

4 Metodologia

4.1. Introdução

Neste capítulo apresentam-se os modelos STVAR-Tree, o principal da dissertação, além dos modelos competidores PAR(p) e Neuro-Fuzzy.

4.2. Modelo STVAR-Tree

O modelo vetorial auto-regressivo de transição suave estruturado por árvores (em inglês, *Smooth Transition Vector Autoregressive-Tree*), denominado aqui por modelo STVAR-Tree, é resultante da combinação do modelo STVAR (*Smooth Transition Vector Autoregressive*) com o algoritmo CART (*Classification and Regression Tree*), ambos discutidos no Capítulo 3.

A construção do modelo STVAR-Tree é uma adaptação do modelo STAR, proposto por Dijk, Teräsvirta e Franses (2002). O ciclo da modelagem consiste em três etapas: especificação, estimação e avaliação do modelo.

A especificação do modelo é feita através de métodos estatísticos, o que permite desenvolver testes de hipóteses. Por esta abordagem, é possível especificar um modelo não-linear paramétrico por meio de uma árvore de decisão. Além disso, a árvore de decisão final pode ser facilmente interpretada, como sentenças lógicas.

A estimação dos coeficientes pode ser vista como uma combinação de diferentes modelos auto-regressivos, podendo utilizar variáveis exógenas no conjunto daquelas que ajudam a compor o espaço de transição.

E, por fim, a avaliação do modelo final dá-se por meio de seu desempenho (habilidade preditiva) em dados fora da amostra (*out-of-sample*).

4.2.1. Formulação matemática

Sejam y_1, y_2, \dots, y_t séries temporais multivariadas de variáveis endógenas, $y_1 = (y_{11}, y_{21}, \dots, y_{K1})'$, $y_2 = (y_{12}, y_{22}, \dots, y_{K2})'$, ..., $y_t = (y_{1t}, y_{2t}, \dots, y_{Kt})'$. Defina $f(\cdot)$ como uma função desconhecida de séries temporais multivariadas x_t , as quais podem ser as variáveis endógenas defasadas e variáveis exógenas, através da expressão:

$$y_t = f(x_t) + \varepsilon_t \quad (4.1)$$

tal que $E[\varepsilon_t] = 0$ (vetor nulo de dimensão K) e $E[\varepsilon_t \varepsilon_t'] = \Sigma_\varepsilon$ (matriz de variância-covariância de dimensão $K \times K$).

Define-se, portanto, a função $H_{\mathbb{J}\mathbb{T}}(x_t; \psi): \mathbb{R}^{q+1} \rightarrow \mathbb{R}$, indexada pelos parâmetros ψ , como o modelo STVAR-Tree através da expressão:

$$H_{\mathbb{J}\mathbb{T}}(x_t; \psi) = \sum_{i \in \mathbb{T}} \Phi_i \tilde{z}_t B_{\mathbb{J}i}(x_t; \theta_i) + \varepsilon_t \quad (4.2)$$

onde,

o vetor $x_t = (x_{1t}, x_{2t}, \dots, x_{qt})'$, tal que $x_t \in \mathbb{S}^q$;

o vetor $z_t = (z_{1t}, z_{2t}, \dots, z_{pt})'$, o qual ($z_t \subseteq x_t$), tal que $z_t \in \mathbb{R}^p, p < q$, contém as variáveis endógenas defasadas e, ainda, pode conter variáveis exógenas e variáveis exógenas defasadas; e o vetor $\tilde{z}_t = (1, \dots, 1, z_t)'$ contém 1's indicando os interceptos acrescentados ao vetor z_t ;

o vetor $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Kt})'$ é K -dimensional contendo ruídos brancos, ou seja, $E[\varepsilon_t] = 0, E[\varepsilon_t \varepsilon_t'] = \begin{cases} \Sigma_\varepsilon, & \text{se } t=s \\ 0, & \text{se } t \neq s \end{cases}$

\mathbb{J} é o conjunto de índices de nós geradores. Define-se \mathbb{J}_i como o subconjunto de \mathbb{J} contendo os índices dos nós geradores que formam o caminho para a folha (nó terminal) i ;

\mathbb{T} é o conjunto de índices de folhas (nós terminais);

\mathbb{S} é o conjunto de índices de variáveis de transição;

As funções $B_{\mathbb{J}i}, 0 < B_{\mathbb{J}i} < 1$, são conhecidas como funções de pertinência, dadas por.

