ALEXANDRE ZANINI

**REDES NEURAIS E REGRESSÃO DINÂMICA: UM MODELO HÍBRIDO PARA
PREVISÃO DE CURTO PRAZO DA DEMANDA DE GASOLINA AUTOMOTIVA
NO BRASIL**

**Dissertação apresentada ao Departamento
de Engenharia Elétrica da PUC/Rio como
parte dos requisitos para a obtenção do
título de Mestre em Engenharia Elétrica:
Teoria de Controle e Estatística**

Orientador: Reinaldo Castro Souza

Departamento de Engenharia Elétrica

Pontifícia Universidade Católica do Rio de Janeiro

Rio de Janeiro, 28 de abril de 2000.

AGRADECIMENTOS

A Deus pela minha existência.

A meus pais, Adelina e Alberto, que com seu amor me deram a vida e com sua dedicação perpetuaram este amor possibilitando tudo o que tenho e sou.

A minha família que com seu carinho e amor nunca me deixaram sentir só.

A mulher da minha vida, Renata, razão primeira de tudo que me faz feliz.

Ao Professor Reinaldo pelas oportunidades oferecidas.

Ao Professor Pedreira pela atenção, sugestões e todo incentivo que me foi dado.

À amiga Ana Paula, por ter sido, acima de tudo, amiga.

Ao mestre e sempre amigo Rogério por toda a força que sempre me deu e por acreditar em mim em todos os instantes.

À amiga, de todas as horas, Fernanda por estar sempre pronta para me ouvir, chorar e sorrir comigo.

Aos amigos Leonardo Macrini e Marcio por dividirem comigo as aflições e por toda a ajuda prestada.

Aos amigos Peterson, Neimar, Joel, Elma, José Américo, Eduardo Campos, Lúcio e Mayte pelo convívio, apoio e amizade.

Ao José Lopes, da ANP, pela atenção que me foi dedicada sempre que precisei.

A Eduardo Correia e João Mattos, da Petrobrás, pelo material bibliográfico cedido.

A todos aqueles que torceram para que este trabalho pudesse ser realizado.

RESUMO

Nesta dissertação é desenvolvido um modelo para previsão de curto prazo da demanda mensal de gasolina automotiva no Brasil. A metodologia usada consiste em, a partir de uma análise exploratória dos dados, construir um modelo usando uma estratégia *bottom-up*, ou seja, parte-se de um modelo simples e processa-se seu refinamento até encontrar um modelo apropriado que mais se adequa à realidade. Partiu-se então de um modelo autoprojeto indo até uma formulação de redes neurais, passando por um modelo de regressão dinâmica.

Os modelos são então comparados segundo alguns critérios, basicamente no que tange à sua eficiência preditiva. Conclui-se ao final sobre a eficiência de se conjugar modelos estatísticos clássicos (como Box & Jenkins e regressão dinâmica) com as chamadas Redes Neurais que, por sua vez, otimizaram os resultados das previsões. Isto é altamente desejável na modelagem de séries temporais e, em particular, neste trabalho, na previsão de curto prazo de gasolina automotiva.

ABSTRACT

In this dissertation a short term model to forecast automotive gasoline demand in Brazil is proposed. From the methodology point of view, data is analyzed and a model using a bottom-up strategy is developed. In other words, a simple model is improved step by step until a proper model that fits well the reality is found. Departing from a univariate model it ends up in a neural network formulation, passing through dynamic regression models.

The models obtained in this scheme are compared according to some criterion, mainly forecast accuracy. We conclude, that the efficiency of putting together classical statistics models (such as Box & Jenkins and dynamic regression) and neural networks improve the forecasting results. This result is highly desirable in modeling time series and, particularly, to the short term forecast of automotive gasoline , object of this dissertation.

SUMÁRIO

LISTA DE QUADROS	V
LISTA DE FIGURAS	V
1 - INTRODUÇÃO	1
2 - SÉRIES TEMPORAIS	6
2.1 - Conceitos básicos	6
2.1.1 - Série Temporal	6
2.1.2 - Processo Estocástico	6
2.1.3 - Domínio do Tempo x Domínio da Frequência	8
2.2 - Modelos de Previsão	10
3 - MODELOS BOX & JENKINS	13
3.1 - Fundamento teórico	13
3.2 - A Metodologia Box & Jenkins	16
3.3 - Modelos SARIMA	18
4 - REGRESSÃO DINÂMICA	20
4.1 - Introdução	20
4.2 - Conceituação	21
4.3 - Estrutura dos Modelos de Regressão Dinâmica	23
4.3.1 - Modelos de Regressão Cochrane-Orcutt Generalizados	25
4.4 - Construção de modelos de regressão dinâmica	26
4.5 - Testes usados nos modelos de regressão dinâmica	29
4.5.1 - Testes de verificação da “dinâmica” do modelo	30
4.5.2 - Testes para a especificação das variáveis causais	32
4.5.3 - Testes baseados na autocorrelação dos resíduos	33
5 - REDES NEURAIS	35
5.1 - Conceituação	35
5.2 - Arquitetura Neural	36
5.3 - Um modelo de neurônio	39
5.4 - Modelagem e aprendizado em redes neurais	41
5.4.1 - Aprendizado supervisionado e o algoritmo de <i>Backpropagation</i>	43

5.4.2 - Generalização	44
6 - MODELOS LINEARES PARA A DEMANDA DE GASOLINA AUTOMOTIVA NO BRASIL	46
6.1 - Modelo Box & Jenkins	46
6.1.1 - Identificação do modelo	46
6.2 - O modelo de regressão dinâmica	53
7 – MODELOS NÃO-LINEARES PARA A DEMANDA DE GASOLINA AUTOMOTIVA NO BRASIL: AS REDES NEURAIIS	61
7.1. - Modelo de Redes Neurais I	61
7.2 - Modelo de Redes Neurais II: O Modelo Híbrido	63
8 – RESULTADOS NUMÉRICOS	66
9 - CONCLUSÃO	73
10 - BIBLIOGRAFIA	76

LISTA DE QUADROS

Quadro 01: Resumo das características teóricas da FAC e da FACP dos modelos AR(p), MA(q) e ARMA (p,q).....	17
Quadro 02: Variação da demanda de gasolina automotiva no Brasil no 2º semestre de 1999	71

LISTA DE FIGURAS

Figura 4.1: Construção de um modelo de regressão dinâmica	28
Figura 5.1: Rede <i>Feedforward</i>	37
Figura 5.2: Modelo de Neurônio	39
Figura 5.3: Sigmóide	41
Figura 6.1: Gráfico da Demanda de Gasolina Automotiva no Brasil (m ³).....	47
Figura 6.2: Função de Autocorrelação (FAC).....	47
Figura 6.3: Função de Autocorrelação Parcial (FACP).....	48
Figura 6.4: Função de Autocorrelação (FAC) após a 1ª diferença	49
Figura 6.5: Função de Autocorrelação Parcial (FACP) após a 1ª diferença.....	49
Figura 6.6: Função de Autocorrelação (FAC) dos Erros.....	52
Figura 6.7: Função de Autocorrelação (FAC) dos erros do modelo RD1	57
Figura 6.8: Gráfico da variação percentual da demanda de gasolina automotiva: 1997-1999.....	58
Figura 6.9: Função de Autocorrelação (FAC) dos erros do modelo RD2	60
Figura 7.1: Par entrada – saída desejada para RN1	62
Figura 7.2: Par entrada – saída desejada para RN2	65
Figura 8.1: Comparação do MAPE (<i>In sample</i>).....	69
Figura 8.2: Comparação do U-THEIL (<i>In sample</i>)	69
Figura 8.3: Comparação do MAPE (<i>Out of sample</i>)	70
Figura 8.4: Comparação do U-THEIL (<i>Out of sample</i>).....	70

1 - INTRODUÇÃO

A década de 90 marcou, no Brasil, os primeiros movimentos para levar o País a uma economia mais aberta, procurando aumentar sua eficiência e competitividade com maior relevância às forças de mercado e à iniciativa privada.

Os setores de energia em geral, e de petróleo e gás em particular, vêm passando por processos de reestruturação institucional abrangentes e complexos. Ao processo de desestatização, somam-se a flexibilização do monopólio do petróleo e a entrada de novos agentes. Neste novo contexto, a concorrência passa a ser parte integrante e de grande relevância.

Em um novo modelo de desenvolvimento econômico e social, o Estado foi levado a uma mudança de atuação. De um Estado empreendedor, o Estado brasileiro passou a exercer, cada vez mais, o papel de regulador e fiscalizador. Neste sentido, são criados e fortalecidos órgãos de regulação e aos poucos se consolida um processo cultural e político envolvendo as diversas esferas de governo, empresas, a imprensa e os consumidores.

Uma mudança institucional de tal magnitude canaliza a atenção dos agentes econômicos para as oportunidades de posicionamento em um novo mercado emergente. Desta forma, o Brasil, em função de seu porte e potencial do mercado, consegue atrair investimentos, a despeito das incertezas dos cenários internacional e nacional.

Sobre o porte do mercado, ressalta-se que o Brasil encontra-se entre os vinte países maiores produtores de petróleo e entre os dez com maior capacidade instalada de refino. A despeito deste porte, o País importa uma parcela considerável de suas necessidades de petróleo e de seus derivados.

A reestruturação de um setor desta magnitude envolve grande complexidade, iniciada com a efetiva flexibilização do monopólio e abertura do setor a outros agentes. Neste sentido, muitas empresas estrangeiras instalaram-se no Brasil nos últimos anos e espera-se que o processo continue. O contexto é, segundo o governo brasileiro, promissor no que se refere ao crescimento dos investimentos em exploração, produção e refino de petróleo e da participação do gás natural na matriz energética. Ainda que sinais de alerta estejam presentes em função dos preços do petróleo e de incertezas sobre o recrudescimento da crise econômica, o petróleo mantém-se como importante vetor de crescimento econômico.

Esta atual fase de abertura da indústria nacional de petróleo e gás natural no Brasil e as tarefas regulatórias do novo órgão regulador do setor, a Agência Nacional do Petróleo (ANP), implicam a necessidade de gerar projeções confiáveis de demanda de combustíveis, em particular, os derivados de petróleo, gás natural e álcool.

Espera-se que uma melhor compreensão do funcionamento do mercado de petróleo e gás natural fornecerá subsídios para estratégias de planejamento e gestão por parte tanto das empresas deste setor quanto do órgão regulador.

Desta forma, nesta dissertação de mestrado, desenvolve-se um modelo de previsão para demanda mensal de gasolina automotiva no Brasil. Este trabalho é resultante de um convênio firmado entre a ANP e a Pontifícia Universidade Católica do Rio de Janeiro onde, entre outros objetivos, procura-se elaborar a modelagem e previsão de oito combustíveis a saber: gasolina automotiva, óleo diesel, álcool hidratado, gás liquefeito de petróleo (GLP), gasolina de aviação, querosene de aviação, óleo combustível e querosene.

Além de analisar o comportamento da série de demanda de gasolina automotiva no Brasil, procurando-se fazer projeções confiáveis que permitam uma melhor compreensão deste mercado,

a escolha pela modelagem desta série específica como tema deste trabalho baseia-se fundamentalmente no seu difícil comportamento ao longo dos anos bem como na existência de poucos trabalhos, no Brasil, com o mesmo intuito.

No início da década de oitenta, tem-se o trabalho de Suyama (1981) que fez uma abordagem econométrica da influência do preço sobre a demanda anual de gasolina para o transporte individualizado, mais precisamente para os “carros de passeio”. O autor dedica atenção especial ao estudo da eficiência média dos carros, tratada, não como um simples coeficiente tecnológico, exógeno ao sistema econômico, mas uma variável sujeita a parâmetros econômicos como preço dos veículos e renda.

Barros e Ferreira (1982) também desenvolveram um modelo econométrico para a demanda de gasolina para os automóveis de passeio. Foram feitos estudos de como a probabilidade de sucateamento de um veículo varia com a sua idade e estimou-se a frota de veículos por categoria e tipo de combustível utilizado. Com base na frota estimada e na definição e estimação da eficiência média dos veículos novos a gasolina, avaliou-se o consumo anual de gasolina e a quilometragem média.

Já no trabalho de Souza *et al* (1984), fez-se uso da metodologia de Função de Transferência¹, desenvolvida por Box & Jenkins, para estimar o consumo mensal de gasolina. Para este fim, Souza *et al* usaram séries de preço e uma *proxy* de renda² como variáveis explicativas.

¹ Box & Jenkins (1994).

² Os autores usaram o Imposto de Circulação de Mercadoria (ICM) como *proxy* para a variável renda.

Buonfiglio (1988), desenvolveu também um modelo econométrico para previsão da demanda anual de gasolina automotiva e álcool hidratado. O autor identificou o PIB per capita, o preço da gasolina e a idade média da frota como principais fatores que afetam o consumo de gasolina.

Na década de noventa, o trabalho de Correia e Mattos (1999), que trabalham com um modelo estrutural³, conjuga a frota de veículos (desagregada por categoria e combustível) e o consumo unitário (consumo/veículo/ano), este por sua vez sendo correlacionado econometricamente com o preço do combustível e o consumo unitário defasado. Foram construídas funções para a distância média percorrida pelos veículos (Km/veículo/ano) e estimadas as frotas e eficiências médias (Km/litro) dos veículos para cada categoria e combustível⁴.

Nesta dissertação, a estratégia de modelagem utilizada foi partir de um modelo univariado ou autoprojetivo [(Montgomery *et al*, 1990) e (Souza & Camargo, 1996)] , ou seja, de um modelo que não faz uso de outras variáveis que não a própria série em estudo, para verificar se a construção de um modelo embasado nas informações extraídas somente da própria série já não seria suficiente para fazer as projeções de forma ótima.

Posteriormente, elaborou-se um modelo de regressão dinâmica [(Goodrich & Stellwagen, 1999); (Barros & Souza, 1995)], objetivando-se explicar o comportamento da série de demanda de gasolina através de outras variáveis denominadas exógenas ou causais. A idéia foi a de verificar se novas causalidades poderiam melhorar a qualidade das previsões.

³ Correia e Mattos (1999) seguem a linha de raciocínio de autores como Kouris (1983) e Eltony (1991) que propõem modelos econométricos desagregados, construindo funções para o estoque de veículos por domicílio, as vendas de veículos novos e sua eficiência, a distância média percorrida e o consumo unitário dos veículos.

