

**UNIVERSIDADE DE CAXIAS DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO – PPGA
CURSO DE MESTRADO**

ANDRÉ MAURO SANTOS DE ESPÍNDOLA

**INTELIGÊNCIA COMPETITIVA E MODELOS DE
SÉRIES TEMPORAIS PARA PREVISÃO DE CONSUMO:
O ESTUDO DE UMA EMPRESA DO SETOR METALÚRGICO**

**CAXIAS DO SUL
2013**

ANDRÉ MAURO SANTOS DE ESPÍNDOLA

**INTELIGÊNCIA COMPETITIVA E MODELOS DE
SÉRIES TEMPORAIS PARA PREVISÃO DE CONSUMO:
O ESTUDO DE UMA EMPRESA DO SETOR METALÚRGICO**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação do Mestrado em Administração da Universidade de Caxias do Sul, como requisito parcial à obtenção do grau de Mestre em Administração.

Área de Concentração: Administração da Produção

Orientadora: Prof^a. Dr^a. Maria Emilia Camargo

Co-Orientadora: Prof^a. Dr^a. Ana Cristina Fachinelli

CAXIAS DO SUL
2013


**“Inteligência Competitiva e Modelos de Séries Temporais para a Previsão de Consumo:
o Caso de uma Empresa do Setor Metalúrgico”**

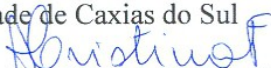
André Mauro Santos de Espíndola

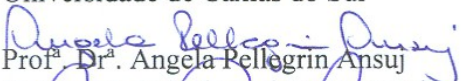
Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Administração da Universidade de Caxias do Sul, como parte dos requisitos necessários para a obtenção do título de Mestre em Administração, Área de Concentração: Administração da Produção


Caxias do Sul, 30 de agosto de 2013.

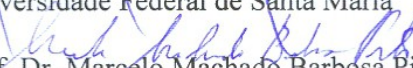
Banca Examinadora:

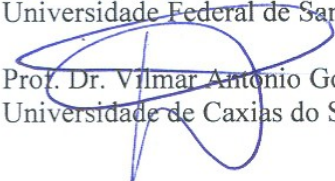

Prof.^a Dr.^a Maria Emilia Camargo (Orientadora)
Universidade de Caxias do Sul


Prof.^a Dr.^a Ana Cristina Fachinelli (Co-Orientadora)
Universidade de Caxias do Sul


Prof.^a Dr.^a Angela Pellegrin Ansuj
Universidade Federal de Santa Maria


Prof. Dr. Luis Felipe Dias Lopes
Universidade Federal de Santa Maria


Prof. Dr. Marcelo Machado Barbosa Pinto
Universidade Federal de Santa Maria


Prof. Dr. Vilmar Antonio Gonçalves Tondolo
Universidade de Caxias do Sul

Dados Internacionais de Catalogação na Publicação (CIP)
Universidade de Caxias do Sul
UCS - BICE - Processamento Técnico

E77i Espindola, André Mauro Santos de
Inteligência competitiva e modelos de séries temporais para previsão
de consumo: o estudo de uma empresa do setor metalúrgico / André Mauro
Santos de Espindola. – 2013.

134 f. : il. ; 30 cm

Dissertação (Mestrado) – Universidade de Caxias do Sul, Programa de
Pós-Graduação em Administração, 2013.

“Orientação: Profª. Drª. Maria Emilia Camargo”

1. Inteligência competitiva (Administração). 2. Big data. 3. Tecnologia
da informação. 4. Empresa metalúrgica – Rio Grande do Sul – Estudo de
caso. I. Título.

CDU 2.ed. : 005.343

Índice para o catálogo sistemático:

1. Inteligência competitiva (Administração)	005.343
2. Big data	005.311
3. Tecnologia da informação	005.51:004
4. Empresa metalúrgica – Rio Grande do Sul – Estudo de caso	658.11(816.5)

Catalogação na fonte elaborada pela bibliotecária
Márcia Servi Gonçalves – CRB 10/1500

AGRADECIMENTOS

A presente dissertação é o resultado de muito trabalho e dedicação, mas a conclusão da mesma só foi possível graças ao apoio de pessoas às quais eu desejo destacar os meus sinceros agradecimentos.

À minha companheira Natacha por todo o apoio, compreensão e carinho dedicado durante a realização deste estudo.

À minha família e amigos por compreender a minha ausência e apoiar o meu trabalho.

À minha orientadora professora Maria Emília Camargo, por todo o apoio, incentivo e dedicação, mas principalmente por permitir que eu fosse orientado por ela que é uma grande pesquisadora.

À minha co-orientadora professora Ana Cristina Fachinelli, por todo apoio e dedicação, mas principalmente por ter lançado o desafio que originou este trabalho.

Aos colegas de mestrado que estiveram sempre ao meu lado na realização de seminários, trabalhos e artigos, sempre apoiando e motivando o desenvolvimento do meu trabalho.

À direção, coordenação e colegas do CETEC, pelo apoio e compreensão.

Aos diretores da empresa objeto deste estudo, que compreenderam os objetivos da pesquisa, disponibilizando os dados e as informações necessárias para realização dos objetivos propostos.

*"Jamais considere seus estudos
como uma obrigação, mas como
uma oportunidade invejável para
aprender a conhecer a influência
libertadora da beleza do reino do
espírito, para seu próprio prazer
pessoal e para proveito da
comunidade à qual seu futuro
trabalho pertencer."*

Albert Einstein

RESUMO

O mundo vive um contínuo e acelerado processo de transformação que envolve todas as áreas do conhecimento. É possível afirmar que a velocidade desse processo tem uma relação direta com a rapidez em que ocorrem as mudanças na área tecnológica. Estas mudanças têm tornado cada vez mais as relações globalizadas, modificando as transações comerciais e fazendo com que as empresas repensem as formas de competir. Nesse contexto, o conhecimento assume, a partir do volume de dados e informações, um papel de novo insumo, muitas vezes com maior importância que o trabalho, capital e a terra. Essas mudanças e a importância da informação fazem com que as empresas busquem um novo posicionamento, procurando identificar no ambiente externo sinais que possam indicar eventos futuros. O grande desafio das empresas passa pela obtenção de dados, extração da informação e transformação dessa em conhecimento útil para a tomada de decisão. Nessa conjuntura este estudo teve como objetivo identificar qual o modelo de previsão de consumo para análise das informações no processo de Inteligência Competitiva em uma empresa do setor metalúrgico localizada no estado do Rio Grande do Sul. No desenvolvimento do estudo foram utilizados os temas *Big Data*, *Data Mining*, Previsão de Demanda e Inteligência Competitiva com a finalidade de responder à seguinte questão: Qual o modelo de previsão de consumo de aço que pode ser usado para análise das informações no processo de Inteligência Competitiva? Na realização do estudo foram analisados dados internos e externos a empresa na busca pela identificação de correlação entre o consumo de aço da empresa e variáveis econômicas que posteriormente foram utilizadas na identificação do modelo de previsão de consumo. Foram identificados dois modelos, um univariado sem intervenção através da metodologia de Box e Jenkins, o segundo modelo foi um modelo de previsão com Função de Transferência. Os dois modelos apresentaram uma boa capacidade de descrever a série histórica do consumo de aço, mas o modelo univariado apresentou melhores resultados na capacidade de previsão.

Palavras-chave: Aço, Box e Jenkins, *Big Data*, *Data Mining*, Previsão de Demanda, Inteligência Competitiva.

ABSTRACT

The world has been in a continuous and rapid process of transformation which involves all the areas of knowledge. It is possible to assert that the speed of this process has a direct relationship with the fast changes in the technological area. These changes have influenced the global relationships even more; modifying the commercial trades and making companies rethink their competitive actions. In this field, knowledge takes on a new role giving more importance to the amount of data and information to the detriment of land, labor and capital. These changes and the importance given to information make companies establish new positions in order to identify signs that anticipate events. Obtaining, extracting and transforming information into useful knowledge to help in the final decision is a challenge. Thus the purpose of this study is determine a model of consumption anticipation to analyze the process of competitive intelligence in a Metallurgy Company located in the state of Rio Grande do Sul. To develop the study the themes Big Data, Data Mining, Demand Prediction and Competitive Intelligence were used aiming to answer the question: Which model to anticipate consumption for iron can be used to analyze information in the process of competitive intelligence? For the study, internal and external data were analyzed to identify the relation between the company iron consumption and the economic variables, which were used in the demand anticipation afterwards. Two models were identified, being one of them univariate and having no intervention through Box and Jenkins methodology. The second model had a transfer function. Both of them demonstrated good capability in describing historical series of iron consumption, however the univariate model has demonstrated better results in the capability of anticipation.

Keywords: Steel, Box & Jenkins, Big Data, Data Mining, Demand Forecasting, Competitive Intelligence.

LISTA DE FIGURAS

Figura 1 – Representação gráfica do modelo teórico	10
Figura 2 – O Ciclo Big Data - Adaptação de FEINLEIB, 2012.....	15
Figura 3 – Fluxograma do ciclo iterativo de Box e Jenkins.....	37
Figura 4 – Etapas da Pesquisa	69
Figura 5 – Relação entre o Consumo de Aço e o Mercado.	88

LISTA DE QUADROS

Quadro 1 - Diferença entre dados, informação e conhecimento.	53
---	----

LISTA DE GRÁFICOS

Gráfico 1 – Sobrecarga: Informação Global criada e capacidade de armazenamento em Exabytes – 2005 / 2011	17
Gráfico 2 – Consumo mensal de aço – Empresa Alfa Brasil – jan. 07 / maio 2012	72
Gráfico 3 – Consumo mensal de aço – Empresa Alfa – Brasil – jan. 07 / maio 2012	91
Gráfico 4 – Histograma de distribuição de frequências Consumo mensal de aço Empresa Alfa Brasil – jan. 07 / maio 2012.....	91
Gráfico 5 – Função de Autocorrelação do Consumo mensal de aço Empresa Alfa – Brasil – jan. 07/ - maio/12.....	93
Gráfico 6 – Função de Autocorrelação Parcial do Consumo mensal de aço Empresa Alfa – Brasil – jan. 07/ - maio/12	93
Gráfico 7 – Função de Autocorrelação dos resíduos do modelo 2.....	98
Gráfico 8 – Função de Autocorrelação Parcial dos resíduos do modelo 2.....	99
Gráfico 9 – Função de Autocorrelação dos resíduos do modelo 8.....	100
Gráfico 10 – Função de Autocorrelação Parcial dos resíduos do modelo 8.....	100
Gráfico 11 – Função de Autocorrelação dos resíduos do modelo ARIMA(1, 1, 0)- Índice de Indicadores da Produção Extrativa Mineral	103
Gráfico 12 – Função de Autocorrelação dos resíduos do modelo ARIMA(2, 0, 0) - Produção de Retroescavadeiras	105
Gráfico 13 – Função de Autocorrelação dos resíduos do modelo ARIMA(1, 1, 1) Índice Indicadores da Produção da Indústria de Transformação.....	107
Gráfico 14 – Função de Autocorrelação dos resíduos do modelo SARIMA(2, 0, 1)(1, 0, 0) ₁₂ Índice de Material de Construção	109
Gráfico 15 – Função de Correlação dos resíduos da Função de Transferência.....	113

LISTA DE TABELAS

Tabela 1 – Trabalhos publicados por Base – 2002/2012.....	57
Tabela 2 – Ocorrência dos assuntos – 2002/2012.	58
Tabela 3 – Ocorrência dos assuntos por Base – 2002/2012.	58
Tabela 4 – Ocorrência dos assuntos na BDTD – 2002/2012.....	62
Tabela 5 – Produção Mundial de Aço Bruto – Unid. 10^6 t.....	75
Tabela 6 – Produção Brasileira de Aço Bruto por Empresa - Unid. 10^3 t.....	75
Tabela 7 – Consumo Aparente Per Capita de Aço – Unid. Kg/hab.	76
Tabela 8 – Distribuição Setorial das Vendas Internas de Produtos Siderúrgicos - Unid. 10^3 t.....	76
Tabela 9 – Estrutura da Base de Dados Relação de Clientes	78
Tabela 10 – Estrutura da Base de Dados Consumo de Aço	78
Tabela 11 – Classificação dos clientes por seção conforme a tabela CNAE.	79
Tabela 12 – Classificação dos clientes por divisão conforme tabela CNAE.....	80
Tabela 13 – Identificação dos Setores Econômicos e Indicadores.....	82
Tabela 14 – Avaliação da Correlação.....	83
Tabela 15 – Teste de Kolmogorov Smirnov.....	84
Tabela 16 – Análise de Correlação entre a variável consumo de aço e as variáveis externas através do Coeficiente de Pearson.	85
Tabela 17 – Classificação das variáveis conforme similaridade e sua correlação.	87
Tabela 18 – Avaliação de Multicolinearidade.....	89
Tabela 19 – Estatísticas descritivas da variável consumo	92
Tabela 20 – Valores de d_λ para variável consumo.....	92
Tabela 21 – Resultados do teste ADF para variável consumo	94
Tabela 22 – Hipóteses de Modelos para variável consumo de aço.	95
Tabela 23 – Modelos estimados e suas estatísticas para variável consumo de aço.....	95

Tabela 24 – Estatísticas de avaliação dos modelos.	100
Tabela 25 – Previsões mensais do consumo de aço – jun./12 – abr./13.....	101
Tabela 26 – Identificação do Modelo Univariado	102
Tabela 27 – Estatísticas do Modelo.....	102
Tabela 28 – Previsões mensais do modelo ARIMA(1, 1,0) Índice de Indicadores da Produção Extrativa Mineral – jun./12 – abr./13	104
Tabela 29 – Identificação do Modelo Univariado	104
Tabela 30 – Estatísticas do Modelo ARIMA (2, 0, 0) - Produção de Retroescavadeiras.....	104
Tabela 31 – Previsões mensais do modelo ARIMA(2, 0, 0) Produção de Retroescavadeiras – jun./12 – abr./13.....	106
Tabela 32 – Identificação do Modelo Univariado	106
Tabela 33 – Estatísticas do Modelo ARIMA(1, 1, 1) - Índice Indicadores da Produção da Indústria de Transformação	106
Tabela 34 – Previsões mensais do ARIMA(1, 1, 1) - Índice Indicadores da Produção da Indústria de Transformação – jun./12 – abr./13.....	108
Tabela 35 – Identificação do Modelo Univariado	108
Tabela 36 – Estatísticas do Modelo SARIMA(2, 0, 1)(1, 0, 0) ₁₂ Índice de Material de Construção.....	108
Tabela 37 – Previsões mensais do SARIMA(2, 0, 1)(1, 0, 0) ₁₂ Índice de Material de Construção – jun./12 – abr./13.....	110
Tabela 38 – Identificação do Modelo Univariado	110
Tabela 39 – Estatísticas do Modelo ARIMA(2, 0, 1) - Consumo de Aço.....	110
Tabela 40 – Previsões mensais do modelo ARIMA(2, 0, 1) Consumo de Aço – jun./12 – abr./13.....	111
Tabela 41 – Identificação do Modelo de Função de Transferência.....	111
Tabela 42 – Estatísticas do Modelo de Função de Transferência para variável Consumo de Aço	112
Tabela 43 – Estatísticas da análise de correlação dos resíduos Modelo de Função de Transferência	112
Tabela 44 – Análise das correlações cruzadas dos resíduos.....	114
Tabela 45 – Previsões mensais do modelo de Função de Transferência Consumo de Aço – jun./12 – abr./2013.....	115

LISTA DE ABREVIATURAS E SIGLAS

ACF	Função de Autocorrelação
ADF	Teste aumentado de Dickey-Fuller
AIC	Critério de Informação de Akaike
AR	Autorregressivo
ARIMA	Modelo Autorregressivo Integrado de Médias Móveis
ARMA	Autorregressivo de Médias Móveis
BACEN	Banco Central do Brasil
BD	<i>Big Data</i>
BIC	Critério Bayesiano
CNAE	Classificação Nacional de Atividades Econômicas
CNPJ	Cadastro Nacional de Pessoa Jurídica
CRISP	<i>Cross Industry Standar Process for Data Mining</i>
DM	<i>Data Mining</i>
IABR	Instituto Aço Brasil
IC	Inteligência Competitiva
IDC	<i>International Data Corporation</i>
MA	Médias Móveis
MAPE	Erro Percentual Médio Absoluto
MDF	<i>Medium Density Fiberboard</i>
MME	Ministério de Minas e Energia
MQO	Mínimos Quadrados Ordinários
PACF	Função de Autocorrelação Parcial
PIB	Produto Interno Bruto
PUC	Pontifícia Universidade Católica

S.A.	Sociedade Anônima
SARIMA	Autorregressivo Integrado de Médias Móveis Sazonal
SGS	Sistema Gerenciador de Séries Temporais
VIF	Fator de Inflação de Variância

SUMÁRIO

1	INTRODUÇÃO	6
1.1	TEMA.....	7
1.2	PROBLEMATIZAÇÃO	7
1.3	OBJETIVOS.....	11
1.3.1	Objetivo geral.....	11
1.3.2	Objetivos específicos.....	11
1.4	JUSTIFICATIVA	12
2	REFERENCIAL TEÓRICO	14
2.1	BIG DATA	14
2.2	DATA MINING	18
2.3	PREVISÃO DE DEMANDA.....	19
2.3.1	Séries Temporais	22
2.3.2	Modelos univariados de Box e Jenkins.....	24
2.3.3	Modelos univariados de Box e Jenkins sem intervenção	27
2.3.3.1	Modelo Geral.....	27
2.3.3.2	Modelo ARIMA(p,d,q).....	29
2.3.3.3	Modelo Autorregressivo de ordem p – AR(p).....	30
2.3.3.4	Modelo de Média Móvel de ordem q – MA(q)	31
2.3.3.5	Modelos Mistos Autorregressivos de Média Móvel – ARMA(p,q).....	32
2.3.3.6	Modelos univariados de Box e Jenkins com intervenção.....	33
2.3.3.7	Etapas para a construção de um modelo de previsão	36
2.3.3.8	Modelo de Função de transferência.....	40

2.3.3.9	Modelos de Função de Transferência de Box e Jenkins com Múltiplas Entradas	41
2.3.3.10	Modelos de Função de Transferência Sazonal de Box e Jenkins	42
2.3.3.11	Identificação dos Modelos de Função de Transferência com Múltiplas Entradas - Procedimento de Liu e Hanssens	43
2.3.3.12	Função de Correlação Cruzada	46
2.3.3.13	Estimação dos Parâmetros dos Modelos de Função de Transferência	46
2.3.3.14	Verificação do Modelo de Função de Transferência	47
2.3.3.15	Testes de Adequação	49
2.3.3.16	Previsão com Erro Quadrático Mínimo	50
2.4	INTELIGÊNCIA COMPETITIVA	51
2.5	ESTUDOS EMPÍRICOS REALIZADOS NA ÁREA DA DISSERTAÇÃO	57
3	METODOLOGIA DA PESQUISA	67
3.1	PESQUISA QUALITATIVA	68
3.2	PESQUISA QUANTITATIVA	69
3.3	ETAPAS DA PESQUISA	69
3.4	SELEÇÃO DO CASO DE ESTUDO	70
3.5	CARACTERIZAÇÃO DO OBJETO DE ESTUDO	71
3.5.1	Empresa Alfa	71
3.6	COLETA DE DADOS	73
4	ANÁLISE E DISCUSSÃO DOS RESULTADOS	74
4.1	ETAPA QUALITATIVA	74
4.1.1	Mercado do Aço Brasileiro	74
4.1.2	Coleta de informações na Empresa Alfa	77
4.2	ETAPA QUANTITATIVA	82
4.2.1	Análise de Correlação	83
4.2.2	Análise dos resultados obtidos na análise de correlação	86
4.2.3	Verificação dos pressupostos básicos para identificação do modelo de previsão...	88
4.2.4	Ciclo de Box e Jenkins para o modelo univariado	90
4.2.4.1	Identificação do modelo univariado sem intervenção para variável Consumo de Aço	90
4.2.4.2	Especificação do modelo para variável Consumo de Aço	94

4.2.4.3	Estimação dos parâmetros do modelo para variável Consumo de Aço.....	98
4.2.4.4	Verificação dos modelos identificados para variável Consumo de Aço	98
4.2.4.5	Previsão do SARIMA (0, 1, 1)(1, 0, 0) para variável consumo de aço	101
4.2.5	Ciclo de Identificação da Função de Transferência	101
4.2.5.1	Identificação do modelo univariado - Índice de Indicadores da produção Extrativa mineral	102
4.2.5.2	Análise dos resíduos e estatísticas do modelo ARIMA(1, 1, 0) - Índice de Indicadores da produção Extrativa mineral.....	102
4.2.5.3	Previsão do modelo ARIMA(1, 1, 0) - Índice de Indicadores da Produção Extrativa Mineral	103
4.2.5.4	Identificação do modelo univariado - Produção de Retroescavadeiras.....	104
4.2.5.5	Análise dos resíduos e estatísticas do modelo ARIMA (2, 0, 0) - Produção de Retroescavadeiras	104
4.2.5.6	Previsão do modelo ARIMA(2, 0, 0) - Produção de Retroescavadeiras	105
4.2.5.7	Identificação do modelo univariado - Índice Indicadores da produção da Indústria de transformação	106
4.2.5.8	Análise dos resíduos e estatísticas do Modelo ARIMA(1, 1, 1) - Índice Indicadores da Produção da Indústria de Transformação	106
4.2.5.9	Previsão do modelo Modelo ARIMA(1, 1, 1) - Índice Indicadores da Produção da Indústria de Transformação	107
4.2.5.10	Identificação do modelo univariado - Índice de Material de Construção.....	108
4.2.5.11	Análise dos resíduos e estatísticas do Modelo SARIMA(2, 0, 1)(1, 0, 0) ₁₂ - Índice de Material de Construção.....	108
4.2.5.12	Previsão do modelo SARIMA(2, 0, 1)(1, 0, 0) ₁₂ - Índice de Material de Construção	109
4.2.5.13	Identificação do modelo univariado – Consumo de Aço da empresa Alfa	110
4.2.5.14	Análise dos resíduos e estatísticas do Modelo ARIMA(2, 0, 1) - Consumo de Aço	110
4.2.5.15	Previsão do modelo ARIMA(2, 0, 1) - Consumo de Aço	111
4.2.5.16	Identificação do modelo de Função de Transferência para variável Consumo de Aço	111
4.2.5.17	Análise dos resíduos e estatísticas do Modelo de Função de Transferência para variável Consumo de Aço.....	112
4.2.5.18	Previsão do modelo de Função de Transferência para o Consumo de Aço	114

4.2.6	Análise dos resultados dos modelos de previsão.....	115
5	CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES.....	117
5.1	LIMITES DA PESQUISA	120
5.2	TRABALHO FUTUROS	120
	REFERÊNCIAS	121
	APÊNDICE A- ANÁLISE DAS CORRELAÇÕES CRUZADAS DOS RESÍDUOS....	130

1 INTRODUÇÃO

O processo de globalização tem modificado as inter-relações dinâmicas entre assuntos internacionais, nacionais, regionais e locais, de um lado; e de outro, todos os campos do interesse humano, incluindo a economia, a ciência, a tecnologia, a política, a religião, a cultura, as comunicações, o transporte, a educação, a saúde e a ecologia (HARRIS, 2002). Na economia a grande modificação está na relação entre as empresas de grande, médio e pequeno porte com o mercado que está cada vez mais exigente e competitivo.

Outro fator que tem modificado as inter-relações é o desenvolvimento da tecnologia da informação, desenvolvimento este que faz com que Drucker (1993) defina sociedade atual como “sociedade do conhecimento” onde a diferença com relação às sociedades do passado está diretamente relacionado com a importância do conhecimento. Para Drucker (1993), a importância do conhecimento é latente quando ele assume o papel do recurso mais significativo atualmente.

Stewart (1998, p. 8), afirma que a informação e o conhecimento são as principais armas competitivas desta época. O autor complementa dizendo “o conhecimento é mais valioso e poderoso do que os recursos naturais, grandes indústrias ou polpudas contas bancárias”. Para Stewart (1998) a força do conhecimento atinge todos os setores e as empresas que têm os melhores resultados são as que têm as melhores informações ou as que têm o melhor controle das informações.

Para Davenport (2007), nos dias atuais existe pouca diferenciação entre empresas dos diversos setores o que faz com que muitas das bases de competição como posição geográfica e regulamentações de mercados não estejam mais disponíveis. O autor complementa afirmando que até mesmo as novas tecnologias e as inovações tanto em produtos como serviços são rapidamente copiáveis. Segundo Davenport (2007), a competição passa pela tomada de decisões inteligentes e rápidas com base nas análises das informações pelos líderes das empresas.

Porter (2009), assegura que a competição tem uma relação de dependência com a capacidade de uma indústria inovar e atualizar-se. Mas neste contexto, como ampliar a capacidade de competir de uma indústria, como inovar sem ser copiado, como atualizar-se sem ser copiado, como sustentar a vantagem competitiva?

Os autores citados trazem a importância da informação e do conhecimento nos dias atuais como uma forma de se obter vantagem competitiva. Nesta conjuntura a busca por dados, a transformação destes em informação e o uso desta informação como conhecimento

através da inteligência competitiva passa a ter um papel fundamental no planejamento estratégico das empresas.

Janissek-Muniz, Lesca e Freitas (2007), afirmam que as rápidas e imprevisíveis evoluções mundiais nas áreas tecnológicas, sociais, comportamentais, legais e relacionais fazem com as empresas busquem uma constante capacidade de adaptação. Para os autores as empresas devem estar atentas ao ambiente externo a fim de identificar os movimentos que possam sugerir eventos futuros. O grande desafio está na capacidade de antecipar os eventos através do desenvolvimento de uma inteligência coletiva no ambiente de negócios.

Segundo Freitas, Freitas e Gensas (2007), a fonte de dados da empresa pode ser tanto o ambiente interno como o externo, a empresa deve trabalhar com os dados oriundos destes ambientes utilizando metodologias adequadas.

Neste contexto as empresas precisam saber onde obter dados, como extrair informação destes dados, como transformar esta informação em conhecimento e como utilizar este conhecimento de forma inteligente no do processo de tomada de decisões.

No sentido de contribuir com elementos de respostas para estas questões, este estudo pretende apresentar aspectos teóricos e práticos utilizados para identificação de um modelo de previsão de consumo de aço. Para tanto a pesquisa apresenta uma possível fonte de dados, os processos de extração de informação desses dados e um modelo capaz de transformar essa informação em conhecimento.

Este estudo está dividido em 5 capítulos. O capítulo 1 apresenta a problematização, os objetivos da pesquisa e a justificativa; o capítulo 2 traz o referencial teórico que fundamenta o estudo; o capítulo 3 descreve a metodologia utilizada na pesquisa, o capítulo 4 mostra os resultados e as discussões a cerca da pesquisa realizada e o capítulo 5, as considerações finais e recomendações com relação a trabalhos futuros.

1.1 TEMA

Big Data, Data Mining, Modelos de Previsão e Inteligência Competitiva aplicada na previsão de consumo.

1.2 PROBLEMATIZAÇÃO

Na busca pela vantagem competitiva em um ambiente onde as bases da competição como posição geográfica e regulamentação de mercados não estão mais disponíveis, a

competição passa pela tomada de decisões inteligentes e rápidas com base nas análises das informações pelos líderes das empresas (DAVENPORT, 2007).

Neste contexto surgem questões fundamentais para a utilização da informação no processo decisório pelos líderes das empresas:

- Onde obter os dados?
- Como extrair informação destes dados?
- Como transformar a informação extraída em conhecimento para a liderança utilizar de forma inteligente no processo decisório?

Para responder à primeira questão, é possível citar como uma grande fonte de dados pessoais, governamentais e não governamentais a *Internet*, nesta os dados podem ser encontrados de forma estruturada ou não estruturada. Além dos dados existentes na *Internet*, existem os dados armazenados em redes corporativas e em computadores pessoais. O armazenamento destes conjuntos de dados tem crescido de forma exponencial, tal fenômeno recebeu a denominação de *Big Data* (BD) (THE ECONOMIST, 2010).

Frankel (2012), define o *Big Data* como conjunto de dados estruturados e não organizados, produzidos como um subproduto das operações de uma empresa. Para o autor, esses dados incluem: informações sobre clientes, finanças, funcionários e operações.

Frankel (2012), afirma que a comunidade financeira já trabalha com este conjunto de dados há anos e que esse não é um conceito novo, mas outros setores devem começar a utilizar a informação existente no *Big Data* para tomada de decisões, assim aumentando a produtividade e reduzindo os riscos.

Segundo Johnson (2012), a economia baseada nos dados do *Big Data* se caracteriza pela quantidade e complexidade dos dados e pela velocidade de mudanças e sua influência na tomada de decisões. As organizações, ao reconhecerem a importância dos dados e a sua competitividade, estão mudando a forma como a informação é acessada, interpretada, relatada e aplicada. Segundo o autor, cada vez mais as organizações buscam vantagem competitiva e usam os dados para tomar melhores decisões aumentando a rentabilidade, melhorando a capacidade operacional e controlando os riscos. Para isto é importante que a empresa adote a análise eficiente dos dados, pois esta pode ser o grande motor do crescimento.

Tendo o *Big Data* uma fonte de dados, é possível responder à segunda questão: como usar as informações contidas no *Big Data*, como explorar os dados presentes neste fenômeno na busca de informações que sejam importantes para análise, tomada de decisão e planejamento das empresas?

Lesca, Freitas e Janissek-Muniz (2003, p. 1), ao afirmarem que existe uma mudança da lógica de armazenamento e busca de dados para a lógica de disponibilização de grande quantidades de dados, fazem o seguinte questionamento:

O problema torna-se, então, como - melhor, de forma mais adequada e mesmo mais rápida - identificar algo útil, pertinente ou de consequências para a organização, tendo em vista suas decisões, oportunidades, negócios, estratégias, enfim (LESCA, FREITAS, JANISSEK-MUNIZ, 2003, p. 1).

Uma das formas possíveis talvez seja o processo denominado de *Data Mining* (DM), que consiste na busca do conhecimento existente em banco de dados. O processo vem se desenvolvendo há algumas décadas e combinado com o fenômeno *Big Data* tem obtido mais destaque.

Marcano (2007), afirma que é crescente a necessidade de melhorar o acesso à informação e que a técnica de *Data Mining* pode ser utilizada como uma das etapas no processo de descoberta de conhecimento em bases de dados. O autor caracteriza o *Data Mining* pela velocidade para identificar, processar e extrair as informações que realmente importam, descobrindo e compreendendo os padrões em bases de dados. Para o autor, o *Data Mining* é uma técnica fácil que pode ser aplicada em qualquer área do conhecimento, ela tem como limitação o esforço de estabelecer medidas na avaliação dos resultados.

Para Cardoso e Machado (2008), coletar e armazenar dados não são suficientes para melhorar a estratégia de uma organização. É fundamental o processo de análise das grandes quantidades de dados para se determinar padrões nestes dados e relações de causa e efeito. Os autores complementam que o processamento e análise correta das informações obtidas nas grandes bases de dados são requisitos fundamentais para a tomada de decisão. Segundo o autor, a mineração de dados é uma técnica que compõe uma das etapas do processo de descoberta do conhecimento em banco de dados, capaz de revelar o conhecimento implícito em grandes quantidades de informações. O *Data Mining* pode assumir um papel muito importante no reconhecimento de padrões existentes em conjunto de dados, padrões estes que podem identificar comportamentos a fim de auxiliar o gestor a tomar decisões.

Os autores citados até aqui trazem a importância da exploração dos dados existentes no *Big Data* e o *Data Mining* como um processo capaz de realizar esta exploração. O tratamento, processamento e análise dos dados foram destacados como fundamentais, ou seja, não basta possuir informação é preciso saber utilizá-la. É nesse momento que surge a importância da Inteligência Competitiva e sua capacidade de análise da informação em busca do conhecimento para auxiliar no processo de planejamento da empresa.

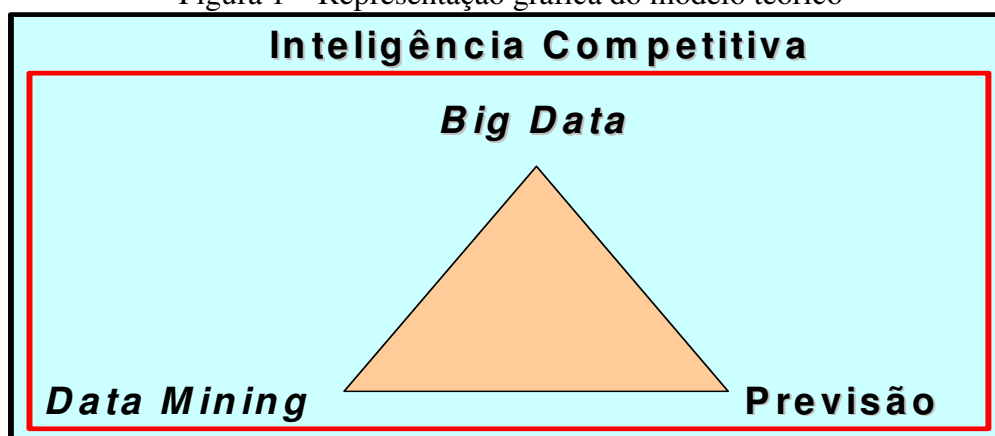
Dentro do processo de planejamento da empresa é possível citar os estudo de previsão de demanda, visto que a informação obtida no *Big Data* através de processos de *Data Mining* pode ser utilizada na construção de modelos de previsão de demanda. A previsão de demanda fornecerá informações à Inteligência Competitiva que terá a função de analisá-los antes de encaminhar para os gestores responsáveis pela tomada de decisões sobre a produção.

Segundo Moreira (1993, p. 317), a previsão de demanda é uma base comum em todo o processo de planejamento. Conhecer a quantidade de produtos que será vendida é o marco inicial para tomadas de decisões e por mais imperfeita que seja uma previsão, ela sempre será necessária.

Autores como Buffa (1972), têm a mesma visão de Moreira sobre a importância da previsão de demanda e complementam afirmando que a previsão é um fator crítico para modelos de decisão. Já Riggs (1976), destaca a importância da informação existente no mercado como um indicador antecipado para a produção.

Este estudo busca combinar o fenômeno *Big Data*, *Data Mining* e modelos de previsão como uma ferramenta de Inteligência Competitiva. Dessa combinação surgiu o modelo teórico apresentado na Figura 1, onde o *Big Data* é a fonte de dados, o *Data Mining* é responsável pela identificação e coleta e os modelos de previsão são instrumentos de análise em um processo de Inteligência Competitiva.

Figura 1 – Representação gráfica do modelo teórico



Fonte: Elaborado pelo pesquisador

O objeto de estudo foi a empresa Metalúrgica Alfa Ltda, com matriz no estado do Rio Grande do Sul. Atualmente emprega cerca de 1.300 funcionários em suas quatro unidades fabris no Rio Grande do Sul, uma unidade produtiva no Estado de Pernambuco, uma unidade produtiva no Estado de Goiás e um centro de distribuição na América Central.

A empresa tem como histórico investir em qualificação no desenvolvimento de líderes com um perfil voltado para a interpretação estratégica e foco na inteligência empresarial.

O estudo concentrou-se na unidade de sistemas de armazenagem, especializada em planejar e desenvolver modernas soluções para estocagem e movimentação de materiais que têm como principal matéria-prima o aço. Esta unidade representou 43% do faturamento do grupo no último período.

Assim, nesta conjuntura, o presente estudo buscou responder à seguinte pergunta de pesquisa:

Qual o modelo de previsão de consumo de aço que pode ser usado para análise das informações no processo de Inteligência Competitiva?

1.3 OBJETIVOS

Na busca por respostas para o problema deste estudo são apresentados o objetivo geral e os objetivos específicos.

1.3.1 Objetivo geral

Este estudo teve como objetivo geral identificar modelos de previsão de consumo de aço para análise das informações no processo de Inteligência Competitiva.

1.3.2 Objetivos específicos

Para atingir o objetivo geral deste estudo, os seguintes objetivos específicos foram definidos:

- I. explorar o fenômeno do *Big Data* e suas possibilidades para a mineração de dados;
- II. utilizar o *Data Mining* como ferramenta de análise de grande volume de informação do *Big Data*;
- III. identificar as variáveis econômicas com influência no mercado do aço;
- IV. detectar as variáveis econômicas relacionadas com os clientes da Alfa;
- V. verificar a correlação entre a consumo de aço da Alfa e o conjunto de variáveis externas;

VI. construir modelos de previsão de consumo de aço através da análise de Séries Temporais.

1.4 JUSTIFICATIVA

As consequências do processo de globalização têm alterado significativamente a sociedade atual atingindo as relações internacionais, nacionais, pessoais, econômicas, tecnológicas, sociais, (HARRIS, 2002).

Este conjunto de transformações tem afetado também a forma com que as pessoas consomem e conseqüentemente afetado as relações comerciais. Essas modificações nas relações comerciais fazem com que as empresas de grande, médio e pequeno porte busquem novas formas de se posicionar em um mercado cada vez mais competitivo. Para isto as empresas cada vez mais têm buscado novas estratégias para obter vantagem competitiva que seja sustentável.

Para Fayard (2000), estratégia é uma arte em que é necessário se combinar, no tempo e no espaço, os meios com a finalidade de obter um determinado resultado. Para combinar estes elementos, o autor afirma que é necessário um profundo conhecimento de si mesmo, do outro e do ambiente onde se está inserido. Segundo Fayard (2000), um dos elementos essenciais da estratégia é a informação, já que para ele é necessário estar atento à informação a fim de compreender as que realmente interessam e a partir delas saber onde e como agir.

Vivemos em uma época onde a importância da informação e do conhecimento se diferenciam de outras épocas, assumindo o papel de recurso mais significativo atualmente (DRUCKER, 1993).

Autores como Stewart (1998) e Davenport (2007), destacam o papel da informação e do conhecimento como as principais armas competitivas desta época, eles afirmam que a competição passa pela tomada de decisões baseada em uma análise inteligente pelos líderes das empresas.