$$B_{\mathbb{J}i}(x_t; \theta_i) = \prod_{j \in \mathbb{J}} G(x_{S_j,t}; \gamma_j, c_j)^{\frac{n_{i,j}(1+n_{i,j})}{2}} \left[1 - G(x_{S_j,t}; \gamma_j, c_j) \right]^{(1-n_{i,j})(1+n_{i,j})} \quad (4.3)$$

onde,

$$n_{i,j} \begin{cases} -1, & \text{se o caminho para a folha } i \text{ não inclui o nó ancestral } j \\ & (B_{\mathbb{J}i} = 1); \\ 0, & \text{se o caminho para a folha } i \text{ inclui o filho direito do nó ancestral } j \\ & (B_{\mathbb{J}i} = 1 - G); \\ 1, & \text{se o caminho para a folha } i \text{ inclui o filho esquerdo do nó ancestral } j \\ & (B_{\mathbb{J}i} = G). \end{cases}$$

Vale ressaltar que $\sum_{j \in \mathbb{J}} B_{\mathbb{J}i}(x_t; \theta_j) = 1, \forall x_t \in \mathbb{S}^q$.

A função $G(x_t; \theta)$ é definida como Função Logística, da forma:

$$G(x_t; \gamma, c) = \frac{1}{1 + e^{-\gamma(x_t - c)}} \quad (4.4)$$

Define-se $\psi = (\Phi_i, \theta_i)$ como a representação de todos os parâmetros do modelo, sendo particionado em dois grupos, classificados como parâmetros lineares $\Phi_i = (\Phi_{1i}, \dots, \Phi_{pi})$ e parâmetros não-lineares $\theta_i = (\gamma_i, c_i)$. A matriz de parâmetros Φ_i é aquela que contém os coeficientes associados às defasagens das variáveis dependentes, e o vetor θ_i é formado pelos parâmetros de suavidade e locação, respectivamente.

O motivo de z_t ser um subconjunto de x_t (conjunto de variáveis de transição) é evitar regressores não estacionários nos modelos lineares locais, uma vez que podemos ter séries não estacionárias como variáveis de transição. Uma ilustração deste caso é dada quando a variável de transição x_{ti} trata-se de uma tendência linear, a qual identifica uma possível quebra estrutural nas séries.

4.2.2. Especificação do modelo

O objetivo, aqui, é obter uma estratégia para o crescimento da árvore, baseado na inferência estatística. Um processo de construção similar ao proposto por van Dijk, Teräsvirta e Franses (2002) para o modelo STAR foi adaptado para

a modelagem STVAR-Tree. Os passos são basicamente os mesmos descritos no Capítulo 3, os quais seguem o princípio “específico para geral”:

1) Seleção de variáveis relevantes para o modelo

Caso o número de variáveis disponíveis para a composição do modelo seja elevado, é importante selecionar as mais relevantes. As candidatas à variáveis de transição podem ser variáveis exógenas ou defasagens da variável dependente. Dois critérios podem ser utilizados nesta fase: a) conhecimento subjetivo do problema em questão, ou seja, usar o conhecimento prévio sobre as séries temporais; ou b) estimar diversos modelos lineares VAR, incrementando a ordem p do modelo, com $p = 1, 2, \dots, p_{max}$. Aquele modelo que minimizar os critérios de informação AIC (Akaike, 1974) ou BIC (Schwarz, 1978) deve ser selecionado e todas as propriedades dos modelos lineares VAR(p) devem ser verificadas. As variáveis que compõem este modelo devem ser selecionadas para a modelagem do STVAR-Tree.

2) Especificação do modelo, baseado em uma sequência de testes de hipótese do tipo LM, o qual avalia a linearidade de um determinado modelo

Esta etapa define o modelo a ser estimado. Aqui, os conceitos de inferência estatística são utilizados para determinar o crescimento da árvore. O procedimento envolve uma sequência de testes do tipo Multiplicadores de Lagrange, conhecido como teste LM (do inglês, *Lagrange Multiplier*), como apresentado em Luukkonen, Saikkonen e Teräsvirta (1988). Na construção do modelo STVAR-Tree, o teste LM verifica a adequação do modelo STVAR-Tree para descrever os dados. No caso em que a hipótese nula de linearidade for rejeitada, seleciona-se o nó a ser dividido e a variável de transição para estimar o modelo.

