5.1. Amortecimento exponencial simples (AES)

5.1.0. Introdução

No modelo constante, supomos que o nível da série não varia ao longo do tempo, ou varia muito lentamente e pode ser (pelo menos localmente) modelado por uma constante,

$$\mu_t = a$$

A série pode ser descrita como:

$$Z_t = a + \varepsilon_t$$
$$E(Z_t) = a$$

A previsão, para qualquer horizonte de previsão *k*, é dada pela estimativa mais recente do nível *a*:

$$\hat{Z}_{T+k} = E(Z_{T+k} \mid \mathbf{Z}_T) = E(a_1 + \varepsilon_{T+k} \mid \mathbf{Z}_T) = \hat{a}_T$$
$$\hat{Z}_{T+k} = \hat{a}_T$$

O problema portanto se resume em obter a estimativa $\hat{a}_T$. No método da *média global*, esta estimativa era obtida a partir da média aritmética simples de toda a série passada, com todos os valores recebendo o mesmo peso. Como foi argumentado, este método não é muito usado, pois em geral parece mais sensato dar um peso maior às observações mais recentes. Já que os modelos não são *globais*, mas apenas *locais* (isto é, válidos apenas num trecho da série que contém as observações mais recentes), é razoável supor que os valores observados recentemente tenham sido gerados por um mesmo modelo, mas não é razoável supor que este modelo seja válido também para todo o passado da série.

Este raciocínio levou a criação dos métodos baseados em *médias móveis*, que descartam as observações mais antigas e consideram na média apenas um subconjunto com as *n* observações mais recentes. Estes métodos contudo ainda não são muito convincentes, pois as observações contidas numa janela móvel recebem todas o mesmo peso, enquanto as observações fora da janela são ignoradas. Se usamos MM(10), por exemplo, os valores Z_t e Z_{t-10} serão incluídos na média com o mesmo peso, mas Z_{t-11} será descartado. Isto não parece muito lógico; por que um valor tão defasado como Z_{t-10} deve entrar na média com o mesmo peso da observação mais recente, Z_t? E, se consideramos que a observação Z_{t-10}, ainda que muito defasada, contém informação tão importante quanto a observação recente Z_t, por que então não usamos também Z_{t-11}, que tem quase o mesmo defasamento?

Se admitimos que as observações mais recentes contêm informação mais útil do que as observações mais antigas, parece que o mais lógico seria criar um método onde as observações sejam ponderadas de acordo com sua idade: quanto mais recente for a observação, maior peso ela deverá receber na média. Na previsão de Z_{T+1}, Z_T deve ter peso maior que Z_{T-1}; Z_{T-1} deve ter peso maior que Z_{T-2}, e assim por diante.

A média M_T poderia então ser expressa na forma:

$$M_T = \alpha_0 Z_T + \alpha_1 Z_{T-1} + \alpha_2 Z_{T-2} + \alpha_3 Z_{T-3} + \ldots$$

onde $\alpha_0 > \alpha_1 > \alpha_2 > \alpha_3 > \ldots$

e $\sum_t \alpha_t = 1$

Usar um método destes seria porém demasiado trabalhoso (teríamos que escolher todos estes pesos α_i). O método de *amortecimento exponencial* resolve o problema de forma simples: usa apenas uma constante α, e os pesos são dados por potências sucessivas desta constante, na forma:

$$M_T = \alpha Z_T + \alpha(1-\alpha) Z_{T-1} + \alpha(1-\alpha)^2 Z_{T-2} + \ldots + \alpha(1-\alpha)^k Z_{T-k} + \ldots + \alpha(1-\alpha)^T M_0 \qquad (\mathbf{1})$$

Como $0 < \alpha < 1$ e $0 < (1 - \alpha) < 1$, os pesos $(1 - \alpha)^k \to 0$ quando $k \to \infty$. Portanto, o maior peso é dado à observação mais recente Z_T, e os pesos seguintes decrescem exponencialmente.

