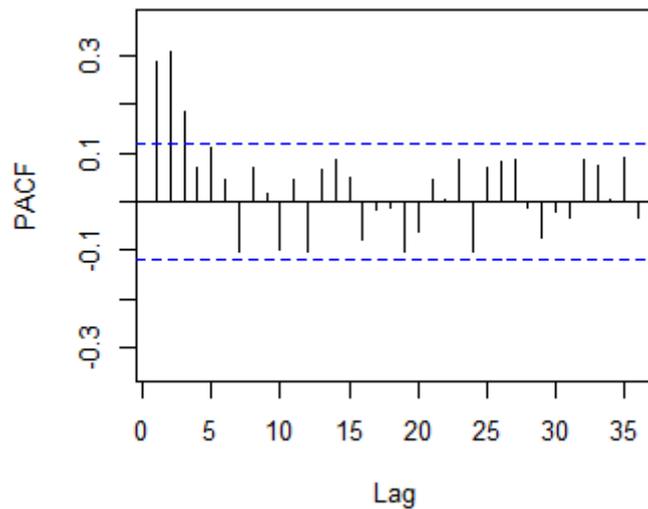


Figura 16 – Gráfico da PACF. Fonte: StackExchange [13]

PACF traçar uma linha azul, estes limites são calculados pela Equação (4.1), utilizando o número de amostras N . Ao identificar os picos de ACF que estão fora do limite da linha pode-se identificar até qual atraso há informações relevantes.

$$Limites = \pm \frac{2}{\sqrt{N}} \quad (4.1)$$

Através do gráfico da PACF pode-se estimar até qual atraso temos relação com o valor presente, pode-se então assumir que este é um bom indicador para a ordem do termo auto-regressivo, que utiliza de valores atuais e passados para calcular os valores futuros, sendo assim, não há sentido em utilizar atrasos que pouco influenciam na sua previsão.

A forma mais simples de identificar a ordem dos termos auto-regressivo através do gráfico da PACF, é identificar quais atrasos tem valores que estão fora da área formada pelos limites, por exemplo na Figura 16, os três primeiros atrasos estão acima do limite superior, indicando que um bom modelo para representa-la é o AR(3).

Mais informações sobre identificação da ordem de termos auto-regressivos por PACF em *Analysis of Financial Time Series* [17].

4.1.2 ACF e termos de médias móveis

A ACF desempenha papel parecido nos termos de médias móveis que a PACF desempenha nos termos auto-regressivos. Lembra-se que as médias móveis utilizam os

valores de erro passados e não os valores propriamente ditos, assim ao visualizar o gráfico da ACF consegue-se notar até que atrasos há erros relevantes.

Se a autocorrelação sofrer uma queda no atraso k , nas primeiras defasagens, isto é um forte indicativo que termos de médias móveis de ordem k devem ser usados na equação de previsão.

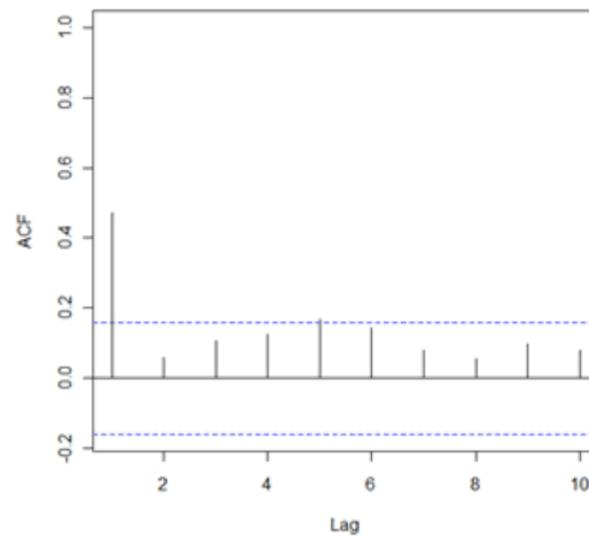


Figura 17 – Gráfico da ACF. Fonte: Pennsylvania State University [14]

Na Figura 17 é possível visualizar o gráfico de ACF de um processo que provavelmente pode ser representado por um modelo com termos de médias móveis de ordem $q = 1$.

Mais informações sobre identificação da ordem de termos de médias móveis por ACF em *Analysis of Financial Time Series* [17].

4.2 Identificação de parâmetros

4.2.1 Ajuste de parâmetros para modelos AR

Para um processo AR de ordem p é necessário definir apenas os parâmetros autorregressivos.

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (4.2)$$

Feita as observações X_1, \dots, X_N , os parâmetros ϕ_1, \dots, ϕ_p podem ser estimados pelo

método de mínimos quadrados. Define-se então:

$$Y = \begin{bmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_n \end{bmatrix} \quad X = \begin{bmatrix} X_p & \dots & X_1 \\ X_{p+1} & \dots & X_2 \\ \vdots & \vdots & \vdots \\ X_{n-1} & \dots & X_{n-p} \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad \Omega = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad (4.3)$$

Pode-se reescrever o modelo na forma matricial como:

$$Y = X\Omega + \varepsilon \quad (4.4)$$

A solução de mínimos quadrados para os coeficientes ϕ é obtida pela equação (4.5).

$$\Omega = (X^T X)^{-1} X^T Y \quad (4.5)$$

4.2.2 Ajuste de parâmetros para modelos ARMA

Os modelos ARMA não podem ter seus parâmetros ajustados pelo método dos mínimos quadrados tradicional, pois estes modelos possuem termos de médias móveis, ou seja, parâmetros que multiplicam o erro.

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon \quad (4.6)$$

Como solução pode-se aplicar uma extensão no método ao se expandir a matriz X e a matriz de parâmetros Ω para incluir os erros e parâmetros dos termos de médias móveis, respectivamente. Como os erros obtidos com o modelo serão utilizados para o cálculo do próprio modelo, se imagina que deve-se usar um algoritmo recursivo, de forma que a cada iteração o erro diminua.

A matriz estendida X^* pode ser calculada a partir da matriz X e dos erros ε , definidos na Equação (4.3). Primeiramente será definido o vetor α e β . O vetor α possui os valores da série temporal e o vetor β os erros.

$$\alpha_p = [X_p \quad X_{p-1} \quad \dots \quad X_1] \quad \beta_q = [\varepsilon_q \quad \varepsilon_{q-1} \quad \dots \quad \varepsilon_1] \quad (4.7)$$

A partir da Equação (4.7) pode-se definir a matriz estendida X^* .

$$X^* = \begin{bmatrix} \alpha_p & \beta_q \\ \alpha_{p+1} & \beta_{q+1} \\ \vdots & \vdots \\ \alpha_{n-1} & \beta_{n-1} \end{bmatrix} \quad (4.8)$$

A matriz estendida de parâmetros Ω^* segue a mesma lógica e pode ser definida ao se agrupar os parâmetros dos termos auto-regressivos ϕ e de médias móveis θ .

$$\gamma = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \delta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} \quad \Omega^* = \begin{bmatrix} \gamma \\ \delta \end{bmatrix} \quad (4.9)$$

Com as matrizes estendidas definida pode-se aplicar o Algoritmo 1 para parametrizar a classe de modelos $ARMA(p, q)$.

