

## 2 Modelos Não-Lineares

### 2.1 Introdução - Modelos Lineares como Inspiração

Um dos papéis mais importantes da estatística é construir modelos para explicar uma variável, usualmente chamada de variável dependente ou resposta, por um conjunto de variáveis explanatórias (independentes)  $(x_1, x_2, \dots, x_p)$ . Este modelo é expresso em termos de uma função matemática  $f$  tal que

$$y \approx f(x_1, x_2, \dots, x_p). \quad (2-1)$$

Quando  $f$  expressa uma função linear de  $y$  em  $x$  (ou uma transformação de  $x$ ), utilizando um conjunto  $\beta_0, \beta_1, \dots, \beta_p$  de pesos, que neste contexto são chamados de parâmetros, o modelo é dito ser linear. Observe os exemplos a seguir, contendo modelos lineares.

#### Exemplo 1

$$y \approx \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (2-2)$$

#### Exemplo 2

$$y \approx \beta_0 + \beta_1 \sin x_1 + \dots + \beta_p \sin x_p \quad (2-3)$$

Devido à grande flexibilidade para especificação de  $f$ , os modelos lineares são utilizados com frequência quando não há conhecimento teórico sobre a forma funcional  $f$ . A tendência até o final dos anos 1970 era a utilização de modelos não-lineares somente quando havia conhecimento teórico sobre a relação entre as variáveis, assim como os parâmetros envolvidos. Deste modo, pode ser considerado relativamente recente o desenvolvimento de modelos não-lineares para regressão e análise de séries temporais, se comparados com a bem sedimentada teoria de modelos lineares gaussianos e modelos lineares generalizados. Como representantes desta última classe, podem ser citados os Modelos de Regressão Linear (MRL) cuja formulação mais tradicional é apresentada na Equação (2-4).

$$y_t = \beta_0 + \sum_{j=1}^p \beta_j x_{j,t} + \varepsilon_t \quad (2-4)$$

A relação linear entre uma variável aleatória dependente  $y_t$  e um vetor de  $p$  variáveis explanatórias  $\mathbf{x}_t = (x_1, x_2, \dots, x_p)$  é descrita através da equação linear em (2-4), que contém um termo estocástico,  $\varepsilon_t$ , e coeficientes determinísticos  $\beta$ 's.

O Modelo de Regressão Linear Simples é obtido nos casos em que a dimensão de  $p$  é igual a 1 e assumindo que os erros são independentes e normalmente distribuídos. Para dimensões maiores do que 1, este é denominado Modelo de Regressão Linear Múltipla.

Outras distribuições de probabilidades para a componente estocástica do modelo são consideradas dentro da Teoria dos Modelos Lineares Generalizados (GLM - Generalized Linear Models) em [68].

Um modelo linear generalizado é caracterizado por:

$$E[y|x_1, x_2, \dots, x_p] = \mu \quad (2-5)$$

$$g(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p. \quad (2-6)$$

Dentro da formulação apresentada nas equações (2-5) e (2-6) a distribuição de  $y$  pertence a classe da família exponencial e  $g(\cdot)$  é chamada função de ligação. Esta vincula a média à um preditor linear. O modelo de Regressão Linear Múltipla é um caso particular quando  $y$  tem distribuição normal e  $g(\cdot)$  é a função identidade.

Quando os dados formam uma série temporal, ou seja, são coletados sequencialmente ao longo do tempo, algumas suposições feitas a respeito da componente aleatória do modelo de regressão linear passam a não ter sentido para a maioria das situações.

O pressuposto de independência entre os erros será constantemente violado. Outros exemplos decorrem de fatos estilizados que são inerentes à algumas séries temporais. O excesso de curtose, por exemplo, é comum em séries financeiras, enfraquecendo a suposição de normalidade.

O uso de modelos lineares para a análise de séries temporais está fortemente associado com a classe dos modelos AutoRegressivos Médias Móveis (ARMA - AutoRegressive Moving Average) desenvolvidos em [18] e representados através de (2-7).

$$y_t = \theta_0 + \sum_{i=1}^p \theta_i y_{t-i} + \sum_{j=1}^q \phi_j \epsilon_{t-j} + \epsilon_t \quad (2-7)$$

Estes modelos podem ser vistos como uma extensão do Modelo de Regressão Linear Múltipla, na qual o conjunto de variáveis explanatórias é formado por valores defasados da série temporal e choques passados. A aplicação destes modelos pressupõe que a série temporal seja fracamente estacionária. No caso de processos estacionários obtidos através de diferenciação, estes modelos recebem a denominação de ARIMA.

O ajuste destes modelos é obtido através de um ciclo iterativo que envolve 4 estágios: identificação (especificação), estimação, diagnóstico e previsão. Até os dias atuais, a etapa de identificação tem sido fértil para o desenvolvimento de novos métodos.

