2
FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os modelos de séries temporais chamados ‘estruturais’, nos quais o valor das observações é visto como composto de uma parte sistemática, modelada por uma equação que é função apenas do período, e de sua parte aleatória, e ~ N(0,σ^2), que se considera independente e identicamente distribuída nos diversos períodos. Em seguida, apresenta os métodos de previsão que se baseiam nestes modelos e as expressões usuais para estimação dos parâmetros (variância e média) do erro de previsão. Descreve, também, os gráficos de controle CUSUM e EWMA, que serão usados nesta dissertação para detecção do viés de previsão.

2.1
Modelos Estruturais de Séries temporais e Métodos de Previsão Correspondentes

2.1.1
Modelo Constante

É o modelo em que não se observa tendência de aumento ou declínio na demanda, a qual varia em torno de um nível. Assim, a cada período t, o valor observado da demanda, $\mathbf{X_t}$, é dado por:

$$\mathbf{X_t = a + e_t} \tag{2.1}$$

Onde $\mathbf{a}$ é o nível da demanda e $\mathbf{e_t}$ é a flutuação aleatória, ou ‘ruído’, que se supõe independente e identicamente distribuída, ~ N(0,σ^2).

Há dois métodos principais de previsão que se baseiam no modelo constante: o de média móvel e o de amortecimento exponencial simples. Estes serão vistos a seguir.

2.1.2
Método de Média Móvel

Este método considera que a média tem certa flutuação e, portanto, os dados antigos não são tão confiáveis quanto os mais recentes. Vem então a ideia de estimar o nível, a, usando apenas as últimas **k** observações (onde k, componente da média móvel, é um parâmetro a escolha do usuário).

A média móvel começa a partir do momento em que haja dados suficientes para calculá-la. Não há 'estimativa inicial'. As previsões seguintes serão obtidas eliminando-se o valor mais antigo e acrescentando-se o valor mais novo da demanda e novamente fazendo-se a média.

Um bom exemplo é mostrado na Tabela 2. Neste exemplo, k=10.

Tabela 2. Simulação da Previsão da Demanda com Média Móvel de 10 semanas.

Semana	Demanda	Previsão	Semana	Demanda	Previsão
1	20		22	24	22.8
2	19		23	18	22.8
3	18		24	19	22.8
4	23		25	26	22.8
5	19		26	24	23.0
6	22		27	26	23.2
7	25		28	21	23.4
8	22		29	30	23.4
9	24		30	20	24.1
10	23		31	24	23.6
11	22	21.5	32	14	23.2
12	24	21.7	33	25	22.2
13	18	22.2	34	25	22.9
14	19	22.2	35	26	23.5
15	24	21.8	36	16	23.5
16	22	22.3	37	23	22.7
17	24	22.3	38	23	22.4
18	21	22.2	39	26	22.6
19	23	22.1	40	24	22.2
20	25	22.0	41		22.6
21	28	22.2			

Fonte: Adaptado de Silver *et al.* (1998).

Para encontrar $\hat{a}_0$ basta somar as 10 primeiras demandas (k=10) e fazer a média aritmética: $\hat{a}_0$ **= (20+19+18+23+19+22+25+22+24+23) / 10 = 21,5**

Para a previsão de demanda da semana 12, basta eliminar a 1ª demanda (20) e incluir a demanda da semana 11 (22) e fazer o cálculo novamente.

$\hat{a}_1$ **= (19+18+23+19+22+25+22+24+23+22) / 10 = 21,7**

E assim sucessivamente, até chegar à semana 41, que é a última deste exemplo.

Podemos inicializar a previsão com a seguinte fórmula:

$$\hat{a}_t = \frac{\sum_{i=(t-k+1)}^{t} X_i}{k} \qquad \textbf{(2.2)}$$

Onde $\hat{a}_t$ é a previsão para o instante t+1 e **k** é o número de observações na média móvel da demanda, X_k.

2.1.3
Método de Amortecimento Exponencial Simples

Quando se usa o modelo de média móvel, dá-se o mesmo peso às últimas k observações e se desconsidera completamente as anteriores.

No amortecimento exponencial simples, isso não ocorre. A cada observação é dado um peso, que decresce, em progressão geométrica, à medida que a observação vai ficando mais antiga.