3) Estimação dos parâmetros lineares (constantes dentro dos nós) e não-lineares (parâmetros da função logística) do modelo selecionado

Dois métodos de estimação são utilizados para produzir as estimativas dos parâmetros do modelo STVAR-Tree, representados por $\psi = (\Phi_i, \theta_i)$. Para estimar os parâmetros classificados como lineares $\Phi_i = (\Phi_{1i}, \dots, \Phi_{pi})$ utiliza-se Mínimos Quadrados Multivariados (ou Mínimos Quadrados Generalizados – MQG) e para estimar os parâmetros não-lineares $\theta_i =$

(γ_i, c_i) utiliza-se Mínimos Quadrados Não-Lineares (MQNL), que é equivalente ao método de Máxima Verossimilhança (MV) sob a normalidade dos erros. Caso os erros não sejam normais, MQNL é equivalente à Quase-Máxima Verossimilhança (QMV).

4) Análise de diagnóstico do modelo

Avalia-se os modelos estruturados por árvores pelo seu desempenho estatístico (habilidade preditiva) em dados fora da amostra (*out-of-sample*).

5) Re-especificação do modelo de acordo com os resultados do diagnóstico

No caso de selecionar um modelo STVAR-Tree que não produza resultados estatísticos satisfatórios, deve-se voltar a etapa 1 e re-especificar um modelo e, assim, seguir todas as etapas.

6) Utilização do modelo com fins descritivos ou de previsão

Com todas as etapas acima verificadas e satisfeitas, determina-se o modelo $H_{\text{JT}}(x_t; \psi)$ como o modelo final que será utilizado de acordo com os seus propósitos.

Esta etapa de especificação engloba a etapa de estimação dos parâmetros, definida na seção posterior, pois o teste LM necessita estimar os modelos sob H_0 e H_1 , o que indica que, para cada divisão, são estimados os parâmetros lineares e não-lineares dos modelos.

4.2.3. Estimação dos parâmetros

Definiu-se que, para a estimação de todos os parâmetros do modelo STVAR-Tree, representados por $\psi = (\Phi_i, \theta_i)$, leva-se em consideração a hipótese de que o vetor $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Kt})'$ é formado por variáveis aleatórias normalmente distribuídas com média zero e matriz de variância-covariância Σ_ε .

Para estimar os parâmetros lineares do modelo inicializa-se os valores dos parâmetros não-lineares $\theta_i = (\gamma_i, c_i)$ através de uma grade de valores (*grid*). Uma vez inicializados, ou seja, dado que os parâmetros não-lineares são conhecidos, a estimação do modelo STVAR-Tree corresponde a estimação de uma regressão multivariada e os parâmetros lineares $\Phi_i = (\Phi_{1i}, \dots, \Phi_{pi})$ são estimados por

Mínimos Quadrados Multivariados (ou Mínimos Quadrados Generalizados – MQG), dados por:

$$\hat{\Phi} = [Z(\theta)'Z(\theta)]^{-1}Z(\theta)'y_t \quad (4.5)$$

onde,

$$y_t = (y_{1t}, y_{2t}, \dots, y_{kt})',$$

$$\theta = (\theta'_1, \dots, \theta'_k)' \text{ e}$$

$$Z(\theta) = \begin{bmatrix} \tilde{z}_1 B_1(x_1; \theta_1) & \cdots & \tilde{z}_1 B_K(x_1; \theta_K) \\ \vdots & \ddots & \vdots \\ \tilde{z}_T B_1(x_T; \theta_1) & \cdots & \tilde{z}_T B_K(x_T; \theta_K) \end{bmatrix}$$

Esse método de estimação é conhecido como Máxima Verossimilhança Concentrada, sob normalidade dos erros, ou Quase-Máxima Verossimilhança Concentrada, caso os erros não sejam normais, e a expressão é dada através de:

$$\hat{\psi} = \underset{\psi \in \Psi}{\operatorname{argmin}} \left\{ \frac{1}{T} \sum_{t=1}^T [y_t - H_{\mathbb{J}\mathbb{T}}(x_t; \psi)]' [y_t - H_{\mathbb{J}\mathbb{T}}(x_t; \psi)] \right\} \quad (4.6)$$

Os parâmetros não-lineares de ψ , ou seja, vetor θ_i são estimados condicionalmente a Φ através de MQNL, o qual faz uso de procedimentos de otimização não-linear. Para a estimação dos parâmetros do modelo STVAR-Tree considerou-se o algoritmo de Levenberg-Marquadt, completando a estimação da iteração. Como o método MQNL é muito sensível aos valores iniciais, é feita uma busca de valores iniciais por grade de valores (*grid*), que busca a maximização da função de log-verossimilhança concentrada.