Algoritmo 1: Estimador estendido dos mínimos quadrados

- 1 Calcule Ω normalmente através do método de mínimos quadrados;
- 2 Calcule o vetor de erros $\varepsilon_1 = Y - X\Omega$;
- 3 **para** $i=2$ até Ω_i^* *convergir* **faça**
- 4 Monte a matriz estendida X_i^* ;
- 5 Calcule $\Omega_i^* = (X_i^{*T} X_i^*)^{-1} X_i^{*T} Y$;
- 6 Calcule novamente o vetor de erros $\varepsilon_i = Y - X_i^* \Omega_i^*$
- 7 **fim**

4.3 Verificação do modelo

A verificação da adequação do modelo é efetuada através da verificação do erro, essa verificação pode ser feita calculando o erro quadrático médio (MSE) pela Equação (4.10).

$$MSE = \frac{1}{n} \sum_{i=0}^n (X_i - \hat{X}_i)^2 \quad (4.10)$$

Onde X_t é o valor real, \hat{X}_t é o valor previsto e n é numero de amostras

MSE é provavelmente a métrica de erro mais utilizada. Ao elevar o erro ao quadrado penaliza os erros maiores gerando um impacto maior do que os erros de menor magnitude.

Existem outras métricas como Critério de Informação de Akaike (AIC) e Critério de Informação Bayesiano (BIC). Estes critérios penalizam modelos com maior número de variáveis, porém devido ao grande poder computacional, o número de variáveis acaba por ficar irrelevante e decide-se apenas reduzir o erro dos modelos.

5 Resultados

Neste capítulo serão aplicados os métodos apresentados no capítulo 4 aos dados do mercado financeiro. Primeiramente será explicado como que os dados foram obtidos e em seguida será aplicada a metodologia à cada classe de modelos.

5.1 Dados

Para obter os dados do ativo EURUSD, definido no capítulo 2, foi utilizado a ferramenta Metatrader5 [15].

Foi realizado o *download* dos dados com amostras semanais no período de 01/01/2015 até 31/12/2019. Este período foi escolhido de forma arbitrária e contém 261 amostras, representadas na Figura 18.

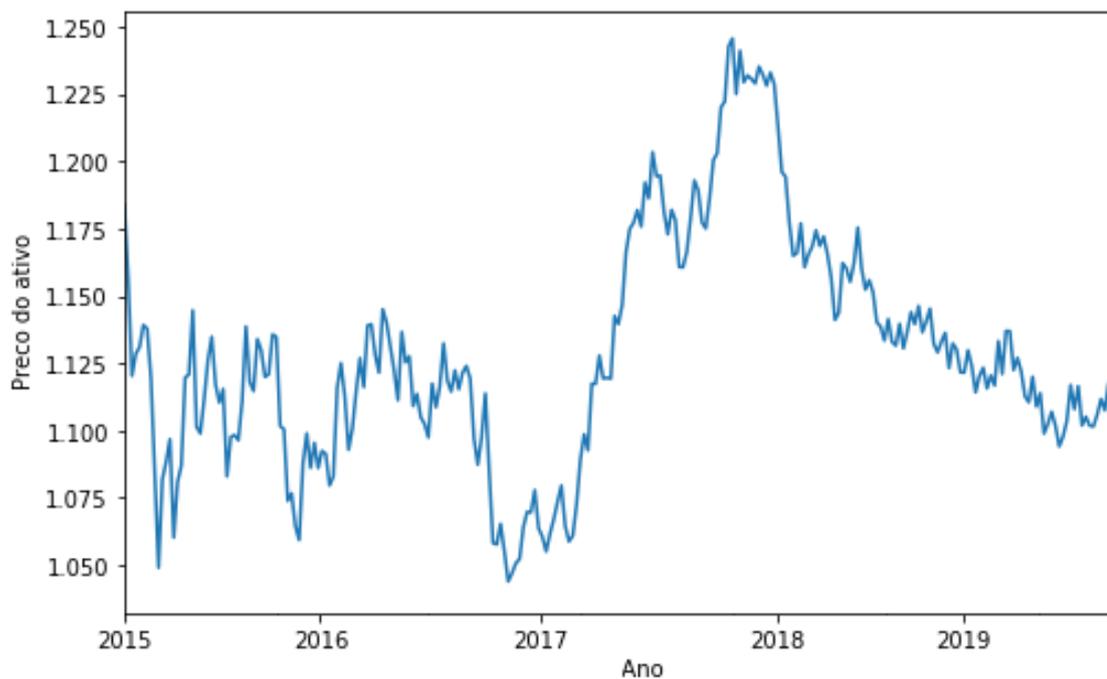


Figura 18 – Ativo financeiro EURUSD. Fonte: Autor

A base será separada em três partes:

1. 80% desta será utilizada para treino dos modelos, totalizando 209 amostras;
2. 13% para auxiliar na seleção dos modelos, totalizando 34 amostras.
3. 7% para previsão, totalizando 18 amostras.

Pode-se fazer uma rápida verificação na autocorrelação em nosso conjunto de dados de séries temporais ao plotar a observação no passo atual t com a observação no próximo passo ($t + 1$) como um gráfico de dispersão.

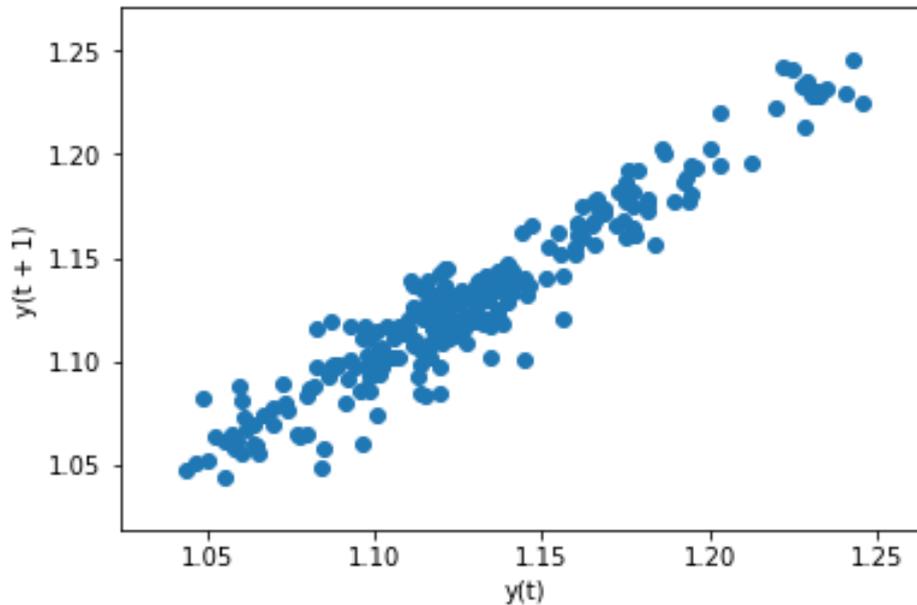


Figura 19 – Gráfico de atraso. Fonte: Autor

Na Figura 19 se observa uma linha diagonal, o que mostra um relacionamento entre as variáveis vizinhas, confirmando que nossos dados são uma série temporal.