Os modelos ARMAX em (2-8) representam uma generalização dos modelos ARMA que inclui variáveis exógenas. Diferentes aplicações destes modelos podem ser encontradas na Econometria e Teoria de Controle (ver Ljung, 1987).

$$y_t = \theta_0 + \sum_{i=1}^p \theta_i y_{t-i} + \sum_{k=1}^m \sum_{s=0}^l \nu_k x_{k,t-s} + \sum_{j=1}^q \phi_j \epsilon_{t-j} + \epsilon_t \quad (2-8)$$

Note que em (2-8), quando  $p, q, l = 0$ , o modelo de regressão linear múltipla se torna um caso particular.

### 2.1.1

#### Especificação e Estimação em Modelos Lineares

Alguns métodos para a especificação e estimação de modelos lineares influenciaram técnicas utilizadas nos modelos não-lineares. Portanto, nesta seção são citados os métodos utilizados nos clássicos modelos de regressão linear sob a suposição de normalidade e também nos modelos ARMA.

Aqui são consideradas duas abordagens para a especificação de um modelo, seguindo os princípios desenvolvidos na Econometria. A primeira destas é a especificação do tipo geral-para-específico na qual o modelo de maior complexidade é escolhido como ponto de partida e testes posteriores são realizados para que este seja simplificado/reduzido. A outra abordagem é a específico-para-geral, na qual o modelo mais simples é especificado e testes para inclusão de variáveis ou aumento da complexidade são realizados gradualmente.

Seja qual for a estratégia de especificação, a Econometria faz amplo uso da inferência estatística - em especial testes estatísticos de hipóteses sobre os parâmetros para encontrar o modelo final.

### 2.1.2

#### Seleção de Variáveis em Modelos de Regressão Linear

Um grande problema encontrado durante a especificação de um modelo de regressão linear múltipla é a seleção de um subconjunto de  $p$  regressores a partir de uma lista de potenciais candidatos. Abaixo, são descritos alguns dos métodos mais conhecidos na literatura. Para maiores detalhes, recomenda-se [72]

#### Método Exaustivo (Todas as Regressões Possíveis)

Com o constante aumento da velocidade de processamento digital, uma das estratégias para especificar o modelo de regressão linear é através da construção de todas as regressões que forem possíveis - mediante a combinação de regressores - e posterior avaliação de critérios de otimalidade.

Os critérios de otimalidade podem estar relacionados com a qualidade do ajuste do modelo, através de medidas como o coeficiente de determinação  $R^2$ , a estatística  $C_p$  de Mallow e/ou a soma dos erros quadráticos.

O número de possíveis regressões cresce exponencialmente com  $k$ , a quantidade de potenciais regressores. Deste fato surgirá uma limitação natural na aplicação deste método.

#### Métodos dos Melhores Subconjuntos

Em situações nas quais a dimensão  $k$  do vetor de potenciais regressores impossibilita o ajuste de todas as regressões possíveis, lança-se mão de algoritmos que investigam apenas parte do espaço de todos os subconjuntos de regressores.

Se há  $k - 1$  regressores, por exemplo, a quantidade de possíveis modelos de regressão é igual a  $2^{k-1}$ . Com 10 preditores, por exemplo, esta quantidade será igual a 1024. Entretanto, existem algoritmos eficientes tais como o *leaps and bounds* em [39] que tratam adequadamente este problema mesmo quando  $k$  é da ordem de 30 ou 40.

#### Método Stepwise

Este método reduz a necessidade de computação intensiva para selecionar o melhor subconjunto de regressores. O seu propósito é incluir ou excluir um regressor a cada iteração através da verificação de um critério estatístico ou de informação. A seleção pode ser feita através do princípio geral-para-específico ou específico-para-geral. No primeiro caso, o modelo é ajustado com a inclusão de todas as variáveis e posteriormente elimina-se uma delas a cada iteração.

Os métodos descritos decidem sobre a inclusão de variáveis no modelo final utilizando critérios de informação ou critérios estatísticos. Abaixo são apresentados alguns destes critérios.

## Critérios de Informação

### 1. O critério de Akaike (AIC - Akaike Information Criterion)

O Critério de Akaike ([2]) é uma medida de qualidade de ajuste que estima<sup>1</sup> o valor esperado da informação de Kullback-Leibler ( $KL$ ) através de:

$$AIC = -2l(\boldsymbol{\theta}) + 2p \quad (2-9)$$

Em (2-9),  $l(\cdot)$  é a verossimilhança do modelo em questão,  $\boldsymbol{\theta}$  é o vetor de parâmetros e  $p$  é a quantidade de parâmetros independentes do modelo. Este critério penaliza o modelo pela quantidade de parâmetros.

### 2. O critério corrigido de Akaike (AICC - Akaike Information Corrected Criterion)

$$AICC = -2l(\boldsymbol{\theta}) + 2p \frac{T}{T-p-2} \quad (2-10)$$

Como em pequenas amostras o vício na estimação do valor esperado de  $KL$  pode ser drástico, conforme mostrado em [54], os mesmos autores sugeriram uma correção. A modificação em (2-10) em relação ao critério AIC tradicional é representada pelo fator de correção  $\frac{T}{T-p-2}$ .

