RODRIGO FORTINI DE LUCCA

UM ESTUDO EXPLORATÓRIO USANDO SÉRIES TEMPORAIS PARA PREVISÃO DE VENDAS NO VAREJO SUPERMERCADISTA

São Paulo
2012

RODRIGO FORTINI DE LUCCA

UM ESTUDO EXPLORATÓRIO USANDO SÉRIES TEMPORAIS PARA PREVISÃO DE VENDAS NO VAREJO SUPERMERCADISTA

Monografia apresentada ao Programa de Educação Continuada da Escola Politécnica da Universidade de São Paulo para obtenção do certificado de conclusão do programa de MBA em Engenharia Financeira

Área de Concentração: Modelos de Previsão de Séries Temporais

Orientador: Prof. Dr. Claudio Garcia

São Paulo
2012

FICHA CATALOGRÁFICA

Lucca, Rodrigo Fortini de

Um estudo exploratório usando séries temporais para previsão de vendas no varejo supermercadista / R.F. de Lucca. São Paulo, 2012.

52 p.

Monografia (MBA em Engenharia Financeira) – Escola Politécnica da Universidade de São Paulo. Programa de Educação Continuada em Engenharia.

1. Previsão de vendas 2. Vendas (Planejamento) 3. Varejo. I. Universidade de São Paulo. Escola Politécnica. Programa de Educação Continuada em Engenharia II. t.

DEDICATÓRIA

Dedico este trabalho à minha esposa, amiga e companheira Angela e à minha amada filha Mariana que tiveram a paciência de me suportar em meus momentos de ansiedade, frustração e busca pela sabedoria, e por me apoiarem nesta caminhada.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por toda saúde, paciência e determinação para finalizar este trabalho, Ele que dá o sentido a todos os nossos desafios e a força necessária para conclusão deles.

Agradeço ao prof. Dr. Claudio Garcia por todo ensinamento, determinação, paciência, dedicação e amizade construída no decorrer deste trabalho.

Pela dedicação e preocupação dos professores coordenadores do MBA em Engenharia Financeira, Dr. Roberto Moura Sales e Dr. Oswaldo Luiz do Valle Costa, em oferecer o melhor conteúdo programático ao curso.

À minha esposa, Angela, e à minha filha, Mariana, pela dedicação que lhes furtei aos finais de semana e longas noites de estudo.

Agradeço a todos os familiares que me apoiaram nos momentos mais difíceis de tempo escasso e cansaço.

Aos amigos do PECE por toda a convivência, apoio e troca de experiências dentro e fora das salas de aulas.

O autor.

EPIGRAFE

"Eu vejo o futuro repetir o passado [...]

"O tempo não para, não para..."

Cazuza/ Arnaldo Brandão

Letra da Música "O tempo não para"

RESUMO

O objetivo principal do trabalho é estruturar um método quantitativo para modelagem e previsão de vendas no varejo supermercadista, entendendo os principais componentes que afetam tal previsão, e a partir deles modelar os dados históricos.

Entender como o comportamento das vendas atuais influencia as vendas futuras está diretamente ligado à quantidade demandada pelo cliente, ao poder de compra do consumidor, ao cenário econômico e até mesmo à execução de sua estratégia de expansão abrangendo aberturas de novos pontos de vendas.

Para tanto, utilizou-se a econometria de séries temporais com testes dos modelos ARMA (Auto Regressivo de Médias Móveis). Os dados do estudo foram extraídos do site corporativo da empresa e se referem aos valores mensais de vendas durante o período de 1968 a 2012 para todas as lojas do grupo ao redor do mundo.

Para realizar as análises, fez-se uso do aplicativo computacional Matlab® versão 7.10.0 R2010a.

Palavras-chave: *Previsão de vendas, planejamento, varejo e supermercado.*

ABSTRACT

The main goal of this paper is to structure a quantitative methodology for sales modeling and forecast in supermarket retail, understanding the main components that affect such prediction, and from them model the historical data.

The understanding of how current sales behavior affects future sales is directly connected to the quantity requested by the client, the consumer's purchasing power, the economic scenario and even the implementation of its expansion strategy comprising the opening of new points of sales.

In order to do so, the econometrics of time series analysis with tests on the ARMA (Auto Regressive Moving Average) model was applied. The data of this study were extracted from the corporate website of the company and refer to sales monthly values between 1968 and 2012 for all the stores of the group around the world.

The analysis was performed by using the software Matlab® version 7.10.0 R2010a.

Key-words: *Sales Forecasting, planning, retail e supermarket.*

LISTA DE FIGURAS

LISTA DE ABREVIATURAS E SIGLAS

AIC	Akaike Information Criteria – Critério de Informação Akaike
AR	Auto Regressive - Autorregressivo
ARMA	Auto Regressive Moving Average – Autorregressivo de Médias Móveis
ARIMA	Auto Regressive Integrated Moving Average – Autorregressivo Integrado de Médias Móveis
BIC	Bayesian Information Criteria – Critério de Informação Bayesiano
Cov	Covariância
FAC	Função de Auto Correlação
FACP	Função de Auto Correlação Parcial
FPE	Final Prediction Error – Erro Final de Predição
IBOVESPA	Índice Bovespa
MA	Moving Average – Médias Móveis
Matlab®	Marca registrada da empresa *The Mathworks, Inc.*
NYSE	New York Stock Exchange – Bolsa de Valores de Nova York
SARIMA	Seasonal Auto Regressive Integrated Moving Average – Sazonal Autorregressivo Integrado de Médias Móveis
Var	Variância

SUMÁRIO

1. INTRODUÇÃO

1.1. COMENTÁRIOS INICIAIS

Ao final do século passado, a palavra de ordem para o mercado mundial passou a ser "globalização". Os efeitos desse processo de integração econômica têm forçado as empresas a se adaptarem continuamente para entender a situação em que se encontram e definirem os rumos para enfrentar os desafios de se manterem eficientes e assegurarem sua participação em seus mercados.

Neste cenário, a competitividade tem feito com que as empresas busquem melhores formas de planejamento com uma gestão eficiente focada em resultados, e que tenham uma visão clara e objetiva das expectativas futuras de sua estratégia de atuação.

No segmento varejista supermercadista, foco deste estudo, as empresas precisam estar aptas a um tempo de resposta curto dado sua concorrência agressiva. Muitas vezes, entender como o comportamento das vendas atuais influenciam as vendas futuras está diretamente ligado à quantidade demandada pelo cliente, ao poder de compra do consumidor, ao cenário econômico e até mesmo à execução de sua estratégia de expansão abrangendo aberturas de novos pontos de vendas. É neste ponto que o planejamento eficaz direciona o gestor a uma correta tomada de decisões.

Um planejamento de vendas ineficaz pode causar resultados negativos desde a falta de produtos nas gôndolas para o consumidor final, redução da expectativa dos lucros associados à venda não realizada e impacto negativo à imagem da empresa.

Segundo MAKRIDAKIS et al (1998), as previsões quantitativas são possíveis quando a informação sobre o passado está disponível, podendo ser quantificada de forma numérica e podendo-se considerar que alguns padrões que aconteceram no passado irão se repetir no futuro.

A econometria de séries temporais tem auxiliado as previsões quantitativas, dado que seu ponto principal é explicar fatos passados, testar modelos e inferir sobre o futuro, ou seja, tirar conclusões sobre o comportamento passado de uma dada variável e que poderão ser úteis para proporcionar informações sobre seu comportamento futuro provável.

Cabe ressaltar que, além da correta previsão quantitativa, a base dados utilizada para aplicar técnicas de previsão inclui ainda informações qualitativas que são obtidas através de especialistas que atuam e conhecem o comportamento das variáveis em estudo, e que consigam interpretá-las de acordo com os negócios e estratégias da empresa.

1.2. RELEVÂNCIA DO TEMA

O tema desta monografia é a previsão de vendas no varejo supermercadista, considerando a abordagem econométrica através de um modelo que melhor aperfeiçoe o planejamento de vendas da rede varejista Walmart.

