

Relatório final

“Modelos de Regressão com Vistas à Análise de Dados
Econométricos”

Orientando: Thiago Moraes Rizzieri

Orientador: Prof. Dr. José Silvio Govone

03 de Maio de 2021 até 29 de Julho de 2022

Thiago Moraes Rizzieri

Sumário

Introdução	4
1 A natureza dos dados econométricos	5
1.1 Metodologia econométrica para teorias econômicas	5
1.1.1 Exposição da teoria ou hipótese econômica	6
1.1.2 Especificação do modelo matemático da teoria;	6
1.1.3 Especificação do modelo estatístico ou econométrico;	6
1.1.4 Obtenção dos dados (organizado, arrumado e limpo);	7
1.1.5 Estimação dos parâmetros do modelo econométrico;	7
1.1.6 Teste de hipóteses;	7
1.1.7 Projeção ou Previsão;	7
1.1.8 Uso do modelo para fins de controle ou de política.	7
1.2 Metodologia para a resolução de um problema	8
1.3 A natureza da análise de regressão	8
1.3.1 Modelagem	8
1.3.2 Correlação e Causação	9
1.3.3 Séries Temporais	9
1.3.4 Regressão Amostral	10
2 Preliminares sobre estatística	11
2.1 Estatística Descritiva	11
2.2 Probabilidade	11
2.3 Inferência estatística	11
3 Análise de regressão clássica	12
3.1 Análise de regressão linear clássica	12
3.1.1 Análise de regressão linear clássica de uma única variável independente . .	13
3.1.2 Propriedades estatísticas dos estimadores de mínimos quadrados	16
3.1.3 Análise de regressão linear clássica múltipla	17
3.2 Análise de regressão não linear clássica	18
4 Econometria de séries temporais	19
4.1 Séries Temporais	19
4.2 Processos Estocásticos	20
4.3 Previsão de séries temporais	21
4.3.1 Mas o que podemos prever?	21

SUMÁRIO

4.3.2	Previsão, meta e planejamento	22
4.3.3	Previsão pontual e previsão intervalar	22
4.4	Autocorrelação	22
4.5	Ruído branco	23
4.6	Decomposição de séries temporais	24
4.6.1	Séries ajustadas sazonalmente	25
4.6.2	Suavização de Média Móvel	25
4.6.3	Decomposição clássica	26
4.7	Intervalo de previsão	26
4.8	Avaliando a previsão pontual	26
4.8.1	Dados para treinamento e dados para testagem	27
4.8.2	Métodos que dependem da escala	27
4.8.3	Métodos para erros percentuais	27
4.8.4	Métodos para erros escalados	28
4.9	Avaliando a previsão intervalar	28
4.9.1	Pontuação dos quantis (ou percentis)	28
4.9.2	Pontuação de Winkler	29
4.9.3	CRPS	29
4.9.4	Skill Score	30
4.10	Validação cruzada para previsões	30
5	Previsões de julgamento	32
5.1	Limitações	32
5.2	Princípios-chave	32
5.3	Método Delphi	33
5.3.1	Especialistas e anonimato	33
5.3.2	Definir as tarefas/problemas	34
5.3.3	<i>Feedback</i>	34
5.3.4	Iteração	34
5.3.5	Previsões finais	34
5.3.6	Limitações e variações do método Delphi	34
6	Suavização exponencial	36
6.1	Suavização exponencial simples	36
6.1.1	Representação por média ponderada	36
6.1.2	Representação por componentes	37
6.1.3	Otimização	37
6.2	Suavização exponencial com tendência	37
6.2.1	Suavização exponencial com tendência amortecida	38
6.3	Suavização exponencial com sazonalidade	38
6.4	Combinação de tendência e sazonalidade	38
6.5	Erros de previsão	38

SUMÁRIO

7	Previsão de séries temporais hierárquicas e séries temporais agrupadas	41
7.1	Diferença entre série temporal hierárquica e agrupada	41
7.2	Abordagem de baixo para cima	43
7.3	Abordagem de cima para baixo	43
7.4	Abordagem intermediária	44
7.5	Representação matricial	44
7.6	Reconciliação ideal	45

Introdução

Este relatório de Iniciação Científica apresenta tópicos iniciais da Econometria, apesar de precisar que o leitor tenha um conhecimento básico de estatística.

No capítulo 1 é abordado sobre a natureza dos dados econométricos, onde é abordado sobre o conceito de modelagem determinística e estocástica, regressão, correlação, de modo introdutório. Também é apresentado duas metodologias distintas, uma para trabalharmos com uma teoria econômica e outra para resolver um determinado problema ou tarefa.

No capítulo 2 é comentado os tópicos que o leitor deve conhecer sobre estatística descritiva, probabilidade e estatística inferencial.

No capítulo 3 é falado sobre a análise de regressão clássica, seja linear ou não linear e seja de uma ou de mais variáveis. São abordados tópicos fundamentais para a abordagem da regressão, como a estimação e a avaliação do modelo.

Com isso, o capítulo 4 traremos uma abordagem sobre uma regressão muito utilizada na Econometria, a regressão de séries temporais. Lá teremos as definições principais, bem como sua formulação como processo estocástico, a importância destes modelos para as previsões, conceitos fundamentais e como avaliar os modelos.

Por fim, temos no capítulo 5 uma abordagem diferenciada, envolvendo previsão de julgamento, em que não há necessidade de se utilizar as ferramentas estatísticas. Veremos abordagens sistemáticas e estruturadas para fundamentar previsões que são baseadas nas experiências profissionais dos envolvidos.

Capítulo 1

A natureza dos dados econométricos

A Econometria, em interpretação literal, significa medição econômica, porém, seu escopo é mais amplo do que apenas a medição. A Econometria trata de um tema interdisciplinar, trazendo estudos da Economia, Economia Matemática, Estatística Matemática e Estatística Econômica (GOLDBERGER, 1964; GUJARATI, 2011).

É importante observar que existem diferentes focos de estudo sobre a Econometria. Seus estudos, e livros, podem ser divididos entre ter um foco na Econometria Teórica ou na Econometria Aplicada. As referências utilizadas nesta iniciação científica foram voltadas à Econometria Teórica, assim, buscamos trabalhar com as teorias que envolvem a Econometria e posteriormente veremos as suas aplicações.

Além disso, por ter grande vínculo com estudos estatísticos, a Econometria pode ser dividida entre clássica e Bayesiana, pois a Estatística se divide nesses dois focos. Aqui estaremos trabalhando com a Econometria através da Estatística clássica.

À seguir, veremos duas metodologias para se trabalhar com a econometria, a primeira para trabalhar com teorias econômicas (GUJARATI, 2011) e a segunda para aplicar em situações problema (HYNDMAN e ATHANASOPOULOS, 2021).

1.1 Metodologia econométrica para teorias econômicas

De acordo com o livro (GUJARATI, 2011), para uma metodologia econométrica clássica no qual temos um modelo à ser testado, podemos fazer o uso dos seguintes passos:

1. Exposição da teoria ou hipótese econômica;
2. Especificação do modelo matemático da teoria;
3. Especificação do modelo estatístico ou econométrico;
4. Obtenção dos dados;
5. Estimação dos parâmetros do modelo econométrico;
6. Teste de hipóteses;
7. Projeção ou Previsão;

8. Uso do modelo para fins de controle ou política.

Vejamos com um exemplo oferecido pelo livro.

1.1.1 Exposição da teoria ou hipótese econômica

A teoria do consumo Keynesiana afirma que “as pessoas estão dispostas, como regra e em média, a aumentar o seu consumo conforme sua renda aumenta, mas não na mesma proporção” (KEYNES, 1936).

Com isso, Keynes postula que a **Propensão Marginal à Consumir (PMC)** que é a taxa de variação da despesa à cada unidade de renda é maior do que zero e menor do que um, pois a despesa tende à aumentar mas sem aumentar mais do que a renda.

Podemos trabalhar em cima dessa teoria para observar como se aplica em uma determinada região.

1.1.2 Especificação do modelo matemático da teoria;

Com a teoria econômica apresentada anteriormente um matemático econômico poderia modelar a seguinte função:

$$Y = \beta_1 + \beta_2 X, \text{ com } 0 < \beta_2 < 1$$

sendo que Y = despesas de consumo, X = renda e β_1 e β_2 são os parâmetros do nosso modelo.

Observe que a restrição de β_2 ser um valor entre zero e um está associado à teoria de Keynes, pois β_2 mede a PMC.

Essa função que associa a despesa com a renda é conhecida como **função consumo** em Economia (GUJARATI, 2011).

1.1.3 Especificação do modelo estatístico ou econométrico;

Um modelo puramente matemático como esse é de pouco interesse para os que trabalham com econometria (chamados econometristas), pois supõe uma relação **determinística** entre o consumo e renda. Porém, em geral as variáveis econômicas costumam **não ser determinísticas** (GUJARATI, 2011).

Além da renda, outros fatores podem afetar o consumo de uma família, como o tamanho da família, região onde vivem, idade dos integrantes, a inflação, entre outros. Assim, para levar em conta esses aspectos, faremos a seguinte alteração:

$$Y = \beta_1 + \beta_2 X + e, \text{ com } 0 < \beta_2 < 1$$

em que e representa o erro da função. Vejamos como interpretar o erro geometricamente. (FIGURA DO ERRO para um modelo de reg)

Este nosso exemplo de modelo se trata de uma Regressão Linear, assunto que será bem trabalhado neste estudo. Lá também exploraremos mais sobre como tratar o erro como uma **variável estocástica** que possuirá propriedades probabilísticas.

1.1.4 Obtenção dos dados (organizado, arrumado e limpo);

Obter os dados é preciso para estimarmos um modelo ideal que se adeque à situação que estamos trabalhando, ou seja, precisamos de dados para obter valores numéricos para os parâmetros β_1 e β_2 , que é o nosso próximo passo. Aqui, é preciso realizar as técnicas de amostragem para obter uma boa coleta de dados, além da organização e limpeza dos mesmos.

1.1.5 Estimação dos parâmetros do modelo econométrico;

Aqui faremos a estimação dos parâmetros com estimativas numéricas através da análise de regressão. O detalhamento deste procedimento está no Capítulo 3, que veremos mais sobre regressão linear.

No livro (GUJARATI, 2011) são utilizados dados de consumo, dado pela Despesa de Consumo Pessoal (DCP) agregada, e renda, dado pelo PIB, tudo em dólares e pelas famílias que vivem nos Estados Unidos. Estimando e aplicando as estimativas no modelo, podemos ter algo como:

$$Y^* = -299,5913 + 0,7218X$$

sendo que Y^* é uma estimativa para o verdadeiro valor Y . Com isso, podemos observar que para cada dólar que se aumenta na renda, *em média*, aumenta-se 0,72 centavos no consumo. É importante ressaltar que esta medida é inexata, portanto, podem haver famílias que aumentaram um pouco mais ou um pouco menos, mas em média a PMC é de 0,72.

1.1.6 Teste de hipóteses;

Neste momento, podemos fazer a inferência estatística, trabalhando com testes de hipóteses para verificar se temos um modelo econométrico adequado. Segundo o economista Milton Friedman, uma teoria ou hipótese que não for verificável com evidências empíricas pode não ser admissível como parte de uma pesquisa científica (FRIEDMAN M., 1953).

Caso o modelo não esteja bem adequado, podemos ter de voltar desde o início. Veremos ao longo do trabalho como avaliar o modelo na prática.

1.1.7 Projeção ou Previsão;

Com um bom modelo, ou seja, com um modelo que não refute os testes de hipóteses e não refute a teoria econômica, podemos utilizar para a previsão de certos eventos relacionados.

Porém, e se no meio do percurso, houver mudanças bruscas que não forem previstas, como por exemplo, interferências políticas. Veremos adiante neste artigo sobre como lidar com estas situações.

1.1.8 Uso do modelo para fins de controle ou de política.

Com as previsões, podemos utilizar para atualizar estratégias, melhorar preparamentos e demais ações. É como saber que irá chover em determinado dia, dessa forma podemos nos preparar para este evento.

1.2 Metodologia para a resolução de um problema

A metodologia acima é ideal caso temos um modelo econômico e queremos adequá-lo à um certo conjunto de dados, que resultará em certos parâmetros.