Saber identificar fonte de dados, extrair informação destes dados e transformar esta informação em conhecimento passa a ser uma fonte de vantagem competitiva muito importante para as empresas. Johnson (2012), já afirmava isto ao escrever que as organizações cada vez mais buscam a vantagem competitiva através da utilização de dados para tomada de decisões. O autor também afirma que a utilização dos dados depende da adoção de uma análise eficiente deles, pois esta pode ser o grande motor de crescimento dos mesmos.

Em uma época em que a informação passa a ser uma fonte importante de busca da vantagem competitiva o estudo da previsão de consumo de aço da empresa Alfa se justifica por quatro motivos.

Primeiro motivo está relacionado com a relevância acadêmica e científica de possibilitar a combinação dos conhecimentos de *Data Mining*, *Big Data* e Modelos de Previsão como um processo da Inteligência Competitiva, combinação essa que não foi encontrada na pesquisas sobre os temas realizadas em bases de dados de publicações científicas.

Segundo, por buscar identificar a correlação entre o consumo de aço da empresa e os indicadores econômicos através do estudo dos dados obtidos no processo de *Data Mining* na empresa e no *Big Data*.

O terceiro motivo tem uma fundamentação social, pois a empresa objeto de estudo é uma das maiores fontes de emprego na sua cidade sede e qualquer vantagem competitiva que esta empresa venha a obter no mercado indireta ou diretamente terá influência sobre a sociedade que dela depende.

O quarto motivo tem origem na importância econômica de desenvolver o trabalho em uma empresa do setor metalúrgico, que faz aquisições periódicas de grande quantidade de aço para produção e conseqüentemente mantém uma relação direta com o mercado do aço. Mercado esse onde o Brasil ocupa a 9ª posição no ranking dos países produtores (MME, 2011, p. 35) exportando indiretamente 2,9 milhões de toneladas de aço e que até maio de 2012 apresentou um crescimento de 14,2% nas importações em relação ao mesmo período de 2011 (IABR, 2012). Segundo o MME(2012), o mercado do aço brasileiro produziu no ano de 2011 $45,1 \times 10^6$ toneladas de aço, consumiu $40,08 \times 10^6$ toneladas de aço, gerou de forma direta 260.000 empregos, teve uma participação no PIB nacional de 2,4% e apresentou um crescimento no faturamento de 17,4% com relação ao ano de 2010.

2 REFERENCIAL TEÓRICO

2.1 BIG DATA

O desenvolvimento da tecnologia da informação através de novos equipamentos, sistemas, segurança e *Internet* tem influenciado significativamente no crescimento do armazenamento e no fluxo de informações. Este tem sido um fator com influência destacada nas mudanças da sociedade e nas organizações através do mundo. Uma das mudanças mais significativas está relacionada com o armazenamento e com a divulgação de dados pessoais ou de organizações governamentais e não governamentais através da *Internet* e outras fontes eletrônicas.

O crescimento de forma exponencial do conjunto de informações armazenadas e divulgadas através de meios eletrônicos exige cada vez mais novas técnicas de gerenciamento e pode fomentar novas fontes de valor econômico, proporcionando novos *insights* para a ciência (THE ECONOMIST, 2010).

Esta explosão de informação estruturada e desestruturada existente nas redes está sendo denominada de *Big Data*. Segundo o *International Data Corporation* (IDC) (2011), este conjunto de dados é composto por informações transacionais, bases de dados, metadados, arquivos de áudio e vídeo, páginas na *web*, mensagens de *e-mail*, *feeds* de mídia social, dados de formulários, ou seja, quaisquer dados disponíveis.

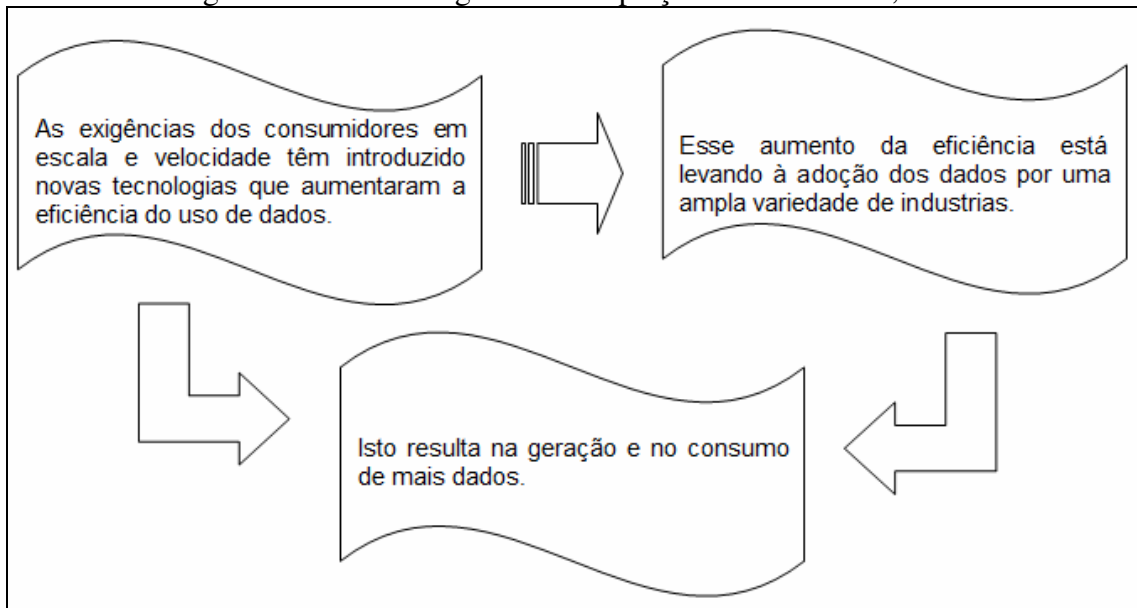
O *Big Data* é um sistema onde os participantes, de maneira consciente ou inconsciente, ofertam dados contínuos sobre si mesmos que, pelo “efeito rede” são coletados constantemente e organizados através de rápidas taxas logarítmicas (MCKINSEY, 2011).

Segundo o EMC (2011), o termo *Big Data* não é um termo exato, ele se caracteriza por um processo interminável de acúmulo de todos os tipos de dados, onde alguns destes dados são estruturados, mas a maior parte são dados desestruturados. Esse conjunto de informações não é composto somente por transações de dados, nele encontram-se vídeos, fotos, conversas, índices de pesquisas, resultados de exames médicos, estas informações são produzidas a cada instante por pessoas através de relações comerciais ou pessoais.

O sistema de organização e captura dos dados, armazenados na rede de forma virtual através de nuvem e as inovações em *software* e ferramentas para análise por meio de atividades dinâmicas em tempo real também fazem parte do *Big Data* (IDC, 2011).

Segundo a FEINLEIB (2012), o ciclo das informações do *Big Data* pode ser representado conforme a Figura 2.

Figura 2 – O Ciclo Big Data - Adaptação de FEINLEIB, 2012



Fonte: Adaptação de FEINLEIB, 2012

No ciclo apresentado pela FEINLEIB (2012), as informações armazenadas na rede estão disponíveis geralmente de forma desestruturada, isto faz com que o conceito do *Big Data* integre-as de forma a gerar um volume mais abrangente de informações. Este conjunto de informações vai possibilitar que a tomada de decisões seja, cada vez mais, baseada em fatos concretos e não apenas em amostragens ou intuição.

Esse processo trará uma mudança grande na forma como as organizações lidam com os dados. As empresas vão precisar reestruturar-se para poder tirar o maior proveito das informações existentes na tomada de decisão. Cada empresa precisará garantir instrumentos para a captura de dados e incentivar a manipulação irrestrita de dados para desvendar *insights* (IDC, 2011).

Para o EMC (2011), o uso correto das informações existentes no *Big Data* pode possibilitar *insights* para direcionar e desenvolver as iniciativas dos negócios através de um melhor entendimento dos clientes, melhoramento de cadeia de suprimentos, desenvolvimento de novos produtos e serviços.

As instituições financeiras há muito tempo têm tirado proveito da tecnologia da informação para lidar com seus clientes e gerenciar seus negócios. Com o surgimento do *Big Data* essas instituições têm aprimorado a sua capacidade de utilizar as informações presentes na rede para analisar o comportamento dos clientes e propor novos negócios. "Essa iniciativa vem ao encontro de uma das prioridades de negócios dos bancos e seguradoras: conhecer melhor seus clientes e oferecer soluções mais assertivas" (IDC, 2011).

O conjunto de informação existe no *Big Data* tem utilidade para os diversos setores da economia, não ficando restrito somente a instituições financeiras. As possibilidades que o *Big Data* traz para as organizações neste momento são incomensuráveis, mas ficam restritas ao posicionamento estratégico das empresas no contexto do fenômeno.

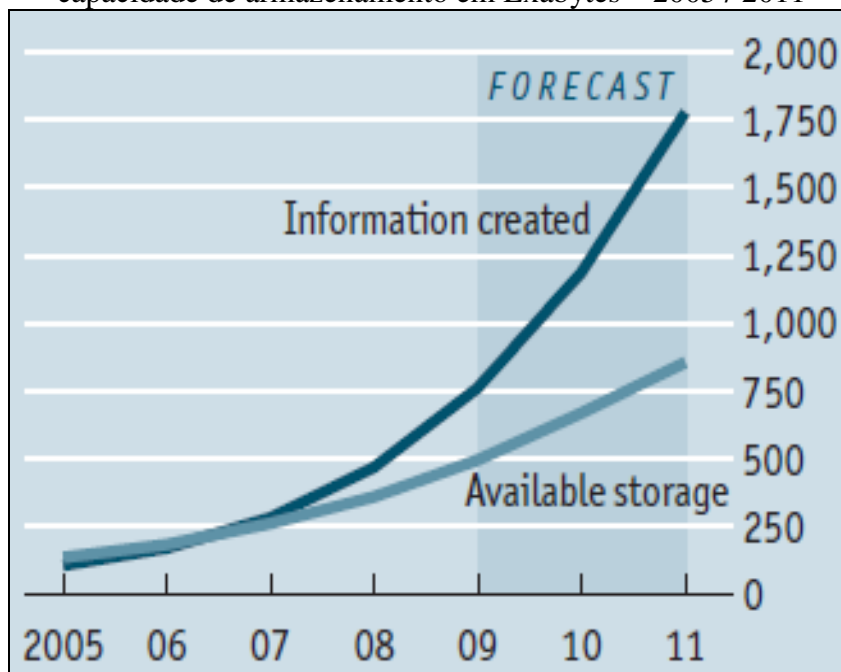
Em análises estratégicas as empresas constantemente devem revisar seus conceitos para permitirem a utilização de toda a informação disponível a fim de obterem bons resultados (FACHINELLI et al., 2007). Assim, segundo a autora, as informações e conhecimentos, se estáticos, não possuem valor em si mesmos.

No contexto apresentado por Fachinelli et al. (2007), não basta ter informação é preciso saber utilizar esta com o objetivo de se obter bons resultados. Saber como utilizar as informações tem sido um problema para as organizações que encontram um gargalo na mão de obra qualificada. Existe uma necessidade latente de formação de pessoal qualificado e de cursos de formação específica para lidar com esse conjunto de informações. Segundo o IDC (2011), a quantidade de informação criada e replicada irá superar 1,8 *zettabytes* (1,8 trilhão de *gigabytes*), crescendo por um fator de nove em apenas cinco anos.

A produção e o crescimento de dados existentes no *Big Data* são uma das suas características mais importantes e assumem papel fundamental quando a capacidade de armazenamento é inferior a essa produção de dados.

Para o EMC (2011), hoje há mais dados sendo produzidos que a capacidade dos computadores têm de armazená-los. O Gráfico 1 apresenta uma comparação entre a capacidade de armazenamento existente ao longo dos anos e a quantidade de informação criada. Como o próprio título do Gráfico 1 afirma, existe uma sobrecarga na capacidade de armazenamento. Nesse gráfico é possível verificar também a diferença entre o comportamento do crescimento da informação e o crescimento da capacidade de armazenamento se dá pela diferença de comportamento das duas curvas. Enquanto a taxa de crescimento da informação criada se aproxima de uma curva exponencial, a taxa de crescimento da capacidade de armazenamento se aproxima do comportamento de uma função linear, fazendo com que nos últimos anos a diferença entre as duas curvas se acentue cada vez mais.

Gráfico 1 – Sobrecarga: Informação Global criada e capacidade de armazenamento em Exabytes – 2005 / 2011



Fonte: Adaptação de The Economist (2010)

O *Big Data* é uma realidade e traz um conjunto de possibilidades que deve ser observado de forma muito crítica pelas organizações governamentais e não governamentais. Para as organizações governamentais ele traz a possibilidade de um melhor gerenciamento dos serviços públicos e de um retorno mais rápido das necessidades da sociedade. Com relação às organizações não governamentais, o *Big Data* pode ser um fator decisivo no planejamento estratégico das empresas e na busca por vantagem competitiva. O aproveitamento das informações existentes encontra uma dificuldade na organização e estruturação dos dados de forma não estruturada na rede e no desenvolvimento de unidades de inteligências nas organizações. As unidades de inteligência teriam a função de mineração e extração das informações existentes, acrescentando profundidade e tratamento aos dados, tornando-os úteis.

No desenvolvimento desse estudo, além de ser a caracterização contextual, o *Big Data* foi uma das fontes de busca de dados, a intenção foi obter séries históricas de indicadores econômicos e financeiros para o estudo de correlação entre estes e o consumo de aço da empresa Alfa. O reconhecimento de padrões nos grandes conjuntos de dados foi feito através da mineração de dados o que facilitou a identificação das séries históricas.

2.2 DATA MINING

O *Data Mining* ou Mineração de Dados é parte de um processo mais amplo denominado busca do conhecimento em banco de dados. Com o desenvolvimento da tecnologia da informação, o armazenamento de informações em banco de dados tem crescido de forma exponencial. Este processo vem se desenvolvendo há algumas décadas, mas tem obtido mais destaque após o fenômeno *Big Data*. As empresas e as instituições governamentais ou não governamentais têm a necessidade cada vez maior de organizar e armazenar as informações em banco de dados.

O grande desafio desse processo está em transformar esse conjunto de informações, existente muitas vezes de forma desestruturada, em conhecimento a partir do desenvolvimento de métodos e técnicas específicas de pesquisa em banco de dados.

A mineração de dados tem se tornado uma ferramenta muito utilizada no processo de busca de informações com o objetivo de desenvolver o conhecimento sobre temas específicos. Hair Jr. et al. (2005), define a mineração de dados como a procura por relações e padrões em grande banco de dados. Neste processo o pesquisador pode operar de forma exploratória, mais estruturada, em torno de uma questão de pesquisa mais ampla, tentando identificar informações ocultas através da identificação de padrões.

Braga (2005, p. 11) define a mineração de dados como “uma coleção de técnicas e métodos facilitadora da aquisição e retenção da parte do mercado que cabe a uma empresa”, ou seja, o autor apresenta a mineração não como uma técnica específica, mas como um conjunto de técnicas e direciona sua aplicação como parte de uma estratégia de obtenção de espaço no mercado. A busca pela informação não tem origem somente no uso de ferramentas genéricas, mas também na capacidade dos analistas identificarem padrões e preverem os resultados com base em dados estratégicos de uma organização (BRAGA, 2005).

Para Braga (2005), a mineração de dados é um método automático para descobrir padrões em dados sem as limitações tendenciosas da intuição humana. Côrtes et al. (2002), discordam do conceito da mineração de dados ser um método automático, afirmam que ele está sendo difundido de forma errada e definem a mineração de dados como:

Mineração de dados é um processo altamente cooperativo entre homens e máquinas, que visa exploração de grandes bancos de dados, com o objetivo de extrair conhecimento através do reconhecimento de padrões e relacionamento de variáveis, conhecimentos esses que possam ser obtidos por técnicas comprovadamente confiáveis e válidas pela sua expressiva estatística (CÔRTEZ et al. 2002, p. 1).

Cardoso e Machado (2008, p. 497), em seu artigo sobre gestão do conhecimento usando *Data Mining*, definem:

Data Mining, ou mineração de dados, é uma técnica que faz parte de uma das etapas da descoberta de conhecimento em banco de dados. Ela é capaz de revelar, automaticamente, o conhecimento que está implícito em grandes quantidades de informações armazenadas nos bancos de dados de uma organização. Essa técnica pode fazer, entre outras, uma análise antecipada dos eventos, possibilitando prever tendências e comportamentos futuros, permitindo aos gestores a tomada de decisões baseada em fatos e não em suposições.

Com base na visão desses autores a mineração de dados pode ser definida como um processo que permite compreender o comportamento dos dados a partir da análise de um grande conjunto de informações existentes em empresas ou organizações, combinando o uso de modelos automáticos e não automáticos com técnicas de inferências estatísticas.

Alguns fatores destacam a necessidade de uso desta técnica, entre eles, o volume de informação disponível, de informação produzida pelas empresas e pela necessidade da busca de conhecimento destas informações. Assim, o acesso a informações pertinentes, interessantes, úteis e selecionadas para se obter vantagem competitiva significa muito mais do que uma quantidade significativa de informações (FREITAS; JANISSEK-MUNIZ, 2006).

A mineração de dados no contexto do *Big Data* tem uma função fundamental levando-se em consideração a desestruturação dos dados existentes. Através da mineração de dados, que permite a cooperação entre homens e máquinas na exploração de grandes bancos de dados (CÔRTEZ et al., 2002), será possível a identificação de padrões, estudo de correlações entre variáveis e desenvolvimento de modelos matemáticos ou estatísticos de previsão.

No desenvolvimento deste estudo a mineração de dados foi um processo responsável pela obtenção de dados, compreensão dos dados existentes no *Big Data*, identificação das correlações com o consumo de aço da empresa Alfa para posterior identificação do modelo de previsão de consumo de aço usado para análise das informações em um processo de Inteligência Competitiva.

2.3 PREVISÃO DE DEMANDA

Planejar “é uma atividade comum a qualquer tipo de empresa, independente de tamanho ou ramo que se dedique” (MOREIRA, 1993, p. 317), por isso toda a empresa que deseja planejar deve buscar métodos e técnicas que ajudem a estruturar o seu planejamento.

Entre os diversos fatores que a empresa deve controlar para preparar o seu planejamento encontra-se a demanda, já que para Moreira (1993):

Não obstante as eventuais diferenças, há pelo menos uma grande e importante base comum a todo o planejamento, que é a previsão de demanda. É necessário saber quanto a empresa planeja vender de seus produtos ou serviços no futuro, pois essa expectativa é o ponto de partida, direto ou indireto, para praticamente todas as decisões. As vendas podem depender de muitos fatores – aumento vegetativo da população, situação da economia mundial, movimentos de mercados internacionais, esforços para aumentar a participação de empresa no mercado, etc. – mas uma previsão, por mais imperfeita que seja, sempre é necessária.” (MOREIRA, 1993, p. 317)

Buffa (1972), afirma que a previsão de demanda é um fator crítico para modelos de decisão no âmbito da administração da produção, relacionados principalmente com estoques, planejamento e programação integradas e gestão da produção.

Riggs (1976) observa a importância das informações existentes no mercado como um indicador antecipado da produção. Ainda, segundo o autor, previsões confiáveis são necessárias, pois diminuem o tempo entre as requisições de compras e as entregas.

Segundo Monks (1987), as previsões são uma avaliação de um futuro de incertezas com o propósito de usar de melhor forma as informações disponíveis para determinar metas da empresa. O autor complementa o conceito afirmando que as boas previsões permitem aos administradores planejar de forma adequada os níveis de pessoal, matéria-prima, capital, estoques etc.

Moreira (1993) define a previsão de demanda como um processo de busca de informações sobre as vendas futuras de um determinado produto ou um grupo de produtos.

Existem diversos métodos de previsão de demanda, Mayer (1977) divide as técnicas de previsão de demanda em:

- Opinião Coletiva: busca de informações junto a vendedores, fornecedores, gerentes sobre estimativas de vendas para suas respectivas áreas;
- Indicadores econômicos: busca junto ao mercado indicadores econômicos e tenta identificar uma relação direta ou correlação entre as vendas de alguns ou de todos seus produtos e esse indicadores;
- Análise de séries temporais: analisa suas vendas passadas, para determinar se existe uma tendência. Essa tendência é, então, projetada para o futuro e as vendas resultantes reveladas são usadas como base para uma previsão.

Moreira (1993) assegura que independente do método utilizado, alguns fatores devem ser observados, entre eles os principais são:

- disponibilidade de dados;
- tempo e recursos;
- horizonte de previsão.

Independente desses fatores Moreira (1993, p. 318) expõe que algumas características são comuns a todos os métodos, e cita as seguintes:

Os métodos de previsão geralmente assumem que as mesmas causas que estiveram presentes no passado, configurando a demanda, continuarão presentes no futuro. Isso quer dizer que o comportamento do passado é base para inferir o comportamento do futuro.

Os métodos conduzem a resultados perfeitos, e a chance de erro é tanto maior quanto mais aprofundamos no futuro, ou seja, quanto maior seja o nosso horizonte de previsão. Isso acontece porque os fatores aleatórios, que nenhuma previsão consegue captar, passam a exercer mais influência (MOREIRA, 1993, p. 318).

Os métodos de previsão de demanda podem ser classificados por diversos critérios, mas segundo Moreira (1993) a classificação mais simples considera o tipo de instrumento e conceitos que foram usados na base da previsão. Nesta linha o método de previsão pode ser classificado em:

- Qualitativo: quando a previsão é feita com base nos julgamentos de pessoas que tenham informações suficientes para opinar sobre o assunto;
- Quantitativos ou Matemáticos: utilizam modelos matemáticos para fazer as previsões, permitindo controle do erro.

Ao se realizar uma previsão de demanda uma variável que deve ser avaliada é o período de tempo da previsão, segundo Buffa (1972) a previsão de demanda pode ser feita levando-se em consideração três períodos diferentes de tempo:

- Futuro imediato: tem base nas operações correntes;
- Futuro intermediário: considera três ou cinco anos próximos;
- Longo prazo: períodos superiores a cinco anos.

Um outro fator que tem influência na previsão de demanda é a classificação do tipo de demanda. Moreira (1993) afirma que existem dois padrões de demanda, um classificado como dependente e outro como independente. A demanda dependente está relacionada à produção interna de outros produtos, basicamente ela tem relação com o consumo de matéria-prima. Já a demanda independente tem relação com o produto acabado, está ligada diretamente ao mercado externo.

Todos os autores trazem a importância da previsão de demanda, os métodos e características que devem ser observados. Mas Corrêa (2007) apresenta um outro elemento que tem grande importância na previsão, a habilidade de prever demanda. Para o autor:

Habilidade de prever a demanda é muito importante para que a empresa saiba utilizar todas as ferramentas disponíveis para conseguir antecipar a demanda futura com alguma precisão. Isso pode envolver, formar e manter base de dados históricos de vendas, assim como informações que expliquem suas variações e comportamento no passado, utilizar modelos matemáticos adequados que ajudem a explicar o comportamento da demanda, compreender como os fatores ou variáveis internas (promoções etc.) e externas (clima, condições econômicas etc.) influenciam o comportamento da demanda, coletar informações relevantes do mercado e ser capaz de derivar daí uma estimativa da demanda futura (CORRÊA, 2007, p. 239).

Nas palavras do autor é possível mais uma vez identificar a importância da previsão de demanda. Ele traz novas variáveis que devem ser pensadas na hora de planejar um processo de previsão de demanda. Essas variáveis são: ferramentas, armazenamento de dados históricos, conhecimento de modelos matemáticos, compreensão das variáveis internas e externas.

Corrêa (2007) afirma que para se fazer uma boa previsão, levando-se em consideração essas variáveis, a empresa precisa manter um canal de comunicação com o mercado e através deste, trazer informações dos clientes e do mercado para a empresa.

A visão de Corrêa sobre previsão de demanda vem ao encontro dos objetivos desse estudo, quando o autor descreve a importância de se obter informações que expliquem as variações de demanda e o uso de modelos que ajudem a compreender a influência de variáveis internas e externas no comportamento da demanda. Nesse estudo a fonte de variáveis externas foi o *Big Data*, o processo para identificação dessas variáveis o *Data Mining*, e o modelo de previsão de consumo de aço, demanda dependente, foi o instrumento usado para a análise das informações em um processo de Inteligência Competitiva.

2.3.1 Séries Temporais

As séries temporais são conjuntos de observações de uma determinada variável ordenada conforme uma unidade de tempo (CAMARGO; SOUZA, 1996), para Souza (1989), uma série temporal é a classe de fenômenos cujo processo observacional e consequente quantificação numérica gera uma sequência de dados distribuídos no tempo.

Pelas definições fica evidente que uma série temporal é uma sequência de valores que uma determinada grandeza assume em intervalos de tempo. A variável independente em

uma série temporal é o tempo, e dependendo da unidade de tempo adotada, a série pode ser classificada em discretas e contínuas.

Uma série é classificada como contínua quando a variável tempo pode assumir qualquer valor em um intervalo infinito não-numerável (CAMARGO; SOUZA, 1996), sendo a série representada por $Z(t)$. Já em uma série discreta, a unidade de tempo deve assumir valores dentro de um conjunto finito e numerável, podendo ser representada por $X^T = \{x_1, x_2, \dots, x_T\}$, sendo que cada observação discreta x_T está associada a um instante de tempo distinto, existindo uma relação de dependência serial entre essas observações (SOUZA, 1989).

O objetivo principal de se estudar uma série temporal está relacionado com a realização de inferências sobre as propriedades da série ou características do mecanismo gerador da mesma (GRANGER, 1977). Conhecendo as propriedades e as características de seu funcionamento é possível construir um modelo matemático que a represente o que ajudará a previsão de valores futuros da série (GRANGER, 1977).

Uma previsão de uma série temporal é a descoberta de valores futuros da série (CAMARGO; SOUZA, 1996, p. 11), ou, “uma previsão é uma estimativa quantitativa (ou conjunto de estimativas) acerca da verossimilhança de valores futuros de eventos futuros baseada na informação atual e passada”.

Morettin e Tolo (1981) afirmam que a previsão não constitui um fim em si, mas é uma forma de obter informações e subsídios para uma consequente tomada de decisão, visando atingir determinados objetivos.

A previsão de uma série temporal é uma estimativa dos possíveis valores que ela pode assumir fora do intervalo de tempo em que a série foi observada.

Camargo e Souza (1996) afirmam que as previsões podem ser classificadas em:

- Prazo imediato: menor que um mês, neste caso o ideal é ter dados diários ou semanais;
- Curto prazo: entre um mês e três meses, neste caso o ideal é ter dados semanais e mensais;
- Médio prazo: entre três e seis meses, o ideal é ter dados mensais e trimestrais;
- Longo prazo: um ano ou mais, o ideal é ter dados mensais, trimestrais ou anuais.

Existem diversos métodos de previsão de séries temporais, entre eles:

- Métodos de Decomposição de Séries Temporais;

- Média Móvel;
- Alisamento Exponencial Simples;
- Alisamento Exponencial Linear;
- Alisamento Exponencial Sazonal e Linear de Winter;
- Modelo Autorregressivo e de Médias Móveis;
- Modelo de Box e Jenkins;

Neste trabalho foi estudada a série temporal de consumo de aço da empresa Alfa, para o período de janeiro de 2007 a maio de 2012, com o objetivo de construir um modelo de previsão de consumo de aço. Para identificação do modelo de previsão de consumo foi utilizado um modelo Box Jenkins com função de transferência.

2.3.2 Modelos univariados de Box e Jenkins

Os modelos univariados de Box e Jenkins são uma metodologia bastante utilizada para construção de modelos paramétricos (MORETTIN; TOLOI, 2006) e são capazes de manipular séries temporais de qualquer natureza além de ter como objetivo a construção de um modelo Autorregressivo Integrado de Médias Móveis (ARIMA).

A metodologia desenvolvida por Box e Jenkins se caracteriza por interpretar uma série temporal como resultado de um vetor aleatório multivariado com dimensão definida pela série (CAMARGO; SOUZA, 1996). O objetivo da metodologia é buscar e detectar o sistema gerador da série através de informações da própria série (CAMARGO; SOUZA, 1996).

O processo de construção de um modelo ARIMA de previsão constitui-se de ciclo interativo (MORETTIN; TOLOI, 2006), onde a estrutura do modelo é construída através dos próprios dados. Segundo Morettin e Toloi (2006), o ciclo de construção do modelo respeita cinco estágios:

- especificação de uma classe geral de modelos para análise;
- identificação de um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- estimação dos parâmetros modelo;
- verificação do modelo ajustado através de uma análise de resíduos.

Se ao final de um ciclo o modelo não estiver adequado, o ciclo é repetido até que se encontre um modelo que melhor se ajuste ao conjunto de dados, ou seja, o modelo que apresenta o menor erro de previsão (MORETTIN; TOLOI, 2006).

Camargo e Souza (1996) afirmam que além do ciclo iterativo a série deve respeitar o princípio de parcimônia, o qual determina que a série deve ser modelada com o menor número de parâmetros possíveis, com isto será obtida uma representação matemática adequada.

Para se construir o modelo de Box Jenkins devemos inicialmente definir as funções de autocorrelação (ACF) e autocorrelação parcial (PACF). Essas vão verificar se a série temporal é estacionária ou não, além disso, vão determinar a ordem dos modelos de Box e Jenkins.

A série temporal é classificada como estacionária quando se desenvolve ao longo do tempo de forma aleatória ao redor de uma média constante, representando um equilíbrio estável (MORETTIN; TOLOI, 1987). Segundo Bueno (2008, p 15), “o conceito de estacionariedade é a principal ideia que se deve ter para estimar uma série temporal”. “Se o processo estocástico que gerou a série de observações é invariante com relação ao tempo, diz-se que o mesmo é estacionário” (CAMARGO; SOUZA, 1996, p. 27).

Camargo e Souza (1996) destacam que a importância de uma série ser estacionária ou não está diretamente relacionada ao fato de se estar na presença de um processo que tem a mesma forma o tempo todo, o que facilita a obtenção de estimativas das características da série de forma bastante simples.

Uma série temporal estacionária apresenta ambos os coeficientes ACF e PACF tendendo a zero, enquanto que as séries não-estacionárias apresentam coeficientes significativamente diferentes de zero para vários períodos de tempo da série temporal (BOX; JENKINS; REINSEL, 1994).

A função de autocorrelação (ACF) determina a autocovariância padronizada da série, que tem como objetivo medir o comprimento da memória do processo, ou seja, ela estima a dependência entre valores tomados em instantes de tempo distintos t e $t + k$, onde k representa a diferença de tempo entre os intervalos. A autocorrelação com defasagem de tempo k é definida como:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{Cov}[Y_t, Y_{t+k}]}{\sqrt{\text{Var}(Y_t) \text{Var}(Y_{t+k})}} \quad (1)$$

onde:

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

$$\rho_0 = 1 \text{ e } \rho_k = \rho_{-k}.$$

A função de correlação é simétrica sobre a $k = 0$ e a partir dela é possível construir um gráfico dos coeficientes de autocorrelação ρ_k versus k que denominamos correlograma.

A função de autocorrelação amostral representa a autocovariância entre as médias de amostras da mesma variável em instantes distintos de tempo e tem com objetivo medir a dependência entre os dados de diversas amostras em momentos distintos. Em uma série temporal Y_1, Y_2, \dots, Y_n , a autocorrelação amostral é dada por:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad k = 0, 1, 2, \dots \quad (2)$$

Para Box e Jenkins (1976), o tamanho da amostra é fundamental para se obter uma boa estimativa da função de autocorrelação. Eles estimam que uma amostra deve ter no mínimo 50 observações e as autocorrelações $\hat{\rho}_k$ seriam calculadas para $k = 0, 1, \dots, j$, onde j não deve ser maior que $\frac{n}{4}$, onde n representa o número de observações.

A autocorrelação parcial (PACF) mede a correlação entre duas observações seriais distintas Y_t e Y_{t+k} onde a dependência entre os termos intermediários, $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$, é eliminada. A PACF é uma medida de ordem k e tem condicionamento ao conhecimento das autocorrelações de ordem inferior a k , a PACF é definida por Φ_{kk} e dada por:

$$\Phi_{kk} = \frac{|\rho_k^*|}{|\rho_k|} \quad (3)$$

Sendo:

ρ_k : é a matriz de autocorrelação;

$|\rho_k|$: é o determinante da matriz de autocorrelação;

ρ_k^* : é a matriz que difere de ρ_k somente pela última coluna, substituída por $\{\rho_1, \rho_2, \dots, \rho_k\}$

$|\rho_k^*|$: é o determinante da matriz ρ_k^* .

Box, Jenkins e Reinsel (2008, p. 8), definem o operador de translação “B” que representa a defasagem para k períodos de tempo para trás e que será usado neste texto amplamente, por:

$$BZ_t = Z_{t-1}$$

$$B^2Z_t = Z_{t-2}$$

⋮

$$B^k Z_t = Z_{t-k} \quad (4)$$

O operador soma é definido por:

$$SZ_t = \sum_{j=0}^{\infty} Z_{t-j} = Z_t + Z_{t-1} + \dots = (1 + B + B^2 + \dots) Z_t \quad (5)$$

Para Box, Jenkins e Reinsel (2008, p. 8), outro operador muito importante é o operador diferença, que consiste em tomar diferenças sucessivas da série original, sendo definido por:

$$\nabla^d = (1 - B)^d \quad (6)$$

Os modelos univariados de Box e Jenkins podem ser classificados em modelos com intervenção e sem intervenção. Nos próximos capítulos serão apresentados esses dois grupos de modelos e suas respectivas definições.

2.3.3 Modelos univariados de Box e Jenkins sem intervenção

Um modelo é classificado como univariado por representar somente o estudo de uma única série temporal. Este tipo de modelo é denominado de sem intervenção quando não existe impacto de um elemento externo no comportamento de uma determinada série temporal. A intervenção pode ser caracterizada por modificações de legislação, crises políticas, crises econômicas, modificações do clima, etc.

Os modelos univariados sem intervenção podem ser:

- Modelo Geral;
- Modelo ARIMA(p,d,q);
- Modelo Autorregressivo de ordem p – AR(p);
- Modelo de Média Móvel de ordem q – MA(q);
- Modelos Mistos Autorregressivos de Média Móvel – ARMA(p,q)

2.3.3.1 Modelo Geral

Uma das ferramentas importantes para o estudo e modelagem de séries temporais é o modelo Autorregressivo Integrado de Médias Móveis Sazonal (SARIMA) de ordem (p,d,q)x(P,D,Q)s, descrito por Box e Jenkins (1976). Este modelo é representado por:

$$\phi(B)\Phi(B^s)(1 - B^s)^D Y_t = \theta(B)\Theta(B^s)a_t \quad (7)$$

onde:

$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$, operador não sazonal autorregressivo (AR) ou polinômio de ordem p tal que as raízes da equação característica $\phi(B)=0$ devem estar fora do círculo unitário ($\phi(B)$ um polinômio em B de grau p , tenha todas as suas raízes $\phi(B)=0$ maiores que 1, em valores absolutos, ou seja, todas as raízes devem estar fora do círculo unitário para garantir a estacionariedade e os $\phi_i, i=1,2,\dots, p$ são os parâmetros autorregressivos não sazonais;

$\nabla^d = (1 - B)^d$, onde ∇^d é o operador diferença não sazonal de ordem d , usualmente $d = 0, 1$ ou 2 ;

B : operador de retardo, tal que, $B^k Z_t = Z_{t-k}$;

$\Phi(B^s) = (1 - \Phi_1 B^s - \dots - \Phi_p B^{ps})$, operador sazonal autorregressivo (AR) ou polinômio de ordem p tal que as raízes da equação característica $\Phi(B^s)=0$ devem estar fora do círculo unitário para garantir a estacionariedade e os $\Phi_i, i=1,2,\dots, p$ são os parâmetros autorregressivos sazonais;

$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$, operador não sazonal médias móveis (MA) ou polinômio de ordem q tal que as raízes da equação característica $\theta(B)=0$ devem estar fora do círculo para garantir a inversibilidade do processo e os $\theta_i, i=1,2,\dots, q$ são os parâmetros médias móveis não sazonais;

$\Theta(B) = (1 - \Theta_1 B^s - \dots - \Theta_q B^{qs})$, operador sazonal de médias móveis (MA) ou polinômio de ordem q tal que as raízes da equação características $\Theta(B^s)=0$ devem estar fora do círculo para garantir a inversibilidade do processo e os $\Theta_i, i=1,2,\dots, q$ são os parâmetros de médias móveis sazonais;

a_t : processo de ruído branco, normalmente e identicamente distribuído, com média zero e variância constante σ_a^2 ou seja, NID $(0, \sigma_a^2)$;

Y_t : é a série em estudo para o período t ;

(AR): operador não sazonal autorregressivo que modela a dependência de um valor atual sobre os valores passados;

(I): operador de integração que remove a tendência de uma série temporal, tornando-a estacionária;

(MA): operador não sazonal de média móvel, onde os valores atuais são dependentes de erros de previsão de períodos passados;

p : ordem máxima dos parâmetros autorregressivos simples;

d : número de diferenciações não sazonais aplicadas para tornar a série temporal estacionária;

q : ordem máxima dos parâmetros de média móvel simples;

P : ordem máxima dos parâmetros autorregressivos sazonais;

D : número de diferenciações sazonais aplicadas para tornar uma série temporal estacionária;

Q : ordem máxima dos parâmetros de média móvel sazonal;

s : comprimento sazonal.

2.3.3.2 Modelo ARIMA(p,d,q)

Os modelos Autorregressivos Integrados de Médias Móveis (ARIMA (p,d,q)) são resultado de um procedimento onde uma série W_t é obtida através da diferenciação de uma série Z_t , portanto pode-se afirmar que Z_t é uma integração de W_t , o que torna o modelo integrado, além de ser autorregressivo (CAMARGO; SOUZA, 1996).

O processo de diferenciação ocorre quando se tem uma série não estacionária e, deve-se, antes de ajustá-la a um modelo estacionário (MA, AR, ARMA), torná-la estacionária pela remoção de padrões não estacionários, por exemplo tendência e sazonalidade, da série temporal analisada. A diferenciação tem como objetivo tornar a série estacionária em relação a sua média e para torná-la estacionária em relação à variância, devemos realizar transformações logarítmica, raiz quadrada ou exponencial (MAKRIDAKIS; WHELLWRIGHT; HYNDMAN, 1998).