Dado que o planejamento de vendas tem como foco principal fortalecer a imagem da marca e aumentar sua participação no mercado em que atua, torna-se evidente a importância da acurácia do modelo para fins de tomada de decisão.

1.3. OBJETIVO

O objetivo principal do trabalho é estruturar um método quantitativo para modelagem e previsão de vendas no varejo supermercadista, entendendo os principais componentes que afetam tal previsão, e a partir deles inferir na modelagem dos dados históricos.

Analisando a série temporal obtida através dos relatórios de administração no endereço eletrônico da empresa http://investors.walmartstores.com, os principais elementos que compõem o estudo são: os dados históricos de vendas líquidas que indicam a tendência, o ciclo e a sazonalidade da série, compostas por todas as lojas abertas durante o período analisado de 1968 a 2012.

A estruturação do modelo será feita através da metodologia ARMA que pode ser considerado o modelo com melhor aplicação neste estudo.

1.4. JUSTIFICATIVA

Dada à volatilidade que o varejo supermercadista apresenta, realizar um planejamento de vendas baseado em premissas qualitativas é um fator errôneo que muitos gestores consideram, e não se importam em analisar os dados passados a fim de entender quais eventos ocorreram e quais irão ocorrer no futuro.

O trabalho visa entender todos os aspectos quantitativos da série temporal analisada e responder questões como:

- Quais influências que um modelo quantitativo pode ter no planejamento de vendas e tomada de decisões estratégicas?

- Qual impacto que o modelo pode gerar na estratégia e plano de expansão de lojas?

- Quais são os impactos financeiros, se o varejo utilizar previsões qualitativas sem considerar o entendimento da tendência, ciclicidade e sazonalidade dos dados históricos?

1.5. MÉTODO DE TRABALHO

A metodologia proposta é baseada no estudo exploratório que é realizado sobre um problema ou questão de pesquisa que geralmente são assuntos com pouco ou nenhum estudo anterior a seu respeito.

O objetivo desse tipo de estudo é procurar padrões, ideias ou hipóteses. As técnicas tipicamente utilizadas para a pesquisa exploratória são estudos de caso, observações ou análises históricas, e seus resultados geralmente fornecem dados qualitativos ou quantitativos.

A pesquisa exploratória avaliará quais teorias ou conceitos existentes podem ser aplicados a um determinado problema ou se novas teorias e conceitos devem ser desenvolvidos.

O tema assume a forma de um estudo de caso, no qual serão utilizadas as técnicas de pesquisas teóricas, trabalho de laboratório e levantamento de dados sobre os conceitos e questões ligadas à econometria de séries temporais.

1.6. LIMITAÇÃO DO TRABALHO

O trabalho se limita em escolher um método de previsão que melhor atenda a integração das previsões. Outra limitação é referente às observações que serão feitas com base no planejamento de vendas que aqui será considerado como vendas líquidas, ou seja, faturamento bruto excluindo os impostos incidentes sobre vendas e outros possíveis descontos.

2. SOBRE SÉRIES TEMPORAIS

2.1. SÉRIES TEMPORAIS

Uma série temporal pode ser definida como um conjunto de observações quantitativas de uma determinada variável aleatória disposta sequencialmente no tempo.

A natureza das séries temporais e suas estruturas de funcionamento estão relacionadas com a ordem de ocorrência dos fatos, e podem ser classificadas em contínuas e discretas.

A série temporal é definida como contínua quando suas observações são caracterizadas em intervalo contínuo de tempo ou quando o conjunto dos possíveis valores for representado por um contínuo de números para qualquer intervalo dos números reais ($\Re$). O conjunto será definido como $T = \{t : t_1 < t < t_2\}$ e a série denotada por $\{X(t) : t \in T\}$. Como exemplos podem ser citados o índice pluviométrico de uma determinada região e o tempo de duração de uma ligação telefônica.

As séries temporais discretas são mais utilizadas e simplificam o estado matemático que suportam os modelos econométricos, pois cada observação está associada a um tempo distinto que demanda uma relação de dependência serial entre essas observações. O conjunto será definido por $T = \{ t_1 , \ldots t_n \}$ e a série denotada por $\{ X_t : t \in T \}$. Como exemplos podem ser citados os valores de vendas diárias de uma empresa e a variação do índice IBOVESPA.

Segundo FISCHER (1982): "a grande maioria dos estudos econômicos que se valem do tratamento de séries temporais utiliza séries discretas, onde as observações são geradas em um intervalo de tempo com amplitude constante".

Uma série temporal pode ser classificada ainda como determinística ou estocástica[1]. As séries temporais com tendência a serem modeladas por uma função e que podem ser previstas com exatidão são denominadas determinísticas. Quando não se pode ter uma previsão exata sobre as observações futuras, elas devem ser descritas através de uma distribuição de probabilidades, e denominam-se estocásticas.

As séries temporais ainda são compostas por quatro elementos comportamentais: a tendência, a sazonalidade, as variações cíclicas e o ruído aleatório ou ruído branco.

A tendência é atribuída a um movimento regular e contínuo no longo prazo, caracterizada por um movimento ascendente ou descendente em um período de tempo. Em termos estatísticos, pode-se representá-la como o valor esperado de Y que varia no tempo $E(Y_t)$.

A sazonalidade é explicada pelas variações periódicas caracterizadas por um movimento ondulatório de curta duração, em geral, inferior a um ano. Esses movimentos poder ser estendidos a qualquer intervalo de curto prazo, como diário, semanal, mensal etc.

As variações cíclicas são as que se referem ao movimento ondulatório de longo prazo que tende a se tornar periódico. Notadamente estão ligadas a ciclos econômicos, mas que não são regulares como a sazonalidade.

O ruído aleatório ou ruído branco (ε) refere-se não só aos movimentos esporádicos gerados por um evento aleatório imprevisível, mas também ao conjunto de todos aqueles movimentos da série que não foram identificados junto aos demais componentes anteriores, dado que não obedece nenhuma relação comportamental que possa ser descrita de forma determinística. O ruído aleatório é estritamente estocástico.

---

[1] O termo estocástico proveniente da palavra grega "stokhos" significa em branco. Diz-se algo que está relacionado com o acaso e a respeito do que só é possível enunciar por probabilidades.

2.2. O SIGNIFICADO DA ANÁLISE DE SÉRIES TEMPORAIS

A análise das series temporais está diretamente ligada à possibilidade de serem realizadas inferências sobre o comportamento passado de uma dada variável e que poderão ser úteis para proporcionar informações sobre seu comportamento futuro. A partir deste estudo, poderão ser identificadas questões como: A série exibiu alguma tendência que possa influenciar seu futuro? Existe evidência de períodos sazonais que possam ser extrapolados para a predição?

Com base nas respostas e diante da interpretação de seus elementos comportamentais, é possível construir um modelo que consiga descrever os movimentos passados de uma variável e que poderão predizer os futuros movimentos da mesma, dado que a suposição tenha sido gerada por um processo estocástico.

2.3. PROCESSOS ESTOCÁSTICOS

Conforme citado anteriormente, se uma série temporal for obtida matematicamente e de forma exata, a série é caracterizada por uma natureza determinística. O processo estocástico deve ser entendido como um modelo que descreve a estrutura probabilística de uma sequência de observações.

O objetivo de um processo estocástico é buscar descrever as características de sua aleatoriedade a fim de proporcionar os instrumentos necessários para uma inferência sobre as probabilidades associadas a um conjunto de valores futuros à série. Ao se conseguir especificar numericamente uma determinada função de distribuição de probabilidades, torna-se viável inferir a probabilidade de outro valor futuro ocorrer.

Considere a variável aleatória Y e que seu valor para o instante t tenha relação serial com seu valor precedente Y_{t-1} conforme a seguir:

$$Y_t = Y_{t-1} + \varepsilon_t$$

onde ε_t ~ N(0,1), lê-se "ε_t tende à normal gaussiana com média zero e variância constante", ou seja, um processo de ruído branco.