Porém, nem sempre temos o modelo para ser trabalhado em mente. Por vezes, temos apenas os dados e queremos utilizá-los para a **resolução de um problema** dentro do seu ofício.

Neste caso, há uma grande alteração na metodologia, pois só poderemos ter o modelo após uma coleta e análise dos dados. Para tal, recomenda-se o seguinte fluxo de trabalho m (HYNDMAN e ATHANASOPOULOS, 2021):

1. Determinando o objetivo;
2. Obtenção dos dados (organizado, arrumado e limpo);
3. Visualização dos dados (plotar o gráfico);
4. Definição um modelo;
5. Estimação dos parâmetros;
6. Avaliação do modelo;
7. Utilização do modelo;

A maioria dos itens são trabalhados de forma semelhante à metodologia anterior, por isso, não estarei escrevendo sobre cada passo novamente.

As diferenças aqui é que como não há uma teoria econômica, podemos precisar definir o nosso objetivo com o modelo e precisaremos plotar o gráfico para avaliar um bom modelo.

A definição e compreensão do problema ou objetivo à ser avaliado deve ser sempre o início do projeto que o Econometrista irá trabalhar (HYNDMAN e ATHANASOPOULOS, 2021).

Depois de obtermos os dados, a visualização torna-se uma etapa fundamental para a compreensão do que estamos trabalhando. Podemos já observar certos padrões por um simples gráfico, nos ajudando a escolher um modelo apropriado (HYNDMAN e ATHANASOPOULOS, 2021).

Ao avaliarmos o modelo, podemos decidir se retornamos ao passo 2 caso o modelo não seja satisfatório. O modelo também dependerá muito do uso destinado, como por exemplo, existem os modelos que são mais adequados para fins de previsão.

1.3 A natureza da análise de regressão

Como mencionado anteriormente, a regressão fará parte de nossos estudos, seja ela linear ou não linear. Neste capítulo, examinaremos um pouco de sua natureza.

1.3.1 Modelagem

De acordo com Gujarati,

“A análise de regressão diz respeito ao estudo da dependência de uma variável, a variável dependente, em relação a uma ou mais variáveis, as variáveis explanatórias, visando estimar e/ou

prever o valor médio (da população) da primeira em termos dos valores conhecidos ou fixados (em amostragens repetidas) das segundas.” (GUJARATI, 2011).

Assim, regressão trata de uma modelagem para uma determinada situação em que existe uma variável dependente, que geralmente chamamos de Y , que depende de uma ou mais variáveis independentes X .

A modelagem matemática em si é bem semelhante à ideia de regressão estudada na estatística. Porém, a diferença entre cada abordagem é que existem os modelos que são determinísticos e os que não são, que chamamos de estocásticos ou estatísticos. Na regressão trabalhamos com o segundo caso.

Um exemplo de um modelo determinístico seria algum que trabalhasse com a física clássica, que é determinística, como por exemplo a velocidade de um carro em decorrência de um tempo. Dados as variáveis, podemos calcular onde exatamente o carro estará.

Porém, quando há dificuldade em sua exatidão, havendo interferências de diversos fatores ou quando lidamos com uma amostra de dados, é conveniente trabalharmos com a modelagem estatística. Aqui, lidaremos com variáveis aleatórias (ou estocásticas) inseridas em nosso modelo de regressão que servem para lidar com essa incerteza.

1.3.2 Correlação e Causação

A **correlação** é uma medida de “força” ou “influência” entre duas variáveis em uma associação linear (GUJARATI, 2011). Podemos ter por exemplo, uma correlação entre venda de agasalhos e a temperatura ambiente do momento. Assim, normalmente podemos obter bons modelos de regressão pelo desempenho do grau de correlação entre as variáveis.

Porém, vale ressaltar que apesar de encontrarmos um modelo que ajuste muito bem as variáveis, não podemos concluir que a causa de uma é por conta da outra.

Por exemplo, suponha que em uma determinada vila tenhamos um modelo que se ajusta muito bem a relação entre pais que tomam bebidas alcoólicas com casos de filhos que são atacados por cobras. Neste caso, podemos ter uma alta correlação entre as variáveis, mas podemos observar também, que um não é uma causa do outro. Esse resultado pode existir pois em regiões que são mais quentes, os adultos costumam beber mais e, além disso, as cobras costumam sair mais, resultando em um número maior de casos de picadas.

É importante já termos noção sobre esse tipo de situação e saber que nosso modelo não necessariamente apresentará uma causação, independente se há correlação.

1.3.3 Séries Temporais

Uma **série temporal** é um conjunto de observações de uma variável em diversos momentos do tempo. Podemos pensar em preços de ações, lucros de vendas, prejuízos e qualquer outro tipo de caso que esteja ocorrendo ao longo do tempo, incluindo situações que não envolvam economia e finanças.

Por haver tantos casos de séries temporais, um estudo aprofundado sobre este tipo de dado pode ser bem frutífero. Porém, requer modelos de regressão específicos para que possamos ter bons ajustes e usos.

Assim, no capítulo 4 damos uma atenção exclusiva ao estudo da Econometria para dados de séries temporais.

1.3.4 Regressão Amostral

Na maioria das vezes, não conseguimos ter acesso aos dados de toda a população do conjunto à ser estudado. Dessa forma, é muito difícil obtermos uma regressão que utilize destes dados, ou seja, uma regressão populacional.

Assim, estaremos sempre lidando com amostras da população e consequentemente, estaremos trabalhando com regressões amostrais. Os parâmetros que forem obtidos devem ser o mais próximos possíveis dos parâmetros da regressão populacional.

Portanto, o objetivo primordial da análise de regressão é estimar uma regressão populacional com base na regressão amostral (GUJARATI, 2011).

Capítulo 2

Preliminares sobre estatística

Para esse documento, é preciso que o leitor esteja familiarizado com alguns conceitos de Probabilidade e Estatística. Os principais conceitos estão descritos abaixo.

2.1 Estatística Descritiva

É preciso que o leitor tenha familiaridade com as medidas de centro: Média, Moda, Mediana, bem como as medidas de variabilidade: Variância, Desvio padrão, desvio médio.

Além disso, a variabilidade entre variáveis também será bem importante, onde se trabalha com covariância e correlação, além das medidas de forma, como a assimetria e a curtose.

Representação em gráficos como o histograma também é fundamental.

2.2 Probabilidade

Para a Teoria de Probabilidade é importante entender a definição e propriedades básicas da função de probabilidade, ter conhecimento sobre esperança amostral e conhecer o conceito de distribuição de probabilidade, bem como as distribuições mais conhecidas.

2.3 Inferência estatística

É importante que se saiba um pouco sobre inferência estatística também, como por exemplo, sobre a estimação, teste de hipóteses e criação de intervalos de confiança.

Capítulo 3

Análise de regressão clássica

Como vimos anteriormente, o modelo de regressão trata de um modelo onde relacionamos uma variável dependente Y com uma ou mais variáveis independentes X_i que buscam explicar a variação de Y . Por tratarmos de modelos não determinísticos, o valor obtido no modelo, Y^* , é uma aproximação ao valor de Y .

De modo geral, podemos escrever o modelo da seguinte forma:

$$Y^* = f(X_1, X_2, \dots, X_n).$$

Onde Y^* é uma aproximação para Y e f é uma função que necessita dos valores de cada variável independente X_i , $i \in \{1, \dots, n\}$. Podemos também denotar a função pelo conceito de esperança, de modo que, $f(X_1, X_2, \dots, X_n) = E(Y|X_1, X_2, \dots, X_n)$, ou seja, é o valor esperado de Y dado os valores X_i .

Por exemplo, a função f poderia ser da seguinte forma:

$$f(X_1, X_2, \dots, X_n) = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_n X_n.$$

onde α_j são **parâmetros** que são desconhecidos e precisam ser estimados, também chamados de coeficientes de regressão, para todo $j \in \{0, \dots, n\}$. Iremos estudar mais a fundo sobre como estimar estes parâmetros.

Nosso objetivo é encontrar um modelo no qual essa aproximação seja a melhor possível (GUJARATI, 2011).

3.1 Análise de regressão linear clássica

Uma função é dita linear quando trata-se de uma função cuja o gráfico seja descrito por uma linha reta. Assim, a função $Y = X^2 + 3$ é não linear, pois graficamente, se trata de uma parábola enquanto que $Y = 3X + 2$ pode ser dito como uma função linear. Assim, uma função é linear, se houver uma relação linear entre as variáveis.

Mas o conceito de linearidade pode não ser proposta apenas para as variáveis, mas também para os parâmetros. Por exemplo, $Y = \alpha_0 + \alpha_1 X_1^2$, é linear nos parâmetros mas não em sua variável.

Pois bem, todo momento que é comentado sobre linearidade no livro do Gujarati, é deixado claro desde o começo que não estará se referindo às variáveis do modelo, e sim em relação aos parâmetros, pois acaba sendo uma divisão mais relevante para o estudo de modelos de regressão. O mesmo será feito aqui, ou seja, todo momento que for citado a linearidade, subentende-se que estamos nos referindo aos parâmetros do modelo.

Primeiramente, iremos abordar sobre a análise de regressão linear clássica, para depois adentrarmos sobre o estudo da análise de regressão não linear clássica.

3.1.1 Análise de regressão linear clássica de uma única variável independente

Iremos nos restringir à modelos com apenas uma variável independente X para ficar mais fácil de entender a ideia de como funciona uma regressão linear.

Como vimos anteriormente, estaremos lidando com um modelo no qual seu resultado, Y^* , seja uma aproximação do valor Y . Assim, para cada Y_i , teremos um valor Y_i^* que assumirá um certo valor de erro e_i . Podemos ter então, que:

$$Y_i = Y_i^* + e_i.$$

O valor e_i é conhecido como **termo de erro estocástico**. Assim, cada valor aproximado pelo modelo Y_i^* somado com seu erro e_i pode resultar no valor original Y .

Significado do termo ‘estocástico’

Esse erro, que na verdade não é exatamente um erro, representa todas as variáveis omitidas no modelo, mas que coletivamente afetam o valor Y . Mas então, por que motivos é interessante o seu uso? Não seria melhor colocar todas as variáveis possíveis? Há algumas razões para isso (GUJARATI, 2011):

1. Podemos desconhecer quais são **todas** as outras variáveis que afetam Y . Como por exemplo, sabemos que a temperatura pode afetar no consumo de energia, mas quais são todas as variáveis que afetam?
2. Podemos ter indisponibilidade nos dados dessas variáveis, mesmo que possamos conhece-las.
3. Podemos ter variáveis que não afetam tanto no valor Y , podendo resultar num trabalho maior e menos recompensador incrementá-las.
4. Há a presença do caráter aleatório do comportamento humano, pois ninguém se comporta da mesma forma.
5. Pode ajustar os erros de medição de certas variáveis.
6. Quanto mais simples e substancial, melhor.
7. Não é fácil determinar o modelo de regressão mais adequado.

Por todas essas razões, o termo de erro estocástico e_i assume um papel fundamental na análise de regressão, como veremos no decorrer do livro.

A estimação

Seja um modelo de regressão linear dado por

$$Y_i^* = \alpha + \beta X_i$$

em que $Y_i = Y_i^* + e_i$.

Como podemos então estimar os parâmetros α e β do nosso modelo de regressão de modo que o nosso resíduo estocástico seja o menor possível? Há dois principais métodos para se estudar inicialmente.

- Método dos Mínimos Quadrados Ordinários
- Método da Máxima Verossimilhança

O primeiro método é atribuído a Carl Friedrich Gauss, um matemático alemão, que é amplamente utilizado por ser matematicamente mais simples. (GUJARATI, 2011).

Sob certas hipóteses, este método tem propriedades estatísticas muito atraentes que estaremos trazendo mais adiante.

Se queremos reduzir os erros do modelo, precisamos pensar em um jeito de obter valores baixos do erro e_i . Neste caso, ele é dado por

$$e_i = Y_i - \alpha - \beta X_i.$$

Assim, podemos pensar em um modelo que ao somar todos os erros, nos retorne o menor valor possível. Porém, há dois fatores que prejudicam este critério. O primeiro, é que como o valor estimado Y_i^* pode tanto ser maior ou menor que o valor real Y , os nossos erros possam ser tanto negativos quanto positivos e ao somarmos todos, há perigo deles se anularem. Além disso, dessa forma há um peso igual se houve uma distância de erro de 5 unidades ou de 40.

