

2

A Modelagem Univariada por Holt-Winters

2.1

Considerações iniciais

Como um dos objetivos do trabalho é aperfeiçoar os métodos univariados “*benchmark*”, a modelagem e a previsão do consumo de carga elétrica pela técnica Holt-Winters serão feitas de três formas distintas. Modificando a formulação original, os modelos irão incorporar dois ciclos: um diário e outro semanal. Em seguida, serão construídos modelos “estendidos” considerando uma constante de amortecimento para a tendência da série, quando o horizonte de previsão é maior que 01 passo-à-frente, e também modelos assumindo erros estruturados. A motivação para focalizar a atenção nestes modelos alternativos surge em razão do método em sua forma convencional não traduzir bem o comportamento de séries horárias. Logo, é válido verificar a qualidade das estimativas obtidas a partir das inovações que estão sendo propostas.

Este capítulo encontra-se subdividido em cinco partes. Primeiro, sucintamente, coloca-se os aspectos teóricos relativos à modelagem Holt-Winters tradicional. Posteriormente, as principais características de suas três formas estendidas são analisadas separadamente. Isto permite que os resultados fiquem mais claros, facilitando as comparações no capítulo 5 fundamentais para o estudo pretendido. Finalmente, faz-se alguns comentários finais.

2.2

A abordagem tradicional dos modelos Holt-Winters

A técnica Holt-Winters faz parte de um conjunto de modelos para a previsão abrangido pelo Método de Amortecimento Exponencial (MAE) que normalmente é empregado para modelar tendência e/ou sazonalidade existentes na série temporal. Portanto, interessa fazer uma breve descrição do MAE e, ao mesmo tempo, justificar a necessidade de trabalhar com a sua forma estendida na tentativa

de obter previsões melhores quando a referência são dados horários como os que estão sendo investigados.

O Método de Amortecimento Exponencial ganhou destaque em 1970 possibilitando realizar a modelagem univariada dos dados, isto é, fazer as previsões de uma série considerando somente os seus valores defasados. São modelos de validade local e, por isso, o horizonte de previsão não deve ser muito grande.

Chama-se a atenção para o fato de que o MAE se distingue dos modelos de regressão linear, pois o seu objetivo não é ajustar uma reta aos dados, não se trata de um ajuste de curva. A todo instante busca-se determinar a melhor reta, ou seja, é uma reta para cada instante de tempo, enquanto, no problema de regressão, pretende-se definir a melhor reta que passa por todos os pontos.

Segundo Taylor (2002) [15], o MAE é um procedimento bastante difundido, especialmente em aplicações cujas séries necessitam de um procedimento automático de atualização, devido a robustez e a exatidão de suas estimativas. Esse método é uma alternativa às formulações matemáticas no tempo que atribuem um peso único a todas as observações disponíveis. O MAE corrige os pesos do conjunto de dados conforme a idade dos mesmos dando pesos maiores às informações mais recentes e pesos menores às observações mais antigas.

No contexto do trabalho, vale a pena introduzir uma rápida evolução dos métodos para estimar os parâmetros. Inicialmente aparece o método ingênuo com o qual se realiza a previsão para qualquer instante de tempo como sendo igual ao último valor observado. Portanto, a modelagem está sempre em função do último dado. Desta forma de estimação se evolui para a média aritmética do conjunto histórico, quando todas as observações ganham o mesmo peso. Em seguida, surge o conceito de médias móveis (MM) selecionando as 'N' informações mais relevantes para a previsão e privilegiando os dados mais recentes. Destaca-se que $N =$ tamanho da média móvel.

A média móvel leva em conta as 'N' últimas observações e depois efetua a média aritmética das mesmas. Pelo método, a cada nova observação que chega, a mais antiga é retirada. Embora o tamanho da MM permaneça fixo, o conjunto está sempre se movendo. Daí o sentido da palavra "*móvel*" na expressão. Aplicando-se

MM sucessivamente há uma suavização dos dados, quer dizer, as MM limpam um pouco e permitem ver a tendência da série. Quanto maior o número de MM aplicadas, mais o valor da estimativa aproxima-se da média aritmética que é a forma de estimação com máxima suavização.

Supondo MM (N=3):

$$\hat{a}(t+1) = \frac{[Z_t + Z_{t+1} + Z_{t+2}]}{3}$$

$$\hat{a}(t+2) = \frac{[Z_{t+1} + Z_t + Z_{t-1}]}{3}$$

Para estimar mais de um parâmetro é necessário aumentar o número de médias móveis. Em modelos lineares, é preciso aplicar MM duplas, para modelos quadráticos deve-se fazer MM triplas, e assim por diante. Reitera-se que em qualquer destes casos, o conceito de MM permanece o mesmo. Isto é mostrado nas equações que seguem (Montgomery (1976) [11])

$$M_t = \frac{[Z_t + Z_{t-1} + \dots + Z_{t-n+1}]}{N}$$

$$M_t^{[2]} = \frac{M_t + M_{t-1} + \dots + M_{t-n+1}}{N}$$

$$M_t^{[3]} = \frac{M_t^{[2]} + M_{t-1}^{[2]} + \dots + M_{t-n+1}^{[2]}}{N}$$

Isto é :

$$\begin{aligned} M_t &= M_t [a_1, a_2] & ; & & \hat{a}_1(t) &= f(M_t, M_t^{[2]}) \\ M_t^{[2]} &= M_t^{[2]} [a_1, a_2] & ; & & \hat{a}_2(t) &= f((M_t, M_t^{[2]}) \\ M_t^{[3]} &= M_t^{[3]} [a_1, a_2, a_3] & ; & & \hat{a}_3(t) &= f((M_t, M_t^{[2]}, M_t^{[3]}) \end{aligned}$$

É fácil notar que o número de equações deve ser idêntico ao número de parâmetros a serem estimados.

Por fim, o MAE, que será detalhado mais adiante, surge atribuindo pesos diferentes às observações e funcionando como uma média móvel ponderada no tempo.

$$Mt = \alpha Z_t + (1 - \alpha) Mt-1 \quad \alpha \in [0,1]$$

$$Mt^{[2]} = \alpha Mt + (1 - \alpha) Mt-1^{[2]} \quad \alpha \in [0,1]$$

$$Mt^{[3]} = \alpha Mt^{[2]} + (1 - \alpha) Mt-1^{[3]} \quad \alpha \in [0,1]$$

O MAE pode ser aplicado tanto em séries não sazonais modelando apenas nível e/ou tendência (método de Brown), quanto em séries sazonais (método de Winters). Também faz parte do modelo geral MAE o método de amortecimento direto. Porém, os pormenores desta última categoria de modelos não serão explorados em virtude dos mesmos não se enquadrarem entre as prioridades definidas no trabalho.

De acordo com o método de Brown, os modelos podem ser constantes (nível médio da série constante no tempo), lineares (quando a série apresenta alguma tendência no tempo crescente ou decrescente) ou quadráticos (o modelo é representado por uma função quadrática do tempo).