⁴ Buonfiglio (1988) propõe um modelo similar, mas agregado, incorporando a idade média da frota e a renda per capita na estimação da distância percorrida pelo veículo.

Pode-se perceber que esta estratégia de modelagem abrange modelos econométricos e técnicas de séries temporais, particularmente modelos ARIMA (Box & Jenkins, 1994) e de regressão dinâmica. Ressalta-se, por sua vez, que as técnicas utilizadas pressupõem linearidade nas causalidades.

Sabe-se, entretanto, que o mercado apresenta, muitas das vezes, um comportamento não linear. Neste caso, aplicou-se uma técnica de computação intensiva, redes neurais (Haykin 1999), de forma a ajudar a modelar estas não linearidades, auxiliando, portanto, na construção de modelos que se aproximem ao máximo da realidade, ou seja, de modelos que levem à minimização dos erros de previsão.

Feitas as consideração acima, salienta-se que, além desta Introdução, esta dissertação encontra-se dividida em mais sete capítulos. No segundo, faz-se uma abordagem de alguns conceitos básicos do tratamento de séries temporais. Estes conceitos são importantes pois serão citados em várias partes da dissertação. Frente à técnica de modelagem deste trabalho, nos capítulos três, quatro e cinco, respectivamente, aborda-se, de forma geral, a metodologia dos modelos de Box & Jenkins (que nos fornecerá o primeiro modelo ajustado para a demanda de gasolina automotiva), de regressão dinâmica (que originará um modelo causal com estimações das elasticidades da demanda) e de redes neurais (que implicarão na construção de mais dois modelos). No sexto capítulo são apresentados os modelos lineares estimados para a demanda mensal de gasolina automotiva, ou seja, o modelo ARIMA e os modelos causais, estimados por regressão dinâmica. Da mesma forma, no sétimo capítulo, apresentar-se-á os modelos não lineares, ou seja, serão apresentados as redes neurais para a previsão mensal da demanda de gasolina. Por fim, no capítulo oito, faz-se uma análise comparativa dos resultados das previsões geradas pelos modelos, partindo-se para a conclusão no capítulo nove.

2 - SÉRIES TEMPORAIS

2.1 - Conceitos básicos

2.1.1 - Série Temporal

Uma série temporal é um conjunto de observações de uma dada variável, ordenadas segundo o parâmetro tempo, geralmente em intervalos equidistantes. Se Z_t representa o valor da variável aleatória Z no instante t , denota-se a série temporal por $Z_1, Z_2, \dots, Z_N$ onde N é o tamanho da série ou número de observações seriais da variável. As séries temporais podem ser classificadas em discretas, contínuas, determinísticas, estocásticas, multivariadas (discretas ou contínuas) e multidimensionais. Para séries temporais discretas as periodizações de coleta de dados usuais são dados diários, semanais, mensais e anuais (Souza & Camargo, 1996).

É importante salientar que nem todo evento observado no tempo consiste numa série temporal. Exemplo mais claro disto são os jogos de loteria. Isto significa que, saber o resultado da Sena hoje e também os resultados anteriores, por exemplo, não nos permite ajustar um modelo estatístico para prever os próximos resultados. Portanto, para caracterizar uma série temporal é necessário que as observações apresentem uma dependência serial, isto é, sejam dependentes no tempo.

2.1.2 - Processo Estocástico

Um processo estocástico caracteriza-se por uma seqüência no tempo de variáveis aleatórias, ou seja, um modelo que descreve a estrutura de probabilidade de uma seqüência de observações é

chamado de processo estocástico. Desta forma, processos estocásticos são sistemas que evoluem no tempo e/ou no espaço de acordo com leis probabilísticas.

Um processo estocástico é uma família $Z = \{Z(t), t \in N\}$ tal que para cada $t \in \mathbf{R}$, $Z(t)$ é uma variável aleatória. Se $N \equiv Z = \{1, \dots, t\}$, diz-se que o processo é de parâmetro discreto, denotando por Z_t . Se $N \in \mathbf{R}$, diz-se que o processo é de parâmetro contínuo, denotando-se por $Z(t)$.

Suponha que Z_t origina-se de um experimento que pode ser repetido sob condições idênticas; a cada experimento obtém-se um registro dos valores de Z_t no tempo. A coleção de todos os registros possíveis, denomina-se *ensemble*. Cada registro particular é uma realização do processo.

Pela definição de série temporal feita no item anterior, conclui-se que uma série temporal pode ser interpretada como uma parte da trajetória ou de uma realização parcial do processo estocástico, isto é, uma amostra finita de observações no tempo.

Segundo Souza & Camargo (1996), um processo estocástico está estatisticamente determinado quando se conhece suas funções de distribuição até a n -ésima ordem. Na prática, ocorrem situações problemáticas por não se conhecer todas as funções de distribuição até a n -ésima ordem e por ter-se apenas uma realização do processo estocástico, a partir da qual deseja-se inferir características do mecanismo gerador da série. Para superar essas dificuldades assumem-se duas restrições: Estacionaridade e Ergodicidade.

Se o processo estocástico que gerou a série de observações é invariante no tempo, diz-se que o processo é estacionário. Embora, a teoria mostre conveniência prática do uso de séries estacionárias, no mundo real, nem todas as séries encontradas podem ser classificadas como estacionárias. Logo, a estacionariedade é uma condição bastante restritiva imposta à série

temporal. O conceito de processo não estacionário homogêneo surge quando se consegue um processo estacionário através de diferenças sucessivas do processo.

Quanto à ergodicidade, um processo estocástico é dito ergódico se apenas uma realização é suficiente para obter todas as suas estatísticas. Logo, todo processo ergódico é também estacionário, já que apenas uma realização de um processo não estacionário não poderá conter todas as informações necessárias para a especificação do processo.

De forma geral, o objetivo do estudo de uma série temporal consiste em: dada uma realidade (processo estocástico) retira-se uma amostra finita de observações equiespaçadas no tempo (série temporal) e através do estudo desta amostra (análise de séries temporais) identifica-se um modelo cujo objetivo é inferir sobre o comportamento da realidade (modelo estocástico).

2.1.3 - Domínio do Tempo x Domínio da Frequência

Uma série temporal pode ser analisada tanto no domínio do tempo quanto no domínio da frequência.

Na análise o **domínio do tempo**, é considerada a evolução temporal do processo, ou seja, mede-se a magnitude do evento que ocorre em determinado instante de tempo. Esta análise é baseada em um modelo paramétrico que utiliza as funções de autocovariância e autocorrelação.

A autocovariância é a covariância entre Z_t e seu valor Z_{t+k} separado por k intervalos de tempo, definida como

$$\gamma_k = \text{cov}[Z_t, Z_{t+k}] = E \{ [Z_t - \mu][Z_{t+k} - \mu] \}, \text{ onde } \mu \text{ é a média do processo}$$

A função de autocorrelação (FAC) é a autocovariância padronizada que serve para medir o comprimento e a memória de um processo, ou seja, a extensão para qual o valor tomado no

tempo t , depende daquele tomado no tempo $t-k$. A autocorrelação de defasagem k é definida como

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{Cov}[Z_t, Z_{t+k}]}{\sqrt{\text{Var}(Z_t)\text{Var}(Z_{t+k})}}$$

onde

$\text{Var}(Z_t) = \text{Var}(Z_{t+k}) = \gamma_0$ = variância do processo

$\rho_0 = 1$

$\rho_k = \rho_{-k}$ (a função de autocorrelação é simétrica em relação à origem $k=0$)

O gráfico dos coeficientes de autocorrelação ρ_k *versus* k é chamado de correlograma. A idéia de autocorrelação pode ser estendida se for medida a correlação entre duas observações seriais Z_t e Z_{t+k} e não se considerar os termos intermediários Z_{t+1} , Z_{t+2} , Z_{t+k-1} , tem o que se denomina função de autocorrelação parcial (FACP). A autocorrelação parcial (ϕ_{kk}) é representada por: $\text{Corr}(Z_t, Z_{t+k} \mid Z_{t+1}, \dots, Z_{t+k-1})$.

Dados estes conceitos, é importante que se defina um tipo de processo estocástico muito importante, o chamado processo estocástico ruído branco. Este caracteriza-se por ser um processo cujas variáveis aleatórias componentes são independentes e identicamente distribuídas (i.i.d), ou seja, não apresentam qualquer dependência serial. Em termos de autocorrelação tem-se que, para um ruído branco,

$$\rho_k = \begin{cases} 1; & k = 0 \\ 0; & k \neq 0 \end{cases}$$

Já a análise no **domínio da frequência**, é uma análise da frequência com que certos eventos ocorrem. Esta representação é útil quando os componentes harmônicos da série têm um significado físico ou os efeitos práticos do processo são analisados por suas componentes de frequência. A função associada é a densidade espectral.

A análise espectral visa estabelecer as propriedades de um processo estocástico em termos de frequência. Logo a pesquisa de periodicidades porventura existentes em uma série temporal pode ser realizada com auxílio da análise espectral.

Se $Z_t = \{(Z_t, t = 0, \pm 1, \pm 2, \dots)\}$ é um processo estacionário com função de autocovariância finita, define-se o espectro de Z_t como

$$f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \rho(k) e^{-i w k}, \quad -\pi \leq w \leq \pi$$

$$\rho(k) = \sum_{w=-\infty}^{\infty} e^{i w k} f(w), \quad k = 0, \pm 1, \pm 2, \dots$$

A função de densidade espectral normalizada é a transformada de Fourier da função de autocorrelação. A variância de Z_t é uma medida da “potência total do processo” e a função de densidade espectral representa a contribuição para a “potência total” de todos os componentes de frequência presentes no processo. Na prática, tem que se estimar o espectro. O estimador mais utilizado é o periodograma com janelas espectrais, para que tenha boas propriedades estatísticas.

2.2 - Modelos de Previsão

A previsão de uma série temporal é simplesmente o estabelecimento dos valores futuros da série. Uma previsão é uma estimativa quantitativa (ou conjunto de estimativas) acerca de eventos futuros, baseados na informação atual e passada.

Nesta dissertação, o objetivo principal é realizar a previsão da demanda mensal de gasolina automotiva no Brasil. Como dito anteriormente, são poucos os trabalhos no País referentes a este tema. Foram encontrados trabalhos relativos à previsão de demanda anual na década de oitenta⁵ e apenas um trabalho na década de noventa⁶. Ainda na década de 80, encontramos um trabalho de análise de elasticidade preço-renda da demanda de gasolina automotiva que trabalha com dados mensais⁷.

A metodologia usada nesta dissertação consiste em, a partir de uma análise exploratória dos dados, construir um modelo usando uma estratégia *bottom-up*, ou seja, parte-se de um modelo simples, refinando-o até encontrar um modelo apropriado. Partiu-se então de um modelo autoprojetoivo indo até uma formulação de Redes Neurais passando por um modelo de regressão dinâmica.

A noção lógica que norteia a metodologia baseia-se no próprio comportamento da série, ou seja, a partir da constatação de um comportamento irregular da demanda de gasolina automotiva, procura-se verificar se, apesar deste comportamento, esta série pode-se auto-explicar, ou seja, se a adoção de um modelo autoprojetoivo já não seria o suficiente para realizar as previsões.

Num segundo momento, procura-se, através de um modelo que usa outras variáveis que não somente a própria série em estudo, aumentar o poder de explicação acerca do comportamento e conseguinte projeções da demanda de gasolina automotiva.

Por fim, adota-se um modelo de Redes Neurais que, a princípio, consegue tratar as não-linearidades intrínsecas ao fenômeno “demanda”, o que é altamente desejável, ainda mais do ponto de vista de eficiência nas projeções.

⁵ Suyama (1981), Barros e Ferreira (1982), Buonfiglio (1988).

⁶ Correia (1999).

⁷ Souza *et al.*, (1984).

Obviamente objetiva-se sempre encontrar o melhor modelo, ou seja, aquele que possibilitará reduzir o erro de previsão, retratando o mais próximo possível a realidade.

É evidente que, tanto para métodos univariados quanto para métodos causais, existem inúmeros modelos que poderiam ser elaborados. No primeiro caso tem-se basicamente os modelos de médias móveis e de amortecimento exponencial e os modelos ARIMA. Já no segundo, encontram-se os modelos de regressão tradicionais, os modelos de regressão dinâmica e até mesmo os modelos de Filtro de Kalman (Hamilton, 1994), entre outros. O mesmo acontece com diferentes tipos de algoritmos de Redes Neurais que poderiam ser utilizados.

Frente a estes fatos, procurar-se-á nos próximos capítulos, abordar os métodos (Box & Jenkins, Regressão Dinâmica e a Rede Neural) de forma que se possa fundamentar um pouco mais o uso destes.

3 - MODELOS BOX & JENKINS

Como dito anteriormente, a tentativa de modelar a série de demanda de gasolina automotiva no Brasil usando um método univariado, consiste em verificar se esta série, por si só, traz informação suficiente que permita construir um modelo usando apenas a série em questão, ou seja, não fazendo uso de variáveis exógenas⁸. Entre os modelos univariados utilizados⁹, os modelos de Box & Jenkins apresentaram melhor desempenho (menores erros de previsão).

Atenta-se que, com as vantagens computacionais existentes hoje, existem vários métodos de estimação automática de modelos ARIMA. Neste trabalho, estima-se manualmente os modelos, através dos procedimentos indicados por Box & Jenkins a fim de se chegar a um modelo ARIMA satisfatório.

Portanto, a seguir, faz-se uma abordagem geral e sucinta da metodologia proposta por estes autores, procurando-se embasar o modelo encontrado, segundo este método, para a série de demanda de gasolina automotiva.

3.1 - Fundamento teórico

Segundo Souza & Camargo (1996), o fundamento teórico do modelo de Box & Jenkins baseia-se na Teoria Geral de Sistemas Lineares que diz que a passagem de um ruído branco por um filtro linear de memória infinita gera um processo estacionário de segunda ordem¹⁰.

Isto é, pela Teoria Geral de Sistemas:

⁸ Em alguns casos, o modelo univariado pode apresentar desempenho melhor do que um modelo que faz uso de outras variáveis causais.

⁹ Foi tentado também um modelo de amortecimento exponencial (Montgomery *et al*, 1990).

¹⁰ Em termos gerais, processo estacionário de 2º ordem é um processo estocástico com média e variância constantes.

