

UNIVERSIDADE FEDERAL DO PARANÁ

COMPARAÇÃO ENTRE MODELOS DE SÉRIES TEMPORAIS PARA PREVISÃO
DE LUCRO DE UMA COMPANHIA BRASILEIRA DE ALIMENTOS

CURITIBA

2008

ALEXANDRE MOREIRA BATISTA
HIGOR FERNANDO MANGINELLI

COMPARAÇÃO ENTRE MODELOS DE SÉRIES TEMPORAIS PARA PREVISÃO
DE LUCRO DE UMA COMPANHIA BRASILEIRA DE ALIMENTOS

Trabalho acadêmico apresentado à disciplina de Laboratório II, do curso de graduação em Estatística da Universidade Federal do Paraná como requisito parcial à obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Joel Corrêa da Rosa.

CURITIBA
JUNHO 2008

RESUMO

As tomadas de decisão em estratégias comerciais têm impacto direto no resultado das empresas do ramo alimentício no Brasil. Obter informações confiáveis e com antecedência auxiliam nos processos decisórios. O objetivo deste trabalho é explorar o novo pacote tsDyn do software estatístico R, comparando os modelos de séries temporais deste, como os da classe STAR e os modelos de RNA – Redes Neurais Artificiais, com os modelos clássicos de Alisamento Exponencial e Auto-regressivos. Para comparação da eficiência de previsão, foram utilizados os dados de volumes de venda, Preço e Custos mensais, de uma empresa do ramo alimentício, que será denominada Empresa ABC. A principal atividade da companhia é Frango, desta forma foi realizada a análise estatística necessária para aplicação dos modelos de previsão de séries temporais lineares e não-lineares citados acima.

Palavras-chave: tsDyn, Holt-Winters, AR, STAR, Redes Neurais.

LISTA DE FIGURAS

| | |
|---|----|
| FIGURA 01 – CICLO DE MODELAGEM | 13 |
| FIGURA 02 – FUNÇÃO LOGÍSTICA | 18 |
| FIGURA 03 – EXEMPLO DE NEURÔNIOS ARTIFICIAIS..... | 19 |
| FIGURA 04 – VOLUME DE VENDA MENSAL EM UNID. DE PESO..... | 26 |
| FIGURA 05 – FUNÇÃO DE AUTOCORRELAÇÃO - VOLUME DE VENDA | 27 |
| FIGURA 06 – VARIAÇÃO DO VOLUME DE VENDA..... | 27 |
| FIGURA 07 – PREÇO MENSAL UNITÁRIO EM UNID. DE PREÇO | 28 |
| FIGURA 08 – FUNÇÃO DE AUTOCORRELAÇÃO - PREÇO | 28 |
| FIGURA 09 – VARIAÇÃO DO PREÇO..... | 29 |
| FIGURA 10 – CUSTO DIRETO UNITÁRIO MENSAL EM UNID. DE CUSTO..... | 29 |
| FIGURA 11 – VARIAÇÃO DO CUSTO DIRETO..... | 30 |
| FIGURA 12 – FUNÇÃO DE AUTOCORRELAÇÃO – CUSTO DIRETO..... | 30 |
| FIGURA 13 – CUSTO INDIRETO MENSAL EM UNID. DE CUSTO..... | 31 |
| FIGURA 14 – VARIAÇÃO DO CUSTO INDIRETO | 32 |
| FIGURA 15 – FUNÇÃO DE AUTOCORRELAÇÃO – CUSTO INDIRETO | 32 |
| FIGURA 16 – PREVISÃO DA SÉRIE VOLUME – 3º PASSO..... | 38 |
| FIGURA 17 – PREVISÃO DA SÉRIE VOLUME – 5º PASSO..... | 39 |
| FIGURA 18 – PREVISÃO DA SÉRIE PREÇO – 3º PASSO..... | 40 |
| FIGURA 19 – PREVISÃO DA SÉRIE PREÇO – 5º PASSO..... | 41 |
| FIGURA 20 – PREVISÃO DA SÉRIE CUSTO DIRETO – 3º PASSO | 42 |
| FIGURA 21 – PREVISÃO DA SÉRIE CUSTO DIRETO – 5º PASSO | 43 |
| FIGURA 22 – PREVISÃO DA SÉRIE CUSTO INDIRETO – 3º PASSO..... | 44 |
| FIGURA 23 – PREVISÃO DA SÉRIE CUSTO INDIRETO – 5º PASSO..... | 45 |
| FIGURA 24 – EXEMPLO DE RNA DA SÉRIE CUSTO INDIRETO..... | 50 |

LISTA DE TABELAS

| | |
|---|----|
| TABELA 01 - RESUMO DOS MODELOS SETAR | 34 |
| TABELA 02 - RESUMO DOS MODELOS LSTAR..... | 36 |
| TABELA 03 - RESUMO DOS MODELOS DE REDES NEURAI..... | 37 |
| TABELA 04 - RMSE 3º PASSO - VOLUME | 38 |
| TABELA 05 - RMSE 5º PASSO - VOLUME | 39 |
| TABELA 06 - RMSE 3º PASSO - PREÇO | 40 |
| TABELA 07 - RMSE 5º PASSO - PREÇO | 41 |
| TABELA 08 - RMSE 3º PASSO – CUSTO DIRETO..... | 42 |
| TABELA 09 - RMSE 5º PASSO – CUSTO DIRETO..... | 43 |
| TABELA 10 - RMSE 3º PASSO – CUSTO INDIRETO | 44 |
| TABELA 11 - RMSE 5º PASSO – CUSTO INDIRETO | 45 |
| TABELA 12 - MODELOS SELECIONADOS POR HORIZONTE DE PREVISÃO..... | 46 |

SUMÁRIO

| | |
|---|-----------|
| 1 INTRODUÇÃO | 8 |
| 1.1 OBJETIVOS..... | 8 |
| 2 REVISÃO DE LITERATURA | 9 |
| 2.1 METODOLOGIA ESTATÍSTICA DE SÉRIES TEMPORAIS..... | 9 |
| 2.1.1 <i>Definição de Séries Temporais</i> | 9 |
| 2.1.2 <i>Objetivos da Análise de Séries Temporais</i> | 9 |
| 2.1.3 <i>Estacionariedade da Série Temporal</i> | 10 |
| 2.1.4 <i>Modelos de Previsão</i> | 10 |
| 2.1.5 <i>Alisamento Exponencial</i> | 11 |
| 2.1.6 <i>Método de Holt-Winters</i> | 12 |
| 2.1.7 <i>Metodologia Box e Jenkins</i> | 13 |
| 2.1.8 <i>Operadores de Séries Temporais</i> | 14 |
| 2.1.9 <i>Modelo Auto-regressivo</i> | 14 |
| 2.1.10 <i>Modelo de Médias Móveis</i> | 15 |
| 2.1.11 <i>Modelos Auto-regressivos e de Médias Móveis</i> | 15 |
| 2.1.12 <i>Modelos ARIMA</i> | 16 |
| 2.1.13 <i>Modelos STAR</i> | 16 |
| 2.1.14 <i>Modelos RNA</i> | 18 |
| 2.1.15 <i>Diagnósticos de Modelos</i> | 20 |
| 2.1.16 <i>Seleção de Modelos - Critério de Informação de AKAIKE – AIC</i> | 20 |
| 2.1.17 <i>Avaliação de Modelos - RMSE</i> | 21 |
| 3 MATERIAL E MÉTODOS | 22 |
| 3.1 MATERIAL..... | 22 |
| 3.1.1 <i>Obtenção dos Dados</i> | 22 |
| 3.1.2 <i>Classificações dos Dados</i> | 22 |
| 3.2 MÉTODOS..... | 23 |
| 3.2.1 <i>Tratamentos dos Dados</i> | 23 |
| 3.2.2 <i>Tratamento Estatístico dos Dados</i> | 23 |
| 3.2.3 <i>Pacote tsDyn</i> | 24 |
| 3.2.4 <i>Fluxo do Trabalho</i> | 24 |
| 4 ANÁLISE DOS RESULTADOS | 26 |
| 4.1 ANÁLISE DESCRITIVA DAS SÉRIES..... | 26 |
| 4.1.1 <i>Análise Descritiva da Série Volume</i> | 26 |
| 4.1.2 <i>Análise Descritiva da Série Preço</i> | 28 |
| 4.1.3 <i>Análise Descritiva da Série Custo Direto</i> | 29 |
| 4.1.2 <i>Análise Descritiva da Série Custo Indireto</i> | 31 |
| 4.3 SELEÇÃO DE MODELOS..... | 33 |
| 4.3.1 <i>Seleção dos Modelos Holt Winters</i> | 33 |
| 4.3.2 <i>Seleção dos Modelos AR</i> | 33 |
| 4.3.3 <i>Modelos SETAR</i> | 33 |
| 4.3.4 <i>Modelos LSTAR</i> | 35 |
| 4.3.5 <i>Modelos RNA</i> | 37 |
| 4.4 PREVISÃO E COMPARAÇÃO ENTRE MODELOS..... | 37 |
| 4.4.1 <i>Previsão da Série Volume</i> | 38 |
| 4.4.2 <i>Previsão da Série Preço</i> | 39 |
| 4.4.3 <i>Previsão da Série Custo Direto</i> | 41 |
| 4.4.4 <i>Previsão da Série Custo Indireto</i> | 43 |

| | |
|--|-----------|
| 5 DISCUSSÃO | 46 |
| 5.1 <i>Discussão dos modelos para a Série de Volume.</i> | 46 |
| 5.2 <i>Discussão dos modelos para a Série de Preço.</i> | 47 |
| 5.3 <i>Discussão dos modelos para a Série de Custo Direto.</i> | 49 |
| 5.4 <i>Discussão dos modelos para a Série de Custo Indireto.</i> | 50 |
| 6 CONCLUSÃO | 52 |
| REFERÊNCIAS | 53 |
| ANEXOS | 54 |

1 INTRODUÇÃO

Neste trabalho foram consideradas as informações de vendas de uma grande empresa do ramo alimentício, que será denotada por Empresa ABC, por motivos de confidencialidade.

Os produtos da empresa ABC são separados em categorias, chamadas de atividades, conforme a matéria-prima que os compõem e o tipo de preparo destes alimentos. Desta forma, os dados foram separados em atividades: Frango, Industrializados, Margarinas, Peru e Suíno.

Será utilizada a atividade Frango, para qual foram considerados os volumes de venda mensais, Preço, Custo Direto e Indireto, em unidades de medida (u.m.) de modo também a manter a confidencialidade dos valores reais.

Com os dados previstos para Volume de Vendas e Preço, chega-se à Receita da companhia, conseqüentemente, com os custos, consegue-se prever o Lucro Bruto.

1.1 OBJETIVOS

2.1 Objetivos Gerais

Analisar as novas funcionalidades do software estatístico R, implementadas pelo Pacote tsDyn, com modelos lineares e não lineares de séries temporais, para futuramente desenvolver ferramentas de previsão que auxiliem o planejamento e gestão da Companhia ABC.

2.2 Objetivos Específicos

Agregar conhecimento técnico em modelos de Séries Temporais aos envolvidos no trabalho, bem como divulgar a comunidade científica as aplicações destes modelos em dados econômico-financeiros.

2 REVISÃO DE LITERATURA

2.1 METODOLOGIA ESTATÍSTICA DE SÉRIES TEMPORAIS

2.1.1 Definição de Séries Temporais

Uma série temporal pode ser considerada como um conjunto de observações ordenadas no tempo. Podem-se ter séries contínuas, como o registro de temperaturas em Curitiba ou o registro de marés em Paranaguá, e séries discretas como número anual de vestibulandos da UFPR ou os Índices Diários da Bolsa de São Paulo.

Muitas vezes uma série discreta é obtida através da amostragem de uma série contínua. Por exemplo, o registro de temperaturas em Curitiba observada no tempo $[0, T]$, pode-se tomar N intervalos de tempo, onde:

$$N = \frac{T}{\Delta t} \quad (1)$$

Há basicamente dois enfoques na análise de séries temporais, sendo que em ambos o objetivo é o mesmo, modelar a série temporal. No primeiro a análise é feita no *domínio temporal* e os modelos propostos são *modelos paramétricos* (com um número finito de parâmetros). No segundo, a série é conduzida no *domínio de freqüências*, e os modelos são *não-paramétricos*.

2.1.2 Objetivos da Análise de Séries Temporais

Tendo uma série $Z(t_1), \dots, Z(t_n)$, observada nos instantes t_1, \dots, t_n , pode-se estar interessado em:

- (a) investigar o processo gerador da série;
- (b) fazer previsões de valores futuros da série;
- (c) apenas descrever o comportamento da série;
- (d) procurar periodicidades relevantes nos dados;

Em todos os casos podem-se utilizar modelos probabilísticos no domínio temporal. Como citado em Moretin e Toloï (2004), os modelos devem ser parcimoniosos, isto é, o número de parâmetros no modelo de ser o menor possível e sua utilização não deve apresentar dificuldades.

2.1.3 Estacionariedade da Série Temporal

Uma suposição importante e bastante usada é de que a série apresente estacionariedade, ou seja, ela evolua no tempo aleatoriamente ao redor de uma média constante. Desta forma a série apresenta certa estabilidade ou equilíbrio.

As séries econômicas e financeiras, objetivo deste estudo, são normalmente não-estacionárias. Em geral estas apresentam certa *tendência*, ou seja, seus valores flutuam ao redor de uma reta, com uma determinada inclinação.

A classe dos modelos ARIMA será capaz de descrever de maneira satisfatória séries estacionárias e não-estacionárias, que não tenham comportamento explosivo. Este tipo de não-estacionariedade é chamado de *homogênea*.

Como a maioria dos procedimentos de análises de séries temporais supõe que estas sejam estacionárias, será usada neste estudo a transformação clássica para tornar a série estacionária: Tomar *diferenças* sucessivas da serie original, até se obter uma série estacionária. Na equação (2), está a fórmula para n diferenças da série:

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)] \quad (2)$$

Em situações normais é necessário tomar uma ou duas diferenças para tornar a série estacionária.

2.1.4 Modelos de Previsão

A maioria dos métodos de previsão de séries temporais baseia-se na idéia que o comportamento futuro depende de seu comportamento passado, ou seja, seu padrão de comportamento deve ser estudado antes de formular previsões.

Os métodos de previsão utilizados na prática podem variar muito, desde modelos simples e intuitivos até mais complexos. A complexidade pode estar no processo gerador da série.

Em Economia existem dois procedimentos predominantes: O Econométrico e o de Séries Temporais. O primeiro é baseado na teoria econômica, enquanto o segundo se baseia “no que os dados dizem”, não se limitando a teorias econômicas.

Como já comentado, este trabalho foi focado em Análise Estatística de Séries Temporais, sem recorrer a alguma teoria econômica adjacente, e serão utilizadas as seguintes classes de modelos:

- AR;
- Hont-Winters;
- STAR: TAR, SETAR e LSTAR;
- RNA: Redes Neurais Artificiais;

2.1.5 Alisamento Exponencial

Para uma série temporal X_1, \dots, X_n , não sazonal e sem tendência sistemática, é razoável tomar a estimativa de $Z_n(1)$ ¹ como uma soma ponderada das observações passadas, como na equação (3):

$$Z_n(1) = a_0 X_n + a_1 X_{n-1} + \dots \quad (3)$$

onde a_i são os pesos e Z_n é o valor predito para X_n . Parece razoável também dar um peso maior às observações mais recentes do que às observações mais distantes no passado, desta forma, $a_0 > a_1 > a_2 > \dots$. Podem-se adotar pesos que decaem geometricamente a uma taxa constante, como expresso pela equação (4):

$$a_i = \alpha(1 - \alpha)^i, \quad i=1, 2, 3, \dots \quad (4)$$

onde $0 < \alpha < 1$ é chamada de constante de alisamento.

