

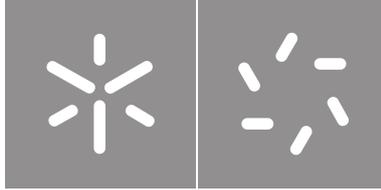
**Universidade do Minho**  
Escola de Ciências

José Lima  
**Análise e Previsão de Séries Temporais - uma  
abordagem com modelos de espaço de estados**

José Francisco Marques de Freitas Lima

**Análise e Previsão de Séries Temporais -  
uma abordagem com modelos de espaço de  
estados**





**Universidade do Minho**

Escola de Ciências

José Francisco Marques de Freitas Lima

**Análise e Previsão de Séries Temporais -  
uma abordagem com modelos de espaço  
de estados**

Dissertação de Mestrado

Mestrado em Estatística para Ciência de Dados

Trabalho efetuado sob a orientação de:

**Professora Doutora Arminda Manuela Andrade  
Pereira Gonçalves**

**Professor Doutor Marco André da Silva Costa**

## **DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS**

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

### ***Licença concedida aos utilizadores deste trabalho***



**Atribuição-NãoComercial-SemDerivações  
CC BY-NC-ND**

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

## **Agradecimentos**

Aos meus orientadores, Professora Doutora Arminda Manuela Gonçalves e Professor Doutor Marco Costa, pela generosa partilha de conhecimento, orientação, constante disponibilidade para ajudar nesta dissertação e pela amizade que construímos.

Aos meus pais, pelo apoio incondicional, motivação e inestimáveis conselhos que me proporcionaram ao longo desta jornada académica.

À minha colega, Fernanda Catarina Pereira, pela sua incansável disponibilidade e paciência em momentos de maior dificuldade.

Aos meus amigos que compartilharam comigo esta etapa final e estiveram ao meu lado em todos os momentos de trabalho árduo. Suas amizades tornaram essa jornada mais rica e memorável.

Quero também agradecer a todos os que de alguma forma contribuíram para o sucesso deste trabalho, mesmo que não estejam aqui mencionados. A todos, o meu sincero reconhecimento e gratidão.

## **DECLARAÇÃO DE INTEGRIDADE**

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

## Resumo

No início de cada mês ocorre uma conferência de imprensa sobre o estudo realizado pelo ISM, *Institute for Supply Management*. Nesta conferência de imprensa são analisadas as sondagens realizadas à produção industrial dos Estados Unidos da América (EUA).

Nas últimas décadas os índices do ISM ganharam relevância como um indicador para avaliar o desempenho e a evolução da produção nos EUA. Os investigadores e analistas económicos não utilizam os dados do ISM como uma fonte de dados para investigação sobre a produção, utilizam apenas dados oficiais fornecidos por instituições governamentais como, por exemplo, o Sistema de Reserva Federal dos Estados Unidos. Estes últimos são os mais economicamente aceites, porque a metodologia para a sua obtenção é considerada mais adequada para a produção de dados mais creíveis do que o índice do ISM. No entanto, o grande valor dos índices do ISM é a sua antecipação em relação aos dados oficiais, uma vez que são publicados previamente. Para o setor privado e as entidades públicas encarregues de propor e implementar políticas é fundamental obter valores próximos da realidade da produção duas ou até quatro semanas antes dos dados oficiais.

O principal objetivo desta dissertação é obter previsões a curto prazo de um índice do ISM através de modelos de previsão, uma vez que este índice tem um horizonte temporal de um mês. Foi proposto aplicar-se modelos lineares, modelos autorregressivos integrados e de médias móveis sazonais (SARIMA) e modelos de espaço de estados, por serem largamente adotados para modelar dados económicos. Enquanto a modelação através dos modelos lineares e dos modelos SARIMA está bem estabelecida na literatura, a modelação através dos modelos de espaço de estados implicou um estudo mais exaustivo e é proposta como um procedimento alternativo de modelação. Nestes últimos, a estimação dos parâmetros desconhecidos é usualmente realizada pela estimação de máxima verosimilhança, com o pressuposto da normalidade dos erros. Contudo, nesta dissertação, consideraram-se estimadores independentes da distribuição dos erros como alternativa aos de máxima verosimilhança, em que foi necessário estabelecer intervalos de confiança para os parâmetros através da metodologia *Bootstrap*, sendo uma contribuição inovadora desta dissertação.

Nesta dissertação foi utilizada a base de dados do ISM, a única disponibilizada por este instituto correspondendo a todos os *scores* mensais divulgados pelo ISM desde 1957, disponível em Bognanni & Young (2018) ou no *website* da instituição.

## Abstract

At the beginning of each month, a press conference is held to discuss the study conducted by ISM, the Institute for Supply Management. During this press conference, surveys conducted on US industrial production are analyzed.

Over the past few decades, ISM indices have gained significance as an indicator to assess the performance and trends in US production. Researchers and economic analysts do not use ISM data as a primary source for production research; they rely solely on official data provided by government institutions such as the Federal Reserve System of the United States. These data are more widely accepted because the methodology of their collection is considered more suitable for producing reliable data compared to ISM indices. However, the great value of ISM indices lies in their anticipation of official data since they are published in advance. For the private sector and public entities responsible for proposing and implementing policies, obtaining values close to the reality of production two to even four weeks before official data is crucial.

The main objective of this dissertation is to forecast an ISM index using short-term statistical models, as this index has a one-month time horizon. It was proposed to apply linear models, autoregressive integrated moving average (ARIMA) models, and state space models, as they are widely adopted for modeling economic data. While modeling through linear models and SARIMA models is well-established in the literature, modeling through state space models presented some challenges. In the latter, the estimation of unknown parameters is usually performed using maximum likelihood estimation, assuming the normality of errors. However, in this dissertation, estimators independent of the underlying distribution were considered as an alternative to maximum likelihood, involving the derivation of parameter confidence intervals using bootstrap methodology, which is an innovative contribution of this dissertation.

The ISM dataset was used in this dissertation, which is the only dataset provided by this institution, encompassing all monthly scores published by ISM since 1957, available at Bognanni & Young (2018) or on the institution's website.

# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Revisão da Literatura</b>	<b>3</b>
<b>3</b>	<b>Abordagem Metodológica</b>	<b>6</b>
3.1	Séries temporais . . . . .	6
3.1.1	Conceito . . . . .	6
3.1.2	Componentes de uma série temporal . . . . .	7
3.1.3	Processo estocástico estacionário . . . . .	9
3.1.4	Funções de autocovariância, autocorrelação e autocorrelação parcial . . . . .	11
3.1.5	Passeio aleatório . . . . .	13
3.1.6	Processos estocásticos não estacionários . . . . .	14
3.1.7	Modelos lineares para séries temporais . . . . .	15
3.2	Modelos de espaço de estados . . . . .	21
3.2.1	Introdução . . . . .	21
3.2.2	Modelos de espaço de estados (MEE) . . . . .	23
3.2.3	Modelos de espaço de estados lineares gaussianos . . . . .	24
3.3	Filtragem, alisamento e previsão de Kalman . . . . .	25
3.3.1	Introdução . . . . .	25
3.3.2	Filtro de Kalman . . . . .	26
3.3.3	Alisamento . . . . .	29
3.3.4	Previsão . . . . .	30
3.4	Estimação de parâmetros . . . . .	32
3.4.1	Estimação pela máxima verosimilhança . . . . .	32
3.4.2	Estimadores distribution-free . . . . .	35
3.5	<i>Bootstrapping</i> . . . . .	38
3.5.1	Conceito de Bootstrap . . . . .	39
3.5.2	Bootstrapping para MEE . . . . .	39
3.6	Avaliação dos modelos . . . . .	41
3.6.1	Medidas de avaliação . . . . .	41
3.6.2	Critérios de seleção . . . . .	45

3.6.3	Análise dos resíduos . . . . .	46
<b>4</b>	<b>Análise e Modelação de Dados Económicos</b>	<b>48</b>
4.1	Dados económicos . . . . .	48
4.2	Análise exploratória dos dados . . . . .	48
4.3	Modelação . . . . .	51
4.3.1	Modelos de regressão linear simples . . . . .	51
4.3.2	Modelos SARIMA - metodologia Box-Jenkins . . . . .	55
4.3.3	Modelos de espaço de estados . . . . .	61
4.3.4	Previsões . . . . .	70
<b>5</b>	<b>Conclusão</b>	<b>75</b>
5.1	Trabalho futuro . . . . .	76

## Índice de Figuras

Simulação de um ruído branco . . . . .	11
Simulação de duas trajetórias do passeio aleatório . . . . .	14
Série Temporal do índice ISM de 197 a 2022 . . . . .	49
Decomposição da Série Temporal ISM . . . . .	50
Histograma e diagrama em caixa de bigodes da Série Temporal do índice ISM. . . . .	51
Modelo RLS com Tendência . . . . .	52
Histograma, diagrama de caixa de bigodes, FAC e FACP do modelo RLS com Tendência . . . . .	53
FAC e FACP da série temporal do índice ISM. . . . .	56
FAC e FACP dos Modelos 1, 2 e 3 . . . . .	59
Histogramas, diagrama de caixa de bigodes e <i>QQplots</i> dos resíduos . . . . .	60
Histograma e <i>QQ-plot</i> das inovações padronizadas . . . . .	63
Diagrama de caixa de bigódes e gráfico das inovações padronizadas . . . . .	64
Previsões a 1 passo do índice do ISM no período de treino e os respectivos intervalos de confiança a 95%, $Y_{t t-1}$ – vermelho; $Y_t$ – preto. . . . .	65
Histograma, diagrama de caixa de bigodes e gráfico das estimativas <i>Bootstrap</i> para $\hat{\mu}$ . . . . .	67
Histograma, diagrama de caixa de bigodes e gráfico das estimativas <i>Bootstrap</i> para $\hat{\phi}$ . . . . .	67
Histograma, diagrama de caixa de bigodes e gráfico das estimativas <i>Bootstrap</i> para $\hat{\sigma}_\varepsilon^2$ . . . . .	68
Gráfico com as previsões para 17 meses do <i>Modelo2</i> e os respectivos intervalos de confiança . . . . .	71
Gráfico com as previsões para 17 meses do Modelo de espaço de estados ( <i>ModeloMV</i> ) com estimadores paramétricos e os respectivos intervalos de confiança . . . . .	71
Gráfico com as previsões para 17 meses do Modelo de espaço de estados ( <i>ModeloNP</i> )com estimadores não paramétricos e os respectivos intervalos de confiança . . . . .	72
Gráfico com as previsões para 17 meses dos 3 Modelos e série observada . . . . .	73

## Índice de Tabelas

Exemplo de valores por sub-indicador para efeito do cálculo do índice ISM . . . . .	2
Padrões teóricos das FAC e FACP do modelo $ARMA(p, q)$ . . . . .	20
Passo máximo da função de autocovariância para a estimação de $\phi$ escolhido em função da dimensão da amostra . . . . .	38
Estatísticas descritivas da série do índice ISM . . . . .	49
Sumário do Modelo de Regressão Linear com variável Tempo . . . . .	52
Sumário do modelo de regressão com coeficientes sazonais e variável tempo . . . . .	54
Estimativas dos parâmetros e respetivos erros padrão com critérios de avaliação do Modelo ARMA(2,0,0) para a série ISM. . . . .	57
Estimativas dos parâmetros e respetivos erros padrão com critérios de avaliação do Modelo SARIMA(2,1,0)(2,0,0)_[12] para a série ISM. . . . .	58
Estimativas dos parâmetros e respetivos erros padrão com critérios de avaliação do Modelo ARIMA(2,1,0) para a série ISM. . . . .	58
Estimativas pela máxima verosimilhança . . . . .	63
valores de prova dos testes estatísticos à normalidade e correlação à série das inovações padronizadas . . . . .	64
Medidas de Avaliação para $Y_{t t-1}$ . . . . .	65
Estimativas pela estimação não-paramétrica para os parâmetros do modelo de espaço de estados	66
Estatísticas descritivas das estimativas <i>Bootstrap</i> . . . . .	68
Teste da Normalidade das estimativas <i>Bootstrap</i> . . . . .	68
Intervalos de confiança e erros-padrão para a estimação paramétrica e estimação não paramétrica	69
Semi-amplitudes dos intervalos de confiança para os parâmetros obtidos pela máxima verosimilhança (MV) e pelos estimadores não-paramétricos com <i>Bootstrap</i> . . . . .	69
Erro quadrático médio e raiz do erro quadrático médio das previsões dos 3 modelos e do modelo de RLS. . . . .	72
Amplitude dos intervalos de confiança das previsões para os 3 modelos . . . . .	74

## Acrónimos

**AIC** – *Akaike's Information Criterion* (ou Critério de informação de Akaike)

**ARIMA** – *Modelo Autorregressivo Integrado e de Médias Móveis*

**ARMA** – *Modelo Autorregressivo e de Médias Móveis*

**BIC** – *Bayesian Information Criterion* (ou Critério de informação Bayesiano)

**EAM** – *Erro Absoluto Médio*

**EEAM** – *Erro Escalado Absoluto Médio*

**EM** – *Erro Médio*

**E-M** – *Expectation Maximization*

**EQM** – *Erro Quadrático Médio*

**EPAM** – *Erro Percentual Absoluto Médio*

**EPM** – *Erro Percentual Médio*

**FAC** – *Função de Autocorrelação*

**FACP** – *Função de Autocorrelação Parcial*

**FK** – *Filtro de Kalman*

**IoT** – *Internet of Things*

**ISM** – *Institute of Supply Management*

**ISO** – *International Organization for Standardization*

**MEE** – *Modelos de Espaço de Estados*

**PMI** – *Purchasing Managers Index*

**RLS** – *Regressão Linear Simples*

**SARIMA** – *Modelo Autorregressivo Integrado e de Médias Móveis Sazonal*

# 1 Introdução

O índice do ISM escolhido para a dissertação foi *Manufacturing PMI* que é um indicador mensal económico dos Estados Unidos da América, construído através de sondagens realizadas a gestores de compras em mais de 300 empresas industriais. É um indicador fundamental para avaliar e monitorizar o desenvolvimento da economia americana. Este índice foi criado pelo "*Institute for Supply Management*", de onde deriva a designação ISM, sendo uma organização não governamental e sem fins lucrativos que se estabeleceu em 1915. Esta organização fornece relatórios sobre desenvolvimento, educação e investigação a indivíduos particulares bem como, também, a empresas ou instituições financeiras com o propósito de ajudar a criar valor e permitindo que estas tenham vantagens competitivas, pois esta informação dá suporte a muitos processos de tomada decisão na gestão.

O índice *Manufacturing PMI* permite analisar as mudanças dos níveis de produção entre meses. Os relatórios são lançados no primeiro dia útil de um determinado mês, sendo, assim, um dos primeiros indicadores económicos disponíveis para os gestores e investidores. Este é composto por 5 outros indicadores com o mesmo peso Bognanni & Young (2018). Os indicadores são: Novas encomendas–refletindo o número de encomendas dos clientes realizadas às empresas; Produção–avalia se a produção das empresas varia ou não em relação a um determinado período anterior (dias, semanas e meses); Empregabilidade–mede a variação da empregabilidade, se esta aumentou ou diminuiu; Entregas–revela se o tempo das entregas entre o fornecedor e a empresa aumentou ou diminuiu em relação a um determinado período anterior; por fim, Inventários–indica quanto aumentaram ou diminuíram os inventários das empresas.

A monitorização do índice do ISM permite aos investidores, nos mercados norte americanos, tirarem conclusões significativas sobre a tendência e o estado da economia. Por exemplo, quando o índice aumenta normalmente, por regra, pode antecipar-se um aumento nos índices das Bolsas e verifica-se um declínio no mercado das Obrigações.

## A construção do índice Manufacturing PMI

Baseado no sistema de classificação NAICS (*North American Industry Classification System*), o índice *Manufacturing PMI* toma em consideração as indústrias com maior percentagem no PIB (Produto Interno Bruto) Americano. Considera-se na dissertação que *Manufacturing PMI* é o índice do ISM.

É realizada uma sondagem a múltiplas empresas em 17 setores industriais (indústria têxtil, química, eletrónica, entre outros). Nesta sondagem são consideradas perguntas sobre as áreas mencionadas,

existindo 3 possíveis respostas, aumentou, diminuiu ou manteve-se (não sofreu alterações). Todos os indicadores utilizados no cálculo variam entre 0 e 100, sendo:

< 50%

indica que o volume de negócios da indústria de produção contraiu em relação ao mês anterior;

> 50%

indica que o volume de negócios da indústria de produção cresceu em relação ao mês anterior;

= 50%

indica que o volume de negócios da indústria de produção manteve em relação ao mês anterior. As repostas são analisadas, sendo contabilizados os números de respostas positivas, negativas ou neutras (= 50%) para cada sub-indicador, e atribuem-se as percentagens conforme as respostas. A pontuação será atribuída através da seguinte fórmula:

$$\%RespostasPositivas + (1/2) \times \%RespostasNeutras. \quad (1.1)$$

Para efeitos do cálculo do índice, todos os sub-indicadores têm o mesmo peso no cálculo, considerando, portanto, a média aritmética dos sub-indicadores.

Tabela 1: Exemplo de valores por sub-indicador para efeito do cálculo do índice ISM

	% melhorou	% mantem	% piorou
Novas ordens	30,00	50,00	20,00
Produção	10,00	60,00	30,00
Empregabilidade	20,00	40,00	30,00
Entregas	20,00	40,00	10,00
Inventários	10,00	60,00	20,00

O valor de 0,59 ou 59% (valor superior a 50), dá uma indicação de que o, setor industrial americano expandiu, por exemplo, 9 pontos percentuais face ao mês anterior, caso no mês anterior o índice tenha sido exatamente igual a 50%.

## 2 Estado de arte

Neste trabalho pretende-se realizar análises e obter previsões de séries temporais utilizando como base os modelos lineares, os modelos SARIMA, e os modelos de espaço de estados. A modelação de uma série temporal implica a identificação e o ajustamento de um modelo aos dados e, por regra, estimar os respetivos parâmetros. No entanto, estes são desconhecidos e obtidos através de métodos de estimação e dos seus intervalos de confiança. Os tais parâmetros dos modelos de espaço de estados são estimados com o método da máxima verosimilhança, tendo sempre em consideração que os erros de observações e dos estados seguem uma distribuição gaussiana. Métodos alternativos também são testados, estimadores de *distribution-free* são desenvolvidos, bem como o método de *Bootstrap* e os erros quadráticos médios.

Proposto por Costa & Alpuim (2010), os estimadores independentes da distribuição de base são uma alternativa aos estimadores da máxima verosimilhança, contudo, não é conhecida a sua distribuição pelo que, neste trabalho utiliza-se o método *Bootstrap* para a obtenção de intervalos de confiança *Bootstrap* para as estimativas não paramétricas.

A série temporal do índice ISM, sendo relativa a dados económicos, apresenta uma correlação temporal e, eventualmente, componentes cíclica e sazonal. Em macroeconomia e econometria existem dois tipos de problemas: o primeiro é o facto de haver variáveis que não se observam, um exemplo seria o *output* da função NAIRU<sup>1</sup> e o segundo problema é que em muitos casos há coeficientes que são inerentemente variáveis indexadas no tempo, traduzindo-se em relações económicas instáveis (curva de *Philips* como exemplo).

Por estas razões os modelos de espaço de estados podem ser úteis e adequados para modelar este tipo de variável, sendo estes modelos bastantes flexíveis para incorporar componentes irregulares e não estacionárias. Há várias extensões deste modelo que permitem, também, analisar sistemas dinâmicos não gaussianos e não lineares. A principal vantagem de modelar séries temporais com modelos de espaço de estados é a sua facilidade de implementação, juntamente com a capacidade de representar de uma forma simples alguns fenómenos relativamente complexos.

Nas últimas décadas os modelos de espaço de estados têm sido cada vez mais utilizados em estudos de macroeconomia e na modelação de problemas financeiros. Estes modelos tiveram origem em Teoria de Sistemas com a sua primeira aplicação no programa *The Appolo and Polaris* Hutchinson (1984). A aplicação destes modelos começou com o intuito de rastrear a localização de objetos no espaço, tirando partido da equação de estado que permite definir equações de movimento para uma determinada posição

---

<sup>1</sup>NAIRU é *None-Accelerating Inflation Rate of Unemployment*, representa a taxa de desemprego que uma economia terá de ter para que os preços dos bens e serviços comecem a decrescer.

(ou estado) de uma aeronave espacial. As variáveis observáveis representam os dados observados por um dispositivo de rastreamento (por exemplo um *Global Positioning System*), como a velocidade e o azimute<sup>2</sup>.

Em 1960 Rudolf E. Kalman publicou o trabalho "*A new approach to linear filtering and prediction problem*". O filtro de Kalman é uma ferramenta muito aplicada na análise e ajustamento de modelos de espaço de estados. A sua eficiência resulta de requisitos computacionais acessíveis, propriedades recursivas bem desenvolvidas e extremamente eficientes, uma representação ótima de um estimador unidimensional para um sistema linear (assumindo que os erros são gaussianos). Permite também obter estimativas e previsões de variáveis não observadas (através das equações recursivas). Pode-se resumir o filtro de Kalman como sendo um algoritmo recursivo que estima estados não observados através de um processo observado.

A flexibilidade da modelação de séries temporais com dados em falta é uma característica que os modelos de espaços de estados possuem. Estes modelos permitem modelar séries temporais que foram observadas de forma irregular durante o tempo, sendo que, no âmbito de economia e macroeconomia, é algo que pode acontecer com certos indicadores e índices, um exemplo é Jones (1980). Shumway & Stoffer (1991) descrevem que com as modificações necessárias pode-se ajustar um modelo de espaço de estados multivariado através do algoritmo EM (*Expectation-Maximization*) com observações omissas.

Um dos métodos para estimar parâmetros desconhecidos dos modelos de espaço de estados é o método de máxima verosimilhança, este analisa todos os valores possíveis que os parâmetros podem assumir, de forma que se torne mais verosímil a ocorrência de uma amostra idêntica à que se observou. Por outras palavras as estimativas de máxima verosimilhança dos parâmetros desconhecidos, são os valores dos parâmetros que tornam máxima a probabilidade de ocorrência de uma amostra idêntica àquela que efetivamente ocorreu.

O *Bootstrapping* é uma abordagem para estimar distribuições de um determinado estimador ou realizar um teste estatístico. Através da reamostragem, o método *Bootstrap* tem como vantagem permitir obter distribuições em situações onde com as metodologias tradicionais não foi possível, substituindo o uso de cálculos analíticos pela eficácia e rapidez da computação. O método *Bootstrap* tem sido aplicado na análise de dados económicos porque tem-se mostrado uma técnica versátil que pode ser utilizada numa ampla gama de aplicações. Por exemplo, pode-se usar este método para estimar a distribuição de retornos de um determinado ativo financeiro ou para avaliar a incerteza associada a uma previsão de vendas no futuro. Em geral, o *Bootstrap* é útil quando há poucos dados disponíveis ou quando a distribuição subjacente dos dados é desconhecida.

---

<sup>2</sup>Azimute é um ângulo de uma dada direção com uma outra direção de referência.

O *Bootstrap* é adequado para analisar dados de séries temporais com um horizonte temporal curto (para o caso do índice do ISM será ideal pois é mensal). Esta metodologia também tem a vantagem de ser aplicável a casos não gaussianos (alguns exemplos são aplicados por Stoffer & Wall (1991)).

Em vários estudos realizados em econometria utiliza-se o modelo de Markov oculto e a autorregressão *switching*. Como já foi referido anteriormente os modelos de espaço de estados são caracterizados por um processo não observável, definido como processo de estado que se assume ser Markoviano, e um processo cujas variáveis são observáveis. A ideia base nestes modelos é que o valor do estado num determinado tempo vai especificar a distribuição da observação nesse mesmo tempo. Estes modelos foram desenvolvidos por Shumway & Stoffer (1991) e são utilizados para modelar séries temporais, em muitas áreas como no reconhecimento de fala e na análise de genomas. Na autorregressão com *switching* os modelos estatísticos para séries temporais permitem que a dinâmica da série mude ao longo do tempo. Isso é feito através de uma combinação de autorregressão (onde as observações futuras são modeladas como uma combinação linear de observações passadas) e modelos de mudança de estado (onde o modelo autorregressivo muda ao longo do tempo).

## 3 Abordagem metodológica

Para a concretização deste trabalho é necessário explorar diversos conceitos e temas relacionados com séries temporais. Na abordagem metodológica é explicado o que são séries temporais, os seus conceitos fundamentais e a sua aplicabilidade para desenvolver o trabalho. Também são apresentados os modelos de espaço de estados, os seus pressupostos e também os modelos estruturais. É realizada também uma introdução ao filtro de Kalman, ao alisamento e à previsão de Kalman. Este capítulo é finalizado com a apresentação da metodologia *Bootstrap* e, por fim, os métodos de estimação dos parâmetros de modelos de espaço de estados.

### 3.1 Séries temporais

#### 3.1.1 Conceito

Uma série temporal é um conjunto de observações medidas sequencialmente no tempo. As medições ocorrem continuamente no tempo ou em momentos específicos, podendo estes ser ou não igualmente espaçados (por exemplo dias, meses ou anos). Chamam-se a estes dois tipos de séries, séries temporais contínuas ou discretas, respetivamente, independentemente da natureza da variável medida, podendo ser esta também discreta ou contínua. Para além destas classificações uma série temporal pode ser univariada caso seja constituída por dados correspondentes a uma variável, ou multivariada quando apresente mais do que uma variável observada no tempo.

Usualmente, quando se analisam séries temporais em tempo contínuo, é comum transformá-las em séries temporais em tempo discreto, sendo que, para isso, elas são amostradas em intervalos de tempo iguais (como por exemplo, medir a temperatura em intervalos de uma hora). Caso o intervalo de amostragem seja pequeno o suficiente, esta técnica, por regra, não leva a uma perda significativa de informação.

No que diz respeito às séries temporais em tempo discreto, elas podem surgir de várias formas, além de poderem ser obtidas a partir de uma série temporal em tempo contínuo (como mencionado anteriormente), também podem surgir quando os valores de uma variável são agregados ao longo de um determinado período (por exemplo, o total de vendas ao longo de meses consecutivos) ou podem até mesmo ser inerentemente observadas em tempo discreto.