Definindo um operador de atraso “B” (*backward shift operator*)

$$B^k Z_t = Z_{t-k}$$

Tem-se que:

$$w_t = a_t - \Psi_1 a_t B - \Psi_2 a_t B^2 - \dots$$

$$w_t = (1 - \Psi_1 B - \Psi_2 B^2 - \dots) a_t = \Psi(B) a_t$$

$$w_t = \Psi(B) a_t \Leftrightarrow a_t = \Psi(B)^{-1} w_t$$

Sendo que:

$$\Psi(B)^{-1} = \pi(B)$$

$$\text{onde } \pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$$

Como $\Psi(B)$ [ou $\pi(B)$] possuem infinitos parâmetros, o que causa um problema, Box & Jenkins mencionam que, sob certas restrições, pode-se afirmar que todo polinômio infinito pode ser expresso pelo quociente de dois polinômios finitos.

Então:

$$\Psi(B) = \theta(B) / \phi(B), \text{ onde:}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \Rightarrow \text{Polinômio MA (q)}$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \Rightarrow \text{Polinômio AR (p)}$$

Desta forma, surgem os denominados Modelos ARMA (p,q) que assumem a seguinte forma:

$$\phi(B) w_t = \theta(B) a_t$$

Entretanto, para se modelar séries não estacionárias na média, produz-se a estacionariedade através da diferenciação da série original, ou seja, produz-se uma série não estacionária homogênea¹¹. Desta forma, se Z_t é uma série não estacionária, procurar-se-á transformá-la na série X_t não estacionária homogênea. Então:

$$X_t = Z_t - Z_{t-1} = Z_t - BZ_t = (1 - B) Z_t = \nabla Z_t$$

Onde: $\nabla = (1 - B) \Rightarrow$ operador de diferença

Isto significa que, aplica-se tantas diferenças quantas forem necessárias para produzir estacionariedade (na média) da série resultante. Genericamente, uma série w_t estacionária é obtida pela aplicação de “d” ($d = 0, 1, 2, \dots$)¹² diferenças na série original: $w_t = \nabla^d Z_t$. Surgem então os chamados Modelos ARIMA (p,d,q) que assumem a seguinte forma:

$$\phi(B)\nabla^d Z_t = \theta(B)a_t$$

A filosofia da modelagem Box & Jenkins fundamenta-se em duas idéias básicas: o princípio da parcimônia (escolher um modelo com o menor número de parâmetros possíveis para uma representação matemática adequada) e a construção de modelos através de um ciclo iterativo (estratégia de seleção de modelos até a obtenção de um modelo satisfatório). Desta forma, atenta-se que a modelagem através da metodologia de Box & Jenkins, abrange várias etapas de análise indo desde a identificação da estrutura do modelo, passando pela estimação paramétrica e por fim, fazendo vários testes de validação do modelo.

¹¹ A não estacionariedade homogênea exclui os processos de comportamento explosivos e/ou altamente não-lineares (Souza & Camargo, 1996).

¹² Na prática, normalmente o grau máximo de diferenciação é 2.

3.2 - A Metodologia Box & Jenkins

O primeiro passo fundamental da metodologia consiste em identificar a ordem de homogeneidade “d”, ou seja, caso seja necessário, trata-se de identificar o número de vezes que a série original deve ser diferenciada para se tornar uma série estacionária. Este procedimento pode ser feito através da observação do próprio gráfico da série ou da função de autocorrelação (FAC)¹³.

O passo seguinte na metodologia é a identificação do modelo, ou seja, da sua ordem (identificação de p e q). Para isso são utilizados os conceitos de função de autocorrelação (FAC) e autocorrelação parcial (FACP), ou seja, utiliza-se os correlogramas¹⁴ definidos no Capítulo 2. De modo geral, para a identificação da ordem, observa-se os comportamentos da FAC e da FACP. No quadro abaixo é feito um resumo das características destas funções para os modelos AR(p), MA(q) e ARMA (p,q).

¹³ A Função de Autocorrelação para uma série não estacionária apresenta um lento decréscimo.

¹⁴ Para mais detalhes ver Souza & Camargo (1996).

Quadro 01: Resumo das características teóricas da FAC e da FACP dos modelos AR(p), MA(q) e ARMA (p,q)

Modelo	Função de Autocorrelação (ρ_k)	Função de Autocorrelação Parcial (ϕ_{kk})
AR(p)	Infinita (Exponencial e/ou senóides amortecidas)	Finita (Corte após o <i>lag</i> “p”)
MA (q)	Finita (Corte após o <i>lag</i> “q”)	Infinita (Exponencial e/ou senóides amortecidas)
ARMA (p,q)	Infinita (Exponencial e/ou senóides amortecidas após o <i>lag</i> “q-p”)	Infinita (Exponencial e/ou senóides amortecidas após o <i>lag</i> “p-q”)

Fonte: Souza e Camargo (1996)

De uma maneira geral, para se identificar a ordem p , de um modelo AR(p), por exemplo, observa-se se a FAC decresce e se a FACP apresenta um corte¹⁵. Se isto acontecer, o *lag* onde este corte ocorre nos fornece a ordem p ($p = \text{lag do corte}$).

Por outro lado, para um modelo MA (q) a FAC e a FACP apresentam comportamento inverso ao de um modelo puramente autoregressivo. Isto quer dizer que, para um modelo MA, a FACP decresce, e a FAC é que apresenta um corte. Da mesma forma, o *lag* onde este corte ocorre fornece a ordem q do modelo MA.

Após a identificação da ordem do modelo, é necessário obter as estimativas dos parâmetros desse modelo. A técnica utilizada para as estimativas é a da máxima verossimilhança (Dudewicz & Mishra, 1988).

¹⁵ Geralmente nos *softwares* específicos para previsão, tanto o gráfico da FAC quando da FACP apresentam intervalos de significância dos *lags*. Os *lags* cujos valores da autocorrelação ultrapassam estes intervalos são ditos significantes (Goodrich & Stellwagem, 1999). Se tenho, por exemplo, a autocorrelação de *lag* 1 significante e, a partir do *lag* 2 (inclusive), as autocorrelações estão todas dentro do intervalo, isto indica um “corte” no *lag* 1.

Por fim, identificado o modelo e estimados os parâmetros, faz-se os denominados testes de aderência para verificar a adequabilidade final do modelo. Nestes testes, encontram-se entre outros, testes para os resíduos e os testes de sobrefixação.

Nos testes para os resíduos, procura-se constatar se, após elaborado o modelo, o resíduo gerado por este modelo é um ruído branco, ou seja, se o modelo foi capaz de explicar satisfatoriamente o comportamento da série de forma que o erro não apresente nenhuma estrutura de correlação. Este fato consiste num dos indicadores de eficiência explicativa do modelo.

O teste de sobrefixação, por sua vez, consiste simplesmente em se gerar modelos de ordem superior ao identificado, de forma que se possa reforçar a pertinência deste.

3.3 - Modelos SARIMA

Os processos encontrados na prática, além de raramente serem estacionários, apresentam muitas vezes componentes sazonais. Assim Box & Jenkins formularam seus modelos para séries temporais com componentes sazonais dando origem aos modelos SARIMA.

Nesse caso a modelagem segue a equação:

$$\phi(B)\Phi(B^S)\nabla_S^D\nabla^d Z_t = \theta(B)\Theta(B^S)a_t \quad (3.1)$$

onde:

- $\phi(B)$: operador não sazonal auto-regressivo
- ϕ_i : parâmetros auto-regressivo não-sazonais
- $\nabla^d = (1-B)^d$: operador diferença não sazonal de ordem d
- $\Phi(B^S)$: operador sazonal auto-regressivo
- Φ_i : parâmetros auto-regressivo sazonais

$\nabla_s^D = (1-B^s)^D$: operador diferença sazonal de ordem D

$\theta(B)$: operador não sazonal de médias móveis

θ_i : parâmetros de médias móveis não sazonais

$\Theta(B^s)$: operador sazonal de médias móveis

Θ_i : parâmetros de médias móveis sazonais

Um modelo com esta estrutura é denominado SARIMA(p,d,q)x(P,D,Q). Ressalta-se que o procedimento de obtenção deste modelo segue os mesmos passos empregados para achar o modelo ARIMA não sazonal. Isto quer dizer que, no SARIMA, faz-se também a observância do comportamento da FAC e da FACP (inclusive observância dos *lags* de “cortes”) entretanto, olha-se para os *lags* sazonais (para uma série mensal, por exemplo, são os *lags* 12, 24, 36, etc).

4 - REGRESSÃO DINÂMICA

4.1 - Introdução

O objetivo principal deste trabalho consiste em elaborar um modelo de previsão da demanda de gasolina automotiva no Brasil. A teoria econômica nos oferece uma gama de análises e interpretações das curvas de demanda¹⁶ tanto a nível micro quanto macroeconômico. Um dos pontos fundamentais da análise da demanda refere-se ao conceito de elasticidade.

O termo “elasticidade” é amplamente empregado por economistas para descrever as características de uma relação entre duas variáveis. A elasticidade, por não depender das unidades de mensuração das variáveis estudadas, permite a comparação das relações entre variáveis medidas em qualquer unidade.

A teoria econômica estuda amplamente a relação da curva de demanda com variáveis como preço e renda que, por sua vez, dão origem às análises de elasticidade-preço e elasticidade-renda da demanda. Normalmente, a estimação destas elasticidades se dá através da estimação de modelos de regressão linear¹⁷. Neste trabalho, entretanto, a tentativa de construir um modelo que gere melhores resultados que o modelo univariado discutido no capítulo anterior, fundamentar-se-á no que se denomina na literatura de regressão dinâmica a qual será abordada a seguir.

¹⁶ Varian (1997).

¹⁷ Gujarati (1995).

4.2 - Conceituação

Nos modelos de regressão linear estudados comumente na literatura, supõe-se que os erros “gerados” pelo modelo possuem algumas características como: média zero, variância constante, distribuição Normal e independência (o que implica na inexistência de correlação serial).

Entretanto, na prática, ao modelar séries econômicas, os resíduos tendem a apresentar correlações positivas, e erros positivos tendem a ser seguidos por outros também positivos (o mesmo comportamento é observado para resíduos negativos). O gráfico dos resíduos *versus* o índice dos tempos revela, nesta situação, que os resíduos tendem a se agrupar em “blocos” de resíduos com o mesmo sinal.

Segundo Barros e Souza (1995), ao tentar 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, segundo os autores, algumas das consequências da autocorrelação dos resíduos são:

- 1) Os estimadores usuais por mínimos quadrados são ainda não tendenciosos, mas não têm variância mínima.
- 2) Os estimadores da variância e 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.
- 3) Os intervalos de confiança para os parâmetros da regressão e os testes de hipóteses relacionados a estes intervalos perdem a validade, como uma consequência direta de 2).

Desta forma, estes três motivos implicam na necessidade de procurar procedimentos para tratar o problema de autocorrelação dos erros, pois ignorá-los leva, em geral, a inúmeras conclusões errôneas.

Em particular, dado que a hipótese de independência dos erros não é realista no contexto de séries temporais, os modelos de regressão dinâmica estendem os modelos usuais de regressão ao levantarem esta restrição.

Os modelos de regressão dinâmica combinam a dinâmica de séries temporais e o efeito de variáveis explicativas. Atenta-se que o termo “regressão dinâmica” não indica que os parâmetros do modelo evoluem no tempo¹⁸. Ao contrário, a palavra “dinâmica” significa aqui um modelo de regressão no qual incluímos a estrutura de dependência de uma série temporal.

Modelos de regressão dinâmica devem ser usados quando existe uma estrutura de dependência entre a variável de interesse e variáveis causais e, ao mesmo tempo, quando a estrutura de correlação da série dependente (série a ser explicada) indicar que não podemos supor a independência dos erros.

Salienta-se que a estimação de parâmetros num modelo de regressão dinâmica é feita através de mínimos quadrados ordinários¹⁹, a exemplo dos modelos de regressão usuais. Entretanto, a estimação em modelos de regressão dinâmica é mais complicada, e envolve um procedimento iterativo com vários estágios.

¹⁸ Caso dos modelos de espaço de estados que usam o Filtro de Kalman (Hamilton, 1994).

¹⁹ Dudewicz & Mishra (1998).

Nos modelos de regressão dinâmica, a variável dependente é explicada por seus valores defasados e pelos valores atuais e passados de variáveis causais ou exógenas. Atenta-se, neste momento, para uma outra distinção entre os modelos de regressão dinâmica e modelos de espaço de estados. Nos modelos de regressão dinâmica, as variáveis exógenas são tratadas como “números fixos” e não como variáveis aleatórias. Já nos modelos de espaço de estados, as variáveis exógenas são tratadas como séries temporais, ou seja, realizações de processos estocásticos. Logo, no contexto da modelagem em espaço de estados, a estrutura de autocovariâncias e autocorrelações das séries de variáveis exógenas é uma informação de interesse, enquanto este aspecto é ignorado nos modelos de regressão dinâmica.

4.3 - Estrutura dos Modelos de Regressão Dinâmica

Os modelos de regressão dinâmica podem ser descritos pela equação:

$$\varphi(B)Y_t = \beta x_t + \varepsilon_t \quad (4.1)$$

onde:

Y_t = variável dependente (endógena) no instante t

β = vetor de coeficientes das variáveis causais, que será estimado por mínimos quadrados

x_t = vetor de variáveis causais (exógenas)²⁰ no instante t

ε_t = ruído aleatório associado ao modelo, onde supomos que os ε_t são independentes e identicamente distribuídos com densidade $N(0, \sigma^2)$

$\varphi(B)$ = polinômio autoregressivo de ordem p , isto é:

$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$ sendo B o operador de atraso

A estrutura do modelo de regressão dinâmica permite considerar como elementos x_t variáveis causais e também suas defasagens.

A presença do polinômio $\varphi(B)$ no modelo traz uma grande flexibilidade desta classe de modelos mas, ao mesmo tempo, dificulta a procura por um modelo adequado. Observa-se que, se $\varphi(B) = 1$, não existem defasagens da variável dependente, e a interpretação do modelo é muito simples, pois as variáveis causais influenciam diretamente a variável endógena. Ao contrário, quando $\varphi(B) \neq 1$, o modelo pode ser usado para representar relações bastante complicadas.

Por fim, salienta-se que uma grande diferença entre os modelos de Regressão Dinâmica e os modelos ARIMA consiste no fato dos modelos de regressão dinâmica incluírem efeitos de variáveis causais através do termo βx_t . Os modelos ARIMA univariados de Box e Jenkins, por sua vez, não incluem tais efeitos, e apenas o passado da série Y_t e os valores defasados da série de erros são usados na modelagem e previsão da série Y_t .²¹

Procura-se, portanto, dentro de metodologia definida, evoluir na modelagem, buscando em outras causalidades, que não somente a própria série de demanda, um melhor modelo para prever esta demanda. Este modelo, como já mencionado, foi o modelo de regressão dinâmica que pode ser considerado como um caso particular do que é conhecido na literatura como modelos de Cochrane e Orcutt generalizados.