¹ É a estimativa da série 1 passo a frente.

2.1.6 Método de Holt-Winters

O procedimento de alisamento exponencial pode ser generalizado para séries que contenham tendência e variação sazonal. Como neste trabalho as observações são mensais e sejam L_t , T_t e I_t o nível, a tendência e o índice sazonal no tempo t , respectivamente. Assim, T_t é o aumento ou redução esperada por mês no nível atual da série.

Supondo que no tempo t os termos (L_1, T_1, I_1) , ..., $(L_{t-1}, T_{t-1}, I_{t-1})$ sejam conhecidos. Então, após observar x_t os termos L_t , T_t e I_t são atualizados via alisamento exponencial. Se a variação sazonal for multiplicativa, isto é, com amplitudes que tendem a crescer ao longo do tempo, as equações de atualização na forma de recorrência são dadas pelas equações (5), (6) e (7), considerando o período sazonal igual a 12:

$$L_t = \alpha(x_t / I_{t-12}) + (1 - \alpha) (L_{t-1} + T_{t-1}) \quad (5)$$

$$T_t = \gamma (L_t - L_{t-1}) + (1 - \gamma) T_{t-1} \quad (6)$$

$$I_t = \delta(x_t / L_t) + (1 - \delta)I_{t-12} \quad (7)$$

e as previsões k períodos à frente são dadas pela equação (8):

$$Z_t(k) = (L_t + kT_t) I_{t-12+k} \quad ,k=1, 2, 3, \dots \quad (8)$$

No caso de sazonalidade aditiva, as equações de atualização para o nível e o índice sazonal são modificadas, como é visto nas equações (9) e (10) respectivamente:

$$L_t = \alpha(x_t - I_{t-12}) + (1 - \alpha) (L_{t-1} + T_{t-1}) \quad (9)$$

$$I_t = \delta(x_t - L_t) + (1 - \delta)I_{t-12} \quad (10)$$

e as previsões k períodos à frente são dadas pela equação (11):

$$Z_t(k) = L_t + kT_t + I_{t-12+k} \quad ,k=1, 2, 3, \dots \quad (11)$$

Os parâmetros de alisamento, α , γ e δ , para cada componente da série, são em geral escolhidos no intervalo $(0,1)$ e podem ser estimados minimizando-se a soma de quadrados dos erros de previsão.

Além disso, estes parâmetros não dependem da escala das observações, mas sim das propriedades temporais do nível, tendência e sazonalidade da série. Estes valores refletem a influência das observações passadas nas previsões de cada componente.

Para o caso particular de séries sem variação sazonal, utilizam-se somente as equações para L_t e T_t (sem o índice I_{t-12}). A previsão k passos à frente no tempo t é simplesmente $L_t + kT_t$.

2.1.7 Metodologia Box e Jenkins

Para a classe de modelos ARIMA, utiliza-se o ciclo de modelagem Box e Jenkins, conforme proposto em Moretin e Toloi(2004) e ilustrado na Figura 01.

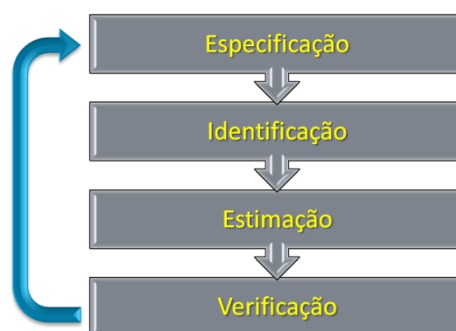


FIGURA 01 – CICLO DE MODELAGEM

Este ciclo engloba os passo:

- (a) uma classe geral de modelos é especificada;
- (b) é realizada a identificação de um modelo com base na análise de gráficos de autocorrelação, gráficos de autocorrelação parcial e outros critérios;
- (c) na fase de estimação, os parâmetros do modelo identificado são estimados;
- (d) na última fase, a verificação dos modelos é realizada através basicamente da análise de resíduos;

Com este procedimento, caso o modelo não seja adequado, repete-se o ciclo, até que o modelo ideal seja escolhido.

A fase crítica do procedimento acima é a identificação, que está sujeita a muita subjetividade. É bem possível então, que dois pesquisadores identifiquem modelos distintos para o mesmo conjunto de dados.

2.1.8 Operadores de Séries Temporais

Está descrito na equação (12) o operador de diferença da série:

$$\Delta Z_t = Z_t - Z_{t-1} = (1 - B)Z_t \quad (12)$$

Também é bastante utilizado o operador de translação ao passado, denotado pela equação (13):

$$BZ_t = Z_{t-1}, B^m Z_t = Z_{t-m} \quad (13)$$

2.1.9 Modelo Auto-regressivo

Um modelo auto-regressivo de ordem p é definido pela equação (14):

$$Z_t = \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \Phi_3 Z_{t-3} + \dots + \Phi_p Z_{t-p} + a_t \quad (14)$$

em que:

$\Phi = (1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p)$ Operador Auto-regressivo

a_t = ruído branco no instante t

O exemplo mais simples seria o AR(1), onde $p=1$, que está descrito na equação (15):

$$Z_t = \Phi_1 Z_{t-1} + a_t \quad (15)$$

De modo que Z_t depende apenas de Z_{t-1} e do ruído no instante t .

2.1.10 Modelo de Médias Móveis

Para um modelo MA(q) tem-se a seguinte notação:

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (16)$$

O exemplo mais simples é o MA(1), dado na equação (17):

$$Z_t = \mu + a_t - \theta_1 a_{t-1} \quad (17)$$

de modo que Z_t depende apenas da média (μ), de Z_{t-1} e do ruído no instante t .

2.1.11 Modelos Auto-regressivos e de Médias Móveis

Os modelos auto-regressivos são bastante utilizados em Economia, quando se acredita que um valor no instante t é a imagem de valores defasados da mesma variável.

Buscando modelos parcimoniosos colocam-se termos de médias móveis e auto-regressivos num mesmo modelo, conforme descrito na equação (18):

$$Z_t = \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \dots + \Phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (18)$$

E novamente, como exemplo simples de um processo ARMA tem-se $p=q=1$, que resulta num modelo ARMA(1,1), conforme a equação (19):

$$Z_t = \Phi_1 Z_{t-1} + a_t - \theta_1 a_{t-1} \quad (19)$$

Novamente, Z_t depende apenas de Z_{t-1} e do ruído no instante t .

2.1.12 Modelos ARIMA

Se $W_t = \Delta^d Z_t$ for estacionária, podemos representar W_t por um modelo ARMA(p, q), ou seja, na equação (20) temos:

$$\phi(B)W_t = \theta(B)a_t \quad (20)$$

Se W_t é uma diferença de Z_t , então Z_t é um integral de W_t , onde Z_t segue um modelo auto-regressivo, *integrado*, de médias móveis, ou seja, ARIMA(p, d, q) pela equação (21):

$$\phi(B)\Delta^d Z_t = \theta(B)a_t \quad (21)$$

Pode-se representar vários modelos com modelos ARIMA, como mostra o Quadro 1.

| | | |
|--------------------|---|----------------|
| ARIMA($p, 0, 0$) | = | AR(p) |
| ARIMA($0, 0, q$) | = | MA(q) |
| ARIMA($p, 0, q$) | = | ARMA(p, q) |

QUADRO 1 - RESUMO DE MODELOS ARIMA

2.1.13 Modelos STAR

Nas últimas décadas, os modelos não-lineares começaram a ser mais desenvolvidos e surgiram os modelos de troca de regime tais como o modelo auto-regressivo com limiar (TAR) e os modelos auto-regressivos com transição suave (STAR).

Estes modelos encontraram um grande número de aplicações de sucesso na literatura, a idéia básica dos mesmos é atribuir modelos lineares distintos em regiões distintas onde se encontram os valores das séries em estudos. Nos casos do STAR as transições entre cada modelo são suaves.

O modelo TAR surgiu com a idéia de aproximar uma função não-linear geral com um modelo linear por partes. Em Chan & Tong (1986) é descrito uma generalização dos modelos TAR com uma transição mais suave, tal generalização ficou conhecida como modelos STAR (*Smooth Transition Autoregressive*). Posteriormente, Teräsvirta (1994) desenvolveu ainda mais este tipo de modelo.

A sua popularidade se deve ao fato da facilidade de especificação, estimação e interpretação quando comparado com outros modelos não-lineares. Todavia pouco se sabe a respeito das propriedades estatísticas de seus estimadores. A idéia principal dos modelos TAR é descrever o seu processo por partes, ou seja, alterar os parâmetros de um modelo auto-regressivo linear dependendo da região em que se encontrem valores de uma determinada variável, isto é, onde os parâmetros de cada modelo dependem da região a que pertence o valor da variável.

A formulação matemática dos modelos da classe STAR é apresentada na equação (22):

$$\begin{cases} Z_t = \Phi_0^{(1)} + \Phi_1^{(1)} Z_{t-1} + \Phi_2^{(1)} Z_{t-2} + \dots + \Phi_{m_1}^{(1)} Z_{t-m_1} & , \quad \text{Se } X_{t-d} < C \\ Z_t = \Phi_0^{(2)} + \Phi_1^{(2)} Z_{t-1} + \Phi_2^{(2)} Z_{t-2} + \dots + \Phi_{m_2}^{(2)} Z_{t-m_2} & , \quad \text{Se } X_{t-d} > C \end{cases} \quad (22)$$

Para transformar esta equação na forma geral do modelo SETAR, basta alterar o termo de limiar X_{t-d} pelo termo de limiar auto-excitante Z_{t-d} .

Já a adaptação da forma STAR para seu caso particular LSTAR, pode ser visto na equação (23):

$$\begin{aligned} Z_t = & (\Phi_0^{(1)} + \Phi_1^{(1)} Z_{t-1} + \Phi_2^{(1)} Z_{t-2} + \dots + \Phi_{m_1}^{(1)} Z_{t-m_1}) [F(Z_{t-d}, \gamma, C)] \\ & + (\Phi_0^{(2)} + \Phi_1^{(2)} Z_{t-1} + \Phi_2^{(2)} Z_{t-2} + \dots + \Phi_{m_2}^{(2)} Z_{t-m_2}) [1 - F(Z_{t-d}, \gamma, C)] \end{aligned} \quad (23)$$

A função F , descrita na equação (24), sendo uma função logística, faz com que a transição seja suave e com inclinação em função do termo γ :

$$F(Z_{t-d}, \gamma, C) = \frac{1}{1 + e^{-\gamma(Z_{t-d} - C)}} \quad (24)$$

Exemplificando, tem-se na Figura 02, a função logística para diferentes valores de γ :

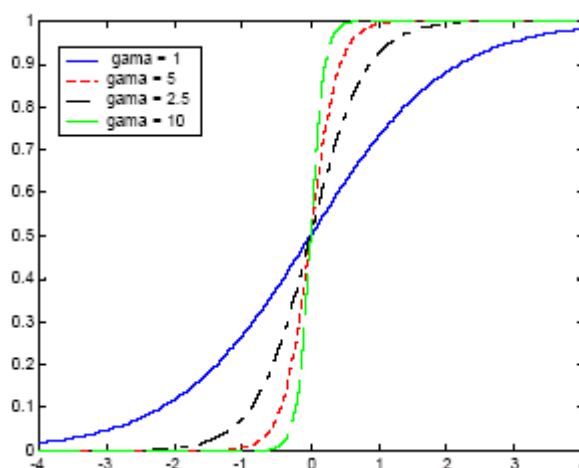


FIGURA 02 – Função Logística

Os modelos da classe STAR podem seguir o mesmo ciclo de modelagem da metodologia Box & Jenkins, ou seja, especificação, estimação dos parâmetros e validação do modelo. É importante dizer que para muitos outros modelos não lineares isso não seria possível.

2.1.14 Modelos RNA

De modo geral, uma rede neural é uma metodologia construída para modelar a forma pela qual o cérebro realiza determinada tarefa ou função (Haykin, 1999). Pode ser implementada por meio de componentes eletrônicos ou de software. Como o cérebro humano, as redes neurais são constituídas de um conjunto de unidades de processamento conectadas entre si, chamadas de neurônios artificiais ou, simplesmente, de neurônios.

A Figura 03 representa um neurônio artificial à direita, em que X_1 , X_2 e X_3 correspondem aos valores de entrada e Y equivale ao valor de saída do neurônio.

Observa-se que, após a multiplicação dos valores de entrada pelos respectivos pesos $W1$, $W2$ e $W3$, os produtos obtidos são somados, resultando no potencial de ativação n . Posteriormente, o valor de n é submetido a uma função de ativação, cujo resultado Y é o valor de saída do neurônio. A função de ativação tem como objetivo limitar, dentro de uma escala de valores, a amplitude do sinal de saída. Portanto, matematicamente, tem-se:

$$x_{t+s} = \beta_0 + \sum_{j=1}^D \beta_j g(\gamma_{0j} + \sum_{i=1}^m \gamma_{ij} x_{t-(i-1)d}) \quad (25)$$

A função de ativação, representada por $g(\cdot)$, é aplicada ao potencial de ativação para a produção do valor de saída x_{t+s} do neurônio artificial. Segundo Haykin (1999), a função mais utilizada na construção de redes neurais artificiais é a sigmoideal.

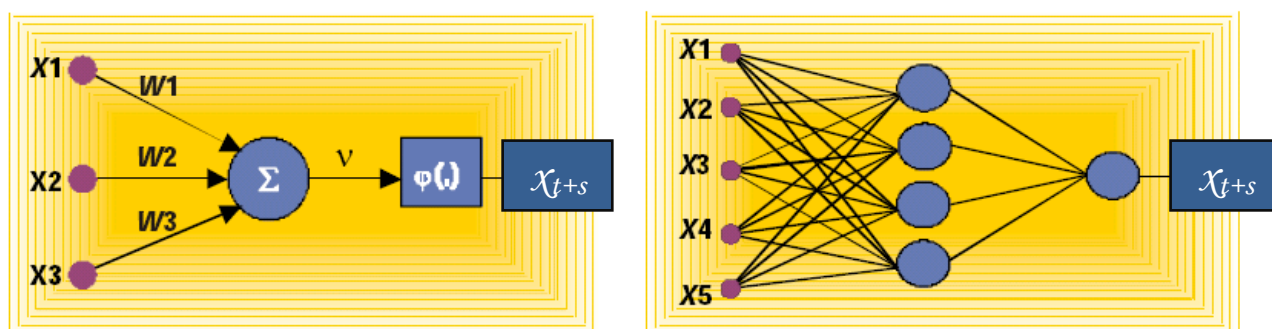


FIGURA 03 – Exemplo de Neurônios Artificiais

Definir a arquitetura de uma rede neural consiste em determinar sua organização estrutural: número de camadas da rede e de neurônios em cada camada, e tipo de conexão entre os neurônios e a topologia da rede (Braga, Ludermir e Carvalho, 2000). Normalmente, as redes neurais são estruturadas em camadas, contendo um ou mais neurônios. Todas elas possuem camada de entrada, que recebe os sinais de entrada, e camada de saída, que armazena o resultado final da rede. As camadas intermediárias localizam-se entre estas duas.

Quanto ao tipo de conexão, as redes neurais podem ser feedforward (acíclica) ou recorrente (cíclica). A recorrente diferencia-se da acíclica por possuir pelo menos

um link de realimentação (feedback), pelo qual a saída de algum neurônio de uma camada i é utilizada como entrada para um neurônio de camada de ordem menor ou igual a i .

As redes neurais podem ser completamente conectadas, quando todos os neurônios estão conectados entre si, ou fracamente conectadas, quando pelo menos dois neurônios não apresentam ligações entre si. A Figura 03 à direita, ilustra uma rede neural acíclica completamente conectada.