Os modelos ARIMA (p,d,q) partem, também, da ideia de que série temporal Y_t é uma realização particular de um processo estocástico gerado pela passagem sucessiva de um processo ruído branco a_t a uma sequência de dois filtros lineares: um estável e outro instável. Não existe uma determinação do número de filtragens que devam ser realizadas, o processo deve ser repetido até que se obtenha, na saída, um processo com as características necessárias para representar o processo não estacionário homogêneo.

Para aplicação de um modelo ARIMA uma condição necessária é que o processo que gerou a série temporal seja estacionário de segunda ordem, ou seja, que sua média e variância sejam constantes no tempo.

No modelo ARIMA(p,d,q) o parâmetro “d” representa o grau de diferenciação necessário à estabilização dos dados e é obtido através da análise do correlograma. Quando o correlograma decresce de forma exponencial, realiza-se o número necessário de

diferenciações na série até que o correlograma mude abruptamente para um valor qualquer, pois neste momento a série será considerada estacionária.

A ordem “p” é determinada pela função de correlação parcial (PACF) da série estudada, se a série for somente autorregressiva ARIMA(p, d, 0), a função de autocorrelação parcial sofrerá uma queda repentina; caso contrário devemos fazer uma análise dos estimadores até que a ordem de defasagem do correlograma desta função indique se ela é estatisticamente autorregressiva e esta será sua ordem autorregressiva.

O parâmetro q representa a ordem máxima dos parâmetros de média móvel simples.

Se $w_t = \nabla^d Y_t$ é estacionária o parâmetro “d” assume o valor nulo já que não é preciso diferenciar a série para torná-la estacionária, portanto pode-se representar w_t por um modelo ARMA (p,q), ou seja:

$$\Phi(B) w_t = \Theta(B) a_t \quad (8)$$

Se w_t é uma diferença de Y_t , então Y_t é uma integral (soma) de w_t , daí diz-se que Y_t segue um modelo Autorregressivo Integrado de Médias Móveis, ou modelo ARIMA(p,d,q) ∇^d (CAMARGO; SOUZA, 1996), assim:

$$\Phi(B)\nabla^d Z_t = \Theta(B)a_t \text{ de ordem (p,d,q)} \quad (9)$$

2.3.3.3 Modelo Autorregressivo de ordem p – AR(p)

O modelo AR(p) é definido pela equação (10), representando uma regressão da variável dependente em função dos valores históricos. Essa equação pode ser representada em termos do operador B, através da equação (11), ou de sua forma simplificada na equação (12), descrita a seguir (BOX; JENKINS; REINSEL, 1994):

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (10)$$

$$\varepsilon_t = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t \quad (11)$$

$$\varepsilon_t = \phi(B) Y_t \quad (12)$$

onde:

μ : constante;

ϕ_i : coeficiente autorregressivo do período i ;

ε_t : erro aleatório do período t ;

$\phi(B)$: polinômio autorregressivo de ordem p .

Fazendo-se $\pi_j = 0, j > p$, no modelo $\pi(B) \tilde{Y}_t = a_t$, obtém-se um modelo chamado de Autorregressivo de ordem p , denotado por AR(p), isto é:

$$\tilde{Y}_t = \phi_1 \tilde{Y}_{t-1} + \phi_2 \tilde{Y}_{t-2} + \dots + \phi_p \tilde{Y}_{t-p} + a_t \quad (13)$$

Definindo-se o operador autorregressivo de ordem p :

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (14)$$

tem-se:

$$\phi(B) \tilde{Y}_t = a_t \quad (15)$$

O modelo AR(p) é invercível ao se verificar que $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é finito. Por outro lado, para $\phi(B)$ convergir, as raízes de $\phi(B) = 0$ devem cair fora do círculo unitário, ou seja, um modelo AR(p) é estacionário, se todas as raízes de $\phi(B) = 0$ caírem fora do círculo unitário.

2.3.3.4 Modelo de Média Móvel de ordem q – MA(q)

O modelo de média móvel MA(q) realiza a regressão da variável Y_t com os erros passados (ε_{t-q}), conforme as equações (16) e (17), sendo que, esta última, representa a forma polinomial da equação (16) (BOX; JENKINS; REINSEL, 1994).

$$Y = \mu - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} + e_t \quad (16)$$

$$Y_t = \theta(B) e_t \quad (17)$$

onde:

θ_i : coeficiente de média móvel do período i ;

$\theta(B)$: polinômio de média móvel de ordem q .

Fazendo-se $\Psi_j = 0, \forall j > q$, no modelo:

$$Y_t = \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \dots \quad (18)$$

Obtém-se o processo de Médias Móveis de ordem q , denotado por $MA(q)$. Usualmente emprega-se a notação $-\theta_j$, ao invés de Ψ_j , ou seja, $\Psi_1 = -\theta_1$, $\Psi_2 = -\theta_2$, ..., $\Psi_q = -\theta_q$ e $\Psi_k = 0$, se $k > q$. Assim, um $MA(q)$ tem a forma:

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (19)$$

ou seja:

$$\tilde{Y}_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t = \theta(B) a_t \quad (20)$$

sendo:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (21)$$

o operador de médias móveis, de ordem q .

Com relação à estacionariedade e à inversibilidade do modelo, dado que $\Psi(B) = \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$, não há restrições sobre os parâmetros θ_j para que o processo seja estacionário. Por outro lado, pode-se verificar que a condição de inversibilidade para um $MA(q)$ é que as raízes de $\theta(B)=0$ caiam fora do círculo unitário.

2.3.3.5 Modelos Mistos Autorregressivos de Média Móvel – ARMA(p,q)

Os modelos autorregressivos e de Média Móvel ARMA(p,q) resultam da combinação das características dos modelos AR(p) e MA(q). Se a série estudada for estacionária, ela pode ser modelada através do modelo ARMA(p,q). As equações (22) e (23) representam o modelo (BOX; JENKINS; REINSEL, 1994):

$$Y_t = \mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (22)$$

$$\phi(B) Y_t = \mu + \theta(B) e_t \quad (23)$$

Os modelos ARMA(p, q) contêm termos autorregressivos e de médias móveis, sendo, portanto, dados por:

$$\tilde{Y}_t = \Phi_1 \tilde{Y}_{t-1} + \dots + \Phi_p \tilde{Y}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (24)$$

onde, considerando-se $\phi(B)$ e $\theta(B)$ como os operadores autorregressivos e de médias móveis, respectivamente, pode-se escrever um ARMA(p, q) na forma:

$$\phi(B) \tilde{Y}_t = \theta(B) a_t \quad (25)$$

2.3.3.6 Modelos univariados de Box e Jenkins com intervenção

Os modelos univariados com intervenção levam em consideração o impacto de variáveis externas no comportamento da série temporal. Esses modelos foram desenvolvidos por Box e Tiao (1975), eles se caracterizam por serem um caso especial dos modelos de função transferência de Box e Jenkins (1970), nos quais as variáveis de entrada são acontecimentos exógenos e de natureza determinística que são denominadas de intervenções quando o momento de ocorrência é conhecido a priori. É possível mencionar como exemplos de intervenções: greves, alterações legislativas, crises econômicas, decisões políticas, perturbações climáticas, campanhas de publicidade e promoção, etc.

Para Box e Jenkins (1970), os modelos de intervenção são uma generalização de métodos usados para a análise de dados, usualmente não expressos na forma de séries temporais, aos quais os estatísticos referem-se pelo título geral de Delineamento e Análise de Experimentos. Uma característica importante dos modelos com intervenção é a sua capacidade de descrever um grande número de efeitos simultâneos, apesar de sua simplicidade.

A equação (26) apresenta o modelo proposto para a análise de intervenção:

$$Y_t = \sum_{i=1}^k v_i(B) I_{j,t} + \eta_t \quad (26)$$

onde:

Y_t é a variável-resposta do modelo;

k é o número de intervenções da série;

$v_i(B)$ é o valor da função e transferência ;

$I_{j,t}$ é a variável binária;

η_t é o ruído do modelo, representado por um modelo ARIMA.

Nos modelos de Box e Jenkins, é necessário que a série seja estacionária, ou seja, não apresente tendência e sazonalidade. Assim, o modelo (26) pode ser expresso pela equação (27) por meio de um modelo ARMA(p,q) com intervenção (BOX; TIAO, 1975).

$$\hat{Y}_t = \sum_{j=1}^k \frac{\varpi_j(B) B^b}{\delta_j(B)} I_{j,t} + \frac{\theta(B)}{\phi(B)} a_t \quad (27)$$

onde:

\hat{Y}_t = variável estimada pelo modelo no período t ;

$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$, é o polinômio Autorregressivo não sazonal de ordem p ;

$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$, é o polinômio de médias móveis não sazonal de ordem q ;

a_t = é uma sequência de ruído branco que segue uma $N(0, \sigma_a^2)$;

$I_{j,t}$: representa as variáveis de intervenção para $j= 1, \dots, k$; para a observação (t) onde ocorreu a intervenção.

$$\omega_j(B) = (\omega_0 - \omega_1 B - \dots - \omega_s B^s) \quad (28)$$

$$\omega_j(B) = (\omega_0 - \omega_1 B - \dots - \omega_s B^s) \quad (29)$$

$$\delta_j(B) = (\delta_0 - \delta_1 B - \dots - \delta_r B^r) \quad (30)$$

Os efeitos iniciais da intervenção são representados na equação pelos ω_j 's e os efeitos permanentes são representados pelos δ_j 's .

A intervenção tem como característica mudar abruptamente a série para cima ou para baixo após um determinado momento ou após um espaço de tempo. Os efeitos dessas mudanças podem ser permanentes ou temporários, se temporários a série tende a ter o mesmo comportamento anterior à intervenção.

Uma forma de caracterizar e quantificar a natureza e a magnitude dos efeitos é distinguir os tipos de variáveis de intervenção. A distinção se dá pela análise do momento da intervenção e pelo intervalo de atuação sobre a série, assim é possível caracterizar os seguintes tipos de variáveis de intervenção:

- Função degrau: quando a intervenção ocorre no instante “T” e permanece atuando na série;
- Função impulso: quando a intervenção ocorre no instante “T” e tem efeito apenas neste instante;
- Função de impulso sazonal: quando a intervenção ocorre em um instante “T” e tem efeito nos instantes múltiplos de “T”.

Quando não é possível identificar os instantes “T” de ocorrência de fenômenos que têm influência sobre as séries *a priori*, os dados desta série recebem o nome de *outliers* ou pontos fora da curva, a identificação de *outliers* em uma série temporal é fundamental para o

desenvolvimento do modelo, pois a sua presença pode distorcer as autocorrelações e as autocorrelações parciais tendo efeito substancial na especificação do modelo (BOX; JENKINS, 2008).

A detecção de *outliers* em séries temporais foi inicialmente introduzida por Fox (1972), ao propor dois modelos paramétricos: *Additive Outlier* e *Innovation Outlier*.

Mais recentemente autores como Tsay (1986,1988), Chang; Tiao; Chen (1988) e Ljung (1993), têm proposto métodos iterativos para a identificação de diferentes tipos de *outliers* e alguns procedimentos para a especificação dos respectivos modelos.

Quatro modelos paramétricos podem ser utilizados para tentar identificar *outliers*, são eles:

- **Additive Outlier:** é usado para função do tipo impulso, ou seja, quando o efeito ocorre apenas no instante “T”, o modelo é definido por:

$$Y_t = X_t + \omega I_t^{(T)} = \frac{\theta_q(B)}{\phi_p(B)} a_t + \omega I_t^{(T)} \quad (31)$$

Onde X_t representa a série sem a presença de *outliers*, que segue um processo ARMA(p,q) invertível e estacionário: $\phi_p(B)Z_t = \theta_q(B)a_t$ e $I_t^{(T)}$ é uma variável binária que indica a presença ou ausência de um *outlier* no momento T, através da relação: $I_t^{(T)} = 1$ se $t=T$ e $I_t^{(T)} = 0$ se $t \neq T$.

- **Innovation Outlier :** pode ser visto como um choque na série residual, cujo efeito se faz sentir em todas as observações a partir do momento t ($Y_t, Y_{t+1}, Y_{t+2}, \dots$) segundo a estrutura do sistema descrito por $\theta_q(B)/\phi_p(B)$. Pode descrever-se através da relação,

$$Y_t = \frac{\theta_q(B)}{\phi_p(B)} (a_t + \omega I_t^{(T)}) = X_t + \frac{\theta_q(B)}{\phi_p(B)} \omega I_t^{(T)} \quad (32)$$

- **Level Step:** quando a variável do tipo “degrau”, cujo efeito ocorre em um instante “T” e mantém-se atuando depois desse momento, sendo definido através da expressão:

$$Y_t = X_t + \frac{\omega}{1-B} I_t^{(T)} = \frac{\theta_q(B)}{\phi_p(B)} a_t + \frac{\omega}{1-B} I_t^{(T)} \quad (33)$$

- **Transient Change:** refere-se a um choque na série para $t \geq T$, cujo efeito tende a diminuir a um ritmo exponencial após um impacto inicial. Este outlier é definido através do modelo:

$$Y_t = X_t + \frac{\omega}{1-B} I_t^{(T)} = \frac{\theta_q(B)}{\phi_p(B)} a_t + \frac{\omega}{1-\delta B} I_t^{(T)} \quad (34)$$

onde $0 < d < 1$.

Generalizando, uma série com k outliers de diferentes tipos pode ser representada pela seguinte relação:

$$Y_t = X_t + \sum \omega_i v_i(B) I_t^{(T_i)} \quad (35)$$

onde $X_t = \theta_q(B)/\phi_p(B)a_t$; $v_i(B) = 1$ para um *additive outlier*, $v_i(B) = \theta_q(B)/\phi_p(B)$ para um *innovation outlier*, $v_i(B) = 1/(1-B)$ para um *level step* e $v_i(B) = 1/(1-\delta B)$ para um *transient change*; e $I_t^{(T_i)} = 1$ se $t = T_i$ e $I_t^{(T_i)} = 0$ se $t \neq T_i$.

2.3.3.7 Etapas para a construção de um modelo de previsão

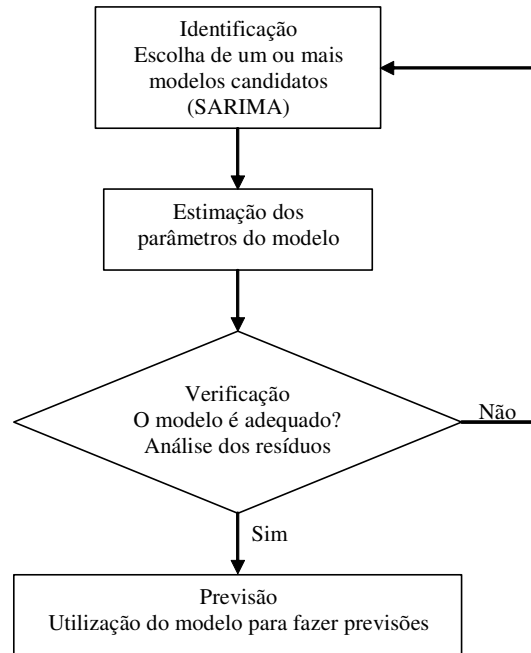
A metodologia Box e Jenkins de construção de modelo de previsão para séries temporais segue quatro etapas (BOX; JENKINS, 1976), conforme representação da Figura 3:

A primeira etapa da construção de modelos de previsão consiste não só na análise como também na preparação dos dados e seleção do modelo que melhor descreve o comportamento da série temporal entre os vários modelos de Box e Jenkins.

Para a seleção do modelo é necessário identificar a ordem dos parâmetros p , q e d :

O primeiro parâmetro a ser identificado é o grau de diferenciação d necessário à estabilização dos dados. A identificação deste parâmetro é feita através da análise do correlograma dos valores das autocorrelações (ACF) em relação aos *lags* k . Se o correlograma descrever uma série não estacionária, ou seja, se as autocorrelações decrescerem de forma linear, devemos realizar diferenciações na série até que o comportamento do correlograma apresente uma mudança abrupta para um valor qualquer de autocorrelação, identificando neste momento uma série estacionária.

Figura 3 – Fluxograma do ciclo iterativo de Box e Jenkins.



Fonte: Adaptado de Box; Jenkins (1970).

O segundo parâmetro a ser identificado é o p que indica a ordem máxima dos parâmetros autorregressivos simples, a determinação do p é dada pela verificação da função de autocorrelação parcial (PACF) da série.

Por fim devemos identificar o parâmetro q que indica a ordem máxima dos parâmetros de média móvel simples.

A identificação do modelo que melhor descreve o comportamento da série pode ser feita por diversos procedimentos, entre os quais:

- Critério de Informação de Akaike (AIC) (AKAIKE, 1973, 1974) para processos ARIMA (OZAKI, 1977), dado por:

$$AIC(p, d, q) = N \ln(\hat{\sigma}_a^2) + \frac{N}{N-d} 2(p + q + 1 + \delta_0 + N \ln(2\pi + N)) \quad (36)$$

Onde:

$$\delta_0 = \begin{cases} 1, d = 0 \\ 0, d \neq 0 \end{cases}$$

N é o número de observações da série em estudo;

$\hat{\sigma}_a^2$ = variância estimada dos resíduos;

$p, q =$ graus dos polinômios $\phi(B)$ e $\theta(B)$;

d é o número de diferenças utilizadas para estacionarizar a série.

- Critério Bayesiano (BIC) (AKAIKE, 1977) também conhecido por BIC (Bayesian Information Criterion), é dado pela seguinte expressão:

$$BIC(p, d, q) = \ln(\hat{\sigma}_a^2) + (p + q) \frac{\ln(N)k}{N} \quad (37)$$

onde:

k é o número de parâmetros.

O objetivo dos critérios é identificar a ordem do modelo SARIMA a ser utilizado, diminuindo a possibilidade do ajuste de modelos que possuem uma grande quantidade de parâmetros.

Os dois critérios se equivalem e buscam medir a verossimilhança dos modelos e a parcimônia paramétrica além de terem base na previsão realizada dentro da amostra. O critério para seleção do modelo está relacionado como o valor do AIC ou BIC, quanto mais próximo de zero for o valor calculado, melhor será o ajuste e maior a parcimônia (HAIR JR. et al., 2005).

A segunda etapa consiste na estimação de parâmetros dos potenciais modelos e a variância dos erros. Deve-se verificar a aleatoriedade dos resíduos, que é feita através do método de máxima verossimilhança, para posterior identificação do modelo e após, estimar os parâmetros.

Para testar se um parâmetro é zero ou não, pode-se usar a estatística t de *student*, ou seja:

$$t = \frac{\text{Estimativa do parâmetro}}{\text{Estimativa do desvio padrão desse parâmetro}} \quad (38)$$

A hipótese nula é que o parâmetro é igual a zero, e deve ser rejeitada quando o valor absoluto dessa estatística é maior que 1,96.

A terceira etapa consiste na verificação da adequação do modelo. Esta etapa avalia o ajuste do modelo através do comportamento da função de autocorrelação dos resíduos $\hat{\rho}_k(a)$, dadas por:

$$\hat{\rho}_k(\hat{a}) = \frac{\sum_{t=1}^{n-k} (a_t - \bar{a})(a_{t+k} - \bar{a})}{\sum_{t=1}^n (a_t - \bar{a})^2} \quad (39)$$

A estatística de teste utilizada para avaliar cada autocorrelação residual é a t , dada por:

$$t = \frac{\hat{\rho}_k(\hat{a})}{s[\hat{\rho}_k(\hat{a})]} \quad (40)$$

sendo $s(\hat{\rho}_k)$ um estimador do desvio padrão de $s(\rho_k)$, que é dado por:

$$s(\hat{\rho}_k) = n^{-1/2} [1 + 2 \sum_{j=1}^{k-1} \hat{\rho}_j^2]^{1/2} \quad (41)$$

onde:

n é o número de elementos da série de resíduos.

A hipótese nula a ser testada é $H_0 : \rho_k(a) = 0$, para cada autocorrelação residual.

Os limites de confiança para o desvio padrão das autocorrelações dos resíduos são dados por:

$$LC(\hat{\rho}_k(a)) = n^{-1/2} [1 + 2 \sum_{j=1}^{k-1} \hat{\rho}_j^2]^{1/2} . (\pm 2) \quad (42)$$

Na prática, se $|t| < 1,25$, nos *lags* 1, 2 e 3 e $|t| < 1,6$ nos demais *lags*, conclui-se que os erros são não correlacionados. Caso contrário, rejeita-se H_0 , ou seja, conclui-se que os erros são correlacionados.

Outro teste utilizado é o teste de *Portmanteau*, proposto por Box e Jenkins (1970), sendo modificado para o teste Q, seguindo uma distribuição de χ^2 (Qui-quadrado), tem-se a hipótese nula: $H_0 : \rho_1(a) = \rho_2(a) = \dots = \rho_v(a) = 0$

A estatística do teste é:

$$Q^* = N(N+2) \sum_{k=1}^v (N-k)^{-1} \hat{\rho}_k^2(\hat{a}) \quad (43)$$

Sendo N é o número de observações usadas para estimar o modelo e v o número de autocorrelações. A distribuição de Q^* é aproximadamente qui-quadrado com $(v-m)$ graus de

liberdade, sendo m o número de parâmetros estimados no modelo SARIMA. Pode-se testar a validade do modelo, dentro de um nível de significância, comparando o valor de Q^* com o valor tabelado da distribuição qui-quadrado.

A última etapa é a da extrapolação dos dados históricos através do modelo encontrado. Nesta etapa devemos avaliar os resultados encontrados, uma das medidas de avaliação muito usada é o *Mean Absolute Percent Error* (MAPE), também conhecido como erro médio percentual absoluto. Esta estatística tem como finalidade indicar o erro percentual absoluto médio das previsões sobre todo o conjunto de teste, sendo expressa pela seguinte equação:

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|}{n} \quad (44)$$

onde:

Y_t é o valor atual da série;

\hat{Y}_t o valor previsto;

n é o número de observações utilizadas para cálculo do MAPE.

Nesta última etapa é utilizado o modelo de Box e Jenkins para verificar se modelo é adequado para a previsão de observações futuras, ou seja, se ele descreve o comportamento dos dados da série o mais próximo possível dos dados reais. Se o modelo não cumprir esta função, o ciclo de Box e Jenkins (1970) deve ser repetido.

2.3.3.8 Modelo de Função de transferência

Os modelos apresentados anteriormente são modelos univariados, ou seja, modelos de previsão que descrevem o comportamento de uma série temporal com base somente em seu passado. Os modelos de função de transferência têm como objetivo descrever o comportamento de uma série com base no seu comportamento passado e no comportamento de outras séries temporais que podem explicar o comportamento da série estudada.

Neste tipo de modelo a série temporal Y_t é causada de forma casual unidirecional pelas variáveis $X_{1,t}$, $X_{2,t}$, ..., $X_{k,t}$.

2.3.3.9 Modelos de Função de Transferência de Box e Jenkins com Múltiplas Entradas

A análise de séries temporais multivariadas é um método que busca descrever as relações entre diversas séries temporais (BOX, JENKINS, REINSEL, 2008). A metodologia adotada consiste em considerar as séries estudadas com componentes de um vetor de séries temporais. O estudo deste vetor busca descrever o estudo das relações existentes entre os componentes dele (BOX, JENKINS, REINSEL, 2008). Este procedimento é necessário quando as séries estudadas têm um relacionamento simultâneo ao longo de um período de tempo.

Para o estudo de processos multivariados, um quadro é necessário a fim de descrever não apenas as propriedades da série individual mas também as possíveis relações transversais entre as séries e as relações possíveis entre transversais da série (BOX, JENKINS, REINSEL, 2008). A compreensão das relações existentes entre as séries se dá através do estudo da estrutura das correlações entre os componentes da série com o objetivo de determinar um modelo que represente estas relações ao longo do tempo para assim melhorar a precisão das previsões das séries individuais (BOX, JENKINS, REINSEL, 2008).

Nos modelos multivariados vamos considerar um vetor de dimensão “k” que representa um conjunto de séries temporais de comprimento n n, $X_{j,t}$, $t = 1, 2, \dots, N$, $j = 1, 2, \dots, k$ são supostas exógenas e a série Y_t , $t = 1, 2, \dots, N$ é suposta endógena. O modelo de função de transferência - ruído que relaciona as séries $X_{j,t}$ e Y_t pode ser escrito como,

$$Y_t - \mu_y = \sum_{j=1}^k v_j(B)(X_{t,j} - \mu_j) + \varepsilon_t \quad (45)$$

onde,

$$v_j(B) = \left[\frac{\omega_j(B)}{\delta_j(B)} \right] B_j^b = v_{0,j} + v_{1,j} B + v_{2,j} B^2 + \dots, \text{ é a função de transferência da } j\text{-ésima série}$$

com os pesos $v_{0,j}, v_{1,j}, \dots$, os quais são chamados de função impulso resposta;

$\omega_j(B) = \omega_{0,j} - \omega_{1,j} B - \omega_{2,j} B^2 - \dots - \omega_{s,j} B^s$ é o operador do numerador de j -ésima função de transferência, tal que as raízes de $\omega_j(B) = 0$ caiam fora do círculo unitário e, $\omega_{i,j}$, $i = 0, 1, 2, \dots, s$, são os parâmetros de $\omega_j(B)$;

$\delta_j(B) = 1 - \delta_{1,j} B - \delta_{2,j} B^2 - \dots - \delta_{r,j} B^r$ é o operador do denominador da função de transferência e, para garantir a estabilidade do sistema, as raízes de $\delta_j(B) = 0$ devem cair fora do círculo unitário e, $\delta_{i,j}$, $i = 1, 2, 3, \dots, r$ são os parâmetros de $\delta_j(B)$;

μ_j : é a média da j-ésima série de entrada;

b_j : é o atraso no tempo para $X_{t,j}$ afetar Y_t ;

$$\varepsilon_t = \frac{\theta(B)}{\phi(B)} a_t \quad (46)$$

é um ruído ARMA

a_t : é uma sequência ruído branco, com média zero e variância σ_a^2 e, $\theta(B)$ e $\phi(B)$ são, respectivamente, os operadores médias móveis e Autorregressivo, como definidos na equação 7.

Se os operadores AR e MA possuírem zeros em comum e/ou o numerador e o denominador possuírem raízes comuns o modelo não é identificável.

É recomendável que se sigam os estágios de identificação, estimação e testes de verificação do modelo quando se deseja um ajustamento a um modelo particular da função de transferência ruído para um determinado conjunto de informações. (BOX; JENKINS, 1976) (VANDAELE, 1983)

2.3.3.10 Modelos de Função de Transferência Sazonal de Box e Jenkins

Como nos modelos univariados é possível que existam parâmetros sazonais em modelo de função de transferência. Neste caso devemos induzir a estacionariedade através da diferenciação e o modelo de função de transferência sazonal multiplicativo mais geral é dado por:

$$y_t = \frac{\omega_j(B)\Omega_j(B^s)}{\delta_j(B)\nabla_j(B^s)} X_{j,t-b_j} + \frac{\theta(B)\theta(B^s)}{\phi(B)\phi(B^s)} a_t \quad (47)$$

com

$$Y_t = \nabla_s D' \nabla d' Y_t \quad (48)$$

$$X_{j,t} = \nabla_s D \nabla d X_{j,t} \quad (49)$$

Os polinômios sazonais são definidos como:

$$\Omega(B^s) = \Omega_0 - \Omega_1 B^s - \Omega_2 B^{2s} - \dots - \Omega_S B^{Ss}$$

$$\nabla(B^s) = 1 - \nabla_1 B^s - \nabla_2 B^{2s} - \dots - \nabla_R B^{Rs}$$

$$\theta(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs}$$

$$\phi(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{ps}$$

Na equação (48) os parâmetros d' e D' representam os operadores de diferença simples e sazonal, respectivamente para o processo Y , já em (49) d e D são os operadores de diferença simples e sazonal para o processo X .

É possível observar que se o numerador possui parâmetros regulares, normaliza-se a função de transferência com $w_0 \neq 1$ e $\Omega_0 = 1$, caso contrário, assume-se que $\Omega_0 \neq 1$.

2.3.3.11 Identificação dos Modelos de Função de Transferência com Múltiplas Entradas - Procedimento de Liu e Hanssens

A identificação de uma função de transferência pode ser feita através de um processo desenvolvido por Liu e Hanssens (1982).

O procedimento de Liu e Hanssens (1982), utiliza as séries estacionárias originais ou as séries de entrada e saída transformadas por um filtro comum. Este filtro é obtido através da estrutura autorregressiva das variáveis de entrada.

Para exemplificar o método é apresentado a seguir um caso que utiliza duas séries de entrada representadas pelos seguintes modelos.

$$Y_t = \frac{\omega_1(B)}{\delta_1(B)} X_{1t} + \frac{\omega_2(B)}{\delta_2(B)} X_{2t} + \varepsilon_t, t = 1, 2, \dots, n \quad (50)$$

onde x_{1t} e x_{2t} são as séries de entrada e Y_t é a série de saída.

Já a função de transferência também pode ser expressa da seguinte forma:

$$\frac{\omega_j(B)}{\delta_j(B)} B_j^b = v_j(B) = v_{j0} + v_{j1} B + v_{j2} B^2 + \dots \quad (51)$$

onde $v_j(B)$ tem um número finito de termos, se $\delta_j(B) = 1$ e um número infinito caso contrário.

Se o modelo (50) é estável, todas as raízes de $\delta_j(B)$, com $j = 1, 2, \dots$ caem fora do círculo unitário e a função de transferência (51) pode ser aproximada por um polinômio $v_j(B)$ com um número finito de termos, denotado por k_j .

O objetivo é encontrar estimativas apropriadas de $v_j(B)$ e expressá-las na forma racional $\left(\frac{\omega_j(B)}{\delta(B)} \right)$, utilizando a técnica de Box e Jenkins, para identificação de r , s e b .

Suponha k_1 e k_2 suficientemente grandes, de forma que possamos reescrever (50) da seguinte forma:

$$Y_t = (v_{10} + v_{11} B + \dots + v_{1k_1} B^{k_1}) x_{1t} + v_{20} + v_{21} B + \dots + v_{2k_2} B^{k_2} x_{2t} + \epsilon_t \quad (52)$$

$$\text{onde: } \epsilon_t = \frac{\theta(B)}{\phi(B)} a_t$$

Fazendo:

$$\underline{\beta} = [v_{10}, v_{11}, \dots, v_{1k_1}, v_{20}, v_{21}, \dots, v_{2k_2}]$$

$$\underline{x} = [x_1^0, x_1^1, \dots, x_1^{k_1}, x_2^0, x_2^1, \dots, x_2^{k_2}]'$$

$$\underline{Y} = [Y_{k+1}, Y_{k+2}, \dots, Y_{k+n}]'$$

$$\text{Onde } \underline{x}^j = B^j \underline{x}^0, \underline{x}^j = [x_j(kj+1), x_j(kj+2), \dots, x_j(kj+n)]$$

$$n = (N - k)$$

$$k = \max(k_1, k_2)$$

Assim, tem-se que o estimador linear uniformemente melhor para $\hat{\beta}$ é dado por:

$$\underline{\hat{\beta}} = (\underline{x}' \Sigma^{-1} \underline{x})^{-1} \underline{x}' \Sigma^{-1} \underline{Y} \quad (53)$$

onde Σ_n , x_n , é a matriz de covariância de $\epsilon_t = \frac{\theta(B)}{\phi(B)} a_t$.

De acordo com Ljung e Box (1979), Σ pode ser aproximada por :

$$\Sigma^{-1} = \underline{H}' \underline{H} = (\underline{D}_\theta^{-1} \underline{D}_\phi) ' \underline{D}_\theta^{-1} \underline{D}_\phi \quad (54)$$

tem-se que:

$$\underline{\hat{\beta}} = (\underline{x}' \underline{H} \underline{x})^{-1} \underline{x}' \underline{H}' \underline{H} \underline{Y} \quad (55)$$

onde $\underline{H} \underline{Y} = \underline{D}_\theta^{-1} \underline{D}_\phi^{-1}$ consiste em filtrar a série de saída por $\phi(B) \theta^{-1}(B)$. Assim, o estimador (55) corresponde a um estimador de mínimos quadrados ordinários (MQO) utilizando as séries de entrada e saída, filtradas pelo modelo do ruído.

Este procedimento pode ser representado em cinco etapas, apresentadas abaixo:

I) Ajustamento das séries de entrada

- i) Diferenciar as séries de entrada de modo a torná-las estacionárias;
- ii) Ajustar, a cada uma delas, um modelo ARMA apropriado;
- iii) Observar se os modelos não apresentaram componentes AR ou se esses componentes tiveram raízes bem maiores que um. Se isto ocorrer, passa-se para o 2º Passo.

Se a parte AR, de pelo menos um dos modelos ajustados contém raízes próximas de um, construa um filtro do tipo $(1-G_1B)(1-G_2B) \dots (1-G_lB)$, onde G_i^{-1} , $i = 1, \dots, l$ são raízes próximas de um. Este filtro é denominado “filtro Comum”. Aplique o filtro comum às séries de saída e entrada.

Esta operação é denominada pré-filtragem e é indicada com o objetivo de inverter mais facilmente a matriz $X'X$ (a pré-filtragem é indicada para obter precisão numérica).

Observa-se que o modelo de função de transferência não se altera quando seu filtro comum é aplicado a todas as séries, se elas forem estacionárias como neste caso. Conforme Priestley (1971) e Granger e Newbold (1977).

II) Estimação preliminar dos pesos da Função de Transferência

Estimam-se os parâmetros do modelo da função de transferência, utilizando o MQO e as séries obtidas no 1º Passo.

Para assegurar que valores adequados de k_j estão sendo utilizados, sugere-se começar com $k_j = 6$ e ir aumentando de 3 em 3 unidades. Um teste F pode ser utilizado para decidir quantos termos devem ser incluídos na análise.

III) Construção do modelo preliminar para resíduos

- i) Ajustar um modelo ARMA aos resíduos obtidos da estimação preliminar da função de transferência, filtrando todas as séries “pré-filtradas” por um modelo ARMA;
- ii) Caso os resíduos sejam ruído branco, obtém-se uma forma racional $(\omega_j(B)) / (\delta_j(B))$ para as séries de entrada x_j .

IV) Estimação dos pesos da Função de Transferência

Para estimar os parâmetros do modelo da função de transferência, utiliza-se o MQO e as séries obtidas no 3º Passo (isto equivale a calcular $\hat{\beta}$ da equação (55)). Recomenda-se excluir todos os termos não-significativos na estimação de MQO.

V) Estimação do modelo de Função de Transferência

- i) Identificar um modelo ARMA para os resíduos a partir da série de saída original (estacionária) e os pesos da função de transferência obtidos na estimação preliminar (4º Passo);
- ii) Reescrever as funções de transferências $v_j(B)$ na forma racional $\omega_j(B) / \delta_j(B)$, obtendo um modelo preliminar da função de transferência.

2.3.3.12 Função de Correlação Cruzada

A análise de dados para identificação do modelo de função de transferência é feita através da função de correlação cruzada entre a entrada e a saída. A dependência normalizada entre as séries X_t e Y_t será medida pela função de correlação cruzada, denotada por $\rho_{xy}(k)$ (BOX; JENKINS, 1976) e (VANDEALE, 1983).

$$\rho_{xy}(k) = \frac{Cov[X_t, Y_{t+k}]}{[\sigma_x^2 \sigma_y^2]^{1/2}} = \frac{\gamma_{xy}(k)}{[\gamma_{xx}(0)\gamma_{yy}(0)]^{1/2}} \quad (56)$$

com $k \in \mathbb{Z}$, sendo σ_x e σ_y os desvios padrões das séries X_t e Y_t , respectivamente.

2.3.3.13 Estimação dos Parâmetros dos Modelos de Função de Transferência

Após a identificação de um modelo de função de transferência através da equação (47) a tarefa é estimar os vetores de parâmetros $\underline{\omega}_j$ ($\omega_{0j}, \omega_{1j}, \dots, \omega_{sj}$), $\underline{\delta}_j$ ($\delta_{0j}, \delta_{1j}, \dots, \delta_{rj}$), $\underline{\phi}$ ($\phi_1, \phi_2, \dots, \phi_p$) e $\underline{\theta}$ ($\theta_0, \theta_1, \dots, \theta_q$). De modo semelhante se o modelo de função de transferência é sazonal, haverá vetores representativos dos parâmetros sazonais que deverão ser estimados e possivelmente haverá diferenças sazonais (47).

Para os modelos univariados ARMA há várias aproximações para a estimação destes parâmetros. Escolhe-se $\underline{\hat{\omega}}_j$, $\underline{\hat{\delta}}_j$, $\underline{\hat{\phi}}$ e $\underline{\hat{\theta}}$, como estimadores de mínimos quadrados de $\underline{\omega}_j$, $\underline{\delta}_j$, $\underline{\phi}$ e $\underline{\theta}$ respectivamente, assim que:

$$S(\underline{\hat{\omega}}_j, \underline{\hat{\delta}}_j, \underline{\hat{\phi}} \text{ e } \underline{\hat{\theta}}) = \sum_{t=1}^n a^2 t \quad (57)$$

é um mínimo. O índice n representa o número de observações possivelmente ajustadas por diferenciação. Neste processo de minimização serão utilizadas as seguintes etapas:

I) Dada a escolha inicial dos valores dos parâmetros dos polinômios $\omega_j(B)$ e $\delta_j(B)$, é possível prever a parte sistemática do modelo de Função de transferência da equação como:

$$\hat{Y}_t = \hat{v}(B) X_{jt-b}$$

II) Calcular os resíduos \hat{a}_t usando:

$$\hat{a}_t = Y_t - \hat{Y}_t$$

Após estas etapas a série \hat{a}_t pode ser usada para avaliar \hat{e}_t usando o modelo univariado ARMA, cuja representação é:

$$\hat{e}_t = \phi(B)^{-1} \cdot \theta(B) \hat{a}_t$$

Quando o parâmetro de atraso b é desconhecido, em princípio podem ser calculados os valores de ω_j , δ_j , ϕ e θ os quais minimizam a soma quadrática dos resíduos (57), para diferentes valores de b , de forma semelhante e seleciona-se o valor de b , que corresponde ao mínimo geral.