2.4. RUÍDO BRANCO

Uma sequência (ε_t) é um processo chamado de ruído branco se cada valor da sequência tiver esperança nula (i), variância constante ou igual a um (ii) e não tiver covariância (iii) com nenhum dos demais valores da sequência, ou seja:

i) $E(\varepsilon_t) = 0$

ii) $Var(\varepsilon_t) = E(\varepsilon^2_t) = \sigma^2 = 1$

iii) $Cov(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t \varepsilon_s) - E(\varepsilon_t)E(\varepsilon_s) = 0$ ▯ $t \neq s$

2.5. AUTOCOVARIÂNCIA

A autocovariância mede o grau de dependência linear entre as observações próximas de um processo estocástico. Pode-se definir o coeficiente de autocovariância com defasagem "*k*" da série como sendo a covariância entre X_t e X_{t+k} onde "*k*" é o número de intervalos de tempo defasados:

$$\gamma_k = \text{cov}\,[X_t, X_{t+k}] = E[(X_t - \mu_t)(X_{t+k} - \mu_{t+k})]$$

A interpretação da equação acima é que a autocovariância entre dois pontos da série é o valor esperado do produto do desvio de cada ponto em relação à média do processo. Se uma observação que se encontra acima do valor médio tende a ser seguida por outra observação acima da média, "*k*" períodos defasados, ou semelhantemente, para observações X_t e X_{t+k} situadas abaixo da média, então a autocovariância será positiva. Caso contrário, apresentará comportamento negativo.

2.6. FUNÇÃO DE AUTOCORRELAÇÃO (FAC)

Uma característica importante à análise de séries temporais é a autocorrelação que determina o quanto o valor de uma realização de uma variável aleatória é capaz de influenciar seus valores próximos. Em geral, a autocorrelação é alta para eventos situados no passado imediato e diminui à medida que as observações vão se distanciando ao longo de vários períodos.

A função de autocorrelação (FAC) de uma série Y_t é dada por:

$$\rho_k = \frac{\sum_{t=k+1}^{n}(y_t - \overline{y})(y_{t-k} - \overline{y})}{\sum_{t=1}^{n}(y_t - \overline{y})^2}$$

2.7. FUNÇÃO DE AUTOCORRELAÇÃO PARCIAL (FACP)

A autocorrelação parcial expressa a correlação existente entre dois pontos y_t e y_{t-k} desconsiderando a influência sobre y_t dos demais pontos, ou seja, eliminam-se as correlações implícitas entre seus elementos intermediários y_{t-1}, y_{t-2}, y_{t-k-1} sobre o valor de y_t.

O coeficiente de autocorrelação parcial de ordem k é usualmente representado por ϕ_{kk} e dado pelo último coeficiente β_{kk} de cada uma das regressões a seguir:

$$Y_t = \beta_{11} y_{t-1} + \varepsilon_t$$

$$Y_t = \beta_{11} y_{t-1} + \beta_{22} y_{t-2} + \varepsilon_t$$

...

$$Y_t = \beta_{k1} y_{t-1} + \beta_{k2} y_{t-2} + \beta_{kk-1} y_{t-k-1} + \beta_{kk} y_{t-k} + \varepsilon_t$$

Os coeficientes de autocorrelação simples e parciais seguem aproximadamente uma distribuição normal com média zero e variância igual a um.

2.8. ESTACIONARIEDADE

Os processos estocásticos podem ser classificados como estacionários ou não estacionários. A estacionariedade de um processo estocástico pode ser interpretada como o equilíbrio estatístico.

Para HAMILTON (1994), se a média μ_t e as autocovariâncias γ_{jt} não dependerem do tempo *t*, então o processo para Y_t será dito estacionário em covariância ou fracamente estacionário:

$E[Y_t] = \mu$ para todo t = 1, 2, ..., n

$E[(Y_t - \mu)^2] = \gamma_0 = \sigma^2$ para todo t = 1, 2, ..., n

$E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j$ para todo t = 1, 2, ..., n com $t \neq j,\ j \in Z$

As duas primeiras condições indicam que a média e a variância da função distribuição de probabilidades não variam no tempo, e a covariância entre as observações existentes em dois períodos distintos de tempo depende somente das distâncias entre esses pontos e não do específico período de tempo das observações.

2.9. ERGODICIDADE

Considerando apenas a propriedade de estacionariedade, não é possível estimar uma série temporal. Essencialmente, quando utilizado uma modelagem de série temporal, esta deve satisfazer a propriedade da ergodicidade.

Um processo é dito ergódico quando com apenas uma realização do processo for possível caracterizá-lo, ou seja, se a média temporal $\bar{y}$ convergir em probabilidade para $E(y_t)$ conforme $T \to \infty$. A representação é dada a seguir:

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

2.10. PROCESSOS ESTACIONÁRIOS

WERNER (2004) conceitua os modelos como estacionários quando assumem o equilíbrio do processo. Os modelos serão ditos fracamente estacionários caso sua média e variância se mantenham constantes ao longo do tempo, e serão fortemente estacionários se todos esses momentos estatísticos permanecerem constantes ao longo do tempo.

A seguir alguns modelos utilizados em processos estacionários:

2.10.1. Modelo de Média Móvel – MA (q)

Um modelo MA (q) descreve como uma observação y_t é dependente de seu erro ε_t e de seu erro imediatamente passado. Considere que o processo estocástico seja representado a seguir:

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

Se o valor de y_t depende apenas de seu erro ε_{t-1} tem-se um processo chamado de média móvel de ordem 1 e denotado por MA (1). Se o processo depender de ε_{t-2} então será chamado de MA (2) e assim por diante.

O processo está sempre associado aos erros que o modelo pode gerar. Ainda assim, os pesos poderão ser diferentes dadas à importância das observações passadas que se quer analisar.

Para que y_t satisfaça a condição de estacionariedade do processo, é preciso calcular a esperança, a variância e as autocovariâncias:

i) $\mathrm{E}(y_t) = \mu + \mathrm{E}(\varepsilon_t) + \theta\,\mathrm{E}(\varepsilon_{t-1}) = \mu$

ii) $\mathrm{Var}(y_t) = \mathrm{E}(y_t - \mu)^2 = (1+\theta^2)\sigma^2$

iii) $\mathrm{E}[(y_t - \mu)(y_{t-1} - \mu)] = \sigma^2\theta$

A intepretação para cada um dos casos segue que: i) a esperança é constante e finita para cada "*t*"; ii) a variância é finita e iii) a autocovariância não depende de t. Como esperança e autocovariância não são funções do tempo tem-se que o processo é fracamente estacionário, independentemente do valor assumido por θ.

Por fim, a autocorrelação do processo só existe para a primeira defasagem e é dada por:

$$\rho = \frac{\theta\sigma^2}{(1+\theta)^2\sigma^2} = \frac{\theta}{(1+\theta)^2}$$

Em um processo MA (1), uma importante propriedade que gera sua função de autocorrelação é que a memória do processo só ocorre durante um período de tempo, ou seja, numa dada observação, Y_t está correlacionada apenas ao seu antecessor Y_{t-1} ou seu sucessor imediato Y_{t+1}, mas não com qualquer outro membro da série, como por exemplo, Y_{t+5}.

Abaixo, generaliza-se o processo de médias móveis para q defasagens MA(q):

$$y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$$

2.10.2. Modelo Autorregressivo - AR(p)

Um modelo AR (p) descreve como uma observação é dependente diretamente de seus valores regredidos no tempo por um parâmetro ϕ e por seu ruído branco ε_t.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t$$

onde os parâmetros correspondentes são:

Y_t observação da série temporal no tempo "t";

c constante do modelo;

ϕ_p parâmetro do modelo AR de ordem p; e

ε_t representa o erro de eventos aleatórios que não podem ser medidos e explicados pelo modelo.