### 3. O critério bayesiano de informação (BIC - Bayesian Information Criterion)

O BIC ([81]) é, também, um critério assintótico cuja adequação está fortemente relacionada com a magnitude do tamanho de amostra. Em relação à penalização aplicada na quantidade de parâmetros, esta será mais pesada do que a do AIC para amostras pequenas onde ocorre grande perda nos graus de liberdade.

$$BIC = -2l(\boldsymbol{\theta}) + p \log T \quad (2-11)$$

<sup>1</sup>Este representa um estimador assintoticamente não viesado

Em (2-11),  $T$  é a quantidade de observações e as demais quantidades são as mesmas referidas em 2-9. Este critério é assintoticamente consistente.

## Critérios Estatísticos

### 1. O Coeficiente de Determinação $R^2$

Este pode ser considerado o critério de qualidade de ajuste mais utilizado. Mede-se através deste, o percentual de variação de  $y$  explicado pelo conjunto de regressores  $\mathbf{x} = (x_1, x_2, \dots, x_p)$ .

$$R^2 = \frac{SQE}{SQT} \quad (2-12)$$

Em (2-12),  $SQT = \sum_{t=1}^T (y_t - \bar{y})^2$  é a Soma de Quadrados Total e  $SQE = \sum_{t=1}^T (\hat{y}_t - \bar{y})^2$  é a Soma de Quadrados Explicada.

Note que esta medida explora a decomposição

$$SQR = SQT + SQE \quad (2-13)$$

onde  $SQR = \sum_{t=1}^T (\hat{y}_t - y_t)^2$  é a soma dos quadrados dos resíduos.

### 2. O Coeficiente de Determinação Ajustado $R_{adj}^2$

O coeficiente de determinação ajustado em (2-14) adiciona à medida anterior uma penalização pela quantidade de parâmetros.

$$R_{adj}^2 = 1 - (1 - R^2) \frac{(T - 1)}{T - p - 1} \quad (2-14)$$

### 3. A Estatística $C_p$ de Mallow (Cp)

A estatística  $C_p$  de Mallow é uma função de duas somas de quadrados: uma para o modelo completo e outro para o modelo reduzido a  $p$  parâmetros.

$$C_p = \frac{SQR_p}{SQR} - (T - 2p) \quad (2-15)$$

Em (2-15),  $SQR_p$  é a soma dos quadrados dos resíduos sob o modelo reduzido e  $SQR$  é calculada sobre o modelo total. Sob modelo correto, o valor de  $C_p$  se aproxima de  $p$

### 4. O Critério FPE (Final Prediction Error)

Este critério, assim como o AIC e BIC, escolhe o modelo através de uma negociação entre a qualidade do ajuste e a complexidade, caracterizada pelo número total de parâmetros.

$$FPE = \left(\frac{SQR}{T}\right)\left(\frac{T+p}{T-p}\right) \quad (2-16)$$

### 2.1.3

#### Estimação de Parâmetros em Modelos Lineares

Os métodos mais tradicionais para estimar os parâmetros dos modelos descritos acima são:

- Método dos Momentos;
- Método dos Mínimos Quadrados Ordinários;
- Método da Máxima Verossimilhança.

O Método dos Momentos é pouco convencional na estimação dos parâmetros dos modelos de regressão linear. Seu fundamento é igualar momentos amostrais a momentos populacionais, o que geralmente resulta em um sistema de equações, dependendo da dimensão do vetor de parâmetros  $\theta$ . Para o modelo de regressão linear múltipla em (2-4), este método consiste em resolver, com base no dados observados  $(\mathbf{x}_t, y_t)$ , o sistema de  $p + 1$  equações:

$$\sum_{t=1}^T x_{i,t} \varepsilon_t = 0 \quad i = 0, 1, \dots, p$$

onde  $x_{0,t} = 1$  para  $t = 1, \dots, T$ .

O método de Mínimos Quadrados Ordinários (MQO) é o mais popular dentre os listados. Os estimadores possuem expressões analíticas que são facilmente compreendidas e calculadas. O Teorema de Gauss Markov, desde que sejam satisfeitas algumas suposições, garante que dentre os estimadores não viesados e lineares, estes possuem variância mínima <sup>2</sup>.

O método da Máxima Verossimilhança (MV) exige que seja determinada uma distribuição de probabilidades para a componente estocástica do modelo. O princípio da verossimilhança consiste em tratar a função densidade de probabilidade como função de  $\theta$ , vetor de parâmetros.

A suposição de erros normalmente distribuídos faz com que o problema de maximizar a verossimilhança, nos modelos de regressão linear, torne-se equivalente à minimização dos quadrados dos erros.

<sup>2</sup>Resultado que é válido, mesmo se relaxada a suposição de normalidade.