Por isso, neste método estaremos elevando ao quadrado cada termo de erro, eliminando os dois pontos abordados anteriormente. Assim, como queremos que este seja o menor valor possível, podemos agora entender melhor o porquê do nome do método.

Queremos encontrar α e β de modo que

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2$$

seja o menor possível. Isso também garante as propriedades que este método tem, que serão abordadas logo adiante (GUJARATI, 2011).

Primeiramente, vamos abrir o lado direito dessa nossa equação.

$$\begin{aligned} \sum_{i=1}^n e_i^2 &= \sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2 \\ &= \sum_{i=1}^n (Y_i^2 + \alpha^2 + \beta^2 X_i^2 - 2Y_i\alpha - 2Y_i\beta X_i + 2\alpha\beta X_i) \\ &= \sum_{i=1}^n Y_i^2 + \sum_{i=1}^n \alpha^2 + \sum_{i=1}^n \beta^2 X_i^2 - \sum_{i=1}^n 2Y_i\alpha - \sum_{i=1}^n 2Y_i\beta X_i + \sum_{i=1}^n 2\alpha\beta X_i \\ &= \sum_{i=1}^n Y_i^2 + n\alpha^2 + \beta^2 \sum_{i=1}^n X_i^2 - 2\alpha \sum_{i=1}^n Y_i - 2\beta \sum_{i=1}^n Y_i X_i + 2\alpha\beta \sum_{i=1}^n X_i \end{aligned} \tag{3.1}$$

Assim, podemos realizar as derivadas em relação a α e β e igualar a soma dos erros a zero. Ao derivar em relação a α obtemos o seguinte:

$$\begin{aligned}
 0 &= 2n\alpha - 2 \sum_{i=1}^n Y_i + 2\beta \sum_{i=1}^n X_i \\
 2n\alpha &= 2 \sum_{i=1}^n Y_i - 2\beta \sum_{i=1}^n X_i \\
 \alpha &= \frac{\sum_{i=1}^n Y_i - \beta \sum_{i=1}^n X_i}{n} \\
 \alpha &= \bar{Y} - \beta \bar{X}
 \end{aligned} \tag{3.2}$$

em que \bar{Y} é a média dos Y_i e \bar{X} é a média dos X_i .

De mesmo modo, podemos derivar em relação a β . Irei omitir o procedimento, por ser da mesma forma. Assim, obtemos o seguinte:

$$\beta = \frac{\sum_{i=1}^n ((X_i - \bar{X})(Y_i - \bar{Y}))}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Assim, temos como estimar os parâmetros para a regressão linear que queremos, tudo baseado nos dados que iremos trabalhar.

Os estimadores de mínimos quadrados possuem propriedades numéricas além das estatísticas que veremos adiante. Eles são **estimadores pontuais**, ou seja, apresenta um único valor estimado, ao contrário de um intervalo de previsão. A regressão linear feita por estes estimadores passa pelas médias amostrais de Y_i e X_i , a média do valor estimado Y_i^* é igual a média de Y_i , a média dos erros é igual a zero e não há correlação dos erros com a variável dependente ou independente, fator que poderia ocasionar em um modelo com viés (GUJARATI, 2011).

O método da máxima verossimilhança é mais complexo de se trabalhar, por isso, será omitido aqui.

Modelo Clássico de Regressão Linear (MCRL)

Na análise de regressão, nosso objetivo não é apenas obter os parâmetros da regressão, mas o de obter inferências dos verdadeiros parâmetros, ou seja, dos parâmetros da regressão populacional. O **modelo clássico de regressão linear** parte de 7 hipóteses:

1. Hipótese: O modelo é **linear** nos parâmetros.
2. Hipótese: Ou os valores de X são fixados ou então, se não forem, devem ser independentes do termo de erro, isto é, $cov(X_i, e_i) = 0$.
3. Hipótese: A média dos termos de erro e_i é zero.
4. Hipótese: Homocedasticidade (variância constante) de e_i .
5. Hipótese: Não há autocorrelação entre os termos de erro.

6. Hipótese: O número de observações n deve ser maior que o número de parâmetros.

7. Hipótese: Os valores de X não devem ser os mesmos.

Nossa discussão sobre as hipóteses subjacentes ao modelo clássico de regressão linear agora está completa.

3.1.2 Propriedades estatísticas dos estimadores de mínimos quadrados

Pelo Teorema de Gauss-Markov, o método dos mínimos quadrados possui a propriedade de ser **o melhor estimador linear não viesado (BLUE, em inglês)** (GUJARATI, 2011).

Ou seja, trata-se de um estimador para regressão **linear** que **não é viesado**, ou seja, a média do parâmetro obtido pela amostra é igual ao parâmetro obtido pela população, e é **eficiente**, que implica que dentro de todos os estimadores não viesados, este é o que possui variância mínima.

Coefficiente de Determinação R^2

Até o momento, estivemos preocupados com a questão da estimação do modelo. Agora vamos trazer um olhar para a qualidade do ajustamento.

O **coeficiente de determinação** R^2 , que varia de zero à um, é dado pelo quadrado do coeficiente de correlação, que varia entre o um negativo e o um positivo. Esse coeficiente é uma medida resumida que diz quanto que a linha de regressão se ajusta aos dados, sendo que 1 indica que há uma relação determinística, pois a regressão passa por todos os pontos, e 0 que não há relação linear entre as duas variáveis, ou seja, não importa o quanto variemos o valor X , a variável Y não receberá influência, e a reta ajustada acaba sendo uma linha horizontal.

Outra forma de se obter este coeficiente é pela razão entre a soma do quadrado da diferença entre os valores ajustados com a média do observado $\sum(y_t^* - \bar{y})^2$ e a soma do quadrado da diferença entre os valores observados com a sua média $\sum(y_t - \bar{y})^2$.

Vale observar que, assim como a correlação, o coeficiente de determinação é uma medida para relações lineares entre as variáveis.

Coefficiente de Determinação R^2 ajustado

O R^2 pode ser interessante para avaliar modelos de regressão que possuem a mesma quantidade de variáveis. Caso não tenham, ele não é aconselhável pois possui uma grande falha.

O coeficiente R^2 tende à aumentar conforme aumentamos a quantidade de variáveis de um modelo.

Por isso, existe o R^2 ajustado que busca lidar com este problema. Este valor é calculado por:

$$\bar{R}^2 = 1 - (1 - R^2) \left(\frac{T - 1}{T - k - 1} \right)$$

onde T representa a quantidade de observações e k a quantidade de variáveis independentes. Novamente, um auto valor de \bar{R}^2 indica um bom modelo estimado e dessa vez, podemos comparar com modelos mais diferentes.

Modelo Clássico de Regressão Linear Normal (MCRLN)

O MCRLN supõe que cada e_i seja distribuído normalmente, com média zero e variância constante σ^2 . Mas por que utilizamos esta hipótese? Por alguns motivos (GUJARATI, 2011):

1. Pelo Teorema Central do Limite, se há um grande número de variáveis aleatórias independentes e com distribuição idêntica, então a distribuição de suas somas tende à uma distribuição normal, conforme o número de variáveis for grande. Esta é uma justificativa teórica para esta hipótese.
2. Com essa hipótese, qualquer função linear de variáveis com distribuição normal, também possui distribuição normal. Assim, os dois parâmetros também possuem distribuição normal.
3. A distribuição normal é conhecida e simples.
4. Ela é fundamental para utilizar testes estatísticos, como t de Student, F e χ^2 .

Claro que na prática ainda é preciso avaliar a normalidade dos erros por algum teste estatístico, para verificar se é uma suposição adequada ou não.

3.1.3 Análise de regressão linear clássica múltipla

Raramente uma teoria econômica é tão simples para ser utilizado apenas duas variáveis. Assim, é imprescindível que se estude modelos com n variáveis.

A ideia é semelhante ao de duas variáveis, temos um Y^* que aproxima o verdadeiro Y em função das variáveis X_i e dos parâmetros α_j .

Modelo Clássico de Regressão Linear (MCRL) atualizado

Para o caso de um modelo com mais variáveis independentes, temos mais duas hipóteses para o MCRL:

1. Não há relação linear exata entre as variáveis independentes;
2. Modelo é corretamente especificado. Isso quer dizer que, caso tenhamos um modelo que seja verdadeiro ao nosso problema e omitimos uma variável importante, este novo modelo ocasionará em viés baseado na especificação do modelo;

A estimação

Os estimadores de mínimos quadrados seguem a mesma lógica feita anteriormente. É preciso derivar a soma do quadrado dos erros em relação à cada parâmetro à ser estimado.

Nota-se que sempre que há um novo parâmetro, os demais precisarão de uma nova fórmula. Para encontrar, basta seguir o procedimento feito anteriormente.

Avaliação

O coeficiente de determinação R^2 é obtido da mesma forma e possui a mesma ideia aqui também.

3.2 Análise de regressão não linear clássica

Consideremos o seguinte modelo de regressão:

$$Y = \beta_1 X_2^{\beta_2} X_3^{\beta_3} + e_i.$$

Não há como transformar a equação de modo que a regressão seja linear. Assim, uma abordagem diferente pode ser necessária para trabalhar com estes casos.

Ao realizar o método dos mínimos quadrados ordinários, obtemos equações que não podem ser resolvidas analiticamente. Por conta disso, precisamos utilizar **métodos iterativos** que são métodos que utilizam de certos algoritmos que aproximam um resultado. Esses algoritmos em geral, podem ser aplicados por algum computador, bastando escolher o método requisitado.

Um exemplo de algoritmo utilizado de regressão não linear é utilizando redes neurais artificiais, que possuem utiliza uma função não linear para transformar os dados e obter bons parâmetros minimizando o erro por várias iterações (HYNDMAN e ATHANASOPOULOS, 2021).

Capítulo 4

Econometria de séries temporais

Veremos neste capítulo sobre como trabalhar com a Econometria utilizando regressão de séries temporais. Para começar, nada melhor do que definir o que é uma série temporal e a importância de seus modelos.

4.1 Séries Temporais

Iremos apresentar formalmente a definição de uma série temporal através do chamado Processo Estocástico, mas o seu conceito em si pode ser facilmente explicado.

Uma série temporal de uma variável é quando estamos observando seus valores sequencialmente ao longo do tempo (HYNDMAN e ATHANASOPOULOS, 2021). Vale ressaltar que iremos trabalhar aqui apenas com intervalos de tempo igualmente espaçados.

Um modelo de série temporal se difere de um modelo explanatório, onde seu intuito é apresentar variáveis que possam explicar a variável dependente, com o foco em responder o porquê daquilo ocorrer. Por exemplo, podemos ter um modelo explanatório que represente o salário de um indivíduo baseado em sua idade, escolarização e sexo. O modelo pode ser descrito da seguinte forma:

$$Renda = f(\text{idade}, \text{escolarização}, \text{gênero}, \text{erro})$$

O modelo de série temporal utiliza os valores passados da própria variável, não se baseando em variáveis externas. Assim, podemos escrever este modelo da seguinte forma:

$$X_{t+1} = f(X_t, X_{t-1}, X_{t-2}, \dots, \text{erro})$$

em que $t + 1$ representa o próximo momento, t o momento atual, $t - 1$ o momento anterior e assim por diante.

Mas também pode haver modelos que combina os dois. Eles são conhecidos como modelos de regressão dinâmica e um exemplo deste tipo de modelo pode ser o seguinte:

$$Renda_{t+1} = f(Renda_t, \text{idade}, \text{escolarização}, \text{gênero}, \text{erro})$$

em que utilizamos o valor de renda passado, junto com a informação da idade, escolarização e gênero.