²⁰ As variáveis exógenas representadas por x_t têm seus valores determinados “fora” do modelo de regressão, enquanto a variável endógena Y_t é determinada a partir das variáveis exógenas e do ruído ε_t .

²¹ A metodologia Box & Jenkins também possui um modelo causal, o chamado modelo de Função de Transferência. Neste trabalho, tentou-se, inicialmente, também a elaboração de um modelo deste tipo, entretanto houve um impedimento pois o *software* Autobox não conseguiu trabalhar com as variáveis na escala logarítmica e, portanto, algumas relações sem sentido foram obtidas. Além do mais, dentro da metodologia deste trabalho, o objetivo não era usar um infinito número de técnicas mas sim seguir uma lógica de trabalho que procura, entre outros objetivos, mostrar que, para determinadas séries, às vezes um tratamento via modelos que abordem as não linearidades existentes, possibilita modelar mais satisfatoriamente uma série como a demanda de gasolina automotiva.

4.3.1 - Modelos de Regressão Cochrane-Orcutt Generalizados²²

O modelo de regressão generalizado de Cochrane e Orcutt (1949) é dado por:

$$\varphi(B)Y_t = \beta x_t + w_t \quad (4.2)$$

$$R(B)w_t = \varepsilon_t \quad (4.3)$$

Onde $R(B)$ = polinômio autoregressivo

Pode-se notar que a equação (4.2) tem a mesma forma da equação (4.1) da regressão dinâmica, entretanto os erros w_t apresentam uma estrutura AR dada pela equação 4.(3)²³.

Este modelo dado pelas equações (4.2) e (4.3) ainda pode ser descrito em termos de uma única equação como a seguir. Note que da equação (4.2):

$$w_t = \varphi(B)Y_t - \beta x_t$$

Substituindo esta última expressão na equação (4.3) tem-se que:

$$R(B) \cdot [\varphi(B)Y_t - \beta x_t] = \varepsilon_t \quad (4.4)$$

Desta última expressão nota-se que o modelo de regressão generalizado de Cochrane e Orcutt introduz defasagens tanto na variável dependente (Y_t) quanto nas causais. A expressão (4.4) indica também que a relação de causalidade entre Y_t e x_t não é afetada pela introdução do polinômio autoregressivo $R(B)$.

²² Barros e Souza (1995).

²³ A formulação original de Cochrane e Orcutt supõe que os ruídos w_t apresentam apenas uma estrutura $AR(1)$, ou seja, a equação (4.3) é dada por: $w_t = \alpha w_{t-1} + \varepsilon_t$ onde os ε_t são i.i.d e apresentam distribuição $N(0, \sigma^2)$.

Esta última equação pode ainda ser escrita em termos de novas variáveis Y^*_t e x^*_t dadas por:

$$Y^*_t = R(B)Y_t \text{ e}$$

$$x^*_t = R(B)x_t$$

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

A equação obtida então usando-se estas novas variáveis é:

$$\varphi(B)Y^*_t = \beta x^*_t + \varepsilon_t \quad (4.5)$$

Logo, o modelo de regressão generalizado de Cochrane e Orcutt reduz-se ao modelo de regressão dinâmica usual ao se considerar as novas variáveis Y^*_t e x^*_t . O modelo original de Cochrane e Orcutt tem como fator comum $R(B) = 1 - \alpha B$ e então a equação (4.5) reduz-se a:

$$\varphi(B) \cdot [Y_t - \alpha Y_{t-1}] = \beta [x_t - \alpha x_{t-1}] + \varepsilon_t \quad (4.6)$$

Salienta-se que, neste modelo, o procedimento de estimação é seqüencial. A estimativa inicial de α é 0, e a partir dela estimamos β e $\varphi(B)$ por mínimos quadrados ordinários. A partir destas estimativas encontra-se um estimador de $R(B)$ através da equação $R(B)w_t = \varepsilon_t$. O polinômio estimado $R(B)$ é então usado para transformar Y_t e x_t e reestimar β e $\varphi(B)$. O processo é repetido até que se alcance a convergência dos parâmetros.

4.4 - Construção de modelos de regressão dinâmica

Geralmente os modelos econométricos têm uma estrutura conhecida, baseada em considerações teóricas e o problema reduz-se ao problema de estimação dos parâmetros do

modelo já conhecido. Entretanto, este é raramente o caso no contexto de séries temporais, onde a estratégia é construir modelos a partir dos dados.

A estratégia usualmente empregada para construir um modelo de regressão dinâmica é uma estratégia *bottom-up*, isto é, partimos de um modelo simples e o refinamos e incluímos novas variáveis até encontrar um modelo apropriado. A elaboração de um modelo de regressão dinâmica é muitas vezes um procedimento difícil, pois é preciso 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. Por exemplo, vendas (ou demanda) de um produto são geralmente afetadas por seu preço. O aumento do preço, por sua vez, tende a diminuir as vendas (ou demanda), e vice-versa. Logo, se o modelo de regressão encontrado para explicar venda pelo preço apresenta um coeficiente positivo para a variável preço, é bom desconfiar por mais bem ajustado que esteja o modelo, pois a relação apontada pelo modelo não é, em geral, verdadeira²⁴. Em síntese, na escolha de um modelo de regressão, não é necessário apenas encontrar um ajuste de parâmetros adequado, mas fundamentalmente faz-se mister verificar se os coeficientes estimados são coerentes.

As previsões geradas por um modelo de regressão dinâmica dependem não só de valores passados da série mas também dos valores previstos para as variáveis causais. Logo, para obtermos as previsões da série Y_t para $T+1$, $T+2$, $T+3$, etc, é necessário fornecer ao modelo os valores futuros do vetor de variáveis causais x_t . Se as previsões destas variáveis exógenas não forem apropriadas, o modelo de regressão dinâmica irá também gerar previsões inadequadas.

²⁴ Na verdade, “casos estranhos” como este ocorrem na prática com certa frequência e uma possível saída é olhar para o coeficiente da variável preço com alguma defasagem, e verificar se é possível encontrar algum resultado coerente.

Isto caracteriza um aspecto importante dos modelos de regressão dinâmica que consiste na possibilidade de consecução de cenários ao se chegar a um modelo relacional de variáveis dependentes em relação a variáveis explicativas. Isto é, surge a possibilidade de montagem de vários cenários para as variáveis causais o que enriquece, por exemplo, qualquer trabalho de planejamento de vendas de um determinado produto.

Feitas estas considerações, o fluxograma a seguir indica, de maneira genérica, os passos usados na construção de um modelo de regressão dinâmica.

Figura 4.1: Construção de um modelo de regressão dinâmica

Como dito anteriormente, a especificação correta de um modelo de regressão dinâmica envolve a precisa especificação da relação causal entre as variáveis e da estrutura dinâmica do modelo.

Nos modelos de regressão dinâmica podem ser usadas também variáveis de intervenção (ou variáveis *dummy*). O objetivo deste tipo de procedimento é considerar situações atípicas como, por exemplo, aumento das vendas de brinquedo no Natal e no Dia da Criança (neste caso é razoável incluir *dummies* para os meses de dezembro e outubro no modelo). O mesmo procedimento pode ser usado para levar em conta os efeitos de situações incomuns como greves e planos econômicos²⁵.

Os modelos de regressão dinâmica incorporam ainda diretamente a sazonalidade da série ao modelo, ao invés de supor que a série será previamente dessazonalizada. Atenta-se que existem duas maneiras de tratar a sazonalidade: via *dummies* sazonais ou diretamente, através de defasagens na variável dependente ou nos erros estruturados.

4.5 - Testes usados nos modelos de regressão dinâmica

Como dito anteriormente, em regressão dinâmica a construção do modelo envolve vários passos até se chegar a um modelo “final”. Diversos testes da adequação de um modelo de regressão podem ser mencionados. Estes testes²⁶ são aplicados em diversos estágios da modelagem da série. Temos, por exemplo:

i) testes como o objetivo de definir a especificação do modelo explicativo;

²⁵ Variáveis *dummies* são geralmente definidas como 1 (no período de ocorrência do fato relevante) e 0 (fora deste período).

²⁶ A maioria dos testes empregados em regressão dinâmica é uma variante dos testes de Multiplicadores de Lagrange (testes LM) e são baseados na distribuição Qui-Quadrado. (Barros e Souza, 1995).

ii) testes visando encontrar a dinâmica do modelo, isto é, a inclusão ou não de variáveis defasadas,

iii) testes para verificar o ajuste do modelo.

Para que se entenda um pouco melhor o modelo de regressão dinâmica que será apresentado para a demanda da gasolina automotiva no Brasil, será abordado de uma forma geral, como se processa o “algoritmo” dos testes na prática:

4.5.1 - Testes de verificação da “dinâmica” do modelo

Como dito anteriormente, a dinâmica de um modelo acontece através dos *lags* da variável dependente e/ou através da presença de erros estruturados num modelo de Cochrane-Orcutt. A cada momento da elaboração do modelo, são realizados testes de hipóteses²⁷ sobre a “dinâmica” do modelo.

Em todos os casos a seguir, a hipótese nula afirma que a dinâmica do modelo está corretamente especificada, ou seja, a inclusão de outros *lags* da variável dependente ou outros erros estruturados não é necessária. A hipótese alternativa, em cada caso, representa a necessidade de inclusão de novos termos. Sendo assim tem-se :

1) Teste de defasagem da variável endógena

Suponha que a variável dependente Y_t e seus *lags* $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p-1}$ estão presentes no modelo atual. A hipótese alternativa consiste em adicionar a variável defasada Y_{t-p} ao modelo, isto é, adiciona-se o primeiro lag ainda não presente no modelo atual. Se esta variável for considerada significativa, a hipótese nula é rejeitada e deve-se adicionar a variável Y_{t-p} ao modelo.

²⁷ Dudewicz e Mishra (1988).

2) Teste da defasagem sazonal da variável endógena

Este teste é semelhante ao anterior. A hipótese alternativa consiste em adicionar ao modelo atual a variável defasada até o primeiro *lag* sazonal Y_{t-pS} ainda não presente ao modelo. Se o coeficiente de Y_{t-pS} for significativo, esta variável deve ser incluída no modelo, e a hipótese nula deve ser rejeitada.

3) Teste da sequência de defasagens da variável endógena

A hipótese alternativa consiste em adicionar todos os *lags* da variável dependente que ainda não estão presentes no modelo.

4) Teste da defasagem dos resíduos

A hipótese alternativa consiste em adicionar ao modelo o primeiro termo defasado ε_{t-p} ainda não incluído no modelo atual.

5) Teste da defasagem sazonal dos resíduos

Na hipótese alternativa adiciona-se ao modelo atual o primeiro *lag* sazonal ε_{t-pS} ainda não presente no modelo.

6) Teste da sequência de defasagens dos resíduos

Na hipótese alternativa adiciona-se às variáveis do modelo atual uma sequência de resíduos defasados ε_{t-1} , ε_{t-2} , ε_{t-S} onde S é o período sazonal. É importante ressaltar que, na hipótese alternativa, inclui-se apenas os resíduos ainda ausentes no modelo atual.

4.5.2 - Testes para a especificação das variáveis causais

O objetivo de todos estes testes é verificar se a inclusão de uma ou mais variáveis ainda não contempladas no modelo resulta numa melhora do ajuste. Atenta-se para o fato de que os testes para a especificação de variáveis causais não se referem à parte dinâmica do modelo, e não tratam da inclusão de *lags* da variável dependente e de erros estruturados²⁸.

1) Teste das variáveis causais excluídas

Neste teste verifica-se a necessidade de inclusão de cada uma das variáveis (escolhidas previamente para análise) mas que ainda não estão presentes no modelo. Se quaisquer destas variáveis são consideradas significantes, deve-se incluí-las no modelo (talvez sequencialmente) e “rodar” a mesma bateria de testes para verificar se a inclusão foi vantajosa.

2) Teste de tendência temporal

Este teste corresponde à inclusão de uma variável do tipo $X_t=t$ no modelo. Esta variável é útil em casos onde a série dependente não é estacionária.

3) Teste da defasagem das variáveis exógenas (causais)

Na hipótese alternativa inclui-se um *lag* adicional das variáveis causais já presentes no modelo atual.

4) Teste para a presença de funções não lineares das variáveis exógenas

Neste teste inclui-se o quadrado de cada variável exógena já presente no modelo. Todos os quadrados das variáveis exógenas são incluídos de uma só vez e, portanto, é necessário buscar quais (ou qual) quadrados são realmente significantes.

²⁸ Estes são realizados nos testes para a “dinâmica”.

5) Teste do fator comum

Este teste é realizado só quando o modelo inclui erros estruturados. Sob a hipótese alternativa, a autorregressão dos erros é eliminada, e todos os *lags* da variável dependente e das causais são adicionados ao modelo. Se a hipótese nula é rejeitada, existe evidência de que um modelo mais geral deveria ser considerado, ao invés do modelo de Cochrane-Orcutt. O grande problema é descobrir em que direção deve-se generalizar o modelo corrente, e não existe uma resposta única para esta questão.

4.5.3 - Testes baseados na autocorrelação dos resíduos²⁹

Como pode ser visto, o processo de construção de um modelo de regressão dinâmica deve levar em conta diversos diagnósticos com o objetivo de verificar se o modelo atual é apropriado. Em particular, deve-se sempre examinar o gráfico das autocorrelações dos resíduos. Se estas são significantes para alguns *lags*, alguma característica da variável dependente não foi capturada pelo modelo atual. Por exemplo, no caso de dados mensais, se a autocorrelação dos resíduos é significativa no *lag* 12, a observação situada num período genérico $t-12$ meses é relevante para explicar a observação no período t , e sua inclusão no modelo possivelmente resultará num decréscimo dos erros de previsão do modelo.

²⁹ Podem ser feitos ainda testes para verificar a existência de variações na variância dos resíduos, ou seja, procura-se detectar a heterocedasticidade da série de resíduos.

Segundo Barros e Souza (1995), a existência de autocorrelações significantes nos resíduos pode então indicar uma das seguintes situações: 1) deve-se incluir mais *lags* da variável dependente ou 2) deve-se incluir *lags* adicionais das variáveis exógenas já presentes no modelo ou incluir novas variáveis causais.