5.2 Modelagem e identificação

Seguindo a metodologia proposta, será feito a identificação dos modelos, estimação dos parâmetros e diagnóstico individualmente para cada classe de modelo. É importante lembrar que esse é um processo baseado em um ciclo iterativo e portanto, caso os resultados não forem adequados será repetido novamente o processo. Ao fim será também feito uma comparação dos resultados entre os melhores modelos de cada classe.

5.2.1 Modelos AR

Inicialmente para identificar a ordem do modelo $AR(p)$, pode-se fazer uma análise da PACF nos dados de treino. Será utilizado 15 atrasos no gráfico e as amostras de treino, observa-se que o limite calculado pela Equação (4.1) é igual a 0.14.

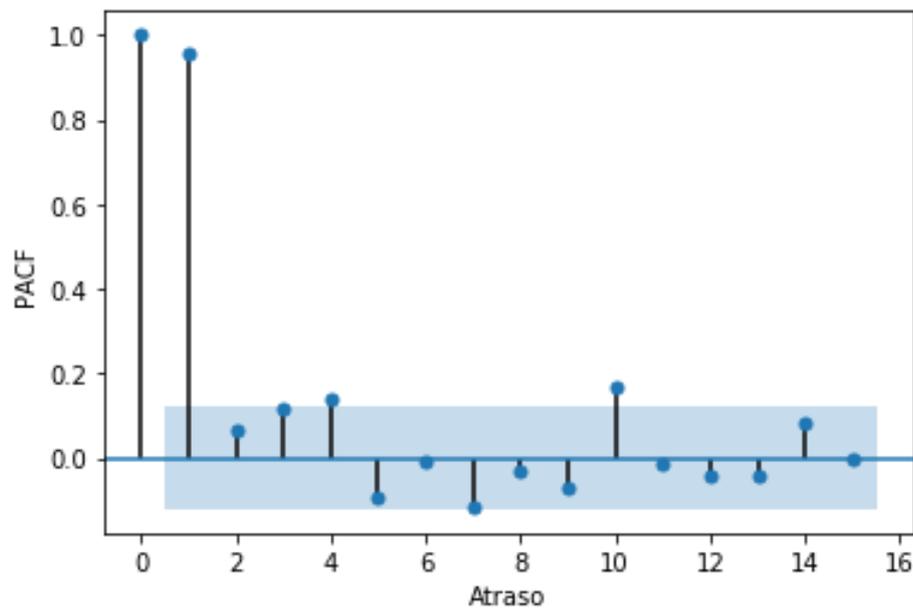


Figura 20 – Gráfico de autocorrelação parcial. Fonte: Autor

Pode-se visualizar na Figura 20 que a partir do segundo atraso, a PACF entra dentro dos limites, sendo assim, um bom modelo para representar estes dados é de ordem $p = 2$, ou seja, o modelo $AR(2)$.

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \quad (5.1)$$

Em seguida, utiliza-se do método dos mínimos quadrados para definir os parâmetros ϕ do $AR(2)$, definido na Equação (5.1). Os parâmetros estão listados na Tabela 4.

Tabela 4 – Parâmetros $AR(2)$

Modelo	Parâmetros
$AR(2)$	$\phi_0 = 0.04896053$
	$\phi_1 = 0.90487437$
	$\phi_2 = 0.05179287$

Com o modelo parametrizado, calcula-se que seu $MSE = 4,2 \cdot 10^{-5}$ e pode-se visualizar sua previsão na amostra de teste na Figura 21.

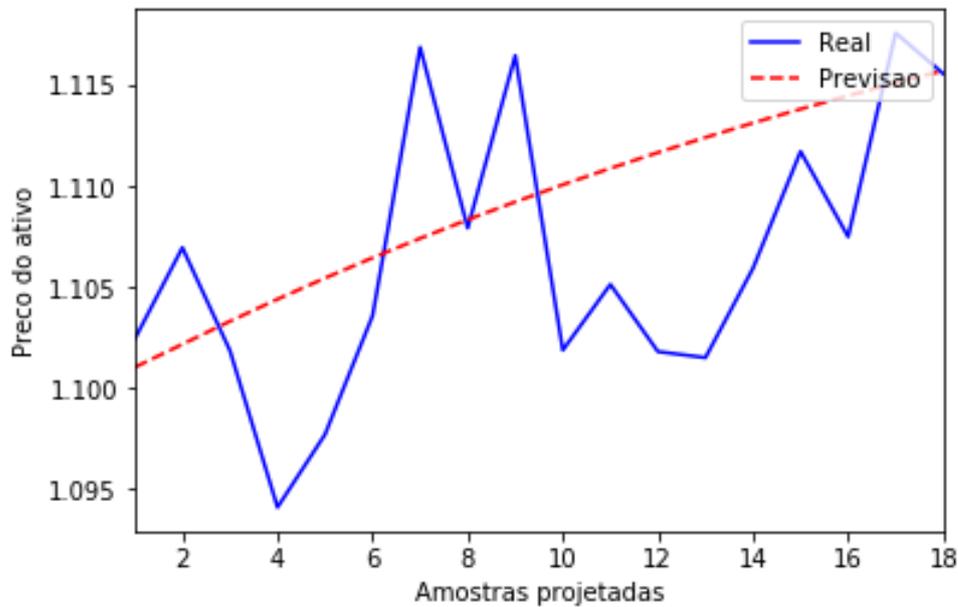


Figura 21 – Previsão do modelo AR(2). Fonte: Autor

O modelo AR(2) é simples e parece seguir a tendência, porém sua variação é muito pequena em relação à série temporal, isso deve-se principalmente a sua ordem ser de $p = 2$. Com base nesta análise pode-se criar e avaliar modelos AR com maior complexidade.

Ao aumentar a ordem dos modelos, pode-se fazê-la indefinidamente. Como critério será utilizado a variação da MSE, ou seja, aumenta-se a ordem do modelo até o momento que esta varie minimamente ou que comece a aumentar.

Tabela 5 – MSE de modelos AR(p) na base de seleção

Modelo	MSE
AR(1)	4,54.10 ⁻⁵
AR(2)	4,03.10 ⁻⁵
AR(3)	3,96.10 ⁻⁵
AR(4)	4,05.10 ⁻⁵
AR(5)	4,09.10 ⁻⁵
AR(6)	4,12.10 ⁻⁵
AR(7)	3,69.10 ⁻⁵
AR(8)	3,86.10 ⁻⁵
AR(9)	3,95.10 ⁻⁵
AR(10)	4,23.10 ⁻⁵
AR(11)	4,73.10 ⁻⁵