Em critérios de **previsão** é mais aconselhável utilizar os modelos de séries temporais ao invés de modelos explanatórios ou dinâmicos, pois (HYNDMAN e ATHANASOPOULOS, 2021):

1. Pode ser difícil atribuir variáveis que expliquem a mudança de uma variável, mas mesmo que sabemos quais são, pode ser difícil medir cada variável.
2. É preciso conhecer os valores futuros de cada variável extra.
3. Geralmente, para situações que necessitam previsões, é mais interessante prever o que vai acontecer com precisão, que é o que a série temporal oferece, do que saber o porquê daquilo ocorrer, que é o que o modelo explanatório oferece.
4. Para previsões, os modelos de séries temporais são mais precisos que os demais.

Por isso, a **Econometria de Séries Temporais é amplamente estudada e trabalhada para a finalidade de previsões** e, por consequência, estarei destinando um capítulo apenas para isso.

4.2 Processos Estocásticos

Para fundamentar nossos estudos sobre séries temporais e obter análises e resultados matemáticos, precisamos dar uma olhada sobre os conceitos que envolvem os chamados Processos Estocásticos.

Definição 4.1. (*Processo Estocástico*) *Seja T um conjunto qualquer. Um processo estocástico é uma família $\{X(t), t \in T\}$ tal que para cada $t \in T$, $X(t)$ é uma variável aleatória.*

Vimos que uma série temporal é um conjunto de observações dos valores que uma variável assume em diferentes momentos do tempo. Assim, podemos definir uma série temporal como um processo estocástico em que T representa o tempo. Esse tempo pode ser representado de forma discreta, usualmente colocando $T \subset \mathbb{Z}$, ou de forma contínua, usualmente colocando $T \subset \mathbb{R}$.

Definição 4.2. *Processo Estocástico Estacionário* *Um processo estocástico $\{X(t), t \in T\}$ é dito estacionário se, e somente se:*

1. $E\{X(t)\} = \mu$, para todo $t \in T$. Ou seja, a média é constante em todo o tempo;
2. $Var\{X(t)\} = \sigma^2$ constante, para todo $t \in T$. Ou seja, a variância é constante em todo o tempo;
3. $Cov\{X(t_1), X(t_2)\} = Cov\{X(t_1 + k), X(t_2 + k)\} = \gamma(k)$ para todo $t_1, t_2, k \in T$. Ou seja, a covariância depende apenas do intervalo entre os dois valores.

Assim, um processo estocástico estacionário que varia no tempo é chamado de **série temporal estacionária**. Ao olhar um gráfico podemos entender melhor este conceito de estacionariedade.

Na figura 4.1, temos três série temporais. A primeira representa uma série temporal estacionária pela definição vista anteriormente. A segunda representa uma série que não é estacionária pela média não ser constante em todo o tempo, pois podemos ver que a série tende à crescer. A última representa uma série que não é estacionária por conta da variância, que muda com o tempo, apesar da média ser constante.

No geral, as séries econômicas e de finanças costumam não ser estacionárias, principalmente pela média não ser constante, apresentando quedas e altas. Assim, os estudos de Econometria de séries temporais são voltados para modelos que possam lidar com essa ausência de estacionariedade ou com métodos que transformam a série para que possibilite essa propriedade, para ser mais fácil de trabalhar.

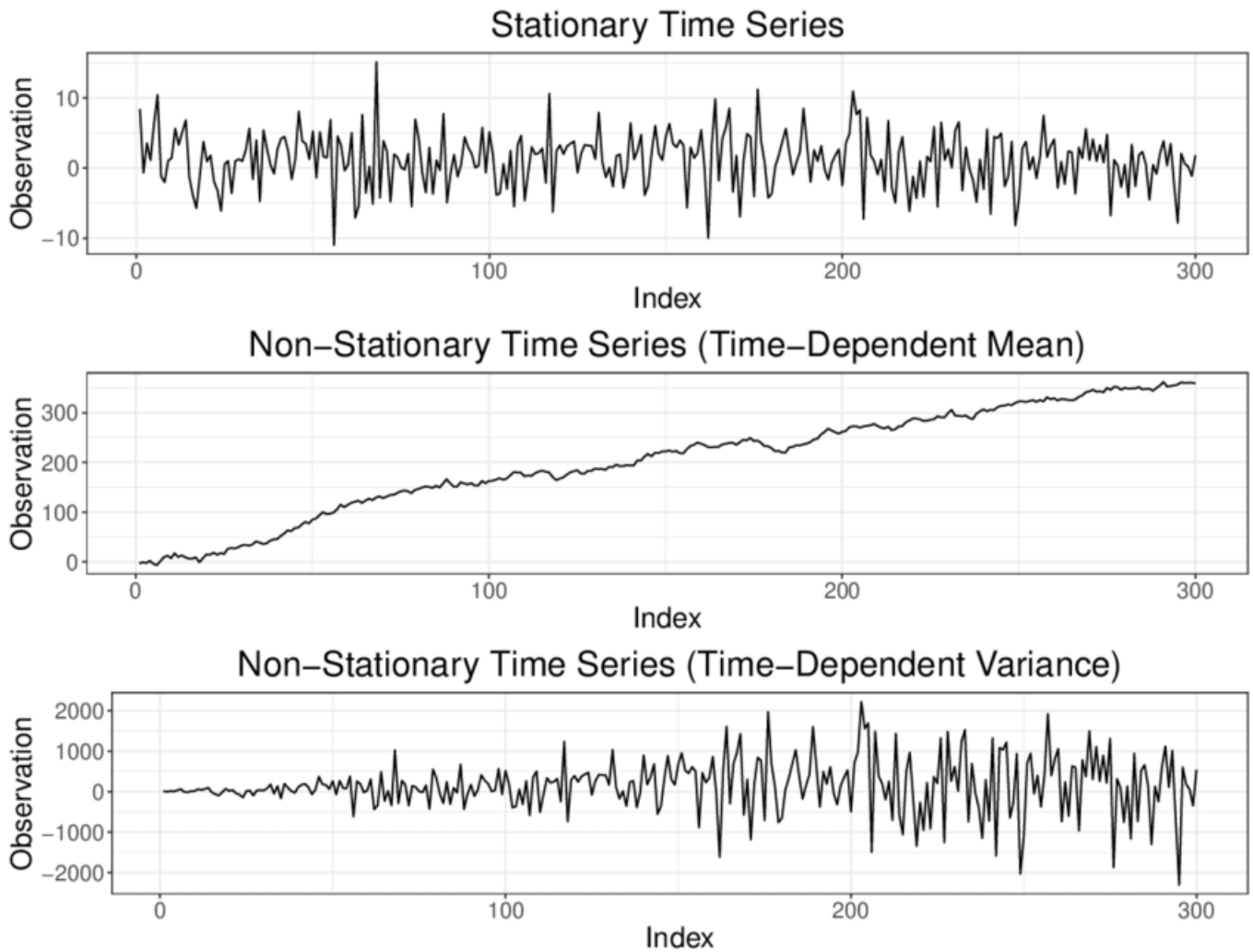


Figura 4.1: Comparação da estacionariedade entre três séries temporais. Fonte: BAUER, A., Automated Hybrid Time Series Forecasting: Design, Benchmarking, and Use Cases, 2020.

4.3 Previsão de séries temporais

As previsões sempre foram algo que fascinava as pessoas por muitos anos. A prática de prever algo era ligado à uma ideia divina, associando um oráculo ou um profeta, e muitas outras vezes, ligado à algo criminoso também.

Apesar disso, a previsão é algo real e possível de se fazer com base nas técnicas matemáticas e estatísticas, mas sempre com uma incerteza presente. De qualquer forma, a possibilidade de prever o que tem grandes chances, ou probabilidade alta, de acontecer é uma vantagem para qualquer área de estudo, principalmente na Economia (HYNDMAN e ATHANASOPOULOS, 2021).

4.3.1 Mas o que podemos prever?

Podemos prever qualquer coisa? Bom, podemos ter a certeza de que o sol irá nascer amanhã, mas não podemos prever os números de uma loteria. Então, como saber o que é possível de se

prever e o que não é possível?

A avaliação de se é possível prever ou não pode depender de vários fatores como por exemplo, o quão bem entendemos os fatores que interferem na variável que queremos prever, quantos dados estão disponíveis sobre o assunto, o quão semelhante o futuro é do passado e se as previsões podem afetar o que estamos querendo prever (HYNDMAN e ATHANASOPOULOS, 2021).

Uma etapa fundamental é avaliar se a nossa previsão não pode ser melhor do que decidir o que vai acontecer lançando uma moeda. Boas previsões costumam detectar **padrões** que ocorreram, desde que as mesmas sejam significativas, e não meras flutuações. Iremos ver como avaliar se um certo padrão é uma flutuação insignificante ou não.

4.3.2 Previsão, meta e planejamento

A previsão pode ser confundida com meta e planejamento, por isso, iremos definir aqui o que cada um significa (HYNDMAN e ATHANASOPOULOS, 2021):

Previsão: prever o futuro com a maior precisão possível, conhecendo os dados do passado e os eventos futuros.

Meta: o que você gostaria que acontecesse. Deve estar ligada à previsão e ao planejamento.

Planejamento: é uma resposta à diferentes previsões e metas. Determina as ações apropriadas para o ocorrido.

4.3.3 Previsão pontual e previsão intervalar

Na Econometria de séries temporais, também estaremos lidando com previsões que podem ser tanto pontuais quanto intervalares. Uma previsão pontual representa um único valor, geralmente dado pela esperança das possibilidades possíveis, ou seja, o valor com mais chance de acontecer.

Já a previsão intervalar, é justamente dada pelo intervalo de confiança, informando uma região de valores sob uma certa probabilidade. Geralmente, costuma-se calcular uma probabilidade entre 0,95 e 0,8. Vale ressaltar, que quanto mais longe queremos prever, maior será o nosso intervalo de previsão, tornando cada vez mais impreciso de se prever. A previsão de chuva, por exemplo, é mais precisa para o próximo dia, do que para dias mais distantes.

4.4 Autocorrelação

A autocorrelação mede a correlação linear entre os valores da mesma variável em períodos diferentes. Podemos obter diversos coeficientes de autocorrelação em uma série temporal, em que a correlação é usualmente calculada entre o valor atual e o valor k vezes anterior. Por exemplo, a autocorrelação r_1 é dado pela correlação entre y_t e y_{t-1} . Assim, r_k é dado por

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k+1}^T (y_t - \bar{y})^2}$$

em que T representa o comprimento da série temporal.

Podemos obter uma função de autocorrelação (ACF - Autocorrelation function) que informa a autocorrelação em função de k , que representa o atraso. Assim, $k = 1$ é o valor imediatamente anterior.

O gráfico da ACF é chamado de **correlograma**. Vejamos o correlograma de uma produção de cerveja na Austrália:

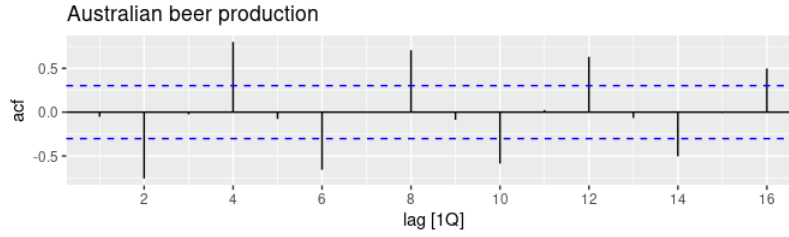


Figura 4.2: ACF de uma produção de cerveja na Austrália. Fonte: Hyndman, R.J., & Athanassopoulos, G., Forecasting: principles and practice [1].

Podemos observar padrões até mesmo no ACF, como neste exemplo mesmo. Pelo ACF, podemos observar que a venda de cervejas possui autocorrelação positiva com o que acontece à um ano, no mesmo mês, e negativa com o semestre anterior. Neste caso, facilmente podemos entender que a temperatura influencia muito na venda de cervejas, bem como as estações do ano.

A venda no verão de um ano possui uma alta assim como a venda no verão do ano anterior. Este padrão será discutido mais adiante.

4.5 Ruído branco

O **ruído branco** é uma série temporal que não apresenta autocorrelação ou caso apresente, a autocorrelação será mínima. Esperamos que as autocorrelações sejam próximas de zero, ou pelo menos, dentro do intervalo $\pm \frac{2}{\sqrt{T}}$ (HYNDMAN e ATHANASSOPOULOS, 2021).