Enfim, é importante saber que, em qualquer das situações mencionadas, o fato dos resíduos apresentarem autorrelações significantes indica que algum tipo de estrutura presente na série Y_t não foi captada pelo modelo em consideração.

5 - REDES NEURAIS

O uso das redes neurais nas mais diversas áreas do conhecimento tem se tornado cada vez mais freqüente. Ganhos significativos têm sido verificados com o uso desta metodologia nos mais diversos campos e um, de particular interesse, é o campo das previsões de séries temporais³⁰. Por ser tratar de uma metodologia de certa forma recente, acredita-se ser relevante fazer uma abordagem conceitual desta.

5.1 - Conceituação

A motivação original desta metodologia foi a tentativa de modelar a rede de neurônios humanos visando compreender o funcionamento do cérebro. Portanto, como o próprio nome da metodologia revela, sua motivação inicial foi a de realizar tarefas complexas que o cérebro executa com elevada efetividade (por exemplo: reconhecimento de padrões, percepção e controle motor) através da simulação de seu funcionamento.

Muito pouco se conhece sobre o funcionamento dos neurônios e suas conexões o que implica na impossibilidade de se obter um modelo da rede de neurônios humanos que reproduza com fidelidade suficiente o funcionamento do cérebro. Desta forma, segundo Pedreira (2000), apenas no início do desenvolvimento das redes neurais a motivação fisiológica desempenhou papel relevante. Atualmente, a pesquisa neste campo do conhecimento não está mais atrelada às peculiaridades do funcionamento do cérebro como um todo ou do neurônio mais especificamente.

³⁰ O uso de redes neurais para previsão de carga elétrica, por exemplo, tem ganhado bastante relevância. Ver, entre outros, Sobral (1999), Lourenço (1998), Pelikan *et al* (1996), Khotanzad *et al* (1995), Lee *et al* (1992), Park *et al* (1991), Hsu (1991).

Hoje, portanto, as redes neurais artificiais são uma metodologia estatística eficiente e capaz de resolver uma gama de problemas importantes. Conceitualmente, uma rede neural artificial é um dispositivo tanto capaz de processar informação de forma distribuída quanto de incorporar conhecimento através de exemplos. Trata-se, portanto, de um processador capaz de extrair conhecimento experimental disponibilizando-o para uso prático (tomada de decisões por exemplo).

Uma grande vantagem de usar uma rede neural é a capacidade de resolver problemas sem a necessidade de definição de listas de regras³¹ ou de modelos explícitos. Isto possibilita tratar de situações onde é difícil criar modelos adequados da realidade ou situações com freqüentes mudanças no ambiente. Atenta-se que, grande parte desta sua adequabilidade funcional deve-se à sua capacidade em inferir relações não lineares complexas. Frente a estas suas propriedades, hoje, podemos observar sua aplicabilidade principalmente nas áreas de classificação de padrões e de previsão³².

5.2 - Arquitetura Neural

O termo “arquitetura” em redes neurais é a forma com que as unidades (ou neurônios) e suas conexões são arranjadas. A arquitetura mais comum é a chamada *feedforward*, onde o sinal é propagado para frente, ou seja, da entrada para a saída.

Pode-se, então, interpretar as redes neurais como sistemas constituídos por unidades de processamento denominados neurônios.

³¹ Como é necessário nos sistemas especialistas e em métodos estatísticos clássicos.

³² Tem-se observado que a aplicação em previsão tem apresentado mais efeito quando conjugada com outras técnicas, por exemplo, métodos estatísticos.

A figura 5.1, por exemplo, mostra um exemplo de uma rede *Feedforward* (propagação direta) com três entradas, uma camada escondida e uma camada de saída (são representados 4 neurônios). Ela está incorporando informações através do vetor $X_k = [X_{1k} \ X_{2k} \ X_{3k}]^T$ para produzir como resultado Y_i . O peso $w_{c,d}$ conecta o neurônio c ao neurônio d da camada imediatamente posterior. Ressalta-se que numa rede *Feedforward* os neurônios não têm realimentação.

Figura 5.1: Rede *Feedforward*

Atenta-se ainda que, dentre os tipos de redes neurais, temos redes em camada totalmente conectadas, redes em camada parcialmente conectadas (normalmente não é usada na prática), redes em malhas (as entradas são ligadas nas malhas, por exemplo, Rede de Kohonen) e redes de Hopfield, entre outras³³.

Segundo Pedreira (2000), ressalta-se que, as redes neurais com neurônios organizados em camadas estão entre as mais bem sucedidas do ponto de vista de aplicações. Em geral, tem-se um nível de entrada, um nível de saída e camadas de unidades ocultas. As chamadas camadas ocultas são aquelas que não estão diretamente ligadas nem ao nível de saída nem ao nível de entrada. A ligação entre as unidades do nível de entrada e da camada oculta é feita através dos pesos. O mesmo se dá entre a camada oculta e o nível de saída. Salienta-se que, em muitas aplicações, basta a utilização de uma camada oculta. São poucas as situações onde é preciso utilizar mais de duas camadas ocultas.

Ainda de acordo com Pedreira, pode-se provar que uma rede neural em camadas com apenas uma camada oculta é capaz de representar qualquer mapeamento contínuo. Como a prova deste resultado não é construtiva, dado um problema (mapeamento contínuo) real, sabemos com certeza que ele pode ser resolvido com uma rede de uma camada oculta, embora não saibamos nem qual rede resolve este problema, nem tão pouco quais os valores de pesos que devem ser associados. Para resolver este problema, faz-se mister projetar uma arquitetura de rede e calcular iterativamente os pesos³⁴.

³³ Haykin (1999).

³⁴ Este procedimento será abordado um pouco mais adiante.

5.3 - Um modelo de neurônio³⁵

Neste momento, para que se entenda um pouco melhor como funciona uma rede neural, será feito, a ilustração de um neurônio específico³⁶ e de sua funcionalidade. Esta ilustração pode ser vista na figura 5.2 a seguir.

Figura 5.2: Modelo de Neurônio

Isto é, tem-se X_p entradas e w_{kp} é o peso que liga a entrada p ao neurônio k , sendo que ϕ é chamada de Função de Ativação.

³⁵ McCulloch & Pitts criaram em 1943 um modelo **simplificado** de neurônio, mas apenas em 1962, Frank Rosenblatt faz, pela primeira vez, simulações de redes neurais em um computador digital. Entretanto, após a publicação do livro de Minsky & Papert em 1969, onde os autores levantam algumas questões que aparentemente inviabilizariam o uso prático das redes neurais, os desenvolvimentos ficam quase que paralisados por mais de uma década. Apenas em 1982, quando John Hopfield propõe uma arquitetura diferente e exemplifica o seu uso potencial através do famoso problema do caixeiro viajante, é que começa a renascer o interesse pelas redes neurais. (Pedreira, 2000).

³⁶ Apenas para exemplificar, é como estivessemos verificando o que existe “dentro” de apenas um dos quatro neurônios da *rede feedforward* expressa na figura 5.1

Então:

$$u_k = \sum_{j=1}^P w_{kj} x_j \quad \text{e} \quad Y_k = \varphi(u_k)$$

É extremamente importante salientar que a Função de Ativação é a parte não linear de cada neurônio³⁷. Podemos então interpretar a rede neural como uma combinação destas não linearidades, o que possibilita a construção de relações não lineares complexas.

São três os principais tipos de funções de ativação utilizadas em redes neurais:

- a) sigmóides e tangentes hiperbólicas;
- b) lineares e
- c) radiais.

A função sigmóide, por sua vez, é representada pela seguinte equação:

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \quad a > 0$$

Atenta-se que, esta é, juntamente com a tangente hiperbólica, a não linearidade mais utilizada. Esta função tem a propriedade de aliar uma região quase linear à propriedade de saturação. É importante saber que a família das sigmóides é diferenciável e não decrescente. O gráfico de uma sigmóide pode ser visto na figura 5.3 a seguir.

A função tangente hiperbólica possui as mesmas características da sigmóide, mas opera na faixa “-a” a “+a” (e não de 0 a “a”). Sua equação é expressa por:

$$\varphi(a) = \tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}} \quad a > 0$$

³⁷ Esta não linearidade é em geral muito simples. Ver figura 5.3 a seguir.

Já a função linear, é muito utilizada em neurônios da camada de saída, quando não é desejável o efeito de saturação³⁸ das funções sigmóides e tangentes hiperbólicas. As funções radiais, por sua vez gaussianas, são o fundamento das redes de bases de funções radiais³⁹.

Figura 5.3: Sigmóide

5.4 - Modelagem e aprendizado em redes neurais

A modelagem em redes neurais pode ser sintetizada basicamente nos seguintes passos:

- 1) Escolher os exemplos (observações) a serem usados para o treinamento, de modo que o ambiente seja bem representado. Salienta-se que os exemplos consistem em pares do tipo (entrada, saída desejada)⁴⁰.

³⁸ Relacionado com o método de otimização usado que, no algoritmos de *Backpropagation*, se baseia na regra da cadeia e no método de máxima descida. Ver Haykin (1999).

³⁹ Haykin (1999) e Pedreira (2000).

EXEMPLOS

- 2) Escolher os exemplos que serão utilizados para validação (avaliação da capacidade de generalização do modelo).
- 3) Escolher a arquitetura apropriada.
- 4) Escolher o algoritmo para treinamento da rede.

Especificada a metodologia de modelagem quando do uso de redes neurais, é importante que se entenda como a rede “adquire” conhecimento. O conhecimento em redes neurais é adquirido através de um processo de aprendizado ou de treinamento. A informação é transformada e armazenada nas “densidades de conexão” ou pesos sinápticos (ou, como visto anteriormente, simplesmente “pesos”).

O conhecimento é então produzido utilizando estes pesos para combinar as não linearidades (o que implica como visto, numa distribuição da não linearidade na rede). Salienta-se que este conhecimento é incorporado sem que se tenha que explicitar modelos *a priori*, ou seja, através de um aprendizado por exemplos, o modelo é construído implicitamente a partir dos dados. Geralmente, com a metodologia tradicional, um modelo do ambiente deve ser pressuposto *a priori*, sendo os dados usados para validá-lo.

⁴⁰ Isto ocorre no chamado “aprendizado supervisionado” que será visto a seguir. No aprendizado “não supervisionado”, como o utilizado nas Redes de Kohonen, são apresentadas apenas as entradas que consistem numa representação qualquer do ambiente.

A rede neural aprende, então, o ambiente através de um processo iterativo de modificação dos pesos de interligação, a partir de estímulos fornecidos pelo ambiente. O tipo de aprendizado é determinado pelo modo com que se promove a adaptação dos parâmetros e isso pode ser feito de dois modos:

- 1) Aprendizado Supervisionado – usa-se um conjunto de pares, entrada e saída, previamente conhecidos que representam a realidade;
- 2) Aprendizado Não Supervisionado – não se usa um conjunto de exemplos previamente conhecidos. Uma medida da qualidade da representação do ambiente pela rede é estabelecida e os parâmetros são modificados de modo a otimizar esta medida. Este tipo de aprendizado é muito utilizado na área de reconhecimento de padrões⁴¹.

5.4.1 - Aprendizado supervisionado e o algoritmo de *Backpropagation*

Neste trabalho, utilizou-se apenas o aprendizado supervisionado. Este é baseado no erro de saída da rede. Obviamente, para que seja possível calcular um erro na saída, é preciso que haja um conjunto de pares (entrada – saída desejada) conhecidos. Desta forma, os pesos são ajustados visando minimizar o erro de saída.

O algoritmo mais freqüentemente empregado no aprendizado supervisionado de redes neurais é conhecido como Retropropagação do Erro (*Backpropagation*). Destaca-se que, este algoritmo tem fundamentos matemáticos baseados na regra da cadeia e no método do gradiente descendente⁴².

⁴¹ Ver Bishop (1995), Kohonen (1997) .

⁴² Haykin (1999).

O algoritmo de *Backpropagation* consiste basicamente em:

- 1) Calcular o erro na saída da rede.
- 2) Retropropagar o erro na rede calculando de que forma modificações nos pesos afetam o erro.
- 3) Modificar os pesos de modo a minimizar o erro.

É importante ressaltar que o objetivo é sempre minimizar o erro médio. Para isto, faz-se modificações nos pesos padrão a padrão. De acordo com Pedreira (2000), a média destas modificações (sobre todos os exemplos) é uma estimativa da modificação que se obteria em se minimizando o erro sobre todo o conjunto.

Desta forma, o mecanismo de aprendizado pode ser então sintetizado em três passos:

Passo 1) a rede é estimulada pelo ambiente⁴³ através de exemplos extraídos do mesmo.

Passo 2) os pesos são modificados através de um procedimento iterativo, como resultado do passo 1.

Passo 3) a rede passa a responder ao ambiente de uma nova forma em decorrência das mudanças efetuadas no passo 2.

5.4.2 - Generalização

Uma rede neural opera em duas fases distintas: treinamento (*in sample*) e testes (*out of sample*), quando se pretende atingir a generalização. Na primeira fase, a rede retira do ambiente o conhecimento. Na segunda, a qualidade da primeira fase é posta em cheque, para que possa ser utilizada para os fins almejados quais sejam: projeção, classificação, segmentação, entre outros.

⁴³ Este processo se dá apresentando-se inúmeras vezes o conjunto entrada-saída para a rede. (Equivale ao número de iterações do processo).

Desta forma, o que se faz, na prática, é escolher um conjunto de dados para o treinamento e separar um outro para a generalização⁴⁴.

É importante salientar que uma rede neural generaliza com sucesso quando a relação entrada-saída aprendida no treinamento é representativa do ambiente gerador. Outro aspecto interessante é que quanto mais simples a arquitetura, maior a probabilidade de uma generalização adequada⁴⁵. A justificativa consiste no fato de que, ao adicionarmos complexidade ao modelo, se está sujeito a uma generalização fraca pois, por ser excessivamente complexo o modelo capta a maioria dos detalhes dos dados de treinamento em detrimento de propriedades mais gerais.

No caso das redes neurais, dado o processo de estimação e ajuste dos pesos, quanto mais camadas e neurônios nas camadas ocultas existirem, mais complexo será o modelo e, portanto, mais “pobre” será a generalização. Isto implica no fato de que, quanto menor a quantidade de dados, mais simples deve ser o modelo.

⁴⁴ O conjunto de teste não deve ser usado para validar o modelo.