2.3.3.14 Verificação do Modelo de Função de Transferência

Após a identificação do modelo e a estimação dos parâmetros, é necessário que se verifique se o modelo é adequado ou não. Para se realizar esta verificação, tomam-se como base os resíduos.

$$\hat{e}_t = Y_t - \hat{Y}_t, \quad (58)$$

onde

$$\hat{Y}_t = \frac{\hat{\omega}(B)}{\hat{\delta}(B)} x_{jt-b} + \frac{\hat{\theta}(B)}{\hat{\phi}(B)} a_t \quad (59)$$

PIERCE (1972) mostra que $\hat{a}_t \xrightarrow{P} a_t$, para todo t , quando $n \rightarrow \infty$.

Se existirem inadequações do modelo (58) elas poderão ser detectadas examinando-se as autocorrelações dos resíduos \hat{a}_t e as correlações cruzadas entre os resíduos e a entrada pré-branqueada α_t , isto porque, se o modelo ajustado for:

$$Y_t = v_o(B) x_{jt} + (B) a_{ot} \quad (60)$$

Enquanto que o modelo correto é (59), então

$$A_{ot} = \psi_o^{-1}(B) \{v(B) - v_o(B)\} x_t + \psi_o^{-1}(B) a_t \quad (61)$$

Logo, dado que o modelo identificado é incorreto, os resíduos a_{ot} serão autocorrelacionados e também correlacionados com x_t (e α_t) e podemos ter dois casos:

i) Modelo de função de transferência correto e modelo do ruído a_t incorreto

Neste caso, $v_o(B) = v(B)$ e $\psi_o(B) \neq \psi(B)$ e (74) fica:

$$a_{ot} = \psi_o^{-1}(B) \psi(B) a_t, \quad (62)$$

de modo que os a_{ot} serão autocorrelacionados e examinado-se as autocorrelações $r_{a_0 a_0}(k)$ poderemos sugerir modificações no modelo. Neste caso, os a_{ot} não serão correlacionados com x_t e α_t .

ii) Modelo de função de transferência incorreto.

Neste caso, como ressaltamos acima, teremos a_{ot} autocorrelacionados e ainda correlacionados com a entrada. Se o modelo ruído estiver correto, $\psi_o(B) = \psi(B)$ e

$$a_{ot} = \psi^{-1}(B) \{ v(B) - v_o(B) \} x_t + a_t, \quad (63)$$

e ainda teríamos as duas situações acima. Portanto, tenta-se primeiro modificar o modelo de função de transferência para depois analisarmos o modelo do ruído.

Considerando a entrada pré-branqueada t dado por:

$$\alpha_t = \theta_x^{-1}(B) \phi(B) x_t \quad (64)$$

β_t dado por

$$\beta_t = v(B) \alpha_t + \epsilon_t, \text{ isto é,} \quad (65)$$

onde $\epsilon_t = \theta_x^{-1}(B) \phi_x(B) \eta_t$, logo

$$\epsilon_{ot} = \beta_t - v_o(B) \alpha_t = [v(B) - v_o(B)] \alpha_t + \epsilon_t \quad (66)$$

Nesta equação a existência de correlação entre ϵ_{ot} e α_t indicará se o modelo de função de transferência está correto ou não. Multiplicando-se (66) por α_{t-k} e tomando a esperança e usando \hat{v}_j obtemos;

$$v_k - v_{ok} = \rho_{\alpha\epsilon_0}(k) \frac{\sigma_{\epsilon_0}}{\sigma_\alpha}, \quad k = 0, 1, 2, \dots \quad (67)$$

A equação fornece uma medida da diferença entre a função resposta de impulsos estimada e a correta. Substituindo $\rho_{\alpha\epsilon_0}(k)$ e σ_{ϵ_0} , por estimativas s_{ϵ_0} , s_{α} , teremos novos valores preliminares para a função resposta de impulso. Depois voltamos a analisar o modelo do ruído, para ver se é adequado.

Existem outros critérios de identificação, mas descrevemos somente o critério Liu e Hanssens, porque será o critério utilizado para analisar séries de múltiplas entradas (TIAO *et al*, 1975) e (BEGUIN *et al*, 1980) e (LII, 1985).

2.3.3.15 Testes de Adequação

Em (58) é possível estimar as autocorrelações dos resíduos através de:

$$r_{\hat{\hat{\delta}}(k)} = \frac{\sum_{t=k+1}^m \hat{\delta}_t \hat{\delta}_{t+k}}{\sum_{t=1}^m \hat{\delta}_t^2} \quad (68)$$

onde m é o número da observação de \hat{a}_t efetivamente disponível. Quando m é grande, a distribuição de $r_{\hat{\hat{\delta}}(k)}$ é aproximadamente a mesma de $r_{\hat{\hat{\delta}}(k)}$, que é dada pelo seguinte resultado (BOX; PIERCE, 1970).

Se o modelo for correto e se os a_t são calculados usando-se os verdadeiros valores dos parâmetros (e portanto esses a_t são não-correlacionados), temos que $r_{\hat{\hat{\delta}}(1)}, \dots, r_{\hat{\hat{\delta}}(k)}$, terão uma distribuição assintótica normais multivariadas, supondo-se k pequeno em relação a m . As autocorrelações $r_{\hat{\hat{\delta}}(k)}$ são dadas por:

$$r_{\hat{\hat{\delta}}(k)} = \frac{\sum a_t a_{t-k}}{\sum a_t^2}, k = 1, \dots, k, \quad (69)$$

são não correlacionadas e $Var\{r_{\hat{\hat{\delta}}(k)}\} = \frac{m-k}{m(m+2)} \cong \frac{1}{m}$

Como os a_t são ruído branco, com $a_t \sim N(0, m-1)$. Assim, segue a estatística:

$$Q = m(m+2) \sum_{k=1}^k (m-k)^{-1} r_{\hat{\hat{\delta}}(k)}^2 \quad (70)$$

Para m grande, tem uma distribuição qui-quadrado com k graus de liberdade. Contudo, tem-se que estimar os a_t e as propriedades de $r_{\hat{\hat{\delta}}(k)}$ para os valores pequenos de k , são afetados. Em particular, a variância de $r_{\hat{\hat{\delta}}(k)}$ pode ser menor que m^{-1} e comparar as autocorrelações dos \hat{a}_t com $\frac{1}{\sqrt{m}}$ pode subestimar sua significância. O teste que leva em conta este fato é o seguinte.

Teste de autocorrelação: se o modelo for correto e k suficientemente grande, a estatística:

$$Q = m \sum_{k=1}^k r_{\hat{\hat{\delta}}(k)}^2 \quad (71)$$

tem uma distribuição χ^2 com $\nu = k-p-q$ graus de liberdade. Portanto, o número de graus de liberdade depende somente do número de parâmetros do modelo do ruído.

Assim, tem-se a seguinte relação: $Corr\{r_{xa}(k), r_{xa}(k + \ell)\} \cong \rho_{xx}(\ell)$, de modo que o mesmo modelo seja correto, ainda existirá correlação entre a entrada x_t e o ruído \hat{a}_t . Portanto, dado que os x_t são autocorrelacionados, um modelo de função de transferência correto poderá resultar em algum padrão das correlações cruzadas $r_{\hat{a}a}(k)$, este efeito é eliminado ao calcular-se as correlações cruzadas entre os resíduos e a entrada pré-branqueada α_t . Ou seja, analisam-se as correlações cruzadas $r_{\hat{a}\alpha}(k)$, onde:

$$r_{\hat{a}\alpha}(k) = \frac{\sum \alpha_{t-k} \hat{a}_t}{\sqrt{\sum \alpha_t^2 \sum \hat{a}_t^2}} \quad (72)$$

O seguinte teste é baseado num resultado de Pierce (1972).

Teste de correlação cruzada: se o modelo for correto e k suficientemente grande, a estatística:

$$S = m \sum_{k=0}^k r^2 \hat{\alpha}(k) \quad (73)$$

tem distribuição aproximada χ^2 com $\nu = k+1-(r+s+1) = k-r-s$ graus de liberdade.

2.3.3.16 Previsão com Erro Quadrático Mínimo

Após o teste de adequação e dados os parâmetros do modelo e as observações $Y_t, Y_{t-1}, Y_{t-2}, \dots$ a previsão no tempo t para um horizonte h , denotada por $\hat{Y}_t(h)$, pode ser recursivamente calculada considerando o seguinte fato:

$$[Y_{t+h}] = Y_{t+h}, \text{ se } h \leq 0$$

$$\hat{Y}_t(h), \text{ se } h > 0$$

$$[a_{t+h}] = a_{t+h}, \text{ se } h \leq 0$$

$$0, \text{ se } h > 0$$

onde, $[Y_t = h]$ e $[a_{t+h}]$ representa o valor esperado da variável aleatória $Y_t = h$ e a_{t+h} , respectivamente.

2.4 INTELIGÊNCIA COMPETITIVA

O mundo vive um contínuo e acelerado processo de transformação que envolve todas as áreas do conhecimento. É possível afirmar que a velocidade desse processo tem uma relação direta com a rapidez em que ocorrem as mudanças na área tecnológica. Estas mudanças têm tornado cada vez mais as relações globalizadas, modificando as transações comerciais e fazendo com que as empresas repensem as formas de competir.

Davenport (1998), com base em opiniões de autoridades e de equipes de venda de empresas de informática, afirma que estamos em uma nova era da informação, que irá revolucionar a maneira como se trabalha, compete e até mesmo como se pensa no mercado.

Fachinelli et al. (2007, p. 162) corroboram com a visão do autor ao afirmar que “o grande desenvolvimento das novas tecnologias de informação provocou profundas modificações no comportamento da sociedade contemporânea, criou novas formas de ação e interação e assim transformou a organização espacial e temporal da vida social”.

Atualmente existem diversas denominações para a sociedade em que vivemos, por exemplo, para Drucker (1993), vivemos na sociedade do conhecimento que se diferencia das épocas passadas pela importância do conhecimento. Segundo Drucker (1993), é claro que o conhecimento assume um papel diferenciado entre os recursos, mas ele não é mais um recurso como trabalho, capital e terra, mas sim um recurso único, o mais significativo e singular nesta sociedade.

Stewart (1998) não classifica a sociedade, mas a época em que vivemos, como a era da informação. Justifica essa denominação afirmando que existe uma nova realidade em que as fontes de riqueza, como recursos naturais e trabalho físico perdem a precedência para o conhecimento e a comunicação.

Tarapanof (2001), denomina a sociedade como sociedade de informação, para ela a nossa sociedade é resultado desses novos referenciais sociais, econômicos, tecnológicos e culturais, os quais também provocam um conjunto significativo de mudanças de enfoque no âmbito das sociedades e de suas organizações. Nesta conjuntura, Tarapanof (2001) destaca o papel da informação, afirma que ela é a principal matéria-prima da sociedade e que esta pode ser tratada até como “um insumo comparável à energia que alimenta os sistemas”.

É possível verificar, na fala de cada um dos autores que na época em que vivemos independente da nomenclatura adotada para a sociedade, a informação e o conhecimento assumem o papel principal tanto nas relações sociais quanto nas econômicas. Isto é tão evidente que autores como Tarapanof (2001) classificam a informação como um insumo.

Segundo Tarapanof (2001), existe uma aceitação universal da importância da informação para as organizações, e a informação é um dos recursos, se não o mais importante, cuja administração e a utilização estão diretamente ligados ao sucesso desejado.

Para Nonaka e Takeuchi (2002), a informação traz a possibilidade de se observar a interpretação de eventos ou situações de forma diferente. Eles consideram a informação como um meio de onde se extrair o conhecimento.

Fachinelli et al. (2007) asseguram que a obtenção de bons resultados passa pela necessidade das empresas revisarem seus conceitos e pela busca de caminhos que permitam a utilização de toda a informação que se encontra espalhada.

Porter (1997) não foge da visão dos demais autores e complementa citando a necessidade de possuir um sistema de informações sobre os concorrentes. Além disso, ele afirma que a informação pode ser utilizada como um instrumento de diferenciação do produto, com impacto nas vantagens competitivas. Segundo Porter e Montegomery (1998), um dos princípios da elaboração de uma estratégia está na capacidade de se lidar com a competição. Assim as empresas precisam conhecer os seus pontos fortes e fracos, seus concorrentes e saber identificar as tendências de mercado.

Na linha de Porter é possível identificar, mesmo que implicitamente, a importância da informação para tomar decisões sobre a forma como a empresa vai adquirir uma vantagem competitiva e sustentá-la.

Fachinelli et al. (2007, p. 173) não citam a informação como o elemento de diferenciação de produto, mas trazem a informação como um elemento chave para boas estratégias, dizendo que a “informação permite que a empresa tenha subsídios para uma decisão rápida e consistente”.

É clara a importância da informação, mas também é evidente que alguns autores além de tratar de informação, também trazem a questão do conhecimento. Davenport (1998), além de tratar da informação e do conhecimento, traz mais um elemento, que são os dados. Para o autor existe uma distinção entre os três, mas a distinção existente é imprecisa. Para tornar a diferenciação entre os três conceitos mais evidente, o autor apresenta o Quadro 1:

Quadro 1 - Diferença entre dados, informação e conhecimento.

Dados	Informação	Conhecimento
<p>Simple informação sobre o estado do mundo</p> <ul style="list-style-type: none"> • Facilmente estruturado; • Facilmente obtido por máquinas; • Frequentemente quantificado; • Facilmente transferível. 	<p>Dados dotados de relevância e propósito.</p> <ul style="list-style-type: none"> •Requer unidade de análise; •Exige consenso em relação ao significado; •Exige necessariamente a mediação humana. 	<p>Informação valiosa da mente humana. Inclui reflexão, síntese, contexto.</p> <ul style="list-style-type: none"> •Difícil estruturação; •Difícil captura em máquinas; •Frequentemente tácito; •Difícil transferência.

Fonte: Adaptação de Davenport (1998)

Observando o quadro 1, fica evidente a diferença entre os três conceitos e fica mais clara ainda a importância de se diferenciar os três, pois na situação em que vivemos, existe um crescimento de dados estruturados ou desestruturados, mas nem todos podem ser transformados em informação ou em conhecimento. O autor complementa afirmando que:

Conhecimento é a informação mais valiosa e, conseqüentemente, mais difícil de gerenciar. É valiosa precisamente porque alguém deu à informação um contexto, um significado, uma interpretação; alguém refletiu sobre o conhecimento, acrescentou a ele sua própria sabedoria, considerou suas implicações mais amplas (DAVENPOR, p. 18, 1998).

Ansoff e McDonnell (1993), descrevem a estratégia como um conjunto de regras de tomada de decisão que serve para orientar uma organização com relação às direções gerais para o desenvolvimento e crescimento da empresa. Nesta linha os autores também destacam que uma empresa deve criar novas estratégias quando existem novas exigências impostas pela sociedade que mudam drasticamente os objetivos da organização.

Fayard (2000) se refere à estratégia como a arte de saber usar as forças do ambiente para gerar uma determinada situação. Para ele a estratégia é a combinação entre meios diferenciados, no tempo e no espaço, para se chegar a um determinado fim. O autor também afirma que o conhecimento de si mesmo e do ambiente são elementos necessários para se identificar o momento de agir para atingir os objetivos. Ainda, segundo Fayard (2000), a informação é um dos elementos essenciais da estratégia.

Nesta conjuntura em que os dados, a informação e o conhecimento surgem como elementos transformadores e fundamentais para o posicionamento estratégico das empresas também surge o conceito de Inteligência Competitiva (IC).

Para Kahaner (1997), a inteligência competitiva consiste em um programa sistemático de coleta e análise de informações sobre as atividades de seus concorrentes e

tendências gerais dos negócios para continuar as metas da sua própria empresa. O autor destaca que a IC deve estar presente em todos os níveis da empresa, não sendo vista como mais uma função na organização, mas sim como um processo que deve aparecer em todos os aspectos do negócio, ajudando a organização a resolver diversos problemas diferentes. Por fim ele destaca que a base da inteligência competitiva está na diferenciação entre inteligência e informação.

Kahaner (1997) defende o uso da inteligência competitiva e justifica sua defesa a partir da existência de diversos fatores como: ritmo dos negócios, sobrecarga de informação, competição global, competição mais agressiva, mudanças de política e mudanças tecnológicas.

Tarapanof (2001, p. 45), define a inteligência competitiva como um “conjunto de ferramentas, o qual têm se despontado como útil para gestão da informação”. Para a autora, além de coletar dados, a inteligência competitiva tem a habilidade de “desenvolver o entendimento das estratégias e da forma de agir de seus competidores chave”. A autora complementa afirmando que a inteligência competitiva faz a “organização manter o foco no externo contínuo”.

Fuld (2007, p. 4), afirma que a IC é uma forma de ver o todo à frente do seu concorrente mais rápido e que essa tem assumido um papel crítico nos negócios, portanto deveria estar presente no trabalho de todos em uma organização. Para Fuld (2007, p. 16), “desenvolver a inteligência competitiva é semelhante a criar uma pintura pontilhada. Seu objetivo não é criar uma imagem perfeita, mas uma imagem representativa da realidade”.

Segundo Fuld (2007), uma das fontes da IC são os dados existentes na Internet, cabe à empresa aprender a obter informação e conhecimento através destes dados, para isso uma das formas consiste em estudar os truques de inteligência utilizados por analistas e estudiosos da área.

Menezes (2005), após uma revisão bibliográfica das principais publicações sobre inteligência competitiva, faz a seguinte observação sobre IC:

Observou-se que inteligência competitiva é um tema que vem sendo tratado sob diferentes abordagens: como o uso da Internet, excesso de informações, processo de coleta da informação formal, informal, automatizada, personalizada, métodos, técnicas e ferramentas para análise e monitoramento da informação, e principalmente da exigência da intervenção humana, no tratamento da informação. Com as constantes mudanças que ocorrem no ambiente competitivo, as decisões são tomadas em espaços de tempo cada vez mais curtos. Daí a necessidade de se dispor de um sistema capaz de fornecer informações previamente analisadas, aí é que se encontra o

diferencial, assegurando a tomada de decisão na hora certa (MENEZES, p. 124, 2005).

Starec et al. (2005, p. 23) consideram a inteligência competitiva como “uma área interdisciplinar e sua constituição epistemológica e aplicada recorre principalmente a conhecimentos de administração, ciências da informação, ciências da computação e economia”. Os autores também destacam a importância da inteligência competitiva como fonte de vantagem competitiva, pois segundo eles tanto a informação formal quanto a informal passam a ter relevância na formulação da estratégia das empresas.

Segundo Starec et al. (2005), o entendimento do campo empresarial sobre a inteligência competitiva é de um processo de coleta, tratamento, análise e disseminação da informação, com a finalidade de dar subsídio à tomada de decisão para atingir metas e estratégias da empresa. Conforme os autores, a inteligência competitiva tem uma orientação estratégica voltada para a busca da vantagem competitiva.

Todos os autores citados trazem suas definições sobre a inteligência competitiva e a importância desta para a estratégia da empresa. Existe uma convergência nas definições, em síntese, é possível definir inteligência competitiva como sendo um processo dinâmico que, combina diversas áreas do conhecimento na coleta e organização de dados, análise e identificação de informações existentes nos dados além de buscar a disseminação do conhecimento existente nas informações com o objetivo de fortalecer a tomada de decisões na busca da vantagem competitiva.

Starec et al. (2005) afirmam que existem versões diferentes para as fases da inteligência competitiva, mas de modo geral a inteligência competitiva pode ser dividida em 4 fases:

- planejamento e coordenação;
- coleta, processamento e armazenagem;
- análise, validação e formatação dos produtos;
- disseminação e utilização.

Planejamento e coordenação: é a primeira etapa do processo de inteligência competitiva, nela é feito um levantamento das informações estratégicas necessárias para estruturar o ciclo da inteligência competitiva. Nesta etapa é muito importante definir quais são as informações que os gestores necessitam a fim de evitar a coleta de dados que não têm valor estratégico para a tomada de decisão.

A coleta, o processamento e o armazenamento dos dados fazem parte da segunda etapa. Esta tem como objetivo identificar as fontes públicas ou privadas e buscar nelas as

informações identificadas como fundamentais na primeira etapa. Após a obtenção das informações, as mesmas são organizadas e armazenadas de forma estruturada para posterior análise.

A terceira etapa consiste na análise, validação e formatação dos dados. Os mesmos são analisados e processados na busca de identificação de relevância para que os mesmos possam servir de informação e posteriormente possam se consolidar como conhecimento. Nesta etapa o profissional de análise da informação tem um papel fundamental.

Fuld (1995), define a análise como o meio para converter a informação em inteligência. A análise não precisa ser um processo complicado, mas deve ser completa e precisa.

Tena Milan e Comai (2005), em seu artigo sobre os modelos de análise utilizados pela Inteligência Competitiva, descrevem o processo de análise:

El análisis es una parte central del círculo de la Inteligencia Competitiva. Puede concebirse como una etapa más para completar la producción de inteligencia que apoye a las decisiones empresariales. Sin embargo, su impacto en el resultado final del esfuerzo de IC es profundo. Del análisis depende la obtención de interpretaciones y conclusiones que puedan convertirse en acción. Quizá se trate de la etapa más compleja porque en ella se reúnen y combinan todos los hallazgos previos buscando escudriñar el futuro y las intenciones y acciones de otros. Por otra parte, disponemos para el análisis de una amplia variedad de recursos y técnicas. Esto facilita y potencia la capacidad de interpretación del analista, pero incrementa la importancia de la elección, a veces sutil, de las técnicas a emplear en cada momento. (TENA MILÁN e COMAI, p. 1, 2005)

Para Bruce e George (2008), análise é a parte decisiva do processo de inteligência, pois através da análise o conhecimento é produzido e podem ser tomadas as decisões. Os autores também citam a importância do analista no processo complexo da inteligência de transformar informações oriundas de várias fontes, sabendo diferenciar as informações boas e ruins, dando relevância e utilidade para a tomada de decisões. Para os autores, a mudança que teve maior impacto sobre a Inteligência foi a proliferação da mídia digital, que hoje traz o desafio da quantidade de informação produzida.

Tena Milan e Comai (2005), Bruce e George (2008) definem de uma forma muito clara a importância da Análise no processo de IC e trazem mais uma contribuição ao identificar a dependência da Análise do processo de obtenção de informação.

A última etapa do ciclo consiste na disseminação e utilização dos resultados obtidos nas etapas anteriores. É nesta etapa que o conjunto de conhecimento adquirido pode se transformar ou não em inteligência. Isto só ocorre, se os mesmos forem utilizados para a

tomada de decisão. Neste momento a participação dos responsáveis pela tomada de decisão estratégica na organização é fundamental.

O ciclo da inteligência nunca se encerra. Enquanto as etapas posteriores estão sendo executadas, o ciclo é reiniciado na busca de mais dados que possam se transformar em informação, conhecimento e, finalmente em inteligência competitiva.

Nas palavras de todos os autores fica evidente a importância da informação, também fica evidente que essa informação pode ser obtida através de dados da própria empresa ou de dados externos à empresa. A grande questão é como transformar essa informação em conhecimento que seja capaz de produzir uma vantagem competitiva. Neste ponto surge a importância da IC, que tem entre suas funções, a busca, o tratamento e a análise de dados como um processo para auxiliar a tomada de decisões.

Dentro do ciclo da IC é possível identificar o papel do fenômeno *Big Data* como uma fonte de dados. A função do *Data Mining* no processamento destes dados e a utilidade dos modelos de previsão para a organização e a identificação de padrões para o processo de análise das informações é fundamental.

2.5 ESTUDOS EMPÍRICOS REALIZADOS NA ÁREA DA DISSERTAÇÃO

Para o levantamento dos estudos realizados na área foi feita uma pesquisa bibliométrica, nas bases de dados como, *Academic One File*, *Science Direct*, *Emerald* e *Scielo* no período de 2002 a 2012. O critério de busca foram artigos que tivessem separadamente as palavras Previsão de demanda, *Data Mining* ou Mineração de dados e *Big Data* como palavras – chaves.

Os resultados obtidos por base de dados podem ser verificados na Tabela 1.

Tabela 1 – Trabalhos publicados por Base – 2002/2012.

Base	Nº	%
<i>Academic On File</i>	24	30,4
<i>Emerald</i>	19	24,1
<i>Scielo</i>	18	22,8
<i>ScienceDirect</i>	18	22,8
Total	79	100,0

Fonte: Elaborado pelo pesquisador (2012)

O levantamento de ocorrência dos assuntos previsão de demanda, *Big Data* e *Data Mining* com base na análise dos resumos e das palavras-chaves são apresentados na Tabela 2.

Tabela 2 – Ocorrência dos assuntos – 2002/2012.

Temas	Nº	%
Previsão de Demanda	36	45,6
<i>Big Data</i>	7	8,9
<i>Data Mining</i>	31	39,2
Previsão de Demanda e <i>Data Mining</i>	1	1,3
<i>Big Data</i> e <i>Data Mining</i>	3	3,8
Previsão de Demanda, <i>Data Mining</i> e <i>Big Data</i>	1	1,3
Total	79	100,0

Fonte: Elaborado pelo pesquisador (2012)

A Tabela 3 é uma tabela de referências cruzadas que combina os resultados das Tabela 1 e Tabela 2 apresentando a ocorrência dos assuntos por bases pesquisadas.

Tabela 3 – Ocorrência dos assuntos por Base – 2002/2012.

Temas	Base de Dados				Total
	<i>Academic On File</i>	<i>Emerald</i>	<i>Scielo</i>	<i>ScienceDirect</i>	
Previsão de Demanda	10	9	10	7	36
<i>Big Data</i>	4	0	0	3	7
<i>Data Mining</i>	9	9	8	5	31
Previsão de Demanda e <i>Data Mining</i>	0	1	0	0	1
<i>Big Data</i> e <i>Data Mining</i>	1	0	0	2	3
Previsão de Demanda, <i>Data Mining</i> e <i>Big Data</i>	0	0	0	1	1
Total	24	19	18	18	79

Fonte: Elaborado pelo pesquisador (2012)

A seguir serão apresentados resumos dos trabalhos que apresentaram alguma contribuição teórica ou técnica para o desenvolvimento da presente pesquisa.

Pellegrini e Fogliatto (2001), apresentaram um procedimento para a estruturação de um sistema de previsão de demanda de produtos e serviços. O objetivo do trabalho apresentado no artigo foi de propor uma metodologia que permitisse a utilização de técnicas de previsão de demanda com apoio à tomada de decisões. No desenvolvimento do artigo os autores fizeram uma revisão dos principais modelos estatísticos utilizados para identificação de padrões e tendências presentes em estudo de séries históricas. Apresentaram a estruturação e as diretrizes de implementação de técnicas de previsão de demanda e exemplificaram a aplicação da metodologia através de um estudo de caso em uma empresa do ramo da alimentação. Os autores concluem afirmando que a metodologia proposta pode contribuir de maneira eficiente à otimização dos processos de previsão.

Werner e Ribeiro (2003), realizaram um estudo de previsão de demanda através da metodologia de Box e Jenkins para analisar dados históricos de uma empresa de assistência

técnica de computadores pessoais a fim de obter previsões do número de atendimentos. Para a realização do estudo de caso a demanda da empresa foi dividida: tipo de cliente em contrato, garantia e avulso. Após a aplicação da metodologia foi possível definir modelos apropriados para os três segmentos de clientes. Os autores encerraram o trabalho afirmando que o estudo foi de grande utilidade para a empresa por possibilitar maior entendimento do comportamento dos clientes, constituindo-se como uma fonte adicional de informação para as tomadas de decisões.

Frankel (2012), em seu artigo discutiu o crescimento exponencial do volume de dados que as empresas coletam, processam e analisam. Ele define esses volumes de dados, estruturados e não organizados produzidos, como um subproduto das operações de uma empresa como *Big Data*. O autor apresentou a importância desses dados para setores financeiros e de seguros e a possibilidade que existe no *Big Data* para executivos de outros setores. Para Frankel (2012), os dados existentes podem melhorar as decisões através da análise da informação o que aumentaria a produtividade e diminuiria os riscos para a empresa. Por fim, o autor destacou o benefício de usar um método padrão para recolher e processar dados e a capacidade para usar as ferramentas analíticas projetadas especificamente trazer o contexto e o significado para os dados para uma análise apropriada e consequentemente tornando-os um recurso para empresa.

Johnson (2012), trouxe o desafio e as oportunidades existente na análise dos dados oriundos do *Big Data*. Segundo o autor, a modelagem e a análise realizada sobre os dados do *Big Data* estão conduzindo os negócios atuais e futuros. Isto fará com que as organizações evoluam, por essa razão a análise dos dados terá significativa influência nos modelos comerciais. Serão necessários esforços para gerenciar as mudanças e para a criação de uma cultura que esteja preparada para novas habilidades, infraestrutura e capacidade de avaliação de desempenho. Para o autor, as empresas, que desenvolverem de forma eficaz a capacidade de analisar dados como algo necessário, poderão estar ganhando vantagem competitiva na economia atual. Em um mundo baseado na cultura da informação, a análise de dados pode ser a força do crescimento.

Fleisher et al. (2008), realizaram um estudo que teve como objetivo a verificação de planejamento e execução da integração das técnicas inteligência competitiva, gestão de relacionamento com clientes, *Data Mining* e pesquisa de mercado. A metodologia adotada foi de estudo de caso exploratório descritivo durante o período de 3 anos. Os autores verificaram que é possível a integração das técnicas, ela gerou melhores estratégias de mercado e o

trabalho forneceu um mapa para a execução bem sucedida das técnicas integradas onde foi possível identificar as barreiras para integração bem como e sugestões.

Çiflikli e Özyirmidokuz (2012), realizaram um estudo para verificar se a utilização da mineração de dados melhora o desempenho do controle de qualidade na atividade de fabricação reduzindo perdas de produtividade. O objetivo dos autores era descobrir padrões em conjunto de dados que possibilitassem a redução de produtos defeituosos e melhorassem a qualidade. A metodologia utilizada explorou diferentes técnicas de *Data Mining*, entre elas árvore de decisões, análise de relevância de atributos e análise e detecção de anomalias. Ao final do estudo foi possível identificar um modelo baseado na árvore de decisão e de regras de decisão. Para os autores o modelo é limitado pela necessidade de ser refeito periodicamente, mas indicam que o estudo pode ser aperfeiçoado através do uso de técnicas de inteligência artificial.

Chye e Gerry (2002), discutiram em seu artigo a potencial utilidade da mineração de dados para a gestão de relacionamento com clientes no setor bancário. No desenvolvimento do artigo os autores apresentaram os conceitos de relacionamento com clientes, a metodologia de mineração de dados, suas ferramentas e sua aplicação em bancos. Em suas conclusões os autores afirmaram que não existe dúvida sobre as potências aplicação da mineração de dados no setor bancário e que esta pode ser uma fonte de vantagem estratégica perante os concorrentes. Uma das limitações apresentadas pelos autores é a da necessidade de usuários que tenham conhecimento da metodologia e das ferramentas de mineração de dados.

Chowdhury (2009), em seu trabalho buscou descobrir se as tecnologias e as ferramentas de mineração de dados são apropriadas para o uso na área empresarial a fim de detectar e prever negócios e práticas fraudulentas, além de tendências de mercado financeiro. No desenvolvimento do trabalho o autor revisa os conceitos de *Data Mining* e de Redes Neurais artificiais e suas aplicações. Em suas conclusões o autor afirmou que na área da saúde as redes neurais têm se desenvolvido de forma atrativa, ele acredita que as redes neurais artificiais podem ter uma grande aplicabilidade nos problemas empresarial reais por identificar padrões ou tendências nos dados.

Bala (2012), realizou um estudo que teve como objetivo a melhora do desempenho dos estoques de varejistas através do desenvolvimento de um modelo de previsão com base na segmentação dos clientes. A metodologia utilizada para a elaboração do modelo foi a de *Data Mining*. Nele se buscou capturar informações sobre a segmentação de clientes para transformar em conhecimento como base para a previsão de demanda. O modelo proposto pelo estudo reduziu o nível de estoque e aumentou o nível de serviço ao cliente, superando

amplamente os resultados obtidos por outros modelos. Para Bala (2012), os comerciantes podem utilizar o modelo para a previsão de demanda para diversos itens melhorando dessa forma a rentabilidade do negócio.

Cardoso e Machado (2008), realizaram um estudo onde desenvolveram, aplicaram e analisaram uma ferramenta de *Data Mining* para extrair conhecimento referente à produção científica dos pesquisadores na Universidade Federal de Lavras. A metodologia utilizada foi a pesquisa bibliográfica, a pesquisa documental e o método do estudo de caso. Os autores concluem o estudo afirmando que uma limitação encontrada foi a falta de padronização no preenchimento do currículo Lattes que leva à dificuldade no refinamento da análise e no estabelecimento de indicadores, mas que através do estudo foi possível gerar um banco de dados estruturado, este pode auxiliar na elaboração de novas políticas de gestão científica, tecnológica e no aperfeiçoamento do sistema de ensino superior brasileiro.

Tarapanoff et al (2000), apresentaram o conceito de inteligência competitiva e discutiram como as unidades de informação devem atuar em uma sociedade baseada na informação. A proposta dos autores é de que a inteligência competitiva deve atuar no monitoramento do ambiente, na busca de informações e na adequação da organização ao contexto. No desenvolvimento do trabalho os autores apresentaram: a caracterização das unidades de informação, a aplicação da inteligência competitiva em unidades de informação, o sistema de inteligência e algumas técnicas para o tratamento das informações, entre elas: redes neurais, *data warehousing* e *Data Mining*, e exemplo de unidade de informação que utiliza sistema de inteligência em seus processos.

Kwon e Sim (2013), apresentaram a necessidade crescente de algoritmos de classificação em técnicas de mineração de dados para a análise de grandes conjuntos de dados. O trabalho buscou avaliar os cenários onde conjuntos de dados afetam o desempenho dos algoritmos de classificação. Para avaliar a causalidade entre as características dos conjuntos de dados com as variáveis independentes foi utilizada a metodologia de regressão múltipla. Com base nos resultados, os autores discutem os requisitos de algoritmos de classificação para a análise de grandes bancos de dados na nova era de inteligência de negócios.

Além dos artigos científicos também foi realizada uma pesquisa nas bases de dados da Biblioteca Digital Brasileira de Teses e Dissertações BDTD (2013), para encontrar teses e dissertações referentes aos temas: Previsão de Demanda, *Big Data* e *Data Mining*.

O critério de busca na base de dados foi através do campo assunto, onde se pesquisaram separadamente os assuntos Previsão de demanda, *Data Mining* ou Mineração de dados e *Big Data*. Os resultados obtidos são apresentados na Tabela 4.

Tabela 4 – Ocorrência dos assuntos na BDTD – 2002/2012.

Temas	Nº	%
Previsão de Demanda	33	10,41
<i>Big Data</i>	0	0
<i>Data Mining</i> ou Mineração de dados	284	89,59
Total	317	100,0

Fonte: Elaborado pelo pesquisador (2013)

A seguir serão apresentados resumos das dissertações ou teses que apresentaram alguma contribuição teórica ou técnica para o desenvolvimento do presente projeto.

Kurrle (2004), realizou um estudo para o desenvolvimento de um sistema formal de previsões de demanda de veículos novos em uma concessionária da marca *Volkswagen*. A pesquisa teve objetivos exploratórios, de abordagem qualitativa e quantitativa e adotou como procedimento o estudo de caso único. Os modelos utilizados no estudo foram Modelos Box e Jenkins e de suavização exponencial com avaliação qualitativa através do método *Delph*. O autor conclui o estudo afirmando que métodos mais acurados para cada série foram considerados apropriados para gerar novas previsões de demanda. Nas sugestões para trabalhos futuros o autor sugere que sejam aplicadas as análises de regressão e correlação sobre séries temporais de demanda com a finalidade de se determinar a influência de uma classe de veículos sobre os demais e a reestruturação do processo de gestão de demanda da montadora.

Lemos (2006), em sua dissertação apresentou uma metodologia para a escolha de métodos de previsão de demanda mais apropriados a diferentes situações. A pesquisa foi de natureza aplicada, objetivos exploratórios, abordagem quantitativa, procedimentos de revisão bibliográfica e estudo de caso. A metodologia apresentada pelo autor demonstrou eficiência na seleção do método mais adequado para a previsão de demanda de produtos. Ao final do trabalho o autor sugere a incorporação de outros métodos, além dos 13 citados no trabalho e de outros fatores de seleção que não foram abordados no trabalho. Sugere também a utilização da metodologia para o desenvolvimento de um software integrado com pacotes computacionais de previsão de demanda em um sistema de apoio à decisão.

Scherrer (2006), apresentou uma dissertação que teve como problema a estimação da demanda de aço. O objetivo da autora é estimar as variáveis econômicas que têm impacto

sobre o consumo de vergalhões no Brasil. Para contextualizar o problema a autora fez um rápido resumo da história da siderurgia no Brasil e comentou rapidamente a situação mundial do setor. Na tentativa de estimar as variáveis econômicas a autora criou diversos modelos econométricos de demanda, utilizando as modelagens de vetor autorregressivo, vetor autorregressivo *Bayesiano* e variáveis instrumentais. A autora conclui que a metodologia de vetor autorregressivo *Bayesiano* foi a que apresentou melhores resultados. Como sugestão para estudos posteriores fica a possibilidade da criação de modelos para aços planos que auxiliem a explicação das possíveis diferenças entre os impactos das variáveis macroeconômicas na indústria de aços longos e planos.

Fragoso (2010), apresentou uma dissertação que teve como objetivo a proposição de um método para a criação de um processo de previsão de vendas, apoiado nas técnicas modernas de previsão de demanda e nas ferramentas de gestão de processos. O autor discutiu como as técnicas de previsão de demanda podem ser analisadas ou adaptadas dentro de uma visão de gestão de processos e propõe um método para estabelecimento de processos de previsão de demanda. Para testar a metodologia proposta, o autor realizou um estudo de caso em uma empresa do setor químico da região de Campinas. O autor conclui o estudo afirmando que o método proposto e aplicado ao estudo de caso foi eficaz. Como sugestão para estudos posteriores fica a possibilidade de um estudo de caso com demandas irregulares para solucionar o problema pela ótica de processos.