Os estimadores de máxima verossimilhança possuem propriedades ótimas para estimar os parâmetros dos modelos de regressão linear, dentre as quais: consistência, eficiência assintótica e variância mínima.

#### 2.1.4

#### **Especificação de Modelos para Séries Temporais.**

Para os modelos ARMA, caso a série seja estacionária, devem ser especificadas na equação (2-7), as ordens  $p$  e  $q$ . As ferramentas propostas inicialmente foram as funções de autocorrelação e autocorrelação parcial. Seus comportamentos teóricos eram comparados com as estimativas obtidas na amostra. Nas décadas de 70 e 80, diversos métodos alternativos foram propostos para realizar esta tarefa.

Para citar alguns importantes desenvolvimentos; [3] propõe a aplicação do critério AIC para identificar a ordem autoregressiva do modelo. Em [75], uma sequência de testes do tipo Multiplicadores de Lagrange é utilizada para selecionar a ordem de processos ARMA. Neste caso, é demonstrada a consistência da ordem selecionada. Mais recentemente, o uso Redes Neurais Artificiais (RNA) na especificação de modelos ARMA pode ser visto em [38] que aplicam a metodologia para o famoso conjunto de dados *airline*.

### 2.2

#### **Abordagem Econométrica**

Os modelos econométricos têm como objetivo a descrição de relações entre variáveis com base na teoria econômica. Um diferencial destes modelos é a possibilidade de testar empiricamente as hipóteses econômicas, além da racionalidade na interpretação dos parâmetros. Literatura sobre o assunto pode ser encontrada em [44], [36] e [50], dentre outros autores.

Muitas relações não-lineares estão pautadas na Teoria Econômica e, por isto, muitos desenvolvimentos na modelagem não-linear de séries temporais e também dados de painel surgem nesta área. Na sequência são apresentados alguns modelos da literatura econométrica que serão relevantes no presente trabalho.



### 2.2.1 Modelos de Transição Suave

Uma importante vertente econométrica na modelagem não-linear surgiu a partir dos desenvolvimentos em [87], através dos modelos autoregressivos com limiar (TAR-Threshold AutoRegressive). Estes modelos constituem uma importante metodologia para análise não-linear de séries temporais pois utilizam desenvolvimentos na construção de modelos lineares e conseguem reproduzir características dos dados que são encontradas no cotidiano, tais como: ciclos-limite, salto de ressonância, amplitude dependente da frequência e caos.

Uma classe mais geral que engloba os modelos TAR surgiu em [26] e posteriormente em [41]. Esta é formada pelos modelos de (auto)regressão com transição suave (STR - Smooth Transition Regression, STAR - Smooth Transition AutoRegression). O desenvolvimento originou do problema colocado em [5] que tratava a relação entre a chuva e vazão do rio através de um modelo de regressão linear simples cujo intercepto, entretanto, variava suavemente de acordo com o nível de precipitação pluviométrica.

Tal como fora feito com os modelos ARMA propostos em [18], o procedimento de modelagem foi colocado em forma de um ciclo iterativo com as fases de especificação, estimação, diagnóstico e previsão (ver [85, 101]).

A formulação matemática destes modelos para uma série temporal univariada observada nos instantes  $t = 1 - p, p, \dots, -1, 0, 1, \dots, T - 1, T$  é apresentada em (2-17) e (2-18).

$$y_t = G(\mathbf{z}_t, s_t; \Psi) + \varepsilon_t \quad (2-17)$$

$$G(\mathbf{z}_t, s_t; \Psi) = F(s_t, \Phi) \alpha_1' \mathbf{z}_t + [1 - F(s_t, \Phi)] \alpha_2' \mathbf{z}_t \quad (2-18)$$

A formulação acima descreve um modelo com 2 regimes cuja transição entre eles é governada por uma função suave, não-linear,  $F$ , que assume valores no intervalo (0,1).  $F$  é usualmente chamada de função de transição.

O vetor  $\mathbf{z}_t = (1, z_{1,t}, z_{2,t}, \dots, z_{p,t})'$  de dimensão  $(p + 1) \times 1$  contém observações de  $p$  variáveis exógenas e/ou autoregressores. Assume-se, em geral, que o termo aleatório  $\varepsilon_t$  é normalmente distribuído com variância  $\sigma^2$ .

O vetor de parâmetros  $\Psi$  é, por diversas situações, particionado da forma  $\Psi = (\Theta, \Phi)$  na qual o primeiro componente  $\Theta = (\alpha_{01}, \dots, \alpha_{p1}, \alpha_{02}, \dots, \alpha_{p2})$  contém os parâmetros relacionados à parte "linear" do modelo, que por isso, são chamados de parâmetros lineares. O segundo componente  $\Phi$  contém os chamados parâmetros não-lineares, ou seja, aqueles da função de transição e a variância de  $\varepsilon$ .