Os dados de uma série temporal caracterizam-se por serem observações sucessivas e, por regra, correlacionadas. É necessário ter em consideração a ordem pela qual as observações são recolhidas.

Particularmente nos casos de séries temporais univariadas, é possível considerar cada observação como bivariada, sendo que a segunda variável corresponde ao tempo no qual é observada.

Em muitas áreas, a previsão de séries temporais é imprescindível. Alguns casos práticos trata-se de previsão de vendas de uma empresa, etc. Essas previsões são usadas para planejar a estratégia de *marketing*, definir objetivos de vendas e prever a receita futura. Muitos investidores usam modelos de previsão de séries temporais para prever os preços futuros de ações, sendo que essas previsões são usadas para tomar decisões de investimento informadas e maximizar os retornos. As empresas da área de energia utilizam modelos de previsão de séries temporais para prever o consumo de energia no futuro, sendo estas previsões usadas para planejar a produção de energia e garantir que haja energia suficiente para responder à procura.

O propósito de analisar uma ou mais séries temporais depende de cada caso, há objetivos específicos como referido anteriormente. No entanto deve ser mencionado os principais propósitos Riani et al. (2023). Para analisar uma série temporal é comum descrever os dados usando estatísticas descritivas e/ou métodos gráficos. É importante compreender o mecanismo que gera a série, ou seja, procurar razões que justifiquem o comportamento da série, monitorizar a sua trajetória e identificar periodicidades relevantes nos dados. Uma vez que a série temporal tenha sido analisada, é necessário encontrar um modelo estatístico adequado para descrever a sua evolução. Existem vários modelos disponíveis, como os modelos SARIMA, os modelos de alisamento exponencial e os modelos de regressão múltipla, que podem ser usados para modelar o comportamento da série. Por fim, a previsão do comportamento futuro de uma série temporal pode ser extremamente útil na construção e execução de planos a curto, médio ou longo prazo, bem como no controlo de um determinado processo. As previsões podem ser de curto, médio ou longo prazo, dependendo do horizonte temporal em que são feitas. As previsões podem ser feitas para um único passo, ou seja, para o próximo instante, ou para vários passos, prevendo para vários instantes futuros. A previsão a múltiplos passos é mais complexa e imprecisa, pois a incerteza aumenta à medida que o horizonte temporal aumenta, no entanto, pode fornecer dados valiosos para o planeamento e para os processos de tomada de decisões.

### **3.1.2 Componentes de uma série temporal**

A variabilidade de uma série temporal pode ser decomposta em quatro componentes, sendo estas a tendência  $T$ , a componente sazonal  $S$ , a componente cíclica  $C$  e, por último, a componente irregular ou residual  $E$ .

A Tendência ( $T$ ) descreve a inclinação que uma série temporal apresenta ao longo do tempo, esta

pode ser linear ou não e crescente ou decrescente. A tendência pode ser consequência do facto dos valores observados dependerem de uma componente determinística que é função monótona do tempo, embora para muitos autores esta possa ser de natureza estocástica.

A componente da Sazonalidade (*S*) corresponde a um padrão de aumento e diminuição que ocorre regularmente na série em períodos específicos, originando oscilações que se repetem (e.g., as vendas a retalho apresentam frequentemente um padrão sazonal mensal, onde cada mês, considerando períodos anuais, mostra constantemente a mesma posição relativamente aos restantes: picos de vendas ocorrem em meses de férias e, nos meses seguintes, verifica-se uma diminuição acentuada). O movimento dentro de um período tem, então, duração fixa e é atribuído a fatores sazonais, i.e., relacionados, em regra, com aspetos do calendário (e.g., os meses ou trimestres de um ano ou os dias de uma semana). A sazonalidade pode ser classificada como aditiva, se não depende do nível da série (tendência), ou multiplicativa, quando seja proporcional ao mesmo. Esta fonte de variação é muitas vezes removida através de um procedimento designado de ajustamento sazonal, uma vez que esta pode omitir outras componentes relevantes, tais como a tendência.

Na componente Cíclica (*C*) há um padrão de flutuação que não apresenta periodicidade definida (i.e., a sua duração não é fixa e, portanto, o seu comprimento varia frequentemente de ciclo para ciclo) nem tem causa atribuída a fatores “sazonais”. Por outras palavras, os ciclos são quaisquer componentes não sazonais que apresentem um padrão reconhecível (por exemplo, ciclos económicos). Como estas flutuações são irregulares e se podem prolongar por vários anos, é frequente ignorar-se esta componente para séries “curtas”. Além disso, como se torna difícil dissociar a componente cíclica da tendência, estas são muitas vezes aglomeradas e referidas como componente de tendência cíclica.

A componente Irregular ou Aleatória (*E*) é a componente que contém qualquer variação não explicada pelas componentes anteriores e representa um ruído aleatório. Quando esta componente é modelada por um processo estocástico de variáveis aleatórias não correlacionadas e identicamente distribuídas, é referida como um ruído branco.

Grande parte dos métodos tradicionais para a análise de séries temporais baseia-se na decomposição da variação nas componentes anteriormente referidas. Na análise de uma série temporal, de modo geral, considera-se os seguintes modelos.

- 1. modelo aditivo

$$Y_t = T_t + S_t + C_t + E_t;$$

- 2. modelo multiplicativo

$$Y_t = T_t \times S_t \times C_t \times E_t;$$

- 3. modelo misto

$$Y_t = (T_t + S_t) \times C_t + E_t.$$

### 3.1.3 Processo estocástico estacionário

Os modelos determinísticos não são os mais apropriados para estudar fenômenos dinâmicos observados no mundo real. A teoria dos processos estocásticos tem como objetivo encontrar o modelo probabilístico que melhor descreva o comportamento de um determinado fenômeno, a fim de obter previsões precisas. Uma série temporal pode ser considerada como uma realização de um processo estocástico, em que o espaço de parâmetros é um conjunto discreto de índices que representam o tempo. O processo estocástico descreve a dinâmica do fenômeno subjacente à série temporal, que é afetado por uma variedade de fatores aleatórios e incertezas. A modelação estocástica pode permitir uma melhor compreensão do comportamento da série temporal e permitir a realização de previsões mais precisas.

Chama-se a  $T$  espaço de parâmetros e o contradomínio de  $Y_t$  por espaço de estados. Quando  $T = \mathbb{Z}$  ou  $T = \mathbb{N}$  o processo é de tempo discreto, no entanto quando  $T = \mathbb{R}$  ou  $T = \mathbb{R}^+$  diz-se que o processo é em tempo contínuo.

**Definição 1** Um processo estocástico é estacionário (ou fortemente estacionário) se e só se a distribuição conjunta de  $(Y(t_1), \dots, Y(t_n))$  é igual à distribuição conjunta de  $(Y(t_1 + \delta), \dots, Y(t_n + \delta))$  qualquer que seja o  $n$   $(t_1, \dots, t_n)$  e para qualquer  $\delta$ , ou seja,

$$F_{(Y(t_1), \dots, Y(t_n))}(y_1, \dots, y_n) = F_{(Y(t_1 + \delta), \dots, Y(t_n + \delta))}(y_1, \dots, y_n).$$

Um processo estocástico estritamente estacionário na prática é difícil de se verificar, uma vez que exige o conhecimento de todas as distribuições marginais, Cordeiro (2003). Uma definição alternativa de estacionaridade é definir a estacionaridade através da igualdade nos momentos e não na igualdade das distribuições. Um processo pode ser não estacionário por a média e/ou variância serem funções do tempo e não constantes. As transformações mencionadas permitem estabilizar a variância e só depois a média.

**Definição 2** Um processo  $Y_t, t \in T$  diz-se estacionário de 2ª ordem (ou fracamente estacionário) se e só se todos os momentos até à 2ª ordem de  $Y_{(t_1)}, \dots, Y_{(t_n)}$  existem e são iguais aos momentos correspondentes até à 2ª ordem de  $Y_{(t_1+h)}, \dots, Y_{(t_n+h)}$ . Logo, num processo fracamente estacionário têm-se:

- 1. o valor médio não depende de  $t$ , i.e.,  $E[Y_t] = \mu(t) = \mu$ ;
- 2. a variância não depende de  $t$ , i.e.,  $Var[Y_t] = \sigma^2(t) = \sigma^2$ ;
- 3. a covariância entre  $Y(t_1)$  e  $Y(t_2)$  depende apenas do desfasamento de  $t_1 - t_2$ , i.e.,  $Cov[Y(t_1); Y(t_2)] = \gamma(|t_1 - t_2|)$ .

Uma vez que a definição de um processo estritamente estacionário é muito exigente, de uma forma geral, apenas se considera a estacionaridade de 2ª ordem em muitos processos de modelação. Assim, sempre que se referir a um processo estacionário, assume-se que se refere à estacionaridade de 2ª ordem.

O processo de ruído branco é um exemplo de um processo estacionário de 2ª ordem. A maior importância deste processo é a sua representação na construção de outros processos estacionários relevantes na modelação de séries temporais. Um processo de ruído branco diz-se um processo puramente aleatório, formado por uma sucessão de variáveis aleatórias não correlacionadas entre si e identicamente distribuídas, com média e variância constantes que satisfaz as seguintes condições:

- 1.  $E[Y_t] = 0$ ;
- 2.  $var[Y_t] = \sigma^2$ ;
- 3.  $cov[Y_{t_1}; Y_{t_2}] = 0$  para todo o  $t_1 \neq t_2$ .

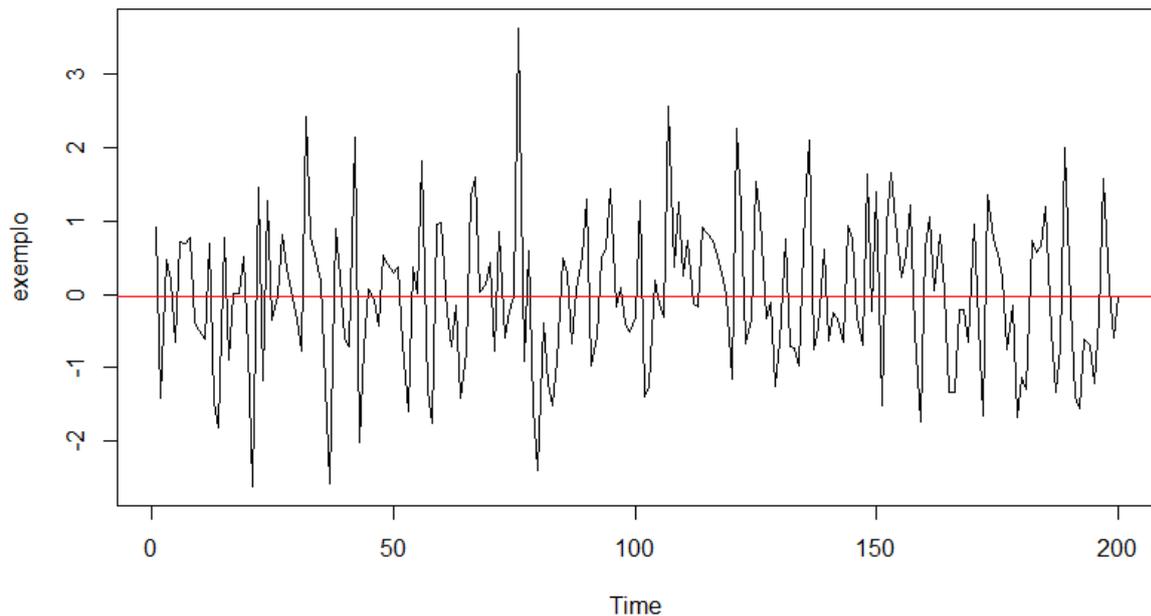


Figura 1: Simulação de um ruído branco

Na Figura 1 está representada uma trajetória de um processo de ruído branco Gaussiano com 200 observações, com média zero e variância unitária.

### 3.1.4 Funções de autocovariância, autocorrelação e autocorrelação parcial

**Definição 3** Para um processo estacionário, define-se a função de autocovariância como

$$\gamma_k = Cov[Y_t; Y_{t+k}] = E[(Y_t - \mu)(Y_{t+k} - \mu)]. \quad (3.2)$$

Esta função mede a intensidade com que as covariâncias acompanham pares de valores do processo separados por um intervalo (lag) de amplitude  $k$ .

A função de autocovariância ou  $\gamma_k$ , é definida para que  $k \in \mathbb{R}$  quando o processo é no tempo contínuo, e para  $k \in \mathbb{Z}$  para o tempo discreto, ou seja,  $k = 0, \pm 1, \pm 2, \dots$ . A função satisfaz as seguintes propriedades:

- 1.  $\gamma_0 = Cov[Y_t, Y_t] = Var[Y_t] = \sigma^2$ ;

- 2.  $\gamma_k = \gamma_{-k}$ , ou seja a função é par, logo dispensa-se a representação gráfica para  $k < 0$ ;
- 3.  $|\gamma_k| \leq \gamma_0$ , como consequência da desigualdade de Cauchy-Schawrz<sup>3</sup>;
- 4. A função  $\gamma_k$  é semidefinida positiva para qualquer conjunto de números reais  $\alpha_1, \dots, \alpha_n$  e instantes de tempo  $t_1, \dots, t_n$ , isto é,

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma(|t_i - t_j|) \geq 0.$$

**Definição 4** Num processo estacionário, a função de autocorrelação (ou FAC) é definida por

$$\gamma_k = \text{Corr}[Y_t, Y_{t+k}] = \frac{\text{Cov}[Y_t, Y_{t+k}]}{\sqrt{\text{Var}[Y_t] \text{Var}[Y_{t+k}]}} = \frac{\text{Cov}[Y_t, Y_{t+k}]}{\text{Var}[Y_t]} = \frac{\gamma_k}{\gamma_0},$$

mede a correlação entre pares de valores do processo separados por um intervalo (lag) de amplitude  $k$ .

Esta função dispõe das mesmas propriedades que a função de autocovariância, com a exceção da propriedade de que  $\rho_0 = 1$ . De modo geral, o aumento do valor de  $k$  implica um decréscimo de  $\rho_k$  e de  $\gamma_k$ . É esperado que a capacidade de retenção no momento  $t + k$  seja reduzida do que se passou no momento  $t$ , Murteira (1993). Logo é de esperar que a capacidade de memória do processo se vá desvanecendo ao longo do tempo. Assim, quando  $k \rightarrow +\infty$ , espera-se que  $\rho_k \rightarrow 0$ .

A função de autocorrelação parcial (ou FACP) tem por interesse quantificar a correlação existente entre duas observações com um desfasamento,  $k$ ,  $Y_t$  e  $Y_{t+k}$ , descartando-se a dependência das outras variáveis.

**Definição 5** O conjunto de autocorrelações parciais de desfasamento  $k$  é dado por  $\phi_{kk}, k = 1, 2, \dots$ , onde

$$\phi_{kk} = \text{Corr}(X_t, X_{t+k} | X_{t+1}, X_{t+2}, \dots, X_{t+k+1}) = \frac{|P_k^*|}{|P_k|},$$

$P_k^*$  é a matriz quadrada (ordem  $k$ ) de autocorrelação, onde a última coluna é substituída por  $[\rho_1, \rho_2, \dots, \rho_k]^T$ . A matriz de  $P_k$  é dada por

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & 1 \end{bmatrix}$$

<sup>3</sup>Desigualdade de Cauchy-Schawrz:  $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$

Seguindo a 5 obtém-se

- 1.  $\phi_{11} = \rho_1$ ;
- 2.  $\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$ ;
- 3.  $\phi_{33} = \frac{\rho_3(1 - \rho_1^2) + \rho_1(\rho_1^2 + \rho_2^2 - 2\rho_2)}{(1 - \rho_2)(1 + \rho_2 - 2\rho_1^2)}$ .

### 3.1.5 Passeio aleatório

Um exemplo clássico de um processo estocástico não estacionário é o passeio aleatório, ver exemplo na Figura 2. Uma partícula em movimento ocupa a posição inicial  $Y_0$ , é observada numa sucessão discreta de pontos,  $t = 1, 2, 3, \dots$ ; para cada momento observado  $t = k$ , essa partícula tem um salto com uma amplitude  $\epsilon_k$ . Descreve-se no momento  $t$  a posição da partícula que é explicada por

$$Y_t = Y_0 + \epsilon_1 + \epsilon_2 + \dots + \epsilon_t,$$

ou

$$Y_t = Y_{t-1} + \epsilon_t,$$

em que o  $\epsilon_t$ ,  $t = 1, 2, 3, \dots$  é uma sequência de variáveis aleatórias não correlacionadas. Ocorre um passeio aleatório na partícula quando  $E[\epsilon_t] = 0$ ,  $t = 1, 2, 3, \dots$ . Neste caso, tanto a média como a variância e covariância dependem do tempo passado, na data em que se inicia o passeio, Murteira (1993).

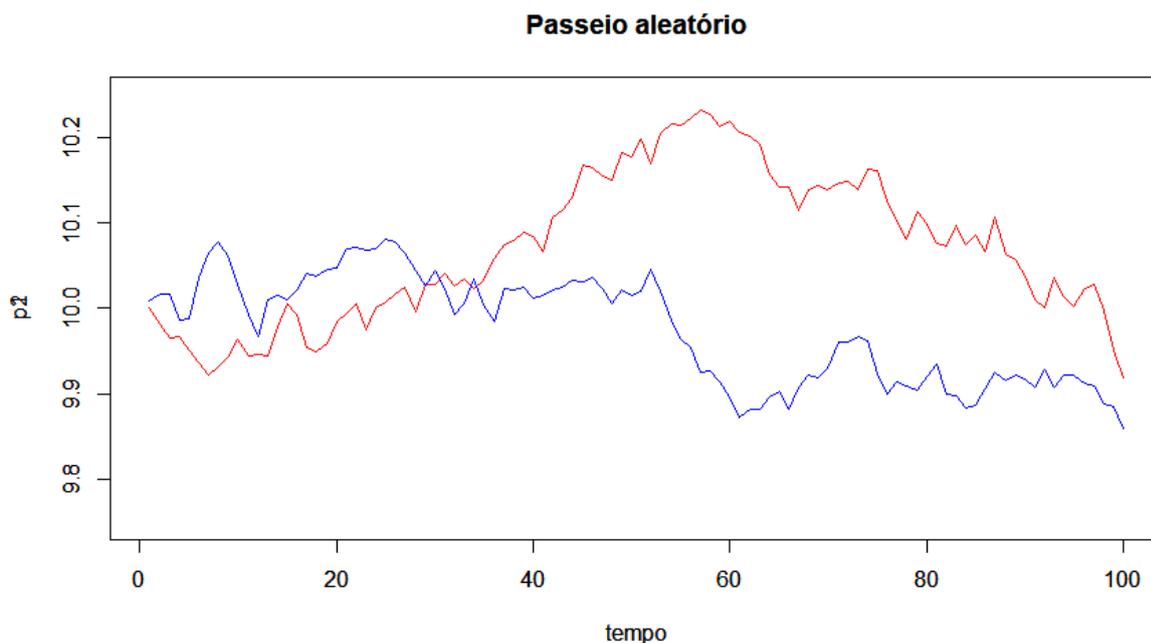


Figura 2: Simulação de duas trajetórias do passeio aleatório

### 3.1.6 Processos estocásticos não estacionários

Numa série estacionária, os valores futuros serão semelhantes aos do passado, tornando-se um pressuposto importante para prever com base em observações passadas. Alguns modelos de previsão de séries temporais assumem que a série é ou pode ser transformada numa série estacionária. No entanto, muitas séries temporais, especialmente as relacionadas com fenómenos ambientais ou económicos, são não estacionárias. Um processo pode ser não estacionário porque a média e/ou a variância são funções do tempo e não são constantes. É importante notar que uma série estacionária em média pode não ser estacionária em variância. Para ultrapassar este problema, é possível recorrer a transformações que estabilizam a média e/ou a variância, transformando uma série temporal não estacionária numa série estacionária. Se uma série temporal for não estacionária em média e variância, deve-se primeiro estabilizar a variância e só depois a média. Esta abordagem é defendida por autores como Murteira (1993).

Uma forma de estabilizar a média é através de um processo de diferenciação (regular), que consiste na aplicação do operador de diferença  $\nabla$ , e define-se  $\nabla Y_t = Y_t - Y_{t-1}$ , e é aplicado à série temporal não estacionária. Aplica-se uma diferenciação de primeira ordem através da seguinte equação,

$$\nabla Y_t = Y_t - Y_{t-1}, t = 2, 3, \dots, n \quad (3.3)$$

caso a diferenciação de ordem 1 não seja suficiente para obter uma série estacionária, pode-se obter as diferenças de ordem 2, que correspondem às diferenças das primeiras diferenças da série original,

$$\nabla^2 Y_t = \nabla(\nabla Y_t) = \nabla(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2}, t = 3, 4, \dots, n. \quad (3.4)$$

O operador de diferenciação de ordem  $d$ , para qualquer inteiro  $d \geq 1$ , consiste em diferenciar a série  $d$  vezes,

$$\nabla^d Y_t = \nabla(\nabla^{d-1} Y_t), t = d + 1, \dots, n. \quad (3.5)$$

No entanto, a diferenciação inapropriada de uma série já estacionária deve-se evitar uma vez que a variância aumenta com a ordem de diferenciação.

## Análise da estacionariedade

Há diversas formas de analisar a estacionariedade de uma série temporal, a mais elementar será representar uma série num gráfico ao longo do tempo. Esta análise baseia-se na subjetividade, pelo que é necessário efetuar determinados testes estatísticos para avaliar se uma série é estacionária. Existem diversos testes para realizar um estudo sobre a estacionariedade, a maioria baseia-se em encontrar uma raiz unitária (deve-se utilizar mais do que um teste para a raiz unitária). Os testes mais utilizados são os testes de Dickey-Fuller (ou DF), Dickey-Fuller aumentado (ou ADF), Philip-Perron (ou PP) e Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Dentro destes testes salientam-se dois tipos: aqueles cujo a hipótese a testar é a presença de uma raiz unitária (ou seja, não estacionariedade) e, no caso desta hipótese ser rejeitada, é fornecida a informação sobre o número de diferenciações necessárias para ser possível atingir a estacionariedade (DF, ADF e PP), para o caso da hipótese ser nula admite-se a estacionariedade.

### 3.1.7 Modelos lineares para séries temporais

Na análise de séries temporais é crucial selecionar um modelo adequado, uma vez que as previsões futuras dependem da estrutura probabilística subjacente à série temporal, que reflete o processo estocástico em questão. Os modelos de séries temporais podem ser classificados como lineares ou não lineares, dependendo se o valor atual da série depende de forma linear ou não das observações

passadas. A seleção do modelo apropriado pode ser feita com base na avaliação de diferentes critérios, como o ajuste aos dados, a capacidade de previsão, a simplicidade e a facilidade de interpretar o modelo.

## a) Processos autorregressivo de ordem $p$ – (AR( $p$ ))

**Definição 6** Um processo  $Y_t, t \in T$  diz-se ser um processo autorregressivo de ordem  $p$ ,  $AR_{(p)}$ , se satisfaz a equação

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t,$$

onde  $\epsilon_t$  é um processo de ruído branco e  $\phi_i, i = 1, \dots, p$  são constantes reais.

Através do operador de atraso, que se representa por  $B$ , podemos escrever a equação anterior da seguinte forma,

$$\Phi_p(B)Y_t = \epsilon_t$$

onde  $\epsilon_t$  é um ruído branco e  $\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ ,  $B^p$  designa-se por polinómio autorregressivo de ordem  $p$ , em que  $p$  representa o número de termos que definem a ordem do processo autorregressivo. Tendo em consideração as  $p$  raízes reais ou complexas,  $G_1^{-1}, G_2^{-1}, \dots, G_p^{-1}$ , da equação característica  $\Phi_p(B) = 0$ , torna-se possível fatorizar o polinómio autorregressivo, com a expressão

$$\Phi_p(B) = \prod_{i=1}^p (1 - G_i B).$$

**Condições de invertibilidade e estacionariedade:** Um processo estocástico é considerado invertível se puder ser representado por meio de uma equação autorregressiva,

$$\epsilon_t = Y_t + \sum_{i=1}^{\infty} \phi_i Y_{t-i}.$$

Pode-se assumir que os processos autorregressivos são sempre invertíveis. Para que sejam estacionários, as raízes do polinómio autorregressivo têm de estar fora do círculo unitário, portanto,  $|G_i| < 1$ , para  $i = 1, 2, \dots, p$ .

**Função de autovariância.** Define-se a função de autovariância através de

$$\gamma_k = E(Y_t Y_{t+k}) = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}, k \geq 1.$$

**Função de autocorrelação (FAC).** Se dividir a função de autocovariância por  $\gamma_0 = \sigma^2$ , obtém-se a expressão para a função de autocorrelação

$$\rho_k = \gamma_k/\gamma_0 = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \dots + \phi_p\rho_{k-p}, k \geq 1.$$

**Função autocorrelação parcial (FACP).** Tendo em conta a definição da função autocorrelação parcial, FACP de um processo autorregressivo (AR) é definido por,

$$\phi_{kk} = \begin{cases} \frac{|P_k^*|}{|P_k|}, & k = 1, 2, \dots, p \\ 0, & k = p + 1, p + 2, \dots \end{cases}$$

Num caso de processo autorregressivo de ordem 1,  $AR(1)$ , tem-se

$$Y_t = \phi Y_{t-1} + \varepsilon_t,$$

ou

$$(1 - \phi B)Y_t = \varepsilon_t,$$

em que  $\phi$  é um número real e  $\varepsilon_t$  é um ruído branco. Além disso,

- Os processos  $AR(1)$  são sempre invertíveis;
- Para que um processo seja estacionário, a variância de  $Y_t$ , tem de se manter constante para qualquer  $t$  e, por isso,  $|\phi| < 1$ , logo  $\mu = E(Y_t) = 0$ ;
- Sendo o processo estacionário, a variância de  $Y_t$  é dada por  $var(Y_t) = \frac{\sigma_\varepsilon^2}{1-\phi^2}$ ;
- Sendo o processo estacionário, a função de autocovariância é dada por

$$\gamma_k = \frac{\phi^k \sigma_\varepsilon^2}{1 - \phi^2}, k \geq 1;$$

- A função de autocorrelação (FAC) é dada por

$$\rho_k = \phi^k;$$

- Ambas as funções de autocovariância e de autocorrelação decrescem exponencialmente para zero à medida que  $k$  aumenta. Quando  $\phi > 0$ , as autocorrelações são sempre positivas, no entanto, quando  $\phi < 0$ , estas alternam de sinal e partem para uma autocorrelação de primeira ordem negativa.