A fórmula para a estimativa do nível constante no instante t é, então:

$$\hat{a}_t = \alpha X_t + (1 - \alpha)\hat{a}_{t-1} \qquad \textbf{(2.3)}$$

Onde α é a constante de amortecimento definida no intervalo (0, 1), X_t é a demanda do período t e $\hat{a}_t$ a estimativa da média da demanda no instante t, que, neste modelo constante, constitui a previsão pontual da demanda para o instante t+1. Assim:

$$\hat{X}_{t+1} = \hat{a}_t \qquad \textbf{(2.4)}$$

O valor ideal para a constante de amortecimento depende de dois fatores: por um lado, valores menores para essa constante filtram melhor o 'ruído' aleatório (correspondem a utilizar mais informação); valores maiores dão peso muito maior às últimas observações de modo que o peso das mais antigas se dissipa mais rapidamente; em contrapartida, valores maiores fazem a estimativa responder mais rapidamente a eventuais mudanças reais no nível da série. Em geral, recomenda-se usar valores de α entre 0,03 e 0,3 (SILVER *ET AL.*, 1998). Uma maneira de escolher o valor 'ideal' de α para uma dada série é buscar o valor que minimize o erro quadrático médio (EQM), que vem a ser uma estimativa da variância do erro de previsão (mais detalhes sobre o EQM na Seção 2.2.1 deste Capítulo). Para isso, partindo-se de uma série histórica, simula-se a aplicação do método de previsão experimentando diferentes valores para α até identificar o que minimize o EQM. Pode-se, por exemplo, usar o "*Solver*" do Excel que minimiza o EQM automaticamente, dado um intervalo para α. Cabe-se registrar que estes procedimentos de escolha de α podem não ser viáveis quando se lida com um número elevado de itens de demanda (muitas séries). Como sugestão de conduta, deve-se, nesses casos, utilizar este procedimento para um grupo reduzido de itens de maior importância e arbitrar um valor, por exemplo, α = 0,10, para os itens de menor importância. Para mais considerações sobre a escolha de α, ver Silver *et al.* (1998, p.106).

Para inicializar o processo é necessária uma estimativa inicial do nível, $\hat{\mathbf{a}}_0$. Esta, usualmente, é simplesmente a média aritmética das observações de uma série histórica.

2.1.4
Modelo Linear

É o modelo em que há uma tendência de aumento ou declínio da demanda. Assim, no instante t, a demanda é modelada por:

$$\mathbf{X_t = a + bt + e_t} \qquad \textbf{(2.5)}$$

Onde **a** é o nível inicial (em t=0) da demanda, **b** é a taxa de aumento ou declínio da demanda e $\mathbf{e_t}$ é a flutuação aleatória (ou 'ruído'), que se supõe independente e identicamente distribuída, ~N(0,σ^2).

Para a previsão usando este modelo, o método mais adequado é o amortecimento exponencial duplo ou método de *Holt*, visto a seguir.

2.1.5
Método de Amortecimento Exponencial Duplo ou de *Holt*

O método de *Holt* estima **a** e **b** a cada instante t por:

$$\hat{a}_t = \alpha X_t + (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}) \qquad \textbf{(2.6)}$$

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1-\beta)\hat{b}_{t-1} \qquad \textbf{(2.7)}$$

Onde α e β são constantes de amortecimento definidas no intervalo (0, 1) e X_t é a demanda do período t.

Este procedimento redefine a origem como estando sempre no instante atual, de tal maneira que a equação 2.5 é substituída por:

$$X_t = a_t + e_t \qquad \textbf{(2.8)}$$

Onde:

$$a_t = a_{t-1} + b_{t-1} \qquad \textbf{(2.9)}$$

Desta forma, a previsão pontual um passo a frente, feita no instante t+1, é obtida por:

$$\hat{X}_{t+1} = \hat{a}_t + \hat{b}_t \qquad \textbf{(2.10)}$$

Para inicializar o processo de previsão, isto é, para obter estimativas iniciais de a e de b, pode-se utilizar a **Regressão Linear por Mínimos Quadrados**, numa série histórica de dados. Tendo-se estimado por regressão linear um termo independente $\hat{a}$ (nível na origem, um período antes do início da série histórica de tamanho n) e uma inclinação $\hat{b}$**,** então, redefinindo a origem dos tempos (t=0) para o último período da série histórica, de modo que t=1 seja o primeiro período para o qual se faz previsão, então, os valores iniciais de nível e tendência para a previsão são dados por:

$$\hat{b}_0 = \hat{b} \qquad \textbf{(2.11)}$$

$$\hat{a}_0 = \hat{a} + \hat{b}n \tag{2.12}$$

Onde, repita-se, n é o número de períodos da série histórica.

