

4 Modelo Híbrido de Previsão de Reativo

A modelagem proposta para previsão de potência reativa, consiste em duas etapas :

- ✓ 1ª Etapa – Classificação dos dados de carga reativa e ativa por subestação, através do método não supervisionado Mapa Auto-Organizável de Kohonen com a finalidade de reduzir dimensionalidade e sazonalidade;
- ✓ 2ª Etapa – Aplicação do Modelo Dinâmico Iterativo em cada grupo obtido na etapa anterior.

Cada etapa do processo de modelagem de previsão de carga reativa é apresentada a seguir.

4.1. Etapa I - Classificação dos Dados

Para evitar problemas com a sazonalidade dos dados, é necessário reduzir a dimensionalidade da amostra com a finalidade de aumentar a precisão dos resultados.

No presente trabalho de pesquisa, optou-se pela implementação de um tipo de rede neural, os Mapas Auto-Organizáveis de Kohonen (SOM), visto que este método, dentre outros objetivos, pode ser utilizado como uma valiosa ferramenta num problema clássico em estatística que diz respeito à análise de agrupamentos. Além disso, esta rede apresenta resultados satisfatórios em aplicações de séries temporais (Debeljak et al, 2001; Aguilera et al, 2001; Jain et al, 2000; Zanini, 2004). Outros métodos foram testados, como o método não-hierárquico conhecido como *K-means*, porém com o SOM os resultados foram mais consistentes.

Os mapas auto-organizáveis de Kohonen caracterizam-se por projeções de um espaço de entrada em um espaço de menor dimensão e que a topologia no espaço de saída reflete características semelhantes no espaço original, ou seja,

entradas com características parecidas serão mapeadas em regiões próximas no espaço de saída.

Neste modelo de Kohonen não se usa um conjunto de exemplos previamente conhecidos. Uma medida da qualidade da representação do ambiente pela Rede é estabelecida, e os parâmetros independentes são modificados de modo a otimizar esta medida. É usada uma regra de competição, uma primeira camada recebe os dados enquanto as unidades de uma segunda camada disputam o privilégio de responder aos estímulos.

Para inicializar o algoritmo do Modelo de Kohonen, escolhem-se pesos de modo aleatório, neste caso, foram escolhidos os mínimos e máximos de cada variável.

Existem diversas arquiteturas e algoritmos de treinamento distintos para as redes neurais artificiais (anexo). Esta seção tem por objetivo descrever apenas as redes de Kohonen (auto-organizadas), que foram utilizadas nesta tese.

As redes que possuem aprendizado competitivo ou auto-organizado (Seção 8.2.2) são chamadas redes competitivas ou auto-organizadas, onde os neurônios de saída competem entre si para estarem ativos ou não. Apenas um neurônio de saída, ou um por grupo, é ativado a cada iteração. As redes auto-organizadas de Kohonen (2000) são caracterizadas pela formação de um mapa topográfico dos padrões de entrada e, por isso, são denominadas de mapas auto-organizáveis de Kohonen (SOM – *self-organizing maps*) (Kohonen, 1997a; Kohonen, 1990a; Kohonen, 1988a).

No SOM, a localização espacial dos neurônios auto-organizados vai indicar características intrínsecas aos padrões de entrada. Seu principal objetivo é a transformação adaptativa e ordenada de um conjunto de dados de entrada em um mapa (*grid*) uni- ou bidimensional de saída. Os dados de entrada que apresentam semelhanças entre si são agrupados em regiões do mapa de saída, formando classes ou agrupamentos denominados *clusters*.

Durante o processo de auto-organização do mapa, a unidade do *cluster* cujo vetor de pesos mais se aproxima do vetor dos padrões de entrada é escolhida como sendo a vencedora. A unidade vencedora e suas unidades vizinhas têm seus pesos atualizados. Além disso, estas redes possuem parâmetros variáveis com o tempo, como a taxa de aprendizagem a , e um raio de vizinhança N_R que indica quais unidades serão atualizadas simultaneamente.

A Figura 3 apresenta arquiteturas típicas de um SOM, considerando configurações de vizinhança unidimensional e bidimensional, embora dimensões mais elevadas possam ser consideradas. Além disso, dada à dimensão, a quantidade de unidades ou neurônios de saída pode ser arbitrada e mantida fixa, ou então definida automaticamente pelo algoritmo de treinamento (Zanini, 2004; Fritzke, 1994; Cho, 1997; de Castro & al., 1999).

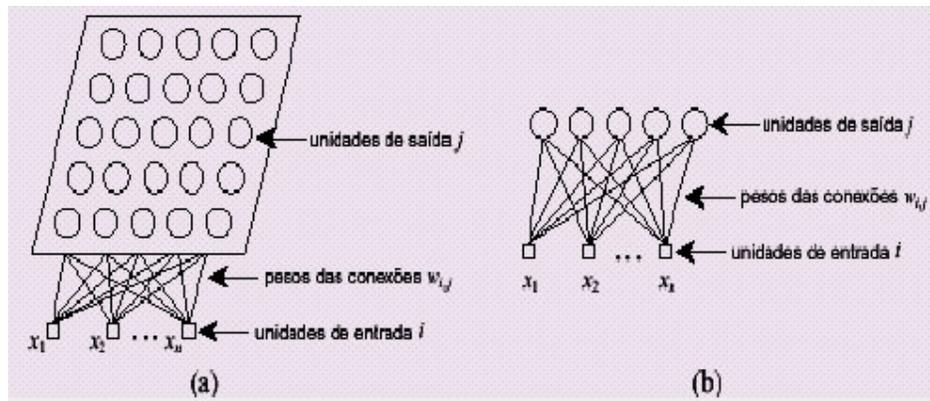


Figura 3 Arquiteturas típicas de um SOM. a) Bidimensional; b) Unidimensional

O Mapa Auto-Organizável produz um ajuste seletivo dos neurônios criando um mapa topográfico dos padrões de entrada. Ainda que esta técnica não tenha sido originalmente concebida para classificação de padrões ou segmentação, é possível tirar-se grande proveito das suas propriedades de auto-organização para tais aplicações.