2.1.15 Diagnósticos de Modelos

Após estimar o modelo, verifica-se se ele se *ajusta* aos dados, se alguma insuficiência no modelo pode levar a um modelo alternativo similar.

Uma técnica utilizada será o *superajustamento*, onde se colocam mais parâmetros no modelo e verifica-se se os mesmos são significativos e se sua inclusão reduz de forma significativa a variância residual.

A verificação pode ser feita analisando os resíduos. Supondo que o modelo ajustado seja o proposto pela equação (26):

$$\phi(B)W_t = \theta(B)a_t \quad (26)$$

Então, sendo este um modelo ideal, os *erros verdadeiros* $a_t = \theta^{-1}\phi(B)W_t$ constituirão um ruído branco.

2.1.16 Seleção de Modelos - Critério de Informação de AKAIKE – AIC

Na fase de seleção dos modelos é normal ter vários modelos candidatos. Desta forma, além de analisar os resíduos e o Erro Quadrático Médio, também é utilizado o Critério de Informação de Akaike – AIC.

A formulação mais comum é dada pela equação (27):

$$AIC = -2 \log \text{verossimilhança maximizada} + 2m \quad (27)$$

Onde m é o número de parâmetros estimado. Em um modelo ARIMA, $m = p + d + q$.

Desta forma, o AIC penaliza a inclusão excessiva de parâmetros, levando em conta não somente o melhor ajuste, mas também o modelo mais parcimonioso.

Existem outros critérios de informação, mas são basicamente transformações do AIC. Assim, neste trabalho será utilizado como critério de informação na fase de seleção de modelos, o AIC.

No pacote tsDyn, este é o critério utilizado para selecionar os 10 melhores modelos, quando é gerada uma auto-especificação.

Neste trabalho, o AIC será utilizado automaticamente na seleção dos modelos SETAR, LSTAR e RNA pelas funções do pacote tsDyn.

2.1.17 Avaliação de Modelos - RMSE

O RMSE, do inglês **Root Mean Squared Error**, define uma medida de erro para o modelo, conforme a equação (28):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Z_t - X_t)^2}{N}} \quad (28)$$

Onde:

Z_t = Valor estimado pelo modelo no tempo t ,

X_t = O valor da série no tempo t ,

N = Número de observações comparadas.

Esta medida avalia a diferença quadrática média entre o valor da série e o valor ajustado ou previsto. Neste trabalho esta medida será utilizada para mensurar os modelos de previsão, definindo qual tem o melhor desempenho, ou seja, o menor RMSE.

3 MATERIAL E MÉTODOS

3.1 MATERIAL

3.1.1 Obtenção dos Dados

Na empresa ABC, o banco de dados do qual foram retiradas as informações, usa tecnologia *Microsoft SQL Server 2005*².

Este banco de dados têm informações desde o volume de vendas, passando pelo preço dos produtos, até as despesas operacionais da empresa.

Com o apoio da equipe de informática da empresa, foi realizada uma consulta ao banco de dados para retornar os dados por produto entre os meses de janeiro de 2003 e fevereiro de 2007.

3.1.2 Classificações dos Dados

Como cada produto tem várias características cadastradas, foi sugerido pela empresa utilizar um agrupamento por *Atividade*.

Atividade é a característica definida pela empresa que classifica seus produtos conforme a matéria-prima e o processo de fabricação. Assim, as atividades estão descritas no quadro 2:

² Este é um software que gerencia grandes quantidades de dados, amplamente usado em grandes empresas. Marca registrada da Microsoft Inc.

| Atividade | Descrição |
|------------------|---|
| Frango | Todos os produtos fabricados com frango, mas que não sofrem processos industriais de processamento. Exemplo: Frango Inteiro Resfriado, Coxas de Frango, File de Peito de Frango, etc. |
| Industrializados | Os produtos que passam por processo de industrialização e tem como origem matéria-prima animal. Exemplo: Almôndegas, Hambúrguer, Salames, Salsichas, Lasanhas, etc. |
| Margarinas | Agrupamento das margarinas produzidas. |
| Peru | Todos os produtos fabricados com peru, mas que não sofrem processos industriais de processamento. Exemplo: Peru Inteiro Resfriado, Coxas de Peru, File de Peito de Peru, etc. |
| Suíno | Todos os produtos fabricados com Suínos, mas que não sofrem processos industriais de processamento. Exemplo: Paleta Suína Resfriada, Pernil Suíno, Bisteca Suína Congelada, etc. |

QUADRO 2 - DESCRIÇÃO DAS ATIVIDADES DA EMPRESA

3.2 MÉTODOS

3.2.1 Tratamentos dos Dados

Com os dados já classificados, os mesmos foram trabalhados no *Microsoft Excel 2007*³, onde foram geradas as tabelas e gráficos apresentados neste trabalho.

3.2.2 Tratamento Estatístico dos Dados

Para o procedimento de modelagem estatística será utilizado o software R (versão 2.7) com uso dos pacotes: Tseries, TsDyn, específicos para séries temporais tanto na parte descritiva da série como para realizar as previsões e comparações entre modelos.

³ Sistema de planilhas eletrônicas, na sua mais recente versão. Marca registrada da Microsoft Inc.

O pacote xlsReadWrite, será utilizado para comunicação entre a base de dados e as saídas para a montagem dos gráficos.

3.2.3 Pacote tsDyn

Foi utilizado neste trabalho o pacote tsDyn, na sua versão 0.6, cujos autores são Antonio Fabio Di Narzo e Jose Luis Aznarte.

Este pacote agrupa a geração e seleção de modelos não-lineares para séries temporais, como os da classe STAR e RNA.

Segundo a vinheta do pacote, os autores citam que o mesmo está em fase inicial, e deve sofrer mudanças significativas em breve, embora esteja totalmente usual nesta versão.

A vinheta tem um estudo de caso da série clássica da quantidade anual de lincas canadenses capturados (*lynx*). Este estudo percorre desde as funções descritivas, passando pelas funções chaves do pacote, que são as de auto-seleção, e compara as previsões com cada modelo aos valores reais da série.

Como já citado, as funções de auto-seleção “selectSETAR”, “selectLSTAR” e “selectnNet”, são as principais do pacote. Estas recebem parâmetros que permitem gerar vários modelos, e retornam os 10 melhores modelos em termos de AIC, ou seja, com o melhor ajuste à série. Embora este trabalho busque os melhores modelos de previsão, estas funções serão utilizadas dado sua simplicidade e também a quantidade infinita de modelos que poderiam ser gerados na classe STAR, por exemplo.

Ressalta-se a aplicação do AIC nos modelos de RNA. O pacote assume uma normalidade nos erros, de forma que consegue estimar a função de máxima verossimilhança e chega assim ao AIC.

3.2.4 Fluxo do Trabalho

Primeiramente, serão analisadas as séries descritivamente, avaliando a estacionariedade e os correlogramas.

Após a análise descritiva, serão retiradas as 5 últimas observações e então gerados os seguintes modelos para cada série:

- 1 modelo Holt-Winters com sazonalidade aditiva;
- 1 modelo Holt-Winters sem sazonalidade;
- 1 modelo Holt-Winters sem sazonalidade e sem tendência (alisamento exponencial simples);
- 5 modelos AR(p), com $p=1, \dots, 5$;
- 10 modelos SETAR (selecionados pelo pacote tsDyn);
- 10 modelos LSTAR (selecionados pelo pacote tsDyn);
- 10 modelos RNA (selecionados pelo pacote tsDyn);

Para cada modelo gerado, será realizada a previsão para 5 passos à frente. Assim pode-se comparar com as 5 observações originais das séries e medir qual o modelo com maior poder preditivo em termos de RMSE.

4 ANÁLISE DOS RESULTADOS

4.1 ANÁLISE DESCRITIVA DAS SÉRIES

4.1.1 Análise Descritiva da Série Volume

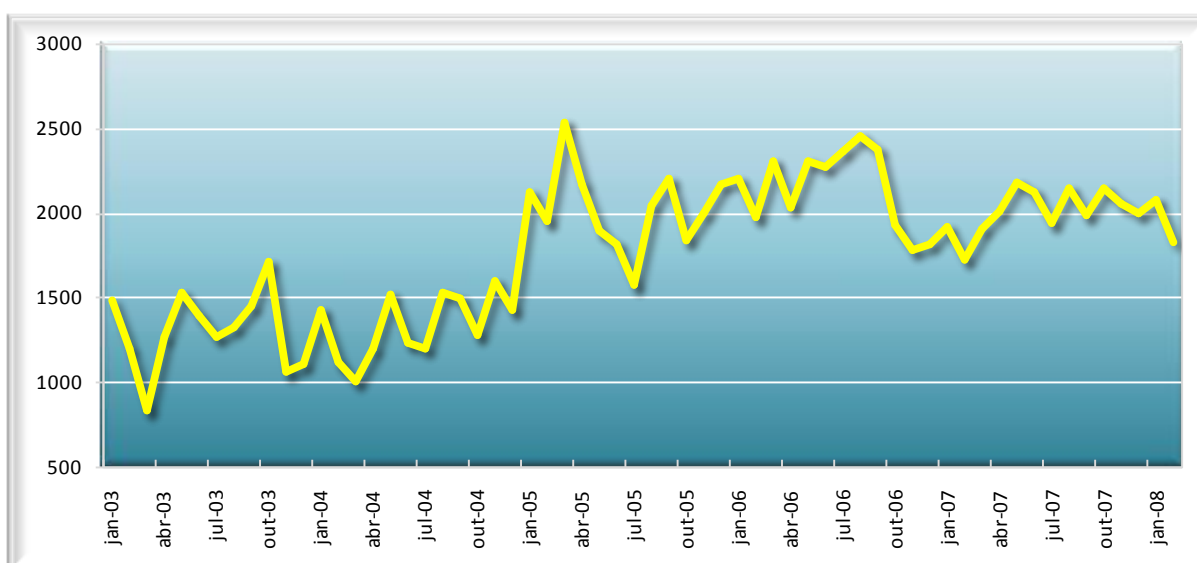


FIGURA 04 – VOLUME DE VENDA MENSAL EM UNID. DE PESO

Na Figura 04, temos a série original, com seus valores expressos em uma unidade de peso. Nota-se claramente uma tendência de crescimento, o que caracteriza um comportamento não-estacionário na média. Desta forma, observa-se na Figura 05, a Função de Autocorrelação da série original e da série de variação, obtida a partir da primeira diferença da série.

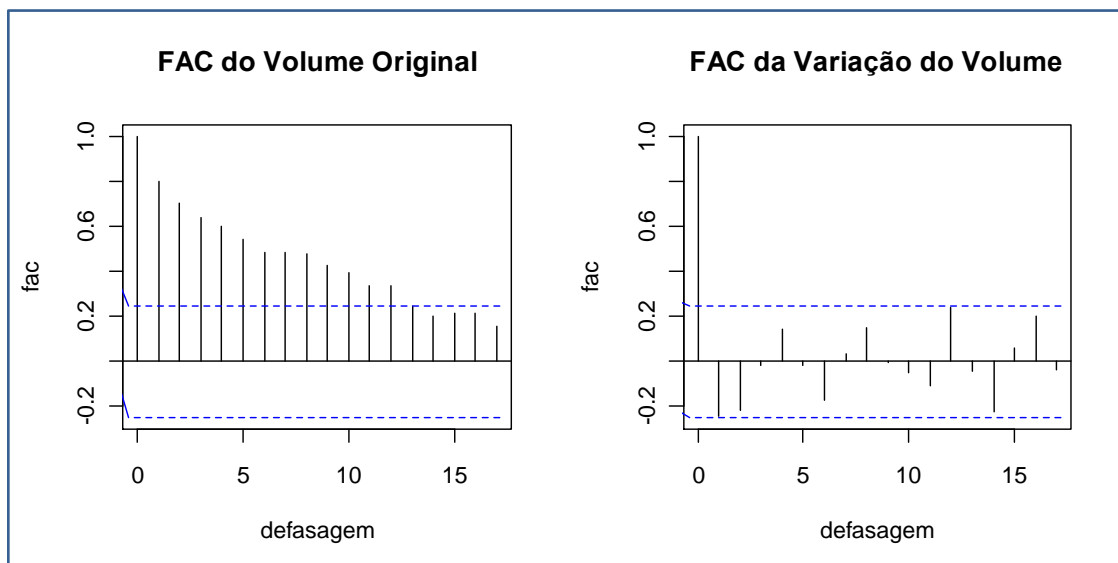


FIGURA 05 – FUNÇÃO DE AUTOCORRELAÇÃO - VOLUME DE VENDA

O comportamento da FAC observado na série original, com decaimento lento e valores relevantes até a defasagem 13, é típico de uma série não-estacionária, conforme visto em exemplos de Moretin e Tolo (2004). Já a série de variação tem um comportamento normal. Desta forma pode-se ver na figura 06 o gráfico da série de variação.

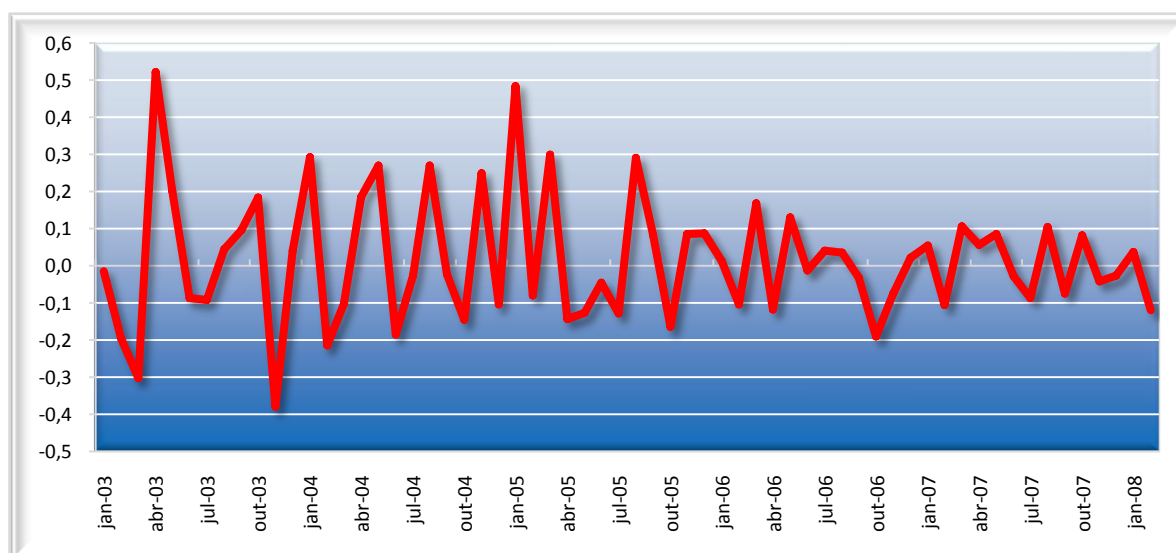


FIGURA 06 – VARIAÇÃO DO VOLUME DE VENDA

Com estacionariedade na média, visto na figura 06 e como os modelos autorregressivos aplicam-se às séries estacionárias, então será utilizada a série de variação de volume neste trabalho, a qual será chamada somente de Série de Volume.