Passari (2003), desenvolveu um estudo com o objetivo de explorar a possibilidade de usar uma metodologia capaz de identificar padrões de relacionamento úteis na previsão de vendas individuais no varejo, com o uso do processo de mineração de dados. Com a utilização de um modelo de previsão de demanda de produtos o autor buscou identificar um composto de marketing adequado que agregasse preços, produtos e promoções. Em suas conclusões o autor afirma que o processo de mineração de dados e de modelagem através de redes neurais artificiais pode ser considerado adequado para a previsão de demanda de produtos para o nível individual, mas ao analisar os erros encontrados nas previsões, indica a necessidade de melhorias significativas na metodologia e na técnica de previsão.

Schaeffer (2003), em sua dissertação apresentou um estudo de caso de mineração de dados no varejo de móveis e materiais de construção. A base da mineração foram os dados gerados pelas negociações geradas pelas vendas em um período de 8 meses e informações cadastrais dos clientes. Foi utilizada a metodologia *Cross Industry Standar Process for Data Mining* (CRISP), e foram avaliados os pontos positivos e negativos das ferramentas de mineração utilizadas. Ao término do trabalho o autor aponta muitas informações de valor que

foram obtidas através da mineração e do uso de ferramentas de mineração de dados em empresas de pequeno ou médio porte, as quais podem representar um diferencial competitivo em um mercado em que está cada vez mais difícil fidelizar clientes.

Ferreira (2005), buscou em sua dissertação propor um sistema de mineração de dados que solucionasse o problema de retenção de clientes na indústria de telefonia celular. A proposta do pesquisador teve como base a identificação, em grandes bases de dados, dos clientes cujos históricos indicam maior rotatividade. No desenvolvimento do trabalho foram testados e comparados os modelos de: redes neurais, sistemas *neuro-fuzzy* hierárquicos, algoritmos genéticos, árvores de decisão e máquinas de vetor de suporte. O autor pretende que a informação existente nas bases de dados possa ser obtida através da mineração de dados e aproveitada pela inteligência empresarial como fonte de vantagem competitiva. O autor conclui o estudo afirmando que os modelos de redes neurais seguido pelo sistema *neuro-fuzzy* obtiveram melhor desempenho identificando os clientes de maior rotatividade e que o trabalho teve como limitação a falta de outras bases de dados relacionados à rotatividade dos clientes para validar o desempenho do sistema de forma mais completa.

Santos (2006), realizou um estudo na busca de compreender a contribuição da descoberta do conhecimento em bases de dados como suporte à decisão de concessão de crédito. O objetivo do autor era a criação de um modelo, utilizando mineração de dados, capaz de identificar informações necessárias para a concessão de crédito. O autor conclui o trabalho afirmando que a mineração de dados junto com os conceitos e definições para escoragem de crédito pode auxiliar no processo de decisão de crédito. Como sugestão para futuros trabalhos o autor propõe a avaliação do grau de acerto do modelo desenvolvido em função da qualidade dos dados analisados.

Fernandes (2007), em sua dissertação trouxe o seguinte problema de pesquisa: “como empresas do setor de serviços utilizam bases de dados para descobrir conhecimento sobre o cliente e embasar o desenvolvimento de estratégias de relacionamento?” Para responder a esta questão o autor buscou na literatura processos de análise de bases de dados. No desenvolvimento do trabalho foram utilizadas, como população de estudo, as empresas do setor de serviços que atuam nas cidades de São Paulo e do Rio de Janeiro, uma pesquisa quantitativa foi realizada por meio da aplicação de um questionário a 67 respondentes. A pesquisa buscava identificar o nível de utilização das etapas do processo de descoberta do conhecimento em bases de dados, as técnicas de mineração utilizadas e as estratégias de relacionamento adotadas com clientes. Nesta pesquisa foi investigado o nível de utilização das etapas dos processos de descoberta de conhecimento em bases de dados, as técnicas de

mineração utilizadas, bem como as estratégias de relacionamento adotadas com clientes. Com os resultados da pesquisa foi possível verificar que o nível de utilização das etapas de descoberta de conhecimento é alto e que somente algumas técnicas de mineração de dados são utilizadas de forma uniforme. Com relação às estratégias de relacionamento foi possível verificar que o nível de utilização de técnicas para obtenção de novos clientes e a identificação de novos clientes é superior ao de técnicas de manutenção de novos clientes. O autor conclui que os resultados obtidos no estudo podem ser muito úteis para os gestores no processo de desenvolvimento de estratégias de relacionamento com clientes, através da análise integrada do negócio, informações sobre clientes e modelos quantitativos de análise sobre estas informações. A utilização desta integração de informações pode ser transformada em conhecimento e dessa forma auxiliar a tomada de decisão.

Amoroso (2009), realizou um trabalho de pesquisa com o objetivo de estudar modelos de mineração de dados como um instrumento de aprendizado para o investidor do mercado de ações. Durante a realização da pesquisa foi utilizado o método CRISP – *Cross Industry Standar Process for Data Mining* – para estudar um modelo de séries temporais baseado em redes neurais artificiais. Foram utilizados dados correspondentes a dez anos de cotações de ações de seis empresas, negociadas na Bovespa. O autor conclui o estudo afirmando que existem evidências de que o modelo serve como importante fonte de aprendizado para investidores. O autor sugere para trabalhos futuros o desenvolvimento de modelos para outros tipos de aplicações como compra e venda de moeda ou Bolsa de Mercadorias e Futuros.

O levantamento dos artigos científicos, dissertações e teses além de contribuir com conceitos, definições e metodologia trouxe uma visão mais ampla sobre as pesquisas e os temas deste projeto. Após o levantamento, a leitura e a elaboração de resumos sobre os trabalhos foi possível verificar que:

- não foi encontrado nenhum trabalho que unisse os assuntos *Big Data*, *Data Mining* e previsão de demanda;
- o conceito de IC aparece de forma implícita em muitos trabalhos desenvolvidos no Brasil, mas em nenhum deles existe a expressão “Inteligência Competitiva”;
- a expressão *Big Data* não apareceu em nenhum dos trabalhos realizados no Brasil, mas alguns trabalhos trazem a importância da mineração de dados em grandes bancos de dados;

- em alguns trabalhos foi possível verificar a combinação dos conceitos previsão de demanda e *Data Mining*, e a combinação dos conceitos de *Data Mining* e *Big Data*.

3 METODOLOGIA DA PESQUISA

O presente estudo buscou identificar modelos de previsão de consumo de aço para análise das informações no processo de Inteligência Competitiva. Na busca pelos modelos foi realizada a observação e a manipulação de variáveis internas e externas à empresa, para Gil (2002) estas são características de um estudo científico. O estudo foi de natureza aplicada, já que teve razão de ordem prática a qual buscou “conhecer com vistas a fazer algo de maneira mais eficiente ou eficaz” (GIL, 2002, p. 17) neste caso, conhecer melhor a consumo de aço da empresa para a análise das informações no processo de Inteligência Competitiva.

Em relação aos objetivos o estudo é classificado como exploratório, pois através dele buscou-se a familiarização com o fenômeno para descobrir novas ideias (CERVO, 2002), e descritivo no momento em que visou descobrir a existência de associação entre as variáveis e a descrição das relações através da construção de modelos de previsão (GIL, 2002).

No desenvolvimento do estudo duas abordagens foram adotadas, a primeira buscou analisar fontes internas e externas na busca de dados primários e secundários para definir o problema com mais precisão, identificando e classificando as principais variáveis a serem estudadas na fase quantitativa. Estas características para Malhotra (2010), definem a abordagem como sendo qualitativa. O autor também afirma que “a pesquisa qualitativa proporciona melhor visão e compreensão do problema. Ela investiga o problema com algumas noções preconcebidas sobre os resultados dessa investigação” (MALHOTRA, 2010, p. 122). “A pesquisa qualitativa exploratória é útil quando o pesquisador não conhece as variáveis importantes a examinar” (CRESWELL, 2007, p. 38). Esta é a situação encontrada neste estudo, pois se buscou a identificação das variáveis externas que têm impacto no consumo de aço da Empresa Alfa para posterior construção de um modelo de previsão de consumo através de modelos de Box e Jenkins com função de transferência.

A segunda abordagem buscou avaliar as variáveis para verificar evidências de correlação e construir modelos de previsão de consumo de aço através de modelos Box e Jenkins com função de transferência, características que Malhotra (2010) atribui a uma pesquisa quantitativa. Creswell (2007, p. 38) contribui para esta classificação quando afirma que “se o problema é identificar os fatores que influenciam um resultado, a utilidade de uma intervenção ou a compreensão dos melhores previsores de resultados, então é melhor usar uma técnica quantitativa”. O autor ainda assegura que a abordagem quantitativa “emprega estratégias de investigação (como experimentos, levantamentos e coleta de dados, instrumentos predeterminados que geram dados estatísticos)” (CRESWELL, 2007, p. 35).

Este estudo será operacionalizado através de um estudo de caso. As fontes de informações serão a Empresa Metalúrgica Alfa e os dados existentes no Fenômeno *Big Data*, para a exploração das fontes de dados foi utilizado o processo de *Data Mining*.

O *Data Mining*, mineração em bases de dados da empresa terá como finalidade mapear dados do mercado com influência na consumo de aço para uma posterior prospecção de índices e indicadores financeiros no site do Banco Central do Brasil (BACEN) (2012). O *Data Mining*, segundo Hair Jr. et al. (2005, p. 540), “é o processo de obtenção de informação a partir de grandes bases de dados”, onde, o pesquisador trabalha em um modo exploratório e organizado em torno de uma questão de pesquisa muito geral (HAIR JR. et al., 2005, p. 541).

Estudo de Caso “é uma investigação empírica que investiga um fenômeno contemporâneo em profundidade e em seu contexto de vida real, especialmente quando os limites entre o fenômeno e o contexto não são claramente evidentes” (YIN, 2010, p. 39). Para Cervo (2002, p. 67), “é a pesquisa sobre um determinado indivíduo, família, grupo ou comunidade que seja representativo do seu universo, para analisar aspectos variados de sua vida”. Já Gil (2002, p. 54) afirma que o estudo de caso “consiste no estudo profundo e exaustivo de um ou poucos objetos, de maneira que permita seu amplo e detalhado conhecimento”. Ao analisar a definição de cada autor fica claro que o uso do estudo de caso é indicado quando deseja-se entender um determinado fenômeno com profundidade, envolvendo variáveis importantes de seu contexto.

O estudo da previsão de consumo da empresa Alfa foi limitado ao período de janeiro de 2007 a maio de 2012.

3.1 PESQUISA QUALITATIVA

A pesquisa qualitativa foi exploratória e desenvolvida através de procedimento de mineração de dados. O objetivo foi o de identificar os setores da economia em que as variáveis ou indicadores de mercado tinham alguma relação com a consumo de aço da Alfa.

A identificação das variáveis seguiu duas frentes, a primeira foi a de uma análise do mercado do aço no Brasil, buscando identificar os setores da economia com relação direta com este mercado. A segunda frente teve desenvolvimento na análise da base de dados dos clientes da Alfa, buscou-se então conhecer os setores da economia do qual fazem parte através da classificação dos mesmos pela Classificação Nacional de Atividades Econômicas (CNAE). A coleta de dados para realização da identificação das variáveis foi feita a através da mineração de dados nos sites do Ministério de Minas e Energia e no Instituto Aço Brasil.

3.2 PESQUISA QUANTITATIVA

A pesquisa quantitativa foi dividida em duas etapas:

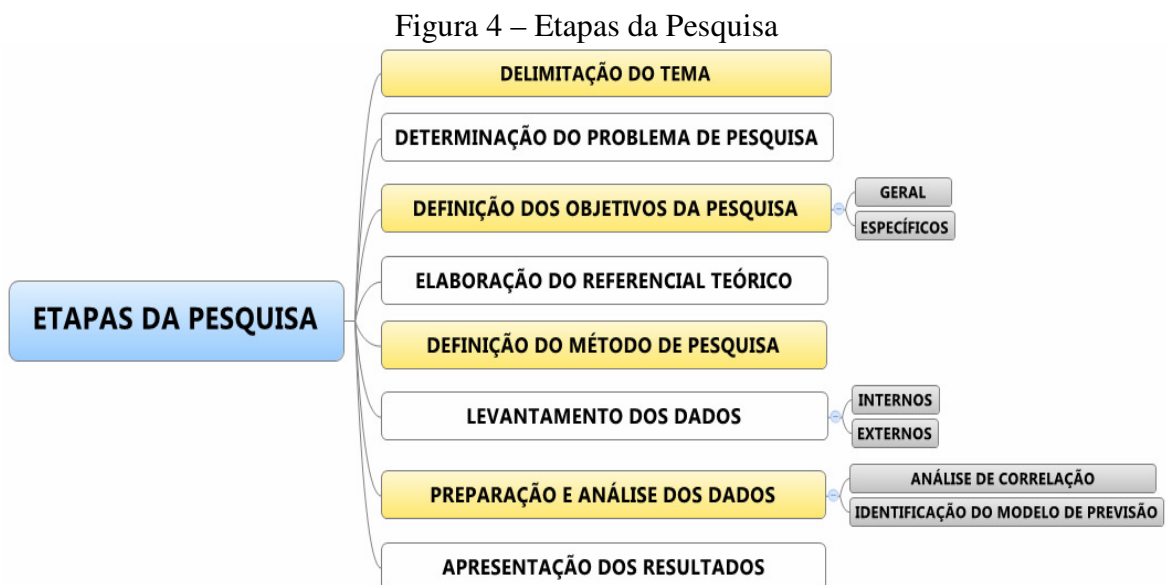
Primeira etapa: foi feita uma análise de correlação entre as variáveis econômicas, pesquisadas na fase qualitativa e a consumo de aço da Alfa no período compreendido entre janeiro de 2007 e maio de 2012.

Segunda etapa: teve como objetivo a identificação dos modelos de previsão de consumo de aço relacionando o consumo com as variáveis econômicas identificadas na etapa anterior. Para construção dos modelos foi utilizada a metodologia univariada e de função de transferência de Box e Jenkins.

A aquisição de dados para a etapa quantitativa foi guiada pelos resultados obtidos na etapa qualitativa de identificação dos setores econômicos relacionados com o consumo de aço. Com os resultados obtidos foi feita uma mineração de dados no Sistema Gerenciador de Séries Temporais (SGS), do Banco Central, em busca de variáveis econômicas relacionadas com os setores da economia pesquisados anteriormente.

3.3 ETAPAS DA PESQUISA

O presente estudo foi realizado conforme as etapas apresentadas pela Figura 4:



Fonte: Elaborado pelo pesquisador

3.4 SELEÇÃO DO CASO DE ESTUDO

Para Yin (2010), a definição do caso de estudo, unidade de análise, é um dos componentes importantes de um projeto de pesquisa que envolva o estudo de caso. Neste trabalho o caso de estudo será a empresa Metalúrgica Alfa Ltda., com matriz em Bento Gonçalves, no Rio Grande do Sul, fundada em 25 de agosto de 1969. Hoje a empresa emprega mais de 1.300 funcionários em suas quatro unidades fabris de Bento Gonçalves no Rio Grande do Sul, uma unidade produtiva no Estado de Pernambuco, uma unidade produtiva no Estado de Goiás e um centro de distribuição na América Central, mais especificamente na Guatemala.

A empresa está subdividida em unidades de negócio focadas em cozinhas de aço para o mercado interno e exportação, sistemas de armazenagem e soluções em logística e armazenagem de produtos, móveis planejados residenciais, móveis exclusivos com foco na produção de produtos especiais para clientes corporativos e exportação e unidade de serviços compreendendo um total de 60.000 m² de área física disponível.

O trabalho terá foco na unidade de sistemas de armazenagem e soluções em logística e armazenagem de produtos. Ele tem como objetivo identificar os modelos mais adequados para a previsão de consumo de aço considerando o impacto de variáveis externas com influência no comportamento do mercado.

A empresa Alfa foi escolhida para realização deste estudo por três motivos:

- importância econômica da empresa para região e comunidade a qual ela pertence;
- importância do setor metalúrgico para a economia local e nacional, importância esta destacada anteriormente pelos dados do mercado do aço no Brasil;
- disponibilidade de dados estratégicos da empresa para realização do estudo.

Em reunião com os diretores da empresa, ficou clara a importância da gestão da consumo de aço para a unidade de sistemas de armazenamento, devido aos custos envolvidos e ao tempo que o aço pode ficar parado em estoque, se não houver demanda do mercado. A solução proposta foi a de se identificar modelos de previsão de consumo de aço para a empresa através da análise de variáveis externas.

3.5 CARACTERIZAÇÃO DO OBJETO DE ESTUDO

Com o objetivo de preservar a confidencialidade das informações específicas da empresa objeto deste estudo, a identificação utilizada para a empresa será de empresa Alfa. Esta necessidade se faz por exigência dos representantes da empresa com o objetivo de preservar as informações divulgadas neste estudo.

3.5.1 Empresa Alfa

A empresa objeto de estudo é do ramo metalúrgico e tem matriz no estado do Rio Grande do Sul, foi fundada no final da década de 60 e iniciou suas atividades em uma área de 80m², com a produção de esquadrias metálicas e outros materiais metálicos do segmento de serralheria como esquadrias e estruturas metálicas a partir da associação de uma família conhecida em sua época por ser talentosa e criativa.

Empreendedora para sua época, já em 1977, contratou pesquisa de mercado para tentar identificar as oportunidades de negócio. Como resultado iniciou a produção de móveis de aço e já investia na aquisição de uma metalúrgica local para obter o maquinário mínimo para iniciar a produção em escala.

A empresa, sempre inovando, criou a primeira máquina do Brasil de tornar a fórmica fosca. Já, em 1995, colhia prêmios nacionais de distinção com a cozinha de aço mais moderna do Brasil, fruto da apropriação constante de novas tecnologias e criação de soluções próprias e personalizadas.

A empresa também foi à primeira do país a produzir industrialmente dormitórios em *Medium Density Fiberboard* (MDF), o que seguiu para as cozinhas que também seriam fabricadas com esse material, partindo assim para nova análise de mercado, seguida pela estruturação de sua rede de lojas exclusivas, visando à comercialização direta ao consumidor.

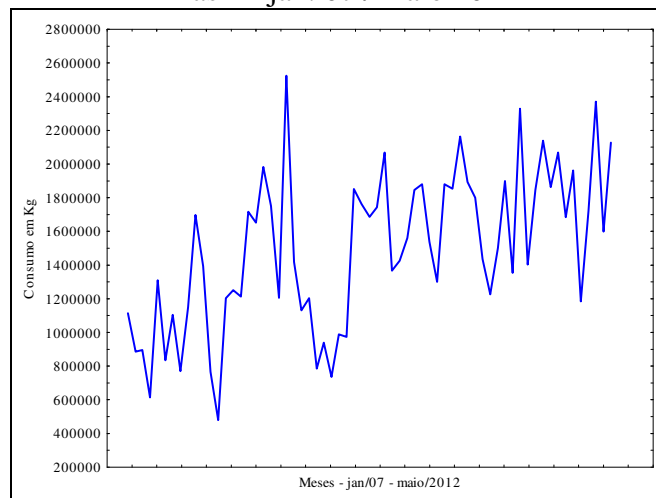
Atualmente, a empresa denomina-se Alfa Sociedade Anônima (S.A.), emprega cerca de 1.300 funcionários em suas quatro unidades fabris no Rio Grande do Sul, uma unidade produtiva no Estado de Pernambuco, uma unidade produtiva no Estado de Goiás e um centro de distribuição na América Central. Está subdividida em unidades de negócio focadas em cozinhas de aço para o mercado interno e exportação, sistemas de armazenagem e soluções em logística e armazenagem de produtos, móveis planejados residenciais, móveis exclusivos com foco na produção de produtos especiais para clientes corporativos e exportação e unidade de serviços compreendendo um total de 60.000 m² de área física disponível.

Nesse cenário, a Alfa investe constantemente em qualificação para o desenvolvimento de líderes com perfil voltado para a interpretação estratégica e foco na inteligência empresarial, mostrando-se assim, preocupada com a longevidade da empresa.

Nesse estudo, concentram-se esforços na unidade de sistemas de armazenagem, esta especializada em planejar e desenvolver modernas soluções para estocagem e movimentação de materiais tendo como linha produtos porta *Pallets*, *Drive in*, *Cantiliver*, *Racks* metálicos e *Intainer*, Divisórias industriais, *Drive In* Dinâmico, Porta Bobinas, *Push Backs*, *Pallets* leves, autoportantes, mezaninos e passarelas, *Drive In* Carro Satélite e *Racks* metálicos *Multiblock* que, em volume de vendas, no ano de 2011, representou 43% do faturamento do grupo.

A principal matéria-prima utilizada pela empresa, na unidade de armazenamento, é o aço. No Gráfico 2 é possível observar o consumo mensal de aço em kg no período entre janeiro de 2007 a maio de 2012.

Gráfico 2 – Consumo mensal de aço – Empresa Alfa Brasil – jan. 07 / maio 2012



Fonte: Elaborado pelo pesquisador com base nos dados da empresa Alfa (2012)

A empresa faz aquisições periódicas de grande quantidade de aço para produção, mas por se tratar de uma *Commodity* a empresa não consegue negociar melhores preços com seus fornecedores na busca de uma vantagem competitiva. Devido à variabilidade do consumo mensal de aço, como é possível verificar no Gráfico 2, a empresa tem dificuldade de prever a demanda de consumo. Uma das soluções pensadas pelos gestores da Alfa para obter alguma vantagem com relação à aquisição do aço consiste em melhorar a previsão da demanda do consumo de aço pela empresa na busca da redução de seus estoques o que acarretará a redução do capital empenhado em estoque.

Em reunião realizada em abril de 2012, com diretores da empresa, foi demonstrado interesse no desenvolvimento de uma pesquisa para a elaboração de um modelo de previsão de consumo de aço. Na mesma reunião os diretores afirmaram que está em estudo um projeto para a implantação de uma unidade de Inteligência Competitiva na empresa e que a construção do modelo de previsão de consumo de aço seria mais um passo em direção à implantação desta unidade. Para isso a empresa disponibilizaria todas as informações necessárias para o projeto, assim como a devida autorização para o estudo desde que a verdadeira razão social fosse alterada para fins de publicações.

3.6 COLETA DE DADOS

A coleta de dados para o desenvolvimento deste estudo seguiu três frentes:

1. Coleta de informações sobre o mercado do aço no Brasil;
2. Coleta de informações na empresa Alfa;
3. Coleta de informações junto ao Sistema gerenciador de séries temporais do BACEN (2012);

A coleta de dados sobre o mercado do aço no Brasil teve como objetivo reconhecer as principais características deste mercado para posterior identificação dos setores da economia que podem influenciar o consumo de aço da empresa Alfa.

A coleta de informações sobre a empresa Alfa teve como finalidade a construção de dois bancos de dados. Um teve a finalidade de identificação e classificação dos clientes da empresa com o propósito de identificar os setores da economia aos quais eles pertencem, o segundo banco teve como finalidade a análise de correlação e a identificação do modelo de previsão de consumo de aço da empresa Alfa.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Neste capítulo serão apresentadas a análise e discussão dos dados obtidos durante o desenvolvimento do estudo. Essa apresentação foi dividida em duas etapas:

- Etapa qualitativa: onde será apresentado o conjunto de informações coletadas na pesquisa qualitativa, assim como o tratamento dessas informações e os resultados obtidos;
- Etapa quantitativa: onde será apresentada a análise de correlação, o processo de identificação dos modelos de previsão e os resultados obtidos.

4.1 ETAPA QUALITATIVA

Na etapa qualitativa da pesquisa foi feito um levantamento de informações necessárias para identificação dos setores econômicos e os seus respectivos indicadores que tinham alguma relação com o consumo de aço da empresa Alfa. Para tanto, o procedimento foi dividido em:

- identificação de informações sobre o Mercado do Aço no Brasil;
- coleta de dados da empresa Alfa;
- análise dos resultados obtidos.

4.1.1 Mercado do Aço Brasileiro

Para estudar o consumo de aço da Empresa Alfa e identificar no mercado variáveis ou indicadores financeiros que possam ter influência sobre esta, foi feito um levantamento das características do mercado do aço brasileiro. Os dados obtidos são referentes à consolidação do mercado até o ano de 2011 e à situação da produção de aço até maio de 2012. As informações foram obtidas junto ao Instituto Aço Brasil (IABR) e ao Ministério de Minas e Energia (MME).

Após a crise de 2008 o mercado mundial do aço vem se recuperando, segundo dados do MME (2011, p. 38), a produção mundial de aço bruto entre os anos de 2006 e 2010 foram respectivamente 1.246,7, 1.346,1, 1.328,5, 1.230,0, 1.413,5 milhões de toneladas, o que representa um crescimento de 13,38 % no período, e um crescimento de 14,92% após a crise de 2008.

O Brasil, segundo MME (2011, p. 35), ocupa a 9ª posição no ranking dos países produtores e 12ª posição no ranking mundial de países exportadores, sendo o 5º maior exportador líquido (exportação – importação), exportando para 100 países, indiretamente o país exporta 2,9 milhões de toneladas em aço contido em outros bens.

Tabela 5 – Produção Mundial de Aço Bruto – Unid. 10⁶ t

PAÍSES	2006	2007	2008	2009	2010
China	419,1	489,3	500,3	573,6	626,7
Japão	116,2	120,2	118,7	87,5	109,6
EUA	98,6	98,1	91,4	58,2	80,5
Índia	49,5	53,5	57,8	62,8	68,3
Rússia	70,8	72,4	68,5	60,0	66,9
Coréia do Sul	48,5	51,5	53,6	48,6	58,4
República Federal da Alemanha	47,2	48,6	45,8	32,7	43,8
Ucrânia	40,9	42,8	37,3	29,9	33,4
Brasil	30,9	33,8	33,7	26,5	32,9
Turquia	23,3	25,8	26,8	25,3	29,1

Fontes: Anuário Estatístico do Setor Metalúrgico-SGM/MME;

Dados obtidos ao IABR (2012) informam que o parque brasileiro produtor de aço é composto por 29 usinas administradas por 11 grupos empresariais, onde a capacidade instalada é de 47,8 milhões de toneladas por ano de aço bruto com produção anual de 35,2 milhões de toneladas. Já o consumo aparente de aço é de 25 milhões de toneladas, gerando um saldo comercial de 3,9 bilhões de dólares que representou 13% do saldo comercial do país.

O Ministério de Minas e Energia apresenta a Tabela 6 contendo a produção de aço brasileira por empresas no período de 2006 a 2010.

Tabela 6 – Produção Brasileira de Aço Bruto por Empresa - Unid. 10³ t

EMPRESAS	2006	2007	2008	2009	2010
Gerdau *	7.698	8.111	8.711	6.105	8.177
Usiminas	8.770	8.675	8.022	5.637	7.298
ArcelorMittal Tubarão	5.136	5.692	6.177	5.334	5.956
CSN	3.499	5.323	4.985	4.375	4.902
ArcelorMittal Aços Longos	3.569	3.739	3.502	3.171	3.394
Votorantim Siderurgia	638	624	712	617	1.041
Aperam	810	797	770	607	771
V & M do Brasil	659	686	655	387	573
CSA	-	-	-	-	458
Sinobras	-	-	42	181	239
Villares <i>Metals</i>	122	135	140	92	119
Total	30.901	33.782	33.716	26.506	32.928

Fontes: Anuário Estatístico do Setor Metalúrgico-SGM/MME;

Notas / Notes: Corresponde à produção de aço em lingotes + produtos de lingotamento contínuo + aço para fundição.

* As estatísticas da Aços Villares estão incorporadas às da Gerdau.

Uma variável que preocupa o setor é o consumo *per capita* de aço, pois em 1980 o consumo *per capita* brasileiro era 70% superior ao chinês, hoje o consumo chinês é 117,76 % superior ao brasileiro (MME, 2011, p. 36). O consumo *per capita* brasileiro em 2011 foi de 145 quilos de aço bruto por habitante. Um histórico do consumo *per capita* pode ser observado na Tabela 7 que traz um comparativo do consumo aparente de aço no período de 1970 a 2010.

Tabela 7 – Consumo Aparente Per Capita de Aço – Unid. Kg/hab.

METAL	1970	1975	1980	1985	1990	1995	2000	2008	2009	2010
Aço (bruto)	50,1	91,6	113	85,7	71,0	86,0	105	143	109	152
Aço Inox	n.d.	n.d.	0,49	0,56	0,55	0,88	1,1	1,9	1,4	1,9

Fontes: Anuário Estatístico do Setor Metalúrgico-SGM/MME; IABR; ABAL; ABC; ICZ; SNIEE; ABRAFE; Grupo Inox.

Conforme o IABR (2012) a produção brasileira de aço bruto até maio de 2012 teve uma queda de 11,7% com relação ao mesmo período de 2011, em contrapartida a produção de aços laminados em maio de 2012 apresentou um crescimento de 1,1% em comparação a maio de 2011.

Com relação às vendas internas o IABR (2012) informa que as vendas acumuladas em 2012 mostram um crescimento de 0,8% em relação ao mesmo período de 2011. As exportações brasileiras até maio de 2012 têm apresentado um declínio em 5,2% em volume e de 8,9% em valor, quando comparados ao mesmo período do ano anterior (IABR, 2012).

Já as importações brasileiras, até maio de 2012, têm registrado um crescimento de 14,2% em relação ao mesmo período de 2011, totalizando 1,6 milhão de toneladas de produtos siderúrgicos importados no ano. O consumo aparente de produtos siderúrgicos até maio de 2012 tem apresentado um crescimento de 0,4% em relação ao mesmo período de 2011, o que totalizaria um crescimento de 1,6% no ano de 2012 com relação a 2011.

A Tabela 8 apresenta a distribuição de vendas de produtos siderúrgicos no Brasil conforme dados obtidos no Anuário Estatístico 2011 do Ministério de Minas.

Tabela 8 – Distribuição Setorial das Vendas Internas de Produtos Siderúrgicos - Unid. 10³ t

SETORES	2006	2007	2008	2009	2010
Automobilístico	1.280	1.552	1.607	1.479	1.554
Autopeça (Inclusive Carrocerias)	2.001	2.292	2.384	1.812	2.375
Ferrovário	91	76	74	27	62
Naval	35	54	85	48	79

(continua)

SETORES	2006	2007	2008	2009	2010
Bicicletas e Motocicletas	13	14	3	8	15
Agrícola e Rodoviário	252	337	417	227	393
Máquinas e Equipamentos Eletro Eletrônicos e de Medidas e Controle	518	540	579	392	543
Máquinas e Equipamentos Industriais	509	600	565	344	583
Construção Civil	2.171	2.720	3.475	2.566	3.203
Utilidades Domésticas e Comerciais	467	549	520	466	510
Embalagens e Recipientes	756	733	710	605	653
Cutelaria	11	16	16	7	17
Parafusos e Rebites (Excl. p/ Ind. Automobilística)	104	105	156	145	239
Derivados de Arames	5	52	52	32	52
Relaminação (Fora do Parque)	613	793	786	496	694
Trefilaria de Barras	56	71	54	44	63
Trefilaria de Arames	784	806	856	674	815
Forjaria Matriz Fechada	499	528	560	287	436
Forjarias (Outras)	26	42	58	18	28
Perfis Conformados a Frio	153	173	168	91	64
Perfis Soldados	14	32	28	-	-
Tubos com Costura de Pequeno Diâmetro	817	1.194	1.061	661	839
Tubos com Costura de Grande Diâmetro	327	517	494	212	255
Distribuidores e Revendedores	5.621	6.359	6.776	5.445	6.942
Outros	408	395	309	259	302
Total	17.531	20.550	21.793	16.345	20.716
Relaminação (p/ Usinas)	1.240	339	486	74	203

Fontes: Anuário Estatístico do Setor Metalúrgico-SGM/MME;

Com base nas informações da Tabela 8 e segundo o IABR (2012), os principais setores consumidores de aço no Brasil são: construção civil, automotivo, bens de capital, máquinas e equipamentos e utilidades domésticas e comerciais.

4.1.2 Coleta de informações na Empresa Alfa

Com o objetivo de construir uma base de dados sobre os clientes da empresa Alfa e sobre o consumo de aço foi solicitado à empresa o acesso às seguintes informações

- tipo de Aço utilizado pela empresa Alfa;
- consumo mensal de Aço;
- pedidos com nomes de clientes;
- solicitação de orçamento com a situação após o orçamento;
- relação de Operadores Logísticos;
- relação de Clientes por faturamento;

- valor pago pelo aço em cada compra.

Após o recebimento das informações da empresa, iniciou-se o processo de avaliação, formatação e construção dos bancos de dados para a pesquisa.

A primeira etapa do trabalho consistiu em uma avaliação dos dados enviados pela empresa Alfa, onde foram realizados os seguintes procedimentos:

- 1º) identificação e exclusão de dados duplicados;
- 2º) identificação de falta de informações;
- 3º) agrupamento de diversas bases de dados em uma única base;
- 4º) exclusão de informações sem relevância para o estudo;
- 5º) padronização de nomes de clientes;
- 6ª) construção de dois novos bancos de dados sendo:
 - Banco de dados com a relação de clientes e seus respectivos pedidos a partir de janeiro de 2007 até 18 de julho de 2012.
 - Banco de dados com o consumo mensal de aço da empresa Alfa, solicitações de orçamentos, entradas de pedidos e a variação entre valor pedido e valor orçado.

O primeiro banco de dados teve como finalidade análise e identificação dos possíveis fatores externos que podem influenciar o consumo de aço da empresa Alfa através da classificação dos clientes conforme tabela de Classificação Nacional de Atividade Econômica (CNAE). O segundo banco de dados teve como objetivo a análise de correlação entre o consumo de aço e as variáveis externas. As estruturas dos bancos de dados são apresentadas nas Tabelas 9 e 10.

Tabela 9 – Estrutura da Base de Dados Relação de Clientes

NOME DO CAMPO	DESCRIÇÃO
V_Ano	Identificação do ano do pedido
V_Cnpj	Identificação do CNPJ do Cliente
V_R_Social	Identificação da Razão Social do Cliente
V_Valor	Identificação do Valor do Pedido

Fonte: Dados elaborados pelo pesquisador

Tabela 10 – Estrutura da Base de Dados Consumo de Aço

(continua)

NOME DO CAMPO	DESCRIÇÃO
V_mes	Identifica o mês e o ano
V_Consumo	Identifica o consumo mensal de aço em Kg
V_Valor_orçado	Identifica o valor do orçamento
V_valor_contratado	Identifica o valor do pedido

(conclusão)

NOME DO CAMPO	DESCRIÇÃO
V_Difer_perc_valor	Identifica a diferença percentual entre o valor orçado e o valor contratado
V_Quantidade	Identifica a quantidade orçada
V_quant_contratada	Identifica a quantidade contratada no pedido
V_difer_perc_quant	Identifica a diferença percentual entre a quantidade orçada e a quantidade contratada

Fonte: Dados elaborados pelo pesquisador

O banco de dados com a relação dos clientes era composto por 10.557 negociações no período de janeiro de 2007 a maio de 2012.

Para realizar a análise do banco de dados foi feita uma amostragem probabilística aleatória simples de uma população finita (PEREIRA, 1979), considerando-se um erro de 5% e uma confiança de 95%, com isto foi obtida uma amostra com 360 clientes.

Com a amostra definida foi realizada uma classificação dos clientes conforme a atividade econômica baseada na tabela CNAE (IBGE, 2012) classificando-os em dois níveis: segundo a seção e com base na divisão, Tabelas 11 e 12 respectivamente. A classificação foi realizada através de uma consulta do Cadastro Nacional de Pessoa Jurídica (CNPJ) no site da Receita Federal para identificar o código e descrição da atividade econômica principal. O código obtido na Receita Federal foi utilizado para identificar os setores da economia e as atividades econômicas dos clientes da empresa Alfa.

Tabela 11 – Classificação dos clientes por seção conforme a tabela CNAE.

(continua)

SEÇÃO	FREQUÊNCIA	%	DESCRIÇÃO DA SEÇÃO
C	141	39,2	Indústrias de Transformação
G	118	32,8	Comércio; Reparação de Veículos Automotores e Motocicletas
H	21	5,8	Transporte, Armazenagem e Correio
PF	15	4,2	Pessoa física
BA	11	3,1	CNPJ baixados
SS	7	1,9	CNPJ suspenso
F	6	1,7	Construção
EX	5	1,4	Exportação
K	5	1,4	Atividades financeiras, de seguros e serviços relacionados
B	4	1,1	Indústrias extrativas
J	4	1,1	Informação e comunicação
N	4	1,1	Atividades administrativas e serviços complementares
M	3	,8	Atividades profissionais, científicas e técnicas
Q	3	,8	Saúde humana e serviços sociais
A	2	,6	Agricultura, pecuária, produção florestal, pesca e aquíicultura

(conclusão)

SEÇÃO	FREQUÊNCIA	%	DESCRIÇÃO DA SEÇÃO
I	2	,6	Alojamento e alimentação
O	2	,6	Administração pública, defesa e seguridade social
P	2	,6	Educação
S	2	,6	Outras atividades de serviços
D	1	,3	Eletricidade e gás
E	1	,3	Água, esgoto, atividades de gestão de resíduos e descontaminação
L	1	,3	Atividades imobiliárias
Total	360	100,0	

Fonte: Dados elaborados pelo pesquisador com base na tabela CNAE (IBGE, 2012)

Tabela 12 – Classificação dos clientes por divisão conforme tabela CNAE.