O objetivo básico dos Mapas de Kohonen é agrupar m elementos de um conjunto de padrões de entrada, X , em J neurônios, ou seja, projeta-se o espaço de entrada em um espaço de menor dimensão (normalmente bidimensional). A “comunicação” entre os ambientes de entrada (subespaço natural dos padrões de entrada) e o de saída do algoritmo (malha) é feita por protótipos, ou seja, a cada neurônio no espaço de saída corresponde um vetor-peso que pode ser visto como um protótipo de características do espaço original.

O algoritmo proposto por Kohonen pode ser dividido em quatro etapas básicas: (i) Cálculo das distâncias aos J protótipos de um elemento sorteado X_k do conjunto de padrões de entrada; (ii) Comparação dos valores das J distâncias e reconhecimento do menor, ou seja, aponta-se o protótipo mais próximo de X_k no subespaço de saída, o que determina o neurônio vencedor; (iii) Ativação, por uma

rede interativa, simultaneamente, do neurônio vencedor e da sua vizinhança; (iv) Diminuição gradativa, através de processo adaptativo, da distância do neurônio vencedor e da respectiva vizinhança a X_k .

A seguir tem-se uma breve descrição do algoritmo de Kohonen:

Passo 1: Escolhe-se de modo aleatório os pesos iniciais $W_j(0)$ para cada neurônio $j=1, 2, \dots, N$ tal que $W_j(0) \neq W_i(0) \forall j \neq i$.

Passo 2: Apresenta-se um vetor de entrada x à rede.

Passo 3: Acha-se o neurônio vencedor $i(x)$ tal que:

$$i(x) = \arg \min_j \|x(n) - W_j\| \quad j=1, 2, \dots, N$$

O neurônio vencedor correspondente ao vetor de peso que está mais próximo ao vetor de entrada corrente.

Passo 4: Atualizam-se os pesos (protótipos) de acordo com:

$$W_j(n+1) = \begin{cases} W_j + \eta(n)[x(n) - W_j(n)] & j \in \Lambda_{i(x)}(n) \\ W_j(n) & \text{caso contrário} \end{cases}$$

Os parâmetros $\eta(n)$ e $\Lambda(n)$ podem variar ao longo do treinamento. O parâmetro de aprendizado $\eta(n)$, tipicamente, deve começar com valores perto de 1 e decair gradualmente até atingir 0.1 por volta da milésima iteração. Posteriormente, $\eta(n)$ deve ser mantido em valores pequenos, próximos a 0.01, por muitas iterações para que seja feito o ajuste fino. O parâmetro de vizinhança $\Lambda(n)$ deve começar atingindo todos os neurônios da rede e gradualmente diminuir a vizinhança para incluir muito poucos vizinhos na fase de ajuste.

Passo 5: Interromper o processo quando a rede atingir estabilização, ou seja, quando não forem detectadas alterações significativas nos protótipos.

4.2.

Etapa II – Modelo Dinâmico Iterativo

Ao tentarmos modelar uma série temporal através de um modelo de regressão, a hipótese de independência dos ruídos não é realista, e os resultados e testes usados nos modelos de regressão não são válidos. Por exemplo, algumas das conseqüências da autocorrelação dos resíduos são:

- a) Os estimadores usuais por mínimos quadrados são ainda não tendenciosos, mas não têm variância mínima;
- b) Os estimadores de variância dos erros padrões dos coeficientes da regressão são subestimados, o que levaria à conclusão de que os estimadores são mais precisos do que na realidade;
- c) Os intervalos de confiança para os parâmetros da regressão e os testes de hipóteses relacionados a estes intervalos perdem a validade.

Em particular, a hipótese de independência dos erros não é realista no contexto de séries temporais, e os modelos de defasagem distribuída autorregressivo (ADL) estendem os modelos usuais de regressão ao levantarem esta restrição.

Os modelos ADL combinam a dinâmica de séries temporais e o efeito de variáveis explicativas. A estimação de parâmetros neste modelo é feita através de mínimos quadrados ordinários, a exemplo dos modelos de regressão usual. Entretanto, a estimação em modelos de regressão dinâmica envolve um procedimento iterativo com vários estágios (Beenstock et al, 1999; Zanini et al, 2001; Fatai et al, 2003; Tang, 2003),

A variável dependente é explicada por seus valores defasados e pelos valores atuais e passados de variáveis causais ou exógenas ou independentes (Barros & Souza, 1995).

Os modelos de defasagens distribuídos autorregressivo $ADL(p, p_1, \dots, p_j)$, também são conhecidos como modelos de regressão dinâmica. Os p e p_j indicam a ordem de defasagem da(s) variável(is) dependente(s) e independente(s), respectivamente. A representação geral do modelo é (Song et al, 2003; Kiviet & Dufour, 1997):

$$y_t = \alpha + \sum_{j=1}^k \sum_{i=1}^{p_j} \beta_{ji} x_{jt-i} + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t \quad (4.1)$$

onde:

y_t : Variável dependente em t

y_{t-1} : Variável dependente em $t-1$

x_{jt-i} : A j -ésima variável independente em $t-i$, sendo $i=\{1, \dots, p_j\}$ e $j=\{1, \dots, p\}$

β_{ji} : Coeficiente da j -ésima variável independente em $t-i$

ϕ_i : Coeficiente da variável dependente em $t-i$

α : constante

ε_t : ruído aleatório associado ao modelo, onde se supõem que são independentes e identicamente distribuídos com densidade $N(0, \sigma^2)$.

A equação (4.1) pode ser escrita da seguinte forma:

$$\phi(B)Y_t = \alpha + \beta(B)X_t + \varepsilon_t \quad (4.2)$$

onde:

Y_t : Variável dependente no instante t

X_t : Variável independente no instante t

$\beta(B)$: $1 - \beta_1 B \dots \beta_p B^p$ polinômio AR(p_j) das variáveis independente

$\phi(B)$: $1 - \phi_1 B \dots \phi_p B^p$ polinômio AR(p)

α : constante

ε_t : ruído aleatório associado ao modelo, onde se supõem que são independentes e identicamente distribuídos com densidade $N(0, \sigma^2)$.