4.1.2 Análise Descritiva da Série Preço

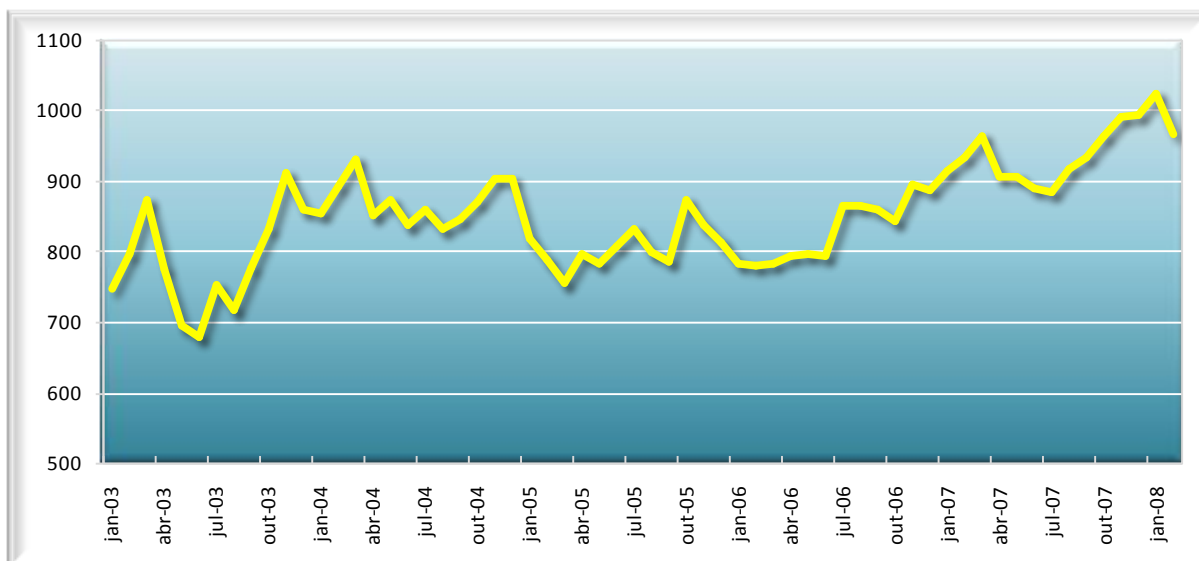


FIGURA 07 – PREÇO MENSAL UNITÁRIO EM UNID. DE PREÇO

Na Figura 07, temos a série original, com seus valores expressos em uma unidade de preço unitário. Tem-se também, uma tendência de crescimento, o que caracteriza um comportamento não-estacionário na média. Desta forma, observa-se na Figura 08, a Função de Autocorrelação da série original e da série de variação, obtida a partir da primeira diferença da série, a exemplo do visto na série de Volume.

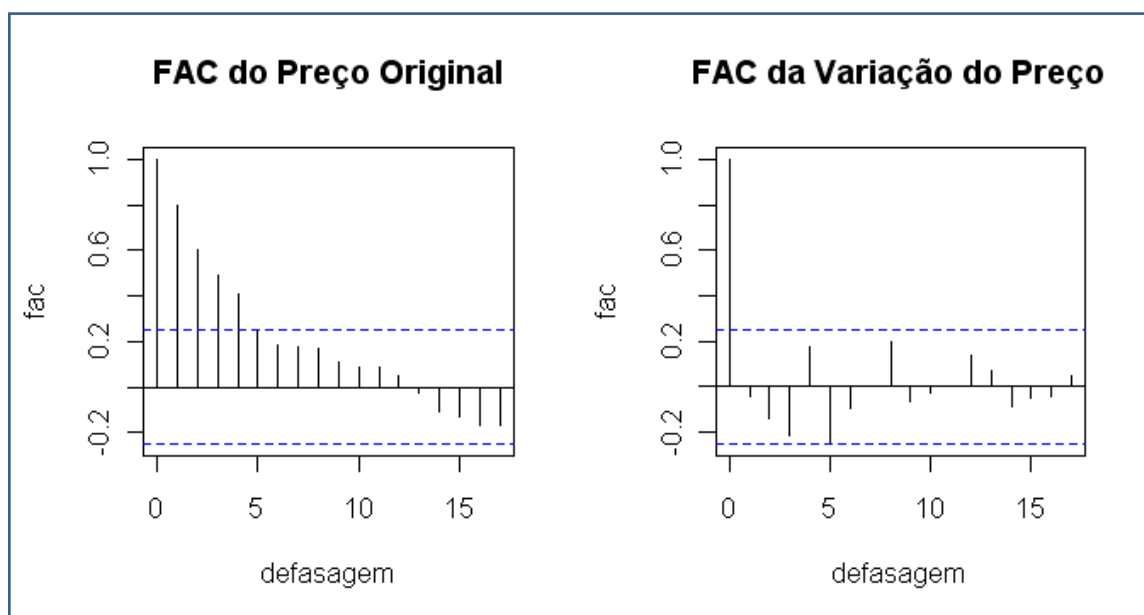


FIGURA 08 – FUNÇÃO DE AUTOCORRELAÇÃO - PREÇO

A exemplo da Série de Volume, na figura 08 nota-se o decaimento lento da autocorrelação. Buscando a utilização de uma série estacionária, na figura 09 tem-se o gráfico da série de variação do Preço.

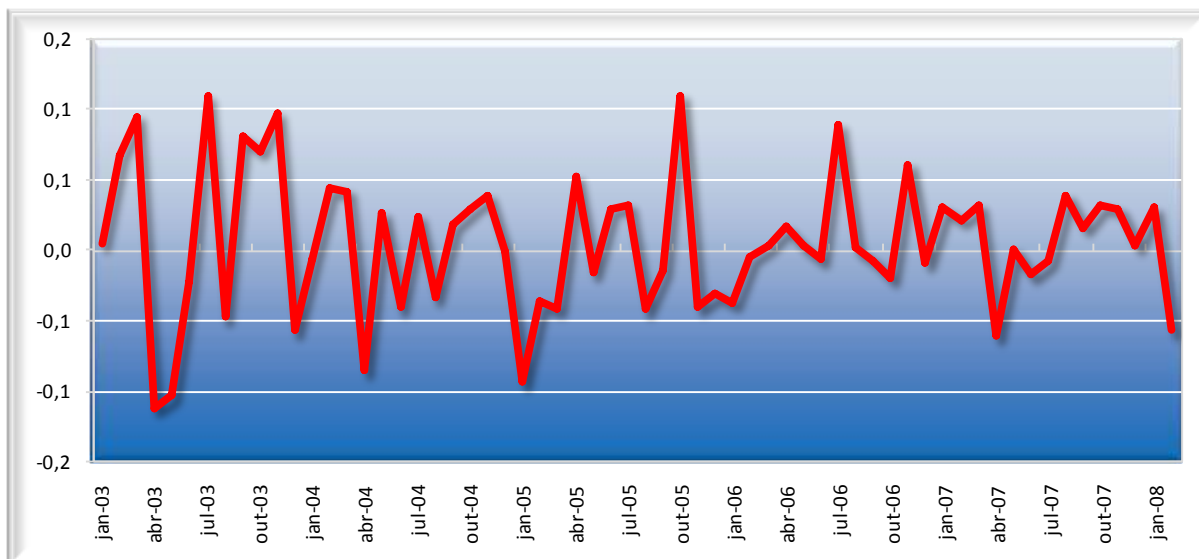


FIGURA 09 – VARIAÇÃO DO PREÇO

Na Figura 09 observa-se uma série estacionária na média. A exemplo da Série de volume, será utilizada a série de variação de Preço neste trabalho e será chamada de Série de Preço.

4.1.3 Análise Descritiva da Série Custo Direto

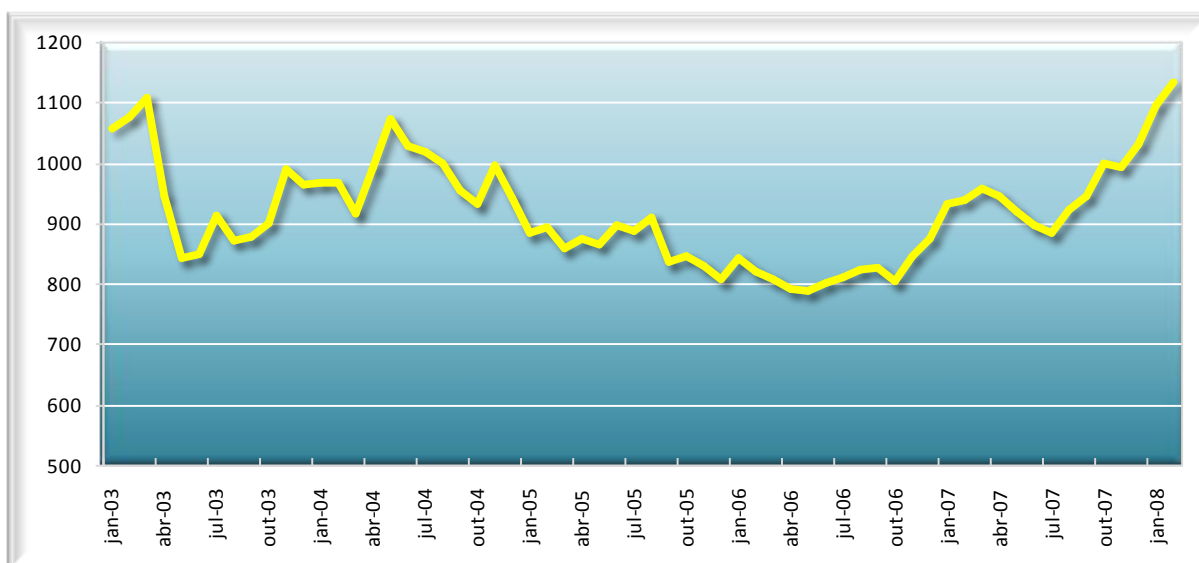


FIGURA 10 – CUSTO DIRETO UNITÁRIO MENSAL EM UNID. DE CUSTO

Na Figura 10, temos a série original, com seus valores expressos em uma unidade de Custo Direto unitário. E na Figura 11, a série de variação do Custo Direto.

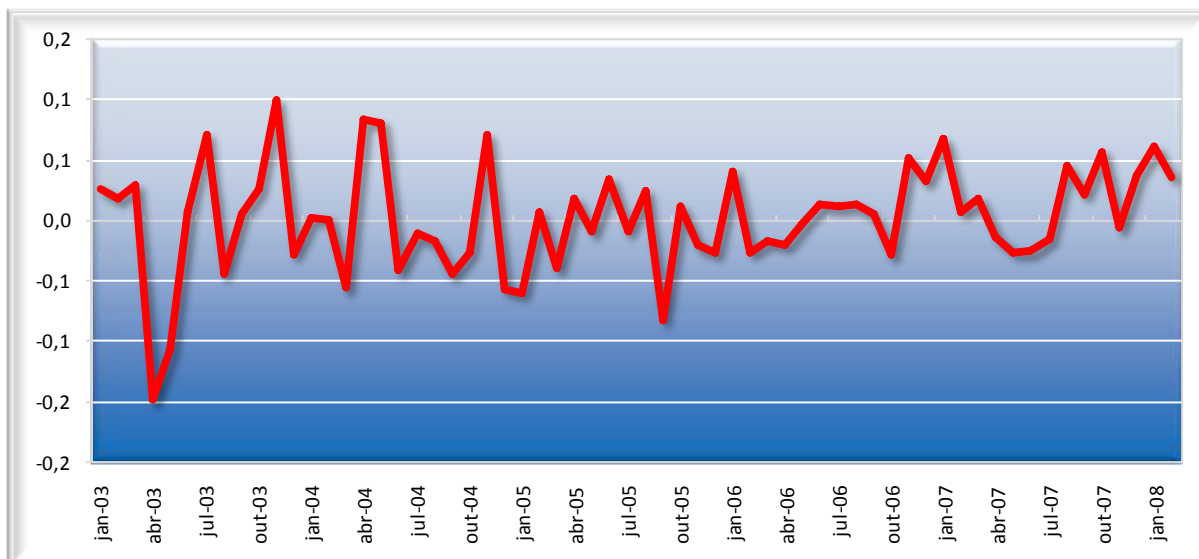


FIGURA 11 – VARIÇÃO DO CUSTO DIRETO

Na Figura 11 tem-se estacionariedade na média. Desta forma, de modo a verificar se será esta série de variação será a série utilizada neste trabalho, a figura 12 mostra os gráficos da FAC da série original e de variação.

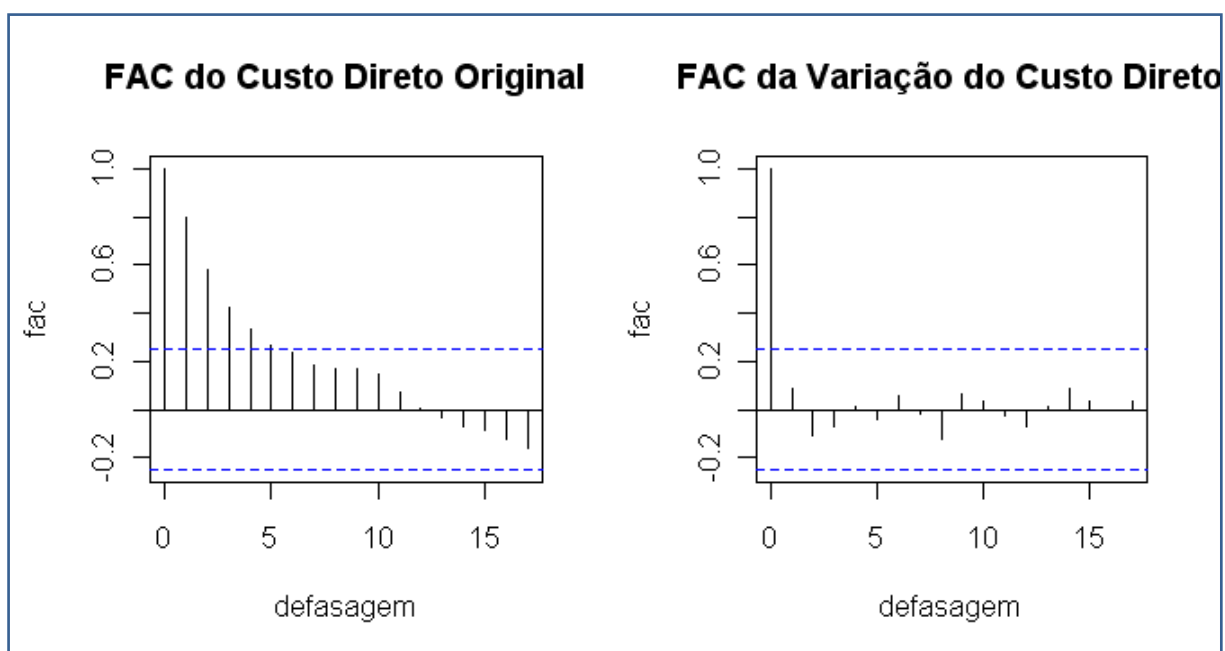


FIGURA 12 – FUNÇÃO DE AUTOCORRELAÇÃO – CUSTO DIRETO

Na Figura 12 tem-se outra vez o comportamento típico da FAC de uma série não-estacionária na série original. Novamente, no trabalho para os dados de Custo Direto, será utilizada também a série de variação que será chamada de Série de Custo Direto.