5.1.1. Formas de calcular as previsões por AES

a) AES em termos de médias amortecidas

A forma em (**1**) parece simples, pois envolve apenas uma constante α. Contudo, ainda seria trabalhoso ter que computar a cada instante a média levando em conta todos os valores passados. Existe uma forma recursiva de cálculo que simplifica o trabalho: a cada instante T o estimador M_T do nível da série é calculado pela média ponderada entre a última observação disponível Z_T e a média M_{T-1} calculada no instante anterior:

$$M_T = \alpha Z_T + (1-\alpha) M_{T-1} \qquad (\mathbf{2})$$

A previsão para qualquer horizonte k será dada pela estimativa mais recente do nível (a_T), calculada pela média amortecida M_T:

$$\hat{Z}_{T+k} = \hat{a}_T = M_T \qquad (\mathbf{3})$$

A equivalência entre as eqs. (**1**) e (**2**) pode ser vista se substituirmos recursivamente os valores das médias na eq. (**2**) por suas estimativas:

$$M_T = \alpha Z_T + (1-\alpha) M_{T-1} \qquad \text{(i)}$$

$$M_{T-1} = \alpha Z_{T-1} + (1-\alpha) M_{T-2} \qquad \text{(ii)}$$

$$M_{T-2} = \alpha Z_{T-2} + (1-\alpha) M_{T-3} \qquad \text{(iii)}$$

Substituindo M_{T-1} em (i) por sua estimativa em (ii), obtemos:

$$M_T = \alpha Z_T + (1-\alpha)\left[\alpha Z_{T-1} + (1-\alpha) M_{T-2}\right]$$

$$M_T = \alpha Z_T + \alpha(1-\alpha) Z_{T-1} + (1-\alpha)^2 M_{T-2} \qquad \text{(iv)}$$

Substituindo M_{T-2} em (iv) por sua estimativa em (iii),

$$M_T = \alpha Z_T + \alpha(1-\alpha) Z_{T-1} + (1-\alpha)^2\left[\alpha Z_{T-2} + (1-\alpha) M_{T-3}\right]$$

$$M_T = \alpha Z_T + \alpha(1-\alpha) Z_{T-1} + \alpha(1-\alpha)^2 Z_{T-2} + (1-\alpha)^3 M_{T-3}$$

Se continuarmos substituindo M_{T-3}, M_{T-4}, etc., por suas estimativas, obteremos ao final:

$$M_T = \alpha Z_T + \alpha(1-\alpha)Z_{T-1} + \alpha(1-\alpha)^2 Z_{T-2} + \alpha(1-\alpha)^3 Z_{T-3} + \alpha(1-\alpha)^4 Z_{T-4} + \\ + \ldots + \alpha(1-\alpha)^{T-1} Z_1 + \alpha(1-\alpha)^T M_0$$

que é idêntica à eq. (**1**).

Portanto, no método AES, ao invés de descartarmos as observações mais antigas e darmos o mesmo peso a um subconjunto de observações mais recentes (como no método das médias móveis), fazemos uma média de todos os valores passados da série, ponderados por pesos que seguem uma progressão geométrica. Assim, todos os valores da série são considerados; contudo, quanto mais recente for um valor, maior peso ele terá na média: a última observação (a mais recente) terá um pouco mais de peso na média que a penúltima, a penúltima um pouco mais de peso que a antepenúltima, e assim por diante.