4.2.4. Divisão dos nós – Seqüência de testes LM

Considere um modelo STVAR-Tree com K nós terminais, escrito como:

$$y_t = H_{\mathbb{J}\mathbb{T}}(x_t; \psi) = \sum_{i \in \mathbb{T}} \Phi_i \tilde{z}_t B_{\mathbb{J}i}(x_t; \theta_i) + \varepsilon_t \quad (4.7)$$

ou

$$y_t = \sum_{i \in \mathbb{T} - \{i^*\}} \Phi_i \tilde{z}_t B_{\mathbb{J}i}(x_t; \theta_i) + \Phi_{2i^*+1} \tilde{z}_t B_{\mathbb{J}2i^*+1}(x_t; \theta_{2i^*+1}) \\ + \Phi_{2i^*+2} \tilde{z}_t B_{\mathbb{J}2i^*+2}(x_t; \theta_{2i^*+2}) + \varepsilon_t \quad (4.8)$$

onde

$$B_{\mathbb{J}2i^*+1}(x_t; \theta_{2i^*+1}) = B_{\mathbb{J}i^*}(x_t; \theta_{i^*})G(x_{i^*t}; \gamma_{i^*}, c_{i^*});$$

$$B_{\mathbb{J}2i^*+2}(x_t; \theta_{2i^*+2}) = B_{\mathbb{J}i^*}(x_t; \theta_{i^*})[1 - G(x_{i^*t}; \gamma_{i^*}, c_{i^*})].$$

A equação pode ser escrita na forma:

$$y_t = \sum_{i \in \mathbb{T} - \{i^*\}} \Phi_i \tilde{z}_t B_{\mathbb{J}i}(x_t; \theta_i) + \phi \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) \\ + \lambda \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*})G(x_{i^*t}; \gamma_{i^*}, c_{i^*}) + \varepsilon_t \quad (4.9)$$

onde $\phi = \Phi_{2i^*+2}$ e $\lambda = \Phi_{2i^*+1} - \Phi_{2i^*+2}$.

Define-se, portanto, a construção do modelo através de testes de hipóteses do tipo Multiplicadores de Lagrange, conhecido como teste LM. Desta forma, o crescimento da árvore está condicionado à existência de não-linearidade nas séries modeladas.

Primeiramente, o teste LM deve ser realizado, baseado nas hipóteses:

H_0 : Linearidade

H_1 : Não-linearidade

Esta etapa consiste em testar se o nó raiz deve ou não ser dividido. Em outras palavras, o teste determinará um modelo constante VAR ou o mais simples modelo STVAR-Tree, com 2 nós terminais.

Sob H_0 , é fácil verificar que, para $\gamma = 0$, a função logística $G(\cdot)$ assume um valor constante igual a 0,5, independente do valor do parâmetro de locação c , pois:

$$G(x_t; 0, c) = \frac{1}{1 + e^{-0(x_t - c)}} = \frac{1}{1 + e^0} = \frac{1}{1 + 1} = \frac{1}{2}$$

Isto implica que a estimação de $H_{\mathbb{T}}(x_t; \psi)$ será simplesmente a média ponderada entre os regimes, sendo então, uma combinação linear de modelos lineares VAR. Consequentemente, o modelo resultante também será linear. Portanto, seria suficiente testar a linearidade das séries considerando as hipóteses:

$$H_0: \gamma = 0$$

$$H_1: \gamma > 0$$

Porém, ao considerar diferentes valores para o parâmetro c , a função de verossimilhança pode permanecer inalterada, o que acarretará num problema de identificação. Como uma solução deste problema, adota-se a proposta de Luukkonen, Saikkonen e Teräsvirta (1988), que aproxima a função $G(\cdot)$ por uma expansão de Taylor de 3ª ordem em torno de $\gamma = 0$. Então, ao reescrever o modelo $H_{\mathbb{T}}(x_t; \psi)$ considerando a expansão de Taylor no lugar da função $G(\cdot)$, chega-se a seguinte expressão:

$$G(x_t; \gamma, c) = \frac{1}{2} + 2\gamma(x_t - c) + \frac{3}{2}\gamma^2(x_t - c)^2 + \frac{1}{2}\gamma^3(x_t - c)^3 + R_3(x_t; \gamma, c) \quad (4.10)$$

onde $R_3(x_t; \gamma, c)$ é o resto da expressão.