Problemas que podem surgir com o uso dos modelos de Defasagem Distribuídas Autorregressivo são as perdas de informação devido ao alto grau de liberdade do modelo (diferença entre o número de dados e o número de variáveis independentes), e a multicolinearidade entre as variáveis independentes. Muito embora, segundo Gujarati (1996), na finalidade de previsão a multicolinearidade não apresenta grande perigo, pois não afeta diretamente os resultados finais. (Song

et al, 2003; Bentzen & Engsted 2001; Garci-Cerrutti, 2000; Duarte & Holden, 2003; Albertson & Aylen, 2003).

Outro problema de grande importância é a dependência temporal dos valores sucessivos dos resíduos, eles são correlacionados entre si. Em termos formais e considerando o modelo 4.1 a autocorrelação serial implica $E(u_i u_j) \neq 0$ para $i \neq j$ (Matos, 1997; Hill et al, 1999). Esta pode ser causada por situações como:

- Omissão de variável explicativa (erro de especificação);
- Má especificação da forma matemática, a qual pode se indicar uma forma diferente da linear;
- Má especificação do verdadeiro termo aleatório, ou seja, efeitos que se propagam ao longo do tempo;
- Ajustes imperfeitos de séries estatísticas causados, por exemplo, por interpolações ou extrapolação do dados, e/ou causados pelas suavizações dos dados.

Quando os resíduos são autocorrelacionados, as estimativas de mínimos quadrados ordinários dos parâmetros não são eficientes, isto é, não apresentam variância mínima, além de seu erro-padrão ser viesado, o que conduz a testes e intervalos de confiança incorretos. Se a autocorrelação for positiva, os erros-padrão serão subestimados e, conseqüentemente, os valores da estatística t , superestimados. Se a autocorrelação for negativa, os erros-padrão serão superestimados e o valor de t , subestimado. Portanto, a autocorrelação positiva é a mais danosa, porque existirá, no caso do teste t , o risco de rejeitar-se a hipótese nula de ausência de efeito, quando se deveria aceita-la.

A autocorrelação é positiva ($r > 0$), quando os resíduos são diretamente relacionados entre si. Se $r < 0$, a autocorrelação é negativa. Tal autocorrelação é dita de primeira ordem, porque o resíduo u_t é relacionado com u_{t-1} , ou seja, o resíduo em t é função do resíduo no período imediatamente anterior ($t-1$), isto é:

$$u_t = ru_{t-1} + \varepsilon_t, \quad (4.3)$$

Na presença de correlação serial, os estimadores de mínimos quadrados ordinários são ineficientes, por isso, é essencial procurar medidas corretivas. Porém, depende do conhecimento que se tem sobre a natureza da interdependência das perturbações, ou seja, do conhecimento que se tem sobre o coeficiente de autocorrelação r da equação 4.3. Na prática, raramente se conhece este coeficiente. Uma alternativa para estimá-lo é o bastante usado método de Cochrane-Orcutt, que emprega os resíduos estimados \hat{u}_t para obter informações sobre o coeficiente r desconhecido.

Para explicitar o método de Cochrane-Orcutt, considere o modelo simplificado de duas variáveis (Barros & Souza, 1995):

$$\begin{cases} Y_t = \beta_1 + \beta_2 X_t + u_t \\ R(B)u_t = \varepsilon_t \end{cases} \quad (4.4)$$

onde: $R(B) = 1 - r_1 B$, isto é, um AR(1)

Se (4.4) for válida no período t , também será válida no período $t-1$. Daí,

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \quad (4.5)$$

Multiplicando ambos os lados de (4.5) por r , obtêm-se:

$$rY_{t-1} = r\beta_1 + r\beta_2 X_{t-1} + ru_{t-1} \quad (4.6)$$

Subtraindo (4.6) de (4.4), tem-se:

$$\begin{aligned} (Y_t - rY_{t-1}) &= \beta_1(1-r) + \beta_2 X_t - r\beta_2 X_{t-1} + (u_t - ru_{t-1}) \\ &= \beta_1(1-r) + \beta_2(X_t - rX_{t-1}) + \varepsilon_t \end{aligned} \quad (4.7)$$

em que no último passo foi feito uso de (4.3) segunda equação. Pode-se expressar (4.7) como,

$$Y_t^* = \beta_1^* - \beta_2 X_t^* + \varepsilon_t$$

sendo: $\beta_1^* = \beta_1(1-r)$; $Y_t^* = R(B)Y_t$ e $X_t^* = R(B).X_t$

$R(\beta)$ é conhecido como fator comum

O procedimento de estimação é sequencial:

1. Faz-se, inicialmente, $r_i = 0$; $i = 1, 2, \dots, n$
2. Estima-se β 's por mínimos quadrados ordinários
3. Calcula-se u_t
4. Estima-se $R(B)$, ou melhor, os coeficientes r' por mínimos quadrados ordinários na equação $R(B).u_t = \varepsilon_t$
5. Calculam-se os valores transformados Y_t^* e X_t^*
6. Re-estimam-se β 's por mínimos quadrados ordinários
7. Como, a priori não se sabe se os r 's estimados são as “melhores” estimativas de $R(B)$, então se substituem os valores dos coeficientes β 's obtidos no passo 6 na regressão original (5.4) e calculam-se os novos resíduos.
8. Repetir passos 4 a 7 até a convergência de r 's, β 's ser alcançada.

Como ε_t satisfaz todas as hipóteses dos mínimos quadrados ordinários (independente e identicamente distribuído), pode-se passar a aplica-los às variáveis transformadas (passo 6) e obter estimadores com todas as propriedades ótimas, ou seja, melhor estimador não-enviesado linear. De fato, equivale a usar os mínimos quadrados generalizados apresentados a seguir:

$$\hat{\beta}_2 = \frac{(1-r^2)x_1y_1 + \sum_{t=2}^n (x_t - rx_{t-1})(y_t - ry_{t-1})}{(1-r^2)x_1^2 + \sum_{t=2}^n (x_t - rx_{t-1})^2} \quad (4.8)$$

Com variância dada por:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{t=2}^n (x_t - rx_{t-1})^2} + D \quad (4.9)$$

em que D é um fator de correção que, na prática, pode ser desconsiderado.