Note que a estimativa atual do nível será dada por uma média ponderada entre as todas as observações da série ($Z_T, Z_{T-1}, \ldots, Z_1$) e um valor inicial M_0. Se a série for longa ($T \to \infty$), o termo do somatório que depende de M_0 tenderá para zero:

$$\alpha(1-\alpha)^T M_0 \to 0$$

O valor inicial M_0, portanto, não terá grande importância na estimativa de M_T.

b) AES em termos do erro de previsão

O método de AES pode ser interpretado de outra maneira. Lembrando que

$$\hat{Z}_{T+1} = M_T$$
$$\hat{Z}_T = M_{T-1}$$

Podemos reescrever a equação (**2**), reproduzida abaixo:

$$M_T = \alpha Z_T + (1-\alpha) M_{T-1}$$

na forma

$$\hat{Z}_{T+1} = \alpha Z_T + (1-\alpha)\hat{Z}_T \qquad (\mathbf{4})$$

Re-organizando os termos,

$$\hat{Z}_{T+1} = \hat{Z}_T + \alpha(Z_T - \hat{Z}_T)$$

O termo entre parênteses é o erro de previsão um passo à frente (Seção **2.3.1**, eq. **1**) :

$$e_T = Z_T - \hat{Z}_T$$

portanto,

$$\hat{Z}_{T+1} = \hat{Z}_T + \alpha e_T \qquad (\mathbf{5})$$

Na forma em (**5**), vemos que a previsão para o próximo valor $\hat{Z}_{T+1}$ será dada pela previsão anterior $\hat{Z}_T$, corrigida por uma fração do último erro de previsão. Se o erro foi positivo (o valor observado foi maior que o previsto), a previsão anterior será aumentada;

se o erro foi negativo (o valor observado foi menor que o previsto), a previsão anterior será reduzida. Desta forma, a previsão será corrigida iterativamente, de acordo com o erro verificado a cada passo. Esta idéia de *autocorreção automática* é simples, mas importante e amplamente usada nos métodos de previsão e controle; por exemplo em mecanismos automáticos para controle de quaisquer processos ou máquinas, nos quais correções são feitas a cada instante com base no erro cometido nos instantes anteriores.

O método pode ainda ser re-escrito como uma combinação entre a observação mais recente e o último erro de previsão:

$$\hat{Z}_{T+1} = Z_T - (1-\alpha)e_T \qquad (6)$$

$$\hat{Z}_{T+1} = \hat{Z}_T + \alpha e_T \qquad \text{(da eq. 5)}$$
$$\hat{Z}_{T+1} = \hat{Z}_T + \alpha(Z_T - \hat{Z}_T)$$
$$\hat{Z}_{T+1} = \hat{Z}_T + \alpha Z_T - \alpha\hat{Z}_T$$
$$\hat{Z}_{T+1} = (Z_T - Z_T) + \hat{Z}_T + \alpha Z_T - \alpha\hat{Z}_T$$
$$\hat{Z}_{T+1} = Z_T - (1-\alpha)Z_T - (1-\alpha)\hat{Z}_T$$
$$\hat{Z}_{T+1} = Z_T - (1-\alpha)(Z_T - \hat{Z}_T)$$
$$\hat{Z}_{T+1} = Z_T - (1-\alpha)e_T$$

Para resumir, reunimos aqui as três formas em que a previsão do método de AES pode ser expressa:

- como combinação da observação mais recente com a previsão anterior (eq. 4):

(i) $\hat{Z}_{T+1} = \alpha Z_T + (1-\alpha)\hat{Z}_T$

- como correção da previsão anterior, feita a partir do erro da previsão anterior (eq. 5):

(ii) $\hat{Z}_{T+1} = \hat{Z}_T + \alpha e_T$

- como combinação da observação mais recente com o erro de previsão anterior (eq. 6):

(iii) $\hat{Z}_{T+1} = Z_T - (1-\alpha)e_T$

As formas mais frequentemente usadas são as em (i) e (ii), mas elas podem ainda ser encontradas re-escritas em termos das médias amortecidas ou de estimativas do nível *a*, em virtude da equivalência (eq. 3, reproduzida abaixo):

$$\hat{Z}_{T+1} = \hat{a}_T = M_T$$

Escrevendo em termos das estimativas do nível *a* da série, teremos as equações (4) e (5) reescritas como:

$$\hat{a}_T = \alpha Z_T + (1-\alpha)\hat{a}_{T-1}$$
$$\hat{a}_T = \hat{a}_{T-1} + \alpha(Z_T - \hat{a}_{T-1})$$