Em linhas gerais:

Z_1, Z_2, \dots, Z_t : série temporal de tamanho t ;

Z_t : idade dos dados;

T : último instante = origem;

$\mu(t)$: nível da série em função do tempo;

Tem-se para a modelagem de Brown:

Modelo geral: $Z_t = \mu(t) + et$

Onde: $et \sim N(0, \sigma_e^2)$; $\sigma_e^2 =$ variância constante.

Modelo constante: $\mu(t) = a_1$; $t = 1, 2, 3 \dots T$

$Z_t = a_1 + et$

Modelos linear: $\mu (t) = a_1 + a_2 (t)$; $t = 1, 2, 3...T$

$$Z_t = a_1 + a_2 (t) + et$$

Modelo quadrático: $\mu (t) = a_1 + a_2 (t) + a_3 (t)^2$; $t = 1, 2, 3...T$

$$Z_t = a_1 + a_2 (t) + a_3 (t)^2 + et$$

Para chegar a equação de previsão, que fornecerá a média das previsões, basta tirar o valor esperado das formulações matemáticas dos modelos. Neste caso, o erro médio quadrático está sendo minimizado.

Equação de previsão: $E(Z_t) = E [\mu (t) + et / Z_1...Z_t]$

Como mencionado, o MAE atribui pesos diferentes aos dados conforme a sua idade. Desta forma, nos modelos existem os parâmetros (medida numérica que descreve alguma característica da série como nível, tendência e sazonalidade e que varia no tempo sendo atualizado a todo instante) e os hiperparâmetros (quantidade fixa e desconhecida que não varia no tempo e que necessita ser previamente estimada para se estimar o parâmetro). Todo hiperparâmetro é obtido minimizando o erro de previsão 01 passo – à – frente.

$$\text{erro} (t) = Z_t - \hat{Z}_t (\tau)$$

Para estimar os parâmetros, o MAE executa uma combinação convexa entre o presente e o passado gerando a estimativa. Essa combinação é feita pelo hiperparâmetro (constante de amortecimento do modelo). O método de Brown considera apenas um hiperparâmetro (α). Ressalta-se que quanto maior o valor de (α), maior o peso dado ao presente da série. Assim, as estimativas para os parâmetros dos modelos são estabelecidas da seguinte forma:

$(\alpha) \in [0,1]$: peso dado ao presente da série.

$(1 - \alpha)$: peso dado ao passado da série.

$$\hat{a}_1(t) = \alpha Z_t + (1 - \alpha) \hat{a}_1(t - 1)$$

Onde:

Z_t : presente da série.

$[\hat{a}_1(t - 1) = \alpha Z_{t-1} + (1 - \alpha) \hat{a}_1(t - 2)]$: passado da série.

Quanto maior a variabilidade da série, maior deverá ser o valor de (α) , pois é preciso acompanhar a variação dando ênfase as informações trazidas pelos últimos dados. Percebe-se uma relação inversa entre o tamanho da série 'T' e a constante de amortecimento (α) . Quanto maior a série, mais informações são necessárias para efetuar uma boa previsão e deve-se dar um peso maior ao passado da série. Da mesma forma, o raciocínio contrário é válido.

Conhecidos os parâmetros têm-se as previsões pontuais para qualquer horizonte.

Um desdobramento dos modelos de Brown é o método de Holt que considera exclusivamente os modelos lineares. Na modelagem de Holt há dois hiperparâmetros: (α) para a atualização seqüencial do nível e (β) para a atualização seqüencial da tendência. Isso se justifica porque a variação do nível pode ser diferente daquela registrada pela taxa de crescimento.

Fazendo: $\mu(t) = a_1 + a_2(t)$

Modelo de Holt: $Z_t = \mu(t) + e_t$

Onde:

α : constante de amortecimento do nível.

β : constante de amortecimento da tendência.

Numa versão mais sofisticada deste método de Holt encontra-se ainda um terceiro hiperparâmetro responsável por amortecer o crescimento à medida que o horizonte de previsão aumenta. Essa formulação será oportunamente explicada na seção 2.3.

Em se tratando de séries sazonais, uma das alternativas do método de amortecimento exponencial para realizar este tipo de modelagem é aplicar a técnica de Winters. Cabe lembrar que sazonalidade refere-se a um tipo de repetição periódica definida que está relacionada às estações do ano.

O ciclo sazonal tem comprimento (L).

Alguns exemplos :

Séries mensais: $L = 12$;

Séries trimestrais: $L = 4$;

Séries quadrimestrais: $L = 3$;

Séries semanais: $L = 52$; entre outros.

Segundo o método de Winters, a modelagem é feita utilizando valores discretos, via fatores sazonais, que caracterizam o período sazonal. Estes fatores podem ser incorporados aos modelos de forma aditiva ou multiplicativa. O primeiro caso retrata séries bem comportadas cuja variabilidade dos fatores é constante. Nos modelos multiplicativos a variância dos fatores sazonais aumenta ou decresce com o nível implícito da série.

Modelo aditivo: $Z_t = \mu(t) + \rho_t + \epsilon_t$

Modelo multiplicativo: $Z_t = \mu(t) * \rho_t + \epsilon_t$

Onde:

$\mu(t)$: nível médio da série

$\rho(t)$: fatores sazonais (parâmetros da parte sazonal do modelo)

ϵ_t : componente aleatório não explicado pelo modelo.

Destaca-se que, como a ênfase deste capítulo é a modelagem por Holt-Winters, é especialmente interessante focar o estudo nos modelos lineares em que:

$$\mu(t) = a_1 + a_2(t).$$

Os estimadores dos fatores sazonais são seqüencialmente definidos utilizando o método de amortecimento exponencial. As equações de atualização passam a considerar três parâmetros e, respectivamente, três hiperparâmetros.

Modelo Holt-Winters convencional:

$$Z_t = [a_1 + a_2(t)] * \rho_t + \epsilon_t$$

ρ_t : fatores sazonais.

Equação de previsão:

$$E(Z_t) = E\{[a_1 + a_2(t)] * \rho_t + \epsilon_t / Z_1 \dots Z_t\}$$

↓

$$\hat{Z}_t(\tau) = [\hat{a}_1(t) + \tau \hat{a}_2(t)] * \rho_m^{t+\tau}(t).$$

A partir desta equação de previsão todos os fatores sazonais são estimados, mas somente aquele correspondente à informação recebida é atualizado. Isto pode ser verificado pelas equações abaixo.

a) Nível: $\hat{a}_1(t)$; hiperparâmetro “ α ”

$$\hat{a}_1(t) = \alpha [Z_t / \hat{\rho}_m^{(t)}(t-1)] + (1 - \alpha) [\hat{a}_1(t-1) + \hat{a}_2(t-1)]$$

b) Tendência: $\hat{a}_2(t)$; hiperparâmetro “ β ”

$$\hat{a}_2(t) = \beta [\hat{a}_1(t) - \hat{a}_1(t-1)] + (1 - \beta) \hat{a}_2(t-1)$$

c) Fatores sazonais $\hat{\rho}_j(t)$ $j = 1, 2, 3, \dots$; hiperparâmetro “ γ ”

$$\hat{\rho}_j^{*(t)}(t) = \gamma [Z_t / \hat{a}_1(t)] + (1 - \gamma) \hat{\rho}_m^{(t)}(t)$$

Onde:

$\hat{\rho}_m^{(t)}(t)$: fator sazonal correspondente ao mês da informação Z_t recebida.