Em aplicações práticas, recomenda-se usar valores de α entre 0,03 e 0,3 e valores de β entre 0,005 e 0,18 (SILVER *ET AL.*, 1998). Os valores mais recomendados são α=0,10 e β=0,05. Pode-se, também, escolher valores de α e β que minimizem o erro quadrático médio (EQM), de modo similar ao observado para o caso do modelo constante. Para mais considerações sobre a escolha de α e β, ver Silver *et al.* (1998).

2.2
Erro de Previsão da Demanda

De modo geral, quaisquer que sejam o modelo de série temporal suposto e o método de previsão empregado, previsões pontuais correspondem a estimativas do componente sistemático, filtrado o ruído aleatório. Em outras palavras, uma previsão pontual é a previsão, para um momento futuro, da média (valor esperado) da variável X (no caso que nos motiva, da demanda). O componente aleatório se manifesta na forma de um erro de previsão, que é a diferença entre o valor previsto e o observado. Os erros de previsão possuem informações valiosas e devem ser analisados cuidadosamente. Como já observamos, sua dispersão permite estabelecer previsões intervalares e seu valor esperado, se diferente de zero, indica inadequação do modelo e possibilidade de haver ocorrido alguma mudança no mecanismo gerador da série.

Em outras palavras: enquanto os erros observados estiverem dentro de uma faixa razoável — coerente com um modelo estatístico dos erros — as empresas podem continuar adotando seu modelo atual de previsão. Caso se observe que um erro está muito além da faixa 'razoável', essa descoberta pode indicar que o modelo de previsão adotado deixou de ser adequado. Se as previsões costumam consistentemente superestimar ou subestimar a demanda, isso também pode ser um sinal de que a empresa deve alterar seu modelo de previsão ou os parâmetros utilizados para avaliar o erro.

O erro de previsão no período t é:

$$\varepsilon_t = \hat{X}_{t-1} - X_t \quad (2.13)$$

Onde $\hat{X}_{t-1}$ é a previsão realizada no período t-1 para o período t, e X_t é a demanda do período t.

2.2.1
Medidas de Dispersão do Erro de Previsão de Demanda

Uma medida da variabilidade do erro da previsão é o erro quadrático médio (EQM), que vem a ser o estimador da variância do erro de previsão, dado pela fórmula:

$$EQM_0 = \frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t)^2 \quad (2.14)$$

A equação acima é um estimador para um número n de observações de ε_t. Porém, quando o número de observações aumenta, isto é, periodicamente obtém-se um novo valor para ε_t, é interessante atualizar a estimativa dinamicamente. Isso pode ser feito pela fórmula recursiva:

$$EQM_n = (n-1)EQM_{n-1} + \varepsilon_n^2 \quad (2.15)$$

Outra opção para atualizar recursivamente a estimativa do EQM é utilizar a sua média móvel ponderada exponencialmente, da seguinte maneira:

$$EQM_t = \omega(\varepsilon_t)^2 + (1-\omega)EQM_{t-1} \quad (2.16)$$

Onde ω é uma constante de amortecimento e $\mathbf{EQM_{t-1}}$ é a estimativa do EQM no período t-1. O valor inicial, $\mathbf{EMQ_0}$, pode ser obtido dos dados históricos: será a variância amostral dos mesmos, no caso do modelo constante, ou a variância dos resíduos da regressão linear, no caso do modelo crescente.

Recomenda-se utilizar valores de **ω** entre **0,01 e 0,1** (SILVER *ET AL.*, 1998).