Escrevendo em termos das médias amortecidas:

$$M_T = \alpha Z_T + (1-\alpha)M_{T-1}$$
$$M_T = M_{T-1} + \alpha(Z_T - M_{T-1})$$

Há portanto várias formas de se escrever as previsões: em termos da série de previsões $\hat{Z}_t$, em termos das médias amortecidas M_t, ou em termos das estimativas da constante $\hat{a}_t$. É importante sempre ter em mente, contudo, que as previsões são apenas estimativas da constante *a*, obtidas através de médias ponderadas M_T, e que todas elas são derivadas uma das outras por meio de manipulações algébricas, e produzem por isso exatamente os mesmo resultados. A escolha entre uma forma ou outra é apenas uma questão de conveniência. Neste *site*, usaremos de preferência a forma em (i).

5.1.2. O efeito do valor da constante de amortecimento α

Se o valor da constante de amortecimento α for grande, será dada muita importância à observação mais recente e pouca importância às informações mais antigas. Os pesos decairão rapidamente; por exemplo, se $\alpha = 0{,}9$:

$$\hat{Z}_{T+1} = 0{,}9Z_T + (1-0{,}9)M_{T-1}$$

aplicando este valor na eq. (**1**) (reproduzida abaixo),

$$\hat{Z}_{T+1} = \alpha Z_T + \alpha(1-\alpha)Z_{T-1} + \alpha(1-\alpha)^2 Z_{T-2} + ... + \alpha(1-\alpha)^k Z_{T-k} + ... + \alpha(1-\alpha)^T M_0$$

$$\hat{Z}_{T+1} = 0{,}9Z_T + 0{,}9(0{,}1)Z_{T-1} + 0{,}9(0{,}1)^2 Z_{T-2} + 0{,}9(0{,}1)^3 Z_{T-3}...$$

$$\hat{Z}_{T+1} = 0{,}9Z_T + 0{,}09Z_{T-1} + 0{,}009Z_{T-2} + 0{,}0009Z_{T-3}...$$

As duas observações mais recentes concentrarão 99 % do peso; as observações restantes praticamente não terão nenhuma influência na previsão. Podemos dizer que a processo tem pouca "memória", e que o grau de amortecimento da série de previsões é pequeno, uma vez que cada observação nova afeta grandemente a previsão. No caso extremo de $\alpha = 1$, teremos um previsor ingênuo (*naïve*), onde só a observação mais recente importa:

$$\hat{Z}_{T+1} = Z_T$$

Por outro lado, se o valor de α for baixo, será dada comparativamente pouca importância à observação mais recente Z_T, e muita ao passado (representado pela média anterior M_{T-1}); o decaimento dos pesos será lento. Por exemplo, se $\alpha = 0{,}1$:

$$\hat{Z}_{T+1} = 0{,}1Z_T + 0{,}1(0{,}9)Z_{T-1} + 0{,}1(0{,}9)^2 Z_{T-2} + 0{,}1(0{,}9)^3 Z_{T-3}...$$

$$\hat{Z}_{T+1} = 0{,}1Z_T + 0{,}09Z_{T-1} + 0{,}081Z_{T-2} + 0{,}073Z_{T-3} + 0{,}066Z_{T-4} + ...$$