## b) Processos de médias móveis de ordem $q$ – $(MA(q))$

**Definição 7** O processo  $Y_t, t \in T$  diz-se um processo de médias móveis de ordem  $q$ ,  $MA(q)$  se satisfaz a seguinte equação

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

onde  $\varepsilon_t$  é um processo de ruído branco e  $\theta_i, i = 1, \dots, q$  são constantes reais.

Através do operador de atraso, representado por  $B$ , a Eq.(7) pode ser reescrita,

$$Y_t = \Theta_q(B) \varepsilon_t,$$

onde  $\varepsilon_t$  é um ruído branco e  $\Theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$  é o polinómio de médias móveis de ordem  $q$ . Um processo de médias móveis de ordem  $q$  define-se em cada instante  $t$ , como a média ponderada das  $q + 1$  observações de um processo de ruído branco.

**Condições de invertibilidade e estacionariedade.** Os processos de médias móveis são sempre estacionários e são invertíveis quando a equação  $\Theta_q(B) = 0$  tem as suas raízes fora do círculo unitário.

**Função de autocovariância.** Considerando que  $E(Y_t) = 0$  e a variância  $\gamma_0 = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma_\varepsilon^2$ , a função de autocovariância é dada por,

$$\rho_k = \begin{cases} \sigma_\varepsilon^2 (-\theta_k + \theta_{k+1} \theta_1 + \dots + \theta_q \theta_{q-k}), & 0 < k \leq q \\ 0, & k > q. \end{cases}$$

**Função de autocorrelação (FAC)**

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_{k+1} \theta_1 + \dots + \theta_q \theta_{q-k}}{1 + \theta_1^2 + \dots + \theta_q^2}, & 0 < k \leq q \\ 0, & k > q. \end{cases}$$

É de notar a queda brusca da FAC para  $k \geq q + 1$ .

**Função de autocorrelação parcial (FACP).** A FACP pode ser difícil de obter-se analiticamente, no entanto, é possível estabelecer que ela é limitada pela soma de duas exponenciais com amortecimento quando as raízes do polinómio de médias móveis são reais, e por uma função sinusoidal amortecida quando as raízes são complexas, Murteira (1993). Na análise de séries temporais, a função de autocovariância parcial (FACP) de um processo de médias móveis ( $MA$ ) de ordem  $q$  tem a mesma

forma que a FAC de um processo autorregressivo ( $AR$ ) de ordem  $p$ , que decai gradualmente para zero. Por outro lado, a FACP de um processo  $AR(p)$  comporta-se como a FAC dos processos  $MA(q)$ , que decai rapidamente para zero. Esta correspondência indica a dualidade existente entre esses dois tipos de processos.

### c) Processos autorregressivo e de médias móveis – $ARMA(p, q)$

Os processos autorregressivos e de médias móveis são modelos que combinam as duas representações, o processo autorregressivo e o processo de médias móveis.

**Definição 8** Um processo  $Y_t, t \in T$  diz-se ser um processo autorregressivo e de médias móveis de ordens  $p$  e  $q$ ,  $ARMA(p, q)$ , se satisfaz a equação

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q},$$

onde  $\varepsilon_t$  é um processo de ruído branco e  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$  são constantes reais.

Utilizando novamente o operador atraso,  $B$ , a equação na definição 8 pode ser reescrita por,

$$\Phi_p(B)Y_t = \Theta_q(B)\varepsilon_t,$$

onde  $\varepsilon_t$  é um ruído branco, independente de  $Y_{t-k}$  para todo o  $k \geq 1$ ,  $\Phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  é o polinómio autorregressivo de ordem  $p$  e  $\Theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  é o polinómio de médias móveis de ordem  $q$ .

**Condições de invertibilidade e estacionariedade.** Para o processo ser estacionário é necessário que as raízes de  $\Phi_p(B) = 0$  estejam fora do círculo unitário. Para que o processo seja invertível, as raízes de  $\Theta_q(B) = 0$  devem estar fora do círculo unitário.

**Função de autocovariância** Se na equação da Definição ( 8) multiplicar-se por  $Y_t$ , tomando os valores esperados, obtém-se

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}, k \geq q + 1.$$

**Função de autocorrelação (FAC).** Quando se divide a função de autocovariância por  $\gamma_0 = \sigma^2$ , resulta na expressão para a FAC,

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}, k \geq q + 1.$$

**Função de autocorrelação parcial FACP.** A função de autocorrelação parcial (FACP) de uma série temporal tem um comportamento que decai gradualmente para zero, sendo limitado superiormente pela soma de exponenciais e/ou sinusoides amortecidas. Este comportamento é muito semelhante ao da FACP dos processos de médias móveis, Murteira (1993).

A Tabela 2 resume os comportamentos teóricos das FAC e FACP para os processos explicados.

Tabela 2: Padrões teóricos das FAC e FACP do modelo  $ARMA(p, q)$

Processo	FAC	FACP
$AR(p)$	Decaimento exponencial e/ou sinusoidal para zero	Anula-se a partir de $k = p + 1$
$MA(q)$	Anula-se a partir de $k = p + 1$	Decaimento exponencial e/ou sinusoidal para zero
$ARMA(p, q)$	Decaimento exponencial e/ou sinusoidal para zero	Decaimento exponencial e/ou sinusoidal para zero

#### d) Processo autorregressivo e de médias móveis integrado sazonal – $SARIMA(p, d, q)(P, D, Q)_s$

Em séries temporais sazonais é esperado que a componente sazonal esteja relacionada, de alguma forma, com as componentes não sazonais, ou seja, se as observações vizinhas de uma série,  $Y_t, Y_1, Y_2, Y_3, \dots$ , estão relacionadas, há uma probabilidade que as observações vizinhas espaçadas em  $s$  unidades temporais,  $Y_t, Y_{t-s}, Y_{t-2s}, \dots$  também estejam relacionadas. Uma diferenciação sazonal é a diferença entre uma observação no instante  $t$  e a observação que ocorre  $s$  momentos antes, isto é, a observação no instante  $t - s$ . Quando uma série temporal apresenta um comportamento periódico repetitivo, aplica-se uma diferenciação sazonal, ou seja,

$$\nabla_s Y_t = Y_t - Y_{t-s} = (1 - B^s)Y_t.$$

A série que resulta desta diferenciação corresponde, então, à mudança entre observações separada por períodos de tempo  $s$ . Semelhante ao que acontece para a diferenciação na Eq.(3.5), a diferenciação sazonal pode ser aplicada a uma série  $D$  vezes, o que dá origem ao operador de diferenciação sazonal de ordem  $D$ , para qualquer inteiro de  $D \geq 1$ , define-se,

$$\nabla_s^D Y_t = (1 - B^s)^D Y_t. \quad (3.6)$$

Para tal, um processo  $Y_t$  é um processo autorregressivo e de médias móveis integrado sazonal,  $SARIMA(p, d, q)(P, D, Q)_s$  quando satisfaz a equação,

$$\Phi_p(B)N_P(B^s)\nabla^d\nabla_s^D Y_t = \Theta_q(B)H_Q(B^s)\epsilon_t, \quad (3.7)$$

em que  $\Phi_p(B), N_P(B^s), \Theta_q(B)$  e  $H_Q(B^s)$  são polinômios já referidos,  $d$  e  $D$  são as ordens de diferenciação das partes regular e sazonal, respetivamente.

De modo geral, quando a FAC de um processo decai lentamente nos *lags* múltiplos de  $s$  e é insignificante nos restantes, existe uma necessidade de proceder a uma diferenciação sazonal.

## 3.2 Modelos de espaço de estados

### 3.2.1 Introdução

Os modelos de espaço de estados (MEE) são uma abordagem matemática para descrever o comportamento dinâmico de sistemas ao longo do tempo, e são amplamente usados em diversas áreas, incluindo a modelação de séries temporais. Em séries temporais, os dados são recolhidos em intervalos regulares ao longo do tempo, e a dependência temporal entre as observações é uma característica fundamental que é considerada na modelação. Os MEE quando considerados para modelar séries temporais descrevem o estado latente (não observado) de um sistema como uma coordenada em um espaço de estados. O estado latente representa a informação não observada, mas que é relevante para a modelar uma série temporal. O estado latente num determinado momento é calculado usando as observações passadas e presentes, bem como um ruído branco. O ruído branco é modelado por meio de um processo estocástico, que é adicionado às equações de estado para tornar o modelo mais flexível e permitir mais uma fonte de variabilidade. Os MEE em séries temporais são usados para realizar previsões, filtragem e suavização. Estes modelos são particularmente úteis para modelar séries temporais com tendência, sazonalidade, ciclos e outras formas de dependência temporal complexa.

Os modelos de regressão linear simples são um caso particular dos MEE, uma vez que, neste caso, os estados são considerados determinísticos, ou seja, são invariantes no tempo.

#### a) Modelos de regressão linear simples

Para obter-se uma previsão de uma variável a partir de uma ou mais covariáveis é necessário que exista uma relação causa-efeito, ou seja, que a variação de uma variável possa ser explicada através da

variação de outras, Shumway & Stoffer (1991). Existem diversos tipos de relação entre variáveis, estas podem ser lineares, exponenciais, logarítmicas, etc. Para melhor identificar o tipo de relação funcional existente, deve começar-se por fazer um diagrama de dispersão de dados. Quando a relação é linear, o diagrama apresenta os pontos dispersados, aleatoriamente, em torno de uma reta no caso da regressão linear simples.

Num modelo de regressão linear simples a relação entre uma determinada variável resposta,  $Y_t$ , e a uma determinada variável explicativa,  $X_t$ , pode ser descrita matematicamente através da equação,

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t, t = 1, \dots, n,$$

em que,

- a variável aleatória  $Y_t$  representa a variável resposta (ou dependente);
- a variável aleatória  $X_t$  representa a variável explicativa (preditor ou independente);
- os coeficientes de regressão  $\beta_1$  e  $\beta_0$ ;
- $\beta_0$  é a interseção da reta com o eixo vertical, representa também o valor esperado da variável aleatória  $Y_t$  quando a variável explicativa é nula,
- $\beta_1$  representa o declive da reta e representa a variação do valor esperado de  $Y_t$  por cada incremento unitário na variável explicativa. Os coeficientes de regressão são constantes e, por norma, desconhecidos;
- $\epsilon_t$  representa o erro aleatório e descreve fatores explicativos de  $Y_t$  que não são explicados através da variável  $X_t$  e pode ainda incluir erros de medição.

## **b) Modelos de regressão linear múltipla**

Em certos casos, a variável  $Y_t$  pode ser influenciada por mais do que uma variável explicativa, sendo estas controladas pelo investigador. Por exemplo, cientistas de dados que trabalham para equipas profissionais de desporto (futebol ou basquetebol) utilizam frequentemente modelos de regressão linear múltipla para medir o efeito que diferentes regimes de treino têm no desempenho dos jogadores. Como um caso prático, os cientistas de dados da NBA (Liga Norte-americana de Basquetebol) podem analisar como diferentes

quantidades de sessões semanais de ioga e levantamento de pesos afetam o número de pontos que um jogador marca. Eles podem ajustar um modelo de regressão linear múltipla usando sessões de ioga e levantamento de pesos como variáveis explicativas e o número total de pontos marcados como variável de resposta, no entanto pode ser afetado por outras variáveis que não são controláveis, como por exemplo a qualidade de sono ou o estado psicológico de um jogador. Quando estão presentes duas ou mais variáveis explicativas, chama-se modelo de regressão linear múltipla e é definido por

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \dots + \beta_p X_{t,p} + \epsilon_t, t = 1, \dots, n.$$

Tal como na equação do modelo de regressão linear simples,  $\epsilon_t$  representa também a variação de  $Y_t$  que não é explicada pelas variáveis explicativas  $X_t$ .  $\beta_0$  representa o valor esperado de  $Y_t$ , no momento em que as variáveis explicativas são simultaneamente nulas e  $\beta_i$ , ( $i = 1, \dots, p$ ) representa a variação do valor esperado de  $Y_t$  por cada incremento unitário na variável  $X_i$ , para quando as restantes variáveis explicativas se mantêm constantes.

Este modelo também é possível ser representado, com  $n$  igualdades, sob a notação matricial. O modelo de regressão linear múltipla, considerando

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{1,1} & \dots & X_{1,p} \\ 1 & X_{2,1} & \dots & X_{2,p} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n,1} & \dots & X_{n,p} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_p \end{bmatrix}$$

pode ser escrito na forma

$$Y = X\beta + \epsilon.$$

Em que  $Y$  é o vetor  $n \times 1$  das variáveis respostas,  $X$  é a matriz  $n \times (p + 1)$  das observações das variáveis explicativas. Uma vez que  $\beta_0$  é um termo independente, a primeira coluna da matriz de  $X$  é constituída por uns. O vetor  $\beta$  tem dimensão  $(p + 1) \times 1$  e contém os coeficientes da regressão.  $\epsilon$  é um vetor com dimensões  $n \times 1$  dos erros aleatórios.

### 3.2.2 Modelos de espaço de estados (MEE)

Os modelos que se baseiam na representação de espaço de estados possuem inúmeras aplicações em diferentes áreas, como a Economia, Biologia, *Machine Learning*, Análise de Séries Temporais, entre outros. Originalmente, os MEE foram desenvolvidos na engenharia na década de 60 do século passado com o objetivo de monitorizar e controlar sistemas dinâmicos. Segundo Hutchinson (1984), Os modelos

foram utilizados nas décadas de 70 e 80 em programas aeroespaciais, como o *Apollo* e o *Polaris*. A formulação de espaço de estados assume que o desenvolvimento de um sistema ao longo do tempo é determinado por um processo de vetores não observados  $\beta_1, \beta_2, \dots, \beta_n$  (estados), associados linearmente, ou não, a uma sequência de variáveis observadas  $Y_1, Y_2, \dots, Y_n$ . Para a análise de séries temporais, a representação de espaço de estados estabelece a relação temporal entre as variáveis que são observadas, as variáveis não observáveis (estados) e os erros estocásticos, (Costa, 2006). Os MEE são modelos flexíveis devido à sua capacidade de poderem incorporar várias características dos dados, permitindo a análise de fenômenos dinâmicos que variam ao longo do tempo. De acordo com Hutchinson (1984), estes são muito mais flexíveis do que os modelos *ARMA* na modelação de séries temporais, em particular, não estacionárias e na modelação de mudanças estruturais, e, na generalidade, com uma melhor interpretação dos parâmetros e da sua estrutura.

### 3.2.3 Modelos de espaço de estados lineares gaussianos

O modelo linear gaussiano de espaço estados, ou modelo dinâmico linear, é composto por duas equações, a equação de observação e a equação de estado. A equação de observação relaciona o vetor  $p$ -dimensional das variáveis observáveis  $Y_t$  com um vetor não observável  $m \times 1$  designado por  $\beta_t$ , a que se chama de vetor de estados. A equação de estados define-se por,

$$\beta_t = \mu + \Phi(\beta_{t-1} - \mu) + e_t, t = 1, \dots, n, \quad (3.8)$$

em que,

- $\Phi$  é uma matriz  $m \times m$  denominada por matriz autorregressiva, e  $\mu$  é o valor médio do estado;
- $e_t$  é um vetor  $m \times 1$  dos erros independentes e identicamente distribuídos e segue uma distribuição Normal de média igual a zero e uma matriz de covariância  $Q$ ;
- $e_t \sim \text{i.i.d. } N_n(0, Q)$ .

A Eq.(3.8) descreve a dinâmica do vetor de estados não observados  $\beta_t$ . Este é atualizado ao longo do tempo através de um vetor autorregressivo de ordem 1, VAR(1).

Define-se a equação de observação como

$$Y_t = W_t \beta_t + \varepsilon_t, t = 1, \dots, n, \quad (3.9)$$

em que,

- $W_t$  é uma matriz  $p \times m$ ;
- $\varepsilon_t$  é um vetor  $p \times 1$  dos erros independentes e identicamente distribuídos e segue uma distribuição Normal multivariada de média igual a zero e uma matriz de variâncias-covariâncias  $H$ ;
- $\varepsilon_t \sim \text{i.i.d. } N_p(0, H)$ .

Os MEE definidos com as equações de observação e de estados permitem estabelecer diversos modelos pelos quais se torna possível tratar dados em falta ou omissos. Permitem também definir modelos com estruturas de efeitos fixos ou aleatórios (podem ser constantes no tempo ou variarem ao longo do tempo).

### 3.3 Filtragem, alisamento e previsão de Kalman

#### 3.3.1 Introdução

A filtragem de Kalman é um método matemático que usa uma série de observações para estimar o estado atual de um sistema dinâmico e prever seu estado futuro. O filtro de Kalman é amplamente utilizado em diversas áreas, como em engenharia, robótica, ciência da computação, sistemas de controle, entre outras. O alisamento de Kalman é uma extensão do filtro de Kalman que utiliza informações passadas e futuras para melhorar a estimativa do estado atual do sistema. Por outras palavras, o alisamento de Kalman é capaz de suavizar as estimativas do filtro de Kalman, tornando-as mais precisas. A previsão de Kalman, por sua vez, é uma aplicação do filtro de Kalman que utiliza a estimativa do estado atual do sistema para prever o seu estado futuro. A previsão de Kalman é pertinente em situações que se deseja estimar o comportamento futuro de um sistema com base em observações passadas e presentes.

Matematicamente o objetivo é encontrar o melhor estimador linear de  $\beta_t$ , vetor de estados não observado, ou seja o estimador que possuir o menor erro quadrático médio tendo por base as observações  $Y_1, Y_2, \dots$ . Divide-se a predição de  $\beta_t$  conforme o problema,

- 1.  $Y_1, Y_2, \dots, Y_{t-1}$ , **problema de previsão**;
- 2.  $Y_1, Y_2, \dots, Y_t$ , **problema de filtragem**;
- 3.  $Y_1, Y_2, \dots, Y_n; (n > t)$ , **problema de alisamento**.

Em ambos os casos as predições são obtidas recursivamente usando um conjunto de equações recursivas de Kalman, Shumway & Stoffer (1991).

### 3.3.2 Filtro de Kalman

O filtro de Kalman foi proposto por Rudolf Emil Kalman em 1960 Kalman et al. (1960) e foi implementado no Projeto Apollo em 1961 para resolver o problema de navegação espacial Grewal & Andrews (2010). O filtro de Kalman proporciona preditores ótimos Kalman et al. (1960) devido à sua capacidade de prever a covariância do erro do sistema e usar a previsão de forma recursiva para melhorar as medições do sistema de tempos em tempos. Como tal, o filtro de Kalman foi implementado em várias aplicações, como em navegações Barczyk & Lynch (2012) e Lim, Lim, & Koo (2012), processamento de imagem Salti, Lanza, & Di Stefano (2014), Bresson, Féraud, Aufrere, Checchin, & Chapuis (2015) e finanças Racicot & Théoret (2010).

Uma das características do filtro de Kalman é que ele consiste em dois processos distintos, nomeadamente, o processo de previsão e o processo de filtragem. Ambos os processos são combinados e operados de forma recursiva para alcançar um processo de filtragem de Kalman ótimo. Outra característica única do filtro de Kalman é a incorporação de erros de previsão e erros de medição no processo global da filtragem de Kalman. É comum que cada processo de previsão e medição seja baseado em erros de natureza aleatória. Esses erros ou ruído são normalmente descritos através de um processo estocástico. Por outro lado, uma aplicação em tempo real pode ser definida como uma aplicação ou programa que reage ou responde dentro de um quadro de tempo predefinido, onde esse quadro de tempo predefinido é um tempo quantificado usando um relógio físico. Do ponto de vista de uma aplicação em tempo real, o tempo contínuo do mundo real é convertido num quadro de tempo discreto. Diferentes aplicações em tempo real têm requisitos de tempo diferentes, que, por sua vez, definem o tempo de resposta das aplicações. A aplicação em tempo real deve reagir dentro do quadro de tempo predefinido para fornecer uma resposta atualizada. Essa restrição de tempo real força a aplicação a concluir a sua rotina dentro do tempo determinado, caso contrário, a saída pode não refletir com precisão o estado atual da entrada.

O filtro de Kalman (FK) é um algoritmo recursivo utilizado em MEE que permite obter as predições ótimas para o vetor de estados não observado  $\beta_t$ . As equações do FK formam um sistema que obtém projeções lineares a cada instante de tempo  $t$ . Desta forma calculam-se estimadores lineares com o menor valor do erro quadrático médio. Através do FK também se obtém predições a 1-passo para os vetores das variáveis observadas. O vetor de estado é atualizado para cada observação nova no instante  $t - 1$ , deduzindo assim a distribuição condicional de  $\beta_t$  ( $t = 1, \dots, n$ ) sempre com baseadas nas observações disponíveis até ao instante  $t - 1$ . Usualmente assume-se que os erros aleatórios seguem

distribuições normais e, neste caso, os preditores do FK são ótimos no sentido de que têm o menor erro quadrático médio dentre todos os preditores, e não apenas dentre os lineares.

Considera-se o modelo com representação em espaço de estado dado pelas Eq.(3.8) e Eq.(3.9) e, sem perda de generalidade, que cujo estado é estacionário de média  $\mu$ ,

$$\beta_t = \mu + \Phi(\beta_{t-1} - \mu) + \varepsilon_t, t = 1, \dots, n,$$

e

$$Y_t = W_t\beta_t + e_t, t = 1, \dots, n.$$

Seja  $\beta_{t|t-1}$  o estimador de  $\beta_t$  com o menor erro quadrático médio com base na informação disponível até ao instante  $t - 1$ , representando o vetor de observações  $\tilde{Y}_{t-1} = (Y_1, Y_2, \dots, Y_{t-1})$ , ou seja,

$$\beta_{t|t-1} = E[\beta_t | \tilde{Y}_{t-1}],$$

com  $P_{t|t-1}$  sendo uma matriz  $m \times m$  (matriz de covariâncias), que por sua vez representa o erro quadrático médio,

$$P_{t|t-1} = E[(\beta_t - \beta_{t|t-1})(\beta_t - \beta_{t|t-1})']. \quad (3.10)$$

Considerando que

$$\beta_{t|t} = E[\beta_t | \tilde{Y}_t] \quad (3.11)$$

e

$$P_{t|t} = E[(\beta_t - \beta_{t|t})(\beta_t - \beta_{t|t})' | \tilde{Y}_t]; \tilde{Y}_t = (Y_1, Y_2, \dots, Y_t)'. \quad (3.12)$$

A previsão de  $Y_t$ , atualizado até aos instante  $t - 1$ , a qual se domina por equação da previsão é dada por

$$Y_{t|t-1} = E[Y_t | \tilde{Y}_{t-1}] = W_t\beta_{t|t-1}. \quad (3.13)$$

No momento que a observação  $Y_t$  é obtida no instante  $t$ , o erro quadrático médio e a respetiva atualização da previsão do vetor de estado  $\beta_t$  pode ser expressada por,

$$\beta_{t|t} = \beta_{t|t-1} + K_t(Y_t - Y_{t|t-1}). \quad (3.14)$$

$$P_{t|t} = [I - K_t W_t] P_{t|t-1}. \quad (3.15)$$

Este processo é chamado de filtragem, Eq.(3.14) e Eq.(3.15) são designadas por equações de atualização. Onde,

$$K_t = P_{t|t-1} W_t' [W_t P_{t|t-1} W_t' + H_t]^{-1}; t = 1, \dots, n. \quad (3.16)$$

Seja  $K_t$  o ganho de Kalman, uma matriz  $m \times p$  e  $I$  é uma matriz identidade de ordem  $m$ .