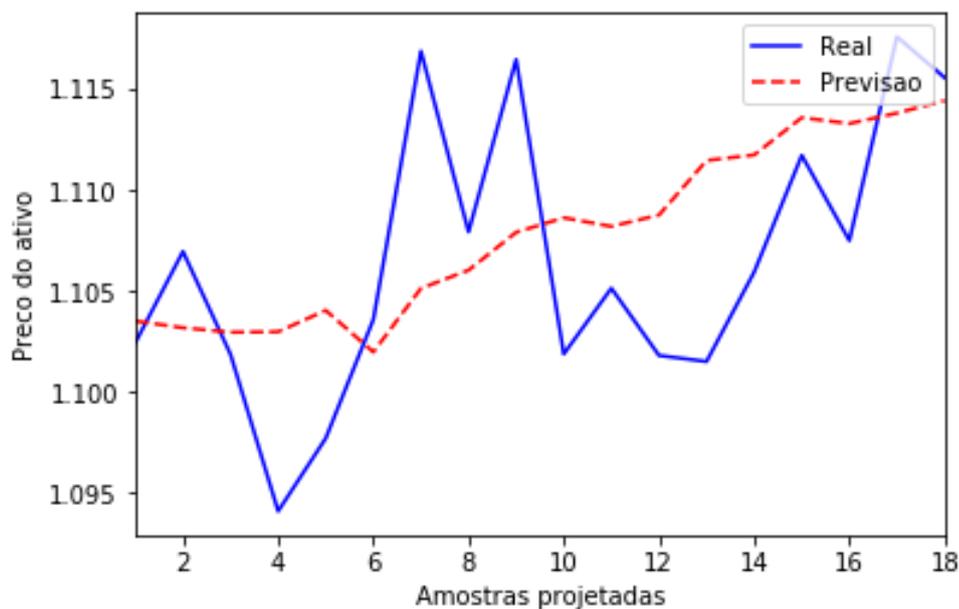
Foi analisado até o modelo AR(11), como visto na Tabela 5.

Comparando os resultados, o melhor modelo auto-regressivo para estes dados é o AR(7). Apesar de ser bem mais complexo que o AR(2), nota-se na Tabela 6 que há muito

mais parâmetros. Na Figura 22 observa-se que a previsão do modelo $AR(7)$ é levemente melhor.

Tabela 6 – Parâmetros $AR(7)$

Modelo	Parâmetros ϕ
AR(7)	$\phi_0 = 0.04509743$
	$\phi_1 = 0.9052481$
	$\phi_2 = -0.03000568$
	$\phi_3 = -0.02205525$
	$\phi_4 = 0.22948854$
	$\phi_5 = -0.11123098$
	$\phi_6 = 0.08214673$
	$\phi_7 = -0.09352934$

Figura 22 – Previsão do modelo $AR(7)$. Fonte: Autor

5.2.2 Modelos ARMA

Já feito o estudo em cima dos modelos auto-regressivos, será feito a identificação da ordem q , referente as médias móveis. Será feito uma análise da ACF com 15 atrasos no gráfico nas amostras de treino.

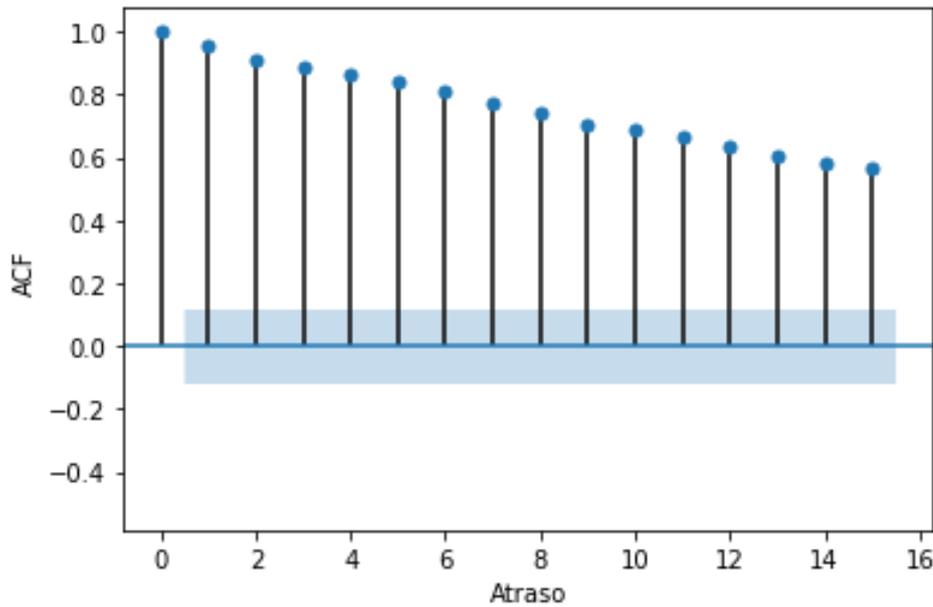


Figura 23 – Gráfico de autocorrelação. Fonte: Autor

Neste caso pode-se observar na Figura 23 que não há nenhum decaimento significativo de ACF cruzando os limites para indicar a ordem de q . Serão então analisados diferentes modelos ARMA, onde a ordem dos termos auto-regressivos será predefinida como $p = 7$ e a ordem dos termos de médias móveis será variada. Assim como o estudo nos modelos AR será utilizado como critério a variação do MSE.

Utilizando dos dados obtidos pelo modelo AR(7), se obtém a matriz de erros, podendo então aplicar o método dos mínimos quadrado estendido para ajustar os parâmetros de média móveis.

Tabela 7 – MSE de modelos ARMA(p,q) na base de seleção

Modelo	MSE
ARMA(7,1)	$3,89 \cdot 10^{-5}$
ARMA(7,2)	$3,74 \cdot 10^{-5}$
ARMA(7,3)	$7,06 \cdot 10^{-5}$
ARMA(7,4)	$7,25 \cdot 10^{-5}$
ARMA(7,5)	$8,31 \cdot 10^{-5}$

Foi analisado até o modelo $q = 5$, como visto na Tabela 7, após o modelo $q = 1$ o MSE só aumentou.

Comparando os resultados, o melhor modelo auto-regressivo de médias móveis para estes dados é o $ARMA(7,2)$. Os parâmetros auto-regressivos ϕ e de médias móveis θ

podem ser visualizados na Tabela 8. Pode-se visualizar a previsão deste modelo na figura 24.

Tabela 8 – Parâmetros ARMA(7,2)

Modelo	Parâmetros θ	Parâmetros ϕ
ARMA(7,2)	$\theta_1 = 0.06982658$ $\theta_2 = -0.37003034$	$\phi_0 = 1.13024315$ $\phi_1 = 0.83800528$ $\phi_2 = 0.39864222$ $\phi_3 = -0.3590107$ $\phi_4 = 0.23201621$ $\phi_5 = -0.07375427$ $\phi_6 = -0.00928413$ $\phi_7 = -0.05757658$