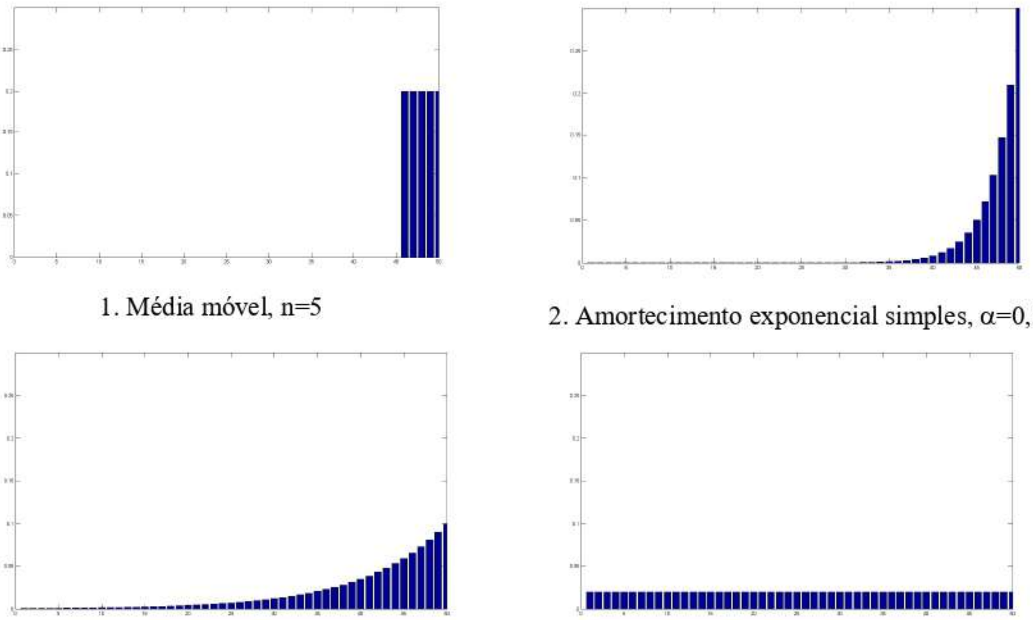
Agora, 99% do peso estará distribuído entre as 22 observações mais recentes; dizemos então que a "memória" do processo é mantida por bastante tempo. Neste caso, o amortecimento é bastante sensível, uma vez que o efeito da informação mais recente Z_T sobre a previsão será muito reduzido.

Portanto, quanto menor for *α*, mais "estável" será a série de previsões, aproximando-se daquela obtida pela média global; quanto maior *α*, mais instável será esta série, aproximando-se daquela do previsor *naïve*. Na prática, o melhor valor de *α* para a previsão de uma série terá que ser determinado experimentalmente, numa amostra de teste.

Para ilustrar o comportamento dos pesos, os gráficos da Fig. **1** abaixo comparam as sequências de pesos em três métodos de previsão – média simples, média móvel e AES, aplicados a uma série com $n = 50$ observações.

No método de *médias móveis* de ordem $n = 5$, as cinco observações mais recentes recebem pesos iguais a 1/5 (gráfico 1). No método de AES, todas as observações da série recebem pesos que decaem geometricamente de acordo com o valor de α (gráficos 2 e 3).

Por fim, no método da *média simples*, todas as observações da série recebem peso iguais a 1/50 (gráfico 4).



1. Média móvel, n=5

2. Amortecimento exponencial simples, α=0,3

3. Amortecimento exponencial simples, α=0,1

4. Média simples

Figura 1. Comparação dos pesos atribuídos às observações passadas nos diferentes métodos

Como exemplo do efeito do valor de α, a Fig. **2** mostra as previsões da série *nhtemp* (temperaturas médias anuais em New Haven, EUA, 1912-1971; disponível no R, no pacote *datasets*) feitas por meio de AES, usando três valores diferentes para α (o valor α = 0,186 foi o valor ótimo encontrado pela otimização do MSE, usando a função `optim`).