Contudo, esta correção se refere somente a autocorrelação serial, outros problemas como a heterocedasticidade (variância não constante) dos dados e a influência de valores extremos continuam presentes no modelo impedindo o bom desempenho da estimação.

Todavia, para lidar com estes problemas, a estimação dos coeficientes por Mínimos Quadrados Reponderados Iterativamente tem obtido bons resultados (Mbamalu & El-Hawary, 1993; El-Hawary & Mbamalu, 1996; Fox, 2002).

O método de estimação por regressão robusta é sugerido para o mercado financeiro brasileiro, dentre outros autores, por Mendes & Duarte (1998), pois nesse mercado as séries temporais apresentam valores extremos em quantidade substancial, capazes de exercer influências indesejáveis sobre as estimativas obtidas com procedimentos usuais. Ademais, as características dessas séries violam hipóteses fundamentais de um modelo de regressão linear estimado por mínimos quadrados ordinários, apresentando forte heteroscedasticidade (variância não constante), desvios de normalidade e correlações seriais que podem acontecer em ordens superiores à defasagem 1 e portanto não são captadas pelo teste de Durbin-Watson (Durbin, 1970).

A limitação da metodologia dos mínimos quadrados conduziu a diversas abordagens alternativas. As técnicas de estimação robusta constituem uma abordagem à estimação que não depende de uma distribuição em particular. O objetivo da estimação robusta é, pois, o de procurar estimadores (i) eficientes sob um certo modelo e (ii) de modo que, pequenas alterações na distribuição da amostra produzam pequenas alterações nas estimativas. Neste sentido, outro conceito diretamente relacionado com a robustez é a resistência.

Na estimação robusta, as estimativas são obtidas no fim de um processo iterativo, devido ao fato dos estimadores não terem uma expressão analítica explícita. Em vez disso está associado a cada um deles um algoritmo para a obtenção da estimativa. Este algoritmo pode ser conduzido baseado no conceito dos Mínimos Quadrados Reponderados Iterativamente (Holland & Welsch, 1977; El-Hawary & Mbamalu, 1996 e Huber, 1981).

Considerando o problema de estimação convencional dos parâmetros envolvendo a equação:

$$Y = \beta X + e_j \quad (4.10)$$

onde Y é um vetor de observações $n \times 1$, X é uma matriz $n \times p$ de coeficientes conhecidos, β é um vetor de parâmetros desconhecidos $p \times 1$, e e_i é um vetor de erros aleatórios $n \times 1$. Um estimador robusto para β , denotado por $\hat{\beta}$, define-se como a solução que minimiza a função objetivo (Birkes, 1993).

$$\sum_{i=1}^n \rho\left(\frac{e_i}{\hat{\sigma}}\right) = \sum_{i=1}^n \rho\left(\frac{Y_i - X_i\beta}{\hat{\sigma}}\right) \quad (4.11)$$

onde $\hat{\sigma}$ é um estimador robusto do parâmetro de escala do ruído, σ , e é introduzido para assegurar no problema de minimização a invariância em relação à escala. E $\rho(u)$ é uma função com valores reais não constantes e tem por objetivo limitar no processo de estimação a influência dos resíduos e_i elevados. Uma função $\rho(u)$ razoável tem as seguintes propriedades (Fox, 2002):

- $\rho(e) \geq 0$
- $\rho(0) = 0$
- $\rho(e) = \rho(-e)$ função par
- $\rho(e_i) \geq \rho(e_i')$ para $|e_i| > |e_i'|$ função crescente

Por exemplo, para estimação Mínimos Quadrados Ordinários, $\rho(e_i) = e_i^2$.

Minimizar a equação (4.11) requer que a derivada de β seja zero:

$$\begin{aligned} \frac{\partial}{\partial \beta} \left\{ \sum_{i=1}^n \rho\left[\frac{(Y_i - X_i\beta)}{\hat{\sigma}}\right] \right\} &= 0 \\ \sum_{i=1}^n \rho' \left[\frac{(Y_i - X_i\beta)}{\hat{\sigma}} \right] X_i' &= 0 \end{aligned} \quad (4.12)$$

Definindo-se uma função ψ como:

$$\psi = \rho'$$

onde (4.13)

$$\sum \psi \left[\frac{Y_i - X_i' \beta}{\hat{\sigma}} \right] X_i' = 0$$

Então, a equação 4.13 pode ser escrita como:

$$\sum_{t=1}^n \psi \left(\frac{e_t}{\hat{\sigma}} \right) X_t' = \sum_{t=1}^n \frac{e_t}{\hat{\sigma}} \frac{\psi(e_t / \hat{\sigma})}{(e_t / \hat{\sigma})} X_t' = 0 \quad (4.14)$$

Os estimadores robustos podem ser vistos como estimadores de mínimos quadrados ponderados por $w_t = \frac{\psi(e_t / \hat{\sigma})}{(e_t / \hat{\sigma})}$. No entanto, as ponderações w_t dependem dos resíduos, os resíduos dependem dos coeficientes estimados e os coeficientes estimados dependem dos pesos. Desse modo, uma solução iterativa (conhecida como mínimos quadrados reponderados iterativamente) será utilizada para estimação:

1. Selecionar inicialmente estimativas $b^{(0)}$, tais como as estimativas por mínimos quadrados.
2. A cada iteração t , calcula-se os resíduos $e_i^{(t-1)}$ e os pesos associados

$$W_t^{(t-1)} = w[e_i^{(t-1)}]$$
3. Calcular as novas estimativas de mínimos quadrados ponderados

$$b^{(t)} = \left[X' W^{(t-1)} X \right]^{-1} X' W^{(t-1)} y$$

onde X é a matriz modelo com x_i' como as i -ésimas linhas, e $W^{(t-1)} = \text{diag}\{w_t^{(t-1)}\}$ é a matriz peso.

4. Repetir os passos 2 e 3 até a convergência dos coeficientes.

A figura 4 compara as funções objetivo, as funções ψ correspondentes e as funções peso para três estimadores robustos: a família dos estimadores Mínimos Quadrados; a dos estimadores Huber; e a dos estimadores Tukey biquadrado (or

biweight). As funções objetiva e peso para estes três estimadores também são apresentadas na tabela 1.