É possível obter uma previsão para o vetor de estados para um instante  $t + 1$ , através de toda a informação disponível até ao instante  $t + 1$ , através da Eq.(3.8), é possível realizar uma previsão sob a forma,

$$\beta_{t+1|t} = \mu + \Phi(\beta_{t|t} - \mu), \quad (3.17)$$

$$P_{t+1|t} = \Phi P_{t|t} \Phi' + Q. \quad (3.18)$$

Denominadas por equações preditivas, para se prever para  $t > n$ , as condições iniciais são  $\beta_{t|t}$  e  $P_{t|t}$ . Denota-se por  $P_{t+1|t}$  o erro quadrático médio do estimador  $\beta_{t+1|t}$  e define-se por

$$\begin{aligned} P_{t+1|t} &= E[(\beta_{t+1} - \beta_{t+1|t})(\beta_{t+1} - \beta_{t+1|t})'] \\ &= E[(\mu + \Phi(\beta_t - \mu) - (\mu + \Phi(\beta_{t|t} - \mu)) + \varepsilon_{t+1})(\mu + \Phi(\beta_t - \mu) - (\mu + \Phi(\beta_{t|t} - \mu)) + \varepsilon_{t+1})'] \\ &= E[\Phi(\beta_t - \beta_{t|t}) + \varepsilon_{t+1}(\Phi(\beta_t - \beta_{t|t}) + \varepsilon_{t+1})]. \end{aligned}$$

Atendendo a que os produtos cruzados têm esperança nula, simplifica-se,

$$P_{t+1|t} = \Phi P_{t|t} \Phi' + \Sigma_\varepsilon. \quad (3.19)$$

As equações previamente introduzidas constituem o FK, no entanto, é necessário definir  $\beta_{1|0}$  e  $P_{1|0}$  para ser possível iniciar o processo recursivo. Sabendo que o vetor  $\beta_t$  é um processo estacionário com média  $\mu$  e com um erro quadrático médio  $\Sigma = E[(\beta_t - \mu)(\beta_t - \mu)']$ , segundo Costa (2006) o processo inicia-se com a previsão de  $\beta_1$ , recorrendo às suas características estocásticas caso não haja nenhuma informação disponível. Assim sendo, o valor inicial  $\beta_{1|0}$  é

$$\beta_{1|0} = \mu. \quad (3.20)$$

Sendo que a matriz de  $P_{1|0}$  será igual à matriz de covariâncias do vetor de estado  $\beta_t$ ,

$$P_{1|0} = \Sigma. \quad (3.21)$$

Após iniciar os valores, pode-se proceder à previsão do vetor de observação  $Y_t$  contendo toda a informação até ao momento  $t - 1$ ,

$$Y_{t|t-1} = W_t \beta_{t|t-1}. \quad (3.22)$$

Atualiza-se a previsão do vetor de estado  $\beta_t$  e do seu respetivo erro quadrático médio no momento que uma determinada observação  $Y_t$  fica disponível,

$$\beta_{t|t} = \beta_{t|t-1} + K_t(Y_t - Y_{t|t-1}) \quad (3.23)$$

e

$$P_{t|t} = [I - K_t W_t] P_{t|t-1}. \quad (3.24)$$

Neste contexto,  $\eta_{t|t-1} = Y_t - Y_{t|t-1}$  é o erro de previsão no instante  $t$ , ou inovação, e  $K_t$  representa o ganho de Kalman Eq.(3.16). Assim, define-se o erro de previsão ou inovação como sendo

$$\eta_{t|t-1} = Y_t - E(Y_{t|t} | \tilde{Y}_{T-1}) = Y_t - W_t \beta_{t|t-1} = Y_t - Y_{t|t-1}, \quad (3.25)$$

com a respetiva matriz de covariâncias dada por

$$\Sigma_{t|t-1} = \text{var}(\eta_{t|t-1}) = \text{var}[W_t(\beta_t - \beta_{t|t-1}) + e_t] = W_t P_{t|t-1} W_t' + H. \quad (3.26)$$

A partir da melhor previsão de  $\beta_t$  no instante  $t$ , isto é, a partir de  $\beta_{t|t}$ , podemos prever  $\beta_{t+1}$  e o calcular o seu respetivo erro quadrático médio através de,

$$\beta_{t+1|t} = \Phi \beta_{t|t}, \quad (3.27)$$

$$P_{t+1|t} = \Phi P_{t|t} \Phi' + Q. \quad (3.28)$$

Depois, recomeça-se o processo na previsão de  $Y_t$  dada toda a informação até ao instante  $t - 1$ .

Quando o FK processa todas as observações ( $n$ ) é devolvido o estimador ótimo do vetor de estados atual, sendo que este estimador contém nesse momento toda a informação que necessita para realizar previsões ótimas dos estados futuros e das observações. Uma vez que o FK é um algoritmo recursivo, ou seja, tem uma função que chama a si mesma repetidamente para chegar a uma solução, devolve assim o valor de  $\beta_{t|t}$  através de  $\beta_{t|t-1}$  apenas quando há uma nova observação  $Y_t$  sem haver necessidade de reprocessar os dados anteriores ( $Y_1, Y_2, Y_3, \dots, Y_{t-1}$ ). A facilidade de implementação do filtro de Kalman e sua estrutura recursiva, juntamente com suas propriedades teóricas, justificam sua ampla aplicação. De facto, os estimadores associados ao filtro de Kalman são as projeções ortogonais que têm o menor erro quadrático médio entre todos os estimadores lineares, para qualquer distribuição dos erros aleatórios. Além disso, como os estimadores  $\beta_{t|t-1}$  e  $\beta_{t|t}$  estimam vetores aleatórios e não parâmetros, estes são considerados estimadores não enviesados, no sentido de que o valor esperado do erro de estimação é igual a zero.

### 3.3.3 Alisamento

Alisamento ou *smoothing* consiste em obter previsões do vetor de estados  $\beta_t$  de variáveis observáveis  $Y_t$ , atualizando as previsões do vetor de estados quando estão disponíveis observações posteriores a estas.

Resumindo, consiste em obter  $\beta_t$  com base em  $n$  observações  $Y_1, Y_2, Y_3, \dots, Y_n$ , representando-se por  $\beta_{t|n}$ , sendo que, neste caso,  $t \leq n$ . Quando se considera a totalidade das observações, a precisão das predições aumenta, e ainda, permite resolver o problema de ausência de observações. Segundo Shumway & Stoffer (1991) esta metodologia designa-se de *smoothing*, uma vez que o gráfico de  $\beta_{t|n}$  é mais liso do que as previsões a 1 passo  $\beta_{t|t-1}$  e pelas predições filtradas  $\beta_{t|t}$ .

Através das equações definidas de um MEE, Eq.(3.8) e Eq.(3.9), substitui-se as condições iniciais definidas nas Eq.(3.14) e Eq.(3.15) por  $\beta_{n|n}$ ,  $P_{n|n}$  e obtém-se,

$$\beta_{t-1|n} = \beta_{t-1|t-1} + J_{t-1}(\beta_{t|n} - \beta_{t|t-1}), \quad (3.29)$$

$$P_{t-1|n} = P_{t-1|t-1} + J_{t-1}(P_{t|n} - P_{t|t-1})J'_{t-1}, \quad (3.30)$$

em que,

$$J_{t-1} = P_{t-1|t-1}\Phi'[P_{t|t-1}]^{-1}, t = n, n-1, \dots, 1. \quad (3.31)$$

Resumidamente, as predições obtidas pelo método de alisamento (*smoothing*) têm menores erros quadráticos médios do que as obtidas através das previsões a 1 passo ou da filtragem. Isso ocorre devido à quantidade de dados usados em cada iteração. O método de alisamento utiliza todos os dados disponíveis, enquanto as previsões usam apenas informações disponíveis no momento anterior e a filtragem usa os dados disponíveis até o momento atual. Como resultado, os intervalos de confiança construídos a partir da predição  $\beta_{t|n}$  num determinado momento apresentará menor amplitude. No cálculo da predição  $\beta_{t|n}$  é incorporado significativamente mais observações do que na filtragem, por um lado aumenta a dificuldade na execução do alisamento, sendo que, deve-se ponderar em que situação as vantagens justificam a complexidade.

### 3.3.4 Previsão

Para realizar uma previsão no FK são utilizadas todas as informações disponíveis para  $t$  dado  $t-1$  (os dados são atualizados sequencialmente sempre que é introduzida uma nova observação). Como no alisamento e na filtragem de Kalman as Eq.(3.9) e Eq.(3.8) que constituem o MEE são consideradas para realizar a previsão. O FK estima o vetor de estados  $\beta_t$  para um instante  $n$  com base nas  $n$  observações. Por outras palavras atualiza-se o vetor de estado através da ultima observação  $Y_n$ .

$$\beta_{n+1|n} = \mu + \Phi(\beta_{n|n} - \mu), \quad (3.32)$$

sabendo que observação  $n+1$  (atualização) é explicada pela seguinte equação,

$$Y_{n+1|n} = W_{n+1}\beta_{n+1|n}. \quad (3.33)$$

Consideram-se os passos por  $h$  como os problemas de previsão, em que,  $h = 2, 3, 4, \dots$  substitui-se na Eq.(3.8) por  $n + h^4$ ,

$$\beta_{n+h} = \Phi^h \beta_n + \sum_{j=1}^{h-1} \Phi^{h-j} \varepsilon_{n+j} + \varepsilon_{n+h}, h = 2, 3, 4, \dots \quad (3.34)$$

observando que,

$$\beta_{n+h|n} = E[\beta_{n+h} | \tilde{Y}_t] = \Phi^h \beta_{n|n}; \quad (3.35)$$

com uma matriz de covariâncias expressada por,

$$P_{n+h|n} = \Phi^h P_{n|n} (\Phi^h)' + \sum_{j=1}^{h-1} \Phi^j Q (\Phi^j)', \quad h = 2, 3, 4, \dots \quad (3.36)$$

É possível obter uma estimativa para  $Y_{n+h}$  através no vetor de estado para a previsão  $\beta_{n+h|n}$ . Para determinar o valor esperado,

$$Y_{n+1|n} = E[Y_{n+1} | \tilde{Y}_t] = W_{n+h} \beta_{n+h|n}, h = 2, 3, 4, \dots, \quad (3.37)$$

a matriz de covariâncias é dada por,

$$\Sigma_{n+h|n} = W_{n+h} P_{n+h} W_{n+h}' + H, h = 2, 3, 4, \dots, \quad (3.38)$$

## a) Intervalos de confiança para as previsões

Em determinados casos a estimação pontual não é suficiente para quantificar a incerteza em relação a uma previsão. Os intervalos de confiança são uma resposta para esta incerteza, como o futuro é de facto desconhecido e as previsões também o são, assim sendo, uma estimação intervalar permite quantificar a incerteza associada às previsões pontuais. Estes intervalos de confiança são baseados na matriz de covariâncias do erro de previsão  $\Sigma_{t|t-1}$  uma vez que permite estimar o erro de previsão a 1 passo. Para se justificar a estimação intervalar de uma previsão é necessário que o modelo ajustado consiga descrever uma série temporal e ainda que os seus respetivos pressupostos sejam válidos. A estatística de teste utilizada para calcular um intervalo de confiança de uma previsão univariada a 1 passo é a seguinte

$$\frac{Y_t - Y_{t|t-1}}{\sqrt{\Sigma_{t|t-1}}} \sim N(0, 1), \quad (3.39)$$

onde  $\Sigma_{t|t-1}$  representa o erro quadrático médio da inovação  $\eta_t = Y_t - Y_{t|t-1}$ .

---

<sup>4</sup> $n$  são as observações e  $h$  os problemas de previsão

Num intervalo de confiança,  $z$  corresponde a um quantil da distribuição Normal padrão,  $1 - \alpha$  corresponde ao nível de confiança do intervalo. O intervalo obtém-se a partir de

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{Y_t - Y_{t|t-1}}{\sqrt{\Sigma_{t|t-1}}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \quad (3.40)$$

obtendo-se

$$P\left(Y_{t|t-1} - z_{1-\frac{\alpha}{2}} \sqrt{\Sigma_{t|t-1}} < Y_t < Y_{t|t-1} + z_{1-\frac{\alpha}{2}} \sqrt{\Sigma_{t|t-1}}\right) = 1 - \alpha.$$

Assim, o intervalo de confiança para o nível  $(1 - \alpha) \times 100\%$  é definido por,

$$\left(Y_{t|t-1} - z_{1-\frac{\alpha}{2}} \sqrt{\Sigma_{t|t-1}}, Y_{t|t-1} + z_{1-\frac{\alpha}{2}} \sqrt{\Sigma_{t|t-1}}\right). \quad (3.41)$$

De modo análogo, existe um conjunto de passos descrito por Hyndman et al. (2014) para obter previsões a  $h$ -passos para um instante  $t + h$ , a partir  $\beta_{t+h|t}$  e  $P_{t+h|t}$  já deduzidos na secção anterior.

## 3.4 Estimação de parâmetros

Definidas as equações do estado Eq.(3.8) e espaço Eq.(3.9) existe um vetor composto por um conjunto de parâmetros desconhecidos,  $\Theta = [\Phi, H, Q]$ . Neste subcapítulo são abordadas duas formas distintas de estimar esses parâmetros, pelo método da máxima verosimilhança gaussiana e através de estimadores *Distribution-Free*.

### 3.4.1 Estimação pela máxima verosimilhança

Para um modelo em espaço de estados por vezes é necessário utilizar técnicas de estimação pontual para a eventualidade de o modelo não especificar os seus parâmetros. Para estes casos a estimação pelo método da máxima verosimilhança é a mais utilizada. A ideia do método da máxima verosimilhança consiste em selecionar, entre todos os valores possíveis dos parâmetros populacionais, aqueles que tornam mais verosímil a ocorrência de uma amostra idêntica àquela que efetivamente se obteve. A probabilidade da ocorrência de uma amostra aleatória ser idêntica à de uma obtida a partir da população chama-se de estimativa de máxima verosimilhança, as estimativas do vetor de parâmetros  $\Phi$  são os valores que maximizam a função verosimilhança, ou seja, são os valores dos parâmetros que tornam máxima a probabilidade de ocorrência de uma amostra ser idêntica à que efetivamente ocorreu. Por outras palavras pretende-se maximizar a probabilidade das observações  $Y_1, Y_2, \dots, Y_n$  em relação aos

componentes do vetor  $\Phi$ , sabendo que o estado inicial  $\beta_t$  segue uma distribuição Normal, bem como a componente aleatória  $\varepsilon_t \sim \text{i.i.d. } N(0, H)$ . A probabilidade é calculada através das inovações,

$$\eta_{t|t-1} = Y_t - W_t \beta_{t|t-1}, t = 1, \dots, n. \quad (3.42)$$

Defini-se a função de verosimilhança condicional por,

$$L(\Theta; \tilde{Y}_n) = \prod_{t=1}^n p(Y_t | \tilde{Y}_{t-1}). \quad (3.43)$$

Note que  $\tilde{Y}_n = (Y_1, Y_2, \dots, Y_n)$  e que  $p(Y_t | \tilde{Y}_{t-1})$  corresponde à densidade de  $Y_t$  dado por  $\tilde{Y}_{t-1}$ , através da Eq.(3.42) pode ser escrita como

$$p(Y_t | \tilde{Y}_{t-1}) = (2\pi)^{-\frac{1}{2}} |\Sigma_{t|t-1}(\Theta)|^{-\frac{1}{2}} \exp[-\frac{1}{2} \eta_{t|t-1}(\Theta)' \Sigma_{t|t-1}(\Theta)^{-1} \eta_{t|t-1}(\Theta)], \quad (3.44)$$

sabendo que  $\Sigma_{t|t-1} = W_t P_{t|t-1} W_t' + H$  expressa o erro quadrático médio das inovações.

Com a aplicação do logaritmo na Eq.(3.43), obtém-se a seguinte função verosimilhança,

$$\log L(\Theta; \tilde{Y}_n) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \log |\Sigma_{t|t-1}(\Theta)| - \frac{1}{2} \sum_{t=1}^n \eta_{t|t-1}(\Theta)' \Sigma_{t|t-1}(\Theta)^{-1} \eta_{t|t-1}(\Theta). \quad (3.45)$$

Note que na função anterior as inovações  $\eta_{t|t-1}(\Theta)$  e a sua respetiva matriz de covariâncias  $\Sigma_{t|t-1}(\Theta)$  dependem do vetor de parâmetros  $\Theta$  que se pretende estimar. Segundo Costa (2006) para certas aplicações, quando o processo  $\beta_t, t = 1, \dots, n$  não é estacionário, pode-se incluir no vetor  $\Theta$  os valores iniciais do FK, isto é,  $\beta_{1|0}$  e  $P_{1|0}$  para serem estimados com base nos dados. Como consequência há um aumento na complexidade dos processos da otimização da função (3.45). Há algumas formas de evitar esta situação, como por exemplo optar por valores iniciais não informativos ou por algoritmos mais específicos.

Uma vez que a função (3.45) é não linear e complexa, a sua otimização implica a utilização de métodos numéricos, sendo os mais comuns o método dos *scores* de Fisher ou o algoritmo EM (*Expectation – Maximization*). Tanto o método dos *scores* de Fisher quanto o algoritmo EM são algoritmos estatísticos usados para estimar parâmetros. No entanto, eles têm abordagens diferentes para a estimativa de parâmetros e são aplicados em contextos diferentes. O algoritmo EM é um algoritmo iterativo que é usado para estimar os parâmetros de modelos com distribuições incompletas ou dados não observados. Desenvolvido por Dempster et al. (1977), este método é composto por duas etapas iterativas: na etapa *E-step*, calcula-se o valor esperado da função log-verosimilhança (3.45); e, na etapa *M-step*, maximiza-se o valor esperado da mesma função (3.45). Uma desvantagem do algoritmo

EM é a necessidade de um grande número de iterações até se verificar os critérios de convergência. Também exige esforço computacional e ainda é sensível à escolha inicial dos parâmetros. A construção da verosimilhança, proposta por Shumway & Stoffer (1991), para o caso de se observarem os estados  $\beta_{1:n} = \{\beta_t\}_{t=1,\dots,n}$  e as observações  $Y_{1:n} = \{Y_t\}_{t=1,\dots,n}$ , a densidade conjunta é dada por,

$$p_{\Theta}(\beta_{0:n}, Y_{1:n}) = p_{\mu_0, \Sigma_0}(x_0) \prod_{t=1}^n p_{\phi, Q}(\beta_t | \beta_{t-1}) \prod_{t=1}^n p_H(Y_t | \beta_t). \quad (3.46)$$

Este algoritmo consiste num processo iterativo em que sucessivamente se maximiza a esperança condicional da log-verosimilhança. Cada uma destas iterações contém dois passos, Dempster et al. (1977). O passo do valor esperado ou *E-step*,

$$Q(\Theta | \Theta^{k-1}) = E[-2 \ln L_{\beta_t, Y_t}(\Theta) | Y_n, \Theta^{k-1}]. \quad (3.47)$$

O passo seguinte é o da maximização, isto é, a minimização do valor médio condicional dado por

$$Q(\Theta | \Theta^{k-1}) = E[-2 \ln L_{\beta_n, Y_n}(\Theta) | Y_n, \Theta^{k-1}], \quad (3.48)$$

corresponde à maximização da função log-verosimilhança.

Outra alternativa de um método numérico ao algoritmo EM é o método de Newton-Raphson. Este método é uma técnica iterativa para encontrar raízes de uma função. No contexto da máxima verosimilhança, o objetivo é encontrar o valor do parâmetro que maximiza a função de verosimilhança, ou seja, o valor do parâmetro que torna a observação dos dados mais provável. Começa-se com um valor inicial para um parâmetro e encontra-se a tangente à curva da função de verosimilhança nesse ponto. A tangente é então estendida até a intersecção com o eixo, o que fornece uma nova estimativa para o valor do parâmetro. Resolvendo a seguinte equação,

$$\partial 1(\Theta) = \frac{\partial \log L(\Theta; \tilde{Y}_n)}{\partial \Theta} = 0, \quad (3.49)$$

através da série de Taylor de primeira ordem obtém-se,

$$\partial 1(\Theta) \simeq \tilde{\partial} 2(\Theta) + \tilde{\partial} 1(\Theta)(\Theta - \tilde{\Theta}), \quad (3.50)$$

para um valor ótimo de  $\tilde{\Theta}$ , onde,

$$\tilde{\partial} 1(\Theta) = \partial 1(\Theta)|_{\Theta=\tilde{\Theta}},$$

$$\tilde{\partial} 2(\Theta) = \partial 2(\Theta)|_{\Theta=\tilde{\Theta}},$$

com

$$\tilde{\partial} 2(\Theta) = \frac{\partial^2 \log L(\Theta; \tilde{Y}_n)}{\partial \Theta \partial \Theta'}. \quad (3.51)$$

Se igualar a Eq.(3.50) a zero, irá obter-se o valor corrigido de  $\tilde{\Theta}$ ,

$$\tilde{\Theta} = \tilde{\Theta} - \tilde{\partial}^2(\Theta)^{-1}\tilde{\partial}^1\Theta. \quad (3.52)$$

Chama-se a  $\partial^1(\Theta)$  gradiente ou vetor *score* e este determina para que direção é esperado que a função log-verosimilhança vá crescer até ao ponto máximo. Este método converge rapidamente quando a estimativa inicial é próxima do valor real do parâmetro e quando a função de verosimilhança é bem sucedida (ou seja, a segunda derivada é sempre positiva). No entanto, o método pode não convergir para valores iniciais distantes ou quando a função de verosimilhança tem uma curvatura complexa ou mais de que um máximo local.

### 3.4.2 Estimadores *distribution-free*

Os estimadores *distribution-free* são uma alternativa para estimar parâmetros de modelos de espaço de estados. Uma possível dificuldade da adoção do método da máxima verosimilhança, é o facto de se ter de assumir que os erros da Eq.(3.9) e da Eq.(3.8) são gaussianos, o que na realidade, em muitos casos, é difícil de se verificar. Outra dificuldade associada ao método da máxima verosimilhança é necessidade da utilização de métodos numéricos complexos como o algoritmo EM, sendo que estes métodos necessitam de valores iniciais, o que na prática, nem sempre é fácil a sua identificação. Costa (2006) ainda refere que em MEE com coeficientes variáveis é colocada a dúvida sobre a unicidade das estimativas de máxima verosimilhança, além de que que essas mesmas estimativas podem corresponder a máximos locais.

Uma alternativa à máxima verosimilhança são os estimadores *distribution-free* ou estimadores independentes da distribuição de base, sendo que estes são uma classe de estimadores estatísticos que não fazem suposições específicas sobre a distribuição subjacente dos erros. Por outras palavras, estes estimadores não exigem que os erros sigam uma distribuição em particular, como uma distribuição Normal ou uma distribuição de Poisson. São frequentemente considerados em situações em que não há informações suficientes sobre a distribuição dos erros, ou quando os erros têm distribuições que são muito complexas para serem modeladas adequadamente por uma distribuição paramétrica.

Neste subcapítulo é abordada a estimação de parâmetros de modelos de espaço de estados univariados, ou seja, as observações e os estados são variáveis unidimensionais. Considera-se o MEE cujo processo de estado segue um processo autorregressivo de ordem  $p$ ,  $AR(p)$ , ou seja,

$$Y_t = w_t\beta_t + e_t, \quad (3.53)$$

$$\Phi(B)(\beta_t - \mu) = \varepsilon_t, \quad (3.54)$$

em que  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ ,  $B$  é um operador de atraso, ou seja,  $B(\beta_t - \mu) = \beta_{t-1} - \mu$  e  $w_t$  representa uma constante não nula conhecida para cada tempo  $t$ . Considera-se ainda as seguintes condições para a Eq.(3.54) e Eq.(3.53) definidas em Costa (2006):

- 1.  $\phi_i$ , com  $i = 1, 2, \dots, p$  são tais que todas as raízes do polinómio  $x^p - \phi_1 x^{p-1} - \dots - \phi_p$  estão dentro do círculo unitário;
- 2. Os erros  $e_t$  e  $\varepsilon_t$  são ruídos brancos não correlacionados com variâncias  $var(e_t) = \sigma_e^2$  e  $var(\varepsilon_t) = \sigma_\varepsilon^2$ ;
- 3. Os erros  $e_t$  e  $\varepsilon_t$  possuem os momentos de 4ª ordem constantes com  $var(e_t^2) = \tau_e^2$  e  $var(\varepsilon_t^2) = \tau_\varepsilon^2$ .

As condições estipuladas estabelecem um MEE em que a equação de transição representa um processo autorregressivo estacionário de ordem  $p$ . É comum que não haja correlação entre os erros nas equações de observação e de transição. A condição 3. é essencial para garantir que os estimadores sejam consistentes. Os modelos estabelecidos nas Eq.(3.54) e Eq.(3.53) podem ainda ser reescritos de outra forma, uma vez que um processo  $AR(p)$  é representado sobre a forma de um vetor autorregressivo de ordem 1. Assim, o MEE pode ser reescrito pelas equações,

$$Y_t = \mathbf{W}_t \beta_t + e_t, \quad (3.55)$$

$$\beta_{t-1} = \mu + \phi(\beta_t - \mu) + \varepsilon_t. \quad (3.56)$$

Neste subcapítulo apresentam-se estimadores com uma forma funcional explícita e que não estão sujeitas ao pressuposto das distribuições dos erros, mas que têm por base as propriedades estocásticas inerentes à estrutura do MEE.

## a) Estimador para a média

O estimador para a média  $\mu$  para uma série temporal de dimensão  $n$ ,  $Y_1, Y_2, Y_3, \dots, Y_n$ , gerada pelo modelo de espaço de estados é dado por

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n \frac{Y_t}{w_t}. \quad (3.57)$$

Se considerar o modelo em que a equação de estado representa um processo  $AR(1)$ , o estimador  $\hat{\mu}$  deverá ser mais eficiente à medida que o parâmetro de transição da equação de estado,  $\phi$ , estiver

mais próximo de zero. Nesse caso, a variância do processo  $\beta_t, t \in \mathbb{N}$  será menor, o que estará em concordância com a abordagem subjacente na construção do estimador.

## b) Estimador para os parâmetros autorregressivos

Costa (2006) refere que uma alternativa para a estimação é estimar os parâmetros autorregressivos de forma a obter menores erros de previsão para  $Y_t$ . Ou seja, os parâmetros  $\phi_i$  serem escolhidos de forma a minimizar a média dos quadrados das inovações. No entanto, esta abordagem é difícil de implementar computacionalmente, uma vez que a estimativa dos parâmetros  $\phi_i$  é a última etapa do processo de estimação, e as variâncias dos erros precisam ser estimadas com base em valores iniciais dos parâmetros autorregressivos. Além disso, a minimização da média dos quadrados das inovações deve respeitar as restrições do espaço de parâmetros, tanto para os  $\phi_i$  quanto para as variâncias.

Em alternativa, os estimadores propostos para os parâmetros autorregressivos da Eq.(3.54) de transição são baseados na estrutura de covariâncias do processo  $\{Y_t/w_t\}_{t \in \mathbb{N}}$ . Sob a condição da não correlação dos erros, a função de autocovariância do processo  $\{Y_t/w_t\}_{t \in \mathbb{N}}$  que designamos por  $\gamma_{w^{-1}Y}$ , coincide com a função de autocovariância do processo  $\{\beta_t\}_{t \in \mathbb{N}}$ , Costa (2006). Um estimador para  $\phi$ , considerando a propriedade  $l > p + 1$  e aplicando a regressão linear múltipla, surge

$$\hat{\phi} = [\hat{G}'(l)\hat{G}(l)]^{-1} \hat{G}'(l)\hat{\gamma}(l), \quad (3.58)$$

com  $l > p + 1$ , uma matriz  $\hat{G}$  definida em Costa (2006) e

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-k} \left( \frac{y_{t+k}}{w_{t+k}} - \hat{\mu} \right) \left( \frac{y_t}{w_t} - \hat{\mu} \right). \quad (3.59)$$

A função de autocovariância pode ainda ser estimada por

$$\hat{\hat{\gamma}}(k) = \frac{n}{(n-k)} \hat{\gamma}(k). \quad (3.60)$$

É sugerido em Costa & Alpuim (2010) que, para obter boas estimativas, a dimensão da amostra deve ser no mínimo 50 e  $k \geq \frac{n}{4}$ . Em Costa (2006) é realizado um estudo de simulação que analisa a influência da estimação para função de autocovariância pelos estimadores (3.59) e (3.60). A simulação conclui uma regra prática para determinar o ótimo valor de  $l$ , para uma determinada amostra de dimensão  $n$ . Os valores escolhidos encontram-se na Tabela 3.