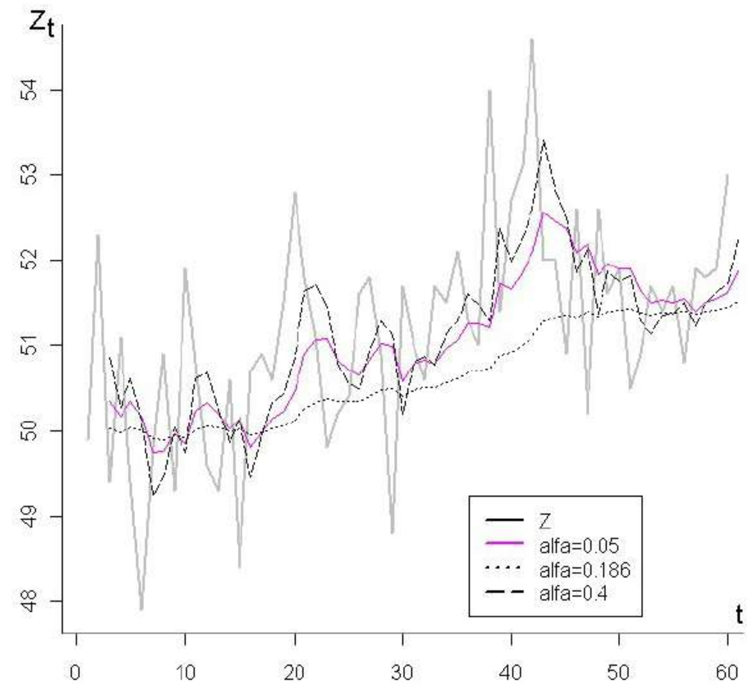


Figura 2. Previsão da série *nhtemp* por AES

5.1.3. Métodos de inicialização

Todos os métodos de amortecimento exponencial exigem a especificação de alguns valores iniciais. Uma vez que estes métodos são recursivos e cada parâmetro é estimado pela combinação da observação mais recente com a estimativa anterior, é preciso começar de alguma estimativa inicial.

Se usarmos a forma (i) do método

$$\hat{Z}_{T+1} = \alpha Z_T + (1-\alpha)\hat{Z}_T$$

As primeiras iterações serão:

$$\hat{Z}_1 = \alpha Z_0 + (1-\alpha)\hat{Z}_0$$
$$\hat{Z}_2 = \alpha Z_1 + (1-\alpha)\hat{Z}_1$$

Como não faz sentido falar em Z_0 ou $\hat{Z}_0$, as previsões serão iniciadas a partir do instante t = 2. Para obter uma estimativa da previsão inicial $\hat{Z}_1$, podemos usar diferentes métodos:

Método 1: tomar simplesmente à primeira observação como estimativa inicial $\hat{Z}_1$:

$$\hat{Z}_1 = Z_1$$

Se fazemos isto, a previsão para a segunda observação será também igual à primeira observação:

$$\hat{Z}_2 = \alpha Z_1 + (1-\alpha) Z_1 = Z_1$$

Método 2: se há muitos dados disponíveis, tomar a média de algumas das primeiras observações, por exemplo:

$$\hat{Z}_1 = \frac{Z_1 + Z_2 + Z_3 + Z_4}{4}$$

É claro que, ao fazer isto, estamos trapaceando um pouco, pois estaremos usando informações do futuro (instantes t =2 a 4) para estimar um valor do passado (Z_1); mas isto deixa de ter importância a partir do instante t = 5.

Se o valor de α for razoavelmente grande, o método de inicialização não faz muita diferença, já que a previsão colocará muito peso na observação mais recente, e pouco peso nas mais antigas; depois de algumas iterações, os valores usados para a inicialização já terão sido "esquecidos". Se porém o α for pequeno e a série for curta, o método de inicialização pode fazer diferença no resultado final, pois a previsão será por muito tempo afetada pelos valores iniciais escolhidos, e não haverá um número suficiente de iterações até que estes valores sejam esquecidos.