(continua)

DIVISÃO	FREQUÊNCIA	%	DESCRIÇÃO DA DIVISÃO
46	54	15,0	Comércio por atacado, exceto veículos automotores e motocicletas
47	50	13,9	Comércio varejista
10	18	5,0	Fabricação de produtos alimentícios
22	15	4,2	Fabricação de produtos de borracha e de material plástico
45	14	3,9	Comércio e reparação de veículos automotores e motocicletas
49	13	3,6	Transporte terrestre
25	12	3,3	Fabricação de produtos de metal, exceto máquinas e equipamentos
29	11	3,1	Fabricação de veículos automotores, reboques e carrocerias
20	10	2,8	Fabricação de produtos químicos
28	10	2,8	Fabricação de máquinas e equipamentos
26	9	2,5	Fabricação de equipamentos de informática, produtos eletrônicos e ópticos
31	9	2,5	Fabricação de móveis
77	3	,8	Aluguéis não-imobiliários e gestão de ativos intangíveis não-financeiros
15	7	1,9	Preparação de couros e fabricação de artefatos de couro, artigos para viagem e calçados
17	7	1,9	Fabricação de celulose, papel e produtos de papel
32	7	1,9	Fabricação de produtos diversos
52	7	1,9	Armazenamento e atividades auxiliares dos transportes
14	6	1,7	Confecção de artigos do vestuário e acessórios
64	5	1,4	Atividades de serviços financeiros
42	4	1,1	Obras de infraestrutura
13	3	,8	Fabricação de produtos têxteis
23	3	,8	Fabricação de produtos de minerais não-metálicos
24	3	,8	Metalurgia
33	3	,8	Manutenção, reparação e instalação de máquinas e equipamentos
86	3	,8	Atividades de atenção à saúde humana

(conclusão)

DIVISÃO	FREQUÊNCIA	%	DESCRIÇÃO DA DIVISÃO
1	2	,6	Agricultura, pecuária e serviços relacionados
7	2	,6	Extração de minerais metálicos
27	2	,6	Fabricação de máquinas, aparelhos e materiais elétricos
30	2	,6	Fabricação de outros equipamentos de transporte, exceto veículos automotores
56	2	,6	Alimentação
58	2	,6	Edição e edição integrada à impressão
61	2	,6	Telecomunicações
84	2	,6	Administração pública, defesa e seguridade social
85	2	,6	Educação
94	2	,6	Atividades de organizações associativas
5	1	,3	Extração de carvão mineral
6	1	,3	Extração de petróleo e gás natural
11	1	,3	Fabricação de bebidas
16	1	,3	Fabricação de produtos de madeira
18	1	,3	Impressão e reprodução de gravações
19	1	,3	Fabricação e estoque de produtos derivados do petróleo e de biocombustíveis
35	1	,3	Eletricidade, gás e outras utilidades
38	1	,3	Coleta, tratamento e disposição de resíduos; recuperação de materiais
41	1	,3	Construção de edifícios
43	1	,3	Serviços especializados para construção
51	1	,3	Transporte aéreo
68	1	,3	Atividades imobiliárias
71	1	,3	Serviços de arquitetura e engenharia; testes e análises técnicas
72	1	,3	Pesquisa e desenvolvimento científico
74	1	,3	Outras atividades profissionais, científicas e técnicas
81	1	,3	Serviços para edifícios e atividades paisagísticas
Total	360	100,0	

Fonte: Dados elaborados pelo pesquisador com base na tabela CNAE (IBGE, 2012)

Com as informações obtidas sobre o mercado do aço no Brasil, sobre a classificação dos clientes da empresa Alfa por atividade econômica conforme a tabela CNAE, e seguindo indicação de Mayer (1990), sobre os indicadores econômicos comumente utilizados nas previsões de venda, foi possível construir a Tabela 13 que traz uma visão sobre as informações obtidas nesta fase da pesquisa. Esta tem como finalidade orientar o levantamento de indicadores econômicos para o estudo de correlação e posteriormente a identificação do modelo de previsão de consumo.

Tabela 13 – Identificação dos Setores Econômicos e Indicadores

SETORES CONSUMIDORES DE AÇO NO BRASIL - IBR (2012)	INDICADORES ECONÔMICOS USADOS NA PREVISÃO DE VENDAS - MAYER (1990)	CLASSIFICAÇÃO DOS CLIENTES POR SEÇÃO - CNAE (2012)
<ul style="list-style-type: none"> • construção civil; • automotivo; • bens de capital; • máquinas e equipamentos; • utilidades domésticas e comerciais. 	<ul style="list-style-type: none"> • renda per capita; • produção de automóveis; • renda agrícola; • nível de emprego; • renda nacional bruta; • preços ao consumidor; • preços de mercadorias no atacado; • depósitos bancários; • produção de aço; • produção industrial. 	<ul style="list-style-type: none"> • Indústrias de Transformação • Comércio; Reparação de Veículos Automotores e Motocicletas • Transporte, Armazenagem e Correio • Pessoa física • Construção • Exportação • Atividades financeiras, de seguros e serviços relacionados • Indústrias extrativas • Informação e comunicação • Atividades administrativas e serviços complementares • Atividades profissionais, científicas e técnicas • Saúde humana e serviços sociais • Agricultura, pecuária, produção florestal, pesca e aquicultura • Alojamento e alimentação • Administração pública, defesa e seguridade social • Educação • Outras atividades de serviços • Eletricidade e gás • Água, esgoto, atividades de gestão de resíduos e descontaminação • Atividades imobiliárias

Fonte: Dados elaborados pelo pesquisador

4.2 ETAPA QUANTITAVA

A etapa quantitativa da pesquisa teve por finalidade avaliar a relação entre às variáveis do mercado através de uma análise de correlação e identificar o modelo de previsão de consumo de aço da empresa Alfa através da metodologia de Box e Jenkins com função de transferência. Esta etapa ficou dividida em:

- análise de correlação;

- ciclo de Box e Jenkins para construção do modelo Univariado;
- construção do Modelo com Função de Transferência;
- análise dos resultados obtidos.

4.2.1 Análise de Correlação

Tendo como referência as informações obtidas na etapa qualitativa e orientação dos dados da Tabela 13 realizou-se um processo manual de mineração de dados no SGS (BACEN, 2012) para identificar variáveis que poderiam ter alguma relação com o consumo de aço da empresa Alfa.

Para identificar a relação entre as variáveis de mercado obtidas através da mineração de dados no SGS, do Banco Central, e o consumo de aço da empresa Alfa, foi realizada uma análise de correlação. Esta tem como propósito saber como se comporta uma variável conhecendo o comportamento de outras variáveis relacionadas (SAMPIERI; COLLADO; LUCIO, 2001, p. 63).

A correlação é medida pela estatística que resume a força e a direção da associação entre duas variáveis (MALHOTRA, 2010, p. 419).

Em um primeiro momento realizou-se uma análise de correlação com todas as variáveis envolvidas e a variável consumo de aço da empresa Alfa, não se considerando os supostos paramétricos, pois o objetivo era identificar o tipo de correlação existente para diminuir o número de variáveis.

O critério utilizado foi a classificação da força de uma correção apresentada por Santos (2007) na Tabela 14, as variáveis com correlação no intervalo de -0,6 e 0,6 foram retiradas do estudo.

Tabela 14 – Avaliação da Correlação

COEFICIENTE DE CORRELAÇÃO	CORRELAÇÃO
$P = 1$	Perfeita Positiva
$0,8 \leq p < 1$	Forte Positiva
$0,5 \leq p < 0,8$	Moderada Positiva
$0,1 \leq p < 0,5$	Fraca Positiva
$0 < p < 0,1$	Ínfima Positiva
0	Nula
$-0,1 < p < 0$	Ínfima Negativa
$-0,5 < p \leq -0,1$	Fraca Negativa
$-0,8 < p \leq -0,5$	Moderada Negativa
$-1 < p \leq -0,8$	Forte Negativa
$P = -1$	Perfeita Negativa

Fonte: Santos (2007) reelaborada pelo pesquisador

Após a primeira análise e a retirada do estudo dos indicadores econômicos, com coeficiente de correlação fora dos do intervalo estipulado foram verificados os supostos paramétricos, estes são requisitos que devem ser respeitados para que os testes paramétricos possam ser aplicados, caso os supostos não sejam respeitados, os testes aplicados devem ser não paramétricos (HAIR JR. et al., 2005).

Os supostos paramétricos são:

- a. a variável dependente é quantitativa contínua, medida pelo menos em uma escala de intervalo;
- b. a amostra é maior que trinta elementos;
- c. a população é normalmente distribuída;
- d. há homoscedasticidade entre os grupos de variáveis.

Ao avaliar o conjunto de dados foi possível verificar que os supostos a e b são respeitados pelo conjunto, pois as variáveis estudadas são contínuas e o tamanho da amostra é maior que 30.

Para verificar a normalidade dos dados foi aplicado o teste de *Kolmogorov-Smirnov* (HAIR JR. et al., 2005 p. 78) já a homoscedasticidade foi avaliada pelo teste *Levene* (HAIR JR. et al., 2005, p.79). Este indicou que esta condição não era respeitada, essa etapa do estudo fica restrita a uma análise de correlação e, para este, o resultado do teste de *Levene* (HAIR JR. et al., 2005, p.79) não terá influência, mas na identificação dos modelos de previsão este resultado foi levado em consideração.

Na avaliação do suposto paramétrico de normalidade da população, onde se avaliou se os dados seguem uma distribuição normal, testou-se a hipótese nula. Esta afirmava que a variável é normalmente distribuída contra a hipótese alternativa que a variável não segue uma distribuição normal. Através dos resultados obtidos na aplicação do teste *Kolmogorov-Smirnov* foi possível verificar que a significância foi maior que 0,05 para todas as variáveis o que nos impede de rejeitar a hipótese nula, ou seja, os dados seguem uma distribuição normal, conforme resultados apresentados na Tabela 15.

Tabela 15 – Teste de Kolmogorov Smirnov

(continua)		
VARIÁVEIS	KOLMOGOROV SMIRNOV Z	SIG.
Consumo de aço (kg)	,730	,661
Produção de caminhões (unidades)	,567	,905
Produção de tratores de rodas (unidades)	,560	,913
Produção de retroescavadeiras (unidades)	,743	,639

VARIÁVEIS	(conclusão)	
	KOLMOGOROV SMIRNOV Z	SIG.
Produção de máquinas agrícolas (unidades)	,830	,496
Produção de derivados de petróleo - Gás natural (barris/dia) 10 ³	,716	,685
Consumo de derivados de petróleo – Óleo diesel (barris/dia) 10 ³	,567	,905
Consumo de energia elétrica - Brasil - Industrial (GWh)	,943	,336
Consumo de energia elétrica - Brasil - Total (GWh)	,956	,320
Índice volume de vendas no varejo - Combustíveis e lubrificantes (índice)	,671	,758
Índice volume de vendas no varejo – Automóveis, motocicletas, partes e peças – Brasil (índice)	,730	,661
PIB mensal - Valores correntes (R\$ milhões)	,824	,505
Vendas de veículos pelas concessionárias - Comerciais leves (unidades)	,603	,861
Vendas de veículos pelas concessionárias – Caminhões (unidades)	,738	,647
Vendas de veículos pelas concessionárias – Ônibus (unidades)	,514	,954
Vendas de veículos pelas concessionárias - Total (unidades)	,525	,946
Indicadores da produção (2002=100) – Geral (índice)	,831	,495
Indicadores da produção (2002=100) - Extrativa mineral (índice)	,756	,618
Indicadores da produção (2002=100) (índice)	,724	,671
Indicadores da produção (2002=100) - Por categoria de uso – Bens de capital (índice)	1,001	,269
Indicadores da produção (2002=100) - Por categoria de uso – Bens de consumo (não-duráveis e semiduráveis) (índice)	,562	,910
Insumos da construção civil (índice)	,707	,699
Material de construção (índice)	,421	,994
Cesta básica (u.m.c)	1,346	,053
Endividamento das famílias com o Sistema Financeiro Nacional em relação à renda acumulada dos últimos doze meses (índice)	,633	,818

Fonte: Dados elaborados pelo pesquisador

Tendo realizado a avaliação dos supostos paramétricos foi possível realizar a análise de correlação entre as variáveis externas e o consumo de aço da empresa Alfa.

Tabela 16 – Análise de Correlação entre a variável consumo de aço e as variáveis externas através do Coeficiente de Pearson.

VARIÁVEIS	(continua)		
	PEARSON	SIG	N
Produção de caminhões	,646**	,000	65
Produção de tratores de rodas	,632**	,000	65
Produção de retroscavadeiras	,686**	,000	65
Produção de máquinas agrícolas (total)	,669**	,000	65
Produção de derivados de petróleo – Gás natural	,608**	,000	65
Consumo de derivados de petróleo – Óleo diesel	,740**	,000	65
Consumo de energia elétrica - Brasil - Industrial	,609**	,000	65
Consumo de energia elétrica - Brasil - Total	,645**	,000	65

(conclusão)

VARIÁVEIS	CONSUMO		
	PEARSON	SIG	N
Índice volume de vendas no varejo – Combustíveis e lubrificantes – Brasil	,707**	,000	65
Índice volume de vendas no varejo - Automóveis, motocicletas, partes e peças - Brasil	,614**	,000	65
PIB mensal - Valores correntes (R\$ milhões)	,668**	,000	65
Vendas de veículos pelas concessionárias - Comerciais leves	,643**	,000	65
Vendas de veículos pelas concessionárias - Caminhões	,661**	,000	65
Vendas de veículos pelas concessionárias – Ônibus	,657**	,000	65
Vendas de veículos pelas concessionárias - Total	,601**	,000	65
Indicadores da produção (2002=100) - Geral	,655**	,000	65
Indicadores da produção (2002=100) - Extrativa mineral	,613**	,000	65
Indicadores da produção (2002=100) – Indústria de transformação	,651**	,000	65
Indicadores da produção (2002=100) - Por categoria de uso - Bens de capital	,680**	,000	65
Indicadores da produção (2002=100) - Por categoria de uso - Bens de consumo (não-duráveis e semiduráveis)	,662**	,000	65
Insumos da construção civil (2002=100)	,797**	,000	65
Material de construção	,715**	,000	61
Cesta básica	,606**	,000	65
Endividamento das famílias com o Sistema Financeiro Nacional em relação à renda acumulada dos últimos doze meses	,620**	,000	65

Fonte: Dados elaborados pelo pesquisador

Analisando a Tabela 16 é possível verificar que todas as variáveis apresentaram um coeficiente de correlação superior a 0,6, o que para Santos (2007) representa uma correlação moderada positiva. Este resultado também confirma os resultados da análise qualitativa, uma vez que a correlação entre as variáveis pode ser vista como um indicativo de correlação entre os setores econômicos aos quais as variáveis pertencem.

4.2.2 Análise dos resultados obtidos na análise de correlação

Com base nos resultados da análise de correlação, apresentados na Tabela 16, foi possível agrupar as variáveis conforme a suas similaridades e com estes grupos construir um modelo que representa a relação entre o mercado e o consumo de aço da empresa Alfa.

A Tabela 17 traz o conjunto de variáveis agrupadas conforme a sua similaridade e a classificação, em ordem decrescente, do grau de correlações existentes entre as variáveis econômicas e o consumo de aço.

Tabela 17 – Classificação das variáveis conforme similaridade e sua correlação.

GRUPO	VARIÁVEIS	PEARSON
Produção de Veículos Pesados	Produção de retroescavadeiras	,686**
	Produção de máquinas agrícolas (total)	,669**
	Produção de caminhões	,646**
Prod. E Consumo de derivados de petróleo e energia elétrica	Consumo de derivados de petróleo – Óleo diesel	,740**
	Consumo de energia elétrica – Brasil – Total	,645**
	Consumo de energia elétrica – Brasil – Industrial	,609**
Índice de vendas	Índice volume de vendas no varejo - Combustíveis e lubrificantes – Brasil	,707**
	Vendas de veículos pelas concessionárias – Caminhões	,661**
	Vendas de veículos pelas concessionárias - Ônibus	,657**
Indicadores da produção	Indicadores da produção (2002=100) - Por categoria de uso - Bens de capital	,680**
	PIB mensal - Valores correntes (R\$ milhões)	,668**
	Indicadores da produção (2002=100) - Por categoria de uso - Bens de consumo (não-duráveis e semiduráveis)	,662**
Construção civil	Insumos da construção civil (2002=100)	,797**
	Material de construção	,715**
Social	Endividamento das famílias com o Sistema Financeiro Nacional em relação à renda acumulada dos últimos doze meses	,620**
	Cesta básica	,606**

Fonte: Dados elaborados pelo pesquisador

O modelo apresentado na Figura 5 representa uma visão das relações existentes entre o mercado e o consumo de aço da empresa Alfa na época em que o estudo foi realizado. É importante destacar que este modelo é temporal, ou seja, só representa o momento do estudo e pode sofrer alterações se o estudo for realizado em outros momentos.

A função do modelo apresentado na Figura 5 é de trazer um panorama das relações entre o mercado e o consumo de aço, e tem como objetivo melhor orientar a vigilância interna e externa da empresa em um processo de inteligência estratégica.

Figura 5 – Relação entre o Consumo de Aço e o Mercado.



Fonte: Dados elaborados pelo pesquisador

A Figura 5, dentro de um processo de inteligência, traz uma visão mais ampla dos grupos de variáveis que devem ser constantemente monitorados pela empresa Alfa na busca de identificação de sinais fracos que possam orientar o seu planejamento estratégico relacionado com a compra e armazenamento de aço. É importante destacar mais uma vez que o modelo representado na Figura 5 está em constante processo de transformação e os elementos presentes neste momento, necessariamente, não precisam estar presentes em momentos posteriores, pois estes dependem de um monitoramento constante por parte da empresa, das relações existentes entre o mercado externo e as suas variáveis internas relacionadas com o consumo do aço. Esse monitoramento consiste na identificação de alterações no mercado e passa pela identificação, dentro do *Big Data*, de novas fontes de dados que auxiliem a identificação de grupos de variáveis relacionadas com o consumo de aço da empresa.

Os dados apresentados na Tabela 17 serão utilizados para identificação do modelo de previsão de consumo da empresa Alfa através da Metodologia de Box e Jenkins com função de transferência.

4.2.3 Verificação dos pressupostos básicos para identificação do modelo de previsão

Segundo Camargo e Souza (1996) e Fischer (1982) as hipóteses básicas para encontrar modelo de previsão são o conjunto de dados seguir uma distribuição normal e a variância desse conjunto ser constante, ou seja, existe homocedasticidade. Se as condições não forem respeitadas, a série estudada deve passar por um processo de transformação que

tem como objetivo estabilizar a variância e fazer com que o conjunto de dados seja supostamente normal.

A condição de normalidade já foi verificada durante a etapa qualitativa através do teste da aplicação do teste *Kolmogorov-Smirnov* e os resultados foram apresentados na Tabela 15.

A homocedasticidade foi avaliada pelo teste *Levene* (HAIR JR. et al., 2005, p.79). Neste teste se verifica se existe uma diferença significativa entre as variâncias, para tanto a hipótese nula afirma que não existe diferença entre as variâncias contra a hipótese alternativa a qual afirma existir diferença entre as variâncias. A hipótese nula deve ser rejeitada, ou seja, existe uma diferença significativa entre as variâncias. Esta pode indicar a necessidade de algum tipo de transformação das variáveis, condição que será avaliada novamente durante o processo de identificação do modelo.

Além dos pressupostos de normalidade e homocedasticidade é preciso avaliar a multicolinearidade, que segundo Hair Jr. et al. (2005, 91), é a medida que avalia a relação entre as variáveis independentes, ou seja, avalia o “grau em que cada variável pode ser explicada pelas outras variáveis da análise”. Uma multicolinearidade elevada pode prejudicar a capacidade de prever efeitos sobre a variável, o fator de inflação de variância (VIF) é utilizado para avaliar a multicolinearidade, o VIF não pode exceder o valor de 10. Na primeira análise de multicolinearidade muitas das variáveis estudadas apresentaram um VIF muito superior a 10, para resolver este problema retiramos do conjunto de variáveis as que apresentaram valores elevados de VIF. A Tabela 18 traz a análise de multicolinearidade e o conjunto de variáveis em que o VIF foi inferior a 10.

Tabela 18 – Avaliação de Multicolinearidade

Variáveis	VIF
1375 – Produção de caminhões – unidades	3,767
1386 – Produção de retroscavadeiras – unidades	3,892
1483 - Índice volume de vendas no varejo - Combustíveis e lubrificantes - Brasil – Índice	8,323
1548 - Índice volume de vendas no varejo - Automóveis, motocicletas, partes e peças – Brasil – Índice	4,289
7387 - Vendas de veículos pelas concessionárias - Ônibus – Unidades	4,455
11065 – Indicadores da produção (2002=100) - Extrativa mineral – Índice	4,910
11066 – Indicadores da produção (2002=100) - Indústria de transformação – Índice	6,427
20105 – Material de construção – Índice	8,474
206 - Cesta básica - u.m.c.	7,737

Fonte: Dados elaborados pelo pesquisador

Após essa análise, o número de variáveis que possuem alguma relação com o consumo de aço da empresa Alfa ficou reduzido às apresentadas na Tabela 18. Estas foram as variáveis utilizadas na identificação do modelo de previsão de consumo.

4.2.4 Ciclo de Box e Jenkins para o modelo univariado

Segundo Morettin e Toloï (1987), a construção dos modelos de Box e Jenkins é um ciclo composto pelas seguintes etapas:

- identificação de uma classe geral de modelos que será analisada;
- especificação do modelo, base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- estimação dos parâmetros do modelo;
- verificação do modelo ajustado, através da análise dos resíduos para medir a adequação do modelo para realização de previsões;
- reiniciar o ciclo se o modelo não for adequado.

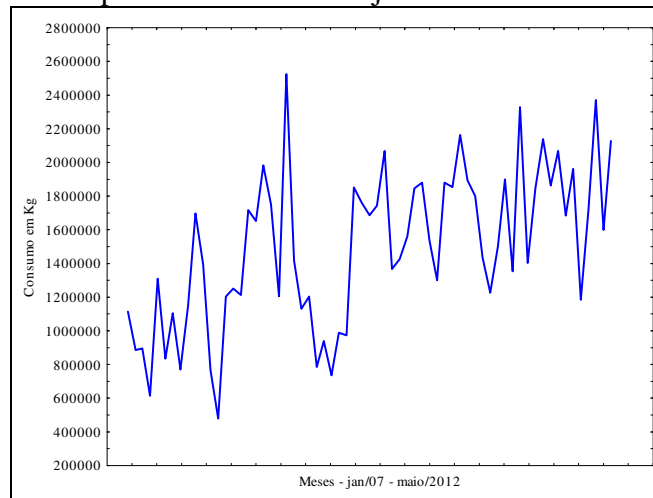
A seguir, o ciclo de Box e Jenkins será realizado para construção do modelo univariado.

4.2.4.1 Identificação do modelo univariado sem intervenção para variável Consumo de Aço

O primeiro passo no processo de identificação do modelo foi a análise do gráfico da série de consumo de aço da empresa Alfa. Essa análise teve como finalidade a identificação informal da presença de tendência, sazonalidade ou alterações na variância. Tais informações podem indicar se a série é ou não estacionária e se a mesma necessita ou não de uma transformação prévia.

Para isto se construiu o gráfico da série e a partir dele foi feita a análise informal das características citadas. O Gráfico 3 representa a série de consumo mensal de aço da empresa Alfa em kg, no período entre janeiro de 2007 a maio de 2012.

Gráfico 3 – Consumo mensal de aço –
Empresa Alfa – Brasil – jan. 07 / maio 2012

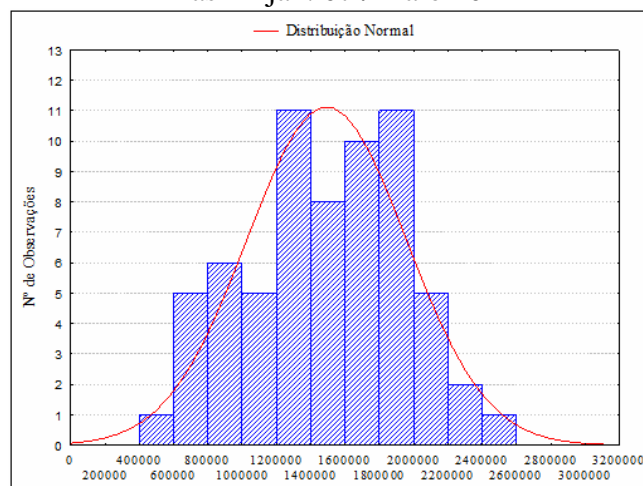


Fonte: Elaborado pelo pesquisador

Em uma análise informal das características do consumo de aço da empresa Alfa através do Gráfico 3, foi possível verificar que a série estudada parece ser estacionária com um conjunto de picos, eles parecem se repetir periodicamente o que sugere uma possível sazonalidade. A confirmação ou não dessas características foi feita posteriormente através das análises da função de autocorrelação (ACF) e função de autocorrelações parciais (PACF) além do teste da raiz unitária.

Também foi feita uma análise do Histograma de distribuição de frequências da variável consumo, apresentado no Gráfico 4.

Gráfico 4 – Histograma de distribuição de frequências
Consumo mensal de aço Empresa Alfa
Brasil – jan. 07 / maio 2012



Fonte: Elaborado pelo pesquisador

A análise do Gráfico 4 indicou que o conjunto de dados se aproxima de uma distribuição normal, condição esta verificada anteriormente no teste de *Kolmogorov-Smirnov*, mas, apresenta uma certa assimetria o que pode indicar a necessidade de algum processo de transformação da variável para construção do modelo. Com o objetivo de confirmar esta análise, foram calculadas as estatísticas descritivas da variável consumo de aço. Os resultados são apresentados na Tabela 19.

Tabela 19 – Estatísticas descritivas da variável consumo

Medidas	Estatísticas	Medidas	Estatísticas
Média	1492617,122	Assimetria	-0,064446274
Mediana	1501001	Amplitude	2045385
Desvio padrão	466028,2411	Mínimo	478780
Curtose	-0,643343072	Máximo	2524165

Fonte: Elaborado pelo pesquisador

Os valores da curtose e da assimetria confirmam um certo grau de assimetria moderada e certo grau de curtose o que indica alguma variabilidade. Segundo Morettin e Tolo (2006), os problemas de simetria podem ser resolvidos através de uma transformação da variável, que torna o conjunto de dados mais simétrico e próximo da distribuição da normal.

A necessidade de transformação da variável foi verificada através do cálculo da estatística d_λ através da expressão:

$$d_\lambda = \frac{|m\u00e9dia - mediana|}{Desvio} \quad (74)$$

O menor valor absoluto da estatística d_λ indica o tipo de transformação que a variável deve sofrer antes do processo de identificação do modelo. Os valores obtidos para d_λ , assim como as sugestões de transformações, são apresentados na Tabela 20.

Tabela 20 – Valores de d_λ para variável consumo

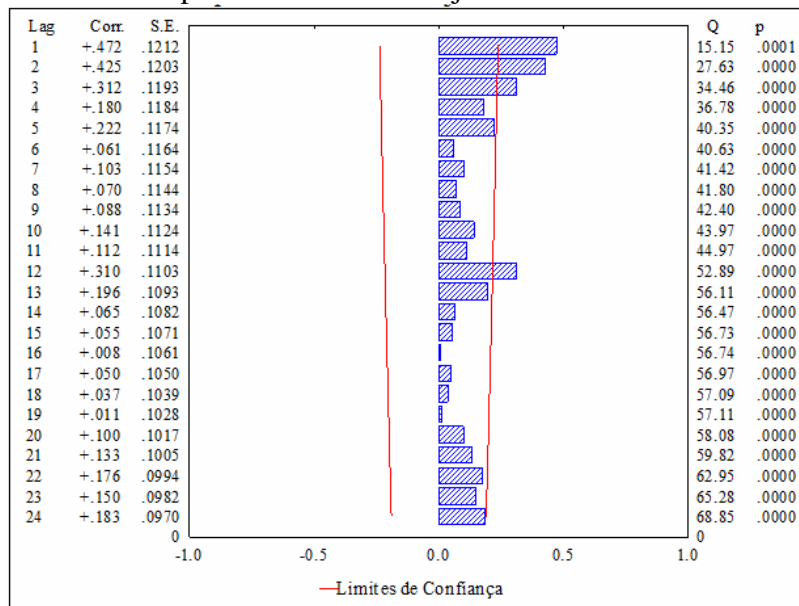
λ	d_λ	Transformação
-1,0	0,2908	$\frac{1}{Z_t}$
-0,5	0,2398	$\frac{1}{\sqrt{Z_t}}$
0	0,1737	$\ln(Z_t)$
0,5	0,0979	$\sqrt{Z_t}$
1	0,0180	Z_t

Fonte: Elaborado pelo pesquisador

O valor encontrado de $d_\lambda=0,0180$ implica em $\lambda=1$, logo, conforme Morettin e Toloi (2006), os valores apresentados na Tabela 20 indicam que a série de consumo de aço não necessita sofrer nenhum processo de transformação, sendo assim, o modelo pode ser identificado a partir dos dados originais.

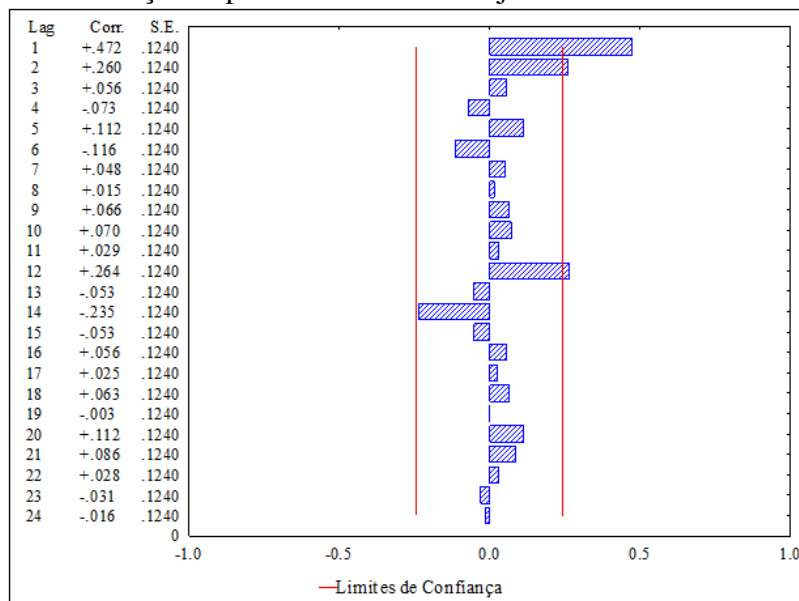
Os Gráfico 5 e 6 trazem os correlogramas das funções de autocorrelações (AFC) e autocorrelações parciais (PACF) da variável consumo.

Gráfico 5 – Função de Autocorrelação do Consumo mensal de aço Empresa Alfa – Brasil – jan. 07/ - maio/12



Fonte: Elaborado pelo pesquisador

Gráfico 6 – Função de Autocorrelação Parcial do Consumo mensal de aço Empresa Alfa – Brasil – jan. 07/ - maio/12



Fonte: Elaborado pelo pesquisador

O comportamento da FAC e da PACF observado em seus respectivos gráficos indicaram que a função consumo decresce rapidamente para zero, o que mais uma vez indica a estacionaridade da série. A confirmação estatística da estacionaridade da série foi verificada através da aplicação do teste aumentado de Dickey Fuller (ADF) para verificar a presença de raiz unitária.

O teste foi realizado no software Gretl versão 1.9.12 e os resultados avaliados conforme valores críticos tabelados por Fuller e citado por Gujarati (1995, p. 815). A hipótese nula avaliada é que a série apresenta uma raiz unitária. Os resultados do teste são apresentados na Tabela 21.

Tabela 21 – Resultados do teste ADF para variável consumo

Teste	τ	P (valor)	Valores críticos de Dickey – Fuller	
			1%	5%
Sem constante	-0,316851	0,5717	-2,66	-1,95
Com constante	-3,03226	0,03201	-3,75	-3,00
Com constante e tendência	-6,26389	0,009359	-4,38	-3,60

Fonte: Elaborado pelo pesquisador

Conforme os resultados do teste ADF, com constante e com constante e tendência, a hipótese nula deve ser rejeitada, portanto a série de consumo de aço é uma série estacionária a um nível de 1%, ou seja, a série respeita a condição de estacionaridade necessária para identificação de um modelo pela metodologia de Box e Jenkins Box. A interpretação do teste também indica que a série deve ser considerada não estacionária se avaliada sem constante.

4.2.4.2 Especificação do modelo para variável Consumo de Aço

Analisando os comportamentos das FAC e PACF através de seus respectivos gráficos e comparando os resultados obtidos com o $2\sigma = 2 \frac{1}{\sqrt{n}}$, onde n representa o número de observações, foi possível verificar que somente as duas primeiras autocorrelações são superiores a $2\sigma = 0,24806$ e sendo assim são significativamente diferentes de zero, levando à suposição de que deve haver processo AR(2). Além disso, o comportamento da FAC aparenta uma mistura de exponencial com ondas senoidais amortecidas característica, que para Morettin e Tolo (2006, p. 153) confirmam o processo AR(2).

Os comportamentos da FAC e da PACF apresentaram a possibilidade de existência de uma componente de sazonalidade. No FAC os lags 12 e 24 são significativamente diferentes de zero e no PACF o comportamento se repete no lag 12. Este comportamento

pode evidenciar um processo SARIMA ou um processo MA(q) com defasagem 12. Com base nestas interpretações foi possível estabelecer algumas hipóteses de modelos apresentadas na Tabela 22.

Tabela 22 – Hipóteses de Modelos para variável consumo de aço.

Modelos
ARIMA (2, 0, 0)
ARIMA (2, 0, 1)
ARIMA (2, 0, 2)
SARIMA (2, 0, 0) (1, 0, 1)
SARIMA (2, 0, 0) (1, 0, 2)
SARIMA (2, 0, 0) (2, 0, 1)
SARIMA (2, 0, 0) (2, 0, 2)
Modelo Automático gerado pelo pacote computacional.

Fonte: Elaborado pelo pesquisador

A Tabela 23 traz os parâmetros e as estatísticas dos supostos modelos para a série consumo do aço, dados estes obtidos através do software Gretl.

Tabela 23 – Modelos estimados e suas estatísticas para variável consumo de aço.

(continua)

Modelo 1: ARIMA (2, 0, 0) usando as observações 2007:01-2012:05				
Parâmetros	Coefficiente	Erro Padrão	Z	p-valor
Const	1,49355e+06	124808	11,9668	<0,00001 ***
Phi_1	0,353363	0,120736	2,9267	0,00343 ***
Phi_2	0,272716	0,120792	2,2577	0,02396 **
Estatísticas do Modelo				
Média var. dependente	1492617	D.P. var. dependente	466028,2	
Média de inovações	8146,899	D.P. das inovações	389916,5	
Log da verossimilhança	-929,2326	Critério de Akaike	1866,465	
Critério de Schwarz	1875,163	Critério Hannan-Quinn	1869,897	
Modelo 2: ARIMA (2, 0, 1) usando as observações 2007:01-2012:05				
Parâmetros	Coefficiente	Erro Padrão	Z	p-valor
Const	1,48749e+06	152055	9.7826	<0.00001 ***
Phi_1	0,330736	0.118094	2.8006	0.00510 ***
Phi_2	0,304647	0,116892	2,6062	0,00915 ***
Theta_12	0,327288	0,111904	2,9247	0,00345 ***
Estatísticas do Modelo				
Média var. dependente	1492617	D.P. var. dependente	466028.2	
Média de inovações	22038,59	D.P. das inovações	364324,3	
Log da verossimilhança	-925,5225	Critério de Akaike	1861,045	
Critério de Schwarz	1871,917	Critério Hannan-Quinn	1865,335	

(continua)

Modelo 3: ARIMA (2, 0, 2) usando as observações 2007:01-2012:05				
Parâmetros	Coefficiente	Erro Padrão	Z	p-valor
Const	1,49391e+06	134593	11,0995	<0,00001 ***
Phi_1	0,757737	0,689382	1,0992	0,27170
Phi_2	0,0123752	0,597028	0,0207	0,98346
Theta_1	-0,424054	0,67862	-0,6249	0,53205
Theta_2	0,0956138	0,383424	0,2494	0,80308
Estatísticas do Modelo				
Média var. dependente	1492617	D.P. var. dependente	466028,2	
Média de inovações	24989,04	D.P. das inovações	388536,7	
Log da verossimilhança	-929,0158	Critério de Akaike	1870,032	
Critério de Schwarz	1883,078	Critério Hannan-Quinn	1875,179	
Modelo 4: SARIMA (1, 0, 0)(1, 0, 1) usando as observações 2007:01-2012:05				
Parâmetros	Coefficiente	Erro Padrão	Z	p-valor
Const	1,4952e+06	194163	7,7008	<0,00001 ***
Phi_1	0,310861	0,114971	2,7038	0,00685 ***
Phi_2	0,318541	0,114123	2,7912	0,00525 ***
Phi_1	0,695034	0,117492	5,9156	<0,00001 ***
Theta_1	-0,337375	0,145065	-2,3257	0,02004 **
Estatísticas do Modelo				
Média var. dependente	1492617	D.P. var. dependente	466028,2	
Média de inovações	56500,99	D.P. das inovações	351334,8	
Log da verossimilhança	-923,9594	Critério de Akaike	1859,919	
Critério de Schwarz	1872,965	Critério Hannan-Quinn	1865,066	
Modelo 5: SARIMA (2, 0, 0)(1, 0, 2) usando as observações 2007:01-2012:05				
Parâmetros	Coefficiente	Erro Padrão	Z	p-valor
Const	1,49494e+06	193251	7,7357	<0,00001 ***
Phi_1	0,309187	0,116372	2,6569	0,00789 ***
Phi_2	0,318644	0,114124	2,7921	0,00524 ***
Phi_1	0,682335	0,120076	5,6825	<0,00001 ***
Theta_1	-0,324785	0,147038	-2,2088	0,02719 **
Theta_2	0,00963069	0,124134	0,0776	0,93816
Estatísticas do Modelo				
Média var. dependente	1492617	D.P. var. dependente	466028,2	
Média de inovações	61892,43	D.P. das inovações	351335,3	
Log da verossimilhança	-923,9589	Critério de Akaike	1861,918	
Critério de Schwarz	1877,139	Critério Hannan-Quinn	1867,923	
Modelo 6: SARIMA (2, 0, 0) (2, 0, 1) usando as observações 2007:01-2012:05				
Parâmetros	Coefficiente	Erro Padrão	Z	p-valor
Const	1,49478e+06	192754	7,7549	<0,00001 ***
Phi_1	0,308277	0,117832	2,6162	0,00889 ***
Phi_2	0,318743	0,114874	2,7747	0,00552 ***
Phi_1	0,630311	0,182402	3,4556	0,00055 ***
Phi_2	0,0309538	0,165997	0,1865	0,85207
Theta_1	-0,272682	0,160916	-1,6946	0,09016 *

(conclusão)

Estatísticas do Modelo				
Média var. dependente	1492617	D.P. var. dependente	466028,2	
Média de inovações	53098,13	D.P. das inovações	351338,3	
Log da verossimilhança	-923,9591	Critério de Akaike	1861,918	
Critério de Schwarz	1877,139	Critério Hannan-Quinn	1867,924	
Modelo 7: SARIMA (2, 0, 0) (2, 0, 2) usando as observações 2007:01-2012:05				
Parâmetros	Coeficiente	Erro Padrão	Z	p-valor
Const	1,4925e+06	186469	8,0040	<0,00001 ***
Phi_1	0,295913	0,114478	2,5849	0,00974 ***
Phi_2	0,323187	0,11242	2,8748	0,00404 ***
Phi_1	-0,356406	0,124988	-2,8515	0,00435 ***
Phi_2	0,64352	0,125078	5,1450	<0,00001 ***
Theta_1	0,75835	0,205989	3,6815	0,00023 ***
Theta_2	-0,238784	0,188507	-1,2667	0,20526
Estatísticas do Modelo				
Média var. dependente	1492617	D.P. var. dependente	466028,2	
Média de inovações	51500,93	D.P. das inovações	342386,5	
Log da verossimilhança	-923,8316	Critério de Akaike	1863,663	
Critério de Schwarz	1881,058	Critério Hannan-Quinn	1870,527	
Modelo 8 : SARIMA (0, 1, 1) (1, 0, 0) usando as observações 2007:01-2012:05				
Observação: Modelo automático gerado pelo pacote computacional				
Parâmetros	Coeficiente	Erro Padrão	Z	p-valor
Phi_1	0,354598	0,125742	2,8200	0,00480 ***
Theta_1	-0,646614	0,0843768	-7,6634	<0,00001 ***
Estatísticas do Modelo				
Média var. dependente	15825,72	D.P. var. dependente	473399,0	
Média de inovações	29758,70	D.P. das inovações	371214,2	
Log da verossimilhança	-912,6574	Critério de Akaike	1831,315	
Critério de Schwarz	1837,791	Critério Hannan-Quinn	1833,866	

Fonte: Elaborado pelo pesquisador

Através da análise dos resultados da Tabela 23 foi possível descartar os modelos 3, 5, 6 e 7 por estes apresentarem parâmetros que não eram estatisticamente significativos ao nível de significância de 5% ($z = 1,96$). A avaliação dos modelos 1, 2, 4 e 8 levou em consideração os critérios de Akaike, Schwarz e Hannan-Quinn. Nestes os modelos 2 e 8 apresentaram os menores valores, o que indicou que os dois eram mais parcimoniosos. Importante destacar que o modelo número 2 foi o primeiro modelo identificado através da análise das FAC e PACF, e que o modelo 8 foi o modelo gerado automaticamente pelo sistema, considerando a série de dados como não estacionária sem constante, classificação esta que foi verificada quando da análise do teste ADF.