$\hat{\rho}_j^{*(t)}(t) = \hat{\rho}_j^{*(t)}(t - 1)$; $j = 1, 2, \dots, L$; $j \neq m(t)$: fatores sazonais estimados e não atualizados.

É importante salientar que para a soma dos fatores sazonais se igualar ao comprimento ‘L’ da sazonalidade, é preciso normalizá-los. A normalização é útil por permitir interpretar os fatores. Sabe-se que para as séries não sazonais todos os valores são iguais a um. Então, supondo o caso da demanda por energia elétrica, no verão, os fatores assumem índices maiores que um (1) indicando um consumo acima da média. No inverno, os valores inferiores à um (1) mostram um consumo abaixo da média esperada. Essa análise só é possível porque os fatores sazonais estão normalizados. Neste caso, a diferença entre $\hat{\rho}_j^{*(t)}(t)$ e $\hat{\rho}_j^{(t)}(t - 1)$ está sendo distribuída entre todos os ‘L’ fatores sazonais.

Normalização:

$$\hat{\rho}_j^{(t)} = \left[\hat{\rho}_j^{*(t)} / \sum_{j=1}^S \hat{\rho}_j^{*(t)} \right] * L$$

A equação acima conta que, matematicamente, o fator normalizado é igual ao fator sazonal estimado em ‘c’ (pág. 27) dividido pelo somatório de (c + fatores anteriores não atualizados) multiplicado pelo comprimento sazonal.

O método de amortecimento exponencial possibilita obter uma modelagem “robusta” devido aos parâmetros serem atualizados a todo instante, o que torna os modelos adaptativos. Apesar destes modelos serem adaptativos, podem ocorrer

situações imprevisíveis que provocam descontinuidade na série como foi registrado com o racionamento de energia em 2001. Salienta-se que o bom modelo é adaptativo, mas nenhum modelo prevê descontinuidade.

Toda a teoria da técnica tradicional Holt-Winters evidencia que o método comporta apenas um único padrão cíclico relacionado às variações observadas na série em virtude de mudanças nas estações do ano.

Segue uma síntese da história da arte.

Síntese da evolução dos métodos:

Método ingênuo



Média aritmética



Médias Móveis



Método de Amortecimento Exponencial (MAE)

As informações exibidas nesta seção estão baseadas na leitura de Montgomery (1976) [11] e Souza (1983) [12].

Na próxima seção, apresenta-se as inovações propostas neste trabalho para a previsão do consumo horário de carga elétrica do sudeste brasileiro cuja série apresenta dois ciclos: diário e semanal.

2.3

A Modelagem Holt-Winters estendida: modelos com ciclos duplo

Em séries horárias verifica-se um comportamento muito similar no perfil do consumo de um determinado dia da semana com o correspondente dia das semanas adjacentes. Essa repetição é também notada com determinada hora do dia quando comparada com a mesma hora de um outro dia no grupo de dias úteis ou de fins de semana.

Considerando que em dados horários não se observa o padrão sazonal, uma maneira alternativa para modelar as séries deste tipo pelo método de amortecimento exponencial é incorporar simultaneamente os ciclos diário (24 horas) e semanal (168 horas). Esta formulação estendida do modelo Holt-Winters convencional pode aparecer tanto na forma aditiva quanto multiplicativa. Neste estudo, a versão multiplicativa será avaliada. Deve ser lembrado que por definição ciclo é uma repetição periódica de um fenômeno que necessariamente não tem haver com sazonalidade (Souza (1983) [12]).

O gráfico 2.1 reporta o consumo de carga diário do 6º, 7º, 8º, 9º e 10º dias úteis de dados. Estes dias referem-se a uma segunda, terça, quarta, quinta e sexta-feira. Como pode ser notado, há uma clara tendência de aumento da demanda por volta das 7 horas em todos os 5 dias e uma diminuição do consumo por volta das 19 e 20 horas.

O gráfico 2.2, retrata os 11º e 12º dias, um sábado e domingo. Observa-se que o comportamento nos dois dias do fim de semana também se parecem. Isto é, há uma tendência de redução na demanda por energia pelas 9 horas e um aumento do consumo entre às 20 e 21 horas.

Pelo gráfico 2.3, que ilustra as 2ª, 3ª e 4ª semanas de dados, é possível constatar que o comportamento nestes 3 períodos considerados é praticamente idêntico.

Assim, fica clara a necessidade de capturar estes padrões de repetição diário e semanal para conseguir previsões consistentes. Embora o trabalho utilize somente dois ciclos, a teoria Holt-Winters estendida, sugerida por Taylor (2002) [15], se aplica a qualquer quantidade de ciclos superior a um.