4.1.2 Análise Descritiva da Série Custo Indireto

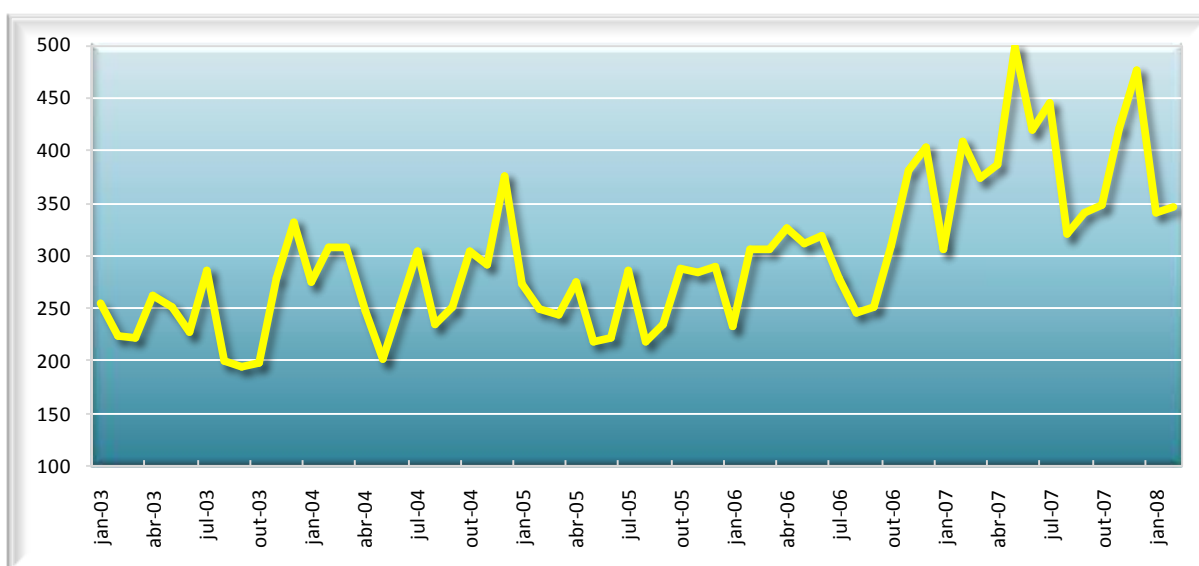


FIGURA 13 – CUSTO INDIRETO MENSAL EM UNID. DE CUSTO

Na Figura 13, temos a série original, com seus valores expressos em uma unidade de custo indireto unitário. E na Figura 14, a série de variação do Custo Indireto.

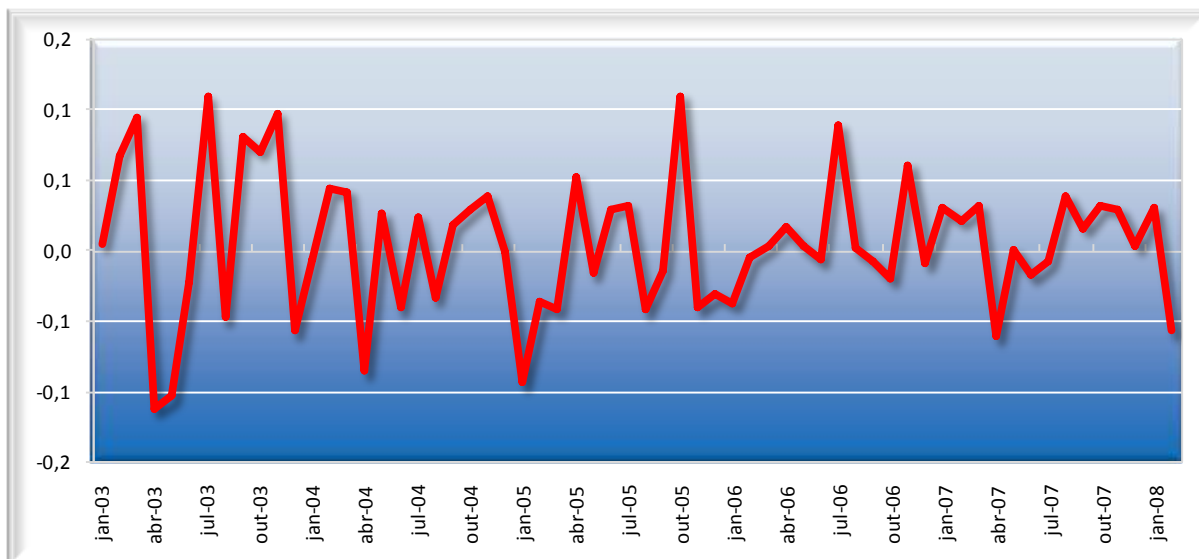


FIGURA 14 – VARIAÇÃO DO CUSTO INDIRETO

Outra vez de modo a decidir qual das séries será utilizada neste trabalho, para melhor verificar a estacionariedade, segue a Figura 15.

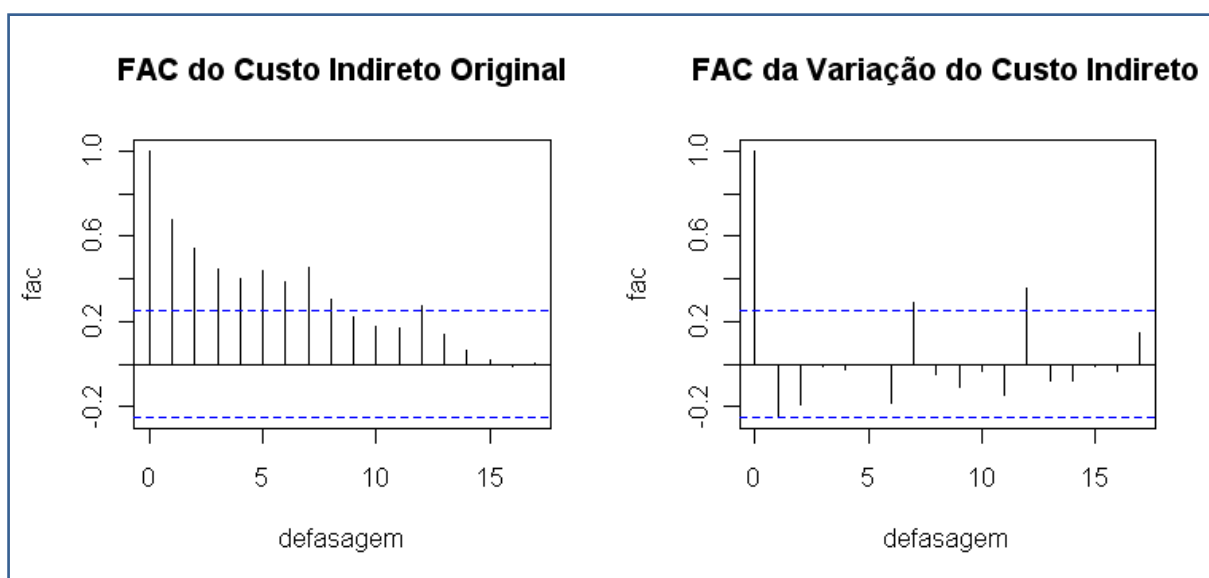


FIGURA 15 – FUNÇÃO DE AUTOCORRELAÇÃO – CUSTO INDIRETO

Pela Figura 15, conclui-se que a série original é uma série não-estacionária. Também será utilizada a série de variação e será chamada de Série de Custo Indireto neste trabalho.

4.3 SELEÇÃO DE MODELOS

4.3.1 Seleção dos Modelos Holt Winters

Para cada série de dados, foram ajustados 3 modelos pela função “*HoltWinters*” do pacote *Tseries* no software *R*, conforme descritos abaixo:

- Holt_winters01: Modelo *HoltWinters* completo, com tendência “ T_t ” e sazonalidade aditiva “ I_{t-12} ”;
- Holt_winters02: Modelo *HoltWinters*, com tendência “ T_t ” e sem sazonalidade “ I_{t-12} ”, de forma que $\delta = 0$, como visto na equação (7);
- Holt_winters03: Modelo *HoltWinters*, sem tendência “ T_t ” e sem sazonalidade “ I_{t-12} ”, de forma que $\gamma = 0$, $\delta = 0$ e com $\alpha \neq 0$, tornando-se um alisamento exponencial simples, como visto na equação (7);

4.3.2 Seleção dos Modelos AR

Para cada série de dados, foram ajustados 5 modelos pela função “*Linear*” do pacote *Tseries* no software *R*. Assim tem-se modelos $AR(p)$ com p variando de 1 à 5 para cada série.

4.3.3 Modelos SETAR

Utilizando para esses modelos o pacote *tsDyn* com a função “*selectSetar*”, foram selecionados 10 modelos. O AIC é utilizado como critério de seleção por esta função de modo a ordenar os modelos e retornar os que têm melhor ajuste.

Pelo critério de parcimônia, ao fazer uso da função acima, os argumentos desta foram limitados, fazendo com que os parâmetros do termo auto-regressivos do Modelo SETAR, m_1 e m_2 variassem entre 0 e 5 para ambos os regimes e o parâmetro d variando entre 0 e 4 (conforme descrito na função 13) no decorrer da auto-seleção da função.

Na tabela 01 observam-se os 10 modelos selecionados pela função descrita acima para cada série.

TABELA 01 - RESUMO DOS MODELOS SETAR

| Série | d | C | m_1 | m_2 | AIC |
|----------------|-----|--------|-------|-------|---------|
| Volume | 4 | -0,061 | 3 | 2 | -224,72 |
| Volume | 4 | -0,061 | 1 | 2 | -224,58 |
| Volume | 4 | -0,061 | 3 | 3 | -223,64 |
| Volume | 4 | -0,061 | 1 | 3 | -223,44 |
| Volume | 4 | -0,061 | 4 | 2 | -223,10 |
| Volume | 4 | -0,061 | 2 | 2 | -222,85 |
| Volume | 3 | -0,061 | 1 | 2 | -222,69 |
| Volume | 3 | -0,061 | 3 | 2 | -222,68 |
| Volume | 3 | -0,061 | 3 | 3 | -222,44 |
| Volume | 3 | -0,061 | 1 | 3 | -222,32 |
| Preço | 0 | -0,024 | 5 | 5 | -361,50 |
| Preço | 1 | -0,024 | 5 | 5 | -361,50 |
| Preço | 2 | -0,024 | 5 | 5 | -361,50 |
| Preço | 3 | -0,024 | 5 | 5 | -361,50 |
| Preço | 4 | -0,024 | 5 | 5 | -361,50 |
| Preço | 0 | -0,024 | 3 | 5 | -361,25 |
| Preço | 1 | -0,024 | 3 | 5 | -361,25 |
| Preço | 2 | -0,024 | 3 | 5 | -361,25 |
| Preço | 3 | -0,024 | 3 | 5 | -361,25 |
| Preço | 4 | -0,024 | 3 | 5 | -361,25 |
| Custo Direto | 4 | 0,039 | 1 | 3 | -368,57 |
| Custo Direto | 4 | 0,023 | 1 | 4 | -368,45 |
| Custo Direto | 4 | 0,012 | 1 | 4 | -368,10 |
| Custo Direto | 4 | 0,012 | 4 | 4 | -367,76 |
| Custo Direto | 4 | -0,025 | 2 | 1 | -367,71 |
| Custo Direto | 4 | 0,039 | 1 | 2 | -367,60 |
| Custo Direto | 4 | -0,025 | 1 | 1 | -367,53 |
| Custo Direto | 4 | 0,023 | 2 | 4 | -367,44 |
| Custo Direto | 4 | 0,023 | 4 | 4 | -367,43 |
| Custo Direto | 0 | 0,023 | 1 | 5 | -367,28 |
| Custo Indireto | 0 | -0,029 | 1 | 5 | -202,58 |
| Custo Indireto | 1 | -0,029 | 1 | 5 | -202,58 |
| Custo Indireto | 2 | -0,029 | 1 | 5 | -202,58 |
| Custo Indireto | 3 | -0,029 | 1 | 5 | -202,58 |
| Custo Indireto | 4 | -0,029 | 1 | 5 | -202,58 |
| Custo Indireto | 4 | 0,230 | 2 | 4 | -201,77 |
| Custo Indireto | 0 | 0,230 | 2 | 4 | -201,75 |
| Custo Indireto | 1 | 0,230 | 2 | 4 | -201,75 |
| Custo Indireto | 2 | 0,230 | 2 | 4 | -201,75 |
| Custo Indireto | 3 | 0,230 | 2 | 4 | -201,75 |

Nota-se na tabela 01 que nas quatro séries, os modelos selecionados automaticamente têm AIC's bem próximos, o que implica que podem existir outros modelos que também tenham um bom ajuste. Estes modelos serão comparados quanto à previsão de 3 e 5 passos em cada uma das séries.

4.3.4 Modelos LSTAR

Utilizando novamente o pacote tsDyn, desta vez com a função “*selectLstar*”, foram selecionados 10 modelos para cada série seguindo o mesmo critério de seleção da função “*selectSetar*”, o menor AIC.

Também pelo critério de parcimônia, ao fazer uso da função acima, os argumentos desta foram limitados de forma com que os termos dos parâmetros auto-regressivos do Modelo LSTAR, m_1 e m_2 variassem entre 0 e 5 para ambos os regimes e o parâmetro d entre 0 e 4, (conforme descrito na função 14).

Na tabela 02 observam-se os 10 modelos selecionados pela função descrita acima para cada série.

TABELA 02 - RESUMO DOS MODELOS LSTAR

| Série | d | m_1 | m_2 | AIC |
|----------------|-----|-------|-------|---------|
| Volume | 2 | 5 | 5 | -238,02 |
| Volume | 2 | 1 | 5 | -235,92 |
| Volume | 2 | 3 | 5 | -231,96 |
| Volume | 2 | 4 | 5 | -230,13 |
| Volume | 2 | 5 | 2 | -229,20 |
| Volume | 2 | 2 | 5 | -227,36 |
| Volume | 2 | 5 | 3 | -227,14 |
| Volume | 2 | 5 | 4 | -225,36 |
| Volume | 2 | 5 | 1 | -222,82 |
| Volume | 2 | 4 | 2 | -221,05 |
| Preço | 4 | 2 | 4 | -361,70 |
| Preço | 4 | 1 | 4 | -359,95 |
| Preço | 4 | 2 | 2 | -359,31 |
| Preço | 0 | 5 | 4 | -359,11 |
| Preço | 4 | 2 | 5 | -358,92 |
| Preço | 4 | 5 | 2 | -358,91 |
| Preço | 3 | 5 | 1 | -358,82 |
| Preço | 4 | 5 | 3 | -357,91 |
| Preço | 4 | 3 | 4 | -357,75 |
| Preço | 4 | 1 | 5 | -357,70 |
| Custo Direto | 0 | 1 | 1 | -347,83 |
| Custo Direto | 0 | 2 | 1 | -347,71 |
| Custo Direto | 0 | 1 | 2 | -347,69 |
| Custo Direto | 0 | 3 | 1 | -347,12 |
| Custo Direto | 0 | 1 | 3 | -347,11 |
| Custo Direto | 2 | 1 | 1 | -346,06 |
| Custo Direto | 0 | 2 | 2 | -345,72 |
| Custo Direto | 0 | 2 | 3 | -345,19 |
| Custo Direto | 0 | 3 | 2 | -345,14 |
| Custo Direto | 2 | 1 | 2 | -344,93 |
| Custo Indireto | 3 | 1 | 3 | -206,80 |
| Custo Indireto | 3 | 2 | 5 | -206,10 |
| Custo Indireto | 3 | 2 | 4 | -205,85 |
| Custo Indireto | 4 | 4 | 5 | -205,43 |
| Custo Indireto | 3 | 3 | 3 | -205,40 |
| Custo Indireto | 4 | 5 | 5 | -205,30 |
| Custo Indireto | 4 | 2 | 5 | -205,15 |
| Custo Indireto | 4 | 5 | 3 | -205,12 |
| Custo Indireto | 4 | 3 | 5 | -204,91 |
| Custo Indireto | 3 | 1 | 2 | -204,67 |

Dentre os modelos acima, os que apresentam melhor desempenho tem AIC's próximos, dentro de cada série. A exemplo dos modelos SETAR, pode-se haver bons modelos em termos de ajuste, que estão fora desta lista.

4.3.5 Modelos RNA

Utilizando novamente o pacote tsDyn, desta vez com a função “*selectnNet*”, foram selecionados 10 modelos para cada série seguindo o mesmo critério de seleção da função “*selectSetar*”, o menor AIC.