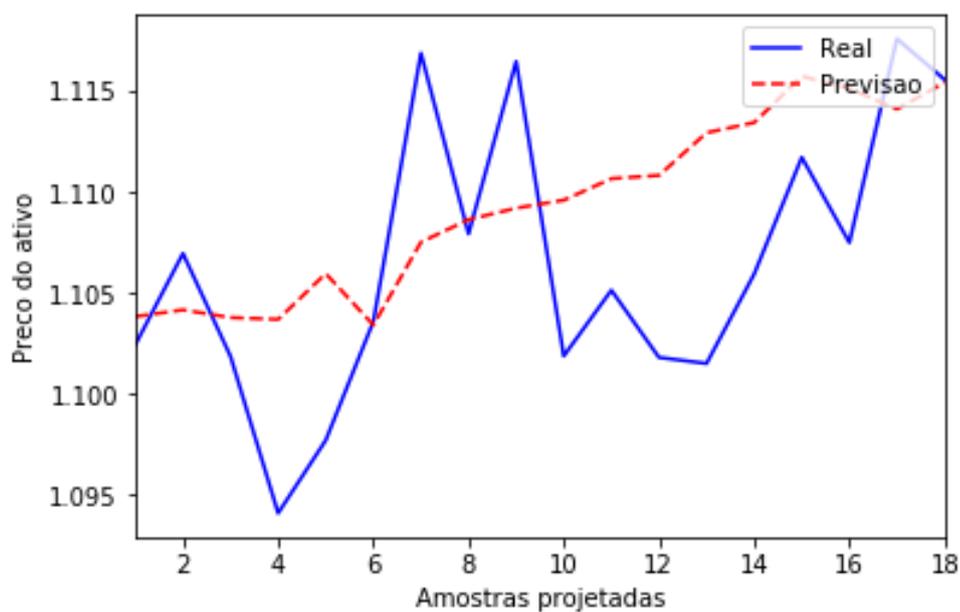


Figura 24 – Previsão do modelo ARMA(7,2). Fonte: Autor

Foi analisado modelos ARMA em que sua ordem auto-regressiva é menor, por exemplo o modelo $ARMA(2, 1)$, para verificar se a parte de médias móveis gerava uma maior influência e possibilitando melhores previsões, entretanto não se obteve bons resultados e a previsão deste modelo é ainda pior do que a previsão do modelo $AR(2)$, como pode ser visto na Figura 25.

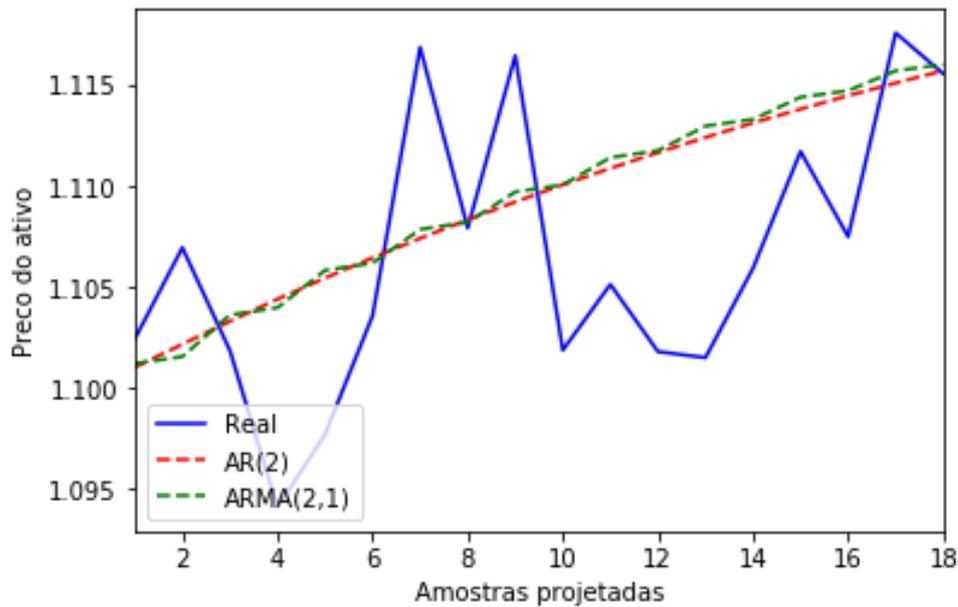


Figura 25 – Comparação ARMA(2,1) e AR(2). Fonte: Autor

5.2.3 Modelos ARIMA

Ao analisar os dados apresentados pela Figura 18, nota-se que não há estacionariedade, por tanto, pode-se aplicar a diferença pela Equação (3.9) nos valores da série X_t com o objetivo de torna-la estacionária.

A série se torna estacionária após a primeira diferença, mostrando que é integrada com ordem $d = 1$. Cria-se então a série temporal \check{X}_t , como pode ser visto na Figura 26

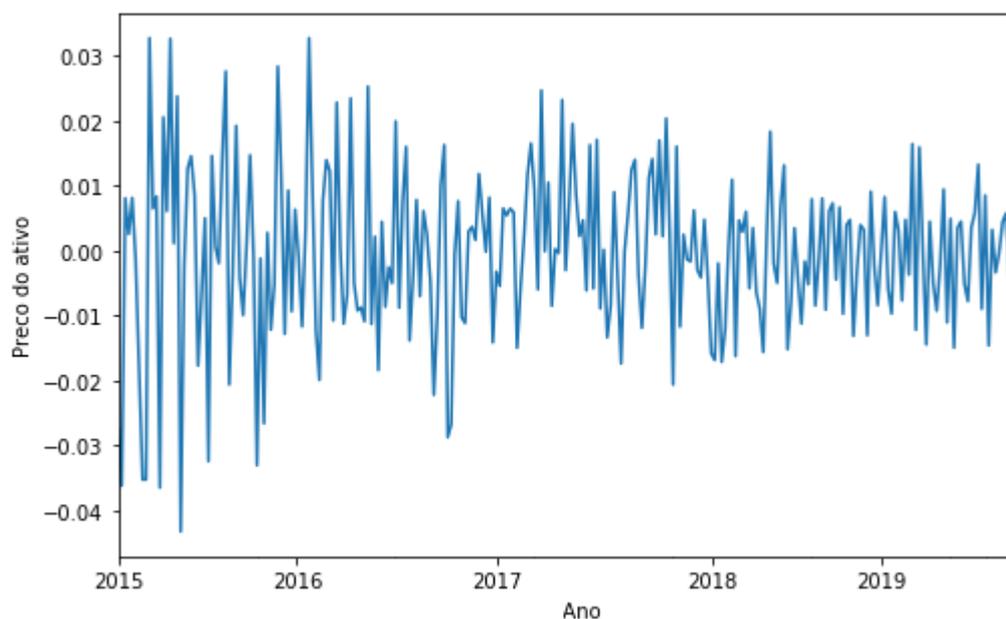


Figura 26 – Série gerada da primeira diferença. Fonte: Autor

Com estes novos dados, será feito novamente a identificação dos termos autorregressivos e de médias móveis.