Um modelo autorregressivo de 1ª ordem, ou AR(1) é a versão mais simples desta classe de modelos e é representado pela equação:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

Dado que o modelo apresente média e variância constantes ao longo do tempo, tem-se um AR (1) fracamente estacionário independentemente do valor assumido por ϕ:

i) $E(Y_t) = \mu + \phi E(\varepsilon_{t-j}) = \mu$

ii) $Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2 / (1+\phi^2)$

Para que a variância de Y_t seja não negativa e finita é necessário que os valores assumidos por $|\phi|$ sejam <1, onde esta condição é conhecida como estacionariedade do modelo.

Sendo assim, o modelo será estacionário ou mesmo fracamente estacionário, se as autocovariâncias não dependerem do tempo *t*. Dessa forma, as autocorrelações serão dependentes apenas de suas defasagens *k* que demonstra a distância entre os valores da série que são dados pela fórmula:

$$\rho = \frac{\gamma_k}{\gamma_o} = \phi_1^k \qquad \text{onde } k = 0, 1, 2 \ldots$$

A função de autocorrelação (FAC) pode ser calculada através das equações de Yule-Walker, substituindo-se *k* = 0, 1, 2... p na equação a seguir:

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \ldots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \ldots + \phi_p\rho_{p-2} \\
\vdots \quad & \quad \vdots \qquad \vdots \qquad \cdots \qquad \vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_{p-2} + \ldots + \phi_p
\end{aligned}$$

Escrevendo-se as equações de Yule-Walker na forma matricial, a solução para os parâmetros considerando as autocorrelações pode ser obtida a seguir:

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \rho_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \quad \rho_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix}$$

2.10.3. Modelo Autorregressivo de Média Móvel – ARMA (p,q)

Um bom modelo deve ser capaz de representar todos os dados de uma série temporal, ou a maior parte deles, com o menor número de parâmetros possível. Considerando-se essa premissa, os modelos autorregressivos de média móvel ARMA (p,q) podem representar séries que contenham características tanto de um modelo AR como de um modelo MA.

O processo autorregressivo de média móvel é simplesmente a combinação dos processos vistos anteriormente. Assim, um ARMA (p,q) é escrito a seguir:

$$Y_t = c + \underbrace{\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p}}_{\text{autorregressivo}} + \underbrace{\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}}_{\text{média móvel}}$$

Utilizando-se o operador atraso chega-se à seguinte proposição:

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots + \phi_p L^p)\, Y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)\varepsilon_t$$

Para se estimar um modelo ARMA, é necessário primeiramente que este apresente estabilidade, significando que as raízes unitárias do modelo acima estejam fora do círculo de raio unitário (processo estável).

Portanto, a estacionariedade de um processo ARMA será dependente apenas dos parâmetros autorregressivos ($\phi_1, \phi_2, \ldots, \phi_p$) e não dos parâmetros de média móvel ($\theta_1, \theta_2, \ldots, \theta_q$).

2.11. PROCESSOS NÃO ESTACIONÁRIOS

Quando um processo estocástico que gerou determinada série for dependente do tempo, diz-se que a série é não estacionária.

A não estacionariedade de uma série tem como características principais que: i) os dados não permanecem ao redor de uma média ao longo do tempo, e/ou ii) a variância está em constante alteração, já que os dados aumentam e diminuem no decorrer do tempo.

Os modelos não estacionários mais comuns são o ARIMA (p,d,q) e o SARIMA (p,d,q)(P,D,Q) que serão vistos na sequência.

A seguir algumas explanações sobre os principais modelos não estacionários.

2.11.1. Modelo Autorregressivo Integrado de Média Móvel – ARIMA (p,d,q)

No caso de modelagem de séries não estacionárias, o procedimento básico consiste em diferenciar a série tantas vezes quantas sejam necessárias para torná-la estacionária.

A primeira diferença de Y_t é definida como:

$$\Delta Y_t = Y_t - Y_{t-1}$$

A segunda diferença segue:

$$\Delta^2 Y_t = \Delta[\Delta Y_t] = \Delta[Y_t - Y_{t-1}] =$$
$$\Delta Y_t - \Delta Y_{t-1} =$$
$$Y_t - Y_{t-1} - Y_{t-1} + Y_{t-2} =$$
$$Y_t - 2(Y_{t-1}) + Y_{t-2}$$

Para MORETIN & TOLOI (1987), em situações normais, será suficiente tomar uma ou duas diferenças para que a série se torne estacionária.

Se y_t torna-se estacionária após "d" diferenças e a série descrita como z_t que resulta destas diferenças for um modelo do tipo ARMA (p,q), então diz-se que y_t será descrito por um sistema ARIMA (p,d,q) do tipo:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

$$z_t = \Delta^d y_t$$

Um modelo ARIMA (p,d,q) pode ser expresso por:

$$\phi_p(B)(1-B)^d y_t = \theta_q(B)\varepsilon_t$$

Ou seja, um ARIMA (p,d,q) é apenas um modelo ARMA (p,q) que sofreu uma transformação que implicou na substituição do termo de $\phi_p(B)y_t$ pelo componente $\phi_p(B)(1-B)^d y_t$, e que portanto possuem as mesmas características.

2.11.2. Modelo Sazonal - SARIMA (p,d,q) x (P,D,Q)s

Algumas séries apresentam componentes sazonais importantes. Considere uma série y_t com um componente sazonal de periodicidade "*s*" que pode apresentar dependência em relação aos seus valores passados para a mesma posição, ou seja, y_{t-s}, y_{t-2s} e assim por diante.

Para os modelos que contemplam as séries com autocorrelação sazonal, atribui-se o nome de SARIMA, cuja nomenclatura segue *Seasonal Auto Regressive Integrated Moving Average.*

O modelo SARIMA é composto de duas partes, sendo uma não-sazonal com parâmetros (p,d,q) e a parte sazonal (P,D,Q)s.

WERNER (2004) cita que o modelo mais comumente utilizado para um SARIMA é dado pela equação a seguir:

$$(1-\phi_1 L-...-\phi_p L^p)(1-\Phi_1 L^s-...-\Phi_P L^{Ps})(1-L)^d(1-L^s)^D Z_t = (1-\theta_1 L-...-\theta_q L^q)(1-\Theta_1 L^s-...-\Theta_Q L^{Qs})\varepsilon_t$$

onde:

$(1-\phi_1 L-...-\phi_p L^p)$	é a parte autorregresiva não sazonal de ordem p;
$(1-\Phi_1 L^s-...-\Phi_P L^{Ps})$	é parte autorregresiva sazonal de ordem P e sazonalidade s;
$(1-L)^d$	é a parte de integração não sazonal de ordem d;
$(1-L^s)^D$	é a parte de integração sazonal de ordem D e sazonalidade s;
$(1-\theta_1 L-...-\theta_q L^q)$	é a parte não sazonal de médias móveis de ordem q;
$(1-\Theta_1 L^s-...-\Theta_Q L^{Qs})$	é a parte sazonal de médias móveis de ordem Q e sazonalidade s.

A sazonalidade do modelo pode ser observada através do comportamento da função de autocorrelação e função de autocorrelação parcial, entretanto, analisam-se as defasagens sazonais, ou seja, se o modelo apresentar uma sazonalidade de periodicidade anual, então as defasagens serão detectadas em k = 12, 24, 36...

2.12. CRITÉRIOS DE INFORMAÇÃO E SELEÇÃO DE MODELOS

Depois de realizadas todas as etapas de estimação dos modelos serão necessárias analisar os resíduos que a série temporal gerou e o resultado estatístico dos parâmetros utilizados. Se o modelo estiver com os parâmetros ajustados, os resultados explicarão a maior parte da relação entre os valores da série e deixarão os resíduos pouco explicáveis com menor evidência.