Também pelo critério de parcimônia, ao fazer uso da função acima, os argumentos desta foram limitados de forma com que a quantidade de neurônios da camada oculta *D*, tivesse variação entre 1 e 10, e a quantidade de neurônios da camada de entrada *m*, estivesse entre 1 e 5, no decorrer da auto-seleção da função.

Na tabela 03 observam-se os 10 modelos selecionados pela função descrita acima para cada série.

TABELA 03 - RESUMO DOS MODELOS DE REDES NEURAIIS

| Volume | | | Preço | | | Custo Direto | | | Custo Indireto | | |
|----------|----------|------|----------|----------|------|--------------|----------|------|----------------|----------|------|
| <i>m</i> | <i>D</i> | AIC | <i>m</i> | <i>D</i> | AIC | <i>m</i> | <i>D</i> | AIC | <i>m</i> | <i>D</i> | AIC |
| 5 | 4 | -228 | 5 | 1 | -353 | 5 | 1 | -360 | 5 | 3 | -212 |
| 5 | 5 | -227 | 4 | 1 | -343 | 4 | 1 | -355 | 5 | 2 | -204 |
| 5 | 7 | -223 | 5 | 2 | -339 | 5 | 2 | -350 | 5 | 4 | -204 |
| 4 | 4 | -221 | 3 | 1 | -338 | 1 | 1 | -346 | 4 | 3 | -196 |
| 5 | 1 | -216 | 2 | 1 | -332 | 2 | 1 | -345 | 2 | 1 | -196 |
| 4 | 1 | -211 | 4 | 2 | -331 | 3 | 1 | -345 | 4 | 1 | -196 |
| 5 | 2 | -210 | 1 | 1 | -331 | 4 | 2 | -343 | 3 | 1 | -196 |
| 5 | 6 | -209 | 3 | 2 | -329 | 4 | 3 | -342 | 5 | 1 | -194 |
| 3 | 1 | -209 | 5 | 3 | -327 | 1 | 2 | -340 | 4 | 2 | -194 |

Na tabela 03, tem-se um resumo da seleção de modelos para as 4 séries através do critério AIC.

4.4 PREVISÃO E COMPARAÇÃO ENTRE MODELOS

De modo a comparar os modelos selecionados quanto à eficiência de previsão, foi utilizada a medida de erro, RMSE, descrita na seção 2.1.17.

Para a companhia ABC, a previsão de 3 e 5 passos à frente, tem impacto direto nas decisões de planejamento. Desta forma foram calculados os RMSE até o 3° e até o 5° passo a frente buscando os melhores modelos para cada um destes horizontes de previsão.

4.4.1 Previsão da Série Volume

Na Figura 16, visualiza-se a previsão para os quatro melhores modelos na Série Volume, quando comparados pelo RMSE até o 3º Passo.

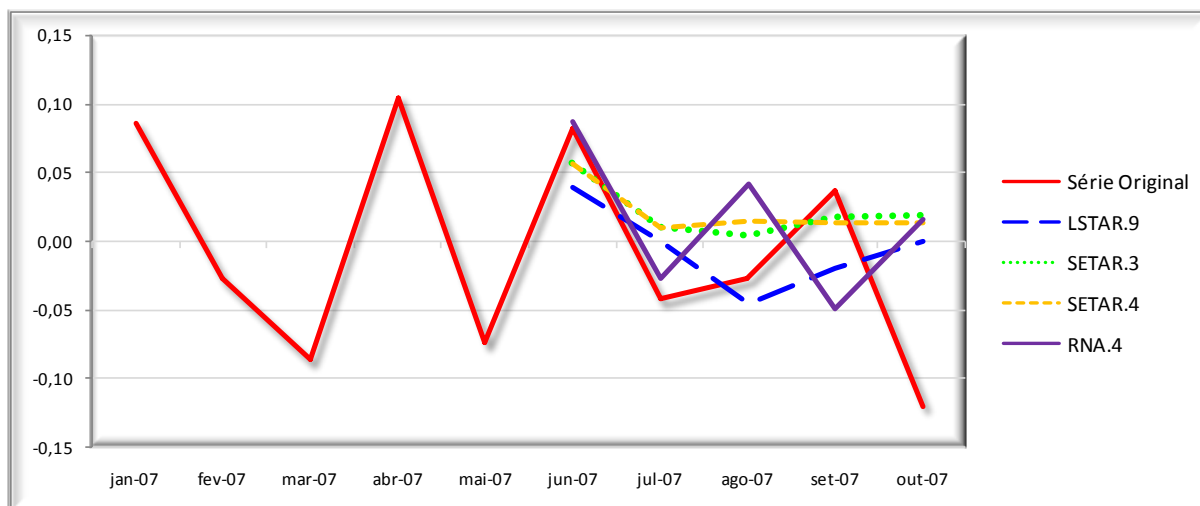


FIGURA 16 – PREVISÃO DA SÉRIE VOLUME – 3º PASSO

Na Figura 16, destacam-se os modelos LSTAR 9 e SETAR 3 e 4, embora o modelo RNA 4, visualmente melhor acompanha a evolução da série original até o segundo passo.

TABELA 04 - RMSE 3º PASSO - VOLUME

| Modelo | RMSE até 3º passo |
|---------|-------------------|
| LSTAR.9 | 0,0361 |
| SETAR.3 | 0,0380 |
| SETAR.4 | 0,0413 |
| RNA.4 | 0,0415 |

Através da tabela 04, nota-se que o menor RMSE até o 3º passo, indica como melhor modelo de previsão o LSTAR 9, seguido pelo Modelo SETAR 3 e 4.

Na Figura 17, visualiza-se a previsão para os quatro melhores modelos na Série Volume, quando comparados pelo RMSE até o 5º Passo.

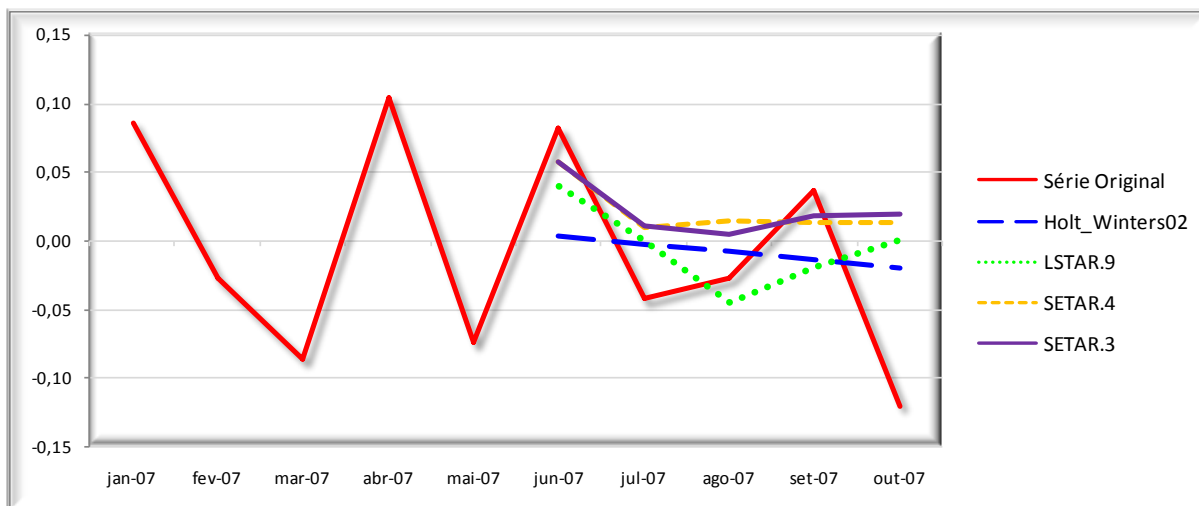


FIGURA 17 – PREVISÃO DA SÉRIE VOLUME – 5º PASSO

Na Figura 17, destacam-se os modelos LSTAR 9 e SETAR 3 que visualmente melhor acompanham a evolução da série original

TABELA 05 - RMSE 5º PASSO - VOLUME

| Modelo | RMSE até 5º passo |
|----------------|--------------------------|
| Holt_winters02 | 0,0643 |
| LSTAR.9 | 0,0655 |
| SETAR.4 | 0,0686 |
| SETAR.3 | 0,0693 |

Através da tabela 05, nota-se que o menor RMSE até o 5º passo, indica como melhor modelo de previsão o Holt_winters02, seguido pelo Modelo LSTAR 9.

4.4.2 Previsão da Série Preço

Na Figura 18, visualiza-se a previsão para os quatro melhores modelos na Série Volume, quando comparados pelo RMSE até o 3º Passo.

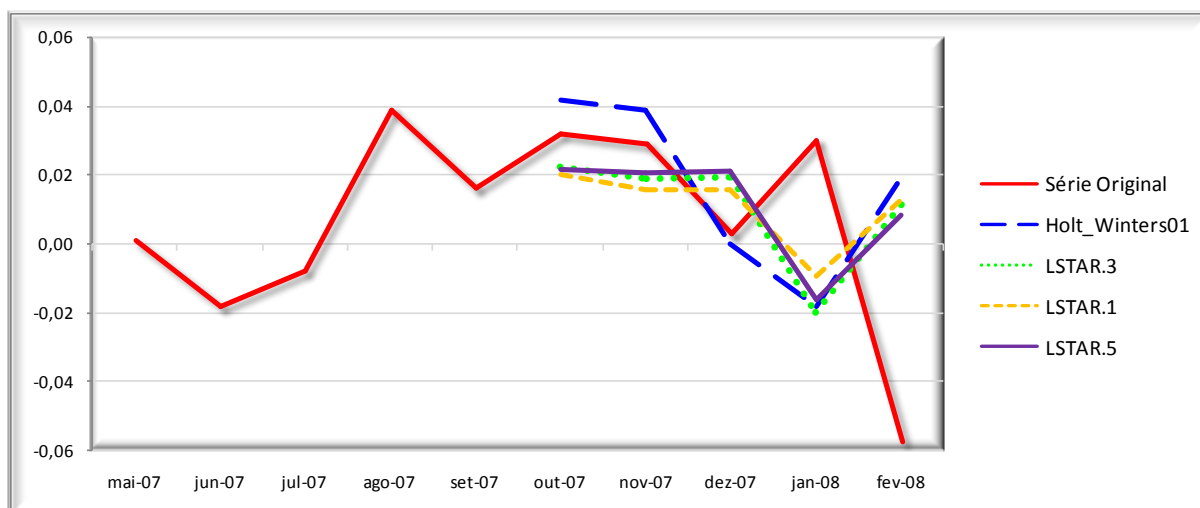


FIGURA 18 – PREVISÃO DA SÉRIE PREÇO – 3º PASSO

Na Figura 18, destaca-se o modelo Holt_winters01, que visualmente melhor acompanha a evolução da série original até o terceiro passo.

TABELA 06 - RMSE 3º PASSO - PREÇO

| Modelo | RMSE até 3º passo |
|----------------|-------------------|
| Holt_winters01 | 0,0080 |
| LSTAR.3 | 0,0126 |
| LSTAR.1 | 0,0129 |
| LSTAR.5 | 0,0130 |

Através da tabela 06, nota-se que o menor RMSE até o 3º passo, indica como melhor modelo de previsão o Holt_winters01, seguido pelo Modelo LSTAR 3.

Na Figura 19, visualiza-se a previsão para os quatro melhores modelos na Série Volume, quando comparados pelo RMSE até o 5º Passo.

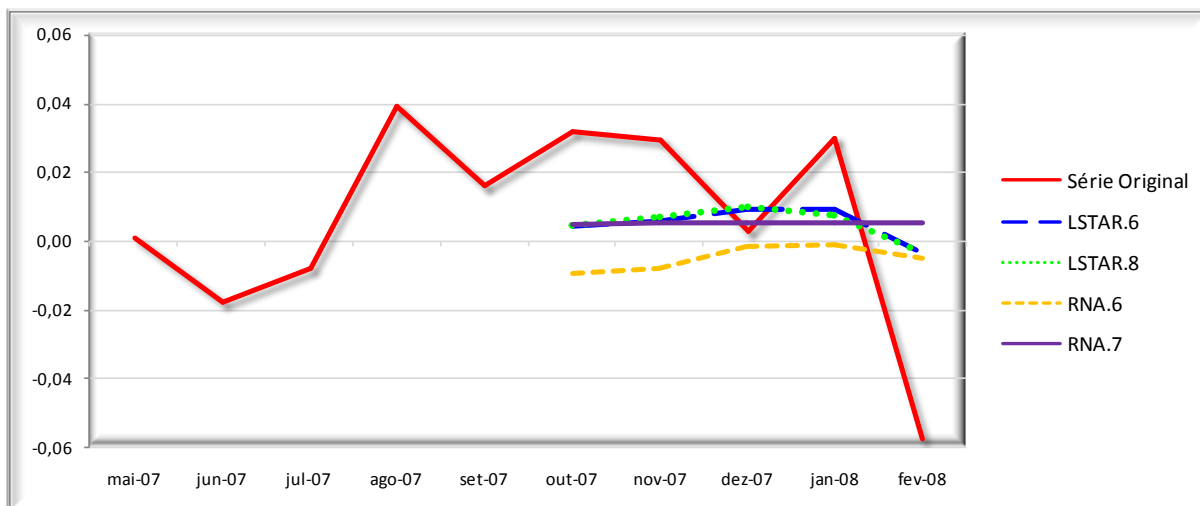


FIGURA 19 – PREVISÃO DA SÉRIE PREÇO – 5º PASSO

Na Figura 19, destaca-se o modelo LSTAR 6, que visualmente melhor acompanha a evolução da série original.

TABELA 07 - RMSE 5º PASSO - PREÇO

| Modelo | RMSE até 5º passo |
|----------------|-------------------|
| LSTAR.6 | 0,0308 |
| LSTAR.8 | 0,0309 |
| Holt_winters03 | 0,0343 |
| RNA.2 | 0,0344 |

Através da tabela 07, nota-se que o menor RMSE até o 5º passo, indica como melhor modelo de previsão o LSTAR 6, seguido pelo Modelo LSTAR 8.

4.4.3 Previsão da Série Custo Direto

Na Figura 20, visualiza-se a previsão para os quatro melhores modelos na Série Volume, quando comparados pelo RMSE até o 3º Passo.

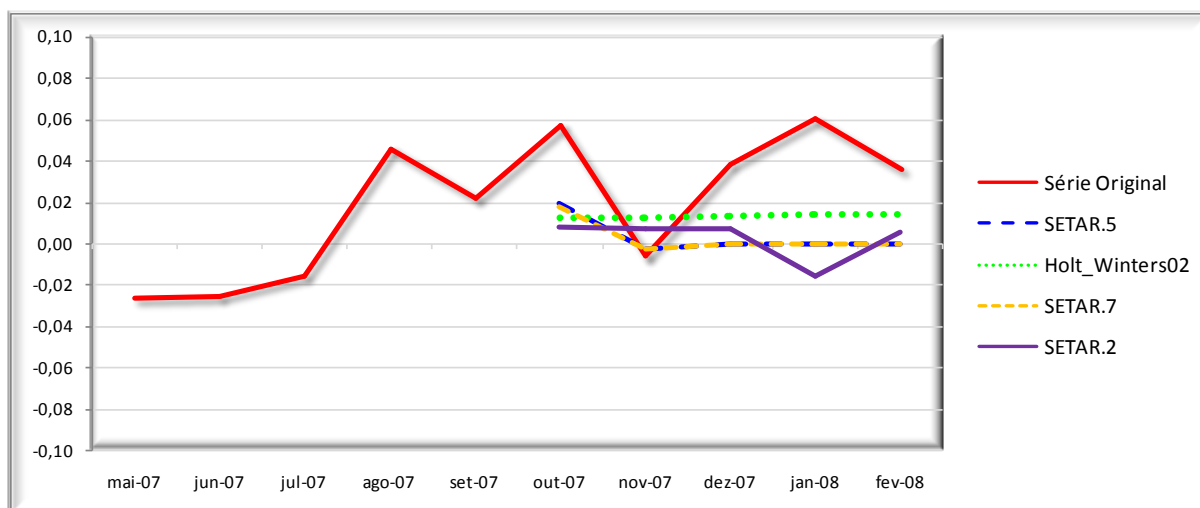


FIGURA 20 – PREVISÃO DA SÉRIE CUSTO DIRETO – 3º PASSO

Na Figura 20, destaca-se o modelo SETAR 5 e 7, que visualmente melhor acompanham a evolução da série original até o segundo passo.