Na figura 4 observa-se que, as funções objetivo dos estimadores por Mínimos Quadrados e Huber crescem quando os resíduos assumem valores maiores que zero, sendo que a função objetivo Mínimos Quadrados cresce mais rapidamente. Em contrapartida, a função objetivo do estimador biquadrado se torna constante quando $|e| > 4$. Os pesos dos mínimos quadrados são constantes para qualquer resíduo; os pesos do estimador de Huber são constante para $|e| < 2$ e os pesos do estimador biquadrado são constantes para $|e| > 4$.

Os estimadores Huber e Biquadrado comparam o módulo dos resíduos a uma constante k conhecida como de *tuning constant* usada para controlar a resistência e eficiência desses estimadores com uma confiança de 95% quando os erros são normais. Um menor valor de k produz maior resistência à valores discrepantes (*outliers*). Em particular, para o estimador de Huber, $k=1.345\sigma$, e para o estimador biquadrado $k=4.685\sigma$ (onde σ é o desvio padrão dos erros) produz.

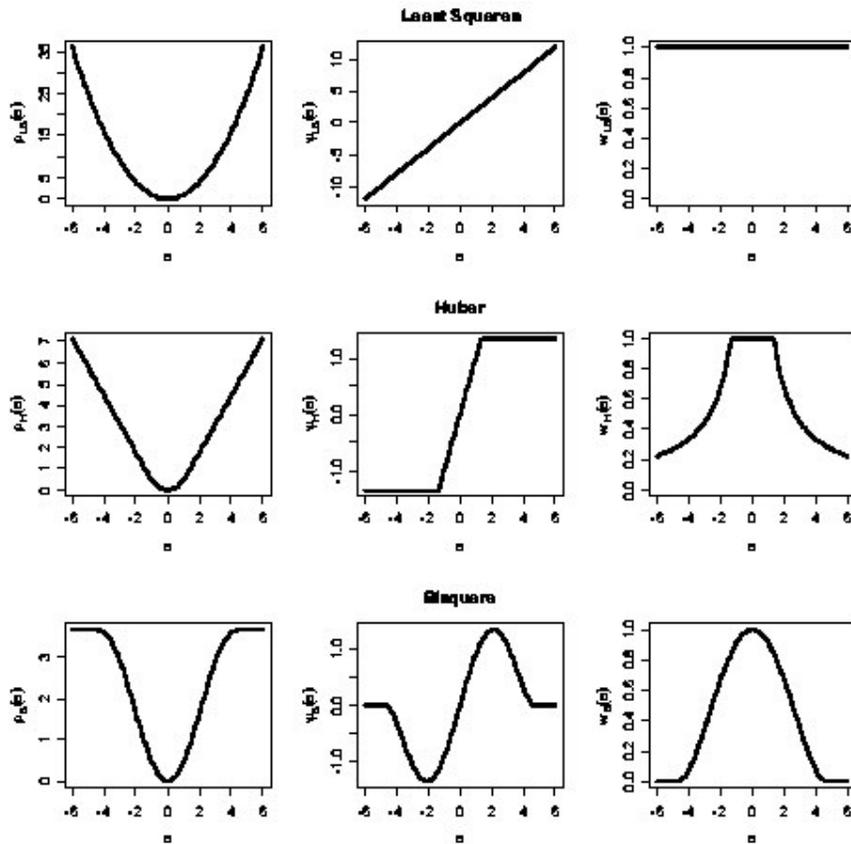


Figura 4– Função objetivo, Função ψ ; e Função peso para os estimadores mínimos quadrados, Huber, e biquadrado, respectivamente.

Tabela 1- Funções objetivo e peso para os estimadores mínimos quadrados, Huber e biquadrado

Method	Objective Function	Weight Function
Least-Squares	$\rho_{LS}(e) = e^2$	$w_{LS}(e) = 1$
Huber	$\rho_H(e) = \begin{cases} \frac{1}{2}e^2 & \text{for } e \leq k \\ k e - \frac{1}{2}k^2 & \text{for } e > k \end{cases}$	$w_H(e) = \begin{cases} 1 & \text{for } e \leq k \\ k/ e & \text{for } e > k \end{cases}$
Bisquare	$\rho_B(e) = \begin{cases} \frac{k^2}{6} \left\{ 1 - \left[1 - \left(\frac{e}{k} \right)^2 \right]^3 \right\} & \text{for } e \leq k \\ k^2/6 & \text{for } e > k \end{cases}$	$w_B(e) = \begin{cases} \left[1 - \left(\frac{e}{k} \right)^2 \right]^2 & \text{for } e \leq k \\ 0 & \text{for } e > k \end{cases}$

O procedimento Mínimos Quadrados Reponderados Iterativamente, para estimação dos parâmetros apresenta resultados eficientes em regressões múltiplas econométricas (ex. MENDES, 1998) e também em modelos autorregressivos (ex. El-Hawary & Mbamalu, 1993 & 1996; Mbamalu & El-Hawary, 1993; Mili & al, 1999; Kuzmic & al, 2004).

Resumidamente, o Modelo Dinâmico Iterativo parte do princípio que na explicação da carga reativa horária são necessárias cargas históricas das mesmas e da carga ativa. Com isso, na estimação dos parâmetros, a correlação entre as variáveis explicativas deve ser considerada, já que se trata de um estudo de séries temporais.

As análises são feitas para cada grupo formado pela classificação dos dados de reativa. Ou seja, modelos diferentes são utilizados para grupos diferentes. Ainda sim, a metodologia de estimação proposta é desenvolvida para cada caso. De maneira geral, o modelo de regressão proposto é mostrado a seguir:

$$\left\{ \begin{array}{l} y_t = \beta_0 + \sum_{j=1}^k \sum_{i=1}^{p_j} \beta_{ji} x_{jt-i} + \sum_{i=1}^p \beta_i y_{t-i} + u_t \\ R(B)u_t = \varepsilon_t \\ R(B) = 1 - r_1 B^1 = AR(1) \end{array} \right. \quad (4.15)$$

onde:

y_t : carga reativa no instante t

x_t : carga ativa no instante t

y_{t-i} : carga reativa no instante $t-i$

x_{t-i} : carga ativa no instante $t-i$

β_0 : constante

β_{ji} : coeficiente da j -ésima variável independente na defasagem i

k : quantidade de variáveis independentes

β_i : coeficiente da variável dependente na defasagem i

p_j : ordem de defasagem da j -ésima variável independente

p : ordem de defasagem da variável dependente

u_t : erro correlacionado

$R(B)$: chamado de “fator comum” e representa a estrutura de correlação presente no erro u_t .