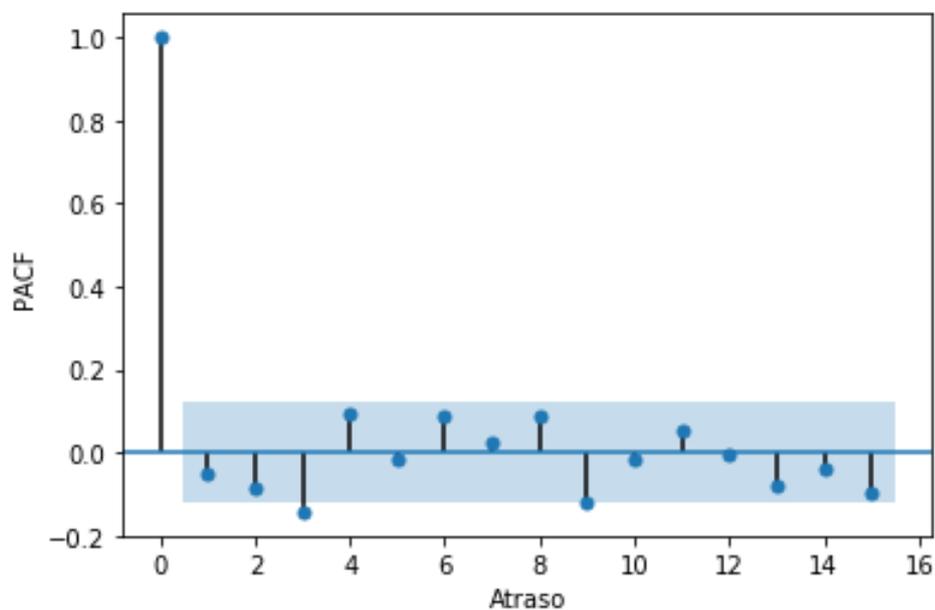


Figura 27 – Autocorrelação parcial da série diferenciada. Fonte: Autor

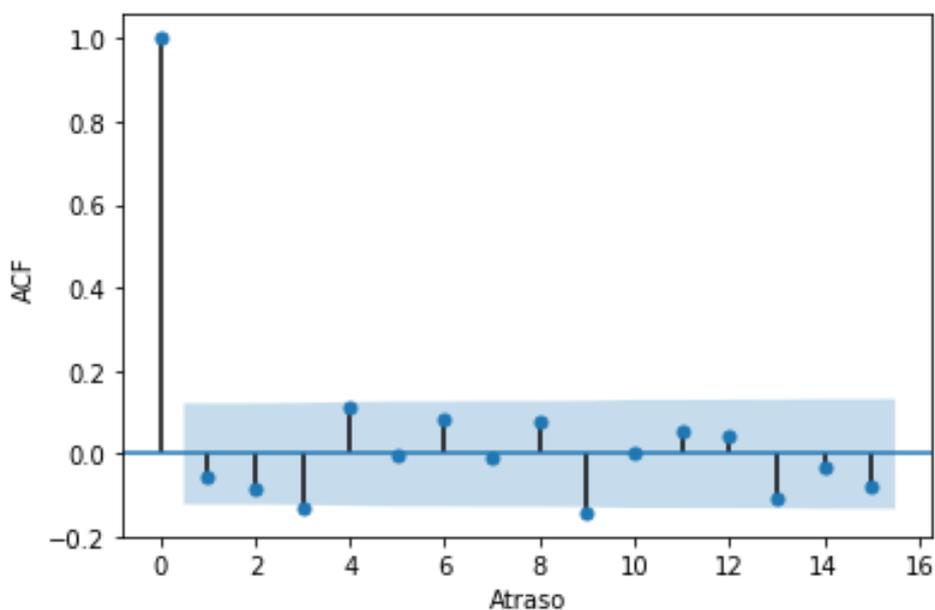


Figura 28 – Autocorrelação da série diferenciada. Fonte: Autor

Analisando a PACF e ACF nas Figuras 27 e 28, respectivamente, se observa que ambas têm rápido declínio logo depois do primeiro atraso. Seguindo a metodologia de

identificação, obtemos que $p = q = 1$ são uma boa escolha para representar estes dados, sendo assim, o modelo identificado é o $ARIMA(1, 1, 1)$.

$$\dot{X}_t = \phi_0 + \phi_1 \dot{X}_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon \quad (5.2)$$

Ajustando os parâmetros do modelo $ARIMA(1, 1, 1)$ pelo método dos mínimos quadrados estendidos, se obtém os parâmetros da Tabela 9.

Tabela 9 – Parâmetros $ARIMA(1,1,1)$

Modelo	Parâmetros ϕ	Parâmetros θ
$ARIMA(1,1,1)$	$\phi_0 = -1.14567679e - 04$ $\phi_1 = -6.9358050e - 01$	$\theta_1 = 3.07865487e - 01$

Com este modelo se atinge um $MSE = 6, 21 \cdot 10^{-5}$, um valor acima dos modelos ARMA e AR, indicando uma péssima previsão, como pode ser visto na Figura 29

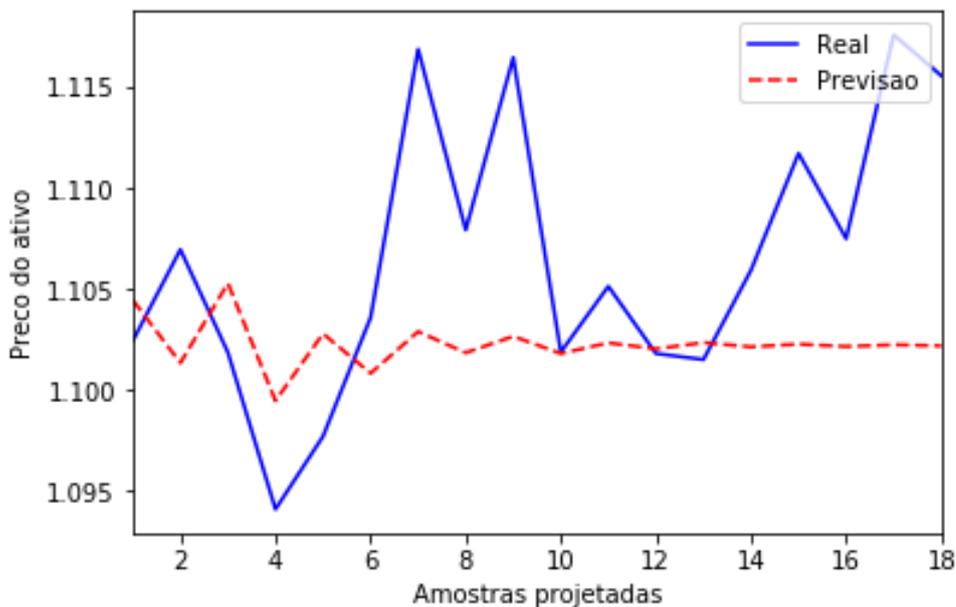


Figura 29 – Previsão do modelo $ARIMA(1,1,1)$. Fonte: Autor

Assim como os testes no modelo AR, pode-se aumentar o grau do modelo em busca de melhores previsões, entretanto não parece ser o que acontece. Ao criar um modelo $ARIMA(7, 2, 1)$, aumentando consideravelmente a complexidade da parte auto-regressiva, ainda se mantém as péssimas previsões, como pode ser observado na Figura 30.

Ao se utilizar a diferença dos dados como base para treinar os modelos, acaba-se por alterar completamente a correlação, esse comportamento fica muito claro ao comparar os gráficos de autocorrelação da série original e série diferenciada, figura 23 e