TABELA 08 - RMSE 3º PASSO – CUSTO DIRETO

| Modelo | RMSE até 3º passo |
|----------------|--------------------------|
| SETAR.5 | 0,0313 |
| Holt_winters02 | 0,0318 |
| SETAR.7 | 0,0321 |
| SETAR.2 | 0,0346 |

Através da tabela 08, nota-se que o menor RMSE até o 3º passo, indica como melhor modelo de previsão o SETAR 5, seguido pelo Holt_winters02.

Na Figura 21, visualiza-se a previsão para os quatro melhores modelos na Série Volume, quando comparados pelo RMSE até o 5º Passo.

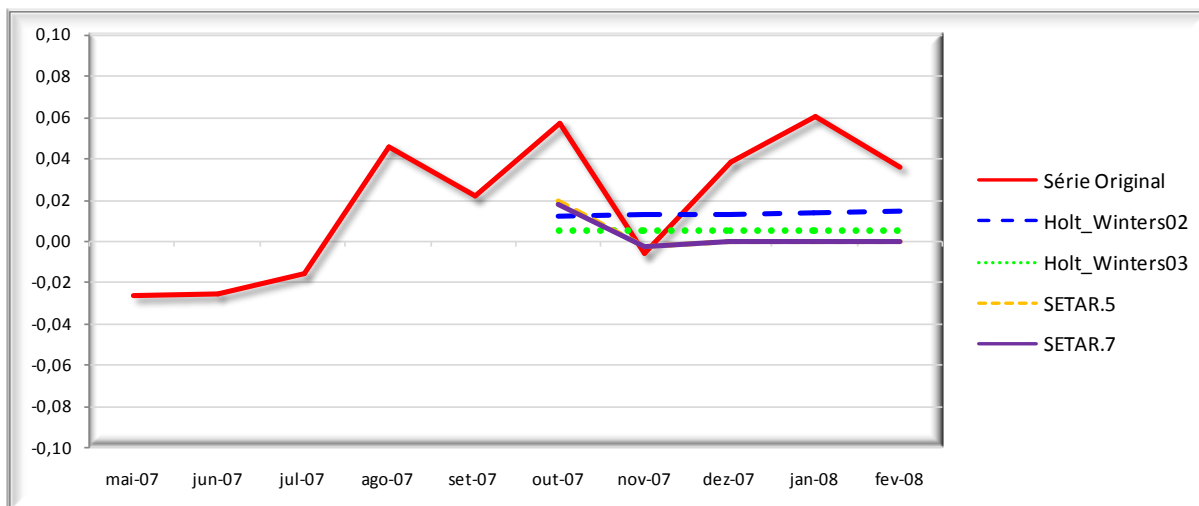


FIGURA 21 – PREVISÃO DA SÉRIE CUSTO DIRETO – 5º PASSO

Na Figura 21, destaca-se o modelo SETAR 7, que visualmente melhor acompanha a evolução da série original até o segundo passo.

TABELA 09 - RMSE 5º PASSO – CUSTO DIRETO

| Modelo | RMSE até 5º passo |
|----------------|-------------------|
| Holt_winters02 | 0,0339 |
| Holt_winters03 | 0,0400 |
| SETAR.5 | 0,0403 |
| SETAR.7 | 0,0407 |

Através da tabela 09, nota-se que o menor RMSE até o 5º passo, indica como melhor modelo de previsão o Holt_winters02, seguido pelo Holt_winters03.

4.4.4 Previsão da Série Custo Indireto

Na Figura 22, visualiza-se a previsão para os quatro melhores modelos na Série Volume, quando comparados pelo RMSE até o 3º Passo.

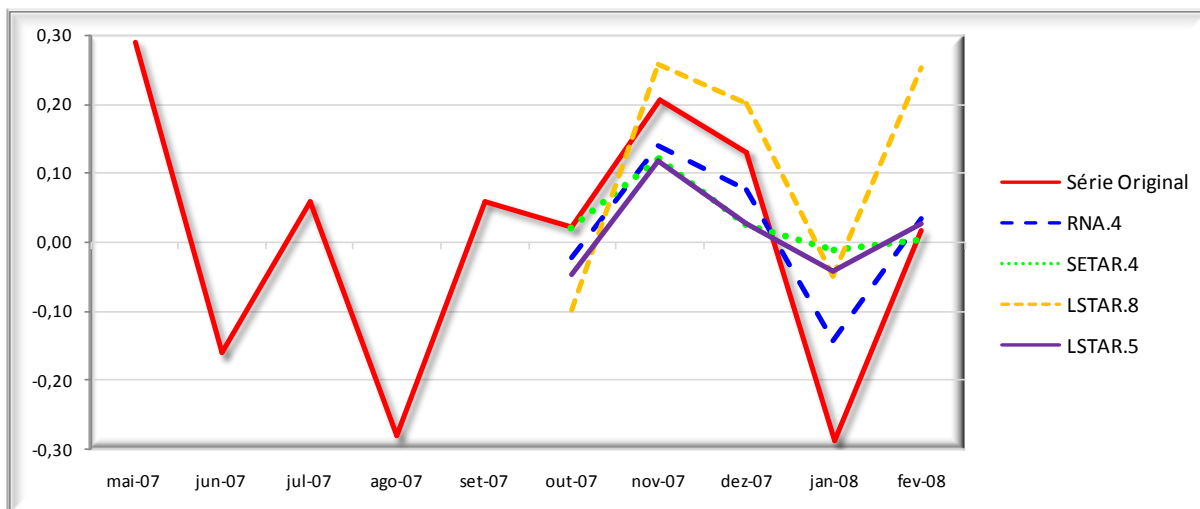


FIGURA 22 – PREVISÃO DA SÉRIE CUSTO INDIRETO – 3º PASSO

Na Figura 22, destaca-se o modelo RNA 4 e SETAR 4, que visualmente melhor acompanham a evolução da série original.

TABELA 10 - RMSE 3º PASSO – CUSTO INDIRETO

| Modelo | RMSE até 3º passo |
|---------|-------------------|
| RNA.4 | 0,0573 |
| SETAR.4 | 0,0792 |
| LSTAR.8 | 0,0860 |
| LSTAR.5 | 0,0888 |

Através da tabela 10, nota-se que o menor RMSE até o 3º passo, indica como melhor modelo de previsão o RNA 4, seguido pelo Modelo SETAR 4.

Na Figura 23, visualiza-se a previsão para os quatro melhores modelos na Série Volume, quando comparados pelo RMSE até o 5º Passo.

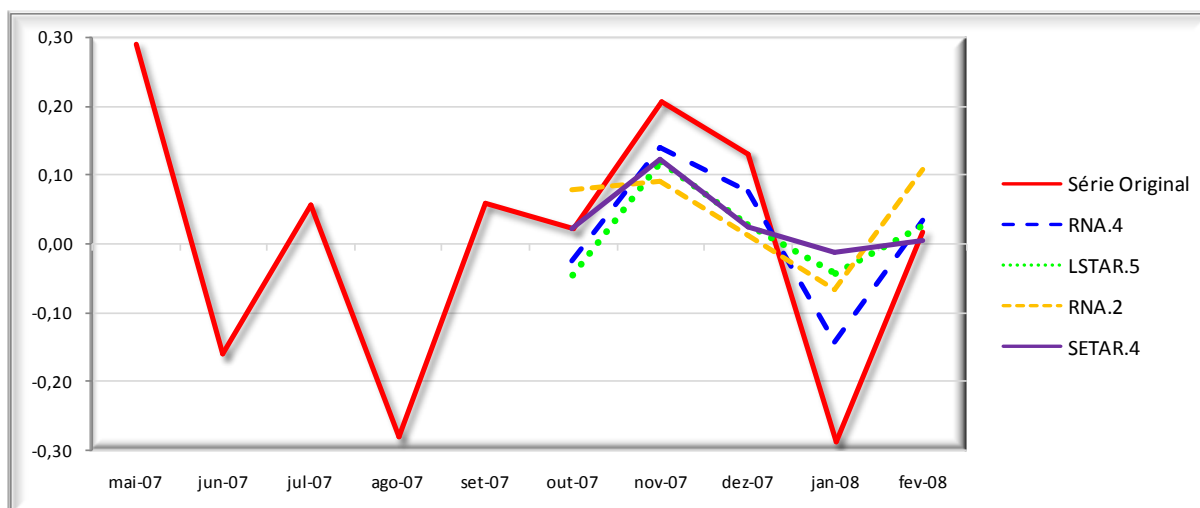


FIGURA 23 – PREVISÃO DA SÉRIE CUSTO INDIRETO – 5º PASSO

Na Figura 23, destaca-se o modelo RNA 4, que visualmente melhor acompanha a evolução da série original.

TABELA 11 - RMSE 5º PASSO – CUSTO INDIRETO

| Modelo | RMSE até 5º passo |
|---------|-------------------|
| RNA.4 | 0,0787 |
| LSTAR.5 | 0,1288 |
| RNA.2 | 0,1326 |
| SETAR.4 | 0,1371 |

Através da tabela 11, nota-se que o menor RMSE até o 5º passo, indica como melhor modelo de previsão também o RNA 4, seguido pelo Modelo LSTAR 5.

5 DISCUSSÃO

A tabela 12 mostra o melhor modelo segundo o RMSE, em cada série estudada.

TABELA 12 - MODELOS SELECIONADOS POR HORIZONTE DE PREVISÃO

| Dados | 3º passo | 5º passo |
|------------------|----------------|----------------|
| Volume de Vendas | LSTAR.9 | Holt_winters02 |
| Preço | Holt_winters01 | LSTAR.6 |
| Custo Direto | SETAR.5 | Holt_winters02 |
| Custo Indireto | RNA.4 | RNA.4 |

Podemos ver entre os modelos da tabela 12, que não houve uma classe de modelo que se destacou de maneira geral. Desta forma não se pode assumir uma única classe de modelo para todas as séries.

Nota-se ainda através da tabela 12, que as séries têm diferentes melhores modelos quando se observa o RMSE até o 3º e 5º passo, com exceção da Série de Custo Indireto.

5.1 Discussão dos modelos para a Série de Volume.

Para a Série de Volume, o modelo com melhor desempenho foi o LSTAR 9 até o 3º passo. Conforme descrito nas equações 13, 14 e 15 têm-se os seguintes parâmetros, segundo cada Regime:

Modelo LSTAR 9 – para previsões até 3º passo:

Regime 01, quando: $Z_{t-2} < -0,0865$

$$\Phi_0 = 0,1958$$

$$\Phi_1 = -0,8209$$

$$\Phi_2 = -0,3405$$

$$\Phi_3 = 0,1626$$

$$\Phi_4 = 0,0187$$

$$\Phi_5 = -0,0953$$

Regime 02, quando: $Z_{t-2} > -0,0865$

$$\Phi_0 = -0,2252$$

$$\Phi_1 = 0,4882$$

E com o parâmetro da função logística de transição:

$$\gamma = 1833$$

Com um valor alto deste parâmetro, como visto na Figura 02, este modelo se aproxima de um modelo SETAR, dado que a transição não é suave entre os dois regimes.

Para previsões até o 5º passo na Série Volume, o melhor modelo em termos de RMSE é Holt_winters02, com tendência e sem sazonalidade, conforme exposto na seção 2.1.6, e com os seguintes parâmetros:

Holt_winters02 – para previsões até 5º passo:

$$L_t = 0,0089$$

$$T_t = -0,0057$$

$$\alpha = 0,3083$$

$$\gamma = 0,3512$$

5.2 Discussão dos modelos para a Série de Preço.

Para a Série de Preço, os modelos com melhor desempenho foram o Holt_winters01 e LSTAR 6, até o 3º e até o 5º passo, respectivamente. Conforme descrito nas equações 13, 14 e 15 têm-se os seguintes parâmetros, segundo cada Modelo:

Modelo Holt_winters01 – para previsões até 3º passo:

$$L_t = 0,0135$$

$$T_t = 0,0$$

$$\alpha = 0,0111$$

$$\gamma = 0,0$$

$$\delta = 0,2106$$

Com os termos de sazonalidade:

$$s_1 = 0,0285$$

$$s_7 = -0,0578$$

$$s_2 = 0,0252$$

$$s_8 = 0,0010$$

$$s_3 = -0,0137$$

$$s_9 = -0,0320$$

$$s_4 = -0,0314$$

$$s_{10} = 0,0193$$

$$s_5 = 0,0057$$

$$s_{11} = -0,0213$$

$$s_6 = 0,0086$$

$$s_{12} = 0,0042$$

Este modelo foi ajustado contendo o termo de tendência T_t , porém sua estimativa foi zero, e conseqüentemente $\gamma = 0,0$.

Modelo LSTAR – para previsões até 5º passo:

Regime 01, quando: $Z_{t-4} < 0,03153$

$$\Phi_0 = 0,0008$$

$$\Phi_1 = 0,1214$$

$$\Phi_2 = 0,1023$$

$$\Phi_3 = -0,0721$$

$$\Phi_4 = 0,1423$$

$$\Phi_5 = -0,3900$$

Regime 02, quando: $Z_{t-4} > 0,03153$

$$\Phi_0 = 0,0278$$

$$\Phi_1 = -0,4227$$

$$\Phi_2 = -0,6633$$

E com o parâmetro da função logística de transição:

$$\gamma = 37995$$

Com um valor alto deste parâmetro, como visto na Figura 02, este modelo se aproxima de um modelo SETAR, dado que a transição não é suave entre os dois regimes.

5.3 Discussão dos modelos para a Série de Custo Direto.

Para a Série de Custo Direto, os modelos não conseguiram reconhecer o padrão da série, fazendo com que suas previsões convergissem para a média da série. Nenhum dos modelos sobressaiu-se aos demais.

Modelo SETAR – para previsões até 3º passo:

Regime 01, quando: $Z_{t-4} < -0,0254$

$$\Phi_0 = 0,0137$$

$$\Phi_1 = 0,2035$$

$$\Phi_2 = 0,0283$$

Regime 02, quando: $Z_{t-4} > -0,0254$

$$\Phi_0 = -0,0006$$

$$\Phi_1 = -0,1234$$

Com este modelo, a série posiciona-se dividida com 26,9% das observações no regime 01.