Gráfico 2.1: Ciclo diário em dias úteis

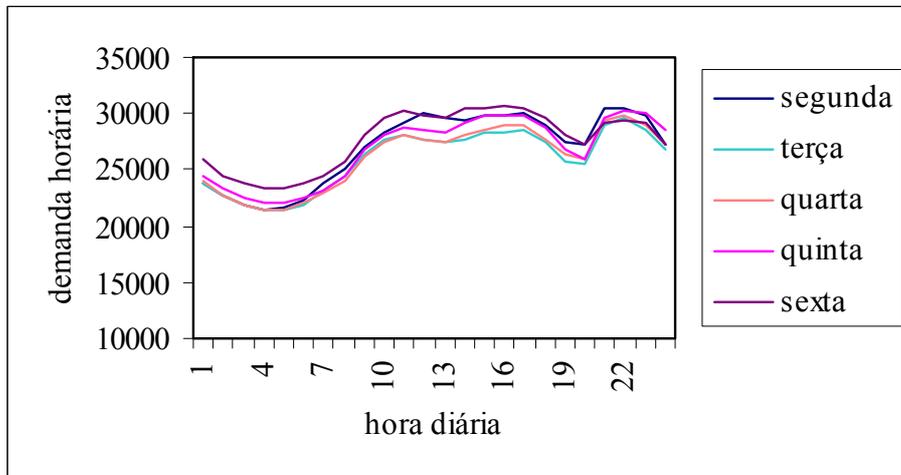


Gráfico 2.2: Ciclo diário em fins de semana

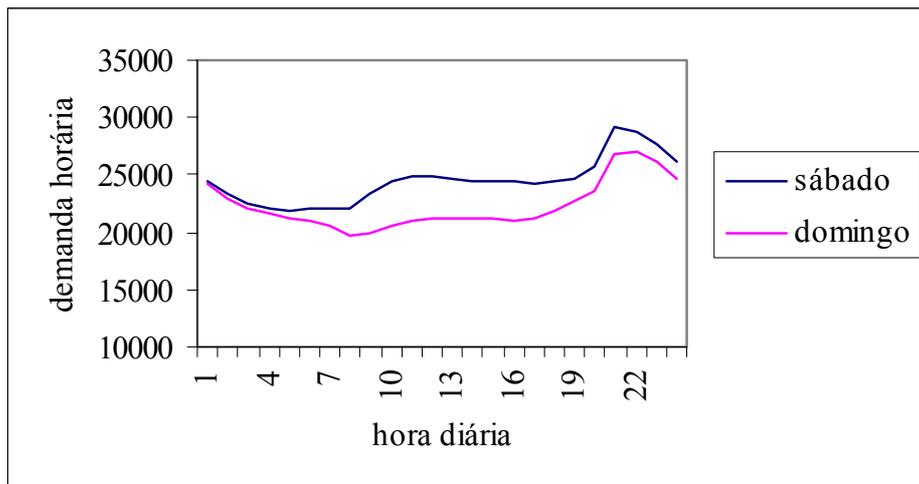
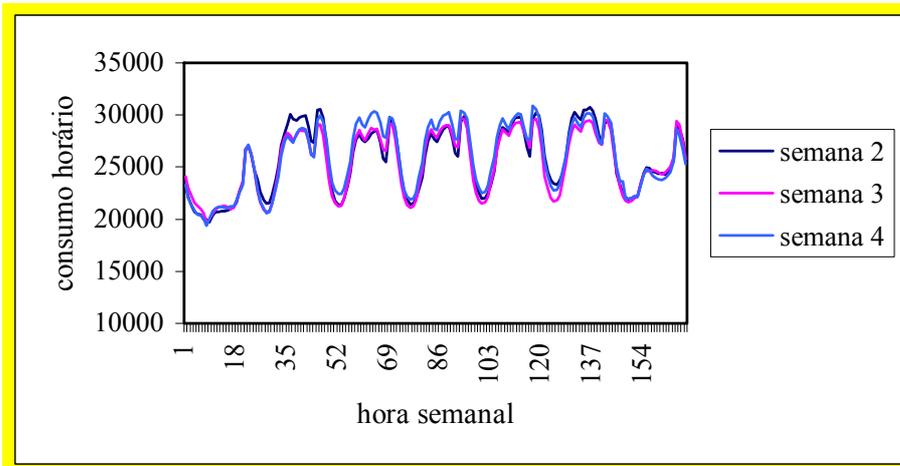
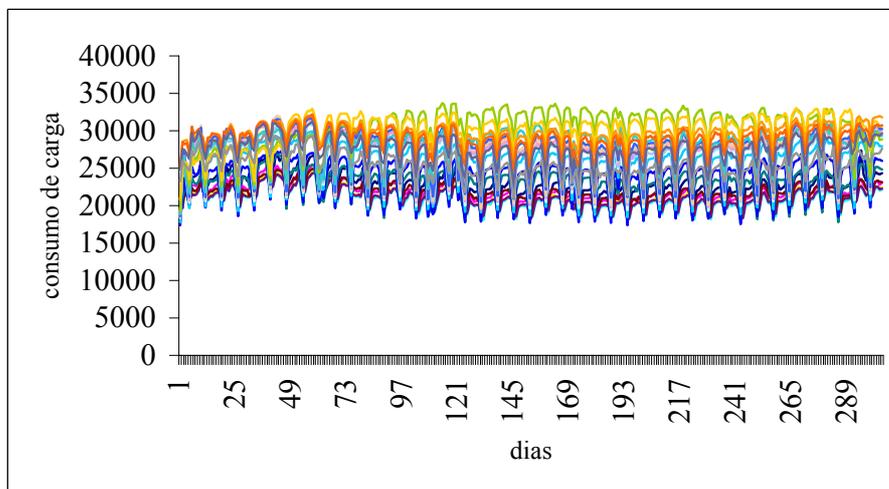


Gráfico 2.3: Ciclo semanal



Com relação ao banco de dados utilizados, pelo gráfico 2.4, nota-se que não há *outliers* e, portanto, nenhuma informação disponível foi retirada.

Gráfico 2.4: Demanda de carga entre 01/01/03 a 31/10/03 (período *in-sample*)

A normalização dos ciclos é dada por:

$$\sum_{j=1}^{L1} \hat{c}_{ij}(\cdot) = L1$$

$$\sum_{j=1}^{L2} \hat{c}_{2j}(\cdot) = L2$$

O modelo Holt-Winters estendido com dois ciclos pode ser escrito como:

$$Z_t = [b_1(t) + b_2(t)] * c_1(t) * c_2(t) + e_t$$

Onde:

Z_t : série temporal de tamanho t

b_1 : nível da série

b_2 : crescimento da série

c_1 : ciclo de comprimento $L1$

c_2 : ciclo de comprimento $L2$

e_t : erro do modelo

Neste modelo existem $(L1 + L2 + 2)$ parâmetros e quatro hiperparâmetros abaixo citados.

α : atualização seqüencial do parâmetro nível.

β : atualização seqüencial do parâmetro tendência.

γ : atualização seqüencial dos 'L1' parâmetros do ciclo 1 (diário).

δ : atualização seqüencial dos 'L2' parâmetros do ciclo 2 (semanal).

Para os ciclos diário e semanal, $L1 = 24$ e $L2 = 168$. Logo, este modelo tem 194 parâmetros a serem estimados. As equações de atualização são dadas por:

a) nível: $\hat{b}_1(t)$; constante ' α '

$$\hat{b}_1(t) = \alpha * [Z_t / \hat{c}_{1 h_1(t)}(t-1) * \hat{c}_{2 h_2(t)}(t-1)] + (1 - \alpha) * [\hat{b}_1(t-1) + \hat{b}_2(t-1)]$$

b) crescimento: $\hat{b}_2(t)$; constante ‘ β ’

$$\hat{b}_2(t) = \beta * [\hat{b}_1(t) - \hat{b}_1(t-1)] + (1 - \beta) * \hat{b}_2(t-1)$$

Observa-se que a equação (b) é idêntica aquela do modelo convencional.

c) ciclo 1: $\hat{c}_{1\ h1}(t)$; constante ‘ γ ’

$$\hat{c}_{1\ h1}(t) = \gamma * [Z_t / \hat{b}_1(t) * \hat{c}_{2\ h2}(t-1)] + (1 - \gamma) * \hat{c}_{1\ h1}(t-1)$$

d) ciclo 2: $\hat{c}_{2\ h2}(t)$; constante ‘ δ ’

$$\hat{c}_{2\ h2}(t) = \delta * [Z_t / \hat{b}_1(t) * \hat{c}_{1\ h1}(t-1)] + (1 - \delta) * \hat{c}_{2\ h2}(t-1)$$

Onde:

h1: hora correspondente ao ciclo 1 no instante ‘t’; $h1(.) \in [1, 2, \dots, L1]$.

h2: hora correspondente ao ciclo 2 no instante ‘t’; $h2(.) \in [1, 2, \dots, L2]$.