Tabela 3: Passo máximo da função de autocovariância para a estimação de  $\phi$  escolhido em função da dimensão da amostra

Dimensão	$l$ escolhido	$\frac{l}{n} \times 100\%$
50	45	90
100	80	80
200	60	30
500	50	10

### c) Estimadores para as variâncias

Os estimadores para  $\sigma_\varepsilon^2$  e  $\sigma_\varepsilon^2$  têm por base o método dos momentos e para o caso particular de  $\sigma_\varepsilon^2$ , a estacionariedade do processo de estados  $\{\beta_t\}_{t \in \mathbb{N}}$ ,

$$\hat{\sigma}_\varepsilon^2 = \frac{(1 - \hat{\phi}^2) \sum_{k=1}^{l_\varepsilon} \hat{\gamma}_k \hat{\phi}^k}{\hat{\phi}^2 (1 - \hat{\phi}^{2l_\varepsilon})}, \quad (3.61)$$

e

$$\hat{\sigma}_\varepsilon^2 = \left\{ \sum_{t=1}^n \left( \frac{Y_t}{w_t} - \hat{\mu} \right)^2 - n \frac{\hat{\sigma}_\varepsilon^2}{1 - \hat{\phi}^2} \right\} \left( \sum_{t=1}^n \frac{1}{w_t^2} \right)^{-1}. \quad (3.62)$$

onde  $l_\varepsilon = 1, 2, \dots$ , por regra, segundo o estudo de simulação realizado em Costa (2006), considera-se  $l_\varepsilon = 1$ .

## 3.5 Bootstrapping

*Bootstrap* é um método estatístico de reamostragem que envolve a criação de múltiplas amostras a partir de uma amostra original, permitindo assim estimar a distribuição de uma estatística amostral sem a necessidade de assumir uma distribuição paramétrica para os dados. Essa técnica foi introduzida por Beran (1992). O termo *Bootstrap* é associado à lenda do Barão *Von Munchhausen*, onde numa situação em que estava preso no pântano, o barão resolveu o problema utilizando os seus próprios recursos. O qual sugere que é possível obter estimativas precisas de um conjunto de dados a partir de uma única amostra, sem a necessidade de grandes pressupostos teóricos. Desde então, o *Bootstrap* tem sido amplamente utilizado em Estatística, Ciência de Dados e *Machine Learning* para estimar incertezas e avaliar a qualidade de modelos e métodos estatísticos. O *Bootstrap* também é utilizado para construir intervalos de confiança e decidir em testes de hipóteses. É, também, aplicado em outras áreas como

Finanças, Biologia e Engenharia. A metodologia anterior ao *Bootstrap* era o *Jackknife* proposto por Beran (1992) que revolucionou todas as técnicas semelhantes até lá existentes.

### 3.5.1 Conceito de Bootstrap

Considere uma sequência de variáveis aleatórias  $X_1, X_2, \dots, X_n$  independentes e identicamente distribuídas com uma determinada distribuição  $F$  desconhecida, com valor médio  $\mu$  e variância  $\sigma^2$  também desconhecidos. Considere ainda  $x_1, x_2, \dots, x_n$  uma amostra designada de população *Bootstrap*. Desta amostra retiramos com reposição um,  $B$ , número de amostras, que designamos de amostra *Bootstrap*. Por outras palavras pode-se calcular  $f^1(x), f^2(x), \dots, f^B(x)$ . Calcula-se a média de todas as previsões obtendo

$$f_{bag}(x) = \frac{1}{B} \sum_{b=1}^B f^{*b}(x). \quad (3.63)$$

O método *Bootstrap* pode ser aplicado a modelos de espaço de estados para avaliar a incerteza nas estimativas dos parâmetros do modelo e nas previsões. Ao aplicar o método *Bootstrap* a um modelo de espaço de estados, podemos gerar múltiplas amostras de dados a partir da amostra original, simulando a variabilidade dos dados observados. Em seguida, ajusta-se o modelo aos dados reamostrados para estimar os parâmetros do modelo e fazer previsões. Repetindo esse processo muitas vezes, podemos obter uma distribuição de estimativas de parâmetros e previsões, permitindo avaliar a incerteza associada a essas estimativas. Ainda é possível comparar as distribuições de estimativas de parâmetros e previsões para diferentes modelos ajustados aos dados reamostrados. Isso pode ajudar a evitar o *overfitting*, um problema comum em modelação estatística, onde o modelo se ajusta demasiado aos dados de treino e, portanto, não generaliza bem para novos dados.

### 3.5.2 Bootstrapping para MEE

Embora na Secção 3.4.1 se tenha discutido o facto de que, em condições gerais, a estimação pela máxima verosimilhança dos parâmetros de um modelo de espaço de estados são consistentes e assintoticamente normais, o número de observações das séries temporais é frequentemente reduzido. Segundo Shumway et al. (2000), vários investigadores encontraram evidências de que as séries temporais devem ser de grande dimensão para que os resultados assintóticos sejam aplicáveis. Determinados problemas ocorrem quando os parâmetros aproximam-se da fronteira do espaço de parâmetros. Neste subcapítulo é discutido um algoritmo para *Bootstrap* de MEE. É pertinente também referir que este algoritmo aplica-se igualmente aos modelos *ARMAX* (*autoregressive moving average*

with exogenous inputs). Atendendo às Eq.(3.8), Eq.(3.9) e Eq.(3.42) do FK define-se a representação *innovation form* dos modelos *ARMAX* da seguinte forma

$$\beta_{t|t+1} = \Phi\beta_{t|t-1} + \Upsilon\mu_{t+1} + K_t\eta_{t|t-1}, \quad (3.64)$$

$$Y_t = W_t\beta_{t|t-1} + \Gamma\mu_t + \eta_{t|t-1}, \quad (3.65)$$

considere  $\Gamma$  uma matriz de dimensões  $p \times r$  e  $\Upsilon$  uma matriz  $q \times r$ ; qualquer uma destas matrizes pode ser uma matriz zero. Estas funções acomodam o algoritmo *bootstrapping*.

É pertinente mencionar que as inovações,  $\eta_{t|t-1}$ , ou os erros de previsão da Eq.(3.42) são não correlacionados mas são heterocedásticos, uma vez que  $\Sigma_t$  pode variar com o tempo  $t$ . Assim, num procedimento de reamostragem, podemos ignorar os primeiros valores de  $\eta_{t|t-1}$  até  $\Sigma_t$  estabilizar, ou utilizar inovações padronizadas

$$\eta_{t|t-1}^* = \Sigma_t^{-\frac{1}{2}}\eta_{t|t-1}. \quad (3.66)$$

onde  $\Sigma_t^{\frac{1}{2}}$  denota a matriz raiz quadrada única de  $\Sigma_t$  definida por  $\Sigma_t^{\frac{1}{2}}\Sigma_t^{\frac{1}{2}} = \Sigma_t$ .

Seguidamente baseia-se a metodologia *Bootstrap* nas inovações estandardizadas (ou padronizadas), tendo em conta, que ignorar os valores iniciais é necessário, Shumway et al. (2000). Os coeficientes do modelo e as suas estruturas de correlação são padronizadas por  $k \times 1$  vetor dos parâmetros  $\Theta_0$ , isto é,

$$\Phi = \Phi(\Theta_0), \Upsilon = \Upsilon(\Theta_0),$$

$$Q = Q(\Theta_0), W_t = W_t(\Theta_0), \Gamma = \Gamma(\Theta_0),$$

e ainda com  $R = R(\Theta_0)$  expressa-se as inovações através da função log verosimilhança Eq.(3.44). Note que não é necessário que o modelo (3.44) seja gaussiano para considerar como um critério para estimar parâmetros. Considera-se  $\hat{\Theta}$  a componente que estima a máxima verosimilhança de  $\Theta_0$ , ou seja,  $\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} L_y(\Theta) + \eta_{t|t-1}(\hat{\Theta})$  e  $\Sigma_t(\hat{\Theta})$  definem os valores da inovação obtidos pelo FK de  $\hat{\Theta}$ .

O método de *Bootstrap* não paramétrico obtém-se através dos seguintes passos,

1. Construir a função de inovações estandardizadas

$$\eta_{t|t-1}^*(\hat{\Theta}) = \Sigma_t^{-\frac{1}{2}}(\hat{\Theta})\eta_{t|t-1}(\hat{\Theta}); \quad (3.67)$$

2. Obter uma amostra com reposição de dimensão  $n$ , a partir do conjunto das inovações estandardizadas  $\eta_{t|t-1}^*$ , com  $t = 1, \dots, n$ , com o propósito de obter a sequência bootstrap  $\{\eta_1^{*B}(\hat{\Theta}), \dots, \eta_n^{*B}(\hat{\Theta})\}$ ;

3. Definir o vetor  $\xi_t = (\beta_{t+1|t}, Y_t)$ , empilhando<sup>5</sup> Eq.(3.65) e Eq.(3.64), obtém-se uma equação de primeira ordem vetorial para  $\xi_t$  dada por,

$$\xi_t = \Phi_t \xi_{t-1} + G \mu_t + W_t \eta_{t|t-1}, \quad (3.68)$$

em que

$$\Phi_t = \begin{bmatrix} \phi & 0 \\ W_t & 0 \end{bmatrix}, G = \begin{bmatrix} \Gamma \\ \Upsilon \end{bmatrix}, W_t = \begin{bmatrix} K_t \Sigma_t^{-\frac{1}{2}} \\ \Sigma_t^{-\frac{1}{2}} \end{bmatrix}, \quad (3.69)$$

para construir um conjunto de dados *Bootstrap*, é necessário resolver (3.69) utilizando  $\eta_{t|t-1}^B = \Sigma_t^{-\frac{1}{2}} \eta_{t|t-1}^{*B}(\hat{\Theta})$ , em vez de  $\eta_{t|t-1}$ . A variável  $\mu_t$  e as condições iniciais do FK permanecem fixas (ou inalteradas) para os valores que lhes foram atribuídos, o vetor de parâmetros é fixo em  $\hat{\Theta}$ ;

4. Através da série temporal *Bootstrap*  $\{Y_t^B\}_{t=1, \dots, n}$  obtém-se as estimativas *distribution-free Bootstrap*;
5. Por fim, volta-se a repetir todos os passos a partir de 2.,  $B$  quantidades de vezes, obtendo um conjunto de estimativas de parâmetros *Bootstrap*  $[\hat{\Theta}_b^*, b = 1, \dots, B]$ . A distribuição da amostra de  $\hat{\Theta} - \Theta_0$  aproxima-se pela distribuição de  $\hat{\Theta}_b^* - \hat{\Theta}, b = 1, \dots, B$ .

## 3.6 Avaliação dos modelos

Em Estatística não existe um modelo definitivo que consiga explicar a realidade e o comportamento dos dados. Por essa razão que é necessário ajustar e testar vários modelos, uma vez que uns conseguem representar melhor a estrutura inerente dos dados que outros, como também, há outros que conseguem explicar de forma mais simplificada a realidade, realizar previsões e incorporar determinadas componentes presentes nos dados.

É necessário existir determinadas métricas que permitam avaliar os modelos. Ao utilizar métricas de avaliação pode-se comparar diferentes modelos, identificar o que apresenta melhores resultados e avaliar o quão eficaz um modelo está a alcançar os objetivos específicos da análise das séries temporais.

### 3.6.1 Medidas de avaliação

A escolha das medidas de avaliação é crucial, uma vez que tem impacto direto na compreensão das vantagens e desvantagens de um modelo. Diferentes métricas são apropriadas para diferentes tipos

<sup>5</sup>O método de empilhamento ou *stacking*, é uma técnica de aprendizagem automática que combina várias previsões de modelos individuais para produzir uma previsão final mais precisa. É uma técnica de modelação estatística que combina diferentes modelos para obter melhores resultados de previsão.

de tarefas de previsão de séries temporais, e a escolha pode depender das características específicas do conjunto de dados e dos objetivos da análise. Estas medidas são utilizadas para verificar o nível de precisão que o modelo de previsão possui, uma vez que, quanto mais próximo as previsões forem aos dados observados de uma série menor será o erro da previsão.

É necessário salientar que nenhuma medida é consensual para ser aplicada a todas as situações de previsão, sendo assim, é necessário utilizar vários critérios para avaliar a qualidade preditiva de um modelo ou método de previsão. Devem ser considerados vários fatores na escolha de um método de previsão, o que inclui, a precisão das previsões, facilidade de utilização, de interpretação e de implementação do método, a poupança de custos, a flexibilidade, entre outros.

Estas medidas de avaliação podem ser classificadas em três categorias, medidas dependentes da escala, medidas baseadas em erros percentuais e medidas de erros escalados.

## a) Medidas dependentes da escala

Conforme o próprio nome indica, as métricas de precisão são sensíveis à escala dos dados utilizados. São amplamente utilizadas e bastante úteis para comparar diferentes métodos aplicados a um mesmo conjunto de dados. No entanto, é importante ter em mente que não devem ser escolhidas para a comparação de dados com escalas distintas. Quando é necessário comparar conjuntos de dados com diferentes escalas, a utilização de métricas dependentes da escala pode levar a conclusões enganosas. Isso ocorre porque os erros absolutos ou quadráticos podem ser influenciados pela magnitude dos valores dos dados. Nesses casos, é mais adequado recorrer a métricas independentes da escala, como aquelas baseadas em erros percentuais.

- **Erro Médio (EM):** Calcula o valor médio dos desvios entre os valores observados  $Y_t$  e as previsões  $\hat{Y}_t$  para os instantes  $t = 1, 2, \dots, n$ ,

$$EM = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t). \quad (3.70)$$

No entanto, o Erro Médio (EM) não fornece uma indicação precisa sobre o tamanho dos erros típicos, uma vez que os erros positivos e negativos tendem a se compensar mutuamente. Consequentemente, é provável que o EM resulte em um valor pequeno. Na realidade, essa métrica apenas indica se há uma subestimação ou sobrestimação sistemática (viés de previsão).

- **Erro Quadrático Médio (EQM):** Representa o valor médio dos desvios ao quadrado entre os

valores observados  $Y_t$  e as previsões  $\hat{Y}_t$  para os instantes  $t = 1, 2, \dots, n$ ,

$$EQM = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2. \quad (3.71)$$

Para além da Eq.(3.71) também existe a raiz do erro quadrático médio, ou,  $REQM = \sqrt{EQM}$ <sup>6</sup>, esta é uma das medidas de avaliação mais utilizadas para medir o desempenho dos modelos em estudo.

- **Erro Absoluto Médio (EAM):** Define o valor absoluto médio dos desvios entre os valores observados  $Y_t$  e as previsões  $\hat{Y}_t$  para os instantes  $t = 1, 2, \dots, n$ ,

$$EAM = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|. \quad (3.72)$$

A medida de avaliação Eq.(3.72) em comparação com as anteriores é relativamente mais fácil de interpretar. O modelo que apresentar um menor valor EAM é considerado o mais adequado.

## b) Medidas baseadas em erros percentuais

O erro percentual é calculado por  $p_t = 100 \times \frac{(Y_t - \hat{Y}_t)}{Y_t}$ . Os erros percentuais possuem a vantagem de serem independentes da escala, tornando-os frequentemente utilizados para comparar o desempenho de previsões entre diferentes séries temporais. No entanto, estas medidas possuem desvantagens, sendo que podem resultar em valores infinitos ou indefinidos se  $Y_t = 0$  para qualquer  $t$ . Além disso, apresentam uma distribuição extremamente assimétrica quando  $Y_t$  está próximo de zero.

- **Erro Percentual Médio (EPM):** Representa o valor percentual médio dos desvios entre os valores observados  $Y_t$  e as previsões  $\hat{Y}_t$  para os instantes  $t = 1, 2, \dots, n$

$$EPM = \frac{1}{n} \sum_{t=1}^n p_t = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t} \times 100. \quad (3.73)$$

Semelhante ao (3.70) é esperado que o EPM seja um valor pequeno, uma vez que os erros percentuais positivos e negativos tendem a compensar-se mutuamente.

- **Erro Percentual Absoluto Médio (EPAM):** Indica o valor percentual absoluto médio dos desvios entre os valores observados  $Y_t$  e as previsões  $\hat{Y}_t$  para os instantes  $t = 1, 2, \dots, n$ ,

$$EPAM = \frac{1}{n} \sum_{t=1}^n |p_t| = \frac{1}{n} \sum_{t=1}^n \left| \frac{(Y_t - \hat{Y}_t)}{Y_t} \right| \times 100. \quad (3.74)$$

<sup>6</sup>REQM é uma medida utilizada para reduzir a grandeza dos valores do erro médio quadrático

Um outro problema das medidas baseadas em erros percentuais é o fato de que estas assumem um zero significativo. Por exemplo, não faz sentido aplicar esse tipo de erro para medir a precisão de previsões de temperatura em escalas Fahrenheit ou Celsius, onde o zero representa apenas um ponto específico na escala, não uma ausência completa da propriedade sendo medida.

### c) Medidas de erros escalados

É uma alternativa às medidas baseadas em erros percentuais propostas por Hewamalage et al. (2021), uma vez que são independentes da escala, e por isso, podem ser utilizadas para comparar métodos diferentes de previsão aplicados a uma série e também para comparar a precisão das previsões entre diversas séries. A ideia por trás dessas medidas consiste em “escalar” os erros de previsão dividindo-os pelo Erro Absoluto Médio (EAM) calculado *in-sample*, ou seja, na amostra de treino usando o método *naïve*<sup>7</sup>. O erro escalado tem a seguinte forma,  $q_t = (Y_t - \hat{Y}_t)/Q_t$ , em que  $Q_t$  representa o erro absoluto médio ou EAM. Para séries temporais sem sazonalidade, pode-se definir o erro escalado com a seguinte expressão:

$$q_t = \frac{Y_t - \hat{Y}_t}{\frac{1}{n-s} \sum_{t=2}^n |Y_t - \hat{Y}_{t-1}|}$$

Séries temporais que tenham presente uma componente sazonal, o erro escalado utiliza previsões de *naïve* sazonais,

$$q_t = \frac{Y_t - \hat{Y}_t}{\frac{1}{n-s} \sum_{t=s+1}^n |Y_t - \hat{Y}_{t-s}|}$$

em que  $s$  representa o período da sazonalidade.

- **Erro Escalado Absoluto Médio (EEAM):** Calcula a média dos erros escalados absolutos,

$$EEAM = \frac{1}{n} \sum_{t=1}^n |q_t|. \quad (3.75)$$

Quando o EEAM é inferior a 1, as previsões obtidas são de maior precisão, em média, do que as previsões do método de *naïve*. Considera-se o modelo mais preciso o que tiver menor EEAM.

O Erro Escalado Absoluto Médio (EEAM) é uma métrica que se presta para comparar o desempenho das previsões entre séries temporais com diferentes escalas. Além disso, possui uma escala significativa e é menos sensível a valores discrepantes. A única situação em que essas medidas se tornam infinitas

<sup>7</sup>O método de *naïve* é uma abordagem de previsão simples que assume que o valor futuro de uma série temporal será igual ao último valor observado no conjunto de dados de treinamento. Em outras palavras, essa técnica pressupõe que não haverá mudanças significativas na série ao longo do tempo e que o próximo valor será uma repetição do valor mais recente disponível, ou seja,  $Y_{t+1} = Y_t, t = 1, 2, \dots, n$ , onde  $n$  é a dimensão da amostra

ou indefinidas é quando todas as observações são iguais. Considerando as limitações dessas métricas de avaliação, é fundamental ter uma atenção especial ao escolhê-las, a fim de evitar tirarem-se conclusões equivocadas.

### 3.6.2 Critérios de seleção

Em qualquer análise estatística de dados, é possível encontrar vários modelos adequados para descrever o fenômeno em estudo. Especificamente, na modelação de uma série temporal, pode haver mais de um modelo que atenda aos diferentes critérios de avaliação do diagnóstico, tornando a tarefa de escolher o melhor modelo uma tarefa desafiadora. Portanto, é importante procurar critérios de seleção de modelos que levem em consideração as estatísticas baseadas nos resíduos do modelo ajustado. Essas estatísticas auxiliam na avaliação da qualidade do ajuste do modelo aos dados e ajudam a identificar o modelo mais apropriado para a série temporal em questão. Existem certos critérios de avaliação que medem a qualidade de ajustamento dos modelos, que têm por base a maximização da função da verosimilhança, esta penaliza os modelos com o maior número de parâmetros.

- **Critério de Informação de Akaike (AIC):** Proposto por Akaike (1974), é um dos critérios de informação mais utilizados,

$$AIC = -2 \log L(\hat{\Theta}) + 2p. \quad (3.76)$$

Em que  $L(\hat{\Theta})$  representa a função de verosimilhança e  $p$  é o número de parâmetros do modelo.

- **Critério de Informação Bayesiano (BIC):** Proposto por Schwarz (1978) é semelhante ao AIC, com a diferença em que a penalização é maior na inclusão de parâmetros adicionais a serem estimados,

$$BIC = -2 \log L(\hat{\Theta}) + p \log(n). \quad (3.77)$$

Em que  $L(\hat{\Theta})$  representa a função de verosimilhança,  $p$  é o número de parâmetros do modelo e  $n$  é a dimensão da amostra.

Ao contrário do AIC, o BIC é dependente da dimensão da amostra  $n$ . Quando  $\log(n) < 2$ , ou seja, quando a amostra possui uma dimensão superior a 7, a penalização do BIC é maior do que a penalização do AIC. Isso significa que, em geral, a minimização do BIC resulta na seleção de modelos com um menor número de parâmetros em comparação com a minimização do critério AIC. Essa abordagem evita, de certa forma, a sobrestimação do número de componentes do modelo, proporcionando uma seleção mais precisa dos parâmetros para a série temporal em questão. Segundo estes critérios seleciona-se o modelo que apresentar menores valores de AIC ou BIC.

### 3.6.3 Análise dos resíduos

A análise dos resíduos permite avaliar a qualidade do ajuste de um modelo e identificar observações que não são bem explicadas pelo modelo. É esperado que um bom modelo, em termos de qualidade de ajuste, produza resíduos com um comportamento semelhante ao de um ruído branco no caso de modelos de regressão linear e resíduos padronizados no caso de MEE. Isso significa que eles devem satisfazer os pressupostos de média zero, variância constante e ausência de correlação. Além disso, é importante verificar se os resíduos seguem uma distribuição aproximadamente Normal. Essa condição é necessária para a construção de intervalos de previsão e para verificar as expressões previamente obtidas para estimadores de máxima verosimilhança e *Bootstrap*. Daqui em diante, sempre que são mencionados os resíduos, estão a ser referidos os resíduos no caso de modelos de regressão linear e os resíduos padronizados no caso dos MEE.

- **Normalidade**

A análise da suposição de normalidade dos resíduos pode ser realizada por meio de representações gráficas, como o histograma, que deve assemelhar-se ao comportamento da função densidade de uma distribuição Normal, e o *QQ-plot*, que relaciona os quantis de uma amostra com os quantis teóricos de uma distribuição Normal, no caso deste ser um dos pressupostos do modelo. No *QQ-plot*, os pontos devem estar próximos de uma linha reta, o que permite visualizar desvios da normalidade. Além da análise gráfica, a normalidade dos resíduos pode ser verificada por meio de testes estatísticos, como o teste de Shapiro-Wilk, mais adequado para amostras de pequenas dimensões (com menos de 30 observações), e o teste de Kolmogorov-Smirnov.

- **Não correlação**

Assim como na avaliação da normalidade, a análise da não correlação dos resíduos pode ser realizada por meio de inspeção gráfica usando a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP) das amostras. A estrutura de correlação dessas funções deve-se assemelhar à de um ruído branco, com autocorrelações que sejam estatisticamente próximas de zero, mantendo-se dentro dos limites do intervalo  $\pm 1,96 \sqrt{n}$ . Além disso, a avaliação do comportamento dos resíduos pode ser feita usando testes estatísticos, como o teste de Ljung-Box. Este teste verifica se as primeiras  $k$  autocorrelações são conjuntamente nulas. A estatística de teste  $Q$  é calculada da seguinte forma,

$$Q = n(n + 2) \sum_{i=1}^k \frac{\hat{\rho}_i^2}{n - i} \quad (3.78)$$

em que  $Q \sim \chi_{k-p}^2$ , onde  $n$  representa o número de observações,  $p$  o número de parâmetros a serem estimados e  $k$  corresponde ao número de autocorrelações testadas. Avalia-se a hipótese nula  $H_0$ : "Os erros não são correlacionados". No trabalho Hassani & Yeganegi (2020) é sugerido que  $k = 10$  para dados sem uma componente sazonal e  $k = 2m$  para dados com uma componente sazonal, em que  $m$  corresponde ao período sazonal.

- **Média nula e variância constante**

Além das duas condições previamente mencionadas que os resíduos devem atender, estes também devem possuir média nula e variância constante, ou seja, devem exibir homocedasticidade. Para testar se a média dos resíduos é igual a zero, pode-se usar o teste t para a média. É importante observar que este teste só pode ser aplicado se os pressupostos de normalidade tiverem sido confirmados. No entanto, se a distribuição dos resíduos for aproximadamente simétrica e a amostra tiver mais de 20 observações, então o teste t pode ser aplicado de forma satisfatória. Para verificar a condição de homocedasticidade, é possível realizar uma análise gráfica observando a representação dos resíduos ao longo do tempo.

## 4 Análise e modelação de dados económicos

### 4.1 Dados económicos

O índice *ISM Manufacturing* é um índice de gestão de compras que mede a atividade da produção industrial dos EUA. É composto por uma sondagem realizada a gestores de compras de empresas de fabricação (indústria de produção) por todo o país.