Considerando a expansão de Taylor de 3ª ordem em torno de $\gamma = 0$, o modelo pode ser escrito como:

$$y_t = \sum_{i \in \mathbb{T} - \{i^*\}} \Phi'_i \tilde{z}_t B_{\mathbb{J}i}(x_t; \theta_i) + \alpha'_0 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) + \alpha'_1 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) x_{i^*t} + \alpha'_2 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) x_{i^*t}^2 + \alpha'_3 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) x_{i^*t}^3 + e_t \quad (4.11)$$

onde $e_t = \varepsilon_t + \lambda' \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) + R_3(x_{i^*t}; \gamma_{i^*}, c_{i^*})$

Os parâmetros α_k , $k = 0, \dots, 3$ são funções dos parâmetros originais lineares e não-lineares do modelo, $(\phi, \lambda, \gamma_{i^*}, c_{i^*})$. Desta forma, as hipóteses do teste LM passam a ser:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$

H_1 : caso contrário

Sob H_0 , o resto da expansão de Taylor desaparece, de forma que as propriedades do erro permanecem inalteradas. E isto gera condições propícias para o uso de inferência estatística assintótica.

O teste LM também pode ser aplicado na sua versão do teste F, seguindo os seguintes passos:

- 1) Estimar o modelo com K regimes
- 2) Regredir os resíduos \hat{u}_t em \hat{h}_t e calcular a soma dos novos resíduos quadráticos $SSR_0 = \sum_{t=1}^T \hat{u}_t^2$. Os novos resíduos \tilde{u}_t serão ortogonais à \hat{h}_t .
- 3) Fazer a regressão de \tilde{u}_t em \hat{h}_t e \hat{v}_t . Calcular a soma dos resíduos quadráticos obtidos a partir desta regressão, $SSR_1 = \sum_{t=1}^T \hat{v}_t^2$.
- 4) Obter a estatística χ^2

$$ML_{\chi^2} = T \frac{SSR_0 - SSR_1}{SSR_0},$$

ou a estatística F

$$ML_F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - n - m)},$$

onde $n = (q + 2)h + p + 1$.

Sob H_0 , a estatística de teste ML_{χ^2} segue uma distribuição χ^2 assintótica com m graus de liberdade, e a estatística de teste ML_F segue aproximadamente uma distribuição F com m e $T - n - m$ graus de liberdade, onde T é o número de observações.

Uma vez definidos a estimação do modelo e o teste ML para o crescimento da árvore, o algoritmo da modelagem torna-se simples.

4.2.5. Controle do crescimento da árvore

Na modelagem STVAR-Tree, o teste LM deve ser executado sequencialmente de modo a decidir a divisão de cada um dos nós sob teste. Sabe-

se que árvores com muitos nós terminais tornam-se complexas e isto compromete a análise do modelo final. Com a finalidade de controlar o erro do tipo I (árvore superestimada), adotou-se o procedimento de diminuir o nível de significância do teste, de acordo com o crescimento da árvore. Assim, ao realizar o teste pela n -ésima vez em um nó da d -ésima profundidade, o procedimento adotado faz com que $\alpha(d, n)$ assumo o valor:

$$\alpha(d, n) = \frac{\alpha}{n^d},$$

onde α é o nível de significância do primeiro teste.

Este procedimento força o teste LM a ser mais rigoroso nas divisões de maiores profundidades e, assim, o uso de técnicas de podagem citadas no Capítulo III não deve ser levado em consideração.

4.2.6. Previsão

A flexibilidade da metodologia STVAR-Tree permite que sejam criados três tipos de previsão diferentes:

1) Combinação de Regimes (RC):

Aplicação direta do modelo estimado. Utiliza-se a equação resultante da estimação para obter a previsão um passo a frente. Assim, a previsão é obtida a partir da soma ponderada dos modelos lineares locais, na qual os pesos são as pertinências das observações em cada regime.

2) Máxima Pertinência (MM):

Observa-se em qual regime a observação apresenta maior pertinência e aplica-se o modelo local correspondente.