As quatro equações (a), (b), (c) e (d) mostram que a atualização é sempre feita pela combinação convexa entre o presente e o passado. Vale ressaltar que os fatores \hat{c}_{1j} e \hat{c}_{2j} devem ser normalizados pela questão da interpretabilidade. Isto porque, à medida que os fatores vão sendo revisados, a soma total deve continuar se igualando ao comprimento do ciclo.

A equação de previsão é dada por:

$$\hat{Z}_t(\tau) = [(\hat{b}_1 + \tau \hat{b}_2(t)) * \hat{c}_{1(t)\ h1}(t + \tau) * \hat{c}_{2(t)\ h2}(t + \tau)]$$

2.3.1

Alguns detalhes da análise da modelagem empírica da técnica Holt-Winters estendida aos dados de consumo de carga horária do sudeste brasileiro

Uma vez apresentados os aspectos teóricos dos modelos que incorporam dois ciclos, pode-se partir para a sua avaliação empírica. Para isso, é preciso definir as estimativas iniciais dos parâmetros $\hat{b}_1(0)$, $\hat{b}_2(0)$, $\hat{c}_{1j}(0)$ e $\hat{c}_{2j}(0)$ e de seus hiperparâmetros (α , β , γ , δ). Com relação aos valores dos parâmetros, o primeiro procedimento é organizar a série temporal a ser estudada na forma matricial.

Para o cálculo do ciclo 1, as linhas da matriz correspondem aos número de dias e as suas colunas representam cada hora do dia. O banco de dados contém os 365 dias de 2003. O cálculo dos valores iniciais será feito considerando os 10 primeiros meses do ano que equivalem a 304 dias *in-sample*. Os últimos 2 meses servirão para a avaliação *out-of-sample* do modelo.

Exemplo:

$$\begin{array}{cccccccc} Z_{1,1} & Z_{1,2} & Z_{1,3} & \dots & Z_{1,24} \\ Z_{2,1} & Z_{2,2} & Z_{2,3} & \dots & Z_{2,24} \\ Z_{3,1} & Z_{3,2} & Z_{3,3} & \dots & Z_{3,24} \\ Z_{4,1} & Z_{4,2} & Z_{4,3} & \dots & Z_{4,24} \\ \dots & \dots & \dots & \dots & \dots \\ Z_{304,1} & Z_{304,2} & Z_{304,3} & \dots & Z_{304,24} \end{array}$$

Esta matriz de dados deve ser interpretada da seguinte forma, como os dados *in-sample* referem-se ao período entre 01/01/03 a 31/10/03, existem 304 dias completos de dados históricos. Logo, cada observação na linha deve ser lida como: $Z_{1,1}$ (dia 01 na hora 1), $Z_{2,2}$ (dia 02 na hora 2), $Z_{3,3}$ (dia 03 na hora 3), e assim por diante até a última observação $Z_{304,24}$ (dia 304 na hora 24). Portanto, a dimensão da matriz para o cálculo do valor inicial do ciclo 1 é (304 X 24)

Com a matriz montada, o próximo passo é calcular as médias diárias. Isso é feito por:

$$\bar{X}_i = \sum_{j=1}^{L1=24} X_{ij} / L1 = 24$$

Na seqüência, calcula-se os fatores diários grosseiros:

$$c^*_{1,i,j} = X_{ij} / \bar{X}_i, \quad \text{para } j = 1, \dots, 24 ; i = 1, \dots, 304.$$

Obtém-se então a seguinte matriz de fatores grosseiros do ciclo 1:

$$\begin{array}{cccccccc} C^*_{1,1,1} & C^*_{1,1,2} & C^*_{1,1,3} & \dots & C^*_{1,1,24} \\ C^*_{1,2,1} & C^*_{1,2,2} & C^*_{1,2,3} & \dots & C^*_{1,2,24} \\ C^*_{1,3,1} & C^*_{1,3,2} & C^*_{1,3,3} & \dots & C^*_{1,3,24} \\ C^*_{1,4,1} & C^*_{1,4,2} & C^*_{1,4,3} & \dots & C^*_{1,4,24} \\ \dots & \dots & \dots & \dots & \dots \\ C^*_{1,304,1} & C^*_{1,304,2} & C^*_{1,304,3} & \dots & C^*_{1,304,24} \end{array}$$

Essa matriz tem a mesma dimensão da matriz de dados original.

Os fatores dos ciclos diários serão:

$$c^*_{1,j} = \sum_{i=1}^{304} c^*_{1,i,j} / 304 ; \quad \text{para } j = 1, \dots, 24$$

Isso equivale a tirar a média das 24 colunas da matriz de fatores grosseiros com dimensão (304 X 24), transformando-a em um vetor linha (1 X 24).

A normalização dos fatores ocorre por:

$$c_{1,j} = \left(c^*_{1,j} / \sum_{j=1}^{L1=24} c^*_{1,j} \right) * 24$$

Todo este método de inicialização que foi descrito pode ser conferido em Montgomery (1976) [11].

Assim, os valores iniciais para os parâmetros são:

$$\hat{b}_1(0) \equiv \bar{X}_2 = 25711,4$$

A estimativa $\hat{b}_1(0)$ refere-se a média da segunda semana de dados, já que a primeira semana de janeiro de 2003 conta com apenas 4 dias e, portanto, está incompleta.

Para a estimativa $\hat{b}_2(0)$ tem-se:

$$\hat{b}_2(0) \equiv (\bar{X}_3 - \bar{X}_2)/168 = (25424,8 - 25711,4)/168 = - 1,706$$

Lembrando que \bar{X}_3 é a média da terceira semana de dados e \bar{X}_2 é a média da segunda semana de dados.

Os fatores iniciais do ciclo1 (diário) estão expostos na tabela 1.1.

Tabela 2.1: Fatores iniciais do ciclo 1

Hora (t)	Fator do ciclo 1 na hora (t)
1	0,8994
2	0,8529
3	0,8263
4	0,8150
5	0,8171
6	0,8413
7	0,8830
8	0,9208
9	0,9879
10	1,0280
11	1,0515
12	1,0464
13	1,0334
14	1,0438
15	1,0529
16	1,0543
17	1,0623
18	1,0789
19	1,1822
20	1,1776
21	1,1566
22	1,1283
23	1,0736
24	0,9866
Soma total	24

* L1=24 sem o arredondamento feito para 4 casas decimais

Os fatores iniciais do ciclo 2 (semanal) são apresentados no quadro 2.2.