ε_t : ruído aleatório associado ao modelo, onde se supõem que são independentes e identicamente distribuídos com densidade $N(0, \sigma^2)$.

Para estimar os coeficientes a metodologia utilizada é o algoritmo de mínimos quadrados ponderados iterativamente com a função de peso do tipo Biquadrada – que apresentou melhor resultado em relação a função de peso do tipo Huber e Mínimos Quadrados Ordinários- associado a uma correção de autocorrelação Cochran-Orcutt. Esse modelo atribui pesos à observação de acordo o quanto essa observação se encontra fora do padrão dos dados, baseando-se no resíduo da observação. Os pesos são posteriormente utilizados para diminuir a influência da observação e revisados após cada iteração até que um ajuste robusto seja encontrado tanto para os coeficientes da regressão quanto para o coeficiente dos resíduos (Neter et al., 1996). O algoritmo é apresentado a seguir:

1. Considerar, inicialmente, $r_i=0$; $i=1,2,\dots,n$
2. Estimar $b^{(0)}$, inicialmente por mínimos quadrados.
3. Calcular os resíduos $u_i^{(t-1)}$ e os pesos associados $w_i^{(t-1)} = w[u_i^{(t-1)}]$. A cada iteração t .
4. Calcular as novas estimativas de $b^{(t)}$ por mínimos quadrados ponderados

$$b^{(t)} = \left[X' W^{(t-1)} X \right]^{-1} X' W^{(t-1)} y_t$$

onde X é a matriz cujas colunas representam a constante, a variável exógena nos instantes $t-i$, e a variável dependente nos instante $t-i$. e

$W^{(t-1)} = \text{diag}\{w_i^{(t-1)}\}$ é a matriz peso.

5. Calcular os resíduos $u_t = y_t - \beta_0 - \sum_{j=1}^k \sum_{i=1}^p \beta_{ji} x_{jt-i} - \sum_{i=1}^p \beta_i y_{t-i}$
6. Estimar $r^{(0)}$, inicialmente, por mínimos quadrados.
7. Calcular os resíduos $\epsilon_i^{(t-1)}$ e os pesos associados $w_{i1}^{(t-1)} = w_1[\epsilon_i^{(t-1)}]$. A cada iteração t .
8. Calcular as novas estimativas de $r^{(t)}$ por mínimos quadrados ponderados

$$r^{(t)} = \left[U_{t-1}' W_1^{(t-1)} U_{t-1} \right]^{-1} U_{t-1}' W_1^{(t-1)} u_t$$

onde U_{t-1} é a matriz dos resíduos no instante $t-1$, u_i como a i -ésima linha, e $W_1^{(t-1)} = \text{diag}\{w_{i1}^{(t-1)}\}$ é a matriz peso.

9. Calcular os valores transformados $Y_t^* = R(B) Y_t$ e $X_t^* = R(B) X_t$
10. Estimar $b^{*(0)}$, inicialmente, por mínimos quadrados.
11. Calcular os resíduos $u_i^{*(t-1)}$ e os pesos associados $w_i^{*(t-1)} = w^*[u_i^{*(t-1)}]$. A cada iteração t .

12. Calcular as novas estimativas de $b^{*(t)}$ por mínimos quadrados ponderados

$$b^{*(t)} = \left[X^{*'} W^{*(t-1)} X^* \right]^{-1} X^{*'} W^{*(t-1)} y_t^*$$

onde X^* é a matriz cujas colunas representam os valores transformados da constante, das variáveis exógenas nos instantes $t-i$, e das variáveis dependentes nos instantes $t-i$. $W^{*(t-1)} = \text{diag}\{w_i^{*(t-1)}\}$ é a matriz peso.

13. Substituir os valores dos β^* obtidos na regressão original (4.15) e calcular os novos resíduos.
14. Repetir os passos 7 a 14 até a convergência dos coeficientes r_i 's, β 's.

4.3. Especificação do Modelo Dinâmico

A elaboração de um modelo de regressão dinâmica é muitas vezes um procedimento difícil, pois se precisa não apenas escolher as variáveis a serem incluídas no modelo, mas também os “lags” (defasagens) destas variáveis.

Na definição do modelo adequado é necessário levar em conta não só a significância dos parâmetros, mas também uma certa estrutura “lógica” do modelo. Em resumo, na escolha de um modelo de regressão não é necessário apenas encontrar um ajuste de parâmetros de adequado, também é necessário verificar se os coeficientes estimados são coerentes.

A estratégia usualmente empregada para construir um modelo de regressão dinâmica é uma estratégia “bottom-up”, isto é, parte-se de um modelo simples refinando-se incluindo novas variáveis até encontrar um modelo apropriado.

O fluxograma a seguir (figura 5) indica, de maneira genérica, os passos usados na construção de um modelo de regressão dinâmica.