As empresas são divididas em 18 categorias diferentes, alimentos e bebidas, produtos químicos, máquinas, equipamentos de transporte, entre outros. Em resumo, os dados provenientes do *ISM Manufacturing* (em particular o PMI) permite realizar uma avaliação abrangente do desempenho da indústria de fabrico dos EUA.

A base de dados nomeada de “ISM” tem 785 observações, com os valores do ISM mensal e as suas respetivas datas. O registo do valor do índice está disponível desde janeiro de 1957 até maio de 2022 (não havendo até ao início da investigação acesso a novos valores). Os valores são apresentados com um periodicidade mensal.

### 4.2 Análise exploratória dos dados

O primeiro passo foi construir uma série temporal com os valores do índice do ISM. A série começa em janeiro de 1975 e acaba em maio de 2022 com um horizonte temporal de 12 meses e 569 observações. As observações anteriores a este ano não são consideradas porque não refletem corretamente o comportamento da indústria de produção dos EUA, o que poderia dificultar o ajuste de um modelo aos dados. Na Figura 3 está representada a série temporal dos valores do índice ISM.

A linha vermelha no gráfico assinala o valor 52,90, o valor médio, sendo que essa observação permite determinar que nos últimos 48 anos a atividade industrial dos EUA tem expandindo apesar de ser visível ciclos em que a indústria contrai, face ao mês anterior, durante determinados períodos para valores abaixo de 40. Estes períodos, de modo geral, representam meses em que a economia americana enfrentou períodos de regressão económica. Com a visualização do gráfico, observa-se que a série não é estacionária em variância. Existem diversas variações por toda a série, mais notoriamente na década de 1980, entre 2007 e 2010, como também entre 2019 e 2022.

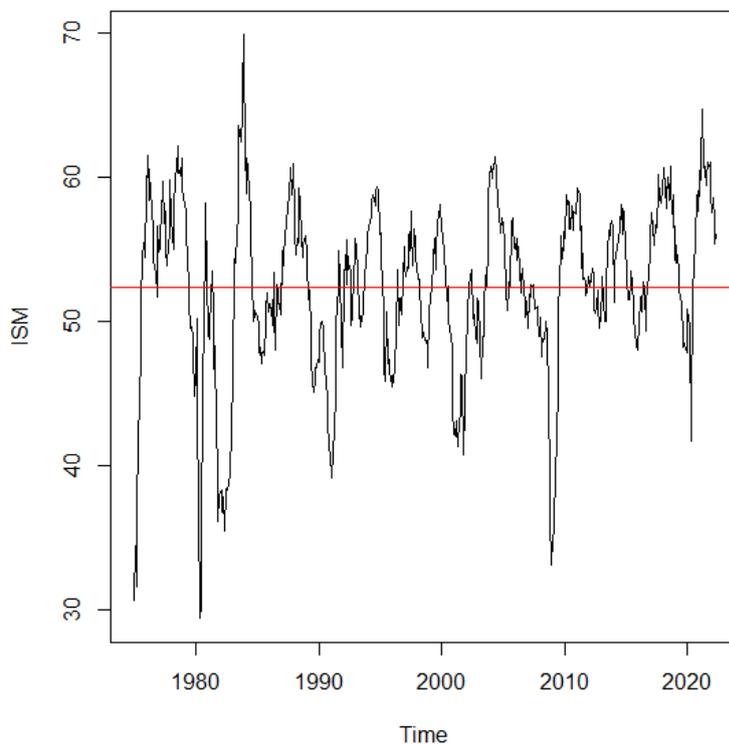


Figura 3: Série Temporal do índice ISM de 197 a 2022

Tabela 4: Estatísticas descritivas da série do índice ISM

Min.:	1ºQ.:	Mediana:	Média:	3º Q.:	Max.:	Desvio Padrão
29,4	49,5	52,9	52,27	56,4	69,9	6,12

O valor mínimo de 29,40 em abril de 1980 reflete um período em que a economia já se encontrava em recessão, nessa década a taxa de desemprego era cerca de 7,5% (Tabela 4). Tanto a recessão de 1980 como as recessões de 1981-82 foram desencadeadas por uma política monetária restritiva na tentativa de combater a crescente inflação. Durante as décadas de 1960 e 1970, economistas e responsáveis políticos acreditavam que poderiam reduzir o desemprego através de uma maior inflação, numa troca conhecida como *Phillips curve*<sup>8</sup>. Esta estratégia afetou severamente as empresas industriais

<sup>8</sup>A curva de Phillips é um conceito económico que demonstra a suposta relação inversa entre a taxa de desemprego e a taxa de inflação numa economia. De acordo com esta teoria, quando o desemprego é baixo, a inflação tende a aumentar, e quando o desemprego é elevado, a inflação tende a diminuir

norte americanas, Goodfriend & King (2005). Já o valor máximo observado de 69,90 é obtido após a recuperação da recessão mencionada.

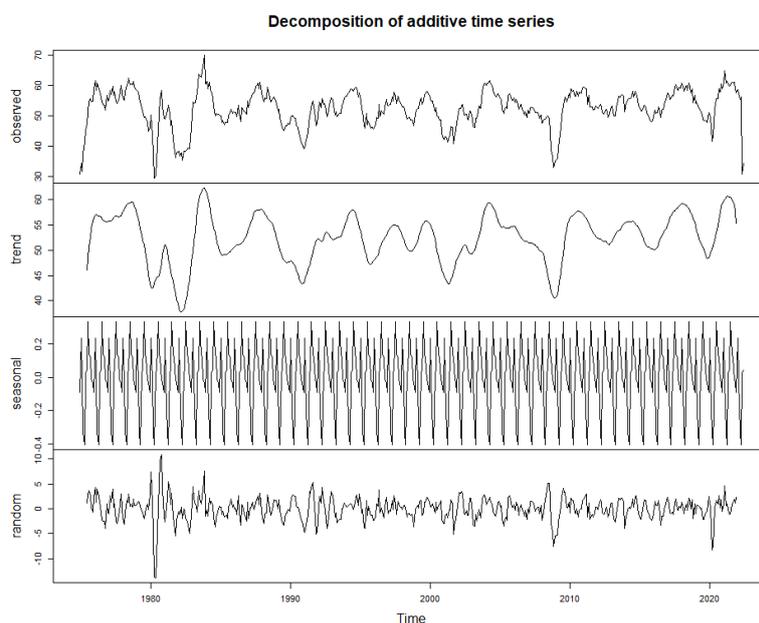


Figura 4: Decomposição da Série Temporal ISM

O gráfico da Figura 4 foi elaborado com o propósito de apresentar a série decomposta através das componentes da sazonalidade (*seasonal*), tendência (*trend*), e irregular (*random*). Da análise gráfica verifica-se que a componente sazonal varia aproximadamente entre  $-0,4$  e  $0,3$ , enquanto que a componente da tendência varia aproximadamente entre 35 e 65, evidenciando que a componente sazonal obtida por este procedimento tem valores de uma ordem de grandeza menor que a componente da tendência. A representação gráfica da componente da tendência não evidencia um padrão claro que seja identificado pela inspeção visual. Contudo, será de explorar a existência de uma tendência linear estatisticamente significativa através, por exemplo, de modelos lineares, uma vez que a economia americana está em constante crescimento Sachs & Warner (1997).

Atendendo ao diagrama em caixa de bigodes e ao histograma na Figura 5 podemos concluir que os valores observados deste índice estão compreendidos entre o mínimo 29,40 e o máximo 69,90. Há um número significativo de observações representadas como *outliers* (32 observações), não tendo qualquer significado concreto neste contexto por este representar valores de uma série temporal. Contudo, e apenas numa perspetiva económica, esta representação permite evidenciar a existência de observações que sendo inferiores a 40, indica, segundo os economistas, uma recessão económica, Goodfriend & King

(2005) e Bognanni & Young (2018).

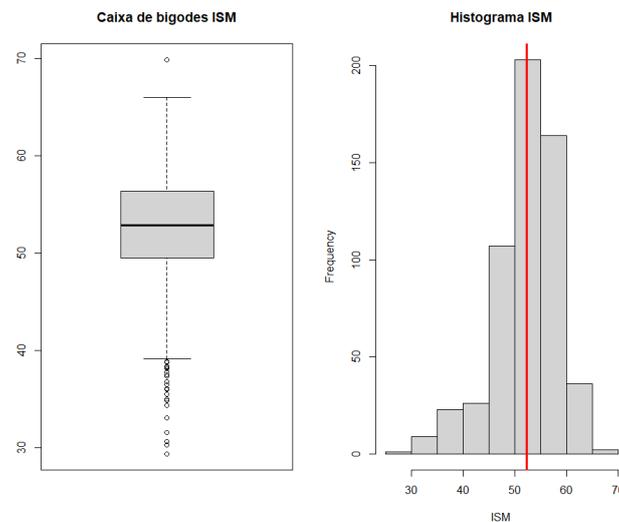


Figura 5: Histograma e diagrama em caixa de bigodes da Série Temporal do índice ISM.

## 4.3 Modelação

Após a análise exploratória dos dados foram observadas determinadas características que têm de ser consideradas no processo de modelação. Através do gráfico da Figura 4 observa-se que há uma eventual tendência, embora reduzida. Apesar de não se observar uma sazonalidade significativa o facto de os dados serem de natureza económica pode colocar-se a hipótese de que haja ciclos ou sazonalidade. Perante as características dos dados serão considerados os vários modelos descritos anteriormente. As 569 observações foram divididas em duas séries, uma série de treino com 552 observações para identificar e estimar modelos, e uma outra série de teste com as últimas 17 observações (meses) para avaliar a qualidade de previsão dos modelos.

### 4.3.1 Modelos de regressão linear simples

O modelo de regressão linear simples é uma forma de estabelecer uma relação linear entre uma variável independente (ou explicativa) e uma variável dependente (ou resposta). Neste contexto, o modelo de regressão linear simples é utilizado para ajustar uma função afim aos dados que represente a tendência subjacente. Além disso, será efetuada uma tentativa para modelar a sazonalidade, com um período de 12 meses, a fim de determinar se essa componente é estatisticamente significativa ou não.

## a) Tendência

Foi considerado um modelo de regressão linear simples (RLS) (o Modelo a)). Foi decidido ajustar um modelo com tendência. Para este modelo  $Y_t$  representa os valores do índice ISM ao longo dos meses (1975 a 2022),  $tempo_t$  é variável explicativa,  $\beta_0$  é a interseção da reta com o eixo vertical e  $\beta_1$  é o coeficiente de regressão. A Tabela 5 representa um sumário do modelo.

Tabela 5: Sumário do Modelo de Regressão Linear com variável Tempo

Coeficientes	$\beta_0$	tempo
Valor de prova	$2 \times 10^{-16}$	0,0046
Estimativa	50,8679	0,0045
Erro de padrão	0,5053	0,0016
Variância Residual	5,98	
EQM	35,0150	

Observa-se na Tabela 5 que o coeficiente da variável  $tempo_t$  é estatisticamente significativo. Logo a série apresenta uma tendência crescente fraca (0,0045). Também através do gráfico da Figura 6 observa-se, de facto, uma tendência crescente ao longo dos anos, embora fraca. O modelo de regressão linear ajustado na Figura 6 indica que essa tendência deve ser considerada. Para melhor avaliar o ajustamento do modelo é necessário uma análise aos resíduos deste.

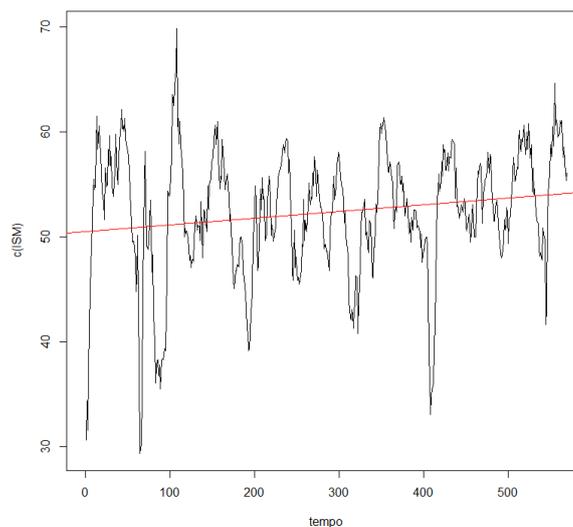


Figura 6: Modelo RLS com Tendência

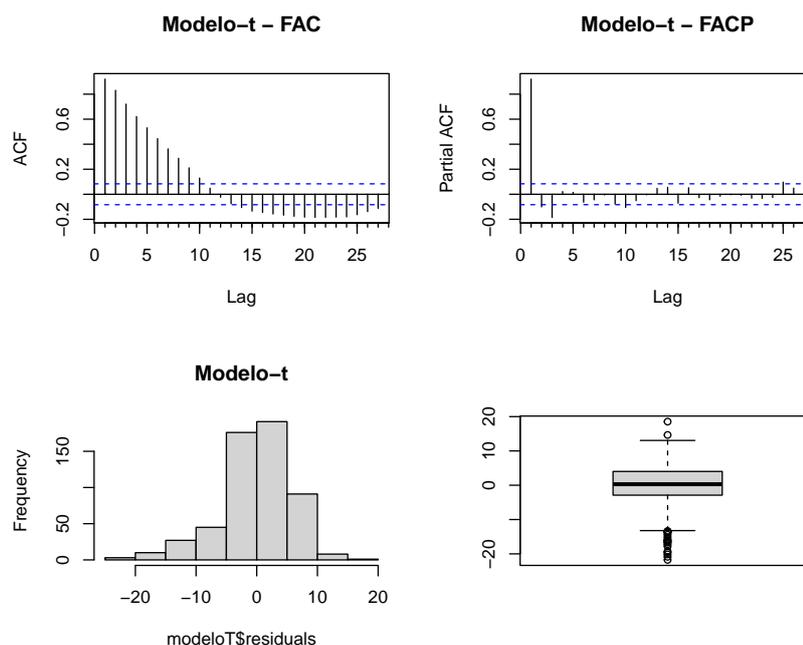


Figura 7: Histograma, diagrama de caixa de bigodes, FAC e FACP do modelo RLS com Tendência

Procede-se a uma análise dos resíduos deste modelo através de gráficos e alguns testes estatísticos. O histograma na Figura 7 indica que os resíduos não apresentam um comportamento semelhante ao de um processo de ruído branco. Os testes à normalidade de Shapiro-Wilk e Kolmogorov-Smirnov rejeitaram a hipótese nula da normalidade dos resíduos. A análise realizada aos gráficos da FAC e FACP revela que os resíduos apresentam, ainda, uma estrutura de correlação temporal compatível, eventualmente com um processo autorregressivo de ordem superior a um.

## b) Sazonalidade

A série de treino do ISM são de natureza económica, uma vez que mede o desempenho da indústria transformadora norte americana, o que permite concluir que, em princípio, poderá haver sazonalidade ou uma componente cíclica. Observando o gráfico da Figura 4 verifica-se que a componente sazonal é extremamente reduzida. Para avaliar a significância da componente sazonal ajustou-se um modelo com um conjunto de variáveis explicativas indicatrizes, para 11 meses, o coeficiente independente, a variável independente tempo (uma vez que já se verificou que há tendência) e a variável resposta a série ISM. As variáveis indicatrizes são também consideradas variáveis *dummies*, ou mudas<sup>9</sup>.

<sup>9</sup>Variáveis *dummy* são variáveis binárias criadas para representar as categorias de uma variável categórica. Para uma variável categórica com  $n$  categorias distintas, normalmente são criadas  $n - 1$  variáveis *dummy*. Isso ocorre porque uma

O modelo de regressão linear múltipla contempla o termo independente  $\beta_0$ , a variável tempo,  $t = 1, 2, \dots, n$ , 11 variáveis mudas,  $d_{t,i}$ , com  $i = 1, \dots, 11$ , que correspondem a variáveis indicatrizes de janeiro até novembro e o erro aleatório

$$Y_t = \beta_0 + \sum_{i=1}^{11} \beta_i d_{t,i} + \epsilon_t.$$

Criar um modelo com variáveis indicatrizes é uma técnica usada na modelação estatística para representar variáveis categóricas como variáveis binárias (0 e 1) num modelo de regressão. Variáveis categóricas representam características qualitativas que não possuem uma relação numérica ou ordem entre os seus valores. Para incorporar variáveis categóricas num modelo de regressão ou preditivo, estas precisam de ser transformadas num formato numérico que possa ser interpretado pelo modelo. As variáveis mudas, para este caso é pertinente uma vez que os dados são mensais, ou seja, é criada uma variável muda para cada mês com o propósito de se verificar se há sazonalidade e em que mês.

Tabela 6: Sumário do modelo de regressão com coeficientes sazonais e variável tempo

	Estimativas	Erro Padrão	Valor-t	Valor de prova
Constante	50,6512	0,9782	51,78	$2 \times 10^{-16}$
tempo	0,0064	0,0015	4,18	$3,45 \times 10^{-5}$
d_1	-0,5033	1,2290	-0,41	0,6823
d_2	-0,0347	1,2290	-0,03	0,9775
d_3	-0,3598	1,2290	-0,29	0,7698
d_4	-0,5662	1,2290	-0,46	0,6452
d_5	-0,5434	1,2290	-0,44	0,6585
d_6	-0,0894	1,2354	-0,07	0,9424
d_7	-0,0511	1,2354	-0,04	0,9670
d_8	0,2808	1,2354	0,23	0,8203
d_9	0,1213	1,2354	0,10	0,9218
d_10	0,1362	1,2354	0,11	0,9123
d_11	0,0234	1,2354	0,02	0,9849

A Tabela 6 representa as estimativas e os respetivos valores de prova relativos aos parâmetros estimados. Para um nível de significância de 5% apenas o termo independente e o coeficiente categoria é utilizado como referência, e as demais categorias são representadas pelas variáveis *dummy*.

associado à variável tempo é que têm valores de prova inferiores a 0,05, o que significa que os coeficientes sazonais não são estatisticamente significativos, quando considerada a sazonalidade anual com os 12 meses. Assim, a modelação através dos modelos lineares será considerada através do modelo de regressão linear simples ajustado anteriormente.

### 4.3.2 Modelos SARIMA - metodologia Box-Jenkins

Os modelos SARIMA são uma classe de modelos importante na análise e modelação de fenómenos que envolvem incerteza e aleatoriedade. Neste tipo de modelos, o foco é a captura da variabilidade e flutuações inerentes aos dados, reconhecendo que nem todos os fenómenos podem ser descritos pelos modelos de regressão linear.

Para a modelação através dos modelos SARIMA foi optada a metodologia de *Box-Jenkins*. Esta abordagem tem por objetivo identificar o melhor modelo  $ARMA(p, q)$ ,  $ARIMA(p, d, q)$  ou  $SARIMA(p, d, q)(P, D, Q)$  de ajustamento a uma série temporal. A metodologia segue um processo iterativo de três fases,

- 1. Identificação do modelo,
- 2. Estimação do modelo,
- 3. Validação do modelo.

A série temporal é dividida numa série de treino e noutra série designada de teste, esta última com o propósito de validar o modelo e realizar comparações para as previsões de uma forma independente da fase do ajustamento e da estimação dos parâmetros que ocorrem com a série de treino. Na identificação do modelo, a metodologia começa por testar se a série é estacionária e verificar se existe sazonalidade. Os gráficos das funções de autocorrelação e autocorrelação parcial são boas ferramentas para esta fase, além da aplicação de testes estatísticos.

Dos gráficos da FAC e da FACP da série temporal do índice ISM (Figura 8) conclui-se que a primeira função tem um decaimento lento para zero e a segunda indica a presença de uma estrutura autorregressiva. Assim, vários modelos terão de ser ajustados e avaliados quanto à qualidade do respetivo ajustamento.

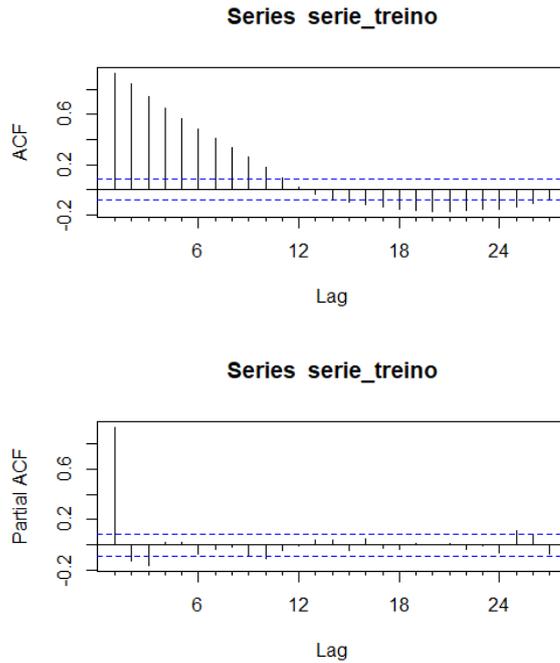


Figura 8: FAC e FACP da série temporal do índice ISM.

### a) Modelo ARIMA

Foram ajustados modelos com e sem sazonalidade, como foi mencionado na Secção 4.3.1. O primeiro passo foi ajustar um modelo *ARMA* aos resíduos do modelo com a tendência e foi testado com parâmetros autorregressivos de ordem 1, 2, e 3. No entanto, concluiu-se que estes modelos não eram adequados. Nos modelos *ARIMA* não está presente a componente sazonal, e representam-se por,

$$Y_t \sim ARIMA(p = ?, d = ?, q = ?).$$

Define-se  $Y_t$  como um processo autorregressivo e de médias móveis integrado *ARIMA*( $p, q, d$ ) quando assume uma expressão

$$\Phi_p(B)\nabla^d Y_t = \Theta_q(B)\varepsilon_t. \tag{4.79}$$

Para tal  $\nabla^d Y_t = (1 - B)^d Y_t$ , com  $d \geq 1$ , é a série estacionária depois de diferenciada  $d$  vezes,  $\Phi_1, \Phi_2, \Phi_3, \dots$  representam os parâmetros autorregressivos,  $\theta_1, \theta_2, \theta_3, \dots$  são os parâmetros de médias móveis,  $\Phi_p(B)$  e  $\Theta_p(B)$  são os polinómios autorregressivos e de médias móveis regulares.

A metodologia de Box-Jenkins exige o ajustamento de diversos modelos até ser descoberto o melhor processo para um modelo ARIMA. O critério da escolha dos modelos foi pelo menor valor médio quadrático, pelo critério AIC, pela qualidade da previsão e exigindo que os parâmetros do modelos sejam estatisticamente significativos.

As Tabelas 7, 8 e 9 representam três modelos diferentes ajustados através da metodologia de Box-Jenkins. Ao *Modelo1* é ajustado um processo ARMA com uma componente autorregressiva de ordem 2 aos resíduos do modelo com tendência. Ao *Modelo2* foi ajustado à série do índice ISM um processo autorregressivo e de médias móveis integrado sazonal (SARIMA) com horizonte temporal de 12 meses. Ao *Modelo3* é ajustado à série do índice ISM um processo autorregressivo e médias móveis integrado. Estes modelos definem-se da seguinte forma:

$$\text{Modelo1} : (1 - 1,0705B + 0,1409B^2)Y_t = \varepsilon_t, \quad (4.80)$$

$$\text{Modelo2} : (1 - 0,0539B - 0,1108B^2)(1 + 0,1345B^{12} + 0,1478B^{24})(1 - B)Y_t = \varepsilon_t, \quad (4.81)$$

$$\text{Modelo3} : (1 - 0,0898B - 0,1226B^2)(1 - B)Y_t = \varepsilon_t. \quad (4.82)$$

Tabela 7: Estimativas dos parâmetros e respetivos erros padrão com critérios de avaliação do Modelo ARMA(2,0,0) para a série ISM.

*Modelo1 = ARMA(2,0,0)*

Coeficientes	$\phi_1$	$\phi_2$
Estimativa	1,0705	-0,1409
Erro de padrão	0,0423	0,0427
Variância Residual	4,557	
Log-verosimilhança	-1204,14	
Critério de informação de Akaike (AIC)	2416,28	
REQM	2,1393	
EM	0,046652	

Tabela 8: Estimativas dos parâmetros e respectivos erros padrão com critérios de avaliação do Modelo SARIMA(2,1,0)(2,0,0)<sub>[12]</sub> para a série ISM.

$$\text{Modelo2} = \text{SARIMA}(2, 1, 0)(2, 0, 0)_{[12]}$$

Coeficientes	$\phi_1$	$\phi_2$	$\Phi_1$	$\Phi_2$
Estimativa	0,0539	0,1108	-0,1345	-0,1478
Erro de padrão	0,0437	0,0429	0,0456	0,045
Variância Residual	4,431			
Log-verossimilhança	-1168,54			
Critério de informação de Akaike (AIC)	2347			
REQM	2,1031			
EM	0,0077			

Tabela 9: Estimativas dos parâmetros e respectivos erros padrão com critérios de avaliação do Modelo ARIMA(2,1,0) para a série ISM.

$$\text{Modelo3} = \text{ARIMA}(2, 1, 0)$$

Coeficientes	$\phi_1$	$\phi_2$
Estimativa	0,0898	0,1226
Erro de padrão	0,0424	0,0425
Variância Residual	4,67	
Log-verossimilhança	-1206,43	
Critério de informação de Akaike (AIC)	2418,86	
REQM	2,159	
EM	0,0432	

Os gráficos das FAC e FACP (Figura 9) sugerem que existe sazonalidade. A diferença entre o Modelo1 e *Modelo2* para além da ordem de diferenciação ( $d$ ), difere na componente sazonal. Contudo, pelo erro quadrático médio e variância residual, o *Modelo2* é preferível. Verifica-se ainda que o Critério de Informação de Akaike é maior no modelo mais simples (Modelo1). O Modelo3 é ajustado através de um processo autorregressivo e de médias móveis integrado (ARIMA), com a ausência da componente sazonal e apresenta um AIC maior, um erro quadrático médio maior e um aumento no REQM. Atendendo apenas à variância residual, o *Modelo2* seria o mais apropriado, no entanto cada um destes modelos tem

vantagens e desvantagens dependendo do que é pretendido. Uma medida de comparação necessária são as respetivas FAC e FACP.