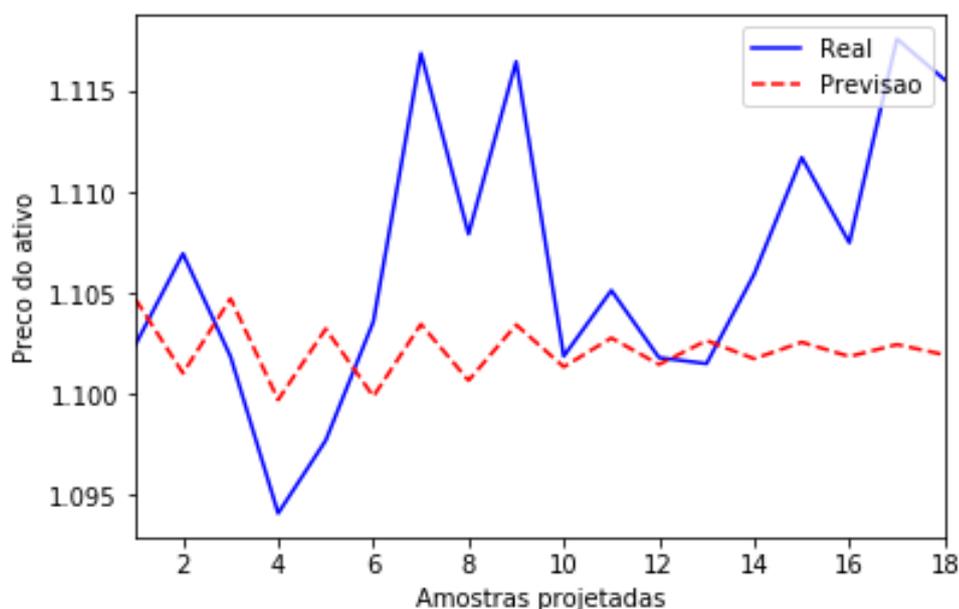


Figura 30 – Previsão do modelo $ARIMA(7,2,1)$. Fonte: Autor

28 respectivamente, onde na série original temos uma maior relação entre o valor atual e valores com maior atraso, essa diminuição de correlação entre as variáveis nos faz extrair menos informações a cada atraso, o que acaba por gerar piores previsões a atrasos maiores.

Considerando esse problema na correlação dos dados com atrasos com períodos distantes, uma boa solução é utilizar de identificação online nos modelos ARIMA. Uma das formas de fazer a identificação online é considerar que será feita projeção somente para uma amostra a frente, a partir do momento que se receber o valor real da amostra, se recalcula o modelo e cria-se a previsão da próxima amostra.

Ao aplicar a identificação online no modelo $ARIMA(1, 1, 1)$, consegue-se gerar um $MSE = 2,91 \cdot 10^{-5}$. Vemos na Figura 31 que a previsão melhora, sendo que agora o valor real anterior a cada amostra tem um forte impacto na previsão.

5.3 Análise entre classes dos modelos

Identificado o melhor modelo de cada classe para nosso conjunto de dados, pode-se comparar suas previsões pela Figura 32 e MSE pela Tabela 10.

Nota-se que as previsões dos modelos AR e ARMA seguem a tendência a um alto número de amostras projetadas, mas dificilmente acertam o valor real. A previsão do modelo ARIMA teve péssimos resultados, não conseguindo passar nenhuma informação relevante, já na sua forma online há melhores resultados. A previsão gerada pelo modelo online é quase que totalmente influenciada pela amostra mais recente obtida, o que reflete em uma previsão idêntica a própria série temporal, só que atrasada em uma amostra.

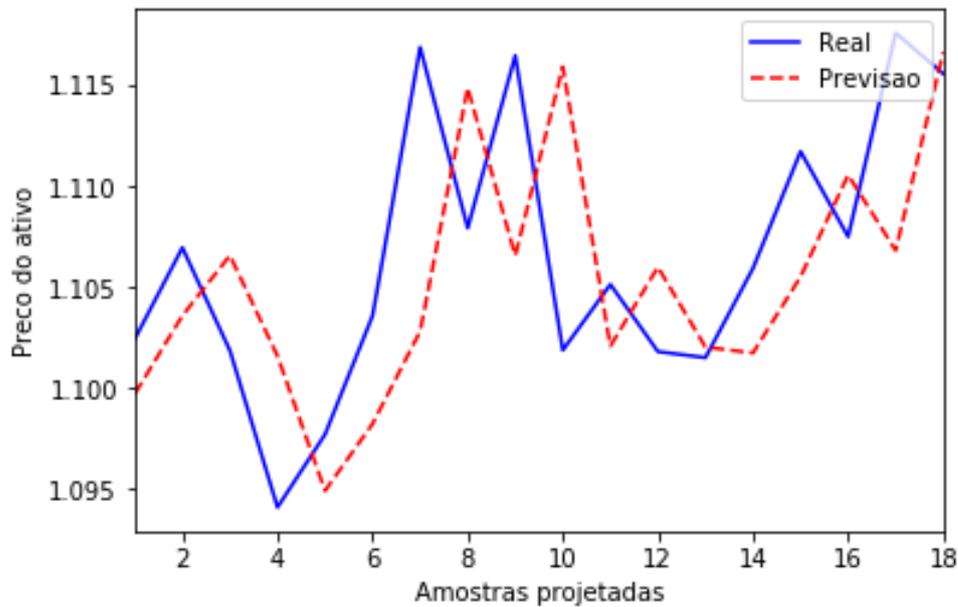


Figura 31 – Previsão do modelo $ARIMA(1,1,1)$ com identificação online. Fonte: Autor

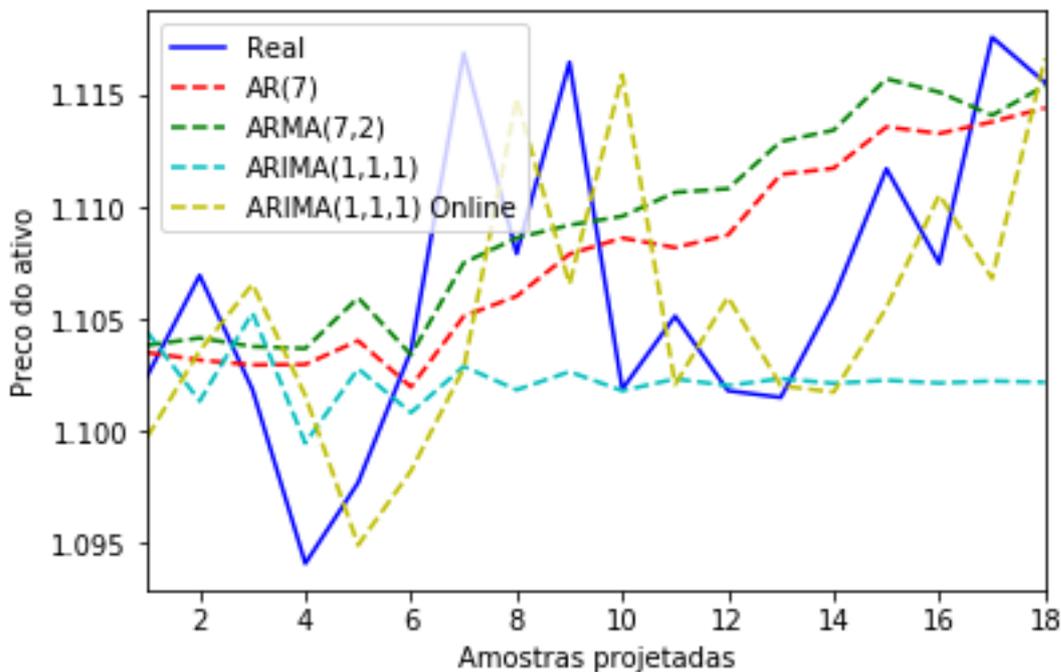


Figura 32 – Previsão dos modelos. Fonte: Autor

A previsão com o menor MSE foi a gerada pelo modelo $ARIMA(1,1,1)$ com identificação online, mesmo sendo um modelo com poucos parâmetros, isso se deve principalmente pela identificação online, que utiliza de amostras recentes para criar seu modelo e deixa a previsão destes sempre com valores próximos dos reais. Entretanto, a identificação online só teve esse grande impacto por a série temporal estudada não varia muito a cada amostra.