Em (2-18), o argumento  $s_t$  da função  $F$  é chamado de variável de transição, ou limiar, e, na presente abordagem, é tratado o caso univariado. Usuais escolhas para a variável de transição:

- a)  $s_t = t$ , a transição é regida pelo tempo;
- b)  $s_t = x_{t-d}$ , a transição é regida por uma variável exógena;
- c)  $s_t = y_{t-d}$ , a transição é regida por uma autoregressor.

Deve ser notado que, nas escolhas b) e c), adiciona-se um parâmetro  $d$  ao modelo que é chamado parâmetro de defasagem (*delay parameter*). O modelo TAR(p) em [89] é obtido como um caso particular quando  $F$  é uma função indicadora do tipo:

$$F(.) = \begin{cases} 1 & \text{se } s_t \leq c \\ 0 & \text{se } s_t > c \end{cases} \quad (2-19)$$

Na situação em que  $s_t = y_{t-d}$ , este modelo é denominado SETAR (Self-Exciting Threshold Autoregressive, [89]) e pode ser considerado como o predecessor do modelo STAR.

Em ambas situações envolvendo os modelos TAR e SETAR, o limiar entre os dois regimes é abrupto e determinado por  $c$ , o parâmetro de limiar.

Uma das grandes vantagens na utilização dos modelos de transição suave é a possibilidade de especificar a função de transição de forma a evitar o problema da busca por um limiar "rígido" entre os regimes. Uma das principais opções à escolha da função que governará a transição entre os dois regimes é a função logística:

$$F(s_t; \gamma, c) = \frac{e^{-\gamma(s_t-c)}}{1 + e^{-\gamma(s_t-c)}} \quad (2-20)$$

Desde a primeira referência sobre os modelos de transição suave em [5], tem sido mostrado que a função logística é uma escolha natural para reger a transição suave entre dois regimes. Os autores utilizaram a função tangente hiperbólica cuja forma funcional está bem próxima da função logística. Discussão sobre este tópico pode ser encontrada em [66], que menciona a idéia de usar a função logística, mas não a aplica. Outras referências são [41] e [85].

Ao escolher (4-2) como função de transição, o vetor de parâmetros não-lineares será formado por  $\Phi = (\gamma, c)$ . O parâmetro  $\gamma$  é o responsável pelo grau de suavidade da função de transição. É importante notar que na situação em que  $\gamma \rightarrow \infty$ , (4-2) se aproxima de uma função do tipo degrau e, por conseqüência, o modelo SETAR torna-se um caso particular quando  $s_t = y_{t-d}$ .

O modelo resultante da substituição de (4-2) em (2-18) é denominado LSTAR (Logistic Smooth Transition AutoRegression), que é abordado com maior profundidade em [64]. Alguns problemas que cercam a estimação dos parâmetros deste modelo serão tratados no Capítulo 4.

Outra possibilidade para a especificação de um modelo com transição suave entre dois regimes é considerar a função exponencial em (2-21).

$$F(s_t; \gamma, c) = 1 - e^{-\gamma(s_t - c)^2}, \gamma > 0 \quad (2-21)$$

Uma das vantagens na utilização da função exponencial preferivelmente à função logística é a possibilidade de associar o mesmo regime a valores baixos e altos da variável de transição  $s_t$ . Estes modelos são denominados ESTAR (Exponential Smooth Transition AutoRegression) e representam uma generalização dos modelos EAR (Exponential Autoregressive) em [43]. Por outro lado, o modelo SETAR não se torna um caso particular quando  $\gamma \rightarrow \infty$ . Este fato pode ser contornado com o uso da função logística quadrática descrita em [57]:

$$F(s_t; \gamma, c_1, c_2) = \frac{1}{1 + e^{-\gamma[(s_t - c_1)(s_t - c_2)]}}, c_1 < c_2, \gamma > 0 \quad (2-22)$$

que proporciona, no limite, um modelo SETAR com 3 regimes ; 2 regimes nas situações em que  $s_t < c_1$  e  $s_t > c_2$ , respectivamente, e um terceiro regime na situação em que  $s_t > c_1$  e  $s_t < c_2$ .

van Dijk & Franses [103] propõem a extensão do modelo STAR de forma que este possa abrigar múltiplos regimes. Estes modelos receberam a denominação MRSTAR(Multiple Regime Smooth Transition AutoRegression). Para uma revisão sobre demais desenvolvimentos obtidos a partir da idéia de modelos de transição suave, é recomendável consultar [101].

### **Ciclo da Modelagem STAR**

Terasvirta [85] discute o processo de construção dos modelos STAR em termos de um ciclo iterativo tal como é feito nos modelos de Box & Jenkins [18]. Os passos deste ciclo são apresentados na sequência :

1. Especificação de um modelo autoregressivo de ordem  $p$ .
2. Teste da hipótese de linearidade contra uma alternativa da família STAR. Em caso de rejeição da hipótese nula, é feita a seleção da variável de transição e a forma da função de transição.
3. Estimação dos parâmetros do modelo STAR selecionado.