Tabela 2.2: Fatores iniciais do ciclo 2

Hora (t)	Fatores do ciclo 2 na hora (t)	Hora (t)	Fatores do ciclo 2 na hora (t)	Hora (t)	Fatores do ciclo 2 na hora (t)	Hora (t)	Fatores do ciclo 2 na hora (t)
1	0,8493	43	1,1761	85	1,1094	127	0,9452
2	0,8070	44	1,1660	86	1,1268	128	0,9975
3	0,7766	45	1,1469	87	1,1439	129	1,0757
4	0,7579	46	1,1220	88	1,1451	130	1,1210
5	0,7490	47	1,0672	89	1,1538	131	1,1460
6	0,7483	48	0,9743	90	1,1513	132	1,1364
7	0,7264	49	0,8825	91	1,2324	133	1,1194
8	0,7196	50	0,8373	92	1,2204	134	1,1373
9	0,7549	51	0,8124	93	1,2002	135	1,1522
10	0,7825	52	0,8036	94	1,1841	136	1,1501
11	0,7996	53	0,8100	95	1,1256	137	1,1465
12	0,8065	54	0,8429	96	1,0309	138	1,1477
13	0,8062	55	0,9100	97	0,9302	139	1,2294
14	0,8004	56	0,9574	98	0,8821	140	1,2203
15	0,7969	57	1,0328	99	0,8567	141	1,1974
16	0,8038	58	1,0726	100	0,8484	142	1,1715
17	0,8238	59	1,0975	101	0,8532	143	1,1250
18	0,8875	60	1,0909	102	0,8847	144	1,0392
19	1,0345	61	1,0737	103	0,9471	145	0,9264
20	1,0449	62	1,0900	104	0,9976	146	0,8749
21	1,0380	63	1,1053	105	1,0739	147	0,8437
22	1,0093	64	1,1064	106	1,1163	148	0,8295
23	0,9582	65	1,1125	107	1,1432	149	0,8277
24	0,8825	66	1,1110	108	1,1351	150	0,8357

25	0,8146	67	1,1935	109	1,1191	151	0,8397
26	0,7786	68	1,1803	110	1,1367	152	0,8669
27	0,7577	69	1,1539	111	1,1535	153	0,9265
Hora (t)	Fatores do ciclo 2 na hora (t)	Hora (t)	Fatores do ciclo 2 na hora (t)	Hora (t)	Fatores do ciclo 2 na hora (t)	Hora (t)	Fatores do ciclo 2 na hora (t)
28	0,7508	70	1,1331	112	1,1552	154	0,9634
29	0,7591	71	1,0797	113	1,1608	155	0,9801
30	0,7992	72	0,9831	114	1,1586	156	0,9695
31	0,8711	73	0,9274	115	1,2386	157	0,9596
32	0,9244	74	0,8794	116	1,2251	158	0,9573
33	1,0035	75	0,8549	117	1,2057	159	0,9520
34	1,0512	76	0,8450	118	1,1811	160	0,9513
35	1,0805	77	0,8498	119	1,1271	161	0,9586
36	1,0779	78	0,8828	120	1,0385	162	1,0021
37	1,0625	79	0,9467	121	0,9429	163	1,1451
38	1,0773	80	0,9931	122	0,8911	164	1,1589
39	1,0909	81	1,0657	123	0,8638	165	1,1367
40	1,0915	82	1,1084	124	0,8534	166	1,0849
41	1,0992	83	1,1340	125	0,8570	167	1,0234
42	1,0962	84	1,1266	126	0,8867	168	0,9497

* A soma dos fatores é igual ao seu comprimento $L2 = 168$.

A programação do modelo Holt-Winters com duplos ciclos em qualquer software, já que o mesmo não se encontra pronto nos pacotes estatísticos atualmente à disposição no mercado, requer o cuidado e a atenção de que a repetição periódica dos ciclos esteja acontecendo. Isto é, como os dados horários podem formar uma série bastante grande, como no caso da série em estudo ($n = 7296$), e os vetores de fatores horários dos ciclo 1 e ciclo 2 têm respectivamente 24 e 168 elementos, que correspondem a cada hora do dia (ciclo 1) e da semana (ciclo 2), é preciso certificar-se de que a observação 169 da série, por exemplo,

está correspondendo a hora 1 na matriz de fatores 1 e 2. Da mesma forma, a observação serial 25 deve corresponder a hora 1 na matriz 1 e a hora 25 na matriz 2 e, assim sucessivamente para todo 'n' da série.

2.4

Modelagem Holt-Winters estendida: modelos com tendência amortecida

Em Taylor (2003) [16] são apresentadas simulações incorporando uma variável de amortecimento extra ao modelo Holt-Winters convencional que aparece associada à tendência aditiva. Com isso pretende-se aumentar a exatidão das previsões. Segundo o autor, os modelos sem este parâmetro adicional, ao estimar a tendência por sucessivas diferenças no nível local, tendem a superestimar a taxa de crescimento no curto prazo. Se essa taxa é positiva, pode levar o modelo a uma tendência explosiva.

Como abordado, o método de amortecimento exponencial realiza uma boa modelagem se a previsão é para curtos horizontes. Para períodos maiores é interessante “amortecer” o modelo aplicando um fator de correção (constante de amortecimento). Este fator permite melhorar as estimativas para horizontes mais distantes. Sendo este o caso, a técnica de Holt torna-se mais acurada ao incorporar o parâmetro de amortecimento cuja função é “brochar” a taxa de crescimento para que a mesma, se positiva, não cresça infinitamente, o que na prática não acontece.

Reitera-se que a modelagem que está sendo pretendida visa prever 01 passo-à-frente e 24 passos-à-frente. A técnica Holt faz *damping* só a partir do horizonte $\tau = 2$. Logo, o modelo com *damping*, para o horizonte 1, equivale ao modelo HW sem a constante de amortecimento. Porém, é válido observar o comportamento das previsões com *damping* para o horizonte de 24 passos.

A equação do modelo Holt-Winters convencional permanece a mesma com ou sem tendência amortecida. Na equação de previsão, se for desejável amortecer a tendência, é que a constante de amortecimento deverá ser incluída como mostrado a seguir.