Nesta etapa, pode-se citar o teste Q de Ljung-Box, que testa a hipótese nula de ausência de autocorrelação, aplicado aos resíduos dos modelos selecionados. Sua equação pode ser verificada a seguir:

$$Q = n\,(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

onde os parâmetros são:

n tamanho da amostra;

$\hat{\rho}_k$ autocorrelação da amostra com *lag k;*

h número de *lags* sendo testados;

Depois de feita esta análise, é possível que mais de um modelo represente o processo estocástico da série gerada. A comparação entre modelos só é válida mediante o uso dos chamados critérios de informação. Entre os diversos critérios existentes, há três bastante utilizados:

2.12.1. Akaike Information Criteria (AIC)

É um teste frequentemente utilizado para a escolha da ótima especificação de um modelo onde não existem variáveis independentes comuns aos dois. O resultado gerado pelo teste é o que melhor apresentar o menor valor do critério de Akaike. Como exemplo, pode ser citado que o número de defasagens (*k*) a serem incluídas numa equação pode ser identificado pela seleção que produza o menor valor para o critério AIC.

GARCIA (2011) define o critério de Akaike (AIC) como:

$$AIC(p,q) = ln(V_T)\frac{2\ (p+q)}{T}$$

onde os parâmetros são:

p ordem de um modelo AR;`

q ordem de um modelo MA;

V_T função custo ou variância de ε_t;e

T número de amostras coletadas para obtenção do modelo;

2.12.2. Bayesian Information Criteria (BIC)

É um teste semelhante ao critério de AIC, no entanto, sua característica principal é a imposição de uma penalidade maior pela inclusão de coeficientes adicionais a serem estimados. GARCIA (2011) define o critério de BIC como:

$$BIC\ (p,q) = ln(V_T) + \frac{ln\ (T)}{T}\ (p+q)$$

Dependendo do comportamento das séries temporais que se queira analisar, vários modelos podem ser empregados na previsão de valores futuros. A escolha do melhor modelo é feita, geralmente, pelo princípio da parcimônia, já que é o modelo que minimizará os critérios adotados.

2.12.3. Final Prediction Error (FPE)

O FPE (Erro Final de Predição) fornece uma medida de qualidade ao modelo, simulando a situação em que ele seja ensaiado com um conjunto de dados diferentes. Após calcular vários modelos diferentes, o FPE pode ser usado na comparação entre eles. É também atribuído a Akaike, e o seu modelo tem maior precisão quando apresentado um menor FPE. Sua fórmula é definida por:

$$FPE\,(p,q) = V_T * \left(\frac{1+\frac{p+q}{T}}{1-\frac{p+q}{T}}\right)$$

Embora haja diversos critérios para seleção de modelos ARMA de ordem (p,q) é necessário ressaltar que não existem conclusões acerca de qual método seja melhor do que o outro. Como se trata de testes empíricos, a escolha da ordem que gere os mínimos valores e a opinião de quem as interpreta, são fatores que devem ser ponderados nesta análise.

3. ANÁLISE SETORIAL

3.1. O VAREJO SUPERMERCADISTA BRASILEIRO

Embora o mundo tenha sofrido com a grande crise econômica no final de 2008, o Brasil tem sido favorecido e experimentado um avanço positivo no consumo, em decorrência, principalmente da melhora no poder aquisitivo das famílias, da elevação da renda e do nível de concessão de crédito. Não obstante, tais fatores foram favorecidos indiretamente pelas políticas governamentais de inclusão social com foco às classes sociais menos favorecidas neste cenário.

Dentre deste segmento, o comércio de veículos e as atividades supermercadistas são os maiores em volume de receitas. Em particular, o ramo de hipermercados e supermercados apresenta forte tendência de expansão, uma vez que a relação existente entre população e território no Brasil é favorável se comparados com alguns países da Europa e Estados Unidos.

O varejo supermercadista apresenta ainda um fator de sazonalidade significativa de demanda e alto nível de giro de produtos, além de forte suscetibilidade a políticas econômicas que afetam a conjuntura macroeconômica e os indicadores de renda e emprego. Portanto, fatores como aumento da população e estabilidade econômica são fatores preponderantes para o crescimento da atividade.

O setor busca constantemente a eficiência operacional e financeira para gerar vantagem competitiva em relação à concorrência, pois entender a estrutura de custos, qualidade e atendimento, impacta diretamente na percepção do consumidor. Esta ação evidencia a fidelização da marca e o conhecimento de suas preferências, além de poder ofertar um número maior de produtos diferenciados.

Com a entrada de novos participantes externos dentro do mercado, os movimentos de fusões e aquisições tiveram destaque, principalmente na busca por aquisição de redes menores e de supermercados populares localizados na periferia, com opções de pontos comerciais, a fim de expandir sua presença local.

Com isso, as grandes redes têm investido na diversificação dos formatos, criando bandeiras de lojas de vizinhança para disputar com supermercados menores nas grandes cidades. Algumas redes optam por modalidades mais simples com foco na população de baixa renda.

Outro movimento bastante forte que vem impulsionando as redes de supermercados são as estratégias de ampliação de vendas em outros produtos além dos alimentos, com receitas e margens diferenciadas. Nos últimos anos foi relevante o crescimento das linhas de eletrodomésticos, eletrônicos e produtos de informática.

Na busca por alternativas e serviços adicionais, as grandes redes entraram em negócios diferenciados como postos de combustíveis, farmácias, laboratórios digitais e comércio eletrônico.

A criação de marcas próprias tem sido outro fator positivo para o varejo supermercadista, pois o crescimento de mais de 30% em 2008, reflete a fidelização dos clientes em relação a produtos mais baratos que possuam diferencial de qualidade.

A tendência do segmento é privilegiar uma posição mais defensiva por meio da manutenção do mix de produtos concentrado em alimentos, além de maximizar as receitas operacionais e minimizar a exposição ao risco de dívidas diante da redução da atividade econômica.

A atitude dos participantes é focar no aumento da competitividade em preço e marcas, mix de produtos, níveis de estoque, diferenciação, segmentação e novos formatos.

3.2. A EMPRESA

Wal-Mart Stores, Inc. conhecida como Walmart desde 2008 é uma multinacional americana de varejo supermercadista. Sua sede global fica em Bentonville, Arkansas, EUA.

A rede foi fundada por Sam Walton nos Estados Unidos em 1962, com a abertura da primeira loja de descontos Walmart em Rogers, no Arkansas. Em 1972, a empresa abriu capital na Bolsa de Valores, o que acelerou sua expansão. No final dos anos 70, já contava com 276 lojas em 11 Estados.

Walmart tem mais de 10.130 lojas em 15 países diferentes, com 55 nomes diferentes. A companhia opera sob seu próprio nome nos Estados Unidos, incluíndo seus 50 estados. Ela também opera sob seu próprio nome em Porto Rico. Opera no México como Walmex, no Reino Unido como Asda, no Japão como Seiyu e como Best Price na Índia.

O Walmart chegou ao Brasil em 1995 e vem crescendo desde então com presença em todo o território nacional com o site de comércio eletrônico e presentes com lojas e clubes em 18 Estados e também no Distrito Federal, nas regiões Nordeste, Centro-Oeste, Sudeste e Sul.

Atualmente conta com cerca de 500 unidades nos diversos formatos hipermercados, supermercados, clubes de compra, lojas de atacado e de vizinhança.

Além da sede em Barueri (SP), mantém escritórios regionais em Porto Alegre (RS), Curitiba (PR), Salvador (BA) e Recife (PE).

4. O TRABALHO

4.1. DEFINIÇÕES DO OBJETO DE ESTUDO

Este capítulo busca apresentar o problema central da empresa do estudo, bem como o objetivo deste trabalho.

Nos últimos anos, o varejo supermercadista tem se tornado cada vez mais dinâmico e numa competitividade mais acirrada com a concorrência instaurada pelas empresas do segmento. Dentro deste contexto, inúmeros são os fatores que passaram a ser levados em consideração numa tentativa de conquistar a preferência dos consumidores na hora da compra. Estratégias de novos pontos de vendas, de preço mais baixo e, de enaltecimento da força da marca são alguns dos exemplos de ações que passaram a ser constantemente utilizadas para fidelizar o cliente.