4. Realização do diagnóstico do modelo.
5. Re-especificação do modelo de acordo com os resultados do diagnóstico.
6. Utilização do modelo com propósitos descritivos ou de previsão.

### **Estimação dos parâmetros dos modelos STAR**

A estimação dos parâmetros nos modelos STAR é feita após a seleção da variável de transição  $s_t$  e da função de transição  $F(\cdot)$ . O vetor é estimado por mínimos quadrados não-lineares e sob certas condições de regularidade ([96]) os estimadores são consistentes e assintoticamente distribuídos de acordo com a distribuição normal. Os procedimentos não-lineares de otimização podem ser encontrados em [76, 48, 51].

### **Diagnósticos do ajuste dos modelos STAR**

Alguns testes são propostos para verificar a qualidade do ajuste do modelo não-linear :

- teste para constância dos parâmetros ao longo do tempo;
- teste de correlação serial dos resíduos;
- teste para verificação de não-linearidade remanescente [35];
- teste de constância da variância (homocedasticidade) ao longo do tempo contra alternativa de evolução suave ([22]).

## **2.2.2 Modelos GARCH**

A modelagem GARCH (Generalized Autoregressive Conditional Heteroskedasticity [37, 19]) surgiu sob a motivação de projetar a variância da série ao invés do nível. A incorporação deste fato estilizado na estrutura do modelo tem grande aplicação na previsão de séries financeiras.

Os primeiros desenvolvimentos destes modelos estão associados com a estrutura ARCH (Autoregressive Conditional Heteroskedasticity) proposta em [37]. Esta abordagem permite a modelagem simultânea da média e variância de uma série temporal.

Considere um processo autoregressivo de ordem  $p$ :

$$y_t = \theta_0 + \sum_{i=1}^p \theta_i y_{t-i} + u_t \quad (2-23)$$

cuja variância condicional possa variar com o tempo. Tal fato pode ser contemplado sob a suposição de que o quadrado de  $u$  siga um processo AR( $m$ ) conforme:

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \dots + \alpha_m u_{t-m}^2 + w_t \quad (2-24)$$

supondo que  $w_t, t = 1, \dots, T$  é ruído branco. Isto implica que  $E(w_t) = 0$ ,  $Var(w_t) = \lambda^2$  e  $E(w_t, w_s) = 0, t \neq s$ . O processo  $u_t$  descrito em (2-23) e (2-24) é chamado de ARCH( $m$ ). É comum adotar, para este processo, a representação alternativa em (2-25).

$$u_t = \sqrt{h_t} v_t \quad (2-25)$$

onde  $h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \dots + \alpha_m u_{t-m}^2$  é a parte determinística da equação (2-24) e  $v_t$  é uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média zero e variância igual a 1,

$$E(v_t) = 0$$

e

$$Var(v_t) = 1.$$

Se  $u_t^2$  é um processo fracamente estacionário e sob outras condições que podem ser vistas em [20], a variância incondicional de  $u_t$  é obtida a partir de:

$$E(u_t^2) = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \dots - \alpha_m} \quad (2-26)$$

O modelo ARCH( $m$ ) é generalizado ao admitir-se que a variância condicional  $h_t$  dependa de uma quantidade infinita de defasagens.

$$h_t = \sum_{i=0}^{\infty} \pi_i u_{t-i}^2 = \pi_0 + \pi(\mathbf{B})u_t^2 \quad (2-27)$$

$$\pi(\mathbf{B}) = \sum_{j=1}^{\infty} \pi_j B^j$$

$$B^k u_t^2 = u_{t-k}^2$$

Representando  $\pi(\mathbf{B})$  como a razão entre dois polinômios de ordem finita, é obtida para  $h_t$  uma forma funcional similar à dos modelos ARMA conforme a equação (2-28).

$$h_t = \alpha'_0 + \delta_1 h_{t-1} + \dots + \delta_r h_{t-r} + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \dots + \alpha_m u_{t-m}^2 \quad (2-28)$$

$$\alpha'_0 = (1 - \delta_1 - \dots - \delta_r) \alpha_0$$

Este modelo recebe a denominação GARCH( $r, m$ ) e as suas propriedades são discutidas em [19].

## 2.3 Abordagem nas Ciências da Computação

### Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) formam uma classe de modelos não-lineares que têm sido aplicada com muito sucesso em diversas áreas do conhecimento. Exemplos podem ser encontrados na psicologia, ciências da computação, engenharia, linguística e economia.

Esta é uma dentre muitas outras técnicas surgidas nas áreas de Inteligência Artificial e Aprendizado por Máquina. Sua gênese está presente em [67, 49, 80].

As RNA procuram modelar sistemas imitando o comportamento biológico dos neurônios. Do ponto de vista matemático, são funções polinomiais.

Embora tenha sido utilizada inicialmente nas Ciências da Computação, sua equivalência com métodos estatísticos já consagrados veio logo à tona. Para tal, cita-se como exemplo o trabalho encontrado em [33] que utiliza a estrutura de uma rede neural para representar um modelo ARMA não estacionário, e também o desenvolvimento em [70] que conjuga ferramentas da inferência clássica com elementos das RNA para especificar modelos não-lineares para análise de séries temporais.