Se as autocorrelações vistas no correlograma passar desse intervalo, então provavelmente **não é** um ruído branco (mais especificamente, se mais de 5% das autocorrelações passarem do intervalo).

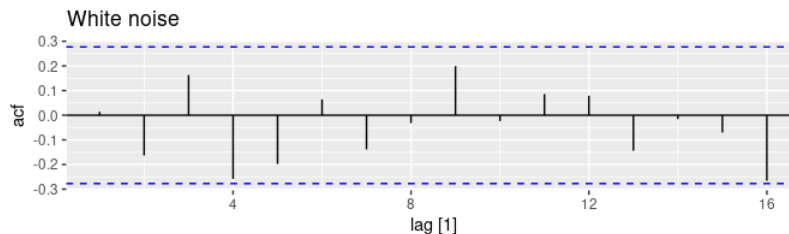


Figura 4.3: Correlograma de um ruído branco. Observemos que o intervalo descrito é mostrado pelas linhas pontilhadas azuis. Fonte: Hyndman, R.J., & Athanassopoulos, G., Forecasting: principles and practice [1].

4.6 Decomposição de séries temporais

Para a análise de séries temporais poderemos estar lidando com vários padrões que nos fornecem informações valiosas para efetuar uma previsão. Há três padrões que possuem destaque nisso, são a tendência, a sazonalidade e o cíclico. Descreveremos sobre cada padrão à seguir (HYNDMAN e ATHANASOPOULOS, 2021).

Tendência: Uma tendência existe quando há um aumento ou diminuição de longo prazo nos dados, não necessariamente linear. Às vezes há alterações no sentido de tendência também, como por exemplo, uma tendência de alta se tornar tendência de baixa. Quando há tendência, a média deixa de ser constante e consequentemente, a série não é estacionária.

Sazonalidade: Um padrão sazonal é aquele que apresenta alteração em uma certa temporada do ano. Assim, em cada ciclo de tempo haverá a presença desse padrão. Por exemplo, a venda de sorvetes é influenciada pelos meses mais quentes, ou seja, no verão. Assim, sempre há alta na venda nos meses do verão em todos os anos. Se há uma alta força de sazonalidade, a variância é constante, que é uma das condições para que a série seja estacionária(?).

Cíclico: Um ciclo ocorre quando há flutuações com muitas subidas e descidas sem haver uma frequência fixa, que é o caso da sazonalidade. Esse padrão também se difere do anterior pela sua duração ser maior e ter magnitudes que variam mais.

Além de poder verificar através do gráfico de séries temporais, também podemos identificar a tendência e a sazonalidade no correlograma.

Há presença de tendência na série temporal quando há correlação alta e positiva para os valores anteriores mais próximos, que vai diminuindo até os valores mais distantes.

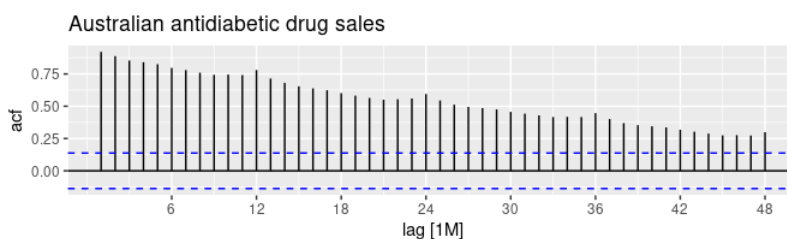


Figura 4.4: Correlograma indicando possível tendência. Fonte: Hyndman, R.J., & Athanasopoulos, G., Forecasting: principles and practice [1].

A sazonalidade é presente quando há alta correlação para valores de períodos sazonais anteriores, indicando a influência. Um exemplo disso é no correlograma da venda de cervejas visto anteriormente.

Com estes padrões, podemos realizar uma **decomposição da série temporal**. A decomposição pode ser feita de forma aditiva ou multiplicativa.

Decomposição aditiva:

$$y_t = S_t + T_t + R_t.$$

Decomposição multiplicativa:

$$y_t = S_t T_t R_t.$$

em que y_t é a variável da série temporal original, S_t representa a componente sazonal, T_t representa o ciclo de tendência, ou seja há o padrão de tendência e de ciclos nesta componente e R_t é o componente residual, todos no tempo t .

Veremos formas diferentes de se trabalhar com a decomposição da série temporal, desde as mais simples até as mais complexas e atuais.

4.6.1 Séries ajustadas sazonalmente

São dadas por $y_t - S_t$, no caso aditivo, e y_t/S_t , no caso multiplicativo. Ou seja, são séries temporais em que é descartado a componente sazonal.

Essa série temporal passa por “dentro” da série temporal original.

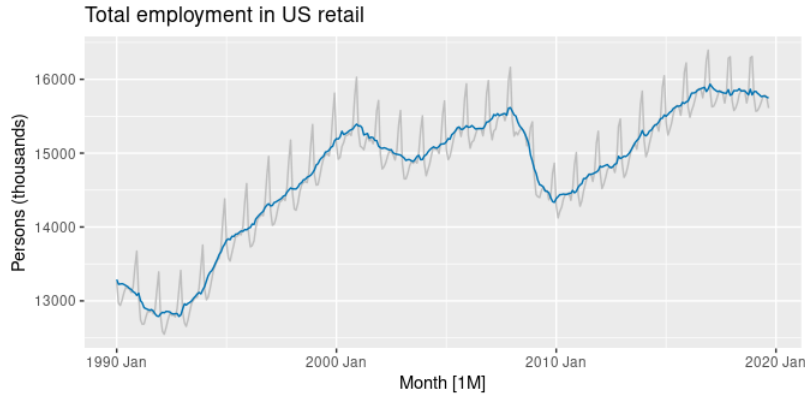


Figura 4.5:

De acordo com o livro [1], os economistas costumam se interessar por variações não sazonais. Porém, isso depende do propósito de estudo também.

4.6.2 Suavização de Média Móvel

Podemos utilizar a média móvel como uma aproximação para a curva de tendência. Uma média móvel de ordem m pode ser escrita da seguinte forma:

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j}$$

sendo que $m = 2k + 1 \Rightarrow k = \frac{m-1}{2}$. Ou também pode ser escrito por:

$$\hat{T}_t = \frac{1}{m} \sum_{j=0}^{m-1} y_{t-j}.$$

Em ambos os casos temos a média de valores passados dos dados, a diferença é que no primeiro formato calculamos a média utilizando o valor atual junto com os próximos k valores e os k valores anteriores, enquanto que no segundo, utilizamos o valor atual junto com os $m - 1$ valores passados para calcular a média.

Uma média elimina alguma das aleatoriedades dos dados, tornando a tendência mais suave (fica menos áspero).

Este método é utilizado para a decomposição clássica de séries temporais que veremos à seguir.

4.6.3 Decomposição clássica

Para a decomposição clássica fazemos os seguintes procedimentos (HYNDMAN e ATHANASOPOULOS, 2021):

1. Calculamos a componente de tendência aproximada T_t^* pela suavização de média móvel;
2. Calculamos a série sem tendência: $y_t - T_t^*$;
3. Para a componente sazonal, calculamos a média de todos os valores sazonais da série sem tendência. Por exemplo, se a sazonalidade é de ano por ano, então calculamos a média dos valores de todos os meses de Março de todos os anos disponíveis da série sem tendência;
4. O resíduo aproximado é dado pela diferença da série com a tendência e sazonalidade aproximados;

Para a decomposição multiplicativa é um processo semelhante, a única diferença é que ao invés de realizar subtrações, são realizadas divisões.

Embora a decomposição clássica ainda seja amplamente utilizada, não é recomendada, pois agora existem vários métodos muito melhores os (HYNDMAN e ATHANASOPOULOS, 2021). Porém, entender sobre como este método funciona, pode nos ajudar a entender os demais também.

4.7 Intervalo de previsão

Um intervalo de previsão equivale ao conceito de intervalo de confiança, só que para valores futuros, que desejamos prever.

Por exemplo, se assumirmos que a distribuição dos valores futuros é normal, e considerando uma previsão Naive em que se repetirá o valor anterior, um intervalo de 95% de confiança do próximo valor pode ser dado por $\hat{y}_t \pm 1,96\hat{\sigma}_{t+1}$, em que $\hat{\sigma}$ é uma estimativa para o desvio padrão no tempo $t + 1$.

4.8 Avaliando a previsão pontual

Aqui veremos como avaliar a previsão pontual de cada modelo, ou seja, a média da distribuição de previsão de cada modelo. Por ser pontual, teremos um único valor para cada previsão, ao invés de um intervalo de valores.

4.8.1 Dados para treinamento e dados para testagem

Para avaliar a nossa previsão é preciso utilizar dados em que já sabemos o que ocorreu, assim, podemos observar o desempenho do modelo. Porém, é importante fazer a seguinte separação nos dados que formos utilizar na modelagem: os dados para treinamento e os dados para a testagem.

Treinamento: Aqui temos os dados que serão utilizados para treinar o modelo, estimando os devidos parâmetros que serão utilizados.

Testagem: Estes, por outro lado, não devem ser usados para treinar o modelo, mas para avaliar o modelo. Ou seja, comparamos os valores destes dados com os valores ajustados pelo modelo. Pela natureza da previsão, o conjunto dos dados de teste deve ser futuro ao conjunto dos dados de treinamento.

É importante fazer essa distinção para que a avaliação seja confiável. Além disso, sempre que dissemos **erros de resíduos** que foram utilizados para verificar se é um bom modelo, se tratam de erros entre os dados de treinamento e o que foi ajustado no modelo. Esses erros se diferem dos **erros de previsão** que são dados pela diferença entre os dados de teste com o que foi ajustado, que servem para avaliar a precisão da previsão do modelo. Ambos se tratam de erros do modelo, mas cada um tratará de uma finalidade diferente.

4.8.2 Métodos que dependem da escala

Medidas de precisão que se baseiam puramente no erro de previsão e_t **não podem** ser utilizados para comparar o desempenho de séries diferentes. Afinal, este erro carrega a escala em que as variáveis são dadas. Os métodos que dependem da escala podem ser úteis para avaliar diferentes modelos para a mesma série temporal, pois assim, não haverá problemas na escala dos dados.

Os métodos mais comuns são o MAE (Mean Absolute Error), que representa a média do erro absoluto médio, e RMSE (Root Mean Squared Error), que representa a raiz do erro quadrático médio. Os métodos são dados por:

$$MAE = \text{média}(|e_t|)$$
$$RMSE = \sqrt{\text{média}(e_t^2)}.$$

4.8.3 Métodos para erros percentuais

Para retirar a unidade dos erros, podemos pensar em erros baseados em porcentagem. Assim, estaremos lidando com erros dados por $p_t = e_t/y_t$, de modo que se $e_t = 0$ então não há nenhum erro e consequentemente $p_t = 0$ também, e quando $e_t = y_t$, ou seja, o erro foi totalmente o valor original dos dados, então $p_t = 1$.

Por serem livres de unidades, podemos realizar comparações entre diferentes séries temporais.

O método mais comum é o MAPE (Mean Absolute Percentage Error), que é o mesmo que o MAE para erros percentuais. Este método é dado por:

$$MAPE = \text{média}(|p_t|).$$

Porém, há duas desvantagens: não podemos utilizar este método quando $y_t = 0$, pelo fato de p_t não existir. E também, valores de y_t muito próximos de 0 podem fazer a porcentagem atingir valores extremos.

4.8.4 Métodos para erros escalados

Proposto por Hyndman e Koehler (2006) como alternativa para os erros percentuais. Eles propõem em escalar os erros baseado no método MAE.

A ideia é comparar os erros do modelo com os erros do modelo NAIVE. Isso é feito da seguinte forma:

$$q_t = \frac{e_t}{MAE(NAIVE)}$$

ou seja, dividimos cada erro do novo modelo pelo MAE do modelo NAIVE. Observemos que:

1. Se $q_t < 1$, então o modelo terá uma previsão melhor que o NAIVE para aquele momento t .
2. Se $q_t > 1$, então o modelo terá uma previsão pior que o NAIVE para aquele momento t .

E para o caso de séries sazonais, podemos utilizar o NAIVE Sazonal no lugar e realizando o mesmo procedimento.