Na busca de identificação do melhor modelo, decidiu-se não escolher um modelo único nesta etapa do trabalho. Trabalhou-se com os modelos 2 e 8 nas etapas posteriores para verificar qual dos dois apresentariam melhores resultados.

4.2.4.3 Estimação dos parâmetros do modelo para variável Consumo de Aço

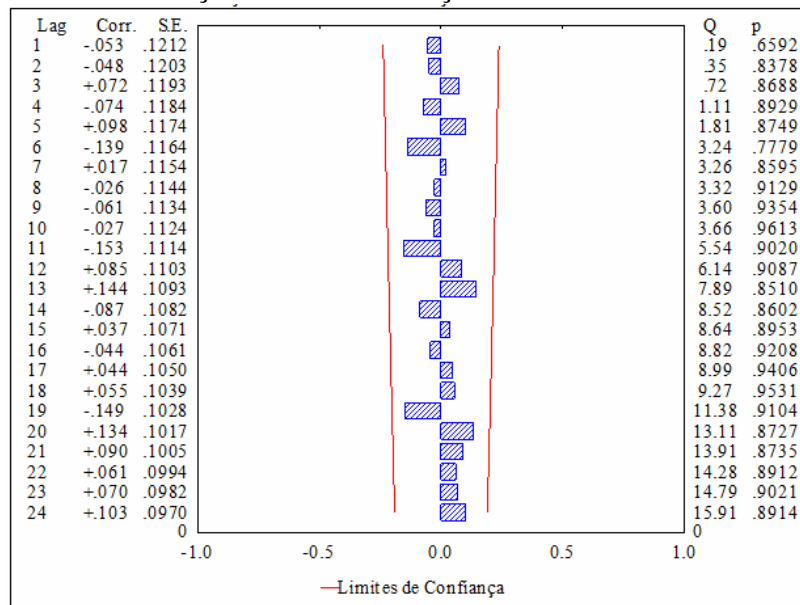
Os parâmetros dos modelos foram estimados nas etapas anteriores e encontram-se na Tabela 23.

4.2.4.4 Verificação dos modelos identificados para variável Consumo de Aço

Nesta etapa foi feita a verificação dos modelos identificados na etapa anterior, para isto foram feitas as análises gráficas da FAC e PACF dos resíduos dos dois modelos, a aplicação do teste de *Portmanteau* e avaliação dos resultados através das estatísticas dos dois modelos.

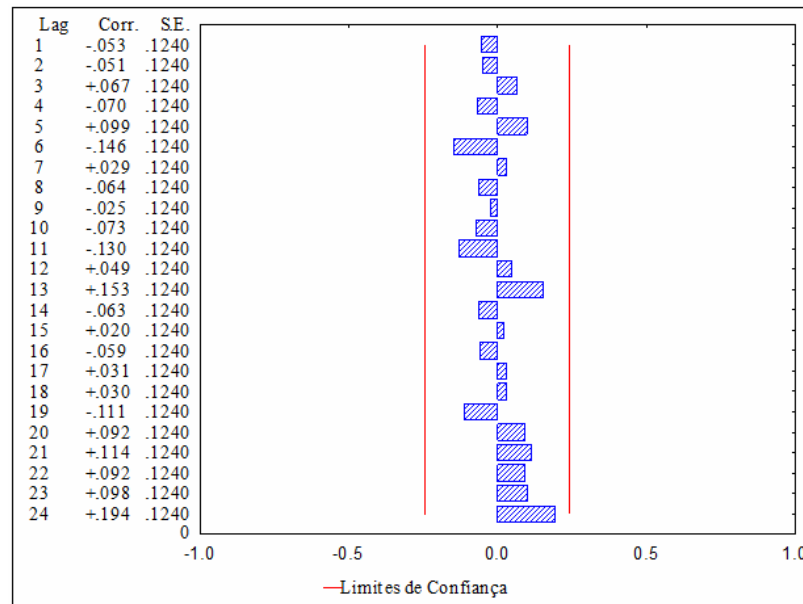
Para análise dos resíduos dos dois modelos foram construídos os FAC e PACF, estes seguem representados nos Gráficos 7, 8, 9 e 10.

Gráfico 7 – Função de Autocorrelação dos resíduos do modelo 2.



Fonte: Elaborado pelo pesquisador

Gráfico 8 – Função de Autocorrelação Parcial dos resíduos do modelo 2.

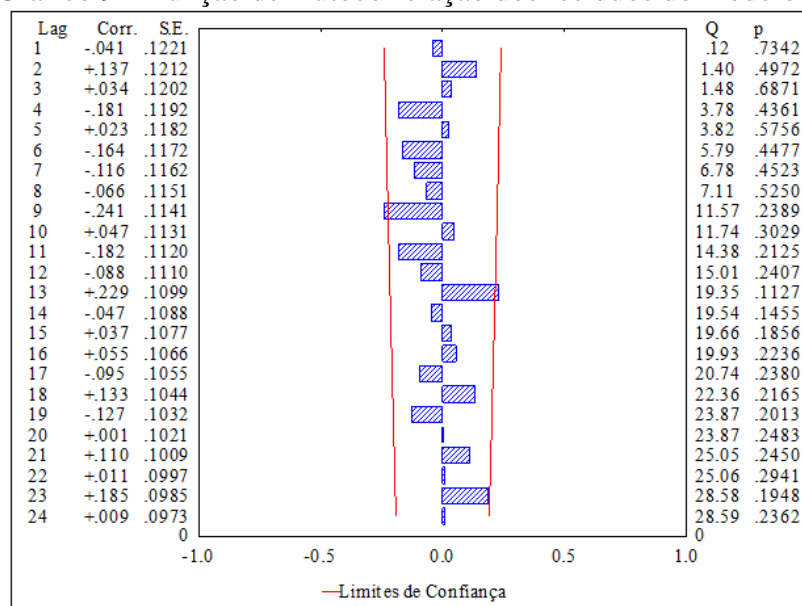


Fonte: Elaborado pelo pesquisador

Avaliando os Gráficos 7 e 8 foi possível verificar que a FAC e PACF dos resíduos permanecem dentro dos limites de confiança. No gráfico da FAC é possível verificar também os valores de p para a estatística Q , avaliando individualmente as correlações como sendo indicativo de um alto grau de significância para hipótese dos resíduos não estarem correlacionados e assim constituírem um ruído branco. Avaliando o conjunto de correlações pelo teste de *Portmanteau*, a estatística Q apresentou um valor de 14,95551, a um nível de significância de 5%, o valor tabelado do Qui-quadrado, para 21 graus de liberdade é de 32,67056, o que confirma a hipótese nula que os resíduos são brancos.

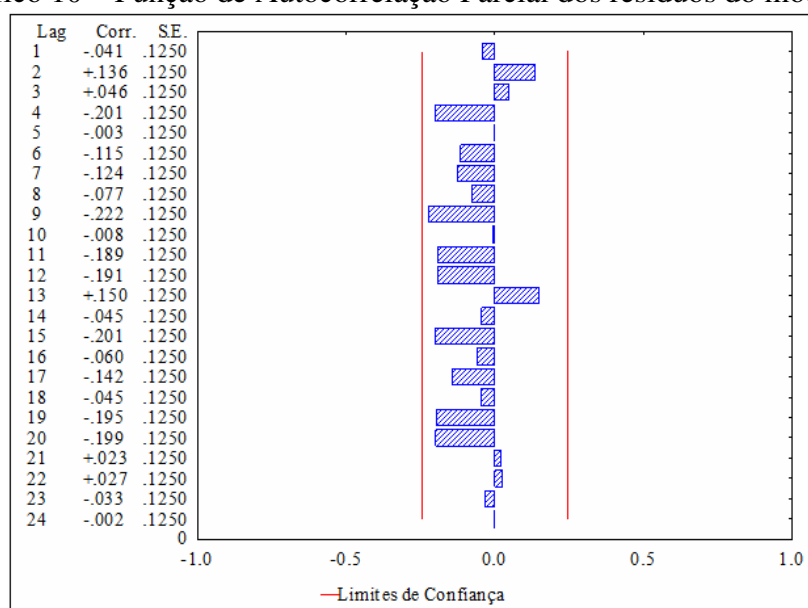
Os Gráficos 9 e 10 trazem a FAC e PACF dos resíduos do modelo 8. Para este modelo alguns valores parecem estar fora dos limites de confiança, mas avaliando os p valores individualmente é possível verificar que todos são significantes. O teste de *Portmanteau* para avaliar o conjunto de correlações apresentou estatística Q igual 27,17847, a um nível de significância de 5%, o valor tabelado do Qui-quadrado, para 22 graus de liberdade é de 33,92446, o que confirma a hipótese nula que os resíduos são brancos.

Gráfico 9 – Função de Autocorrelação dos resíduos do modelo 8.



Fonte: Elaborado pelo pesquisador

Gráfico 10 – Função de Autocorrelação Parcial dos resíduos do modelo 8.



Fonte: Elaborado pelo pesquisador

A etapa seguinte consistiu na avaliação das estatísticas dos dois modelos. A Tabela 24 traz os resultados das estatísticas de avaliação da previsão para os dois modelos

Tabela 24 – Estatísticas de avaliação dos modelos.

Estatísticas de avaliação Modelo 2		Estatísticas de avaliação Modelo 8	
Erro Percentual Médio	-6,1111	Erro Percentual Médio	-2,6223
Erro Percentual Médio Absoluto	23,301	Erro Percentual Médio Absoluto	24,19
R ²	37,86%	R ²	75,32%

Fonte: Elaborado pelo pesquisador

Após a avaliação da FAC e PACF das séries de resíduos dos dois modelos, pode-se perceber que as séries dos resíduos dos dois modelos apresentam comportamento de uma série de resíduos brancos, ou seja, as séries podem ser consideradas aleatórias. As análises das estatísticas dos dois modelos indicaram que o modelo 2 possui um Erro percentual médio absoluto (MAPE) inferior ao do modelo 8, mas em contrapartida o modelo 8 apresenta um R^2 superior ao do modelo 2 e por este motivo foi escolhido como o modelo de previsão da variável consumo de aço.

4.2.4.5 Previsão do SARIMA (0, 1, 1)(1, 0, 0) para variável consumo de aço

Para realizar as previsões do consumo de aço através do modelo SARIMA (0, 1, 1)(1, 0, 0)₁₂ optou-se por 11 períodos adiante do período utilizado para identificação do modelo, ou seja, as previsões são relativas ao período de junho de 2012 a abril 2013. Este conjunto de previsões apresentou um MAPE de 15,26%. A Tabela 25 traz as previsões mensais do consumo feitas através do modelo identificado.

Tabela 25 – Previsões mensais do consumo de aço – jun./12 – abr./13

Ano / Meses	Previsto	Intervalo de Confiança (95%)		Valor Real	% de Erro
2012:06	1769409,3	1041842,7	2496976	1988447,0	11,0%
2012:07	1926607,4	1154947,1	2698268	2184883,0	11,8%
2012:08	2030349,9	1216982,7	2843717	2968178,0	31,6%
2012:09	1932339,8	1079302,4	2785377	1818111,0	-6,3%
2012:10	2005348,3	1114405,3	2896291	2239756,0	10,5%
2012:11	1869176,4	941875,93	2796477	2188962,0	14,6%
2012:12	1967756,3	1005471,2	2930041	1581586,0	-24,4%
2013:01	1691946	695904,14	2687988	1914027,0	11,6%
2013:02	1882627,7	853936,35	2911319	1819222,0	-3,5%
2013:03	2112353,2	1052017,1	3172689	1697094,0	-24,5%
2013:04	1839057,2	747993,93	2930121	2244515,0	18,1%

Fonte: Elaborado pelo pesquisador

4.2.5 Ciclo de Identificação da Função de Transferência

Esta etapa da pesquisa consistiu na construção de um modelo para previsão de consumo de aço através de metodologia de Função de Transferência de Box e Jenkins. Para a construção do modelo foi utilizado o pacote estatístico Autobox (versão 1.2), este gerou de forma automática os modelos com base nas variáveis informadas. O software Autobox apresenta uma limitação quanto ao número de variáveis independentes, este só trabalhava com 5 variáveis independentes por vez, por este motivo foi necessário rodar mais que um modelo.

Com base nas variáveis informadas, o software identificou quatro para o modelo de função de transferência, sendo elas: Índice de Indicadores da produção - Extrativa mineral; Produção de Retroescavadeiras; Índice de Indicadores da produção - Indústria de transformação; Índice de Material de construção.

Após identificar as variáveis o *software* iniciou o processo de pré-branqueamento das variáveis, que consistiu em identificar modelos univariados para cada uma das variáveis selecionadas para fazerem parte do modelo com função de transferência. A seguir serão apresentados os resultados do processo de identificação de pré-branqueamento.

4.2.5.1 Identificação do modelo univariado - Índice de Indicadores da produção Extrativa mineral

A Tabela 26 traz a identificação do modelo univariado para variável Índice de Indicadores da produção Extrativa mineral feita pelo *software* Autobox.

Tabela 26 – Identificação do Modelo Univariado

Variável: Índice de Indicadores da Produção Extrativa Mineral			
Número de Observações: 65			
Diferenciação: Sim, de primeira ordem			
Backcasting: Não			
Parâmetros do modelo univariado			
Elemento	Lag	Coefficiente	Estatística t
AUTOREGRESSIVE 1	12	0,58721E+00	5,46

Fonte: Elaborado pelo pesquisador

4.2.5.2 Análise dos resíduos e estatísticas do modelo ARIMA(1, 1, 0) - Índice de Indicadores da produção Extrativa mineral

A Tabela 27 traz as estatísticas do modelo gerado no processo de pré-branqueamento da variável.

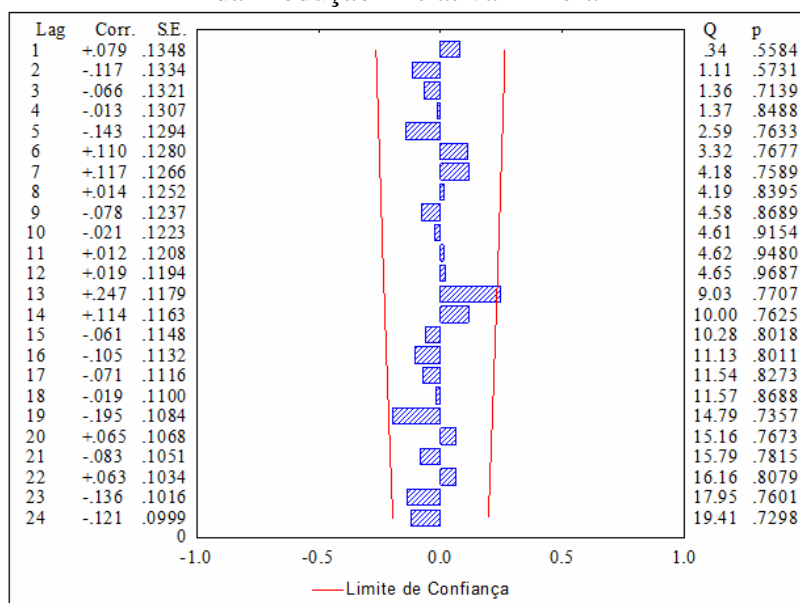
Tabela 27 – Estatísticas do Modelo

SOMA DOS QUADRADOS: .19053E+00	GRAUS DE LIBERDADE: 51
MEAN SQUARE : 37359E-02	NÚMERO DE RESÍDUOS : 52
R ² : 0,71868E+00	CRITÉRIO DE AKAIKE (AIC): -0,55707E+01
CRITÉRIO DE BAYES (BIC): -0,55332E+01	

Fonte: Elaborado pelo pesquisador

Para fazer a verificação do modelo identificado foi feita a análise gráfica da FAC dos resíduos do modelo e posteriormente o teste de *Portmanteau*. O Gráfico 11 traz a FAC dos resíduos do modelo.

Gráfico 11 – Função de Autocorrelação dos resíduos do modelo ARIMA(1, 1, 0)- Índice de Indicadores da Produção Extrativa Mineral



Fonte: Elaborado pelo pesquisador

A análise do gráfico da FAC da série de resíduos do modelo parece indicar que o modelo não tem um comportamento de uma série de ruído branco, porque alguns valores individuais parecem estar fora do limite de confiança, mas os p valores da FAC são significativos, o que indica em um primeiro momento que a série tem comportamento de ruído branco. Para confirmar o comportamento da série foi feito o teste de *Portmanteau* que apresentou estatística Q igual 17,76516, a um nível de significância de 5%, o valor tabelado da qui-quadrado, para 22 graus de liberdade é de 33,92446, o que confirma a hipótese nula de que os resíduos são brancos.

4.2.5.3 Previsão do modelo ARIMA(1, 1, 0) - Índice de Indicadores da Produção Extrativa Mineral

A Tabela 28 traz as previsões do modelo ARIMA(1, 1, 0) para variável Índice de Indicadores da produção Extrativa mineral, no período de junho de 2012 a abril de 2013. Os resultados obtidos nesta previsão tiveram um MAPE de 6,62% e posteriormente serão utilizados para previsão da variável consumo de aço através da função de transferência.

Tabela 28 – Previsões mensais do modelo ARIMA(1, 1,0)
Índice de Indicadores da Produção Extrativa Mineral – jun./12 – abr./13

Ano / Meses	Previsto	Intervalo de Confiança (95%)		Valor Real	% de Erro
2012/06	154,63	142,65	166,61	148,6	-4,1%
2012/07	157,33	140,39	174,27	153,8	-2,3%
2012/08	157,39	136,64	178,14	153,9	-2,3%
2012/09	154,81	130,85	178,77	145,0	-6,8%
2012/10	155,92	129,14	182,71	159,1	2,0%
2012/11	155,86	126,52	185,21	147,5	-5,7%
2012/12	157,27	125,58	188,97	157,0	-0,2%
2013/01	148,11	114,23	182,00	142,8	-3,7%
2013/02	148,58	112,64	184,52	126,2	-17,8%
2013/03	152,40	114,52	190,28	133,5	-14,2%
2013/04	150,81	111,08	190,54	132,6	-13,7%

Fonte: Elaborado pelo pesquisador

4.2.5.4 Identificação do modelo univariado - Produção de Retroescavadeiras

A Tabela 29 apresenta os resultados da identificação do modelo univariado para variável Índice de Indicadores da produção Extrativa mineral feita pelo *software* Autobox.

Tabela 29 – Identificação do Modelo Univariado

Variável: Produção de Retroescavadeiras			
Número de Observações: 65		Diferenciação: Não	backcasting: Não
Parâmetros do modelo univariado			
Elemento	Lag	Coefficiente	Estatística t
Média		0,50603E+01	6,07
Autoregressivo 1	1	0,79707E+00	8,87
Autoregressivo 2	12	0,28846E+00	1,96

Fonte: Elaborado pelo pesquisador

4.2.5.5 Análise dos resíduos e estatísticas do modelo ARIMA (2, 0, 0) - Produção de Retroescavadeiras

A Tabela 30 apresenta as estatísticas do modelo gerado no processo de pré-branqueamento da variável.

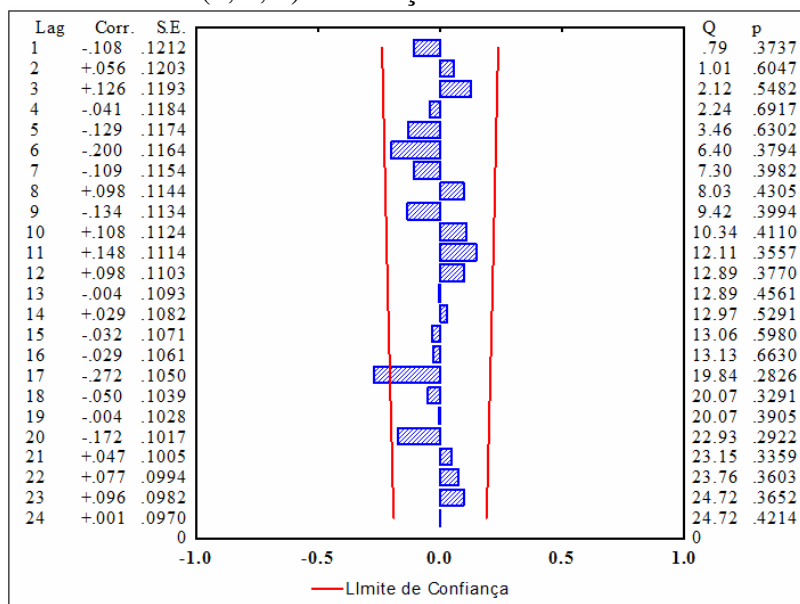
Tabela 30 – Estatísticas do Modelo ARIMA (2, 0, 0) - Produção de Retroescavadeiras

Soma dos Quadrados: .34899E+02	Graus de Liberdade: 49
Média dos Quadrados : .71223E+00	Número de Resíduos : 52
R ² : .64272E+00	Critério de Akaike (AIC): -.28339E+00
Critério de Bayes (BIC): -.17082E+00	

Fonte: Elaborado pelo pesquisador

Para fazer a verificação do modelo identificado foi feita a análise gráfica da FAC dos resíduos do modelo e posteriormente o teste de *Portmanteau*. O Gráfico 12 traz a FAC dos resíduos do modelo.

Gráfico 12 – Função de Autocorrelação dos resíduos do modelo ARIMA(2, 0, 0) - Produção de Retroescavadeiras



Fonte: Elaborado pelo pesquisador

A análise do gráfico da FAC da série de resíduos indicou que somente o *lag* 17 tem comportamento de ruído não branco por estar fora dos limites de confiança, mas os p valores são todos significativos e o teste de *Portmanteau* para o conjunto de autocorrelações indicou a estatística Q igual 22,96401, a um nível de significância de 5%, o valor tabelado da qui-quadrado, para 22 graus de liberdade é de 33,92446, o que confirma a hipótese nula de que os resíduos são brancos.

4.2.5.6 Previsão do modelo ARIMA(2, 0, 0) - Produção de Retroescavadeiras

A Tabela 31 apresenta as previsões do modelo ARIMA(2, 0, 0) para variável Produção de retroescavadeiras, no período de junho de 2012 a abril de 2013. Os resultados obtidos nesta previsão tiveram um MAPE de 17,01% e posteriormente serão utilizados para previsão da variável consumo de aço através da função de transferência.

Tabela 31 – Previsões mensais do modelo ARIMA(2, 0, 0)
Produção de Retroescavadeiras – jun./12 – abr./13

Ano / Meses	Previsto	Intervalo de Confiança (95%)		Valor Real	% de Erro
2012/06	640,14	474,73	805,55	628	-1,9%
2012/07	637,2	425,67	848,72	639	0,3%
2012/08	618,7	382,51	854,89	696	11,1%
2012/09	582,21	331,61	832,81	566	-2,9%
2012/10	565,3	305,95	824,64	665	15,0%
2012/11	552,61	287,86	817,35	623	11,3%
2012/12	523,16	255,04	791,28	317	-65,0%
2013/01	558,47	288,22	828,71	525	-6,4%
2013/02	550,91	279,32	822,49	638	13,7%
2013/03	583,29	310,86	855,73	744	21,6%
2013/04	566,58	293,61	839,55	914	38,0%

Fonte: Elaborado pelo pesquisador

4.2.5.7 Identificação do modelo univariado - Índice Indicadores da produção da Indústria de transformação

A Tabela 32 traz os resultados da identificação do modelo univariado para variável Índice de Indicadores da produção da Indústria de Transformação feita pelo *software* Autobox.

Tabela 32 – Identificação do Modelo Univariado

Variável: Índice Indicadores da Produção da Indústria de Transformação			
Número de Observações: 65 Diferenciação: Sim, primeira ordem backcasting: Não			
Parâmetros do modelo univariado			
Elemento	Lag	Coefficiente	Estatística t
Autoregressivo 1	12	0,79647E+00	10,48
Média Móvel 1	3	-0,24055E+00	-1,73

Fonte: Elaborado pelo pesquisador

4.2.5.8 Análise dos resíduos e estatísticas do Modelo ARIMA(1, 1, 1) - Índice Indicadores da Produção da Indústria de Transformação

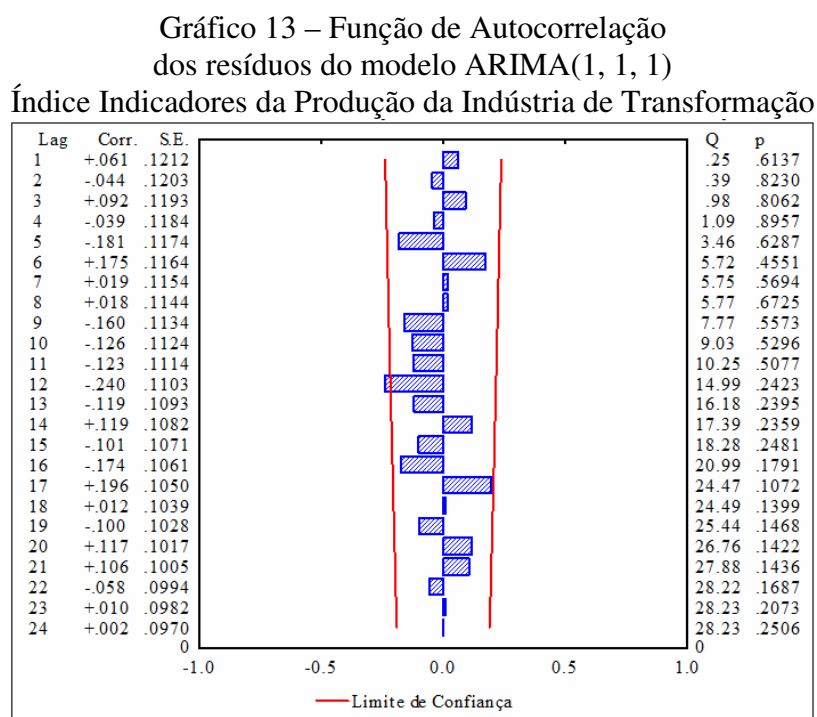
A Tabela 33 traz as estatísticas do modelo gerado no processo de pré-branqueamento da variável.

Tabela 33 – Estatísticas do Modelo ARIMA(1, 1, 1) - Índice Indicadores da Produção da Indústria de Transformação

SOMA DOS QUADRADOS: 0,12899E+00	GRAUS DE LIBERDADE: 50
MEAN SQUARE : 0,25799E-02	NÚMERO DE RESÍDUOS : 52
R ² : 0,78210E+00	CRITÉRIO DE AKAIKE (AIC): -0,59223E+01
CRITÉRIO DE BAYES (BIC): -0,58473E+01	

Fonte: Elaborado pelo pesquisador

Para fazer a verificação do modelo identificado foi feita a análise gráfica da FAC dos resíduos do modelo e posteriormente o teste de *Portmanteau*. O Gráfico 13 traz a FAC dos resíduos do modelo.



Fonte: Elaborado pelo pesquisador

A análise do gráfico da FAC da série de resíduos indicou que somente o *lag* 12 tem comportamento de ruído não branco por estar fora dos limites de confiança, mas os *p* valores são todos significativos e o teste de *Portmanteau* para o conjunto de autocorrelações indicou a estatística *Q* igual 26,22624 a um nível de significância de 5%, o valor tabelado da qui-quadrado para 21 graus de liberdade é de 32,67056, o que confirma a hipótese nula de que os resíduos são brancos.

4.2.5.9 Previsão do modelo Modelo ARIMA(1, 1, 1) - Índice Indicadores da Produção da Indústria de Transformação

A Tabela 34 traz as previsões do modelo ARIMA(1, 1, 1) - Índice Indicadores da Produção da Indústria de Transformação, no período de junho de 2012 a abril de 2013. Os resultados obtidos nesta previsão tiveram um MAPE de 3,63 % e posteriormente serão utilizados para previsão da variável consumo de aço através da função de transferência.

Tabela 34 – Previsões mensais do ARIMA(1, 1, 1) - Índice Indicadores da Produção da Indústria de Transformação – jun./12 – abr./13

Ano / Meses	Previsto	Intervalo de Confiança (95%)		Valor Real	% de Erro
2012/06	123,88	113,93	133,83	120,94	-2,4%
2012/07	126,28	112,21	140,36	128,61	1,8%
2012/08	131,79	114,55	149,02	136,48	3,4%
2012/09	125,91	104,71	147,11	127,28	1,1%
2012/10	127,42	102,89	151,95	137,03	7,0%
2012/11	125,6	98,136	153,06	130,13	3,5%
2012/12	115,21	85,103	145,32	113,32	-1,7%
2013/01	110,38	77,844	142,92	118,33	6,7%
2013/02	112,45	77,645	147,25	110,99	-1,3%
2013/03	123,32	86,392	160,25	123,67	0,3%
2013/04	114,59	75,655	153,52	128,27	10,7%

Fonte: Elaborado pelo pesquisador

4.2.5.10 Identificação do modelo univariado - Índice de Material de Construção

A Tabela 35 apresenta os resultados da identificação do modelo univariado para variável Índice de Material de Construção feita pelo *software* Autobox.

Tabela 35 – Identificação do Modelo Univariado

Variável: Índice de Material de Construção		Número de Observações: 65	
Diferenciação: Não		Backcasting: Não	
Parâmetros do modelo univariado			
Elemento	Lag	Coefficiente	Estatística t
MEAN		0,11929E+01	4,16
AUTOREGRESSIVE 1	1	0,37393E+00	2,69
AUTOREGRESSIVE 1	2	0,41594E+00	3,02
AUTOREGRESSIVE 2	12	0,83195E+00	9,60
MOVING AVERAGE 1	3	-0,53937E+00	-3,82

Fonte: Elaborado pelo pesquisador

4.2.5.11 Análise dos resíduos e estatísticas do Modelo SARIMA(2, 0, 1)(1, 0, 0)₁₂ - Índice de Material de Construção

A Tabela 36 traz as estatísticas do modelo gerado no processo de pré-branqueamento da variável.

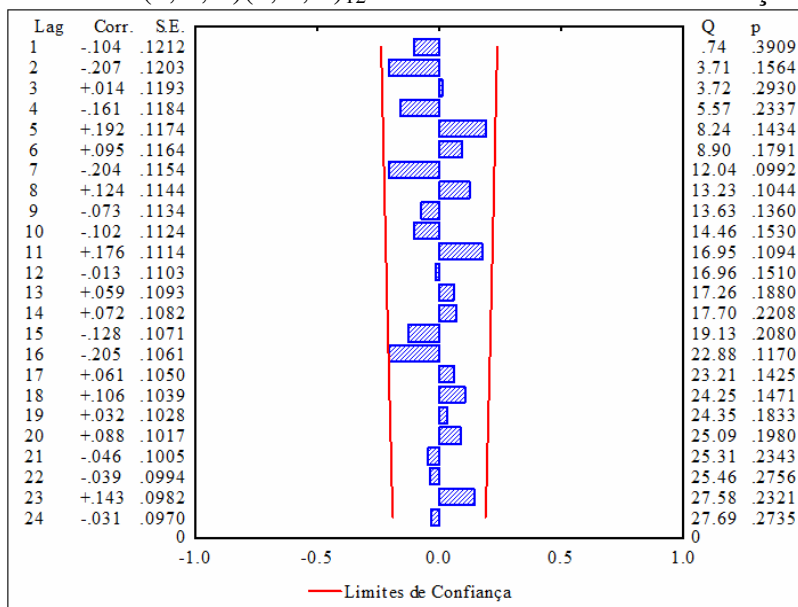
Tabela 36 – Estatísticas do Modelo SARIMA(2, 0, 1)(1, 0, 0)₁₂
Índice de Material de Construção

SOMA DOS QUADRADOS: 0,90022E-01	GRAUS DE LIBERDADE: 46
MEAN SQUARE : 0,19570E-02	NÚMERO DE RESÍDUOS : 51
R ² : 0,87584E+00	CRITÉRIO DE AKAIKE (AIC): -0,61434E+01
CRITÉRIO DE BAYES (BIC): -0,59540E+01	

Fonte: Elaborado pelo pesquisador

Para fazer a verificação do modelo identificado foi feita a análise gráfica da FAC dos resíduos do modelo e posteriormente o teste de *Portmanteau*. O Gráfico 14 traz a FAC dos resíduos do modelo.

Gráfico 14 – Função de Autocorrelação dos resíduos do modelo SARIMA(2, 0, 1)(1, 0, 0)₁₂ Índice de Material de Construção



Fonte: Elaborado pelo pesquisador

A análise do gráfico da FAC da série de resíduos indicou que todos os *lags* estão dentro do limite de confiança para serem considerados com o comportamento de ruído branco, os resultados são todos significativos e o teste de *Portmanteau* para o conjunto de autocorrelações indicou a estatística Q igual 25,70741 a um nível de significância de 5%, o valor tabelado da qui-quadrado, para 21 graus de liberdade é de 31,41042, o que confirma a hipótese nula de que os resíduos são brancos.

4.2.5.12 Previsão do modelo SARIMA(2, 0, 1)(1, 0, 0)₁₂ - Índice de Material de Construção

A Tabela 37 apresenta as previsões do modelo SARIMA(2, 0, 1)(1, 0, 0)₁₂ - Índice de Material de Construção, no período de junho de 2012 a abril de 2013. Os resultados obtidos nesta previsão tiveram um MAPE de 4,94 % e posteriormente serão utilizados para previsão da variável consumo de aço através da função de transferência. Os valores reais para o período de dezembro de 2012 a maio de 2013 não tinham sido atualizados no SGS até a data de realização da pesquisa.

Tabela 37 – Previsões mensais do SARIMA(2, 0, 1)(1, 0, 0)₁₂
Índice de Material de Construção – jun./12 – abr./13

Ano / Meses	Previsto	Intervalo de Confiança (95%)		Valor Real	% de Erro
2012/06	113,01	104,34	121,68	103,1	-9,6%
2012/07	110,01	100,75	119,27	110,74	0,7%
2012/08	113,7	103,27	124,14	119,25	4,7%
2012/09	111,45	98,407	124,49	108,09	-3,1%
2012/10	109,29	95,35	123,24	119,48	8,5%
2012/11	113,22	98,373	128,07	116,78	3,0%
2012/12	102,12	86,758	117,49	-	-
2013/01	106,01	90,223	121,79	-	-
2013/02	99,491	83,427	115,55	-	-
2013/03	111,29	95,011	127,56	-	-
2013/04	103,45	87,021	119,87	-	-

Fonte: Elaborado pelo pesquisador

4.2.5.13 Identificação do modelo univariado – Consumo de Aço da empresa Alfa

A Tabela 38 apresenta os resultados da identificação do modelo univariado para variável Consumo de Aço da empresa Alfa feita pelo *software* Autobox.