Tal fato tem levado a busca continua de soluções para o entendimento de qual seria a melhor previsão de vendas, já que esta está diretamente ligada à demanda e oferta de bens de consumo que pode oferecer.

Com base nisso, o objeto principal de estudo é a estruturação de um método quantitativo para modelagem e previsão de vendas, entendendo os principais componentes que afetam tal previsão, e a partir deles modelar os dados históricos.

4.2. COLETA DE DADOS

Após levantamento e revisão bibliográfica, para o melhor entendimento das técnicas de previsão, deu-se início a etapa da coleta de informações obtidas através dos relatórios de administração no endereço eletrônico da empresa http://investors.walmartstores.com.

As informações extraídas dos relatórios de administração compreendem as vendas líquidas anuais no período de 1968 a 2012 de todas as 10.130 unidades que compõem a rede atualmente. A série temporal totaliza 45 observações que serviram de base para o desenvolvimento do modelo de previsão.

Como todo planejamento de vendas é feito anualmente, os valores extraídos do site corporativo foram agrupados nesta mesma unidade de tempo. Cabe ainda ressaltar que por consistirem informações de uso público, já que a empresa é listada na NYSE (New York Stock Exchange) desde 1972, as mesmas serão apresentadas conforme publicação nos relatórios de administração.

Utilizando a série temporal, a previsão foi analisada com o *software* Matlab® versão 7.10.0 R2010a. Os dados foram distribuídos graficamente de forma anual para uma melhor visualização e identificação de comportamentos que pudessem ser inferidos sobre as análises geradas, tais como padrões, tendências e sazonalidades que podem estar presentes.

Devido ao comportamento do varejo supermercadista, é possível prever que pelo menos a tendência estará presente. A análise gráfica preliminar fornece assim subsídios auxiliares na escolha dos modelos quantitativos a serem utilizados. Serão então aplicados os modelos quantitativos de previsão que forem julgados mais adequados para a atividade fim do trabalho.

Por fim, uma comparação dos resultados foi feita mediante análise dos valores obtidos, com base nos critérios de informação e seleção de modelos citado anteriormente, chegando a uma conclusão de qual método é o mais apropriado para previsão de vendas do Walmart.

4.3. ANÁLISE DA SÉRIE TEMPORAL E ESTUDO DO MODELO

Analisando a série temporal com o auxílio do *software* Matlab® versão 7.10.0 R2010a, vale mencionar os passos e etapas necessários que constituem o modelo proposto.

A validação parcial do método proposto neste trabalho, por sua vez, será feita através de um estudo de caso.

O modelo proposto está estruturado em seis etapas, listadas e abordadas a seguir. Estas dez etapas consistem em descrever como foi feita a obtenção das previsões através do modelo estudado:

1. Verificação dos dados da série histórica original;
2. Identificação e análise dos dados preliminares;
3. Transformação da série histórica original em logarítmica;
4. Comparação entre a série original e a parábola interpolada;
5. Comparação do modelo ARMA com a diferença da série (original menos parábola interpolada); e
6. Obtenção do modelo final de previsão e seleção de critérios.

4.3.1. Etapa 1 – Verificação dos dados da série histórica original

Esta etapa consiste em verificar a disponibilidade dos dados que permitirão a construção do modelo. Os dados devem estar dispostos de forma histórica em relação à variável em estudo, neste caso as vendas líquidas. A seguir o gráfico gerado pelo *software* Matlab® versão 7.10.0 R2010a, representa a série histórica original:

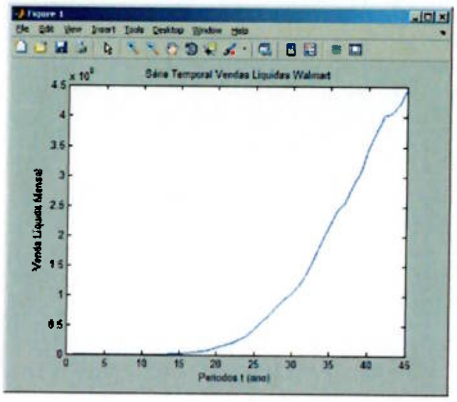

Figura . Série Temporal Vendas Liquidas Original

[Fonte: elaborado pelo autor, utilizando o software Matlab® versão 7.10.0 R2010a]

A Figura 1 registra todas as vendas líquidas e não apresenta dados espúrios, uma vez que os dados são oficiais e publicados aos acionistas da empresa.

4.3.2. Etapa 2 – Identificação e análise dos dados preliminares

Analisando o resultado da figura 1, nota-se forte tendência crescente apontando que a série é não estacionária. A curva indica que a média e variância variam em relação ao tempo, e que sua covariância depende especificamente do tempo das observações. Tal fato pode ainda ser atribuído ao resultado da abertura de novos pontos de vendas, característica marcante deste segmento.

Nesta etapa poderiam ser aplicadas as análises das funções de autocorrelações (FAC) e de autocorrelações parciais (FACP) que auxiliam na verificação da estacionariedade e, também, com base nos seus comportamentos, na proposição de um modelo. No entanto, apenas pela análise gráfica, já é possível identificar a não estacionariedade da série.

Fatores como sazonalidade e ciclicidade não podem ser notados na série dado que os períodos sazonais ocorrem no período mensal e não são percebidos nesta análise que trata os dados anualmente.

4.3.3. Etapa 3 – Transformação da série histórica original em logarítmica

Em certas circunstâncias, os dados parecem evidenciar uma relação linear estreita entre as médias e os desvios padrão da série. Evidenciado tal fato, a transformação logarítmica desses dados é a espécie de transformação que conduz à estabilização da variância.

Com essa transformação, a variância da variável transformada passa a ser constante independente da média. Se os dados são tais que a transformação garante os itens enumerados de forma satisfatória, a inferência efetuada através da análise da variância dos dados transformados é perfeitamente válida.

Para isso, o resultado da curva original transformada em logarítmica é apresentado abaixo:

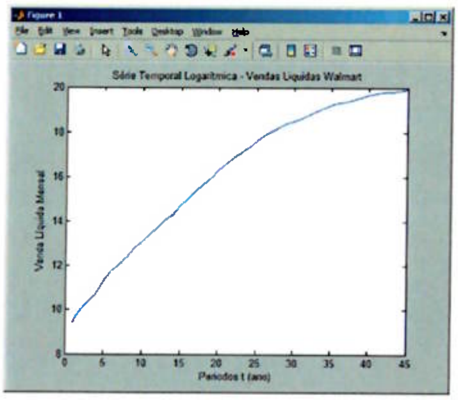

Figura . Série Temporal Vendas Líquidas Original transformada em Logarítmica

[Fonte: elaborado pelo autor, utilizando o software Matlab® versão 7.10.0 R2010a]

4.3.4. Etapa 4 – Comparação entre a série original e a parábola interpolada

Na figura 3, tem-se o resultado da interpolação da parábola. Isto ocorre quando não se pode calcular rapidamente a função nos pontos intermediários desejados.