Depois disso, podemos calcular o MAE e o RMSE desses valores, que denotaremos por MASE (Mean Absolute Scaled Error), que representa a média dos erros escalados absolutos, e RMSSE (Root Mean Squared Scaled Error), que representa a raiz do erro escalado quadrático médio.

$$MAE = \text{média}(|q_t|)$$
$$RMSE = \sqrt{\text{média}(q_t^2)}.$$

4.9 Avaliando a previsão intervalar

Vimos alguns métodos para avaliar a previsão pontual dos nossos modelos. Vejamos agora como estar avaliando os intervalos de previsão desses modelos, ou seja, o desempenho da previsão intervalar.

4.9.1 Pontuação dos quantis (ou percentis)

Suponha que temos um intervalo de previsão de 80% de confiança para os valores futuros. Assim, o limite inferior desse intervalo separa 10% da distribuição de probabilidade à esquerda, ou seja, é o 10º percentil (ou o 0,1 quantil). Assim, temos uma esperança maior de que os dados estejam acima deste quantil.

Generalizando, a probabilidade do valor futuro y_t ser menor que o valor do p quantil no tempo t , que denotaremos por $f_{p,t}$, é dado por p . Ou seja, há uma probabilidade p de que $y_t < f_{p,t}$. Por exemplo, com o 0.1 quantil, temos 0,1 de probabilidade de que $y_t < f_{0.1,t}$.

Assim, a nossa **Pontuação de quantil** é dada por:

$$Q_{p,t} = \begin{cases} 2(1-p)(f_{p,t} - y_t), & \text{se } y_t < f_{p,t}, \\ 2p(y_t - f_{p,t}), & \text{se } y_t \geq f_{p,t}. \end{cases} \quad (4.1)$$

também conhecido como “função de perda pinball” devido o seu formato.

Quanto menor $Q_{p,t}$, melhor a nossa estimativa. Observemos que quando $p = 0,5$, ou seja quando estamos avaliando a nossa mediana da distribuição, o $Q_{p,t}$ equivale à um erro absoluto.

Quando $p > 0,5$, ou seja, quando a probabilidade do dado ser menor que o esperado for maior que 0,5, há uma maior penalidade quando o dado for maior que o esperado, penalizando o erro cometido. O contrário acontece quando $p < 0,5$.

4.9.2 Pontuação de Winkler

Pode ser mais interessante avaliar um intervalo de previsão inteiro, ao invés de apenas os quantis.

Definição 4.3. Se $100(1 - \alpha)\%$ do intervalo de previsão no tempo t é dado por $[l_{\alpha,t}, u_{\alpha,t}]$ então a **Pontuação de Winkler** é definida utilizando o seguinte critério:

O comprimento do intervalo de previsão é uma penalidade, caso a observação esteja fora do intervalo. Caso esteja dentro do intervalo, é contado apenas o intervalo.

4.9.3 CRPS

O CRPS (Continuous Ranked Probability Score) significa Pontuação de Probabilidade Classificada Contínua. Supomos que em um determinado tempo t , o valor $y(t)$ assumiu o valor a nos dados apresentados. O modelo de previsão terá um intervalo de previsão, de modo que, cada valor terá uma certa probabilidade. A ideia é que a probabilidade de ser o valor a seja a maior possível, enquanto que para demais valores sejam as menores possíveis.

1. Colocamos a probabilidade contínua do dado que realmente ocorreu. Ou seja, é uma função que atribui o valor 0 para todos os valores antes de a , e valor 1 para todos os que aparecem à partir de a .
2. Fazemos o gráfico da probabilidade contínua do nosso modelo.
3. Calculamos a área entre as duas funções.

Quanto menor a área, o modelo estará trazendo maiores probabilidades próximas do que realmente ocorreu, ou seja, haverá um melhor desempenho em seu intervalo de confiança.

Definição 4.4. O CRPS, ou Pontuação de Probabilidade Classificada contínua, é calculada pela discrepância quadrática entre a Distribuição de Probabilidade Acumulada da previsão e a Distribuição de Probabilidade Acumulada do valor observado. (Matheson e Winkler 1976 ; Hersbach 2000)

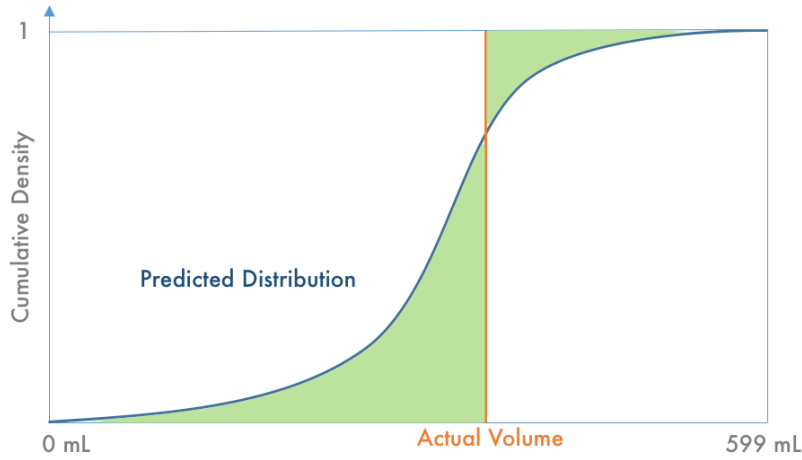


Figura 4.6: CRPS. Fonte: Feito por Booz Allen Hamilton no site Kaggle.

4.9.4 Skill Score

Assim como na avaliação da previsão pontual do modelo, podemos querer um método que possa avaliar desempenhos de diferentes escalas.

Com os Skill Scores, podemos calcular um desempenho que é livre de escala. Neste método, utilizamos o CRPS visto anteriormente do modelo a ser testado e o CRPS de algum método de referência (Benchmark), como por exemplo o método Naive.

Assim, o nosso Skill Score seria dado por:

$$\frac{CRPS_{Naive} - CRPS_{Modelo}}{CRPS_{Naive}}$$

Neste caso, quanto maior o nosso Skill Score, mais se distanciou do CRPS do método Naive, indicando um melhor desempenho que este método. Caso o Skill Score seja negativo, então o CRPS do modelo acaba sendo maior que o do método Naive, indicando que o modelo testado possui pior desempenho.

4.10 Validação cruzada para previsões

Uma versão considerada mais sofisticada para trabalhar com o conjunto de dados para testagem é utilizando a validação cruzada. É feito da seguinte forma:

1. Primeiramente, utilizamos os primeiros n dados do conjunto de testes para aplicar no modelo e realizar a previsão de um valor futuro que ainda esteja no conjunto de teste, mas que seja omitido. Calculamos o erro através da diferença entre o que foi previsto com o dado original.
2. Em seguida, fazemos o mesmo procedimento, mas considerando os $n + 1$ primeiros valores do conjunto de teste.
3. Repetimos o aumento até não haver mais valores conhecidos para utilizarmos.

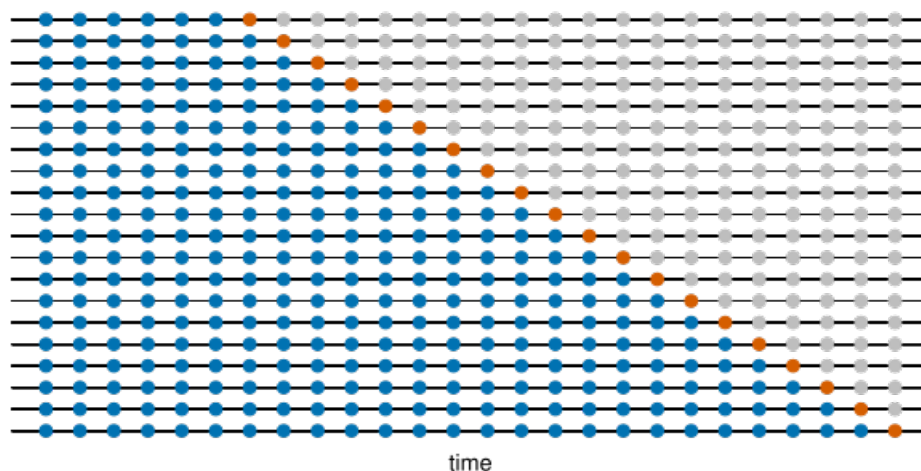


Figura 4.7: Exemplo de uma validação cruzada para prever o dado imediatamente futuro. Os valores em azul representam o que está sendo utilizado para o modelo e os valores em laranja representam os valores que o modelo irá prever. Cada linha é um procedimento, onde haverá um valor de erro. Fonte: Hyndman, R.J., & Athanasopoulos, G., Forecasting: principles and practice [1].

4. Calculemos o erro quadrático médio de todos os erros.

É interessante que a distância da previsão seja igual ao que se pretende utilizar no modelo, como por exemplo, se queremos utilizá-lo para fazer uma previsão para a próxima semana, é interessante que na validação cruzada também seja testado previsões para a próxima semana à partir do último dado utilizado no teste, desde que o mesmo seja conhecido. Na imagem à seguir, foi realizado uma validação cruzada para uma previsão de 4 passos à diante.

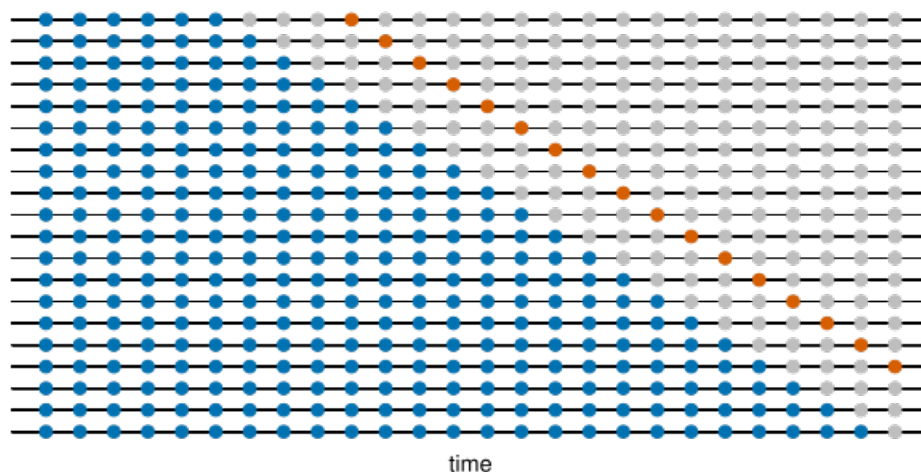


Figura 4.8: Validação cruzada para prever um valor após 4 períodos de tempo. Fonte: Hyndman, R.J., & Athanasopoulos, G., Forecasting: principles and practice [1].

Capítulo 5

Previsões de julgamento

Na prática, a previsão de julgamento é algo bem comum de se ocorrer, como por exemplo, se há uma completa falta de dados históricos, quando um novo evento ocorre, entre outros. O julgamento da previsão é baseada na experiência que o profissional tem do assunto, seja um médico, um analista financeiro ou qualquer outro especialista no assunto à ser abordado.

Pesquisas nesta área nos mostram que a precisão de uma previsão de julgamento melhora quando o previsor tem:

1. Bom domínio no conhecimento;
2. Informações atualizadas e oportunas.

Com o passar dos anos, se aumentou a aceitação desta previsão como ciência (Lawrence, M. et al, 2006), assim como o reconhecimento de sua necessidade. Sua qualidade melhorou muito com abordagens mais estruturadas e sistematizadas (Lawrence, M. et al, 2006).

Porém, é importante reconhecer que ela é subjetiva e possui suas limitações também. Podemos utilizar quando não há dados ou quando há dados também, neste ultimo caso servindo de apoio para ajustes nas previsões estatísticas feitas, combinando as duas técnicas. Afinal, quando há a presença de dados, a análise estatística sempre é a preferível (HYNDMAN e ATHANASOPOULOS, 2021).

5.1 Limitações

Existe inconsistência nas previsões, que é algo que não ocorre pelas fórmulas estatísticas e matemáticas. Afinal, um ser humano pode apresentar uma previsão totalmente diferente de acordo com seu ânimo, estado de espírito, e demais fatores que podem afetá-lo. Pode haver até mesmo manipulação da informação para algo que possa ser mais conveniente ao mesmo. Além disso, uma visão muito otimista pode extrapolar o horizonte de previsão que se tem do assunto também, acarretando em uma informação tendenciosa.