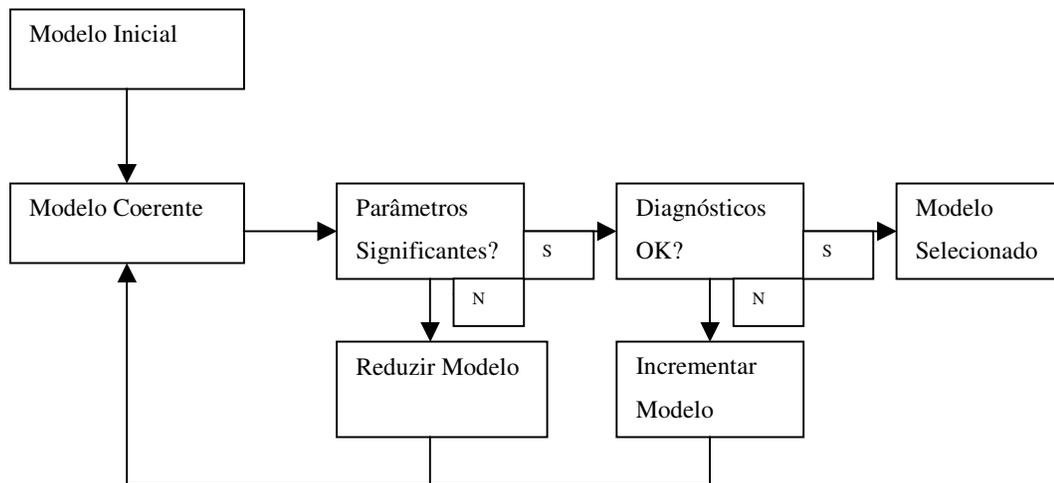


Figura 5 – Fluxograma da construção de um modelo de regressão dinâmica.

A especificação correta de um modelo de regressão envolve a precisa especificação de ambas a relação causal entre as variáveis e da estrutura dinâmica do modelo.

4.4. Testes Estatísticos

Alguns testes estatísticos são necessários para que o modelo híbrido de previsão de reativo tenha um desempenho favorável. Foram escolhidos alguns testes dentre os existentes para medir normalidade (Komogorov Smirnov), homocedasticidade (Levene), autocorrelação dos erros (Durbin e Durbin-Watson) e erros de especificação do modelo (Multiplicador de Lagrange). Todos são considerados com nível de significância de 5%.

4.4.1. Teste de Normalidade

Testes de aderência, como o Kolmogorov-Smirnov (Goodness of fit) servem para comparar as probabilidades empíricas de uma variável com as probabilidades teóricas estimadas pela função de distribuição em teste, verificando se os valores da amostra podem razoavelmente ser considerados como provenientes de uma população com aquela distribuição teórica. Nos testes de aderência, a hipótese nula (H_0) admite que a distribuição seja a especificada (normal, log-normal, gama

e outras), com os seus parâmetros estimados com base nos dados amostrais. O teste de aderência qui-quadrado apresenta limitações. Por exemplo, a frequência de uma classe não pode ser inferior a cinco e os dados são agrupados em classes perdendo informações, o que não ocorre no teste de Kolmogorov-Smirnov, que além de poder ser realizado com os dados agrupados, pode também ser realizado com os dados isoladamente, sendo normalmente mais eficiente. O teste de Kolmogorov-Smirnov é baseado no módulo da maior diferença entre a probabilidade observada e a estimada, que é comparada com um valor tabelado de acordo com o número de observações da série sob teste. (Campos, 1983).

4.4.2. Teste de Homocedasticidade

Alguns testes estatísticos referem que as variáveis apresentam certas características das distribuições mais comuns como a distribuição normal, simétrica ou pelo menos a distribuição de uma variável dentro de cada grupo de outra variável.

O teste de Levene tem como hipótese nula que a variância de uma dada variável é igual dentro de cada grupo de outra variável, em outras palavras, a hipótese nula consiste nas duas variâncias serem iguais e portanto existe uma relação de independência entre as variáveis. Se o nível de significância do teste é inferior a 0,05 é possível rejeitar com 95% de confiança a hipótese nula (Carvier, 2004).

O resultado do teste de Levene condiciona a escolha do teste t a adoptar. Deste modo, quando o valor p para o teste Levene é superior a 0,05 usam-se os resultados para o teste t que correspondem à assunção de variâncias iguais para ambos os grupos. Caso contrário, usam-se os resultados que assumem, para ambos os grupos, que as variâncias não são iguais.

4.4.3. Teste de Autocorrelação dos Erros

O mais usual teste para detectar a correlação serial de primeira ordem foi desenvolvido por Durbin e Watson. É popularmente conhecido como estatística d de Durbin-Watson definida como:

$$d = \frac{\sum_{t=2}^{t=n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^{t=n} \hat{u}_t^2} = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \quad (4.16)$$

Como $\sum \hat{u}_t^2$ e $\sum \hat{u}_{t-1}^2$ diferem em apenas uma observação, elas são aproximadamente iguais. Logo, estabelecendo que $\sum \hat{u}_t^2 = \sum \hat{u}_{t-1}^2$, (4.16) pode ser escrita

$$d \cong 2 \left(1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \right) \quad (4.17)$$

Definindo como o coeficiente de autocorrelação de primeira ordem na amostra, um estimador de r

$$\hat{r} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \quad (4.18)$$

Substituindo (4.18) em (4.17), tem-se

$$d = 2(1 - \hat{r}) \quad \text{onde} \quad -1 \leq r \leq 1 \quad \text{e} \quad 0 \leq d \leq 4 \quad (4.19)$$

Fica evidente pela equação (4.19) que, se $\hat{r} = 0, d = 2$; ou seja, se não houver correlação serial (de primeira ordem), é esperado que d seja cerca de 2. Se $\hat{r} = +1$, indicando uma correlação positiva perfeita nos resíduos, $d \cong 0$. Portanto, quanto mais próximo de zero estiver d , maior o indício de correlação serial positiva. E, vice-versa, quanto mais próximo de 4 estiver d , maior o indício de correlação serial negativa.

Uma grande vantagem da estatística de Durbin-Watson é que ela se baseia nos resíduos estimados, que são calculados rotineiramente na análise de regressão.

Embora usada rotineiramente, é importante observar as hipóteses que fundamentam esta estatística:

1. O modelo de regressão inclui um termo de intercepto.
2. As variáveis explicativas são não-estocásticas, ou fixadas em amostragem repetida.
3. As perturbações são geradas pelo esquema auto-regressivo de primeira ordem como ilustrado em (4.3).
4. O modelo de regressão não inclui valor(es) defasado(s) da variável dependente como uma das variáveis explicativas, ou seja, este teste não é aplicável a modelos auto-regressivos.
5. Não há observações faltantes na amostra.