Outras medidas de dispersão do erro de previsão são:

O desvio absoluto no período t, A_t, dado por:

$$\mathbf{A_t = |\varepsilon_t|} \tag{2.17}$$

O desvio absoluto médio, DAM, como sendo:

$$\mathbf{DAM_n = \frac{1}{n}\sum_{t=1}^{n} A_t} \tag{2.18}$$

O DAM pode ser usado para estimar o desvio-padrão do componente aleatório. Se este componente for distribuído normalmente, com média zero e desvio-padrão σ, então, observa-se a seguinte relação entre DAM e $|\varepsilon_t|$:

$$\boldsymbol{\sigma = 1,25\, E(|\varepsilon_t|)} \tag{2.19}$$

Portanto, nesse caso, pode-se estimar o desvio-padrão do erro de previsão por:

$$\boldsymbol{\hat{\sigma} = 1,25\, \mathrm{DAM}} \tag{2.20}$$

Ou ainda:

$$\boldsymbol{\hat{\sigma} = \sqrt{\mathrm{EQM}}} \tag{2.21}$$

2.2.2
Viés da Previsão de Demanda

Há indicação de viés quando, em média, as previsões efetuadas estão substancialmente acima ou abaixo da demanda real. Quando existe indicação de viés, tem-se evidências de que o modelo é incorreto ou que os parâmetros do modelo não estão sendo estimados adequadamente.

Pode-se utilizar a soma dos erros de previsão para avaliar o viés da previsão, tal como:

$$B_n = \sum_{t=1}^{n} (\varepsilon_t) \quad \textbf{(2.22)}$$

O somatório dos erros de previsão (B), para n observações, oscilará em torno de zero se o erro for verdadeiramente aleatório e não enviesado.

Ao se lançar todos os erros da previsão num gráfico, pode-se visualizar quando as previsões efetuadas ficam acima ou abaixo da demanda real. O gráfico de Soma Cumulativa (SC) do erro de previsão é mais adequado para indicar viés do que o gráfico simples do erro e é definido, recursivamente, da seguinte forma:

$$SC_t = SC_{t-1} + \varepsilon_t \quad \textbf{(2.23)}$$

Sendo o erro acumulado inicial:

$$SC_0 = 0 \quad \textbf{(2.24)}$$

Contudo, ao se utilizar o gráfico da soma cumulativa, a decisão é baseada em inspeção visual subjetiva, sem um critério objetivo, como um limiar (valor crítico) em um teste estatístico.

Como critério objetivo para identificação de viés, só se consegue encontrar nos textos sobre previsão de demanda (ver, por exemplo, Chopra e Meindl, 2004) o cálculo, a cada instante, da razão de viés TS_t dado por:

$$TS_t = \frac{B_t}{DAM_t} \tag{2.25}$$

Considera-se que há viés se TS_t estiver fora do intervalo (-6, 6). Se $TS_t < -6$, a previsão subestima a demanda, ou se $TS_t > 6$, a previsão superestima a demanda. Nesses casos, deve-se escolher um novo modelo de previsão.

2.3
Gráficos de Controle para Análise do Erro de Previsão de Demanda

Buscando na literatura um critério objetivo para análise do erro de previsão da demanda, o mais usual que se encontra é o **TS**, citado acima (ver equação 2.25).

Esse critério é inadequado, pois TS_t é uma soma de t variáveis aleatórias, sendo que t cresce com o tempo e portanto, mesmo se não houver viés, a variância do TS_t é proporcional a t. Assim, a probabilidade de que TS_t saia do intervalo (-6, 6) cresce continuamente.

Dessa forma, para possibilitar uma maior eficácia na análise do erro de previsão da demanda, este trabalho propõe outros métodos de análise do erro da previsão: **os gráficos de controle CUSUM e EWMA**.

Esses gráficos foram desenvolvidos no contexto de CEP – Controle Estatístico de Processos – para detectar pequenas alterações na média de variáveis que representam características de qualidade de processos. São gráficos mais eficientes para sinalizar alterações na média de uma variável (possuem maior sensibilidade) do que os gráficos de *Shewhart*, quando tais alterações são pequenas.

O problema de detecção de viés é análogo ao de CEP, pois, trata-se de uma alteração na média da variável ‘erro de previsão’. Se não houver viés, o valor esperado do erro é zero; caso contrário, é diferente de zero.