⁴⁵ Segundo Pedreira (2000), provavelmente haverá generalização se $N > W/\epsilon$, onde N = número de exemplos usados no treinamento; W = número de pesos da rede e ϵ = erro admitido na fase de teste. Isto implica que, se o erro admitido é de 10% o número de exemplos deve ser aproximadamente 10 vezes o número de pesos da rede. (Pedreira, 2000).

6 - MODELOS LINEARES PARA A DEMANDA DE GASOLINA AUTOMOTIVA NO BRASIL

6.1 - Modelo Box & Jenkins

6.1.1 - Identificação do modelo

Seguindo os fundamentos teóricos da metodologia Box & Jenkins, partiu-se para a modelagem da série de demanda mensal de gasolina automotiva no Brasil segundo este método.

Primeiramente, fez-se o gráfico da série original (figura 6.1), juntamente com os gráficos das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) – figuras 6.2 e 6.3 respectivamente. Percebe-se, na figura 6.2, que a FAC decresce muito lentamente o que corrobora a constatação da não estacionaridade da série, observada na figura 6.1. Há, portanto, necessidade de diferenciá-la. *A priori*, constata-se também que a série apresenta um certo padrão sazonal, com picos sazonais em dezembro.

Figura 6.1: Gráfico da Demanda de Gasolina Automotiva no Brasil (m^3)

Figura 6.2: Função de Autocorrelação (FAC)

Figura 6.3: Função de Autocorrelação Parcial (FACP)

Após tirar a primeira diferença, obteve-se novamente o gráfico da FAC e da FACP (figuras 6.4 e 6.5 respectivamente). Observa-se, neste momento, que a FAC apresenta um corte nos *lags* 1 e 12 (1º *lag* sazonal) e a FACP apresenta também um decréscimo nos dois primeiros *lags* não sazonais (1 e 2) e um corte nos *lags* sazonais⁴⁶.

⁴⁶ O *lag* 3 significativo na FAC (figura 6.4) pode ser definido, a princípio, como espúrio mesmo porque está muito próximo do limite da região de significância. O mesmo pode ser dito do *lag* sazonal 12 na FACP (figura 6.5). Isto deixa claro porque foi falado no Quadro 01 de características **teóricas** das funções de autocorrelação. Na prática, existem variações em tornos deste padrão teórico, o que exige uma grande atenção no exercício da modelagem.

Figura 6.4: Função de Autocorrelação (FAC) após a 1ª diferença

Figura 6.5: Função de Autocorrelação Parcial (FACP) após a 1ª diferença

Enfim, seguindo o procedimento Box & Jenkins de identificação de um modelo, através da visualização dos gráficos da Função de Autocorrelação e Autocorrelação Parcial, pode-se perceber que, observando os *lags* não sazonais, verifica-se que a FAC apresenta um corte no *lag* 1 e a FACP decresce. Identifica-se então a ordem $q = 1$, ou seja um MA(1). Como já foi feito uma primeira diferença, o que dá a ordem de diferenciação $d = 1$, tem-se um modelo ARIMA (0,1,1).

Observando-se agora os *lags* sazonais (12, 24, 36, etc), vê-se que, na FACP, as autocorrelações nestes *lags* são não significantes, ao passo que na FAC pode-se assumir que existe um corte no *lag* 1 sazonal (ou *lag* 12). Percebe-se aqui que há uma variação em torno do padrão teórico definido no Quadro 01. Neste caso, assumiu-se a seguinte interpretação para o comportamento da FAC e FACP: dado que existe um corte no *lag* 1 sazonal e a FACP é quase não significativa nos *lags* sazonais, assume-se uma aproximação em relação ao padrão teórico “decréscimo na FAC e corte na FACP” o que nos dá um AR (1) na parte sazonal. Frente a estas constatações, identificou-se⁴⁷ portanto um SARIMA (0,1,1)x(1,0,0)₁₂.

Pela equação (1) que representa a formulação geral de um SARIMA, tem-se que um SARIMA (0,1,1)x(1,0,0)₁₂ indica a seguinte forma:

$$(1 - \Phi B^{12}) \nabla Z_t = (1 - \theta B) a_t$$

$$(1 - \Phi B^{12})(1 - B)Z_t = a_t - \theta a_{t-1}$$

$$(1 - \Phi B^{12})(Z_t - Z_{t-1}) = a_t - \theta a_{t-1}$$

$$Z_t - \Phi Z_{t-12} - Z_{t-1} + \Phi Z_{t-13} = a_t - \theta a_{t-1}$$

⁴⁷ Salienta-se que é comum na metodologia Box & Jenkins, chegar-se na identificação de mais de um modelo para um mesmo fenômeno. Sendo assim, alguns analistas poderiam identificar, neste caso, um decréscimo a partir do *lag* 1 tanto na FAC quanto na FACP o que implicaria num modelo ARIMA (1,0,1)₁₂ na parte sazonal (isto pois existe um corte no *lag* 1 sazonal (ou *lag* 12) na FAC e este mesmo *lag* está perto do intervalo de significância na FACP).

$$Z_t = Z_{t-1} + \Phi Z_{t-12} - \Phi Z_{t-13} + a_t - \theta a_{t-1}$$

Já que este foi o modelo SARIMA identificado para a demanda de gasolina automotiva no Brasil, é importante interpretar esta formulação matemática que diz que: a demanda de gasolina no mês atual é influenciada pela demanda do mês anterior, pela demanda de 12 meses atrás e pela demanda de treze meses atrás. Existe uma coerência lógica nesta formulação tendo em vista que a demanda no mês anterior nos dá pelo menos um nível da demanda no mês seguinte, caracterizando uma certa inércia no fenômeno demanda o que é fundamentado na teoria econômica. Já a influência da demanda de 12 meses atrás se deve à sazonalidade que podia ser verificada no gráfico da série, ou seja, existe um padrão sazonal na série. Por sua vez, existe também uma correlação entre a demanda de 12 meses atrás, com o seu mês antecedente que é a demanda de 13 meses atrás. Quanto ao termo a_{t-1} presente na formulação, trata-se de uma influência do erro cometido pelo modelo no mês anterior. Esta relação, apesar de não haver uma explicação econométrica clara, advém do uso deste tipo de metodologia que conjuga tanto uma parte autorregressiva (AR) quanto uma parte de “média móvel” (MA) no erro.

Identificado este modelo, passou-se à fase de estimação dos parâmetros Φ e θ . Os valores estimados foram respectivamente: 0,8265 e 0,7487 (ambos significantes estatisticamente)⁴⁸. O modelo identificado para demanda de gasolina então assume a seguinte forma⁴⁹:

$$Z_t = Z_{t-1} + \underset{[0,06]}{0,8265} Z_{t-12} - \underset{[0,06]}{0,8265} Z_{t-13} + a_t - \underset{[0,06]}{0,7487} a_{t-1}$$

⁴⁷ Atenta-se que, neste trabalho, este modelo – SARIMA (0,1,1)x(1,01)₁₂ – também foi estimado, mas apresentou desempenho pior que o daquele realmente identificado por mim (que é inclusive mais parcimonioso).

⁴⁸ Como o parâmetro Φ é o mesmo para Z_{t-12} e Z_{t-13} , o que interessa então é a diferença entre as demandas doze e treze meses atrás.

⁴⁹ O valor entre colchetes abaixo do parâmetro refere-se ao seu erro padrão.

Onde:

Z_t = demanda de gasolina no tempo t

Z_{t-1} = demanda de gasolina no tempo $t-1$

Z_{t-12} = demanda de gasolina no tempo $t-12$

Z_{t-13} = demanda de gasolina no tempo $t-13$

a_t = ruído no tempo t

a_{t-1} = ruído no tempo $t-1$

Finalmente, estimados os parâmetros, fez-se os testes de aderência onde verificou-se que, como desejado, o ruído gerado por este modelo não apresenta estrutura de correlação, ou seja, é um ruído branco. Pode-se verificar este fato ao olhar o gráfico da autocorrelação do erros (FAC dos erros) expresso na figura 6.6. Vê-se que as autocorrelações são insignificantes para todos os lags⁵⁰.

Figura 6.6: Função de Autocorrelação (FAC) dos Erros

⁵⁰ Novamente pode parecer uma contradição o que foi dito, pois os lags 9 e 20 estão na fronteira de significância, entretanto na literatura, autocorrelação nestes lags são raros e, portanto tratadas como correlações espúrias.

Enfim, percebe-se que, neste capítulo, atingiu-se um dos objetivos primordiais de análise de uma série temporal, segundo a metodologia Box & Jenkins, que consiste no fato de: dada a existência de uma série temporal não “branca” (isto é, que apresenta uma estrutura de dependência serial), achar o modelo matemático que descreva esta dependência serial e a transforme num ruído branco⁵¹.

Elaborou-se, portanto, um modelo autoprojeto para a série de demanda de gasolina automotiva no Brasil. O próximo passo da metodologia de modelagem desenvolvida nesta dissertação consiste em buscar novas causalidades para explicar o comportamento desta série. Isto será feito via regressão dinâmica. Posteriormente, as eficiências destes modelos serão comparadas, de forma que se possa escolher o modelo que ofereça maior capacidade preditiva.

6.2 - O modelo de regressão dinâmica

Neste momento apresenta-se o modelo que foi encontrado para a série de demanda de gasolina automotiva no Brasil. Baseando-se na teoria econômica procurou-se verificar a elasticidade-preço e a elasticidade renda da demanda de gasolina automotiva.

O termo elasticidade é empregado para descrever as características de uma relação entre duas variáveis. Neste sentido, a elasticidade é definida como o limite da razão entre a variação proporcional em uma variável em relação à variação proporcional em outra variável.

⁵¹ Se uma série já é branca, então não existe um modelo univariado para ela (o que não implica na inexistência de outros modelos).

Ao considerar, por exemplo, uma equação de demanda envolvendo preço (x) e quantidade demandada (y) onde demanda é função do preço tem-se que: $y = f(x)$. Assumindo que x é um número real não negativo e que a função de demanda é contínua, se x muda por uma quantia Δx , então y muda por uma quantia Δy . A variação relativa em y é então $\Delta y/y$ e a variação relativa em x é $\Delta x/x$.

Isto nos dá que a variação relativa em y (quantidade demandada) por unidade de variação relativa em x (preço) é dada por:

$$\frac{\Delta y}{y} \div \frac{\Delta x}{x} \quad \text{ou equivalente} \quad \frac{x}{y} \div \frac{\Delta y}{\Delta x} \quad (6.1)$$

Como $\Delta y = f(x + \Delta x) - f(x)$, (1) pode ser reescrita como:

$$\frac{x}{y} \cdot \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad (6.2)$$

No limite, quando Δx se aproxima de zero, tem-se que

$$\frac{x}{y} \cdot \frac{dy}{dx} \quad (6.3)$$

Este limite dá a variação percentual aproximada na demanda que corresponde a uma variação de 1% no preço. Ele é denominado **elasticidade-preço** da demanda⁵².

⁵² Obviamente o mesmo raciocínio vale para a elasticidade-renda da demanda.

Esta rápida demonstração é considerada importante, pois ao estimar o modelo de regressão dinâmica para a demanda de gasolina automotiva no Brasil, trabalhou-se com o logaritmo neperiano das variáveis envolvidas na equação. Matematicamente pode-se mostrar que dado $y = f(x)$, ao se diferenciar $\ln(y)$ com respeito a $\ln(x)$, o coeficiente de regressão estimado já dá a medida da elasticidade.

$$\text{Portanto, dado } \ln(y) = \beta \ln(x) \quad (6.4)$$

Definindo primeiro que $u \equiv \ln(y)$ e $v \equiv \ln(x)$. Pode-se então observar uma cadeia de relações ligando u a y e x a v da seguinte maneira:

$$u \equiv \ln(y) \quad y = f(x) \quad x \equiv e^{\ln(x)} \equiv e^v \quad (6.5)$$

Logo, a derivada de $\ln(y)$ com respeito a $\ln(x)$ é:

$$\frac{d(\ln y)}{d(\ln x)} = \frac{du}{dv} = \frac{du}{dy} \cdot \frac{dy}{dx} \cdot \frac{dx}{dv} \quad (6.6)$$

Então de (6.5) e (6.6) tem-se que:

$$\frac{d(\ln y)}{d(\ln x)} = \left(\frac{d}{dy} \ln y \right) \cdot \left(\frac{dy}{dx} \right) \cdot \left(\frac{d}{dv} e^v \right) = \frac{1}{y} \cdot \frac{dy}{dx} \cdot e^v = \frac{1}{y} \cdot \frac{dy}{dx} \cdot x = \frac{x}{y} \cdot \frac{dy}{dx}$$

Isto é:

$$\frac{d(\ln y)}{d(\ln x)} = \frac{x}{y} \cdot \frac{dy}{dx} \quad (6.7)$$

Portanto, de (6.3) (6.4) e (6.7) mostra-se que, ao se trabalhar com as variáveis em escala logarítmica, o coeficiente de regressão estimado β já fornece a medida da elasticidade entre a variável a ser explicada e a variável causal.

Feitas estas considerações e já explicitado como é construído um modelo de regressão dinâmica (capítulo 4), serão apresentados a seguir, dois modelos que foram criados para a série de demanda de gasolina.

O primeiro modelo de Regressão Dinâmica (RD1) reflete a seguinte relação⁵³:

$$\ln(\text{gasolina})_t = -0,25.\ln(\text{preço})_t + 0,32.\ln(\text{renda})_t + 0,50.\ln(\text{gasolina})_{t-1} + 0,40.\ln(\text{gasolina})_{t-12} \quad (6.8)$$

[0,06] [0,09] [0,07] [0,08]

Deve-se salientar o aspecto lógico expresso na relação acima. Primeiramente deve-se observar os sinais do coeficientes para as variáveis preço e renda⁵⁴. Pode-se observar assim uma elasticidade-preço negativa e uma elasticidade-renda positiva como realmente deveria acontecer⁵⁵.

Quanto à presença das defasagens t-1 e t-12 no modelo também tem-se uma explicação coerente. A presença de primeira defasagem (ou seja da gasolina em t-1) baseia-se na teoria econômica que sugere haver uma certa inércia (no curto prazo) nos hábitos de consumo, que conferem uma dinâmica própria à demanda. Já a presença do termo t-12 deve-se à característica sazonal desta série⁵⁶.