5.2 Princípios-chave

Uma abordagem bem estruturada pode contornar as limitações citados anteriormente.

1. Defina a tarefa de previsão de forma clara e concisa, evitando emoções e informações irrelevantes;
2. Implemente uma abordagem sistemática, como por exemplo, sobre fatores que deve levar em consideração para o problema;
3. Documente e justifique tudo;
4. Avalie sistematicamente a previsão;
5. Separe as previsões dos usuários/clientes para eliminar o otimismo;

À seguir, veremos um método sofisticado para utilizar previsões de julgamento que contempla os princípios-chave acima.

5.3 Método Delphi

O método Delphi foi desenvolvido por Olaf Helmer e Norman Dalkey da Rand Corporation na década de 1950 com o propósito de abordar um problema militar específico (HYNDMAN e ATHANASOPOULOS, 2021).

Este método se baseia no aspecto de que as previsões de um grupo são geralmente mais precisas do que as individuais. Para este método, é preciso que exista um facilitador, pessoa que passará as informações entre os especialistas e prosseguir com as etapas do método.

Seguem os passos do método:

1. Um painel de especialistas é montado e colocados todos em anonimato.
2. As tarefas/problemas são definidas e distribuídas aos especialistas.
3. Cada especialista informa a previsão de determinada tarefa/problema. Os resultados são compilados e resumidos pelo facilitador, gerando um *feedback*.
4. Os especialistas recebem o *feedback*, que agora revisam suas previsões podendo modificá-las ou não. Essa etapa poderá ser iterada até que um nível satisfatório de consenso seja alcançado.
5. As previsões finais são obtidas.

De acordo com HYNDMAN e ATHANASOPOULOS, podemos observar que cada uma das etapas tem os seus desafios e facilidades.

5.3.1 Especialistas e anonimato

Primeiramente, o facilitador deve entrar em contato com um grupo de especialistas que possam contribuir para a previsão. Uma sugestão de número, de acordo com HYNDMAN e ATHANASOPOULOS, é de algo entre 5 à 20 especialistas. Todos apresentam suas previsões e informam uma justificativa para suas decisões.

O anonimato é importante para o método Delphi pois elimina qualquer tipo de influência, como por exemplo, por pressão política. Além disso, todas as previsões são atribuídas com o mesmo peso, garantindo que todos contribuam de forma igual.

Por ser anonimato há a vantagem de não precisar de uma reunião com todos os especialistas para conversarem ao mesmo tempo e no mesmo lugar. Isso evita que exista alguma influência pela personalidade de algum dos membros, que pode ser mais influente que outros ou pelo ambiente, pois estudos indicam que o ambiente em grupo pode promover entusiasmo e otimismo, trazendo excesso de confiança e até mesmo o posicionamento das cadeiras pode influenciar no convencimento (BUEHLER et al., 2005).

5.3.2 Definir as tarefas/problemas

Antes de definir e informar a tarefa de previsão ao especialista, pode ser interessante ter uma rodada para coletar informações. Alternativamente, o facilitador também poderá coletar informações valiosas do especialista que não foi compartilhado aos demais, que poderão ser utilizadas no *feedback*.

5.3.3 Feedback

Aqui, o *feedback* deve conter um resumo estatístico das previsões e as justificativas para tais.

Como é algo controlado pelo facilitador, o mesmo pode fazer direcionamentos para que exista atenção em algo, seja uma previsão diferente ou para alguma justificativa.

5.3.4 Iteração

Ao reverem as previsões junto ao *feedback*, os especialistas podem fazer alterações. A ideia é repetir até haver um consenso entre os especialistas, mas isso não quer dizer que deve haver uma convergência completa para uma única previsão. A ideia é diminuir a variabilidade das respostas, para que torne a previsão mais precisa.

De acordo com HYNDMAN e ATHANASOPOULOS, duas ou três iterações já podem ser suficientes, pois se houver muitas, os especialistas podem desistir de prosseguir.

5.3.5 Previsões finais

Por fim, é a previsão final é obtida através das previsões dos especialistas, tomando igualmente os pesos para cada um. Porém, o facilitador deve estar atento aos *outliers* também, dando devido tratamento à eles.

5.3.6 Limitações e variações do método Delphi

O método Delphi pode ser algo bem demorado, do contrário de uma reunião em grupo, que pode ser resolvido e esclarecido por ali. Dependendo do que queremos, um método demorado pode não ser muito útil.

Uma variação ao método, é o chamado “estima-conversa-estima” onde os especialistas podem conversar e interagir entre cada iteração, mas a previsão se mantendo em anonimato. Podem haver esclarecimentos melhores com isso, assim como em uma reunião em grupo, porém também

5.3. MÉTODO DELPHI

há desvantagens, como por exemplo, a influência indevida de um especialista que é mais barulhento sobre suas visões.

Capítulo 6

Suavização exponencial

Proposto na década de 50, a suavização exponencial é um dos melhores métodos de previsão já utilizados através de uma ideia bem simples.

O método de suavização exponencial utiliza observações passadas, cujo os pesos decaem exponencialmente à medida que o tempo se distancie do momento atual. Assim, os dados mais recentes possuem um peso maior do que os mais antigos.

Para isso, existem diferentes formas de realizar essa ideia, veremos algumas delas.

6.1 Suavização exponencial simples

Este método é bem adequado quando não há um claro padrão de tendência ou de sazonalidade. Apesar de ser mais simples, este método serve de base para os métodos apresentados em sequência.

Pois bem, antes de partirmos para o método, lembremos sobre os métodos de referência Naive e o da Média. No método Naive, o peso é aplicado totalmente na última observação, já no método da Média, todos os valores utilizados apresentam o mesmo peso.

Geralmente, a suavização exponencial estará entre estes dois extremos. Podemos escrever a nossa previsão como:

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots$$

onde $0 \leq \alpha \leq 1$ é o parâmetro de suavização. Ou seja, a taxa na qual cada peso decai é dada pelo parâmetro α

Exemplo 6.1. *Consideremos, por exemplo, que $\alpha = 0,2$. Assim, os valores em sequência seriam 0, 16, 0, 128...*

Exemplo 6.2. *Observemos também que se $\alpha = 1$, então o método equivale ao método Naive.*

No geral, quanto menor o α , mais tolerante o método será com valores passados, em quanto que com maiores valores de α , os valores mais recente serão ainda mais priorizados.

6.1.1 Representação por média ponderada

A previsão para o valor no tempo $t + 1$ é igual à média móvel entre a observação mais recente e a previsão anterior:

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1 - \alpha)\hat{y}_t$$

sendo que \hat{y}_t é dado por $\alpha y_{t-1} + \alpha(1 - \alpha)\hat{y}_{t-1}$, e assim por diante.

Observemos que este processo precisa partir de algum ponto inicial, pois se não teríamos uma repetição infinita de valores. Tomemos $\hat{y}_1 = l_0$ como um valor inicial que precisa ser estimado. Então:

$$\begin{aligned}\hat{y}_2 &= \alpha y_1 + (1 - \alpha)l_0 \\ \hat{y}_3 &= \alpha y_2 + (1 - \alpha)\hat{y}_2 \\ &\dots \\ \hat{y}_{t+1} &= \alpha y_t + \alpha(1 - \alpha)\hat{y}_t.\end{aligned}$$

Substituindo cada equação na próxima, obtemos a equação inicial, sendo o último termo do somatório $(1 - \alpha)l_0$.

6.1.2 Representação por componentes

Uma outra representação bem útil é dada via componentes. No caso de uma suavização exponencial simples, a única componente incluída é o nível l_t . Métodos que apresentem tendência ou sazonalidade, incluem mais componentes.

Assim, neste caso temos uma equação de previsão e uma equação de suavização:

$$\begin{aligned}\text{Equação de previsão: } \hat{y}_{t+h} &= l_t \\ \text{Equação de nível: } l_t &= \alpha y_t + (1 - \alpha)l_{t-1}\end{aligned}$$

Observe que no caso da suavização exponencial simples a previsão é constante, ou seja, o valor de previsão é o mesmo para diferentes valores de h . Isso só será adequado se a série não tiver tendência nem sazonalidade.

6.1.3 Otimização

Vale ressaltar que precisaremos estimar os valores de α e l_0 para a suavização exponencial. Isso pode ser feito tanto de forma subjetiva, escolhendo valores com base na experiência, ou através de uma otimização, reduzindo a soma do quadrado dos resíduos nos dados de teste.

Diferentemente do caso da regressão linear, este envolve um problema de minimização não linear, dependendo de métodos mais sofisticados para isso.

6.2 Suavização exponencial com tendência

Holt (1957) estendeu a suavização exponencial simples para séries que possuem tendência. Este método envolve uma equação de previsão e duas de suavização, um para o nível e outro para a tendência. Deixaremos para colocar as componentes todas agrupadas em uma única imagem, junto com os demais métodos.

A ideia aqui é que conforme o valor h de \hat{y}_{t+h} aumente, a nossa previsão pode tanto aumentar ou diminuir, com base na tendência da série.

Assim, temos $\hat{y}_{t+h} = l_t + hb_t$, sendo b_t a estimativa da tendência (inclinação) da série.

Além disso, na componente da tendência também temos a presença do parâmetro de suavização da tendência β^* que estará presente mais adiante.

6.2.1 Suavização exponencial com tendência amortecida

Observemos que no método anterior temos uma tendência constante aplicada no modelo de suavização. Porém, para horizontes de previsão mais distantes, este método pode extrapolar nos resultados.

Diante disso, Gardner & McKenzie (1985) propõem a utilização de um parâmetro que “amortece” a tendência com o tempo, tornando-a uma linha reta horizontal. Por isso, o nome de tendência amortecida.

Assim, é introduzido um parâmetro ϕ de amortecimento que varia de 0 à 1, sendo que quando $\phi = 1$ temos a suavização com tendência sem amortecimento.

6.3 Suavização exponencial com sazonalidade

Holt (1957) e Winters (1960) estenderam o método de Holt para capturar a sazonalidade. O método sazonal de Holt-Winters compreende a equação de previsão e três equações de suavização. Existem duas variações para este método que diferem na natureza do componente sazonal.

O método aditivo é preferido quando as variações sazonais são aproximadamente constantes ao longo da série, enquanto o método multiplicativo é preferido quando as variações sazonais estão mudando proporcionalmente ao nível da série.

Com o método aditivo, o componente sazonal é expresso em termos absolutos na escala da série observada, e na equação de nível a série é ajustada sazonalmente subtraindo-se o componente sazonal. Dentro de cada ano, o componente sazonal somará aproximadamente zero.

Com o método multiplicativo, o componente sazonal é expresso em termos relativos (porcentagens), e a série é ajustada sazonalmente dividindo-se pelo componente sazonal. Dentro de cada ano, m

6.4 Combinação de tendência e sazonalidade

Colocaremos todos os métodos vistos até o momento com suas respectivas componentes numa única imagem.

Na imagem 6.1 teremos a combinação dos métodos com tendência, presente em cada linha, e com sazonalidade, presente em cada coluna.

As tendências estão divididas entre sem tendência (N), tendência (A) e tendência amortecida (A_d). Já a sazonalidade está dividida entre sem sazonalidade (N), sazonalidade aditiva (A) e sazonalidade multiplicativa (M).

6.5 Erros de previsão

Vimos como descrever os dados observados com diferentes componentes, acrescentando a tendência e a sazonalidade. Os últimos modelos apresentados servem para realizar previsões

6.5. ERROS DE PREVISÃO

Trend	Seasonal		
	N	A	M
N	$\hat{y}_{t+h t} = \ell_t$	$\hat{y}_{t+h t} = \ell_t + s_{t+h-m(k+1)}$	$\hat{y}_{t+h t} = \ell_t s_{t+h-m(k+1)}$
	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}$
		$s_t = \gamma(y_t - \ell_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(y_t/\ell_{t-1}) + (1 - \gamma)s_{t-m}$
A	$\hat{y}_{t+h t} = \ell_t + hb_t$	$\hat{y}_{t+h t} = \ell_t + hb_t + s_{t+h-m(k+1)}$	$\hat{y}_{t+h t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$
	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$
		$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1 - \gamma)s_{t-m}$
A _d	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t$	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$	$\hat{y}_{t+h t} = (\ell_t + \phi_h b_t)s_{t+h-m(k+1)}$
	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$
		$s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1 - \gamma)s_{t-m}$

Figura 6.1: Combinação entre os métodos que envolvem tendência e os que envolvem sazonalidade

pontuais dos dados, mas podemos utiliza-los para realizar intervalos de previsão.