Tabela 10 – MSE na base de teste

Modelo	MSE
AR(7)	$3,43.10^{-5}$
ARMA(7,2)	$3,89.10^{-5}$
ARIMA(1,1,1)	$7,65.10^{-5}$
ARIMA(1,1,1) Online	$3,04.10^{-5}$

Analisando a facilidade de identificação dos modelos, deve-se dar destaque para a classe de modelos AR, pois nestes modelos a parametrização é simples e suas previsões são muito próximas dos modelos mais complexos.

6 Considerações finais

Este estudo aplicou a metodologia de Box e Jenkins para determinação de modelos matemáticos de séries temporais do mercado financeiro, através da análise de autocorrelação, estacionariedade e MSE foi possível identificar os modelos de cada classe com as melhores previsões.

Os objetivos foram cumpridos, para cada modelo identificado e parametrizado, foi possível gerar previsões e análises. Foi encontrado o melhor modelo para os dados do problema, de acordo com a métrica especificada, o modelo $ARIMA(1, 1, 1)$ com identificação *online*. Este modelo obteve a melhor previsão e o menor valor do MSE, em sua identificação pode-se também analisar a grande melhoria que a identificação *online* traz para as previsões.

As previsões geradas pelos modelos obtiveram erros significativos, sendo que, dificilmente era predito o valor correto. O alto erro se deve principalmente a origem dos dados ser do mercado financeiro, onde há extrema aleatoriedade, para contornar esse problema se procurou ao máximo diminuir os erros das previsões através de uma boa parametrização. O método aplicado possibilitou otimizar cada modelo proposto, entretanto nota-se que na etapa de identificação da ordem dos modelos, o método de identificação por ACF e PACF priorizou modelos de baixa ordem, como o AR(2) nos modelos AR, o que nem sempre gera o modelo com melhores previsões.

O maior desafio enfrentado durante a realização do trabalho foi o aprofundamento em assuntos na área de estatística. As literaturas que explicam ACF, PACF e Processos estocásticos apresentam um alto nível de complexidade matemático, onde que para o entendimento destes é necessário aprender diversos novos operadores e teorias probabilísticas.

Devido às simplificações feitas nos dados e o ativo *EURUSD* ser muito volátil, pode-se buscar em trabalhos futuros gerar previsões, com erros de menor magnitude, ao se identificar modelos a partir de outros dados do mercado financeiro, por exemplo: Dados de ativos menos voláteis, períodos diferentes e informações além do preço, como o volume de negociações. Considera-se também que foram utilizados simples algoritmos de aprendizado de máquina focados em relações lineares, há uma vasta gama de algoritmos que podem continuar este estudo, sendo os principais:

1. Modelos vetoriais (VAR, VARMA): Modelos para séries temporais multivariadas, onde pode-se utilizar mais dados obtidos do mercado financeiro, como o máximo e mínimo de cada período, para gerar previsões.

2. Modelos de volatilidade estocástica (ARCH, GARCH): Estes modelos fornecem uma maneira de modelar séries com volatilidade variante com o tempo, como é o caso de séries temporais financeiras.
3. Redes Neurais: Algoritmo de otimização para encontrar um conjunto de pesos para mapear a relação de entradas com saída, inspirados pelo sistema nervoso central de um animal são bons para o reconhecimento de padrões, a dificuldade deste modelo está em não poder calcular diretamente o conjunto ideal de pesos, nem obter garantias de convergência global.

Referências

- 1 BOLSA PT- GRÁFICO PETR4. Disponível em: <<http://www.bolsapt.com/resumo/PETR4.SA/10-anos/>>. Citado 2 vezes nas páginas 13 e 21.
- 2 COELHO, A. A. R. *Identificação de Sistemas Dinâmicos Lineares*. [S.l.]: Editora UFSC, 2004. Citado 3 vezes nas páginas 13, 27 e 28.
- 3 SYSTEM-IDENTIFIKATION. Disponível em: <<https://www.slideshare.net/MohamedOubbati/system-identifikation>>. Citado 2 vezes nas páginas 13 e 28.
- 4 RESUMO WOOLDRIDGE - INTRODUÇÃO ÀS SÉRIES TEMPORAIS. Disponível em: <<https://edisciplinas.usp.br>>. Citado 2 vezes nas páginas 13 e 29.
- 5 ANÁLISE DE SÉRIES TEMPORAIS. Disponível em: <https://www.maxwell.vrac.puc-rio.br/4244/4244_5.PDF>. Citado 2 vezes nas páginas 13 e 30.
- 6 SAZONALIDADE. Disponível em: <<http://www.portalaction.com.br/series-temporais/23-sazonalidade>>. Citado 2 vezes nas páginas 13 e 30.
- 7 ANOMALY DETECTION IN TIME SERIES DATA. Disponível em: <https://www.renom.jp/notebooks/tutorial/time_series/lstm-anomalydetection/notebook.html>. Citado 2 vezes nas páginas 13 e 31.
- 8 MORETTIN PEDRO A E TOLOI, C. *Análises de séries temporais*. 2º. ed. [S.l.]: Blucher, 2006. Citado 6 vezes nas páginas 13, 32, 33, 34, 36 e 39.
- 9 CARACTERÍSTICAS DAS SÉRIES TEMPORAIS. Disponível em: <<https://www.abgconsultoria.com.br/blog/caracteristicas-das-series-temporais/>>. Citado 2 vezes nas páginas 13 e 34.
- 10 MODELAGEM ARIMA - ESTATÍSTICA UFMG. Disponível em: <<ftp://est.ufmg.br/pub/glaura/SeriesTemporais/ARIMA/modelosARIMA.pdf>>. Citado 2 vezes nas páginas 13 e 35.
- 11 WHITE NOISE. Disponível em: <https://en.wikipedia.org/wiki/White_noise>. Citado 2 vezes nas páginas 13 e 36.
- 12 HOW DOES ACF AND PACF IDENTIFY THE ORDER OF MA AND AR TERMS. Disponível em: <<https://stats.stackexchange.com/questions/281666/how-does-acf-pacf-identify-the-order-of-ma-and-ar-terms>>. Citado 2 vezes nas páginas 13 e 40.
- 13 ESTIMATE ARMA COEFFICIENTS THROUGH ACF AND PACF INSPECTION. Disponível em: <<https://stats.stackexchange.com/questions/106038/estimate-arma-coefficients-through-acf-and-pacf-inspection>>. Citado 2 vezes nas páginas 13 e 41.
- 14 MOVING AVERAGE MODELS (MA MODELS). Disponível em: <<https://online.stat.psu.edu/stat510/lesson/2/2.1>>. Citado 2 vezes nas páginas 13 e 42.

-
- 15 METATRADER5. Disponível em: <<https://www.metatrader5.com/pt/download>>. Citado 3 vezes nas páginas 15, 23 e 45.
- 16 BOVESPA TEM MAIOR NEGOCIAÇÃO DIÁRIA. Disponível em: <<https://www.infomoney.com.br/onde-investir/bovespa-tem-maior-negociacao-diaria-em-2018-e-sobe-29-em-relacao-a-2017/>>. Citado na página 22.
- 17 TSAY, R. S. *Analysis of Financial Time Series*. 3º. ed. [S.l.]: Wiley, 2010. Citado 3 vezes nas páginas 40, 41 e 42.