A proposta do presente trabalho é testar experimentos utilizando gráficos de controle **CUSUM** e **EWMA** e compará-los com o método do **TS**. Será verificado, também, quais deles são mais eficientes para cada caso experimentado, isto é, quais **minimizam o risco α** (rejeitar o viés igual a zero, quando ele é zero, ou seja, achar que o modelo de previsão é inadequado quando ele é adequado) **e o risco β** (aceitar viés igual a zero, quando ele é diferente de zero, ou seja, achar que o modelo é adequado quando, na verdade, ele não é adequado).

As duas Seções seguintes descrevem concisamente os gráficos de CUSUM e EWMA. Para mais detalhes, pode-se consultar Costa, Epprecht e Carpinetti (2005) ou Montgomery (2004).

2.3.1
Gráfico de CUSUM

O gráfico ou algoritmo de CUSUM utiliza duas estatísticas em paralelo: uma sensível a aumentos na média da variável que está sendo monitorada (por exemplo, a média de uma variável de controle) e outra, sensível a diminuições na média dessa variável:

$$\mathbf{S_t^+ = Max[0, Z_t - d + S_{t-1}^+]} \qquad \textbf{(2.26)}$$

$$\mathbf{S_t^- = Max[0, -d - Z_t + S_{t-1}^-]} \qquad \textbf{(2.27)}$$

Onde **Z** é o valor padronizado da variável e os valores iniciais para $\mathbf{S_t^+}$ e $\mathbf{S_t^-}$ são: $\mathbf{S_0^+}$ = $\mathbf{S_0^-}$ = **0**.

O algoritmo limita $\mathbf{S_t^+}$ e $\mathbf{S_t^-}$ a valores positivos ou nulos, zerando $\mathbf{S_t^+}$ sempre que $\mathbf{Z_t - d + S_{t-1}^+ < 0}$ e zerando $\mathbf{S_t^-}$ sempre que $\mathbf{-d - Z_t + S_{t-1}^- < 0}$.

Recomenda-se usar o valor de referência, **d**, igual a:

$$\mathbf{d} = \delta/2 \tag{2.28}$$

Onde $\boldsymbol{\delta}$ é o módulo da magnitude do deslocamento da média de Z que é relevante detectar com rapidez.

O algoritmo de CUSUM produz um sinal sempre que S_t^+ ou S_t^- > $\mathbf{K_C}$ (limite de controle ou intervalo de decisão para o gráfico de CUSUM).

Segundo Costa *et al.*, 2005, como não existe uma expressão simples que relacione o risco α com o parâmetro K_C para um dado valor de d, uma recomendação geral é adotar:

$$\mathbf{K_C = 5} \tag{2.29}$$

Esta é a versão do algoritmo CUSUM para variáveis padronizadas (com média nula e desvio-padrão igual a um). Caso se trabalhe com variáveis não padronizadas, as fórmulas são alteradas.

2.3.2
Gráfico de EWMA

O gráfico de controle da Média Móvel Ponderada Exponencialmente (ou EWMA, do inglês *Exponentially Weighted Moving Average*) é outro gráfico para detectar pequenos deslocamentos na média do processo. Apresenta desempenho similar ao do CUSUM e é geralmente utilizado com observações individuais (COSTA *ET AL.*, 2005). A estatística EWMA é calculada recursivamente a partir de uma variável Z, por:

$$\mathbf{Y_t = \lambda Z_t + (1-\lambda) Y_{t-1}} \tag{2.30}$$

Onde λ é uma constante de amortecimento, definida no intervalo (0, 1) e $Y_0 = \mu_0$ (média em controle de Z), no caso, zero.

A variância da variável Y_t é dada por:

$$\sigma_{yt}^2 = \sigma^2(\lambda/2-\lambda)[1-(1-\lambda)^{2t}] \qquad \textbf{(2.31)}$$

Onde σ^2 é a variância da variável Z.

Os limites de controle são:

$$LSC = K_E\sigma_0\sqrt{(\lambda/2-\lambda)[1-(1-\lambda)^{2t}]} \qquad \textbf{(2.32)}$$

$$LIC = -K_E\sigma_0\sqrt{(\lambda/2-\lambda)[1-(1-\lambda)^{2t}]} \qquad \textbf{(2.33)}$$

Onde K_E é o fator de abertura dos limites de controle do gráfico de EWMA e σ_0 é o desvio-padrão da variável Z quando o processo está em controle. O termo $[1-(1-\lambda)^{2t}]$ converge rapidamente para 1 quando t cresce.