O grande potencial do uso das redes neurais para a modelagem não-linear pode ser credenciado ao fato desta ser um aproximador universal, resultado derivado do Teorema de Weierstrass cujo enunciado é apresentado em 2.1.

**Teorema 2.1** *Suponha  $f$  uma função contínua em  $\mathbb{R}$ , definida no intervalo  $[a, b]$ . Para cada  $\epsilon > 0$ , existe uma função polinomial  $p$  com coeficientes reais tais que para todo  $x \in [a, b]$ ,  $|f(x) - p(x)| < \epsilon$ .*

Este teorema é importante ponto de partida para mostrar que a RNA pode ser especificada de forma a aproximar qualquer função do tipo Borel-mensurável, com qualquer grau de acurácia desejado (ver [53]).

### Formulação Matemática

As RNA são usualmente representadas pela arquitetura de interconexão entre os neurônios que a constituem. Entretanto, prefere-se ilustrar um caso particular, as redes com uma camada escondida cujos parâmetros são comumente ajustados pelo método *feed-forward*.

$$y_t = \beta_0 + \sum_{j=1}^h \beta_j F(\gamma_j' \mathbf{w}_t - c_j) \quad (2-29)$$

A equação (2-29) representa uma RNA com  $k$  neurônios na camada de entrada e  $h$  neurônios na camada escondida. A camada de saída apresenta apenas um neurônio. Em (2-29)  $\mathbf{w}_t$  é um vetor de dimensão  $k \times 1$  que contém  $k$  variáveis de entrada. Estas variáveis são combinadas linearmente através dos pesos que constituem o vetor  $\gamma_j = (\gamma_{1j}, \dots, \gamma_{kj})$ , de mesma dimensão. A combinação linear das variáveis de entrada, subtraída de  $c_j$  forma o argumento de  $F(\cdot)$ , chamada função de ativação. A escolha mais comum para esta função é a logística. O parâmetro  $c_j$ ,  $j = 1, \dots, h$  determina o limiar da ativação do  $j$ -ésimo neurônio da camada escondida. As saídas da camada escondida, que possui  $h$  neurônios, serão combinadas linearmente utilizando os pesos  $\beta_1, \dots, \beta_h$ .

Note que o lado direito da expressão mostrada em (2-29) não contém componentes estocásticas, ao contrário dos tradicionais modelos de regressão.

### Lógica Nebulosa

A lógica nebulosa ([98]) é uma técnica de inteligência artificial que formula um sistema não-linear de mapeamento de um vetor de entradas em uma resposta, em geral escalar, que torna possível a incorporação tanto do conhecimento objetivo quanto do conhecimento subjetivo.

O conceito de incerteza está presente em grande parte dos métodos estatísticos que auxiliam a tomada de decisão. Entretanto, a lógica nebulosa lida também com a imprecisão que caracteriza algumas medidas. Caso um indivíduo informe apenas que sua renda é baixa, ao invés do valor monetário, tal fato limita o uso de ferramentas estatísticas, porém a lógica nebulosa possui procedimentos para lidar com esta situação.

Alguns dos desenvolvimentos desta tese possuem uma forte correspondência com elementos da lógica nebulosa, em especial, com sistemas de inferência nebulosa. A inferência nebulosa está diretamente ligada com os conceitos de conjuntos nebulosos, regras do tipo se-então e raciocínio nebuloso.

### Sistemas de Inferência Nebulosa (FIS - Fuzzy Inference Systems)

Um sistema de inferência nebulosa é formado por 3 elementos:

- base de regras (nebulosas);
- funções de pertinência;
- procedimentos inferenciais.

Um mecanismo de inferência nebulosa codifica uma base de regras do tipo:

**SE** um conjunto de condições é satisfeito **ENTÃO** um conjunto de consequentes é inferido”.

A Figura 2.1 ilustra a seqüência de operações realizadas durante a inferência nebulosa.

Entrada	$\mathbf{x} = (x_1, x_2, \dots, x_T)$ $\mathbf{y} = (y_1, y_2, \dots, y_T)$
Regras	$\text{SE } x_i \in \mathbb{A}_1 \text{ ENTÃO } y_i \in \mathbb{B}_1$ $\dots$ $\text{SE } x_i \in \mathbb{A}_i \text{ ENTÃO } y_i \in \mathbb{B}_i$ $\dots$ $\text{SE } x_i \in \mathbb{A}_N \text{ ENTÃO } y_i \in \mathbb{B}_N$
Saída	$\hat{y}_i = d(x_i)$

Figura 2.1: Sistema de Inferência Nebulosa

Após o sistema receber os vetores de entrada, que no contexto de regressão representam as variáveis explanatórias e a variável resposta, são formuladas regras (nebulosas) que envolvem um antecedente (SE) e um consequente (ENTÃO).