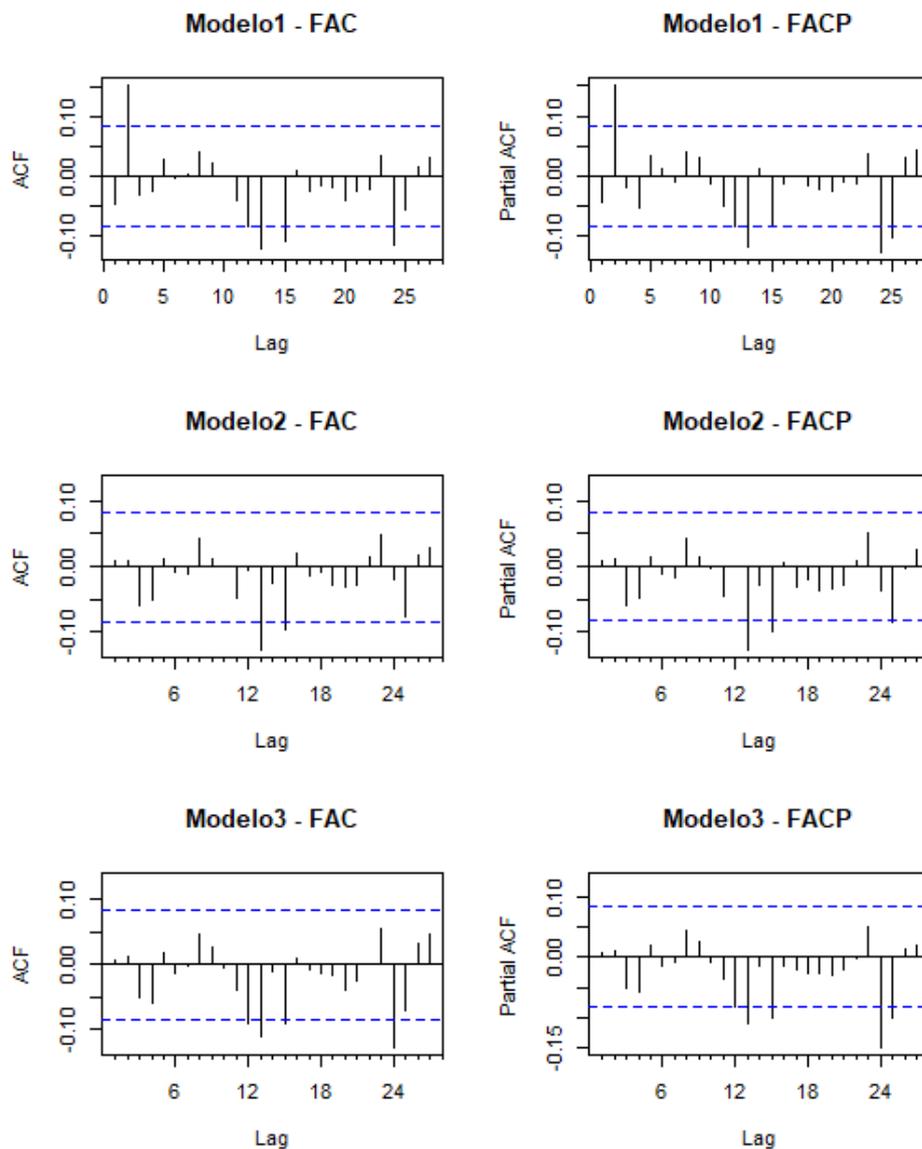


Figura 9: FAC e FACP dos Modelos 1, 2 e 3

A análise da Figura 9 indica a presença de uma componente sazonal entre os *lags* 12 ou 13 e 24, com a exceção do Modelo 2. Falta apenas realizar uma análise aos resíduos de cada um destes modelos.

No âmbito da validação dos modelos procedeu-se à análise dos resíduos para a verificação dos pressupostos, incluindo a análise dos gráficos que permitem tirar conclusões pertinentes sobre a distribuição dos erros, além das características do processo de ruído branco.

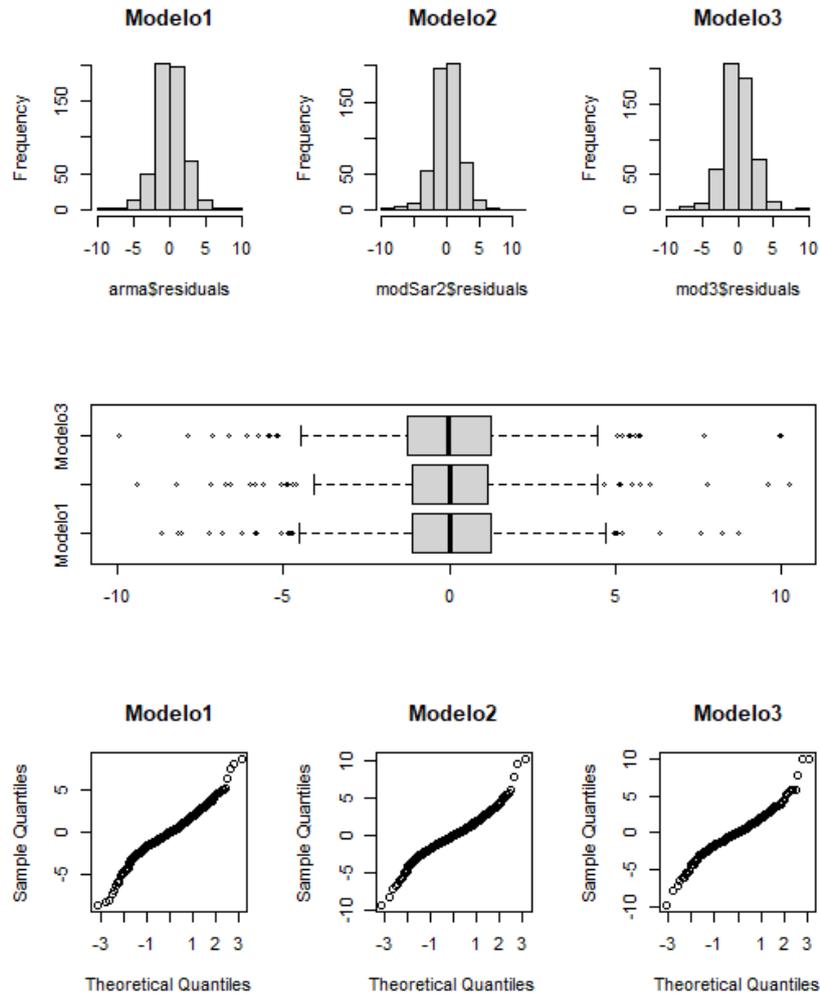


Figura 10: Histogramas, diagrama de caixa de bigodes e *QQplots* dos resíduos

Os resíduos dos três modelos comportam-se de forma semelhante. Com especial atenção aos histogramas, observa-se que estes não apresentam uma distribuição semelhante à Normal. Este facto é analisado através dos testes de hipóteses de Shapiro-Wilk e de Kolmogorov-Smirnov que rejeitam a hipótese nula nos três modelos, ou seja há evidência estatística para rejeitar a hipótese de que os erros tenham uma distribuição Normal.

O teste de correlação Ljung-Box é um teste estatístico que avalia se existe autocorrelação nos resíduos de um modelo de séries temporais. Este verifica se os resíduos exibem algum padrão de dependência temporal. Se os resíduos forem independentes ao longo do tempo (hipótese nula), o teste não encontrará evidências de autocorrelação significativa. No entanto, se a autocorrelação for detetada, isso sugere que o modelo não captura completamente a estrutura temporal dos dados. O Modelo 1 e Modelo 3 rejeitam a hipótese nula deste teste, no entanto o Modelo 2 não rejeita. Foram considerados *lags* de dimensão 26

para o Modelo 1 e Modelo3, e 24 para o Modelo 2.

Numa análise crítica à metodologia Box-Jenkins, concluiu-se que o modelo mais adequado é o *Modelo2*, no entanto não quer dizer que este seja o melhor modelo para descrever e realizar previsões da série em estudo. Embora seja uma abordagem útil para modelar séries temporais, pode não ser a melhor escolha para este caso em específico. Apesar da metodologia de Box-Jenkins ser uma ferramenta poderosa para análise de séries temporais, esta não é uma solução única para todos os casos. É necessário considerar a natureza dos dados e as características específicas da série temporal ao escolher a abordagem mais adequada para ajustar modelos e realizar previsões.

### 4.3.3 Modelos de espaço de estados

Os modelos de espaço de estados são utilizados para representar sistemas dinâmicos e suas relações entre estados não observados ao longo do tempo. Nestes modelos, as variáveis de interesse são os estados não observados, que representam a informação subjacente e não diretamente observável, e as observações, que são medidas disponíveis. Lembra-se o que já foi definido anteriormente na Eq.(3.8) e Eq.(3.9), agora na versão de um modelo univariado tanto relativamente ao estado como em relação à variável observada:

#### **Equação de observação,**

$$Y_t = W_t\beta_t + e_t,$$

#### **Equação de estados,**

$$\beta_t = \phi\beta_{t-1} + \varepsilon_t.$$

Este modelo pode ser interpretado como um modelo linear sem termo independente, no qual a variável independente,  $W_t$ , é multiplicada por um coeficiente estocástico, e não determinístico como na regressão linear, ao qual designamos por *ModeloT*, para  $t = 1, \dots, n$ , em que  $Y_t$  é a variável observada dos valores registados pelo índice do ISM no mês  $t$ ,  $W_t$  corresponde aos valores ajustados da tendência linear do modelo de regressão já ajustado,  $\beta_t$  é o estado que representa um fator de correção/calibração do nível da tendência  $W_t$ , e o erro  $e_t$  é uma variável aleatória que representa o erro aleatório. No modelo que será adotado para esta série em particular vai-se considerar que o processo dos estados  $\beta_t$  segue um processo AR(1) estacionário com média  $\mu$ ,

$$\beta_t = \mu + \phi(\beta_{t-1} - \mu) + \varepsilon_t. \quad (4.83)$$

Assume-se que,

- O processo  $\{\beta_t\}_{t=1,\dots,n}$  é um processo autorregressivo estacionário de ordem 1, logo,  $\{\beta_t\}_{t=1,\dots,n} \sim AR(1)$  e  $|\phi| < 1$ ;
- O processo  $\{\beta_t\}_{t=1,\dots,n}$  tem média igual a  $\mu$ ,  $E(\beta_t) = \mu$ , e variância  $var(\beta_t) = \frac{\sigma_\varepsilon^2}{1-\phi^2}$ ;
- $e_t$  corresponde ao erro aleatório de um processo de ruído branco, logo,  $E(e_t) = 0$  e  $E(e_t e_s) = 0$ ,  $\forall t, s \in \{1, \dots, n\}, t \neq s$ ;
- $\varepsilon_t$  corresponde ao erro aleatório de um processo de ruído branco, logo  $E(\varepsilon_t) = 0$  e  $E(\varepsilon_t \varepsilon_s) = 0$ ,  $\forall t, s \in \{1, \dots, n\}, t \neq s$ ;
- Os erros  $e_t$  e  $\varepsilon_t$  não são correlacionados, logo  $E(e_t \varepsilon_t) = 0, \forall t, s \in \{1, \dots, n\}, t \neq s$ .

Considera-se  $\Theta = \{\phi, \mu, \sigma_e^2, \sigma_\varepsilon^2\}$  o conjunto dos parâmetros desconhecidos do modelo que correspondem respetivamente ao coeficiente autorregressivo, ao valor médio dos estados e às variâncias dos erros das equações de observação e de estado, respetivamente. Adicionalmente, poder-se-á assumir que ambos os erros seguem uma distribuição Normal. Com o pressuposto da normalidade pode recorrer-se ao método da máxima verosimilhança gaussiana da Secção 3.4.1 para estimar os parâmetros desconhecidos. Para implementar o método numérico necessário à obtenção das estimativas de máxima verosimilhança são necessários valores iniciais para os parâmetros para inicializar o processo iterativo. Este conjunto de valores iniciais é crucial, uma vez que pode condicionar a convergência do processo numérico.

O procedimento de estimação pela máxima verosimilhança visa encontrar os valores dos parâmetros que maximizam a probabilidade de observar os dados. Isso é alcançado através da otimização iterativa, ajustando os parâmetros para melhor se adequarem aos dados observados. Os valores iniciais dos parâmetros desempenham um papel fundamental nesse processo, uma vez que determinam o ponto de partida da otimização. Se os valores iniciais dos parâmetros forem bem escolhidos e estiverem próximos dos valores reais, o processo de otimização tende a convergir para as estimativas de máxima verosimilhança dos parâmetros. No entanto, se os valores iniciais estiverem longe dos valores reais, o processo de otimização pode não convergir ou produzir estimativas às quais não correspondem o máximo global da função de log-verosimilhança.

Tabela 10: Estimativas pela máxima verosimilhança

	$\hat{\phi}$	$\hat{\mu}$	$\hat{\sigma}_\varepsilon^2$	$\hat{\sigma}_e^2$	Log-Verosimilhança
Estimativa	0,9224	1,0118	0,0016	0	-700,5805
Erro de Padrão	0,0154	0,0228	0,0001	0,1890	

Após o ajustamento do modelo com as estimativas de máxima verosimilhança obtidas (Tabela 10), foram calculadas as inovações padronizadas, Eq.(3.66), para a necessária análise da verificação dos pressupostos. A Figura 11 representa o histograma e o QQ-plot das inovações padronizadas obtidas no ajustamento do modelo à série de treino.

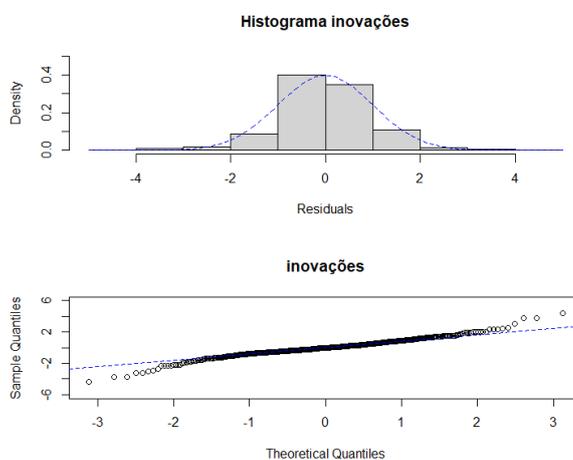


Figura 11: Histograma e *QQ-plot* das inovações padronizadas

A partir destes gráficos podemos concluir que as inovações padronizadas não apresentam um comportamento compatível com uma distribuição Normal, o que, neste caso, significa a não verificação do pressuposto da normalidade assumida na estimação pela máxima verosimilhança. As inovações padronizadas têm uma média de  $-0,0019$  sendo que o seu primeiro quartil é  $Q_1 = -0,5327$  e terceiro  $Q_3 = 0,5713$ . Uma análise ao diagrama de caixa de bigodes e ao gráfico da série das inovações padronizadas, Figura 12, verifica-se que existem *outliers*, com valores extremos de  $-4,3207$  e  $4,3558$ . A Tabela 11 representa os resultados obtidos dos testes da normalidade de Shapiro-Wilk e Kolmogorov-Smirnov, bem como os testes à correlação Ljung-Box e Box-Pierce.

Tabela 11: valores de prova dos testes estatísticos à normalidade e correlação à série das inovações padronizadas

Teste	valores de prova	Hipótese Nula
Shapiro-Wilk	$3,052 \times 10^{-09}$	Rejeitada
Kolmogorov-Smirnov	0,02283	Rejeitada
Ljung-Box	$5,303 \times 10^{-06}$	Rejeitada
Box-Pierce	$1,585 \times 10^{-06}$	Rejeitada

**Boxplot Inovações**

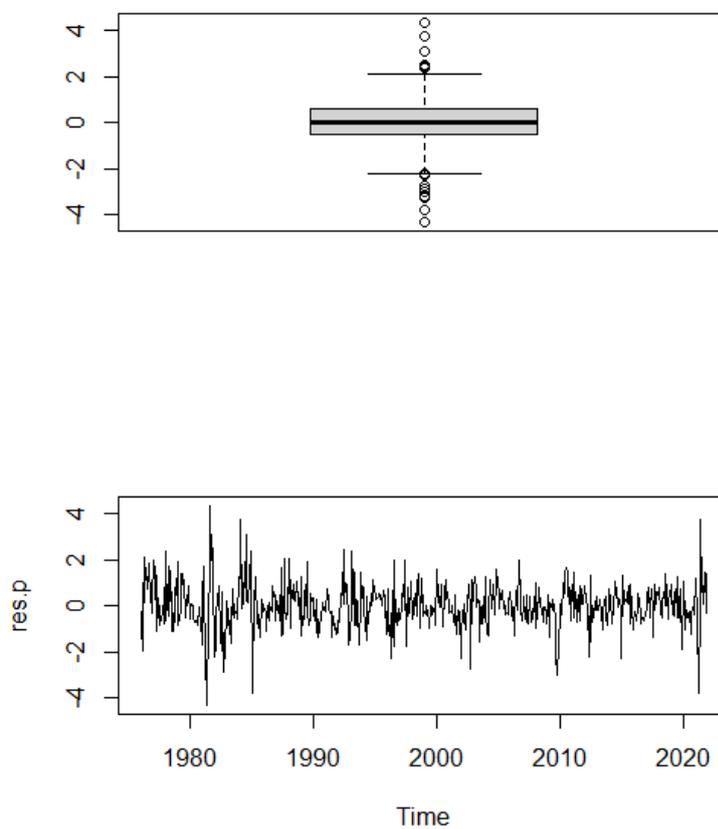


Figura 12: Diagrama de caixa de bigódes e gráfico das inovações padronizadas

Na Tabela 12 estão representadas as medidas de avaliação relativas às previsões a 1 passo,  $Y_{t|t-1}$ , do índice do ISM no período de teste.

Tabela 12: Medidas de Avaliação para  $Y_{t|t-1}$

Medidas de Avaliação	Valores
EM	0,0358
EQM	2,3042
EAM	1,6016
EPM	0,0668
EPAM	3,1689

Compara-se  $Y_{t|t-1}$  com a série de treino através de um gráfico e uma análise às estimativas.

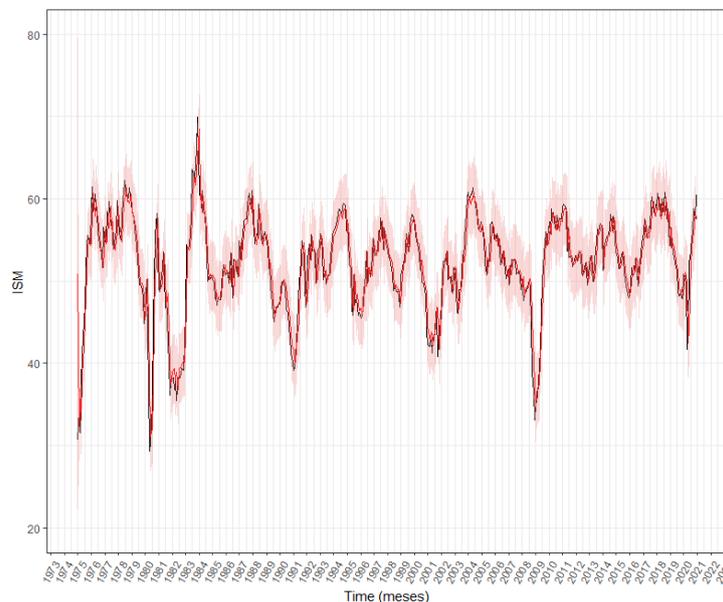


Figura 13: Previsões a 1 passo do índice do ISM no período de treino e os respectivos intervalos de confiança a 95%,  $Y_{t|t-1}$  – vermelho;  $Y_t$  – preto.

Conforme a Tabela 12 e o gráfico da Figura 13, o modelo apresenta uma boa capacidade de descrever o comportamento dos dados, observa-se ainda que os intervalos de confiança captam corretamente toda a evolução da série, uma vez que a taxa de cobertura destes intervalos é de 95%.

Contudo, apesar do bom ajustamento do modelo de espaço de estados com as estimativas de máxima verossimilhança, o pressuposto da normalidade das inovações foi rejeitado. Por isso, consideram-se os estimadores independentes da distribuição de base para obterem-se estimativas sem esse pressuposto. No entanto, não são conhecidas as distribuições desses estimadores pelo que os respectivos intervalos de confiança serão obtidos através da metodologia *Bootstrap*. A justificação para esta escolha reside no facto

de que a estimação pela máxima verosimilhança encontrou desafios relacionados com os pressupostos subjacentes ao modelo.

Consequentemente, e em face destes desafios, optou-se por recorrer à estimação não-paramétrica (Costa, 2006) apoiada pelo método de *Bootstrap*. Esta abordagem permite contornar as limitações associadas à estimação paramétrica pela máxima verosimilhança, uma vez que não se baseia na imposição de pressupostos específicos sobre a distribuição dos erros (Costa, 2006). Em vez disso, a estimação não paramétrica complementada pelo método de *Bootstrap* é mais flexível e capaz de fornecer estimativas robustas, mesmo quando os pressupostos paramétricos não são atendidos.

Tabela 13: Estimativas pela estimação não-paramétrica para os parâmetros do modelo de espaço de estados

	$\hat{\phi}$	$\hat{\mu}$	$\hat{\sigma}_\varepsilon^2$	$\hat{\sigma}_e^2$
Estimativa	0,8998	1,0013	0,0019	0

As estimativas não paramétricas são semelhantes às estimadas pela máxima verosimilhança (Tabela 13). A estimativa  $\hat{\sigma}_e^2$  obtida não pertence ao espaço do parâmetro o que é concordante com a estimativa de máxima verosimilhança ser zero para esta variância. Neste sentido, será considerado para o resto da abordagem que a estimativa não-paramétrica é igual a zero.

No âmbito da investigação, optou-se por utilizar a estimação não paramétrica, seguida da aplicação do filtro de Kalman. Após a aplicação do filtro de Kalman, foram obtidas as estimativas dos estados no tempo atual, representadas por  $\hat{\beta}_{t|t-1}$ , as previsões dos valores observados no tempo atual, representadas por  $\hat{Y}_{t|t-1}$ , e as inovações do modelo, que foram padronizadas. Com base nas estimativas dos  $\hat{\beta}_{t|t-1}$ ,  $\hat{Y}_{t|t-1}$  e inovações, procedeu-se à implementação de uma estratégia de *Bootstrap* com reamostragem.

A aplicação do *Bootstrap* envolveu a geração de 1000 amostras aleatórias simples com reposição das inovações padronizadas, a partir das quais obtiveram-se 1000 séries temporais  $Y_t^*$ . A cada nova série *Bootstrap* são estimados os parâmetros através dos estimadores não paramétricos. Após a execução do *Bootstrap*, foram obtidas 1000 estimativas para cada um dos parâmetros  $\hat{\phi}$ ,  $\hat{\mu}$ ,  $\hat{\sigma}_\varepsilon^2$ . Com base nesse conjunto de estimativas, é possível calcular e construir intervalos de confiança de confiança *Bootstrap* para o nível de confiança pretendido. Nas Figuras 14, 15 e 16 encontram-se representadas as distribuições das estimativas *Bootstrap* destes parâmetros. Na Tabela 14 apresentam-se as medidas descritivas para as estimativas *Bootstrap*.