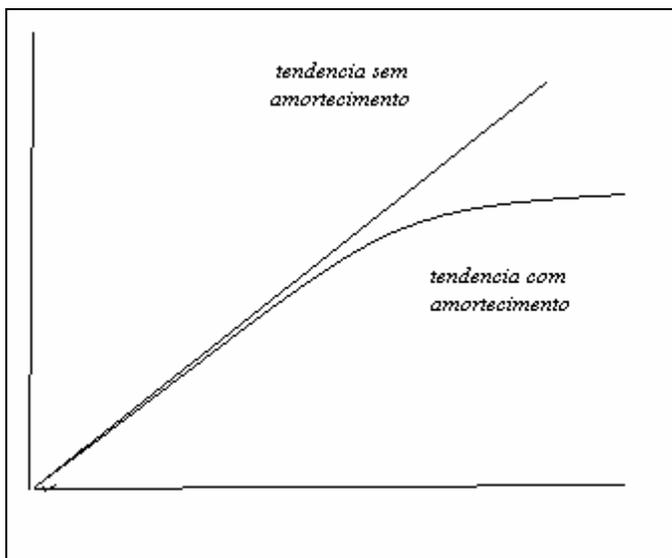
$Z_t = a_1(t) + a_2(t) + et$ [modelo Holt convencional com ou sem *damped trend*]

$\hat{Z}_t(\tau) = \hat{a}_1(t) + \hat{a}_2(t)\tau$ [equação de previsão **sem** a constante de amortecimento]

$\hat{Z}_t(\tau) = \hat{a}_1(t) + \left[\sum_{j=1}^{\tau} \phi^{j-1} \right] \hat{a}_2(t)$ [equação de previsão **com** a constante de amortecimento]

O gráfico 2.5 ilustra como se comporta a tendência amortecida e o comportamento da taxa de crescimento quando não existe a constante de amortecimento incorporada na modelagem.

Gráfico 2.5: Tendência com amortecimento e tendência sem a constante de amortecimento



Em teoria, $\phi \in [0,1]$, mas quase sempre a constante de amortecimento assume valores altos.

Para ilustrar como a variável de amortecimento ϕ atua no modelo faz-se a simulação que segue.

Simulação:

$$\hat{Z}(\tau) = \hat{a}_1 + \left[\sum_{j=1}^{\tau} \varphi^{j-1} \right] \hat{a}_2(t) \quad \text{[equação de previsão com a constante de amortecimento]}$$

Tabela 2.3: Simulação do impacto do fator de correção no modelo HW

(τ) horizonte	a- $\hat{Z}_t(\tau)$ sem correção	b- $\hat{Z}_t(\tau)$ com correção
1	$\hat{a}_1 + \hat{a}_2(t)$	$\hat{a}_1 + \hat{a}_2(t)$
2	$\hat{a}_1 + 2\hat{a}_2(t)$	$\hat{a}_1 + (\varphi^{1-1} + \varphi^{2-1}) \hat{a}_2(t) = \hat{a}_1(t) + (1 + \varphi) \hat{a}_2(t)$
3	$\hat{a}_1 + 3\hat{a}_2(t)$	$\hat{a}_1 + (\varphi^{1-1} + \varphi^{2-1} + \varphi^{3-1}) \hat{a}_2(t) = \hat{a}_1(t) + (1 + \varphi + \varphi^2) \hat{a}_2(t)$
4	.	.
.	.	.
.	.	.
.	.	.

Supondo $\varphi = 0.9$ (geralmente um valor alto), pela tabela 2.3, para o horizonte 2 em 'b' (com constante de correção) o coeficiente associado à tendência (\hat{a}_2) é igual a 1.9 e em 'a' (sem constante de correção) o coeficiente junto à tendência assume valor igual a 2. Da mesma forma, para o horizonte 3, tem-se em 'b' (com constante de correção) um coeficiente à tendência (\hat{a}_2) igual à 2.7 e para 'a' (com constante de correção) um coeficiente igual à 3.

Sabendo que a função de previsão é a soma do nível com uma projeção de crescimento, como ($1.9 < 2$ e $2.7 < 3$), verifica-se que o fator de correção ' φ ' está amortecendo a série.

- Quando $\varphi \in [0,1]$, a tendência é amortecida e as previsões convergem assintoticamente para: $a_1(t) + \varphi / 1 - \varphi$.
- Se $\varphi > 1$, então, a função de previsão tem um comportamento exponencial.

As equações de atualização para o modelo HW com *damped trend* na equação de previsão são dadas por:

$$\hat{a}_1 = \alpha Z_t + (1 - \alpha) [\hat{a}_1(t-1) + \varphi \hat{a}_2(t-1)] \quad (1)$$

$$\hat{a}_2 = \beta [\hat{a}_1(t) - \hat{a}_1(t-1)] + (1 - \beta) \varphi \hat{a}_2(t-1) \quad (2)$$

Para o modelo HW alternativo com dois ciclos e a constante de amortecimento:

Equação do modelo:

$$Z_t = [a_1(t) + a_2(t)] * c_1 t * c_2 t + \epsilon_t$$

Equação de previsão:

$$\hat{Z}_t(\tau) = \hat{a}_1(t) + \sum_{j=1}^{\tau} [\hat{a}_2(t) * \hat{c}_{1(t)h_1}(t+\tau) * \hat{c}_{2(t)h_2}(t+\tau)]$$

Com *damped trend* as equações de atualização para o nível e a tendência são as apresentadas em (1) e (2). As equações de atualização para os ciclos 1 e 2 são as mesmas dispostas na seção 2.3.

Ciclo 1: $\hat{c}_{1h_1}(t)$; constante ‘ γ ’

$$\hat{c}_{1h_1}(t) = \gamma * [Z_t / \hat{b}_1(t) * \hat{c}_{2h_2}(t-1)] + (1 - \gamma) * \hat{c}_{1h_1}(t-1)$$

Ciclo 2: $c_{2h_2}(t)$; constante ‘ δ ’

$$\hat{c}_{2h_2}(t) = \delta * [Z_t / \hat{b}_1(t) * \hat{c}_{1h_1}(t-1)] + (1 - \delta) * \hat{c}_{2h_2}(t-1)$$

Taylor (2003) [16], levanta a discussão acerca das vantagens e desvantagens de modelar a tendência aditivamente, como ocorre com o modelo Holt padrão que está sendo usado no estudo, ou de forma multiplicativa como sugerida por Pegels (1969) *apud in* Taylor (2003) [16]. Sem a preocupação de entrar em maiores detalhes desta questão, para Taylor os modelos aditivos geram estimativas mais eficientes em se tratando de processos automáticos de previsão e, além disso, as formas aditivas são mais comuns na vida real. Por outro lado, os modelos, como apresentados por Pegels, exponenciais, possibilitam trabalhar com dados de tendência agregados o que não é possível de ser feito com modelagens aditivas.

Um exemplo colocado por Taylor para esclarecer essa declaração é imaginar a entrada de uma nova marca no mercado de um determinado produto. Esta nova marca irá repartir o mercado com as demais existentes e podem ser classificadas juntas. As previsões de vendas para a nova marca, inicialmente, serão feitas sem dados. Desta maneira, seria útil recorrer às informações disponíveis acerca dos dados agregados do produto considerando todas as empresas.

É claro que o volume de consumo da nova marca pode ser totalmente diferente do consumo das outras, contudo, a tendência de mercado para todas juntas é a mesma. A formulação proposta por Pegels permite então inicializar a modelagem com a tendência apresentada pelos dados no agregado e, assim, realizar a previsão. Esse procedimento não é possível em se tratando de modelos aditivos.