O que torna nebuloso este processo é o fato de existir uma função  $\mu_j(x_i)$  que irá medir a pertinência de  $x_i$  ao conjunto  $\mathbb{A}_j$ ,  $j = 1, 2, \dots, N$ . Da mesma forma,  $y_i$  irá pertencer ao conjunto  $\mathbb{B}_j$  com pertinência  $\delta_j(y_i)$ . Em geral, as funções de pertinência  $\mu_j(\cdot)$  e  $\delta_j(\cdot)$  assumem valores no intervalo  $[0, 1]$ .

A etapa final da inferência nebulosa é chamada de *defuzzificação*. Na Figura 2.1, a defuzzificação está representada pela função  $d(\cdot)$ . Nela é escolhido um valor (predição) para representar o conjunto nebuloso.



## 2.4 Abordagens Híbridas

Diante da potencialidade das RNA e a interpretabilidade dos modelos econométricos, surgiram recentemente propostas de modelos híbridos que conjugam elementos da duas abordagens.

### O modelo AR-ANN (AutoRegressive Artificial Neural Network)

A combinação da teoria clássica na análise de séries temporais com métodos emergentes é bem ilustrada nos modelos AR-ANN em [99].

$$y_t = G(\mathbf{x}_t; \Psi) + \varepsilon_t \quad (2-30)$$

$$G(\mathbf{x}_t; \Psi) = \beta_0 + \alpha' \mathbf{x}_t + \sum_{i=1}^h \beta_i F(\omega_i' \mathbf{x}_t - \theta_i)$$

O modelo exibido em (2-30) descreve um processo autoregressivo cujo nível se altera dependendo da posição do vetor de autoregressores  $\mathbf{x}_t = (y_{t-1}, \dots, y_{t-p})$  no  $\mathbb{R}^p$ .

A saída de uma rede neural com a mesma especificação da apresentada em (2-29) irá determinar a mudança no nível do processo autoregressivo no instante  $t$ .

O vetor  $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_p)$  é formado por  $p$  coeficientes autoregressivos que são constantes ao longo do tempo. A função  $F(\cdot)$  possui as mesmas características já discutidas para a equação (2-29) e a escolha usual também é a função logística.

Ao contrário das RNA, é incluída uma componente estocástica  $\varepsilon$  normalmente distribuída com média 0 e variância  $\sigma^2$ . Isto possibilita que os parâmetros, não somente do modelo autoregressivo como também os da RNA sejam estimados utilizando o princípio da máxima verossimilhança.

Entretanto, o modelo acima, sob vários aspectos, não é identificável conforme discutido em [55] e [79]. Isto significa que caso não haja algumas restrições, alguns dos parâmetros podem variar livremente sem que isto altere a função de verossimilhança. Estabelecidas as devidas restrições e sob certas condições de regularidade, discutidas em [79], o estimador de máxima verossimilhança para o vetor de parâmetros do modelo é consistente e assintoticamente normal.

## O modelo NCSTAR

O modelo NCSTAR em [69] é uma generalização dos modelos STAR e AR-ANN. Neste enfoque, os parâmetros autoregressivos variam ao longo do tempo conforme a saída de uma rede neural com a mesma arquitetura considerada em (2-29). Este é especificado a partir de uma sequência de testes do tipo ML. A formulação matemática deste modelo é obtida a partir de:

$$y_t = G(\mathbf{z}_t, \mathbf{x}_t; \Psi) + \varepsilon_t \quad (2-31)$$

$$G(\mathbf{z}_t, \mathbf{x}_t; \Psi) = \alpha' \mathbf{z}_t + \sum_{j=1}^h \lambda_j' \mathbf{z}_t F(\omega_j' \mathbf{x}_t - c_j)$$

Em (2-31), a série  $y_t$  é descrita como uma função não-linear das defasagens e/ou variáveis exógenas que irão compor o vetor  $\mathbf{z}_t$ .  $\{\varepsilon_t, t = 1 \dots, T\}$  é uma seqüência de variáveis aleatórias independentes e normalmente distribuídas.

A especificação do modelo NCSTAR segue o princípio específico-geral. Inicialmente, efetua-se um teste do tipo ML (Multiplicadores de Lagrange) para verificar a hipótese nula associada a um modelo linear contra a alternativa do modelo NCSTAR mais simples. Caso a hipótese nula seja rejeitada, aplica-se daí para a frente uma seqüência de testes que colocam na hipótese alternativa modelos com grau de complexidade imediatamente superior ao da hipótese nula.

Após a especificação, procede-se a estimação cuja metodologia é descrita em [69]. Supondo que os parâmetros foram estimados consistentemente, os modelos são diagnosticados com as ferramentas propostas em [71]. Dentre os aspectos a serem investigados estão: constância dos parâmetros, independência serial e homocedasticidade<sup>3</sup>.

<sup>3</sup>Testa-se neste caso a hipótese de que a variância muda suavemente ao longo do tempo