Entretanto, a partir de uma análise mais aprofundada, verificou-se que, “estranhamente”, a inclusão do *lag* 3 (em substituição ao *lag* 1) da demanda de gasolina trazia melhorias significativas no resultado. Ao se analisar a figura 6.7 que expressa a função de autocorrelação

⁵³ Os valores em colchete são os erros padrões dos coeficientes.

⁵⁴ Tanto renda quanto preço são expressos em termos reais (valores de dezembro de 1999), ou seja, usou-se o “Rendimento médio nominal, trabalho principal, das pessoas ocupadas de 15 anos ou mais” (dado do IBGE) e o preço nominal ambos deflacionados pelo Índice de Preços ao Consumidor (IPC) da Fundação Getúlio Vargas.

⁵⁵ Sabe-se que estes coeficientes informam a elasticidade no ponto (Chiang, 1982). Para a elasticidade-preço, por exemplo, temos que um aumento de 1% no preço, causa um decréscimo de 0,25% na demanda de gasolina automotiva.

⁵⁶ Inclusive já verificada quando da elaboração do modelo Box & Jenkins que originou o SARIMA (0,1,1)x(1,0,0).

dos erros gerados pelo modelo RD1 (equação 6.8), vê-se realmente que há uma significância do $lag\ 3$ ⁵⁷.

Figura 6.7: Função de Autocorrelação (FAC) dos erros do modelo RD1

Apesar de, a princípio, não haver uma explicação para este acontecimento, neste caso, os dados estão “informando” que a demanda três meses atrás influencia a demanda atual. Buscou-se nos dados então, verificar o que estava acontecendo.

Como pode ser observado na figura 6.8 a seguir, verificou-se que, principalmente a partir de 1997, a variação percentual da demanda apresentava uma queda, no tempo $t-3$, acompanhada de duas altas, e depois novamente uma queda⁵⁸.

⁵⁷ Temos também o $lag\ 10$ significativo mas, neste caso, podemos interpretá-lo como espúrio.

⁵⁸ Nos demais anos isto não fica muito claro.

Ao olhar para o segundo semestre⁵⁹, por exemplo, verifica-se que ocorre uma queda da demanda em agosto (até explicada pelo comportamento sazonal em julho onde normalmente há um aumento nas vendas devido às férias), seguida de um aumento da demanda em setembro e outubro e de novamente uma queda da demanda em novembro.

Figura 6.8: Gráfico da variação percentual da demanda de gasolina automotiva: 1997-1999

Como este fenômeno não está presente em toda a série temporal, acredita-se ser “meio” forte falar numa sazonalidade a cada três meses. Entretanto, algo realmente acontece a partir do ano de 1997 que, a princípio, não tem uma explicação clara⁶⁰.

⁵⁹ Excetuando o ano de 1999, no primeiro semestre, podemos observar fenômeno semelhante. Isto é, uma queda da demanda em fevereiro, acompanhada de altas em março e abril, com queda em maio.

⁶⁰ Foi feito um contato com a ANP para apresentar esta constatação, mas não se chegou a uma razão clara para o ocorrido.

Enfim, frente a esta constatação, elaborou-se um modelo onde, além da presença do *lag* 12 da demanda de gasolina, entrou-se também com o *lag* 3. Chegamos então ao segundo modelo de Regressão Dinâmica (RD2) expresso a seguir⁶¹:

$$\ln(\text{gasolina})_t = -0,26.\ln(\text{preço})_t + 0,35.\ln(\text{renda})_t + 0,40.\ln(\text{gasolina})_{t-3} + 0,49.\ln(\text{gasolina})_{t-12} \quad (6.9)$$

[0,05] [0,08] [0,07] [0,07]

Novamente este modelo mantém a coerência lógica. Todos os parâmetros são significativos e pode-se também observar o sinal dos coeficientes para preço e renda (negativo e positivo respectivamente). Temos ainda a presença do *lag* 12 (sazonal) e a presença do *lag* 3. Vale a pena ressaltar que este modelo, foi o melhor modelo de regressão dinâmica obtido neste trabalho⁶². Inclusive, observando-se a figura 6.9 a seguir, vê-se que o gráfico da função de autocorrelação dos erros, mostra claramente que o erro gerado por este modelo caracteriza-se por ser um ruído branco.

⁶¹ Neste momento, exclui-se a demanda em t-1, pois este modelo apresentou um pior resultado em termos de eficiência preditiva.

⁶² Além de menor erro médio de previsão (que será abordado no capítulo 8), este modelo apresentou maior poder de explicação ($R^2 = 93,56$ contra 92,33 do modelo RD1) além de apresentar menor BIC (*Bayesian Information Criterion*) que consiste num critério de escolha entre dois modelos. De forma geral, é um dos critérios baseados na teoria da informação que sugerem como modelo “definitivo” aquele que minimiza a variância dos resíduos, ao mesmo tempo em que penalizam o excesso de parâmetros no modelo.

Figura 6.9: Função de Autocorrelação (FAC) dos erros do modelo RD2

Enfim, acredita-se que, dentro da estratégia *bottom-up*, conseguiu-se evoluir na modelagem, ou seja, chegou-se a um modelo que busca, além da própria série de demanda, novas causalidades nas variáveis preço e demanda.

Acredita-se que este modelo é importante na medida em que, além de apresentar melhores resultados preditivos (como será visto mais adiante), fornece maiores possibilidades de análise pois traz, por exemplo, as relações de elasticidade-preço e elasticidade-renda da demanda de gasolina automotiva no Brasil.

Sabe-se, entretanto, que os modelos desenvolvidos até então assumem pressupostos de linearidade nas causalidades. Sendo assim, o próximo passo da metodologia deste trabalho fundamentar-se-á na tentativa de se obter um modelo ainda melhor, ou seja, será feito uso de um modelo que faz um tratamento das não-linearidades existentes no mundo real. Este será um modelo de redes neurais.

7 – MODELOS NÃO-LINEARES PARA A DEMANDA DE GASOLINA AUTOMOTIVA NO BRASIL: AS REDES NEURAIIS

Dentro da metodologia empregada nesta dissertação, busca-se encontrar um modelo para previsão de gasolina automotiva que, mesmo sendo uma simplificação da realidade, possa se aproximar o máximo possível desta. Obviamente, o resultado disto é a elaboração de um modelo que possa compreender o dinamicismo da série de demanda de gasolina automotiva no Brasil, gerando previsões ótimas desta.

Como dito em momento anterior, o uso de redes neurais para previsão de séries temporais tem, comprovadamente, gerado ótimos resultados. Portanto, buscou-se a aplicação deste método para prever a demanda de gasolina automotiva. Atenta-se que foram elaborados dois modelos de redes neurais⁶³ para previsão desta série.

7.1. - Modelo de Redes Neurais I

De acordo com o conceito de aprendizado supervisionado, foram escolhidos os exemplos (entrada - saída desejada) para o treinamento, separando-se um conjunto de dados para se testar o poder de generalização do modelo. Como trabalhamos com dados mensais de janeiro de 1991 a dezembro de 1999, usou-se os dados de demanda até junho de 1999 para treinamento, e o restante (julho a dezembro de 1999) para generalização.

⁶³ Na verdade foram elaborados vários modelos, sendo apresentados aqui apenas dois.

É importante lembrar que, como os dados têm periodicidade mensal, são feitas previsões de curto prazo, ou mais precisamente, um mês a frente. Entretanto para prever a demanda de gasolina no mês $(T + 1)$ ⁶⁴, existem várias possibilidades para o conjunto de entrada que será apresentada à rede neural para treinamento.

Mantendo-se uma seqüência lógica utilizada em todo o trabalho, buscou-se inspiração numa informação oferecida pelo primeiro modelo criado na metodologia de modelagem desta dissertação, ou seja, no modelo de Box & Jenkins. Como visto, o modelo encontrado foi um SARIMA $(0,1,1) \times (1,0,0)_{12}$ que informa que, segundo esta metodologia, o valor da demanda de gasolina no mês T , depende da demanda em $(t - 1)$, $(t - 12)$ e $(t - 13)$ ⁶⁵. Portanto, para prever a demanda em fevereiro de um determinado ano, por exemplo, necessita-se da demanda em janeiro deste mesmo ano, da demanda em fevereiro do ano anterior e, por fim da demanda em janeiro do ano anterior.

Para o primeiro modelo de redes neurais (RN1), tem-se então o seguinte esquema funcional:

Figura 7.1: Par entrada – saída desejada para RN1

⁶⁴ Com origem de previsão no tempo T .

Ressalta-se que buscou-se utilizar uma arquitetura de rede que fosse a mais simples possível. Como pode-se observar, usou-se três entradas. Foi utilizada uma rede multicamadas, com uma camada escondida (com três neurônios) e um neurônio na camada de saída.

7.2 - Modelo de Redes Neurais II: O Modelo Híbrido

No segundo modelo de redes neurais, procurou-se construir um **modelo híbrido**, ou seja, um modelo que trabalhasse com métodos estatísticos, entretanto somado à eficiência no tratamento das não linearidades que uma rede neural oferece. O modelo estatístico escolhido foi o segundo modelo de regressão dinâmica (RD2). As razões pela escolha fundamentam-se nos bons resultados preditivos deste modelo quando comparado ao modelo de Box & Jenkins e ao primeiro modelo de regressão dinâmica. Atenta-se que este modelo expressa a seguinte relação de causalidade entre a variável a ser explicada (demanda de gasolina automotiva) e as variáveis exógenas:

$$\ln(\text{Demanda})_t = f[\ln(\text{Preço}_t; \text{Renda}_t; \text{Demanda}_{t-3}; \text{Demanda}_{t-12})]$$

A lógica utilizada na elaboração no segundo modelo de redes neurais (RN2) foi, em síntese, a de usar uma rede neural para prever o erro originado por este modelo de regressão dinâmica de forma que se pudesse otimizar o seu resultado.

Observando a equação a seguir vê-se que:

$$\ln(D)_t = \underbrace{-\beta_1 \ln(P)_t + \beta_2 \ln(R)_t + \beta_3 \ln(D)_{t-3} + \beta_4 \ln(D)_{t-12}}_{\text{DETERMINÍSTICO}} + \underbrace{\varepsilon_t}_{\text{ALEATÓRIO}}$$

⁶⁵ Além do ruído em $(t - 1)$ que não foi usado na rede.

Onde:

D_t = logaritmo da demanda no tempo t

P_t = logaritmo do preço no tempo t

R_t = logaritmo da renda no tempo t

D_{t-3} = logaritmo demanda no tempo $t - 3$

D_{t-12} = logaritmo demanda no tempo $t - 12$

ε_t = ruído no tempo t

É claro que todo modelo matemático representa uma simplificação da realidade e, portanto, por mais elaborado que seja, sempre apresentará um componente errático. Baseado nisto, trilha-se a seguinte linha de raciocínio: pode-se tomar o lado esquerdo da equação acima e denominá-lo Y (valor real da demanda). A parte determinística do modelo fornece, então, um valor estimado Y^* . Este valor entretanto, é estimado por um modelo que assume pressupostos de linearidade. Deste modo, a diferença $Y - Y^*$ obviamente daria um erro (ε^*) que, teoricamente, representaria o aspecto não linear da causalidade.

Em outras palavras, este erro representa um desvio em relação à realidade e, dado que ele foi gerado através de uma estimativa obtida por um modelo linear, teoricamente ele é representante da parte que falta explicar, ou seja, dos “fenômenos não lineares”. Portanto, a motivação é a de usar uma rede neural (que, como visto, tem todas as vantagens de tratar não linearidades) para prever este “erro”.

Espera-se portanto que, o valor desta previsão do erro (feita pela rede neural), somada à estimativa Y^* (gerada pelo modelo de regressão dinâmica) terá que resultar numa estimativa melhor da demanda (Y^{**}), ou seja, uma estimativa que ao ser comparada com o valor real Y gere

menor um menor erro de previsão. Ou seja, se a rede neural for usada com coerência, o erro resultante de $Y - Y^{**}$ tem que ser menor, *a priori*, do que o erro de $Y - Y^*$.

Obviamente, para previsão do erro, é preciso que se defina novamente o conjunto de entrada – saída desejada.

Figura 7.2: Par entrada – saída desejada para RN2

Neste caso, para prever o erro no tempo t , foram utilizados o erro no instante $t - 1$ e no instante $t-12$ (informação fornecida à rede quanto ao aspecto sazonal). Podemos observar que fez-se uso de duas entradas para a rede. Foi utilizada novamente uma rede multicamadas, com uma camada escondida (com três neurônios) e um neurônio na camada de saída.

Enfim, salienta-se que, no próximo capítulo, será feita uma análise comparativa dos resultados de todos os modelos gerados seguindo a metodologia de trabalho expressa nesta dissertação.

8 – RESULTADOS NUMÉRICOS

Chega-se, neste momento, a um passo importante da metodologia desenvolvida neste trabalho. Após partir de um modelo mais simples (univariado) para um mais complexo (redes neurais) passando por um modelo de regressão dinâmica (que forneceu algumas informações adicionais no estudo da demanda, como os coeficientes de elasticidade), realiza-se neste capítulo uma análise comparativa dos resultados obtidos através destes modelos.

Atenta-se que para comparar os resultados preditivos alcançados por tais modelos, empregou-se as seguintes medidas de erro:

1- Erro de previsão:

$$e(i) = Z(i) - \hat{Z}(i)$$

2- Erro médio:

$$\bar{e} = \frac{\sum_{i=1}^N e(i)}{N}$$

3- Erro absoluto médio:

$$\left| \bar{e} \right| = \frac{\sum_{i=1}^N |e(i)|}{N}$$

4- Erro quadrático médio:

$$EQM = \frac{\sum_{i=1}^N e^2(i)}{N}$$

5- Erro percentual:

$$erro \text{ percentual}(i) = \frac{|e(i)|}{Z(i)} \times 100$$

6- MAPE:

$$MAPE = \frac{\sum_{i=1}^N erro \text{ percentual}(i)}{N}$$

7- U-Theil:

$$\hat{Z}(i) = Z(i-1) \longrightarrow \text{Método } Naive \text{ (Ingênuo)}$$

$$U - THEIL = \sqrt{\frac{\frac{1}{N-1} \sum_{i=1}^{N-1} e_i^2}{\frac{1}{N-1} \sum_{i=1}^{N-1} (Z(i) - Z(i-1))}}$$

Ressalta-se que, comumente, quando da comparação de modelos, são apresentados o MAPE destes. Entretanto, é fundamental que se apresente conjuntamente a medida de erro U-THEIL que permite fazer uma comparação em relação ao “método ingênuo” de previsão (ou método Naive)⁶⁶. Este, por sua vez indica que, a melhor previsão para o $Z(i)$, seria o $Z(i-1)$. Obviamente comparado com este método, o modelo o qual se verifica a adequabilidade precisa ser melhor. Isto significa que quanto menor o U-THEIL, melhor é a eficiência do modelo (sendo que ele precisa ser também menor do que 1).