3) Combinação Adaptativa de Regimes (do inglês, *Adaptive Regime Combining* – ARC):

Metodologia similar à RC, porém os parâmetros lineares são re-estimados a cada passo da previsão, utilizando as últimas observações, correspondentes a um ano. Para séries diárias, as últimas 365 (ou 252, no mercado financeiro). Séries semanais, as últimas 52 observações. E para séries mensais, as últimas 12 observações. E assim por diante.

4.3. Modelo PAR

Essencialmente, qualquer estrutura de dependência temporal pode ser reproduzida por modelos de séries temporais lineares do tipo PAR(p), sendo este tipo de modelo uma abordagem bastante flexível, e bastante popular para a modelagem estocástica de vazões fluviais (Hipel e McLeod, 1994). Na terminologia de séries temporais, a tendência hidrológica é conhecida como estrutura de dependência temporal, sendo quantificada pela função de autocorrelação estimada do registro de vazões.

A análise deste tipo de séries pode ser feita pelo uso de formulações auto-regressivas cujos parâmetros apresentam um comportamento periódico. A esta classe de modelos costuma-se denominar modelos auto-regressivos periódicos de ordem p - PAR(p), Salas et al. (1980). Em geral, p é um vetor, $p = (p_1, p_2, \dots, p_{12})$, onde cada elemento fornece a ordem de cada período.

4.3.1. Formulação matemática

O modelo PAR(p_1, p_2, \dots, p_{12}) pode ser descrito matematicamente por:

$$\left(\frac{Z_t - \mu_m}{\sigma_m} \right) = \phi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + a_t \quad (4.12)$$

onde,

Z_t é uma série sazonal de período s ;

s é o número de período ($s=12$ para séries mensais)

t é o índice do tempo, $t=1,2,\dots,sN$, função do ano T ($T=1,2,\dots,N$) e do período m ($m=1,2,\dots,s$);

N é o número de anos;

μ_m é a média sazonal de período s ;

σ_m é o desvio-padrão sazonal de período s ;

ϕ_m^i é o i -ésimo coeficiente auto-regressivo do período m ;

p_m é a ordem do operador auto-regressivo do período m ;

a_t é a série de ruídos independentes com média zero e variância .

E a correlação do modelo PAR(p) é dada por:

$$\rho^m(k) = E \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \quad (4.13)$$

4.4. Sistema Neuro-Fuzzy

Os Sistemas Neuro-Fuzzy (SNF) são ditos “inteligentes” e associam a capacidade de aprendizado das Redes Neurais e sua tolerância a falhas à interpretabilidade dos Sistemas Fuzzy (Zadeh, 1965). Ao se combinarem duas ou mais técnicas, cria-se, muitas vezes, um sinergismo que pode levar a um sistema mais poderoso. Verifica-se que SNF tem aplicação em diversas áreas, como reconhecimento de padrões, previsão, classificação, controle, etc., com a obtenção de bons resultados.

A razão para o seu bom desempenho é que os mesmos implementam um sistema de inferência fuzzy através de uma arquitetura de redes neurais (paralelamente distribuída), permitindo assim a integração de conhecimentos implícitos (própria base de dados) e explícitos (conhecimento de especialistas).

Uma série de arquiteturas de SNF tem sido proposta na literatura nos últimos anos. Dentre essas, pode-se citar o sistema ANFIS (Jang,1993), um dos modelos mais conhecidos e utilizados na prática. Passa-se a descrever o sistema ANFIS, o qual será utilizado no próximo capítulo para modelar o preço spot de energia elétrica no Brasil.

4.5. ANFIS: Adaptive Neuro-Fuzzy Inference System (Sistema Adaptativo de Inferência Neuro-Fuzzy)

Por simplicidade, considere um sistema de inferência fuzzy com duas entradas x e y e uma saída z . Para um modelo Fuzzy Sugeno de primeira ordem, um conjunto usual de regras “se-então” é a seguinte:

Regra 1: Se x é A_1 e y é B_1 , então $f_1 = p_1x + q_1y + r_1$

Regra 2: Se x é A_2 e y é B_2 , então $f_2 = p_2x + q_2y + r_2$

A partir deste modelo Fuzzy Sugeno de primeira ordem com duas entradas e duas regras ilustra-se a arquitetura ANFIS como aquela da Figura 4.1, onde os nós da mesma camada têm funções similares.