Para isso, precisamos conhecer o termo de erro da nossa previsão.

Para cada método existem dois modelos: um com erros aditivos e outro com erros multiplicativos. As previsões pontuais produzidas pelos modelos são idênticas se usarem os mesmos valores dos parâmetros de suavização. Eles irão, no entanto, gerar diferentes intervalos de previsão.

Para distinguir entre um modelo com erros aditivos e um com erros multiplicativos (e também para distinguir os modelos dos métodos), adicionamos uma três letras na classificação de um modelo de suavização exponencial.

Rotulamos cada modelo de espaço de estado como ETS (Erro, Tendência, Sazonalidade). As possibilidades para cada componente são: Erro = $\{A, M\}$, Tendência = $\{N, A, A_d\}$ e Sazonalidade = $\{N, A, M\}$.

Para a modelagem, assumiremos que os erros utilizados nos modelos são ruído branco normalmente distribuídos com média 0 e variância σ^2 . O mesmo ocorre para tanto para os modelos com erros aditivos e multiplicativos, só que no segundo caso, queremos o erro relativo, ou seja, dividimos pela previsão obtida.

Vale ressaltar que estes erros podem diferir do simples resíduo entre valor observado e valor estimado, dependendo do tipo de modelo.

Temos então uma tabela (Figura 6.2) que apresenta todos os modelos possíveis, tanto com erro aditivo, quando com erro multiplicativo.

6.5. ERROS DE PREVISÃO

ADDITIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$
A _d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1} (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} (1 + \alpha \varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha (\ell_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + s_{t-m}) \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} (1 + \alpha \varepsilon_t)$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1}) (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1}) \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1}) \varepsilon_t$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$
A _d	$y_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1}) \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1}) \varepsilon_t$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$

Figura 6.2: Tabela com todos os modelos ETS

Capítulo 7

Previsão de séries temporais hierárquicas e séries temporais agrupadas

As séries temporais podem ser naturalmente desagregadas por vários atributos de interesse.

Por exemplo, podemos desagregar uma série temporal de vendas de bicicletas em seus diferentes tipos, como vendas de bicicletas de estrada, de montanha, infantis, etc.

Além disso, cada um pode ser desagregado em mais categorias, como por exemplo, dentre as bicicletas de estrada, podemos dividir entre as cores quentes e cores frias.

Assim, quando temos este tipo de desagregação com uma certa hierarquia, temos uma agregação hierárquica. Assim, uma série temporal que pode ser desagregada dessa forma é chamada de **série temporal hierárquica**.

Outro exemplo bem comum de desagregação é via divisões geográficas, como região, estado, cidade ou bairro.

Observemos que uma série pode ser desagregada em dois tipos de desagregação também. Por exemplo, no caso das vendas de bicicletas, que além de poderem ser desagregadas pelos tipos, também podem ser desagregadas pela região, garantindo duas hierarquias diferentes.

Quando temos uma múltipla hierarquia, chamamos a série temporal de **série temporal agrupada**.

O ponto chave deste capítulo está na ideia da coerência da desagregação. É comum produzir previsões desagregadas com base nas séries temporais desagregadas e exigir que a soma retorne a previsão agregada das séries.

Quando isto ocorrer, diremos que temos uma **previsão coerente** da nossa série temporal.

7.1 Diferença entre série temporal hierárquica e agrupada

Ao introduzir o capítulo, já introduzimos a ideia de série temporal hierárquica e série temporal agrupada, bem como suas diferenças.

Nessa seção, lançaremos uma ideia mais formal da diferença entre cada uma delas. Começando com a formalização da série temporal hierárquica.

A imagem 7.1 representa uma hierarquia de nível 2. No topo da hierarquia (que chamamos de nível 0) está o “Total”, o nível mais agregado dos dados. Este total é desagregado em duas séries

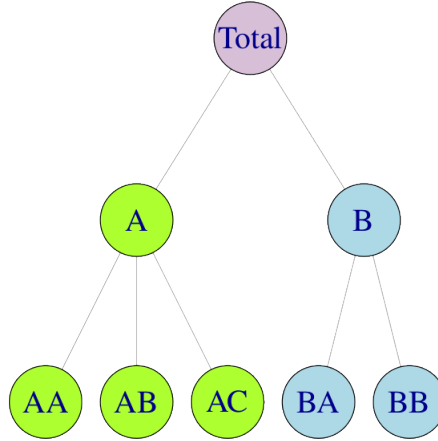


Figura 7.1: Exemplo de estrutura hierárquica

no nível 1, que por sua vez são divididas em três e duas séries respectivamente no nível inferior da hierarquia.

Assim, para qualquer momento t da série temporal, temos:

$$y_t = y_{A,t} + y_{B,t} = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BB,t}$$

Podemos representar por uma notação matricial:

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix}$$

ou então,

$$Y_t = S b_t$$

em que Y_t é a matriz de todas as séries da hierarquia, S é a matriz de soma e b_t é a matriz das séries do nível inferior.

Já a estrutura de série temporal agrupada oferece um caso mais generalizado.

Podemos ter, por exemplo, duas hierarquias de nível 2 como no caso da imagem 7.2. Neste caso, podemos desagregar a série tanto em A e B , como em X e Y . De todo modo, observemos que o nível inferior se mantém.

Na forma matricial, podemos escrever da seguinte forma:

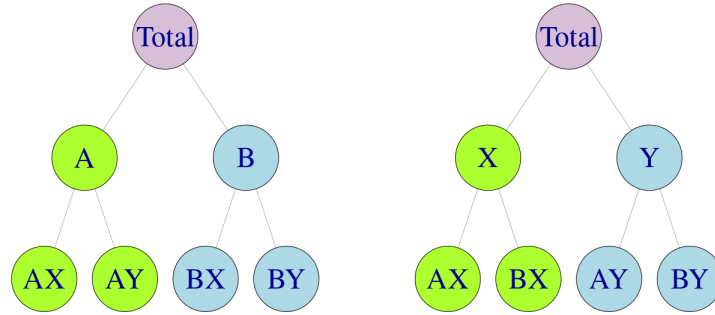


Figura 7.2: Exemplo de estrutura agrupada

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{bmatrix}$$

ou então, podemos novamente reescrever por:

$$Y_t = S b_t$$

assim como na série hierárquica.

7.2 Abordagem de baixo para cima

Este é um dos métodos mais simples de se obter previsões coerentes.

O método consiste em gerar as previsões para cada série do nível inferior e depois realizar as devidas somas para obter as séries dos níveis acima. Ou seja,

$$\tilde{y}_h = S \hat{b}_h.$$

7.3 Abordagem de cima para baixo

Segundo HYNDMAN e ATHANA-SOPOULOS, esta abordagem só funciona para as estruturas estritamente hierárquicas, não para os agrupados.

Como o nome sugere, este método consiste em primeiro realizar as previsões da série y_t e depois desagregá-los na hierarquia através de uma proporção.

Sejam p_1, \dots, p_m um conjunto de proporções da desagregação.

Assim, cada série do nível inferior pode ser representada pela série total multiplicada pela devida proporção.

Em notação matricial, temos:

$$\tilde{b}_t = p\hat{y}_t$$

Assim, obtemos o nível inferior da nossa hierarquia. Depois, estas previsões podem ser agregadas para obtermos por fim, as previsões coerentes.

Assim, podemos escrever de modo matricial:

$$\tilde{y}_h = Sp\hat{y}_t$$

As proporções de desagregação costumam ser dadas com base nas proporções históricas dos dados.

Como obtê-los?

Há varias formas de se fazer na literatura, desde mais simples até mais complexas. Uma das formas é obtendo a média da proporção entre cada nível inferior com o total.

7.4 Abordagem intermediária

Esta abordagem combina as duas abordagens anteriores.

Primeiramente um “nível médio” é escolhido. Acima deste nível é feito a abordagem de baixo para cima, e abaixo deste nível é feito a abordagem de cima para baixo.

Ou seja, utilizaremos as previsões deste nível médio para obter as proporções do nível inferior, que serão agregados até alcançar este nível médio e dali para cima basta continuar agregando.

7.5 Representação matricial

Todos os métodos apresentados podem ser expressos por uma notação comum.

Definição 7.1. *Seja \hat{y}_h o vetor das previsões de cada série da hierarquia sem uso da agregação. Chamaremos os valores de \hat{y}_h como **previsões de base**.*

Todas as abordagens de previsão para estruturas hierárquicas ou agrupadas podem ser representadas como:

$$\tilde{y}_h = SG\hat{y}_h$$

sendo G a matriz que leva as previsões de base até o nível inferior e S a matriz de soma que faz a agregação para obter as previsões coerentes \tilde{y}_h .

Enquanto que a matriz S é obtida de acordo com a estrutura da hierarquia do nosso problema, a matriz G é definida de acordo com a abordagem que será utilizada.

Por exemplo, numa abordagem de baixo para cima, temos algo parecido como:

$$G = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

que neste caso, representa a matriz G do exemplo dado na abordagem de baixo para cima, pois temos exatamente 8 séries ao todo e 5 séries inferiores.

Já na abordagem de cima para baixo, teríamos algo como:

$$G = \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

em que cada p representa uma proporção que ao ser multiplicado pela série total, obtém-se as séries inferiores.

E na abordagem intermediária, a matriz seria uma combinação das duas de cima.

7.6 Reconciliação ideal

Definição 7.2. *Seja $P = SG$. Chamaremos P de **matriz de reconciliação**.*

Assim, podemos escrever $\tilde{y}_h = P\hat{y}_h$.

Nos métodos discutidos anteriormente, **nenhuma** reconciliação real foi feita porque os métodos foram baseados em previsões de um único nível de estrutura.

Podemos então procurar a melhor matriz G para fornecer previsões reconciliadas mais coerentes. Ou seja, a matriz G que minimize os erros das previsões coerentes.

Diferentemente de outras aproximações, a previsão por reconciliação ótima é gerada usando toda informação disponível da estrutura hierárquica ou agrupada.

Referências Bibliográficas

- [1] HYNDMAN, R.J., ATHANASOPOULOS, G., *Forecasting: principles and practice*, 3^a edição, OTexts: Melbourne, Australia. Disponível em <https://otexts.com/fpp3/>. Acesso em: 26 de outubro de 2021.
- [2] GUJARATI, D., *Econometria Básica*. Rio de Janeiro: Campus Elsevier, 2011.
- [3] KEYNES, John Maynard, *The general theory of employment, interest and money*. Nova York: Harcourt Brace Jovanovich, 1936
- [4] FRIEDMAN, Milton, *Essays in positive economics*. Chicago: University of Chicago Press, 1953.
- [5] GOLDBERGER, Arthur S. *Econometric theory*. Nova York: John Wiley & Sons, 1964.
- [6] BAUER, A., *Automated Hybrid Time Series Forecasting: Design, Benchmarking, and Use Cases*. Fakultät für Mathematik und Informatik, 2020.
- [7] LAWRENCE, M., GOODWIN, P., O'CONNOR, M., & ONKAL, D., *Judgmental forecasting: A review of progress over the last 25 years*. International Journal of Forecasting, 2006.
- [8] BUEHLER, R., MESSERVEY, D., & GRIFFIN, D., *Collaborative planning and prediction: Does group discussion affect optimistic biases in time estimation? Organizational Behavior and Human Decision Processes*, 2005.
- [9] HOLT, CE., *Forecasting seasonals and trends by exponentially weighted averages* (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA, 1957.
- [10] GARDNER, E. S., & MCKENZIE, E., *Forecasting trends in time series*. Management Science, 1237–1246, 1985.
- [11] WINTERS, P. R., *Forecasting sales by exponentially weighted moving averages*. Management Science, 324–342, 1960.