Holt_winters02 – para previsões até 5º passo:

$$L_t = 0.0118$$

$$T_t = 0.0005$$

$$\alpha = 0.2314$$

$$\gamma = 0.0868$$

5.4 Discussão dos modelos para a Série de Custo Indireto.

Para a Série de Custo Indireto, o modelo com melhor desempenho foi o RNA 4. A equação do melhor modelo em termos de RMSE pode ser exemplificada pela Figura 24, onde a RNA conta com 4 neurônios de entrada (Z_{t-i}), com 3 neurônios na camada oculta e com 1 neurônio de saída (Z_t):

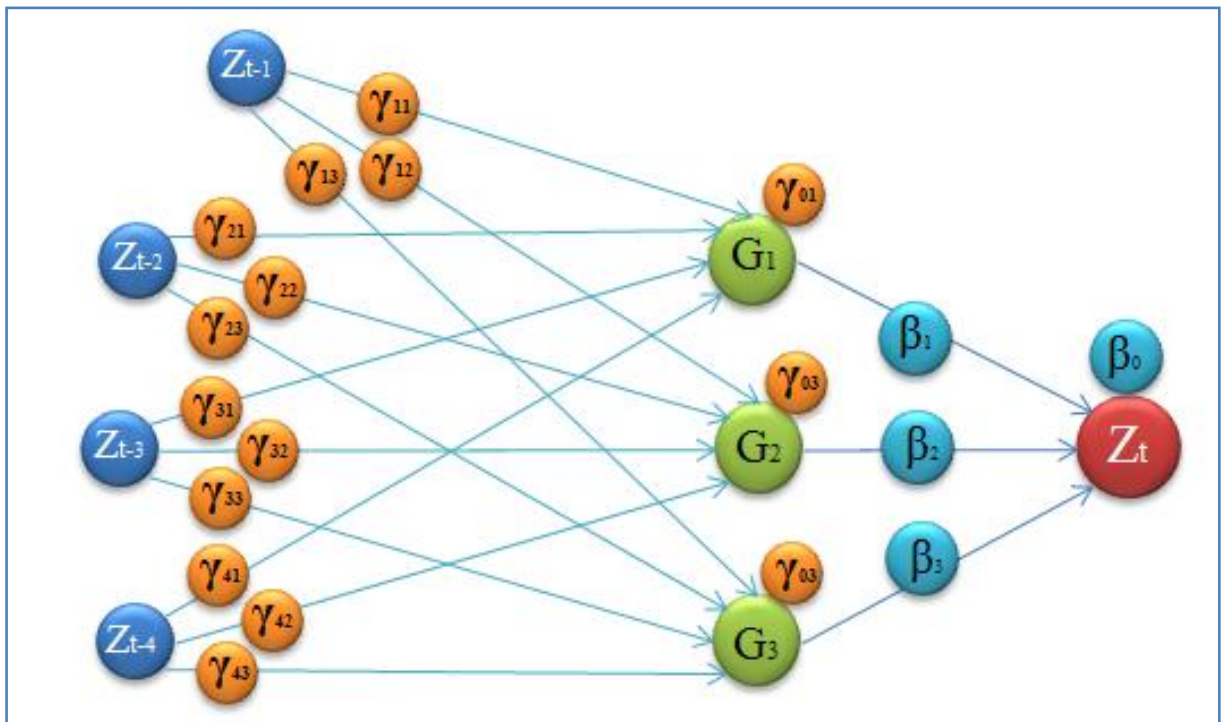


FIGURA 24 – EXEMPLO DE RNA DA SÉRIE CUSTO INDIRETO

Onde:

γ_{ij} = peso dado ao valor do i -ésimo neurônio de entrada, quando este valor é passado ao j -ésimo neurônio da camada intermediária, com $i=1:4$ e $j=1:3$;

γ_{0j} = constante na entrada do j -ésimo neurônio da camada intermediária, com $j=1:3$;

G_j = função não linear de ativação do j -ésimo neurônio da camada intermediária, com $j=1:3$;

β_j = peso dado à saída do j -ésimo neurônio da camada intermediária, com $j=1:3$;

β_0 = constante na entrada do neurônio de saída;

Encontram-se abaixo os 19 pesos da RNA, conforme resultados no R, para os pesos dos 15 elementos de γ e para os 4 elementos de β :

-1,566; -4,102 ; -5,592 ; -4,836 ; 5,144 ; 2,455 ; 4,561 ; 4,001 ; 3,775 ;
0,567 ; -1,424 ; 3,779 ; 1,443 ; 0,327 ; 8,741 ; -3,995 ; 1,977 ; 4,249 ;
-1,091 .

Na função “*selectnNet*” do pacote *tsDyn*, não é possível identificar o peso relacionado-o ao parâmetro.

6 CONCLUSÃO

De modo geral, este trabalho mostra a aplicabilidade de séries temporais lineares e não-lineares, bem como Redes Neurais Artificiais, em previsões futuras na área de vendas varejistas no Brasil.

Para uma implantação prática na empresa ABC, sugere-se gerar vários modelos dentre as classes propostas neste trabalho, não se atendo somente aos selecionados pelas funções automáticas do pacote tsDyn. Pois como é utilizado o AIC para selecionar os melhores modelos pelo tsDyn, e este é uma medida de ajuste aos dados, isto não implica que o modelo com o melhor ajuste terá a melhor previsão. Isto pôde ser verificado em todas as séries, pois em nenhuma delas o modelo com menor AIC, foi o que previu valores com maior eficiência. Desta forma, podem-se gerar vários modelos e compara-los somente no quesito previsão, ou seja, em termos de RMSE.

Pode-se ainda, com a criação de um fator de ponderação, realizar previsões compostas, ou seja, previsões utilizando uma média ponderada dos melhores modelos.

O próximo passo será o aprofundamento das técnicas de ajuste de modelos, de modo a automatizar a comparação do erro de previsão (RMSE), buscando encontrar sempre o modelo com maior eficiência preditiva. Desta maneira subsidiar as previsões de Lucratividade com embasamento estatístico, auxiliando assim o Planejamento e Gestão da empresa ABC, com metodologias já academicamente consagradas.

REFERÊNCIAS

BRAGA, A., LUDERMIR, T., CARVALHO, A.; **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro: Livros Técnicos e Científicos, 2000.

BRESSAN, A.A.; **Tomada de Decisão em Futuros Agropecuários com Modelos de Previsão de Séries Temporais**. RAE-eletrônica, v. 3, n. 1, Art. 9, jan./jun. 2004 . Disponível em <<http://www.rae.com.br/electronica/index.cfm?FuseAction=Artigo&ID=1914&Secao=FINANÇAS&Volume=3&Numero=1&Ano=2004>>. Acesso em: 13/08/2007.

CHAN, K. S.; TONG, H.; **On estimating Thresholds in autoregressive models**. *Journal of Time Series Analysis*, 7, pág.179–190, ano 1986.

EHLERS, R. S.; **Análise de Séries Temporais**: Apostila. 3.ed. Curitiba: 2005. Disponível em < <http://leg.ufpr.br/~ehlers/notas/stemp.pdf>>. Acesso em: 12/08/2007.

HAYKIN, S.; **Neural Networks: A Comprehensive Foundation**, Prentice Hall, EUA: 1999

MATOS, O. C. de; **Econometria Básica**: Teoria e Aplicações. 3.ed. rev. e ampl.. São Paulo: Atlas, 2000.

MORETIN, P. A.; TOLOI M. C. **Análise de Séries Temporais** ABE – Projeto Fischer. 1. ed. São Paulo: Edgard Blücher, 2004.

PEREZ, L.H.; PINO, F.A.; **Preço Recebido pelo Produtor de Bananas em SP: Uma Análise de Séries Temporais**. Revista Agricultura em SP, n. 42, Art. 1, 1995 . Disponível em < <http://www.iea.sp.gov.br/out/verTexto.php?codTexto=1083> >. Acesso em: 13/08/2007.

TERÄSVIRTA, T.; **Specification, estimation, and evaluation of smooth transition autoregressive models**. *Journal of the American Statistical Association*, 89(425), pág. 208–218, ano 1994.

ANEXOS

Anexo 01 – Resumo dos comandos utilizados

```
##### Descritiva #####
```

```
## Arquivos Base p/ saídas do R
```

```
arq01="X:/01_UFPR/01_CE228 - Lab 02/Saidas_R/vol_dec_01.xls"
```

```
arq02="X:/01_UFPR/01_CE228 - Lab 02/Saidas_R/vol_dec_02.xls"
```

```
arq03="X:/01_UFPR/01_CE228 - Lab 02/Saidas_R/vol_dec_03.xls"
```

```
arq13="X:/01_UFPR/01_CE228 - Lab 02/Saidas_R/vol_acf.xls"
```

```
arq14="X:/01_UFPR/01_CE228 - Lab 02/Saidas_R/vol_pacf.xls"
```

```
arq21="X:/01_UFPR/01_CE228 - Lab 02/Saidas_R/vol_proj.xls"
```

```
arq25="X:/01_UFPR/01_CE228 - Lab 02/Saidas_R/mod_lin.xls"
```

```
arq26="X:/01_UFPR/01_CE228 - Lab 02/Saidas_R/mod_setar.xls"
```

```
arq27="X:/01_UFPR/01_CE228 - Lab 02/Saidas_R/mod_istar.xls"
```

```
arq28="X:/01_UFPR/01_CE228 - Lab 02/Saidas_R/mod_nnet.xls"
```

```
## Carregando pacotes
```

```
require(tsDyn)
```

```
require(xlsReadWrite)
```

```
base = read.table(file="X:/01_UFPR/01_CE228 - Lab 02/Comandos/base_geral2.txt",
h=T)
```

```
vol = ts(base$vol_mi[base$mercado=="MI"],frequency=12,start=c(2003,1));vol
```

```
vol_var = ts(base$vol_mi_var[base$mercado=="MI"],frequency=12,start=c(2003,1));vol
_var
```

```
### Descritiva
```

```
par(mfrow=c(2,1))
```

```
plot(vol,main="Volume de Vendas")
```

```
plot(vol_var,main="Variação do Volume de Vendas")
```

```
### Decomposição Clássica
```

```
vol_var_dec = decompose(vol_var)
```

```
par(mfrow=c(3,1))
```

```
plot(vol_var,main="Volume de Vendas")
```

```
lines(vol_var_dec$trend)
```

```
plot(vol_var_dec$random)
```

```
plot(vol_var_dec$seasonal)
```

```
#Gerando as saídas
```

```
write.xls(vol_var_dec$trend , arq01, colNames = TRUE, sheet = 1, from = 1)
```

```
write.xls(vol_var_dec$random , arq02, colNames = TRUE, sheet = 1, from = 1)
```

```
write.xls(vol_var_dec$seasonal , arq03, colNames = TRUE, sheet = 1, from = 1)
```

```
# Correlogramas FAC e FACP
```

```
par(mfrow = c(3,1))
```

```
plot(vol_var,xlab='tempo',ylab='observações',main="Volume de Vendas")
```

```
acf (vol_var,xlab='defasagem',ylab='fac',main='FAC do Volume')
```

```
pacf(vol_var,xlab='defasagem',ylab='facp',main='FACP do Volume')
```

#Gerando as saídas

```
write.xls(acf(vol_var, plot=F)$acf , arq13, colNames = TRUE, sheet = 1, from = 1)
write.xls(pacf(vol_var, plot=F)$acf , arq14, colNames = TRUE, sheet = 1, from = 1)
```

Normalidade

```
hist(vol_var, br = 15, prob=T, main="Histograma do Volume de Vendas")
curve(dnorm(x,mean(vol_var),sd(vol_var)),
      min(vol_var)*0.5,max(vol_var)*1.5,add=T)
```

Seleção de Modelos

```
vol_var_base = window(vol_var, start=c(2003,1), end=c(2007,9)); vol_var_base
vol_var_fim = window(vol_var,start=c(2007,1)); vol_var_fim
```

Modelo Linear

```
mod_linear_vol = list()
mod_linear_vol[["Volume_AR01"]] = linear(vol_var_base,m=1)
mod_linear_vol[["Volume_AR02"]] = linear(vol_var_base,m=2)
mod_linear_vol[["Volume_AR03"]] = linear(vol_var_base,m=3)
mod_linear_vol[["Volume_AR04"]] = linear(vol_var_base,m=4)
mod_linear_vol[["Volume_AR05"]] = linear(vol_var_base,m=4)
mod_linear_vol
```

```
sapply(mod_linear_vol,AIC);sapply(mod_linear_vol,MAPE)
summary(mod_linear_vol[["Volume AR04"]]) #Este é o modelo por parcimoniedade
linear = matrix(nrow = 10, ncol=4, byrow=TRUE)
linear[,1] = as.matrix(c(sapply(mod_linear_vol,AIC),sapply(mod_linear_vol,MAPE)))
linear[,2] =
as.matrix(c(sapply(mod_linear_preco,AIC),sapply(mod_linear_preco,MAPE)))
linear[,3] =
as.matrix(c(sapply(mod_linear_custo_dir,AIC),sapply(mod_linear_custo_dir,MAPE)))
linear[,4] =
as.matrix(c(sapply(mod_linear_custo_ind,AIC),sapply(mod_linear_custo_ind,MAPE)))
saida_linear_xls = as.data.frame(linear, colnames = c("Volume", "Preco", "Cto_D",
"Cto_Ind"))
write.xls(saida_linear_xls , arq25, colNames = TRUE, sheet = 1, from = 1)
```

Setar

```
mod_setar_vol = selectSETAR(vol_var_base, m=5, d=1, thSteps=5, criterion="AIC")
mod_setar_vol
saida_setar_xls = as.data.frame(c(
mod_setar_vol,mod_setar_preco,mod_setar_custo_dir,mod_setar_custo_ind))
write.xls(saida_setar_xls , arq26, colNames = TRUE, sheet = 1, from = 1)
```

#LSTAR

```
mod_lstar_vol = selectLSTAR(vol_var_base, m=5, d=1, thDelay=0:4)
mod_lstar_vol
saida_lstar_xls = as.data.frame(c(
mod_lstar_vol,mod_lstar_preco,mod_lstar_custo_dir,mod_lstar_custo_ind))
write.xls(saida_lstar_xls , arq27, colNames = TRUE, sheet = 1, from = 1)
```

#Redes Neurais

```
mod_nnet_vol = selectNNET(vol_var_base, m=5, size=1:10, maxit=10000)
mod_nnet_vol
```

```

mod_nnet_vol_prev1 = nnetTs(vol_var_base, m=5,
size=4);summary(mod_nnet_vol_prev1)
saida_nnet_xls = as.data.frame(c(
mod_nnet_vol,mod_nnet_preco,mod_nnet_custo_dir,mod_nnet_custo_ind))
write.xls(saida_nnet_xls , arq28, colNames = TRUE, sheet = 1, from = 1)

```

teste de predição

```

mod_test_vol = list()
mod_test_vol[["Linear_AR04"]] = mod_linear_vol[["Volume_AR04"]]
mod_test_vol[["Setar 01"]] = mod_setar_vol_prev1
mod_test_vol[["Setar 02"]] = mod_setar_vol_prev2
mod_test_vol[["Setar 03"]] = mod_setar_vol_prev3
mod_test_vol[["Setar 04"]] = mod_setar_vol_prev4
mod_test_vol[["Lstar 01"]] = mod_lstar_vol_prev1
mod_test_vol[["Lstar 02"]] = mod_lstar_vol_prev2
mod_test_vol[["Lstar 03"]] = mod_lstar_vol_prev3
mod_test_vol[["Lstar 04"]] = mod_lstar_vol_prev4
mod_test_vol[["nNet"]] = mod_nnet_vol_prev1
prev_test_vol = lapply(mod_test_vol, predict, n.ahead = 5)

```

```

plot(vol_var_fim, type = "l")
for (i in 1:10) lines(prev_test_vol[[i]], lty = i+1, col=i+1)
legend(2007.2,-0.0, lty=1:11, col=1:11,legend=c("Valores
Reais",names(prev_test_vol)))

```

```

erro_abs_vol=list()
for (i in 1:10) erro_abs_vol[[names(prev_test_vol[i])] =
abs(as.matrix(window(vol_var_fim,start=c(2007,10))) -
as.matrix(as.data.frame(prev_test_vol[(names(prev_test_vol)[i]))]))))
erro_abs_vol_media = mean(as.data.frame(erro_abs_vol))
erro_abs_vol_media
summary(mod_setar_vol_prev3)

```

Gerando os arquivos de Previsão

```

write.xls(as.data.frame(prev_test_vol), arq21, colNames = TRUE, sheet = 1, from = 1)

```