5.1.4. Escolha do valor de α

O valor de α deve escolhido de forma a minimizar uma medida de erro, geralmente o MSE, mas não há como fazê-lo de forma analítica. Em geral, a escolha é feita por uma *busca em grade* (isto é, escolhemos alguns valores para α, por exemplo, 0.1, 0.2, etc., experimentamos o método com cada um deles, e verificamos qual deles dá melhores resultados), ou por meio de algoritmos de otimização não-linear, como o `optim` ou o `optimize`, ambos partes do pacote `stats` do R.

A escolha do valor de α pode também ser afetada por outros fatores, como o nível de confiança que temos nos valores iniciais atribuídos aos parâmetros, e o tamanho das séries de dados disponíveis, como mencionado na seção 5.1.3. Se α é baixo, os valores iniciais influenciam as previsões por muito tempo; por isso, se as séries disponíveis são curtas, uma possibilidade é fazer as primeiras iterações com valores de α relativamente altos, para que um erro nos valores iniciais não se propague por muito tempo nas série de previsões, e reduzir o α nas iterações posteriores. Por outro lado, se as séries disponíveis são longas, será possível conseguir boas estimativas finais dos parâmetros mesmo que os valores iniciais sejam ruins; neste caso, pode ser melhor usar um α relativamente baixo em toda a série (por exemplo 0,01 ou 0,02), para que a série de previsões reaja de forma lenta às mudanças que ocorrerem, e não seja muito afetada pelo ruído presente nos dados.

5.1.5. Variância do erro de previsão

Para um modelo de nível constante estimado por AES, pode ser demonstrado que a variância do erro de previsão um-passo-à-frente é dada por:

$$\sigma_e^2 = \text{var}(e_{T|T-1}) = \text{var}(\hat{Z}_{T|T-1} - Z_T)$$

$$\sigma_e^2 = \frac{2}{2-\alpha}\sigma_\varepsilon^2$$

Note que variância do erro de previsão σ_e^2 é maior que a variância da série, σ_ε^2. As duas variâncias se aproximam quando $\alpha \to 0$; este valor de α corresponde à previsão por média global, que seria a previsão ideal se o nível do modelo fosse realmente constante todo o tempo.

Em teoria, intervalos de confiança (ICs) para as previsões poderiam ser calculados a partir desta estimativa da variância. Na prática, porém, esta variância estará provavelmente subestimada, o que levaria a ICs demasiado estreitos. Não é realista supor que o nível seja constante em todo o tempo (justamente por isso é que foram criados os métodos de MM e AES, que utilizam modelos locais!), e que as previsões para qualquer horizonte tenham a mesma variância e possam por tanto ser representadas por ICs de mesma amplitude.

Em geral, nas previsões por métodos de amortecimento exponencial estaremos interessados em obter apenas previsões *pontuais*; previsões por meio de ICs são feitas quando trabalhamos com modelos estatisticamente mais complexos, como os ARIMA (Cap. **6**), ou os modelos em espaço de estados, que não serão discutidos neste livro.

5.1.6. Conclusão

Se uma série realmente tem o nível constante todo o tempo, a melhor estimativa do nível seria a obtida pelo método da média global (Seção **4.1.1**). Na prática, porém, na maioria das séries o nível oscila um pouco, e não se mantém constante por muito tempo. É melhor então usar um método *local*, que supõe que o nível seja constante apenas numa pequena janela de tempo, e considere nos cálculos apenas as observações mais recentes; métodos assim são o das médias móveis (seção **4.1.2**) e o de AES. Estes dois métodos, apesar de muito simples, são os métodos mais usados para previsões no dia-a-dia de empresas e na administração pública. Qual dos dois é melhor? Não é possível criar regras para prever qual teria o melhor desempenho, num problema específico; o melhor é fazer testes empíricos, usado os dados disponíveis.

Note que ambos os métodos (MM e AES) só funcionam em séries de nível constante; se a série tem uma tendência crescente (ou decrescente), as previsões por estes métodos estarão sempre subestimando (ou superestimando) o valor real. Nestes casos, será melhor usar um método próprio para séries com tendência, como os que serão vistos na Seção **5.2**.