⁶⁶ Pode acontecer de, na prática, um modelo apresentar uma MAPE pequeno mas ter um U-THEIL maior do que 1.

Adotou-se como método também, calcular o MAPE e o U-THEIL tanto no *in sample* quanto no *out of sample*⁶⁷. Observando a figura 8.1, na próxima página, por exemplo, podemos comparar a eficiência preditiva dos modelos gerados. O MAPE, como visto anteriormente, fornece o erro médio absoluto percentual. Cabem aqui então algumas considerações.

Pode-se observar que, no *in sample*, o segundo modelo de regressão dinâmica (RD2) apresenta melhor resultado do o que primeiro modelo de regressão dinâmica (RD1) assim como também se desempenha melhor do que o modelo de Box & Jenkins (BJ). Este resultado já era esperado, dado que se acredita que um modelo causal, *a priori*, se elaborado com coerência, teria que apresentar um resultado melhor que o de um modelo univariado tendo em vista que, ao levar em conta outras causalidades, ele dispõe de maior poder de explicação⁶⁸.

Nota-se, entretanto, que os modelos de melhor eficiência preditiva no *in sample* são os dois modelos onde se usam redes neurais, com o modelo híbrido (RN2) reduzindo significativamente o erro de previsão.

O mesmo acontece quando olha-se para o gráfico do U-THEIL de cada um dos modelos. Novamente, as redes neurais apresentam melhor desempenho, ficando mais nítida agora a diferença entre o primeiro modelo de redes neurais (RN1) e o modelo RD2.

Analisando agora os gráficos do MAPE e U-THEIL no *out of sample* (figuras 8.3 e 8.4 respectivamente) pode-se novamente constatar um melhor desempenho dos modelos que fazem uso das redes neurais.

⁶⁷ Sempre o período de julho a dezembro de 1999.

⁶⁸ Entretanto nem sempre isto acontece na prática.

Figura 8.1: Comparação do MAPE (*In sample*) Figura 8.2: Comparação do U-THEIL (*In sample*)

Figura 8.3: Comparação do MAPE (*Out of sample*)

LEGENDA:

BJ = Modelo Box & Jenkins

RD1 = Modelo de Regressão Dinâmica 1

RD2 = Modelo de Regressão Dinâmica 2

RN1 = Modelo de Redes Neurais 1

RN2 = Modelo de Redes Neurais 2

Figura 8.4: Comparação do U-THEIL (*Out of sample*)

LEGENDA:

BJ = Modelo Box & Jenkins

RD1 = Modelo de Regressão Dinâmica 1

RD2 = Modelo de Regressão Dinâmica 2

RN1 = Modelo de Redes Neurais 1

RN2 = Modelo de Redes Neurais 2

Novamente vê-se também que o segundo modelo de regressão dinâmica apresenta melhor eficiência preditiva que o modelo RD1 e do que o modelo BJ. Entretanto uma observação merece destaque. Os modelos que usam as redes neurais se houveram melhor, sendo que o primeiro modelo de redes neurais (RN1), reduz pela metade o MAPE no *out of sample*, apresentando também menor U-THEIL. Vê-se também uma diferença significativa do MAPE dos outros modelos para o erro do modelo de RN2.

Atenta-se também que, uma informação advinda dos dados torna ainda este resultado mais relevante. O conjunto de dados escolhido para validação dos modelos, ou seja, os meses de julho a dezembro de 1999 apresentaram um comportamento atípico se comparados ao mesmo período do ano anterior (Quadro 02).

Quadro 02: Variação da demanda de gasolina automotiva no Brasil no 2º semestre de 1999

Período (<i>Out of sample</i>)	Variação (%) em relação ao mês anterior	Variação (%) em relação ao mesmo do mês do ano anterior
Julho	4,28	4,66
Agosto	-8,56	-0,13
Setembro	0,43	-2,22
Outubro	0,86	-8,65
Novembro	-4,60	-0,64
Dezembro	18,59	-5,68

Fonte: Elaboração própria

Percebe-se que, em relação ao mês anterior, os dados apresentam um comportamento sazonal típico com aumentos em julho e dezembro⁶⁹. Entretanto, ao observar a variação percentual em relação ao mesmo período do ano anterior, vê-se que, no segundo semestre de 1999 há uma queda na demanda⁷⁰. Este fato provavelmente é resultante do aumento de preços ocorridos no segundo semestre deste ano e que fez com a demanda se processasse em patamar inferior ao do ano antecedente.

Frente a estes fatos, ou seja, dada a atipicidade deste ocorrido em relação aos dados usados para a validação dos modelos e, tendo em vista o melhor resultado obtido pelos modelos que fizeram uso das redes neurais justamente no *out of sample*, tem-se reforçada a idéia da eficiência das redes para trabalhar em ambientes não lineares o que possibilita tratar de situações onde é difícil criar modelos adequados da realidade ou em situações com freqüentes mudanças no ambiente.

⁶⁹ As quedas em agosto e novembro já foram abordadas no capítulo 6 quando se falou da necessidade de inclusão do *lag t - 3* da demanda no modelo de regressão dinâmica.

⁷⁰ O mesmo não ocorre nos anos anteriores.

9 - CONCLUSÃO

É sabido que séries econômicas como séries de demanda, preço, inflação entre outras, muitas das vezes apresentam um comportamento que dificulta uma modelagem através de modelos puramente lineares quando abordada a questão da eficiência preditiva dos modelos.

A metodologia desenvolvida nesta dissertação contemplou esta realidade através de uma estratégia *bottom-up* de construção de modelos que originou um modelo que apresentou fundamentalmente “melhores previsões” para a série de demanda de gasolina automotiva no Brasil.

O ponto de partida da modelagem foi a construção de um modelo autoprojeto através da metodologia Box & Jenkins. Ressalta-se que esta é fortemente abordada na literatura de séries temporais ainda hoje, tendo apresentado excelentes resultados de previsão nos mais variados contextos.

Num segundo momento, buscando argumentos na teoria econômica, construiu-se um modelo de regressão mas através de um método que contemplasse e contornasse, por exemplo, as hipóteses de independência dos erros assumidas pelos modelos de regressão tradicionais. Chegou-se assim a um modelo que expressa as relações de elasticidade-preço e elasticidade-renda e que ainda contempla as características sazonais da série de demanda de gasolina. Lembra-se que este modelo ainda forneceu informação sobre a significância da demanda no *lag* 3 relacionando-a com a demanda atual o que merece maior investigação.

Por fim, elaborou-se uma rede neural para prever a demanda mensal de gasolina automotiva, na tentativa de encontrar um modelo que tratasse os aspectos não lineares intrínsecos ao mundo real e aos fenômenos econômicos de tal forma que pudesse otimizar os resultados das previsões.

Acredita-se que estes objetivos foram alcançados na medida que foram encontrados não apenas um, mas dois modelos de redes neurais para realizar a previsão da demanda de gasolina no Brasil. No primeiro deles, onde foi fornecida para a rede somente os valores passados da demanda, obteve-se o melhor resultado, ou menor erro, na generalização e o segundo melhor desempenho na fase de estimação *in sample*, “perdendo” apenas para a outra rede neural elaborada.

Esta segunda rede neural, por sua vez, aparece na forma de um modelo híbrido, prevendo o erro resultante do melhor modelo de regressão dinâmica supracitado. Este modelo híbrido, além de reduzir de forma significativa o erro de previsão *out of sample*, também apresenta uma outra importante vantagem: mantém a possibilidade do tomador de decisões, tanto público quanto privado, controlar melhor a realidade na medida que lhes possibilita a consecução de cenários para as variáveis causais renda e preço. Este modelo terá sempre junto de si, a presença de uma rede neural que, pelo que foi verificado neste trabalho, é capaz de contribuir de forma que o valor estimado para a demanda de gasolina automotiva seja mais próxima do valor real.

Tendo em vista estas considerações, pode-se sugerir ainda, para trabalhos futuros, uma tentativa de se criar níveis para medir a incerteza das previsões realizadas pela rede neural. Como visto, também, neste trabalho usou-se apenas o aprendizado supervisionado para treinar a rede. Entretanto, frente a aspectos, por exemplo, regionais da demanda, o uso de redes que têm como umas de suas finalidades o reconhecimento de padrões, como os mapas auto-organizáveis de

Kohonen, pode constituir ferramenta valiosa para a previsão agregada de curvas de demanda que apresentem padrões similares.

Por fim ressalta-se que a metodologia de construção de modelos utilizada para prever a demanda mensal de gasolina automotiva no Brasil, pode ser aplicada, em princípio, a qualquer série temporal de interesse.

Além disso, comprovou-se que, em termos de eficiência preditiva, o uso das redes neurais para previsão em séries econômicas, como séries de demanda, por exemplo, pode constituir uma poderosa ferramenta auxiliar na tomada de decisão, ainda mais quando combinado aos modelos estatísticos clássicos.

10 - BIBLIOGRAFIA

Barros, R.P.de.; Ferreira, S.S.. *Um modelo econométrico para a demanda de Gasolina pelos Automóveis de Passeio*. IPEA, 1982.

Barros, M.; Souza, R.C. *Regressão Dinâmica*. Núcleo de Estatística Computacional. PUC-Rio, 1995 .

Bishop, C.M.. *Neural Networks for Pattern Recognition*. Oxford : Clarendon Press, 1995.

Box, G. E. P., Jenkins, G. M., *Time Series Analysis, Forecasting and Control*, San Francisco, Holden-Day, 1994.

Buonfiglio, A.. *Modelo de Previsão de Consumo de Gasolina e Álcool Hidratado*. Petrobrás, 1988.

Chiang, A. *Matemática para Economistas*. São Paulo : McGraw Hill, 1982.

Cochrane, D.; Orcutt, G.H.. Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms. *Journal of the American Statistical Association*, 44, p. 32-61, 1949

Correia, E.; Mattos, J.B.de.. *Metodologia de Previsão de Demanda de Carburantes automotivos para o Ciclo Otto*, Petrobrás, 1999.

Dudewicz, E.J.; Mishra, S.N.. *Modern Mathematical Statistics*. Wiley, 1988.

Eltony, M. (1991) - *A model for passenger car gasoline demand in Canada*. Surrey Energy Economics Centre, Seeds 58, junho.

Gujarati, D.N.. *Basic Econometrics*. McGraw Hill, 1995.

Goodrich, R.L.; Stellwagen, E.A.. *Forecast Pro for Windows*. Business Forecast Systems, 1999.

Hamilton, J.D.. *Time Series*, Princenton, 1994.

Haykin, S., *Neural Networks: A Comprehensive Foundation*, New Jersey, Prentice Hall, 2^o Edição, 1999.

Hsu, Y., Yang, C., *Design of Artificial Neural Networks for Short-Term Load Forecasting. Part I: Self-Organising Feature Maps for Day Type Identification*, IEE Proc., Generation, Transmission and Distribution, vol. 138, n^o 5, sep. 1991, pp. 407-413.

Kouris, G. (1983) - *Fuel Consumption for road transport in the USA*. Energy Economics, vol. 5, n2, pp. 89-99.

Kohonen, T. *Self organizing maps*. Springer, 1997

Khotanzad, A., Hwang, R.C., Abaye, A. & Maratukulam, D., *An Adaptive Modular Artificial Neural Network Hourly Load Forecaster and its Implementation at Electric Utilities*, IEEE Trans. on Power Systems, vol. 10, n^o 3, aug. 1995, pp. 1716-1722.

Lee, K. Y., Cha, Y. T. & Park, J. H., “*Short-Term Load Forecasting Using an Artificial Neural Network*”, IEEE Trans. on Power Systems, vol. 7, nº 1, feb. 1992, pp. 124-132.

Lourenço, P.M., *Um Modelo de Previsão de Curto Prazo de Carga Elétrica Combinando Métodos Estatísticos e Inteligência Computacional*, PUC - Tese de Doutorado, junho, 1998.

Montgomery, D.C., Johnson, L.A., *Forecasting and Time Series Analysis*, New York, McGraw-Hill Book Co., 1990.

Park, D.C., El-Sharkawi, M.A., Marks II, R.J., Atlas, L.E., Damborg, M.J., *Electric Load Forecasting an Artificial Neural Network*, IEEE Trans. on Power Systems, vol. 6, nº 2, may. 1991, pp. 442-448.

Pedreira, C.E. *Redes Neurais*. (Notas de aula). PUC-Rio, 2000.

Pelikán, E., Mateija, P., Sláma, M., Vinkler, K., *Interactive Forecasting of the Electric Load Using Kohonen Self-Organizing Feature Map*, Proc. of World Congress of Neural Networks, San Diego, California, sep. 1996, pp. 15-20.

Sobral, A.P.. *Modelo de Previsão Horária de Carga Elétrica para Light*. PUC-Rio, Dissertação de Mestrado, 1999.

Souza, R.C.; Iachan, R.; Da Silva, N.de M.. *Aplicação da Modelagem de Função de Transferência em Séries de Consumo de Derivados de Petróleo*, 1984.

Souza, R. C., Camargo, M. E., *Análise e Previsão de Séries Temporais : Os Modelos ARIMA*, SEDIGRAF, 1996.

Varian, H.R.. *Microeconomia. Princípios Básicos*. Campus, 1997.

**REDES NEURAIS E REGRESSÃO DINÂMICA: UM MODELO
HÍBRIDO PARA PREVISÃO DE CURTO PRAZO DA DEMANDA
DE GASOLINA AUTOMOTIVA NO BRASIL**

Dissertação de Mestrado apresentada por *Alexandre Zanini* em 28 de abril de 2000 ao Departamento de Engenharia Elétrica da PUC-Rio e aprovada pela Comissão Julgadora, formada pelos seguintes membros:

Prof. Reinaldo Castro Souza
DEE/PUC-Rio (Orientador)

Prof. Carlos Eduardo Pedreira
DEE/PUC-Rio

Profa. Monica Barros
DEE/PUC-Rio

Prof. Rafael Schechtman
UFRJ/ANP

Visto e permitida a impressão
Rio de Janeiro, 12/05/2000

Prof. Ney Augusto Dumont
Coordenador dos Programas de Pós-Graduação e
Pesquisa do Centro Técnico Científico