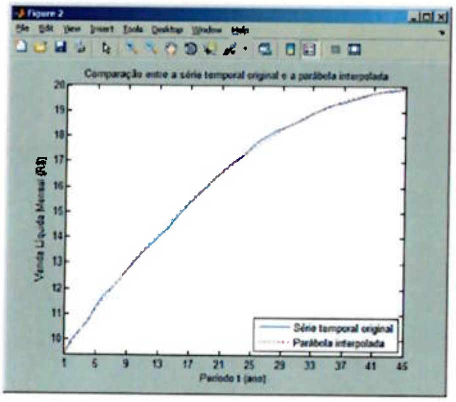

Figura . Comparação entre a série original e a parábola interpolada

[Fonte: elaborado pelo autor, utilizando o software Matlab® versão 7.10.0 R2010a]

A seguir, na figura 4, podemos notar a as diferenças entre a série original e a parábola interpolada. A diferença entre elas será modelada através de um processo ARMA:

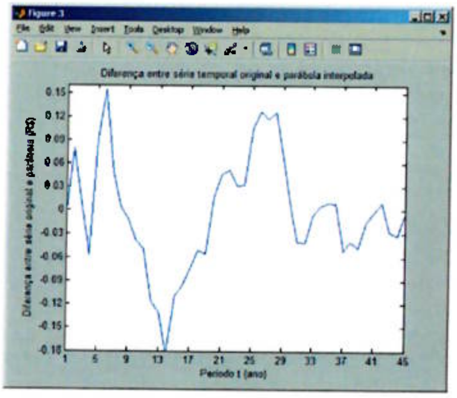

Figura . Diferença entre série temporal original e parábola interpolada

[Fonte: elaborado pelo autor, utilizando o software Matlab® versão 7.10.0 R2010a]

4.3.5. Etapa 5 – Comparação do modelo ARMA com a diferença da série (original menos parábola interpolada)

Nesta etapa, realiza-se o teste entre o modelo ARMA e as diferenças entre a série original e a parábola interpolada. Seguindo as tendências, o modelo é razoável conforme mostrado na figura 5:

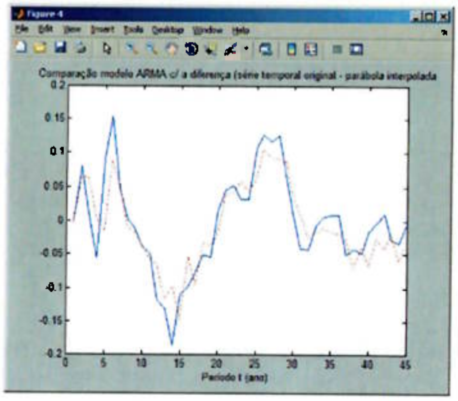

Figura . Comparação modelo ARMA com a diferença da série (original - parábola interpolada)

[Fonte: elaborado pelo autor, utilizando o software Matlab® versão 7.10.0 R2010a]

4.3.6. Etapa 6 – Obtenção do modelo final de previsão e seleção de critérios

As previsões foram realizadas *ex-post* onde são realizadas gerando dentro do período amostral. Quanto melhor forem essas últimas, mais eficiente será o modelo estimado.

O modelo que melhor representa a previsão de vendas para o varejo supermercadista é o ARMA (3,5).

O modelo ARMA (3,5) está baseado somente no passado da sua própria variável para fins de previsões, ou seja, não está baseado em nenhuma teoria econômica, portanto, seus coeficientes não são interpretados. Assim, examina-se a plausibilidade do modelo como um todo se este descreve os dados bem e se produz boas previsões.

A figura 6 compara o modelo ARMA com a série original do Walmart:

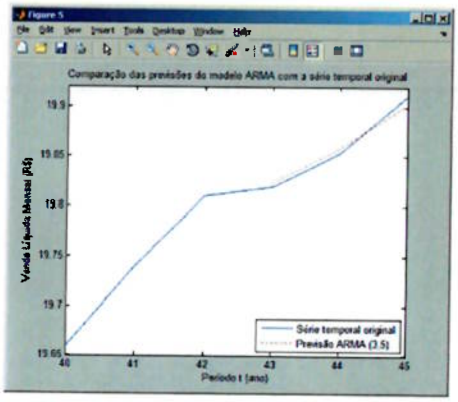

Figura . Comparação das previsões do modelo ARMA com a série temporal original

[Fonte: elaborado pelo autor, utilizando o software Matlab® versão 7.10.0 R2010a]

Através dos critérios de informação e seleção de modelos, os critérios “p” e “q” que melhor apresentaram resultados para o FPE foram p = 3 e q = 5 conforme abaixo:

	FPE	AIC	MSE	FIT
ARMA (3,4)	0,0011	-6,7327	0,000814	60,24%
ARMA (3,5)	0,0008	-6,9661	0,001278	50,18%
ARMA (4,2)	0,0011	-6,7901	0,001060	54,64%
ARMA (4,3)	0,0010	-6,9254	0,001190	51,91%
ARMA (4,5)	0,0006	-7,2799	0,001166	52,40%

Neste caso, a equação que melhor simula o modelo estudado é apresentada a seguir:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} + \theta_5 \varepsilon_{t-5}$$

$$Y_t = 2{,}104\ Y_{t-1} - 1{,}382\ Y_{t-2} + 0{,}2276\ Y_{t-3} - 1{,}348\ \varepsilon_{t-1} + 0{,}1308\ \varepsilon_{t-2} - 0{,}4685\ \theta_3 \varepsilon_{t-3} + 1{,}365\ \varepsilon_{t-4} - 0{,}6063\ \varepsilon_{t-5}$$

5. CONSIDERAÇÕES FINAIS

5.1. CONCLUSÕES

Realizar previsões de vendas é uma atividade fundamental a toda empresa, pois revelam as tendências de mercado, estratégias de crescimento e posicionamento do *market share* dentro do segmento. As previsões auxiliam ainda nas estratégias da área comercial e objetiva entender o quanto deverá vender para gerar o resultado esperado para cada um dos momentos futuros.

Existem diversas técnicas de previsão de vendas. Algumas empresas utilizam-se de técnicas ultrapassadas como análise do histórico e projeções de crescimento baseado em cenários qualitativos ao invés de quantitativos, o que pode gerar no futuro desvios entre as análises de resultado real *versus* orçamento.

Este trabalho apresentou uma modelagem para estimar previsões de vendas com base em dados históricos. O modelo proposto é o uso da metodologia ARMA (Auto Regressivo de Médias Móveis), que auxiliou entender os dados qualitativos do passado e proporcionou informações sobre seu comportamento futuro provável.

Diante disso, as conclusões do trabalho são:

- um modelo quantitativo pode ter uma forte influência no planejamento de vendas e na tomada de decisões estratégicas já que não se utiliza apenas dos fatores qualitativos no processo decisório;

- com o modelo de previsão de vendas é possível inferir na questão da estratégia e plano de expansão de lojas, dada que a tendência pode ser observada no passado e nas possíveis previsões futuras; e

- baseado em fatos do passado, podemos mitigar os impactos financeiros que o varejo supermercadista pode vir a ter, dado que com um modelo quantitativo é possível entender os comportamentos da tendência, ciclicidade e sazonalidades dos dados históricos;

5.2. PESQUISAS FUTURAS

O modelo testado e descrito neste trabalho apresenta uma maneira de obter uma previsão de vendas mais acurada. As informações foram testadas em Matlab® e analisadas em planilha de Microsoft Excel.

Neste trabalho foi testada a modelagem ARMA, mas existem muitas outras metodologias que podem ser mais robustas em outros casos. Fica em aberto a possibilidade de continuação e exploração dessas modelagens com o uso de outras técnicas de previsões.

O trabalho testou a modelagem sobre os resultados consolidados da companhia. Por fim, propõe-se utilizar a modelagem considerando os vários formatos de lojas que a companhia atua (hipermercados, supermercados e atacado) a fim de entender como se comportam cada um em separado.

REFERÊNCIA BIBLIOGRÁFICA

BARROS, A. J. S. e LEHFELD, N. A. S. **Fundamentos de Metodologia: Um Guia para a Iniciação Científica.** 2ª Ed. São Paulo: Makron Books, 2000.

BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. **Time series analysis: forecasting and control**. 3rd ed. New York: Prentice Hall, 1994.

BUENO, RODRIGO DE LOSSO DA SILVEIRA. **Econometria de Séries Temporais.** São Paulo: Cengage Learning, 2011.

FISCHER, SÉRGIO. **Séries Univariantes de tempo – Metodologia de Box & Jenkins**. Porto Alegre, Fundação Economia e Estatística, 1982.

GARCIA, CLAUDIO. **Apostila EGF-004 Modelos de Previsão de Séries Temporais do Curso de MBA em Engenharia Financeira do Programa de Educação Continuada da Escola Politécnica da Universidade de São Paulo**, 2011.