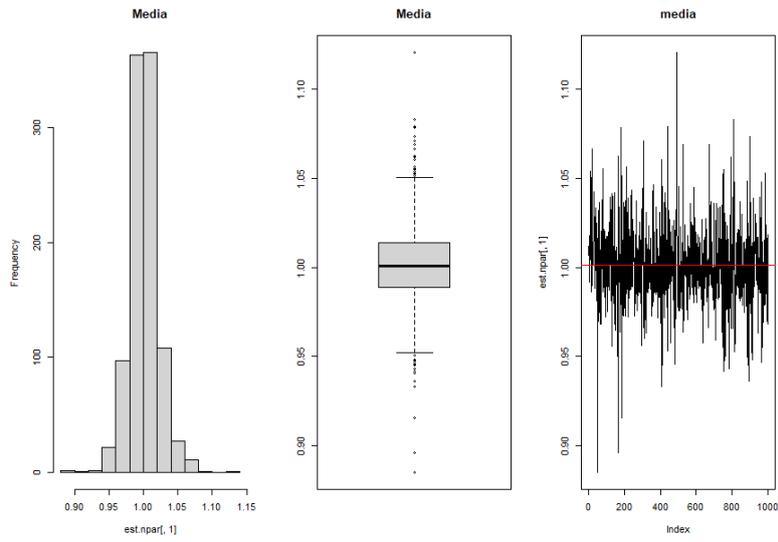


Figura 14: Histograma, diagrama de caixa de bigodes e gráfico das estimativas *Bootstrap* para  $\hat{\mu}$

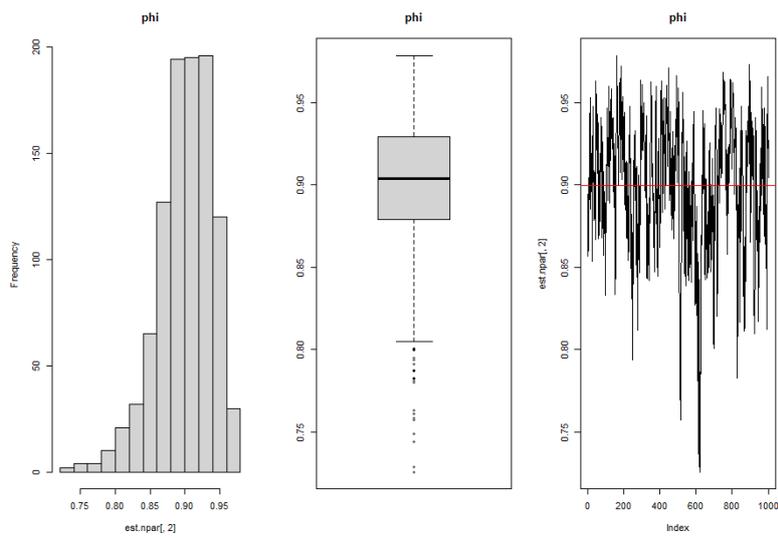


Figura 15: Histograma, diagrama de caixa de bigodes e gráfico das estimativas *Bootstrap* para  $\hat{\phi}$

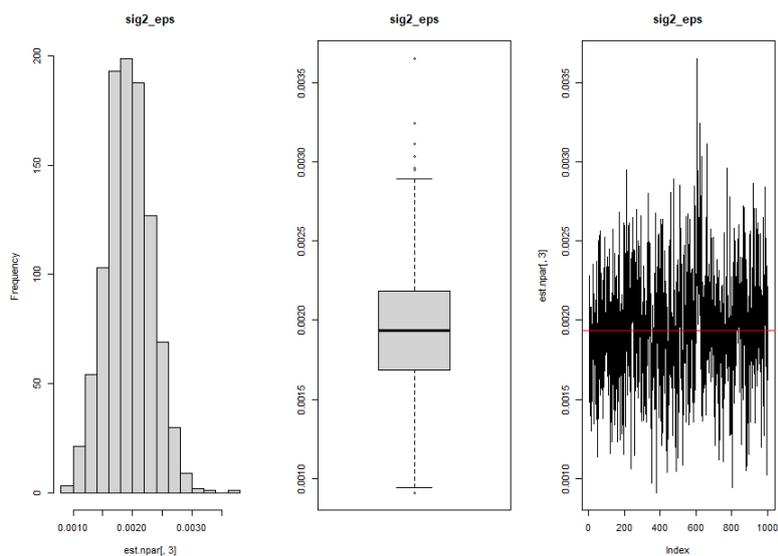


Figura 16: Histograma, diagrama de caixa de bigodes e gráfico das estimativas *Bootstrap* para  $\hat{\sigma}_\varepsilon^2$

Tabela 14: Estatísticas descritivas das estimativas *Bootstrap*

	Min.	1° Q.	Mediana	Média	3° Q.	Max.	Variância
$\hat{\mu}$	0,8850	0,9887	1,0008	1,0013	1,0137	1,1204	0,0004
$\hat{\phi}$	0,7255	0,8790	0,9038	0,8998	0,9293	0,9789	0,0016
$\hat{\sigma}_\varepsilon^2$	0,0009	0,0016	0,0019	0,0019	0,0021	0,0036	$1,414 \times 10^{-07}$

Tabela 15: Teste da Normalidade das estimativas *Bootstrap*

Teste Shapiro-Wilk	$\hat{\mu}$	$\hat{\phi}$	$\hat{\sigma}_\varepsilon^2$
valores de prova	$1,262 \times 10^{-13}$	$2,2 \times 10^{-16}$	0,04602
Hipótese Nula	Rejeita	Rejeita	Rejeita
Teste Kolmogorov-Smirnov			
valores de prova	0,0046	0,0049	0,9315
Hipótese Nula	Rejeita	Rejeita	Não Rejeita

Para além do teste de Kolmogorov-Smirnov não rejeitar a hipótese nula da normalidade para a distribuição das estimativas obtidas a partir da abordagem *Bootstrap* para  $\hat{\sigma}_\varepsilon^2$ , o valor de prova do teste de Shapiro-Wilk está próximo dos 5% (Tabela 15). Os restantes testes à normalidade para os outros parâmetros rejeitam a normalidade a um nível de significância de 5%. Neste caso específico, a hipótese da normalidade da distribuição do estimador da variância do erro na equação dos estados não foi rejeitada. Estes resultados sugerem que os estimadores não-paramétricos, mesmo em séries temporais de dimensão considerável, não têm distribuições Normais por amostragem, ou que, a convergência assintótica para a Normal, caso exista, é muito lenta.

Após a execução da abordagem *Bootstrap*, foram obtidas 1000 estimativas para cada parâmetro, como já foi referido. Com base nesse conjunto de estimativas, calculou-se os quantis *Bootstrap* de 2,5% e 97,5% para que se construam os respetivos intervalos de confiança *Bootstrap* a 95% (Tabela 16) e considera-se a média dos valores dos parâmetros nas 1000 réplicas como os valores estimados. Esse procedimento permitiu avaliar a incerteza das estimativas não-paramétricas e fazerem-se comparações com os respetivos valores de estimação de máxima verosimilhança.

Tabela 16: Intervalos de confiança e erros-padrão para a estimação paramétrica e estimação não paramétrica

	Máxima verosimilhança				Não Paramétrica			
	Estimativas	erros-padrão	Lim. Inf.	Lim. Sup.	Estimativas	Quantil 2,5%	Quantil 97,5%	Erros-padrão
$\hat{\mu}$	1,0017	0,0251	0,9525	1,0509	1,0013	0,9505	1,0572	0,0278
$\hat{\phi}$	0,93003	0,0159	0,8989	0,9612	0,8998	0,8105	0,9698	0,0418
$\hat{\sigma}_\varepsilon^2$	0,0016	0,0001	0,0383	0,0431	0,0019	0,0012	0,0027	0,0004
$\hat{\sigma}_\varepsilon^2$	0							

Tabela 17: Semi-amplitudes dos intervalos de confiança para os parâmetros obtidos pela máxima verosimilhança (MV) e pelos estimadores não-paramétricos com *Bootstrap*

Parâmetros	Limite-inferior	Limite-superior
$\hat{\mu}$	0,0492	0,0533
$\hat{\phi}$	0,0312	0,0796
$\hat{\sigma}_\varepsilon^2$	0,0024	0,0008

Da análise dos resultados podemos concluir que ambos os métodos de estimação apresentam estimativas pontuais dos parâmetros muito semelhantes. No entanto, a diferença mais significativa é relativamente a  $\hat{\sigma}_\varepsilon^2$ , com estimação pela máxima verosimilhança a estimar esta variância em cerca de

15 vezes o valor da estimativa não-paramétrica. Essa diferença, embora clara, não afeta significativamente as conclusões gerais da análise. A estimação pela máxima verosimilhança, embora dependente de pressupostos sobre a distribuição dos erros, ainda ofereceu resultados interessantes do ponto de vista da modelação, e as diferenças em relação à estimação não-paramétrica não foram significativas. Isso sugere que, apesar dos desafios observados nos pressupostos da estimação pela máxima verosimilhança, este método continua sendo uma ferramenta robusta para a análise de modelos de espaço de estados, pelo menos na aplicação específica deste estudo.

A abordagem não paramétrica, por sua vez, permitiu contornar as limitações associadas ao pressuposto da normalidade. No entanto, a discrepância na estimativa de  $\hat{\sigma}_\varepsilon^2$  em relação à abordagem pela máxima verosimilhança destaca a necessidade de se considerar cuidadosamente a escolha da abordagem de estimação em função dos objetivos da análise e das características dos dados.

Da Tabela 17 pode-se concluir que o método da máxima verosimilhança produziu intervalos de confiança com amplitudes inferiores aos estimadores não-paramétricos, com a exceção da estimação da variância  $\sigma_\varepsilon^2$ . A principal consequência da diferença das estimativas da variância de estado pela máxima verosimilhança e pelo estimador não-paramétrico será observada através da amplitude dos intervalos de confiança para as previsões a 1 passo para o índice do ISM, sem prejuízo na taxa de cobertura.

Em última análise, destaca-se a importância de avaliar a adequação das abordagens de estimação em relação aos pressupostos e objetivos específicos do estudo. A escolha entre a estimação pela máxima verosimilhança e a estimação não paramétrica com *Bootstrap* deve ser feita com base numa compreensão aprofundada das limitações e vantagens de cada método e nas *nuances* do problema em análise.

#### 4.3.4 Previsões

Neste secção apresenta-se a obtenção de previsões num contexto de modelação estatística, centrando-se nos três modelos ajustados. O período para qual se pretende obter as previsões a 1 passo corresponde à série de teste, isto é, estende-se desde janeiro de 2021 até maio de 2022, compreendendo um total de 17 observações.

As previsões serão obtidas por meio dos três modelos diferentes: um modelo *SARIMA* (*Modelo2* da Tabela 8), um modelo de espaço de estados cujos parâmetros foram estimados utilizando a máxima verosimilhança (Tabela 10) e um segundo modelo de espaço de estados, no qual os parâmetros foram estimados através de um método não paramétrico complementado pelo uso de *Bootstrap* para estimar os erros padrão e os intervalos de confiança (Tabela 13). Uma característica relevante deste estudo é a consideração de intervalos de confiança a 95% para as previsões a 1 passo obtidas a partir de cada um

dos modelos. Esses intervalos de confiança fornecem uma medida da incerteza associada às previsões e são fundamentais para avaliar os resultados.

As previsões serão avaliadas com base em medidas de avaliação amplamente conhecidas e já referidas, em particular a Raiz do Erro Quadrático Médio (REQM). Essas medidas de avaliação são essenciais para a comparação do desempenho dos modelos, permitindo a identificação daquele que melhor se ajusta aos dados e proporciona previsões mais precisas.

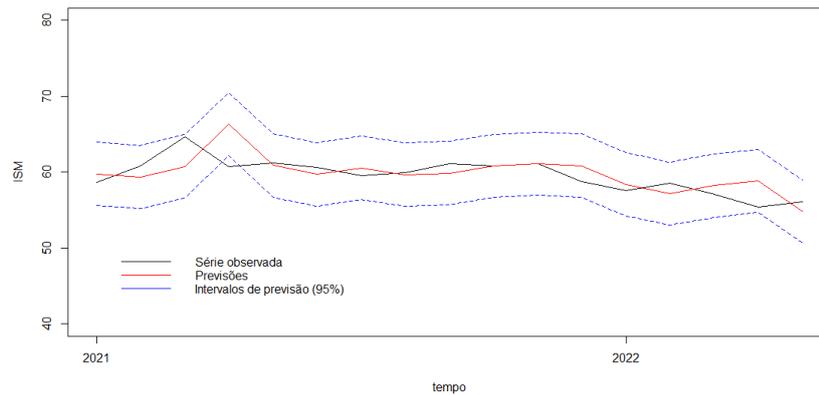


Figura 17: Gráfico com as previsões para 17 meses do *Modelo2* e os respectivos intervalos de confiança

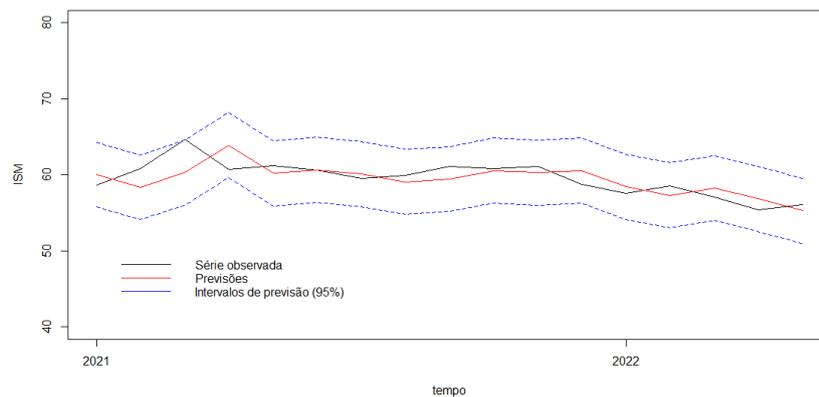


Figura 18: Gráfico com as previsões para 17 meses do Modelo de espaço de estados (*ModeloMV*) com estimadores paramétricos e os respectivos intervalos de confiança

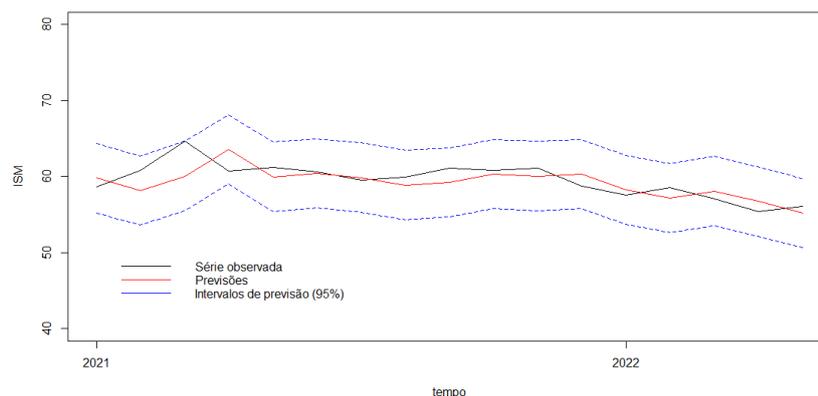


Figura 19: Gráfico com as previsões para 17 meses do Modelo de espaço de estados (*ModeloNP*) com estimadores não paramétricos e os respectivos intervalos de confiança

Tabela 18: Erro quadrático médio e raiz do erro quadrático médio das previsões dos 3 modelos e do modelo de RLS.

Modelos	EQM	REQM
<i>ModeloMV</i>	3,0957	1,7594
<i>ModeloNP</i>	3,1495	1,7747
<i>Modelo2</i>	4,4802	2,1167
<i>RLS</i>	42,9779	6,5557

Nas Tabelas 18 e 19 complementadas pela análise dos gráficos nas Figuras 17, 18 e 19, encontra-se a informação para comparar a qualidade de previsão dos três modelos. A Tabela 5 apresenta o RLS, uma vez que ele representa a tendência ajustada aos dados por meio de um modelo de regressão linear. Estes valores foram utilizados para ajustar os dois modelos de espaço de estados. Presta-se atenção aos valores do EQM e REQM, sendo que o modelo que prevê com maior precisão é o modelo de espaço de estados cujos parâmetros foram estimados pela máxima verosimilhança. No entanto, também se pode considerar o *ModeloNP*, uma vez que a diferença entre as medidas não é significativa. Já as previsões do *Modelo2* apresentam um erro quadrático médio e uma raiz do erro quadrático médio maiores.

Para além dos erros do *ModeloMV* serem menores, ao analisar a amplitude dos intervalos de confiança na Tabela 19, verifica-se que a um nível de confiança de 95%, existe uma pequena diferença entre os limites (gráfico da Figura 18), com uma média de 4,2592. Este valor reforça a qualidade das previsões a um passo deste modelo. Por outro lado, no *ModeloNP*, a amplitude é semelhante (gráfico

da Figura 19), com uma média da amplitude dos intervalos de confiança de 4,5616, o que é aceitável. A taxa de cobertura dos intervalos de confiança para os 3 modelos é de 94,1177%, o que quer dizer que apenas um valor previstos a 1 passo não é observado dentro dos limites dos intervalos de confiança.

Na análise do *Modelo2*, a amplitude dos intervalos comporta-se de forma semelhante à dos outros modelos, no entanto é o modelo com a menor diferença entre os limites, com uma média de 4,1742 (gráfico da Figura 17). Para os meses entre março, abril e maio de 2021, há uma sobrestimação dos valores de previsão comparando com os outros dois modelos e a série original, mas nos meses seguintes, a qualidade das previsões são aceitáveis, com uma ligeira sobrestimação para os meses entre março e maio de 2022 (entre as observações 15 e 17, gráfico da Figura 20). Com base nesta análise, conclui-se que o modelo que produz previsões de maior qualidade é o *ModeloMV*. É igualmente importante analisar os erros de  $W_t$ , uma vez que é através deles que os modelos de espaço de estados são ajustados. Com um valor considerado elevado de 42,9779, verifica-se que a tendência não segue corretamente os valores dos dados reais. Isso prejudica a qualidade das previsões dos MEE, uma vez que o modelo de regressão linear não é suficiente para captar o comportamento dos dados. Em resumo, o *Modelo2* proporcionou a estimação intervalar das previsões com menores amplitudes, mas o *ModeloMV* e *ModeloNP* proporcionaram as previsões com maior precisão.

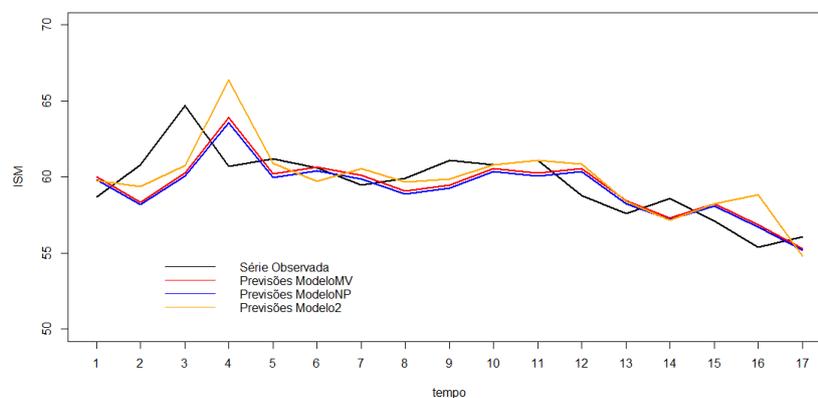


Figura 20: Gráfico com as previsões para 17 meses dos 3 Modelos e série observada

Tabela 19: Amplitude dos intervalos de confiança das previsões para os 3 modelos

Meses previstos	<i>ModeloMV</i>	<i>ModeloNP</i>	<i>Modelo2</i>
janeiro-2021	4,2564	4,5585	4,1676
fevereiro-2021	4,2567	4,5589	4,1648
março-2021	4,2571	4,5593	4,1626
abril-2021	4,2575	4,5597	4,1718
maio-2021	4,2578	4,56	4,1940
junho-2021	4,2582	4,5604	4,1903
julho-2021	4,2585	4,5608	4,1872
agosto-2021	4,2589	4,5612	4,1841
setembro-2021	4,2592	4,5616	4,1807
outubro-2021	4,2596	4,562	4,1782
novembro-2021	4,26	4,5623	4,1744
dezembro-2021	4,2603	4,5627	4,1707
janeiro-2022	4,2607	4,5631	4,1705
fevereiro-2022	4,2610	4,5635	4,1673
março-2022	4,2614	4,5639	4,1653
abril-2022	4,2618	4,5643	4,1627
maio-2022	4,2621	4,5647	4,1686

## 5 Conclusão

Esta dissertação teve como principal objetivo contribuir para a análise e previsão de séries temporais, adotando uma abordagem baseada em modelos de espaço de estados, com foco na base de dados do índice ISM. Uma das etapas cruciais na construção de modelos de séries temporais envolve a estimação dos parâmetros destes, uma vez que essa estimação é fundamental para a obtenção de previsões precisas. Esta etapa é crucial para obter-se previsões precisas sejam estas pontuais ou intervalares.

Neste contexto, consideraram-se abordagens diversas para a estimação dos parâmetros de modelos de espaço de estados. Inicialmente, recorreu-se ao método da máxima verosimilhança, pressupondo que tanto os erros de observação quanto os de estados seguem uma distribuição Normal. Contudo, os erros não seguem uma distribuição Normal, optou-se então por alternativas não paramétricas, como os estimadores do tipo *distribution-free* como em Costa (2006). Nesse âmbito, no trabalho obtiveram-se estimativas robustas, incluindo os erros padrão e intervalos de confiança, aplicando a técnica de *Bootstrap* (Stoffer & Wall, 1991) aos estimadores não paramétricos, conforme proposto por Costa (2006).

A base de dados estudada, *Manufacturing PMI* (índice do ISM), tem um caráter económico e avalia o desempenho de várias indústrias nos Estados Unidos da América, em intervalos mensais. Essa base de dados desempenha um papel crucial para economistas e analistas de mercados financeiros, uma vez que fornece indicadores de grande relevância, sendo um dos primeiros indicadores disponibilizados no início de cada mês, Bognanni & Young (2018). Para obter-se as estimativas dos estimadores propostos por Costa (2006), no âmbito desta dissertação desenvolveu-se uma função específica em linguagem R, possibilitando a obtenção das estimativas de parâmetros de modelos de espaço de estados, tanto no contexto paramétrico pelo método da máxima verosimilhança, assumindo normalidade dos erros, quanto no contexto dos estimadores *distribution-free*. Essas estimativas, sejam elas pontuais ou intervalares, foram subsequentemente aplicadas a séries temporais reais, representadas pelos dados económicos do ISM.

É importante mencionar que a construção dos modelos de espaço de estados foi considerado como covariável as estimativas obtidas pelo modelo de regressão linear que representa a tendência subjacente dos dados. Dessa forma, os modelos de espaço de estados consideraram como  $W_t$  o nível da tendência linear, descrita na Eq.(4.83). O propósito foi comparar as previsões a 1 passo estimadas por três modelos diferentes: um modelo *SARIMA*, um modelo de espaço de estados cujos parâmetros foram estimados via máxima verosimilhança, ou seja, uma abordagem paramétrica, e um terceiro modelo de espaço de estados no qual os parâmetros foram estimados por meio de um método estimação *distribution-free*, sendo os erros padrão e os intervalos de confiança dessas estimativas obtidos por meio da técnica

de *Bootstrap*. A avaliação das previsões foi realizada considerando medidas de avaliação propostas no Subcapítulo 3.6.1.

A análise comparativa entre estas abordagens e a avaliação das previsões estimadas proporcionam uma visão abrangente sobre a eficácia e robustez dos modelos. Com base na análise a todos os modelos considerados na Secção 4.3.4, o modelo de espaço de estados, no qual os parâmetros foram estimados através de um método não paramétrico (proposto por Costa (2006)) complementado pelo uso de *Bootstrap* para estimar os erros padrão e os intervalos de confiança, oferece um equilíbrio entre a precisão e os intervalos de confiança comparando com os outros modelos. De todas as abordagens e modelos esta foi a mais adequada para o índice do ISM.

## 5.1 Trabalho futuro

À luz das complexidades enfrentadas ao longo desta dissertação e das questões que surgiram no decurso da investigação, identificaram-se várias oportunidades para investigações futuras, propondo com maior prioridade as seguintes:

- Desenvolver um estudo de simulações mais abrangente para examinar o desempenho dos estimadores de máxima verosimilhança e não paramétricos, com foco particular na precisão por meio da análise dos erros-padrão. No caso dos estimadores não paramétricos, a avaliação deve incluir a utilização de métodos *Bootstrap* para estimar os erros padrão e a taxa de cobertura dos intervalos de confiança (Lima J.F. (submetido a 13 de novembro de 2023));
- Considerar o desenvolvimento de um *package* para o software estatístico R com o propósito de simplificar o ajustamento de modelos de espaço de estados. Esse *package* facilitaria a obtenção de estimativas de máxima verosimilhança e não paramétricas, juntamente com a avaliação dos erros-padrão e a construção de intervalos de confiança, tornando mais acessíveis futuros trabalhos nessa área;
- Relativamente à modelação do índice do ISM, a consideração de modelos alternativos é uma possibilidade. Por exemplo, podem ser explorados modelos GARCH (Francq & Zakoian, 2019), que são apropriados para lidar com a volatilidade, ou modelos de memória longa. Os modelos GARCH são usados para capturar a variação condicional na volatilidade de séries temporais, sendo úteis na análise de fenómenos financeiros e económicos com variações irregulares de volatilidade ao longo do tempo.

## Bibliografia

- Beran, R. J. (1992). Introduction to efron (1979) bootstrap methods: Another look at the jackknife. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in statistics: Methodology and distribution* (pp. 565–568). New York, NY: Springer New York.
- Bognanni, M., & Young, T. (2018). An assessment of the ism manufacturing price index for inflation forecasting. *Economic Commentary*(2018-05).
- Cordeiro, C. (2003). *Modelos de previsão em séries temporais. aplicação da metodologia bootstrap*. (Doctoral dissertation). doi: 10.13140/RG.2.1.1290.5848
- Costa, M. (2006). *Estimação dos parâmetros de modelos em espaço de estados – uma aplicação à calibração do radar meteorológico* (Doctoral dissertation). doi: 10.13140/RG.2.1.3191.1840
- Costa, M., & Alpuim, T. (2010). Parameter estimation of state space models for univariate observations. *Journal of Statistical Planning and Inference*, 140(7), 1889-1902.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1), 1–22.
- Françq, C., & Zakoian, J.-M. (2019). *Garch models: structure, statistical inference and financial applications*. John Wiley & Sons.
- Goodfriend, M., & King, R. G. (2005). The incredible volcker disinflation. *Journal of Monetary Economics*, 52(5), 981–1015.
- Grewal, M., & Andrews, A. (2010). How good is your gyro [ask the experts]. *IEEE Control Systems Magazine*, 30(1), 12–86.
- Hassani, H., & Yeganegi, M. R. (2020). Selecting optimal lag order in ljung–box test. *Physica A: Statistical Mechanics and its Applications*, 541, 123700.
- Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1), 388–427.
- Hutchinson, C. E. (1984). The kalman filter applied to aerospace and electronic systems. *IEEE transactions on aerospace and electronic systems*(4), 500–504.

- Hyndman, R. J., Athanasopoulos, G., et al. (2014). Optimally reconciling forecasts in a hierarchy. *Foresight: The International Journal of Applied Forecasting*(35), 42–48.
- Jones, R. H. (1980). Maximum likelihood fitting of arma models to time series with missing observations. *Technometrics*, 22(3), 389–395.
- Kalman, R. E., et al. (1960). A new approach to linear filtering and prediction problems [j]. *Journal of basic Engineering*, 82(1), 35–45.
- Lima J.F., G. A. C., Pereira F.C. (submetido a 13 de novembro de 2023). Bootstrapping state-space models: distribution-free estimation in view of prediction and forecasting. *Forecasting, Manuscript ID: forecasting-2743723*.
- Murteira, B. (1993). Estatística descritiva-análise exploratória de dados. *McGraw-Hill*.
- Riani, M., Atkinson, A. C., & Corbellini, A. (2023). Automatic robust box–cox and extended yeo–johnson transformations in regression. *Statistical Methods & Applications*, 32(1), 75–102.
- Sachs, J. D., & Warner, A. M. (1997). Fundamental sources of long-run growth. *The American economic review*, 87(2), 184–188.
- Shumway, R. H., & Stoffer, D. S. (1991). Dynamic linear models with switching. *Journal of the American Statistical Association*, 86(415), 763–769.
- Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3). Springer.
- Stoffer, D. S., & Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the kalman filter. *Journal of the american statistical association*, 86(416), 1024–1033.