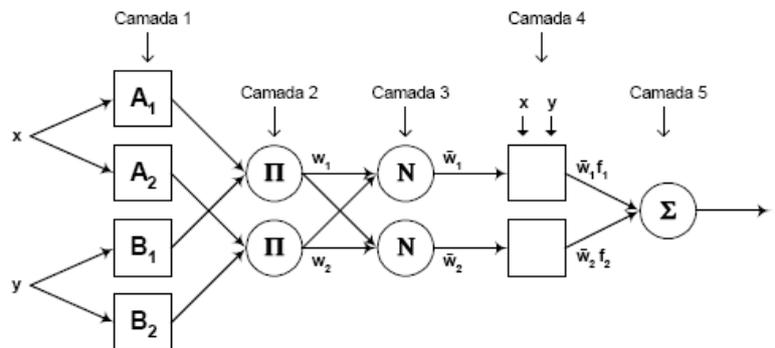


Figura 4.1: Arquitetura ANFIS equivalente

A seguir, passa-se a descrever cada camada do modelo.

Camada 1. Cada nó nesta camada (A_1, A_2, B_1, B_2) é um nó adaptativo com uma função denominada função de pertinência de um conjunto fuzzy, a qual gera a saída do tipo:

$$O_{1,i} = \mu_{A_i}(x), \text{ para } i = 1,2, \text{ ou}$$

$$O_{1,i} = \mu_{B_{i-2}}(y), \text{ para } i = 3,4$$

onde x (ou y) é a entrada para o nó i e A_i (ou B_{i-2}) é um rótulo lingüístico (como “alto” ou “baixo”) associado a este nó. Esta função especifica o grau, pertencente ao intervalo $[0,1]$, em que a entrada x (ou y) satisfaz este conjunto fuzzy,.

Camada 2. Cada nó nesta camada é um nó fixo rotulado Π , representa o nível da regra, cuja saída é o produto de todos os sinais de entrada:

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y).$$

Camada 3. Cada nó i nesta camada é um nó fixo rotulado N . A partir dessa camada temos o processo defuzzificador. A saída desse nó é dada pela razão entre o i -ésimo nível da regra e a soma de todas os níveis:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1,2.$$

Camada 4. Cada nó i nesta camada é um nó adaptativo com uma função nó:

$$O_{4,i} = \bar{w}_i \cdot f_i = \bar{w}_i(p_i x + q_i y + r_i) \quad i = 1,2.$$

onde w_i é um nível da Camada 3 e $\{p_i, q_i, r_i\}$ é o conjunto de parâmetros deste nó. Tem-se, então, um produto entre os níveis o valor do consequente da regra em si. Por isso, parâmetros nesta camada são denominados parâmetros consequentes.

Camada 5. Cada nó i nesta camada é um nó fixo rotulado Σ , que calcula a saída geral do sistema como a soma de todos os sinais de sua entrada:

$$Saída = O_{s,i} = \sum_i \bar{w}_i \cdot f_i = \frac{\sum_i w_i \cdot f_i}{\sum_i w_i}$$

Finalmente, a Figura 4.2 mostra uma arquitetura ANFIS que é equivalente a um modelo Sugeno de primeira ordem de duas entradas com nove regras, onde cada entrada tem três funções de pertinência associadas.

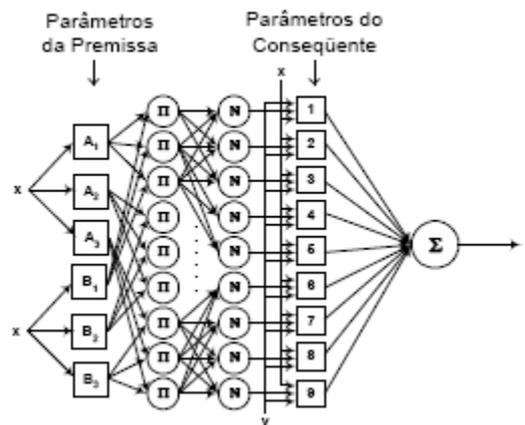


Figura 4.2: Arquitetura ANFIS para o modelo Sugeno

A identificação dos parâmetros do modelo ANFIS é realizada, em geral, empregando algoritmos de aprendizagem híbrida, isto é, combinando estimação de mínimos quadrados com retro-propagação.