HAMILTON, JAMES D. **Time Series Analysis**. New Jersey: Princeton University Press, 1994.

HARVEY, ANDREW C. & SHEPARD, NEIL. **Structural time series models.** Handbook of Statistics, Vol. 11. Elsevier, 1993.

MAKRIDAKIS, S.; WHEELWRIGHT, S.C.; HYNDMAN, R.J. **Forecasting: methods and applications.** 3rd edition. New York: John Wiley & Sons, 1998.

MORETTIN, P.A & TOLOI, C.M de C. **Previsão de Séries Temporais**. Atual Editora, 2ª edição, São Paulo, 1987.

SHUMWAY, ROBERT H. & STOFFER, DAVID S. **Time Series Analysis and its applications.** Springer; 3rd edition. New York, 2011.

TSAY, RUEY. **Analysis of Financial Times.** John Wiley & Sons Inc. Hoboken, New Jersey. 2nd edition. 2005

WERNER, LIANE. **Um modelo composto para realizar previsão de demanda através da integração da combinação de previsões e do ajuste baseado na opinião.** 2004. 166p. Tese (Doutorado) – Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

ZELLNER, ARNOLD & PALM, FRANK C. **The structural econometric time series analysis approach.** Cambrigde University Press, UK.

ANEXO A – CÓDIGO MATLAB

```
% Estudo da série temporal das vendas líquidas do Walmart Corp. no período % de 1968 a 2012

clear all; close all; clc

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Teste de estacionariedade da série temporal.
load Walmart.txt
T = length(Walmart);
figure(1); plot(1:T,log(Walmart))
axis([1 T 9.4 20])
set(gca,'XTick',1:4:T,'YTick',10:1:20);
xlabel('Período t (ano)')
ylabel('Venda Líquida Mensal (R$)')
title ('Série temporal logarítmica das vendas líquidas da Walmart Corp.')

% Observa-se pela curva de vendas líquidas que se trata de uma série não-estacionária com tendência temporal. Essa tendência mostra a evolução das vendas líquidas baseadas no crescimento da companhia.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
t=(1:T)';     % Gera vetor de tempos no formato de coluna.
coef = polyfit(t,log(Walmart(1:T)),2); % Ajuste polinomial por uma parábola
parabola = coef(1)*(1:T).*(1:T)+coef(2)*(1:T)+log(Walmart(1))-coef(1)-coef(2);
figure(2); plot(t,log(Walmart(1:T)),t,parabola,'r:')
axis([1 T 9.4 20])
set(gca,'XTick',1:4:T,'YTick',10:1:20);
xlabel('Período t (ano)')
ylabel('Venda Líquida Mensal (R$)')
title ('Comparação entre a série temporal original e a parábola interpolada')
legend ('Série temporal original','Parábola interpolada',4)

% Observa-se uma boa aderência da parábola ajustada à curva original.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Cálculo da diferença entre a série temporal original e a parábola interpolada.
difer = log(Walmart(1:T)) - parabola';
figure(3); plot(t,difer)
axis([1 T -0.18 0.16])
set(gca,'XTick',1:4:T,'YTick',-0.18:0.03:0.150);
xlabel('Período t (ano)')
ylabel('Diferença entre série original e parábola (R$)')
title ('Diferença entre série temporal original e parábola interpolada')
% A diferença entre elas será modelada como um processo ARMA.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
% Geração de modelo ARMA para a diferença entre a série temporal original e a parábola interpolada.
mi = mean(difer);
dados = iddata(difer - mi); % Dados da série temporal menos a média
p = 3; q = 5; % Definindo o modelo ARMA de ordem (2,3)
mod_ARMA = armax(dados, [p q]);
AIC_ARMA = aic(mod_ARMA); % Calcula o critério AIC (Akaike) para validação do modelo ARMA
FPE_ARMA = fpe(mod_ARMA); % Calcula o critério FPE (Akaike) para validação do modelo ARMA

% Através da análise destes critérios pode-se selecionar qual a melhor ordem para o modelo obtido.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Comparação da saída do modelo ARMA(2,3) com a diferença entre a série temporal original e a parábola interpolada.
y_sim = predict(mod_ARMA,dados,1);
figure(4); plot(1:T,difer,1:T,y_sim.outputdata+mi,'r:')
%axis([1 T -2.04 1.23])
%set(gca,'XTick',1:4:T-2,'YTick',-2:0.25:1);
title ('Comparação modelo ARMA c/ a diferença (série temporal original - parábola interpolada')
xlabel('Período t (ano)')

% Verifica-se que o modelo testado é razoável, seguindo as tendências dos dados originais.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Previsão 1 a 3 períodos à frente com modelo ARMA obtido
HP = 3;        % Horizonte de predição da série temporal
prev_difer = predict(mod_ARMA,dados,HP);
prev = prev_difer.outputdata(T-HP+1:T)' + mi + parabola(T-HP+1:T)
figure(5); plot(T-HP-2:T,log(Walmart(T-HP-2:T)),T-HP+1:T,prev(1:3),'r:')
axis([T-HP-2 T 19.65 19.92])
set(gca,'XTick',T-HP-2:1:T);
xlabel('Período t (ano)')
ylabel('Venda Líquida Mensal (R$)')
title ('Comparação das previsões do modelo ARMA com a série temporal original')
legend ('Série temporal original','Previsão ARMA (2,3)',4)
```

ANEXO B – AVALIAÇÃO E SELEÇÃO DE CRITÉRIOS

Abaixo os resultados da análise dos critérioes "p" e "q", para decisão do melhor modelo que apresentasse o menor critério FPE (Final Prediction Error):

	FPE	AIC	MSE	FIT
ARMA (1,0)	0,0022	-6,1246	0,002041	37,03%
ARMA (1,1)	0,0022	-6,1284	0,001892	39,38%
ARMA (1,2)	0,0023	-6,0679	0,001864	39,83%
ARMA (1,3)	0,0020	-6,1897	0,001764	41,47%
ARMA (1,4)	0,0015	-6,4603	0,001511	45,83%
ARMA (1,5)	0,0012	-6,7266	0,001186	52,01%
ARMA (2,0)	0,0022	-6,0950	0,001956	38,36%
ARMA (2,1)	0,0023	-6,0676	0,001865	39,81%
ARMA (2,2)	0,0024	-6,0046	0,001837	40,28%
ARMA (2,3)	0,0017	-6,3427	0,001364	48,54%
ARMA (2,4)	0,0015	-6,4446	0,001176	52,21%
ARMA (2,5)	0,0010	-6,8419	0,000994	56,07%
ARMA (3,0)	0,0020	-6,2235	0,001640	43,56%
ARMA (3,1)	0,0018	-6,2766	0,001639	43,59%
ARMA (3,2)	0,0027	-5,8822	0,001840	40,22%
ARMA (3,3)	0,0026	-5,9140	0,001642	43,52%
ARMA (3,4)	0,0011	-6,7327	0,000814	60,24%
ARMA (3,5)	0,0008	-6,9661	0,001278	50,18%
ARMA (4,0)	0,0020	-6,1999	0,001556	45,03%
ARMA (4,1)	0,0020	-6,2087	0,001492	46,17%
ARMA (4,2)	0,0011	-6,7901	0,001060	54,64%
ARMA (4,3)	0,0010	-6,9254	0,001190	51,91%
ARMA (4,4)	0,0014	-6,4906	0,001117	53,41%
ARMA (4,5)	0,0006	-7,2799	0,001166	52,40%
ARMA (5,0)	0,0015	-6,4893	0,001076	54,29%
ARMA (5,1)	0,0015	-6,4595	0,001054	54,75%
ARMA (5,2)	0,0013	-6,5963	0,000882	58,61%
ARMA (5,3)	0,0012	-6,6650	0,000920	57,73%
ARMA (5,4)	0,0010	-6,8738	0,000651	64,44%
ARMA (5,5)	0,0007	-7,1619	0,001128	53,19%