Para informação do leitor segue o modelo, a equação de previsão e as equações de atualização dos parâmetros pela modelagem Pegels:

Modelo de Pegels:

$$Z_t = a_1 * (a_2)^t + e_t \quad (\text{modelo com tendência exponencial})$$

Equação de previsão do modelo proposto por Pegels:

$$\hat{Z}_t(\tau) = \hat{a}_1 * [\hat{a}_2(t)]^\tau$$

Equações de atualização:

$$\hat{a}_1(t) = \alpha * (Z_t) + (1 - \alpha) * [\hat{a}_1(t-1) * \hat{a}_2(t-1)]$$

$$\hat{a}_2(t) = \beta * [\hat{a}_1(t) / \hat{a}_1(t-1)] + (1 - \beta) * [\hat{a}_2(t-1)]$$

Para o modelo assumindo um amortecimento na tendência utiliza-se o mesmo programa computacional feito para otimizar os parâmetros e estimar o modelo no caso anterior (sem *damping*). Na programação para efetuar a previsão é que a constante de amortecimento deve ser incluída na equação de previsão da forma como mostrada nesta seção.

2.5

Modelagem Holt-Winters estendida: modelos com erros estruturados

O modelo Holt-Winters estendido considerando erros estruturados foi apresentado por Taylor, Menezes e Mcsharry, em Simpósio acontecido em Sidney, na Austrália, em julho de 2004. O trabalho exposto pretendeu comparar métodos univariados para a previsão 01 dia- à- frente da demanda elétrica de curto prazo. (Taylor & Menezes & Mcsharry (2004)) [17].

Segundo os seus autores, a motivação da pesquisa se deve a necessidade de aumentar a acurácia dos modelos preditivos usados no planejamento energético, essencial para o controle e programação da capacidade do sistema fornecedor.

Com esta finalidade foram estudadas 4 técnicas univariadas:

- 1) Box & Jenkins SARIMA adotando dois ciclos multiplicativos: um diário e um semanal. Este método será melhor investigado no próximo capítulo.
- 2) MAE com duplos ciclos diário e semanal. A razão deste procedimento estar entre os quatro selecionados é devido a robustez de sua metodologia que o torna atraente para fazer a previsão *on-line* do consumo.

- 3) Rede neural. Previsões obtidas com dados similares aos usados no trabalho a partir da implementação de uma rede neural já havia sido feita por Darbellay and Slama (2000) [6] com bons resultados, sugerindo que este método deveria estar entre os quatros a serem pesquisados.
- 4) Análise das componentes principais. Esta técnica reduz os dados e captura a essência da similaridade encontrada no dia-a-dia. Deste modo, 'n' semanas de dados serviram para estimar as componentes e então 'n' componentes principais foram escolhidas para modelar o comportamento da série.

Estes quatros procedimentos mencionados foram aplicados em duas situações:

- 1) demanda energética horária no estado do Rio de Janeiro.
- 2) consumo de energia de 30 min em 30 min da Inglaterra e país de Gales.

Para o Rio contou-se com 3.360 observações para estimar o modelo e 1.680 dados para a avaliação do seu poder de generalização. Para o segundo caso a quantidade é dobrada, 6.720 observações para efetuar a modelagem e 3.360 dados a fim de verificar a capacidade dos modelos. Estes números correspondem a 20 semanas de dados para modelagem e 10 semanas para averiguação e validação.

O comprimento dos fatores dos ciclos diário e semanal são respectivamente:

caso1: $L1 = 24$ e $L2 = 168$ (dados horários);

caso 2: $L1 = 48$ e $L2 = 336$ (dados de 30 min em 30 min);

Das quatro técnicas avaliadas, o método Holt-Winters levando em conta os ciclos diário e semanal de forma multiplicativa foi o que teve melhor desempenho e o menor MAPE. Este fato foi notado principalmente para o caso1 (dados brasileiros). Este modelo, bem como suas equações de atualização, estão colocados na seção 2.3 deste capítulo.

Ainda, de acordo com os pesquisadores, a exatidão das estimativas melhora quando a modelagem é ajustada para incorporar a autocorrelação existente entre os termos de erro. Logo, na pesquisa exposta em Sidney, o componente de erro do modelo HW com dois ciclos foi modelado como um AR (1):

$$e_t = \lambda e_{t-1} + \varepsilon_t$$

Nesta situação o modelo pode ser escrito do seguinte modo:

$$Z_t = (b_1 + b_2 t) * c_1 t * c_2 t - \lambda e_{t-1} + \varepsilon_t$$

As previsões k-passos à frente na origem τ são feitas com um termo a mais: $\lambda^\tau \varepsilon_t$.

Por conseqüência, as previsões desse modelo serão distintas daquelas fornecidas pelo modelo que não incorpora o termo de erro estruturado.

A equação de previsão do modelo com erros estruturados é dada por:

$$\hat{Z}_t(\tau) = [(\hat{b}_1(t) + \tau \hat{b}_2(t)) * \hat{c}_1^{(t)} h_1(t + \tau) * \hat{c}_2^{(t)} h_2(t + \tau) - \lambda^\tau \varepsilon_t$$

Diante dos argumentos levantados em Sidney por Taylor, Menezes e Mcsharry, e da pretensão do presente estudo, considera-se válido averiguar também os resultados da modelagem HW estendida com correlação serial dos erros amostrais.

2.6

Comentários finais

Com base no modelo HW tradicional tão comentado e debatido, este capítulo apresentou uma abordagem alternativa para adequá-lo a tratar problemas que, originalmente, da forma como foi matematizado, não seria capaz. Essa nova abordagem, aqui exaustivamente discutida, embora com algumas publicações existentes na literatura disponível, vem recebendo pouca atenção e, talvez, por

isso, o potencial destas modelagens aperfeiçoadas, especialmente para HW que são modelos robustos, ainda não seja totalmente conhecido.

Dessa forma, este capítulo é um estímulo para que outras pesquisas sejam realizadas assumindo as três formulações “estendidas” apresentadas em distintas séries e com variados ciclos. Estes estudos devem ser incentivados em virtude dos primeiros trabalhos feitos com esses modelos alternativos terem apresentado bons resultados como foi visto em Taylor & Menezes & Mcsharry (2004) [17] e em Esteves (2004) [7] e, portanto, merecendo mais interesse por parte da comunidade acadêmica.

Neste momento em que o mundo passa por uma febre decorrente da era computacional, é normal o deslumbramento com modelagens que usam essencialmente a computação intensiva, tais como: redes neurais, lógica *fuzzy* e algoritmos genéticos. Porém, os métodos “*benchmark*”, como Holt-Winters, não podem ser desprezados e esquecidos. Ao contrário disso, a proposta levantada neste capítulo e nessa dissertação é buscar formas de aperfeiçoá-los e deixá-los em condições de competir com as novas metodologias que vêm naturalmente surgindo com o avanço tecnológico.