Como muitas regressões envolvendo séries temporais são modelos auto-regressivos, Durbin (1970) propôs um outro teste para detectar a presença de correlação no modelo de variável dependente defasada. Esse teste deve ser aplicado quando há uma variável dependente defasada no membro direito da equação, pois, nesse caso, o teste Durbin-Watson usual é tendencioso, já que normalmente não detecta qualquer correlação. O novo teste é válido em grandes amostras e se baseia na estatística (Hill,1999),

$$h = \left(1 - \frac{d}{2}\right) \sqrt{\frac{n-1}{1 - (n-1)[dp(b)]^2}}$$

onde d é estatística de teste de Durbin-Watson usual para erros AR(1), n é o tamanho da amostra, e $dp(b)$ é o desvio padrão do coeficiente estimado da variável dependente defasada em $t-1$. Mesmo que existam outros valores defasados da variável dependente, para o cálculo da estatística Durbin, somente o desvio padrão do coeficiente estimado da variável dependente defasada em $t-1$ é considerado. Em grandes amostras a estatística h tem distribuição normal padronizada se u_t não é autocorrelacionado. Se $|h| \geq 1,96$, rejeita-se a hipótese nula, de nenhuma correlação, em favor da alternativa, de que existe correlação positiva.

4.4.4. Teste de Erro de Especificação

O teste com multiplicador de Lagrange (LM) para adição de variáveis tem como distribuição assintótica uma distribuição Qui-Quadrado com k graus de liberdade.

O “score” é definido como a primeira derivada da log -verossimilhança em relação ao vetor de parâmetros, ou seja,

$$S(\theta) = \frac{\partial l(\theta)}{\partial(\theta)}$$

Seja $S(\theta_0)$ o “score” avaliado sob a hipótese nula. Considere a seguinte estatística de teste:

$$LM = T.S'(\theta_0).I^{-1}.S(\theta_0)$$

onde

S' = o transposto do vetor “score”.

I^{-1} = a inversa da matriz de informação de Fisher.

O “score” mede a inclinação da log -verossimilhança, e então a estatística LM calcula a taxa de crescimento da log-verossimilhança quando se afasta da hipótese nula.

O teste LM não envolve o cálculo da log-verossimilhança sob a hipótese alternativa.

4.5. Algoritmo Conjunto

A figura 6 mostra o algoritmo do Modelo Híbrido de Previsão de Reativa proposto. Este algoritmo foi implementado no software MATLAB 6.0 utilizando algumas de suas Toolbox.

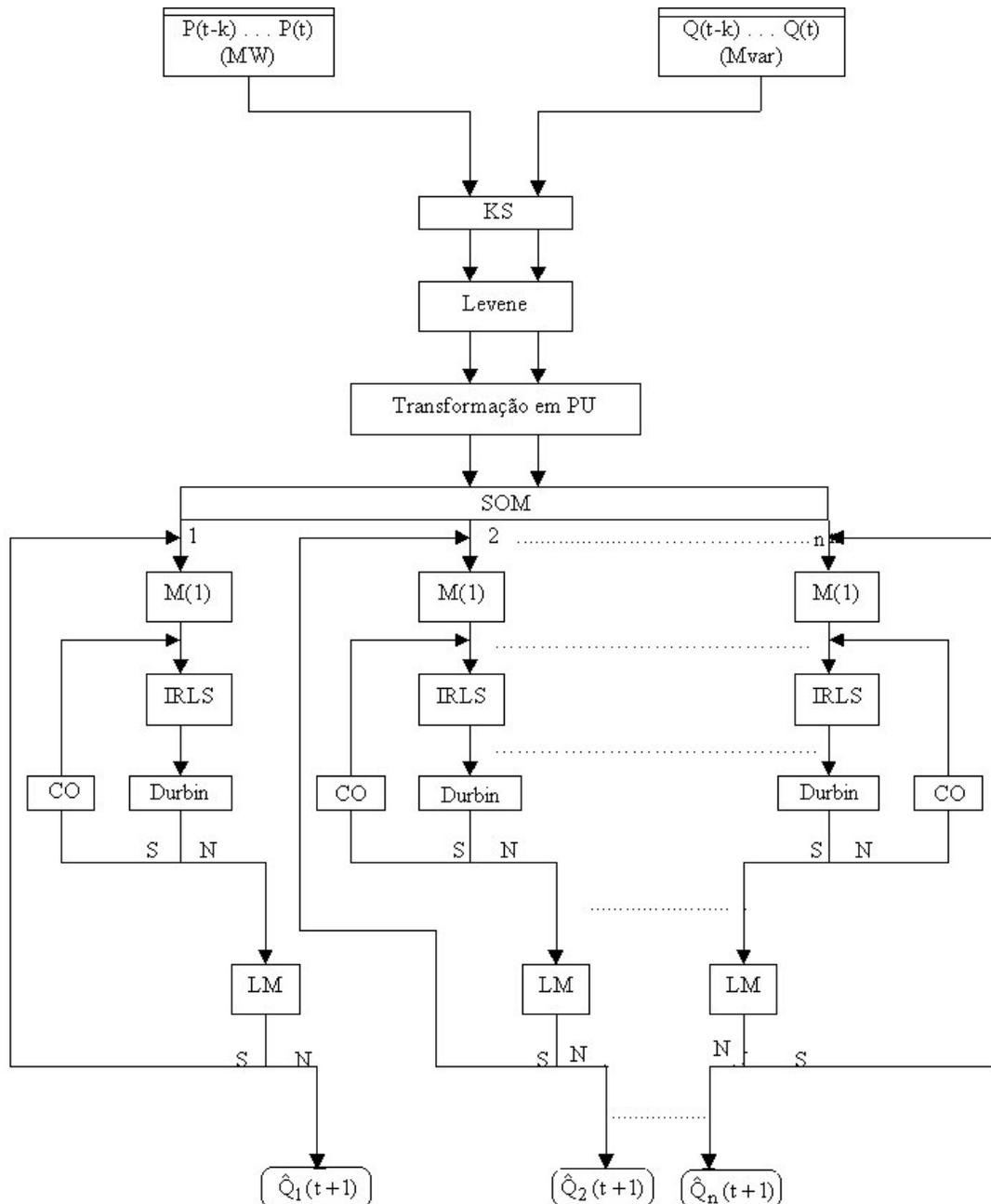


Figura 6 – Algoritmo conjunto