Tabela 38 – Identificação do Modelo Univariado

Variável: Consumo de Aço			
Número de Observações: 65			
Diferenciação: Não		Backcasting: Não	
Parâmetros do modelo univariado			
Elemento	Lag	Coefficiente	Estatística t
MEAN		0,15234E+01	9,04
AUTOREGRESSIVE 1	1	0,34561E+00	2,78
AUTOREGRESSIVE 1	2	0,30630E+00	2,49
MOVING AVERAGE 1	12	-0,34105E+00	-2,49

Fonte: Elaborado pelo pesquisador

4.2.5.14 Análise dos resíduos e estatísticas do Modelo ARIMA(2, 0, 1) - Consumo de Aço

A Tabela 39 traz as estatísticas do modelo gerado no processo de pré-branqueamento da variável.

Tabela 39 – Estatísticas do Modelo ARIMA(2, 0, 1) - Consumo de Aço

SOMA DOS QUADRADOS: 0,85632E+01	GRAUS DE LIBERDADE: 46
MEAN SQUARE : 0,14514E+00	NÚMERO DE RESÍDUOS : 51
R ² : 0,35970E+00	CRITÉRIO DE AKAIKE (AIC): -0,18687E+01
CRITÉRIO DE BAYES (BIC): -0,17326E+01	

Fonte: Elaborado pelo pesquisador

4.2.5.15 Previsão do modelo ARIMA(2, 0, 1) - Consumo de Aço

A Tabela 40 traz as previsões do modelo ARIMA(2, 0, 1) - Consumo de Aço, no período de junho de 2012 a abril de 2013. Os resultados obtidos nesta previsão tiveram um MAPE de 18,94 %. Os resultados obtidos por esta previsão não serão utilizados para previsão da variável consumo de aço através da função de transferência.

Tabela 40 – Previsões mensais do modelo ARIMA(2, 0, 1)
Consumo de Aço – jun./12 – abr./13

Ano / Meses	Previsto	Intervalo de Confiança (95%)		Valor Real	% de Erro
2012/06	1667800	921520	2414200,00	1988447	16,1%
2012/07	1770200	980570	2559800,00	2184883	19,0%
2012/08	1814900	963700	2666000,00	2968178	38,9%
2012/09	1667400	795490	2539200,00	1818111	8,3%
2012/10	1730400	843460	2617200,00	2239756	22,7%
2012/11	1620800	726640	2515000,00	2188962	26,0%
2012/12	1717900	819270	2616500,00	1581586	-8,6%
2013/01	1467500	566490	2368600,00	1914027	23,3%
2013/02	1626000	723540	2528400,00	1819222	10,6%
2013/03	1805000	901840	2708200,00	1697094	-6,4%
2013/04	1605900	702240	2509500,00	2244515	28,5%

Fonte: Elaborado pelo pesquisador

4.2.5.16 Identificação do modelo de Função de Transferência para variável Consumo de Aço

Após o processo de pré-branqueamento das séries o software realizou o processo de identificação do modelo de função de transferência. A Tabela 41 apresenta os resultados obtidos.

Tabela 41 – Identificação do Modelo de Função de Transferência

(continua)

Variável Dependente: (Y) - Consumo de Aço			
Número de Observações: 65		Diferenciação: Não	
Série do Ruído		Diferenciação sobre o ruído: Nenhum	
Parâmetros do modelo de Ruído			
Elemento	Lag	Coefficiente	Estatística t
1 – Média		0,15851E+01	41,13
Primeira Série de Entrada			
Variável Independente: (X ₁) - Índice de Indicadores da Produção Extrativa Mineral			
Número de Observações: 65		Diferenciação: Sim, uma de primeira ordem	
Valor do parâmetro de atraso: 5			
Parâmetros do modelo de Função de Transferência			
Elemento	Lag	Coefficiente	Estatística t
2 – Lag de Entrada	1	-0,56866E+00	-2,27
3 – Lag de Saída	0	0,11654E+01	2,35

(conclusão)

Segunda Série de Entrada			
Variável Independente: (X_2) Produção de Retroescavadeiras			
Número de Observações: 65			
Diferenciação: Não – (Média presumida da série 0,43642E+01)			
Valor do parâmetro de atraso: 0			
Parâmetros do modelo de Função de Transferência			
Elemento	Lag	Coeficiente	Estatística t
4 – Lag de Entrada	0	0,18580E+00	6,29
Terceira Série de Entrada			
Variável Independente: (X_3) - Índice Indicadores da Produção da Indústria de Transformação			
Número de Observações: 65 Diferenciação: Sim, uma de primeira ordem			
Valor do parâmetro de atraso: 4			
Parâmetros do modelo de Função de Transferência			
Elemento	Lag	Coeficiente	Estatística t
5 – Lag de Entrada	0	0,21039E+01	2,65
6 – Lag de Entrada	12	0,13305E+01	1,84
Quarta Série de Entrada			
Variável Independente: (X_4) – Índice de Material de Construção			
Número de Observações: 65 Diferenciação: Não - (Média presumida da série 0,88120E+00)			
Valor do parâmetro de atraso: 6			
Parâmetros do modelo de Função de Transferência			
Elemento	Lag	Coeficiente	Estatística t
7 – Lag de Entrada	0	0,13980E+01	2,24
8 – Lag de Entrada	1	0,19229E+01	2,79

Fonte: Elaborado pelo pesquisador

4.2.5.17 Análise dos resíduos e estatísticas do Modelo de Função de Transferência para variável Consumo de Aço

As Tabela 42 e 44 apresentam, respectivamente, as estatísticas do modelo e as estatísticas da análise de correlação dos resíduos do modelo de função de transferência para consumo de aço.

Tabela 42 – Estatísticas do Modelo de Função de Transferência para variável Consumo de Aço

SOMA DOS QUADRADOS: 0,27036E+01	GRAUS DE LIBERDADE: 40
MEAN SQUARE : 0,67591E-01	NÚMERO DE RESÍDUOS : 48
R ² : 0,66670E+00	CRITÉRIO DE AKAIKE (AIC): -0,25433E+01
CRITÉRIO DE BAYES (BIC):-0,22314E+01	

Fonte: Elaborado pelo pesquisador

Tabela 43 – Estatísticas da análise de correlação dos resíduos Modelo de Função de Transferência

MÉDIA DA SÉRIE DE RESÍDUOS: -0,25183E-05	
DESVIO PADRÃO : 0,23733E+00	Nº DE OBSERVAÇÕES : 48
MÉDIA DIVIDIDA PELO ERRO PADRÃO : -0,73515E-04	

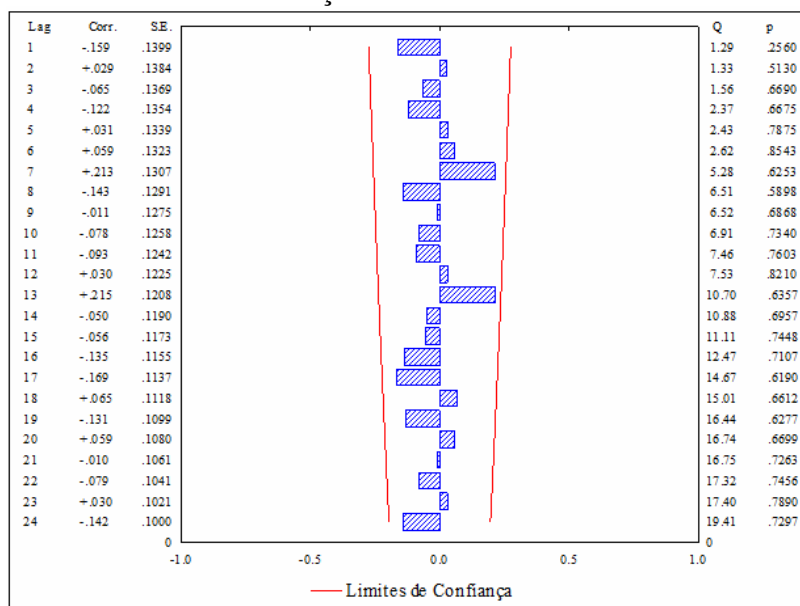
Fonte: Elaborado pelo pesquisador

Avaliando as estatísticas apresentadas foi possível verificar que o modelo apresenta valores para AIC e BIC baixos, o que indica um modelo parcimonioso. Além disso também é possível verificar que o coeficiente de determinação (R^2) tem um valor de 0,6759, o que significa que 67,59% da variável consumo de aço pode ser explicada pelo elementos deste modelo.

A verificação do modelo identificado foi realizada através da análise dos gráficos da FAC dos resíduos da função de transferência e análise das correlações cruzadas entre os resíduos do modelo e os resíduos de cada uma das variáveis independentes. O Gráfico 15 traz a FAC dos resíduos da função de transferência.

A análise do FAC dos resíduos da função de transferência para variável consumo de aço indica que todos os *lags* encontram-se dentro dos limites de confiança com p valores significativos. Para confirmar esta primeira interpretação foi realizado o teste de *Portmanteau*, que apresentou uma estatística Q igual 17,85808, a um nível de significância de 5% o valor tabelado do Qui-quadrado, para 16 graus de liberdade foi 26,29622, o que confirma a hipótese nula de que os resíduos são brancos.

Gráfico 15 – Função de Correlação dos resíduos da Função de Transferência



Fonte: Elaborado pelo pesquisador

A análise das correlações cruzadas entre os resíduos dos processos de pré-branqueamento das variáveis do modelo e os resíduos da função de transferência para variável consumo de aço foi realizada através de saídas do *software* Autobox que encontra-se no Apêndice A. A avaliação dos resultados obtidos é apresentada na Tabela 44.

Na avaliação das correlações cruzadas foi utilizada a estatística S, equação (73), com um nível de significância de 5%. Os resultados obtidos foram comparados com o valor tabelado da distribuição qui-quadrado. Todos os resultados mostraram-se significativos indicando que as correlações cruzadas entre os resíduos do pré-branqueamento e os resíduos da função de transferência são ruídos brancos. Além disso, os resultados obtidos confirmam a casualidade unidirecional entre os resíduos.

Tabela 44 – Análise das correlações cruzadas dos resíduos

Variáveis		Gl	Estatística S	χ^2
Entrada	Saída			
VAR1065	F. de Transferência	17	14,59378	27,5871
F. de Transferência	VAR1065	17	13,34597	27,5871
VAR1386	F. de Transferência	17	16,82645	27,5871
F. de Transferência	VAR1386	17	13,83494	27,5871
V11066	F. de Transferência	17	22,55002	27,5871
F. de Transferência	V11066	17	16,82645	27,5871
VAR20105	F. de Transferência	17	14,10187	27,5871
F. de Transferência	VAR20105	17	19,76645	27,5871

Fonte: Elaborado pelo pesquisador

Como última etapa de avaliação do modelo, foi calculado o MAPE das previsões relativo ao período de janeiro de 2007 a maio de 2012 que apresentou o valor de 9,95%, o que representa uma boa capacidade de descrever o comportamento da série no passado.

Após a realização da etapa de validação do modelo de função de transferência para o consumo de aço, este mostrou-se adequado para a realização de previsões.

4.2.5.18 Previsão do modelo de Função de Transferência para o Consumo de Aço

A Tabela 45 traz as previsões do consumo de aço do modelo de função de transferência no período de junho de 2012 a abril de 2013. Os resultados obtidos nesta previsão tiveram um MAPE de 17,28%.

Tabela 45 – Previsões mensais do modelo de Função de Transferência
Consumo de Aço – jun./12 – abr./2013

Ano / Meses	Previsto	Intervalo de Confiança (95%)		Valor Real	% de Erro
2012/06	1451000	855950	2046100,00	1988447	27,0%
2012/07	2237600	1594100	2881100,00	2184883	-2,4%
2012/08	1647700	975240	2320200,00	2968178	44,5%
2012/09	2039900	1349600	2730100,00	1818111	-12,2%
2012/10	1733800	1001900	2465700,00	2239756	22,6%
2012/11	1797900	1046300	2549500,00	2188962	17,9%
2012/12	1786700	1017200	2556200,00	1581586	-13,0%
2013/01	1568600	784180	2353100,00	1914027	18,0%
2013/02	1736700	950220	2523200,00	1819222	4,5%
2013/03	1699900	912040	2487700,00	1697094	-0,2%
2013/04	1619900	827150	2412600,00	2244515	27,8%

Fonte: Elaborado pelo pesquisador

4.2.6 Análise dos resultados dos modelos de previsão

Após a identificação dos modelos univariado e de função de transferência, para a variável consumo de aço, foi possível verificar que os dois modelos apresentam resultados significativos tanto na capacidade de descrever a série quanto na capacidade de previsão.

Com relação à capacidade dos modelos descreverem a série de consumo de aço no período de janeiro de 2007 a maio de 2012, o modelo univariado SARIMA (0, 1, 1)(1, 0, 0) apresentou um R^2 igual a 75,32% e MAPE de 24,19% e o modelo de função de transferência apresentou um MAPE igual a 9,95% e R^2 de 67,59%, ou seja, o modelo de função de transferência apresenta um menor erro e um coeficiente de determinação inferior, o que indica que o modelo univariado, mesmo tendo um MAPE superior, tem uma capacidade de explicar o comportamento da série melhor que o modelo de função de transferência.

Quanto à avaliação da capacidade dos dois modelos preverem o futuro, o modelo univariado SARIMA (0, 1, 1)(1, 0, 0) apresentou um MAPE de 15,26% e o modelo de função de transferência apresentou o MAPE de 17,28%, ou seja, o modelo univariado obteve melhores resultados de previsão.

No contexto de um processo de inteligência competitiva acredito que os dois modelos podem assumir funções muito importantes. O modelo univariado pode ser uma ferramenta para avaliar as séries históricas do consumo de aço, ou seja, através dele a empresa pode estudar os seus dados históricos e encontrar informações que podem ser transformadas em conhecimento fundamental dentro do processo decisório.

O modelo de função de transferência combinado como um monitoramento do mercado, para identificação das variáveis externas, pode assumir um papel equivalente ao do

univariado, mas com base na combinação de informações internas e externas da empresa para a empresa prever o possível consumo de aço.

A combinação dos resultados obtidos pelos dois modelos será capaz de trazer mais dados e informações para os gestores da empresa e subsidiar as decisões referentes ao planejamento de compra de Aço.

5 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES

Este capítulo apresenta considerações, limitações do trabalho e sugestões de trabalhos futuros. As considerações foram realizadas comparando os resultados obtidos em cada uma das etapas com os objetivos específicos.

O objetivo geral desta pesquisa foi identificar modelos de previsão de consumo de aço para análise das informações no processo de Inteligência Competitiva (IC) em uma empresa do setor metalúrgico. Para realizar este objetivo uma grande quantidade de informações referentes à empresa objeto de pesquisa, indicadores econômicos e do mercado do aço foram levantadas e avaliadas com o intuito de cumprir os objetivos específicos. A seguir serão apresentadas as considerações referentes à realização de cada um dos objetivos específicos.

Considerando o objetivo específico de explorar o fenômeno do *Big Data* (BD) e suas possibilidades para a mineração de dados, a pesquisa demonstrou que o BD é realmente uma fonte importante de dados, mas que devido a suas características de dimensão, diversidade de dados, estrutura dos dados, velocidade de crescimento dos dados e principalmente a limitação para armazenamento destes, exige que as empresas ou instituições governamentais mantenham uma estrutura física e humana muito bem preparada para o monitoramento dos dados existentes. Essa estrutura passa pelo desenvolvimento do processo de Inteligência Competitiva dentro das empresas, pois através da IC a empresa poderá manter o foco nas informações externas característica que para Trapanof (2001, p. 45) torna a IC um conjunto de ferramentas útil para a gestão da informação. É importante destacar que não basta reconhecer o BD como uma fonte de informações. A empresa deve saber explorar esta fonte através de um processo de IC, o que vai exigir que ela possua pessoal com qualificação para realizar este processo. Esta ideia vai ao encontro da visão de Fuld (2007), ele afirma que empresa precisa aprender a obter informações e conhecimento na Internet e uma forma de se realizar isto é estudar a inteligência através da utilização de analistas e especialistas da área.

O segundo objetivo específico foi o de utilizar o *Data Mining* (DM) como ferramenta de análise de grande volume de informação do BD. Mesmo que realizado de forma artesanal, o DM se apresentou como uma ferramenta muito importante no desenvolvimento do trabalho, pois através dele foi possível identificar as variáveis internas e externas para posterior identificação do modelo de previsão. A exploração do processo de DM confirmou a definição de Côtres (2002). Este afirma que o DM é um processo capaz de extrair conhecimento através da identificação de padrões e relacionamentos entre variáveis em grandes bancos de dados.

Também ficou evidente que com mais tempo e recursos financeiros uma empresa pode desenvolver técnicas de DM voltadas para suas necessidades específicas. O desenvolvimento destas técnicas, assim como a identificação das necessidades da empresa, mais uma vez passa pelo desenvolvimento de um processo de IC na empresa, onde o DM assumirá a função de fazer uma análise antecipativa dos eventos, prevendo tendências e comportamentos futuros. Com isto o DM permitirá que os gestores tenham mais informação para tomada de decisões (CARDOSO; MACHADO, 2008).

A identificação de variáveis econômicas com influência no mercado do aço foi o terceiro objetivo específico a ser realizado. A identificação destas variáveis foi feita através de uma pesquisa sobre o mercado do aço no Brasil, suas características e a verificação dos principais setores consumidores de aço no Brasil. A obtenção destas informações foi facilitada exatamente pela combinação do fenômeno BD com os processos de DM utilizados para obtenção das informações sobre o mercado do Aço no Brasil.

O quarto objetivo específico foi o de detectar as variáveis econômicas relacionadas com os clientes da Alfa. Este objetivo demandou o desenvolvimento de uma pesquisa muito grande e um processo, quase que artesanal, de mineração de dados em fontes de dados governamentais, mas ao final foi possível identificar um conjunto de variáveis que apresentavam relação com os clientes da empresa Alfa e que poderiam ter influência sobre o consumo de aço da empresa. O reconhecimento das variáveis é uma informação que pode permitir que a empresa planeje suas atividades com subsídios que facilitem a tomada de decisão, para Fachinelli et al. (2007, p. 173) este é um elemento chave para boas estratégias. Esta identificação também vai ao encontro das ideias de Kahaner (1997), ele afirma que programa de coleta e análise de informações sobre atividades de seus concorrentes ou tendências gerais dos negócios nada mais é do que um processo de inteligência competitiva. Se ele estiver presente em todos os níveis da empresa, pode ajudar a organização a resolver diferentes problemas.

O quinto objetivo foi o de verificar a correlação entre o consumo de aço da Alfa e o conjunto de variáveis externas. Para realizar este objetivo foi feito um cruzamento entre as informações obtidas sobre o mercado do aço brasileiro, as variáveis econômicas relacionadas com os clientes da empresa Alfa e literatura sobre modelos de previsão. Com base no cruzamento das informações foi realizado um estudo de correlação entre as variáveis identificadas e o consumo de aço da empresa. Foi possível identificar um conjunto de variáveis que apresentaram correlação com o consumo de aço entre 0,6 e 0,8. Segundo Tabela 14, desenvolvida por Santos (2007), em um intervalo de variação entre -1 e 1 pode ser

classificada como uma correlação moderada. A verificação da correlação entre as variáveis econômicas e o consumo de aço proporcionou o desenvolvimento de um modelo, representado pela Figura 5, que poderá servir de base de monitoramento do mercado em um processo de IC na empresa.

O último objetivo específico da pesquisa era o de construir modelos de previsão de consumo de aço através da análise de Séries Temporais. Este objetivo foi realizado de forma mais ampla, pois não se construiu um modelo e sim dois modelos. O primeiro modelo construído foi um modelo univariado sem intervenção. O objetivo da construção deste modelo foi o de permitir que a empresa monitore a sua demanda de aço com base nos dados históricos da própria empresa, ou seja, o uso das informações internas em um processo de IC. O modelo identificado apresentou bons resultados tanto na sua capacidade de descrever o passado da série de consumo de aço quanto na capacidade de previsão. O segundo modelo consistiu em um modelo de previsão com Função de Transferência. A construção deste teve como objetivo a combinação das informações internas da empresa com variáveis econômicas externas que tivessem relação com os clientes da empresa, e correlação com o consumo de aço. Este modelo apresentou melhores resultados na capacidade de descrever o comportamento passado da série de consumo de aço da empresa comparado com o modelo univariado, porém um resultado um pouco inferior na capacidade de previsão. A utilização dos modelos encontrados pode ser uma ferramenta muito importante em um processo de IC, pois possibilitarão uma previsão de consumo de aço através de um monitoramento conjunto das séries históricas da empresa e de informações provenientes do mercado externo. Na realização deste último objetivo, mais uma vez ficou caracterizada a Inteligência Competitiva, pois para atingi-lo foi necessário um processo dinâmico de combinação de diversas áreas do conhecimento para coletar, organizar e disseminar o conhecimento nas informações obtidas na empresa e no mercado com o intuito de auxiliar a tomada decisões.

Por fim, o estudo atingiu o objetivo geral ao combinar os resultados obtidos em cada um dos objetivos específicos. Ao atingir este objetivo foi possível confirmar o modelo teórico apresentado na Figura 1, onde fenômeno *Big Data*, *Data Mining* e modelos de previsão mostram-se como instrumentos de análise em um processo de Inteligência Competitiva. A relevância deste estudo está exatamente no fato de que a utilização do modelo teórico possibilitará tanto para a área acadêmica quanto para os profissionais de áreas afins um novo instrumento de Inteligência Competitiva.

5.1 LIMITES DA PESQUISA

A presente pesquisa apresentou alguns limitações de ordem teórica, técnica e operacionais, entre elas podem-se destacar:

- literatura que apresentasse a combinação dos conceitos de Inteligência Competitiva, *Big Data*, *Data Mining* e Previsão de Demanda;
- conhecimento teórico de estatística bastante superior ao conhecimento desenvolvido em cursos de graduação e pós-graduação, o que demandou muitas horas de estudo e pesquisa. A aplicação desta metodologia em uma empresa vai exigir que a mesma tenha profissionais com alto nível de conhecimento, o que é muito difícil de encontrar no mercado hoje em dia;
- necessidade de séries históricas com um número mínimo de observações;
- pacote computacional que realizasse todos os processos e testes necessários para o desenvolvimento da pesquisa. Para superar esta dificuldade, foi necessário utilizar um conjunto de pacotes computacionais que suprisse as necessidades da pesquisa.
- impossibilidade de implantar e testar o modelo na empresa devido ao tempo necessário para a realização desta tarefa e os limites de tempo para realização do estudo.

5.2 TRABALHO FUTUROS

A presente pesquisa abriu um conjunto de possibilidades de trabalhos futuros relacionados com o tema abordado. A seguir, apresentam-se algumas dessas possibilidades:

- implantar o modelo desenvolvido na empresa objeto de pesquisa;
- planejar e desenvolver cursos de qualificação de mão de obra;
- aplicar a metodologia de pesquisa em empresas de outras áreas para comparar os resultados obtidos e confirmar a sua viabilidade;
- combinar a metodologia aplicada utilizando o conceito de redes neurais ou inteligência artificial no aperfeiçoamento do processo de *Data Mining*;
- utilizar a metodologia apresentada na construção de um modelo de previsão para mercado de capitais.

REFERÊNCIAS

AKAIKE, H., **A new look at the statistical model identification**. IEEE Transactions on Automatic Control, AC-19, p.716-723, 1974.

AKAIKE, H., **Maximum likelihood identification of Gaussian autoregressive moving average models**. Biometrika, v. 60, p. 255-265, 1973.

AKAIKE, H., **On entropy maximization principle**. In: Applications of Statistics, Amsterdam. Ed. P.R. Krishnaiah. p. 27-41, 1977.

AMOROSO, E. Devanir. **Um estudo de caso sobre mineração de dados como instrumento de aprendizado para o investidor do mercado de ações**. 2009. 121 f. Dissertação (Mestrado) - Universidade Católica de Brasília, Brasília, 2009. Disponível em: <http://www.bdttd.ucb.br/tede/tde_busca/arquivo.php?codArquivo=1023>. Acesso em: 18 dez. 2012.

ANSOFF, H. Igor; MCDONNELL, Edward J.. **Implantando a administração estratégica**. 2.ed. São Paulo: Atlas, 1993. 590 p.

BACEN, Banco Central do Brasil. **SGS - Sistema Gerenciador de Séries Temporais**. Disponível em:<<https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>>. Acesso em: 10 jun. 2012.

BALA , P. Kumar. Improving inventory performance with clustering based demand forecasts. **Journal of Modelling in Management**, 2012, Vol. 7 Iss: 1 pp. 23 – 37. Disponível em: <<http://dx.doi.org/10.1108/17465661211208794>>. Acesso em: 25 nov. 2012.

BEGUIN, J. M.; GOURIEROUX, C. C.; MONFORT, A.. Identification of a mixed autoregressive-moving average process: The corner method. **Time Series**, Amsterdam, p. 423-436. 1980.

BIBLIOTECA DIGITAL BRASILEIRA DE TESES E DISSERTAÇÕES BDTD (Brasil) (Org.). **Biblioteca Digital Brasileira de Teses e Dissertações**. Disponível em: <<http://bdtd.ibict.br/>>. Acesso em: 01 jan. 2013.

BOX, G. E. P. & JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. Revised edition. San Francisco, 1970.

BOX, G. E. P. & JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. Revised edition. San Francisco: Holden-Day, 1976.

BOX, G. E. P. & Pierce, D. A. Distribution of residual autocorrelations in autoregressive integrated moving average time series models. **Journal of the American Statistical Association**. 65, 1509-1526. 1970.

BOX, G. E. P. & TIAO, G.C. Intervention analysis with applications to economic and environmental problems. **Journal of the American Statistical Association**, 70, 70-79, 1975.

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

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. 4. ed. Hoboken, N.J.: Wiley, 2008. xxiv, 746 p.

BOX, George E. P.; JENKINS, Gwilym M.; REINSEL, Gregory C. **Time series analysis: forecasting and control**. 4.ed. Hoboken, N.J.: Wiley, 2008. xxiv, 746 p.

BRAGA, L. P. V.. **Introdução à mineração de dados**. 2. ed. ampl. e rev. Rio de Janeiro: E-Papers, 2005. 211 p.

BRASIL. Ministério da Fazenda. - (Org.). **Emissão de Comprovante de Inscrição e de Situação Cadastral**: Receita Federal. Disponível em: <http://www.receita.fazenda.gov.br/PessoaJuridica/CNPJ/cnpjreva/Cnpjreva_Solicitacao.asp> . Acesso em: 20 maio 2012.

BRUCE, J.; GEORGE, R.. *Analyzing Intelligence: origins, obstacles and innovations*. Washington: **Georgetown University Press**, 2008. 340 p.

BUENO, R. D. L. S.. **Econometria de séries temporais**. São Paulo: Cengage Learning, 2008. xviii, 299 p.

BUFFA, E. S.. **Administração da produção**. 3. ed. Rio de Janeiro: Livros Técnicos e Científicos, 1972. 2 v.

CAMARGO, M. E.; SOUZA, R. C.. **Análise e previsão de Séries Temporais: Os Modelos ARIMA**. Ijuí: Sedigraf, 1996. 242 p.

CARDOSO, O. N. P.; MACHADO, R. T. M.. Gestão do conhecimento usando Data Mining: estudo de caso na Universidade Federal de Lavras. **Rev. Adm. Pública**, Rio de Janeiro, v. 42, n. 3, jun. 2008 . Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-76122008000300004&lng=pt&nrm=iso>. Acesso em 01 fev. 2013. <http://dx.doi.org/10.1590/S0034-76122008000300004>.

CERVO, A.; BERVIAN, P. **Metodologia científica**. São Paulo: Prentice-Hall, 2002.

CHANG, I.; TIAO, G. C.; CHEN, C. Estimation of Time Series Parameters in the Presence of Outliers, **Technometrics**, v.30, n. 2, 193-204. 1988.

CHOWDHURY, S.. Data Mining tools and technologies for competitive business advantage. **European Journal of Management** 9.2 (2009): 96+. Academic OneFile. Disponível em: <<http://go.galegroup.com/ps/i.do?id=GALE%7CA208535146&v=2.1&u=capes2&it=r&p=ACADEMICONE&sw=w>>. Acesso em: 25 nov. 2012.

CHYE, K. H., GERRY, C. K. L.. Data Mining and customer relationship marketing in the banking industry. **Singapore Management Review** 24.2 (2002): 1+. Academic OneFile. Disponível em: <<http://go.galegroup.com/ps/i.do?id=GALE%7CA119370565&v=2.1&u=capes2&it=r&p=ACADEMICONE&sw=w>>. Acesso em: 25 Nov. 2012.

ÇIFLIKLI, C., ÖZYIRMIDOKUZ, E. K.. Enhancing product quality of a process. **Industrial Management & Data Systems**, Vol. 112 Iss: 8, pp.1181 – 1200. Disponível em: <<http://dx.doi.org/10.1108/02635571211264618>>. Acesso em: 25 nov. 2012.

CORRÊA, H. L.; GIANESI, I. G. N.; CAON, M.. **Planejamento, programação e controle da produção: MRP II/ERP**. 5. ed. São Paulo: Atlas, 2007. xx, 434 p.

CÔRTEZ, S. C.; PORCARO, R. M.; LIFSCHITZ, S.. **Mineração de Dados: Funcionalidades, Técnicas e Abordagens**, PUC-Rio, maio 2002. Disponível em: <http://www.dbd.puc-rio.br/depto_informatica/02_10_cortes.pdf>. Acesso em: 04 ago. 2012.

CRESWELL, J. Projeto de pesquisa: métodos qualitativo, quantitativo e misto. 2. ed. Porto Alegre: Bookman, 2007.

DAVENPORT, T. H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, [1998]. 316 p.

DAVENPORT, T. H.; HARRIS, J. G. **Competição analítica**: vencendo através da nova ciência. Rio de Janeiro: Elsevier, c2007. 241 p.

DIAS, M. A. P.. **Administração de materiais**: uma abordagem logística. 4. ed. São Paulo: Atlas, 1993. 399 p.

DICKEY, D. A.; FULLER, W. A.. Distribution of the estimation for autoregressive time series with a unit root. **Journal Of de American Statistical Association**,, p. 427-431. 01 jun. 1979. Disponível em: <<http://www.deu.edu.tr/userweb/onder.hanedar/dosyalar/1979.pdf>>. Acesso em: 30 maio 2013.

DICKEY, D.A.; FULLER, W.A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49:1057-1072, 1981

DRUCKER, P. F.. **Sociedade pós-capitalista**. São Paulo, SP: Thomson, c1993. 186 p.

EMC - Big Data: Big Opportunities to Create Business Value. **EMC Corporation, Hopkinton, Massachusetts**, nov. 2011. Disponível em: <<http://www.emc.com/microsites/cio/articles/big-data-big-opportunities/LCIA-BigData-Opportunities-Value.pdf>>. Acesso em 03 ago 2012.

FACHINELLI, A.C.; MATTIA, O.M.; RECH, J.; ROVEDA, V.. **Vigília e inteligência estratégica: ferramentas metodológicas para a definição de estratégias de relações públicas**. Conexão - Comunicação e Cultura, Vol. 6, No 11 (2007). Acesso em: 15 out. 2012.

FAYARD, P.. **O jogo da interação: informação e comunicação em estratégia**. Caxias do Sul: Educs, 2000. p. 59.

FEINLEIB, D., **Big Data Trends**, Forbes, 2012. Disponível em <http://www.forbes.com/sites/davefeinleib/2012/07/24/big-data-trends/>. Acesso em: 05 ago 2012.

FERNANDES, M. P.. **Descoberta de conhecimento em bases de dados e estratégias de relacionamento com clientes: um estudo no setor de serviços**. 2007. 122 f. Dissertação (Mestrado) - Curso de Administração de Empresas, Universidade Presbiteriana Mackenzie, São Paulo, 2007. Disponível em: <http://tede.mackenzie.com.br//tde_busca/arquivo.php?codArquivo=702>. Acesso em: 27 jan. 2013.

FERREIRA, J. B.. **Mineração de Dados na Retenção de Clientes em telefonia celular**. 2005. 93 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Puc-rio, Rio de Janeiro, 2005. Disponível em: <<http://www.maxwell.lambda.ele.puc-rio.br/acessoConteudo.php?nrseqoco=20054>>. Acesso em: 15 nov. 2012.

FISCHER, S. L.; **Séries univariantes de tempo - metodologia de Box & Jenkins**. Porto Alegre: Fundação de Economia e Estatística, 1982. 186p. (Teses 4)

FLEISHER, C. S., WRIGHT, S., ALLARD H. T. The role of insight teams in integrating diverse marketing information management techniques, **European Journal of Marketing**, Vol. 42 Iss: 7 pp. 836 – 851. Disponível em: <<http://dx.doi.org/10.1108/03090560810877187>>. Acesso em: 25 nov. 2012.

FOX, A. J. Outliers in Time Series. **Journal of the Royal Statistical Society**, Ser. B, 34, 350-363. 1972.

FRAGOSO, B. B.. **Método para a criação de um processo de previsão da demanda de vendas**. 2010. 138 f. Dissertação (Mestrado) - Curso de Engenharia Mecânica, Faculdade de Engenharia Mecânica, Universidade Estadual De Campinas, Campinas, Sp, 2009. Disponível em: <<http://www.bibliotecadigital.unicamp.br/document/?code=000771699>>. Acesso em: 05 out. 2012.

FRANKEL, D. A.. Big Data and risk management. **Risk Management.**, New York, v. 8, n. 59, p.13-13, 1 out. 2012.
[Http://go.galegroup.com/ps/i.do?action=interpret&id=gale%7ca305454213&v=2.1&u=capes2&it=r&p=aone&sw=w&authcount=1](http://go.galegroup.com/ps/i.do?action=interpret&id=gale%7ca305454213&v=2.1&u=capes2&it=r&p=aone&sw=w&authcount=1). Disponível em:
 <<http://go.galegroup.com/ps/i.do?action=interpret&id=GALE%7CA305454213&v=2.1&u=capes2&it=r&p=AONE&sw=w&authCount=1>>. Acesso em: 1 fev. 2013.

FREITAS, H. M. R. e JANISSEK-MUNIZ, R. **Uma Proposta de Plataforma para Inteligência Estratégica**. In: Congresso Ibero Americano de Gestão do Conhecimento e Inteligência Competitiva (GeCIC), 2006, Curitiba. Anais... 2006.

FREITAS, P.; FREITAS, H.; GENSAS, G. A Inteligência Estratégica no apoio à Decisão de uma Empresa de TI. **Revista REAd**. Vol. 13 No. 2, mai - jun de 2007.
http://www.ufrgs.br/gianti/files/artigos/2007/2007_205_READ.pdf. Acesso em: 01 jan. 2013.

FULD, L. M. **Inteligência competitiva**: como se manter à frente dos movimentos da concorrência e do mercado. Rio de Janeiro: Elsevier, 2007. 235 p.

FULD, L.M. **The new competitor intelligence**: the complete resource for finding, analyzing, and using information about your competitors. New York, US: J. Wiley, c1995. xiii, 482 p.

FURASTÉ, P. A.. **Normas técnicas para o trabalho científico**: explicitação das normas da ABNT. 16.ed. atual. e ampl. Porto Alegre: [s.n.], 2012. 230 p.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GRANGER, C. W. J.; NEWBOLD, Paul. **Forecasting Economic Time Series**. New York : **Academic Press**, 1977.

GUJARATI, D. N.. **Basic econometrics**. 3.ed. New York, US: McGraw-Hill, 1995. 838 p.

HAIR JR., J. F. Jr.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. **Análise Multivariada de Dados**. 5. ed. Porto Alegre: Bookmann, 2005.

HARRIS, R. L. **Globalization and globalismo in Latin America**: contending perspectives. *Latin American Perspectives*, v. 29, 2002.

HÜBNER, M. L. F.; BAPTISTA, M. M.; BERTÉLI, M. O. (Org.). **Guia para elaboração de trabalhos acadêmicos. Caxias do Sul**: UCS, 2012. Disponível em: <http://biblioteca.ucs.br/Manualtrabalhosacademicos_2012.pdf>. Acesso em: 01 out. 2012.

IABR, Instituto De Aço Brasil. **Dados do Mercado**. Disponível em: <<http://www.acobrasil.org.br/site/portugues/numeros/numeros--mercado.asp>>. Acesso em: 01 jun. 2012.

IBGE, Instituto Brasileiro de Geografia e Estatística. **Comissão Nacional de Classificação: Tabela CNAE**. Disponível em: <<http://www.cnae.ibge.gov.br/>>. Acesso em: 04 jun. 2012.

IDC - International Data Corporation. **THE 2011 Digital Universe Study**: Extracting Value from Chaos – Big Value from Big Data. Disponível em: <<http://brazil.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm>>. Acesso em 05 ago. 2012.

JANISSEK-MUNIZ, R.; LESCA, H.; FREITAS, H. Desenvolvimento da capacidade de antecipação pela identificação e captação de indícios antecipativos em contexto de Inteligência Estratégica Antecipativa. In: 4º. IFBAE Congresso do IFBAE, 2007, Porto Alegre. **Anais do 4. IFBAE** – Congresso do IFBAE, 2007.

JOHNSON, J. E.. Big Data + big analytics = big opportunity: Big Data is dominating the strategy discussion for many financial executives. As these market dynamics continue to evolve, expectations will continue to shift about what should be disclosed, when and to whom. **Financial Executive**, ., v. 6, n. 28, p.51-53, 01 jul. – ago.. 2012. Bimestral. Disponível em:

<<http://go.galegroup.com/ps/i.do?id=GALE%7CA298854598&v=2.1&u=capex2&it=r&p=ONE&sw=w>>. Acesso em: 12 dez. 2012.

KAHANER, L., *Competitive Intelligence: How to Gather, Analyse, and Use Information to Move your Business to the top*. New York: Touchstone, 1997.

KAISER, H., F. *A Second – Generation Litte Jiffy*. 1. ed. New York: Psychochometrika 35, 1970.

KURRLE, M. A.. **Métodos para previsão de demanda de veículos novos**: estudo de caso em uma concessionária de automóveis. 2004. 115 f. Dissertação (Mestrado) - Curso de Mestrado Profissionalizante em Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/5266>>. Acesso em: 05 out. 2012.

KWON, O., SIM, J. M.. Effects of data set features on the performances of classification algorithms. **Expert Systems With Applications**, N.c., n. , p.1847-1857, 1 abr. 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417412010718>>. Acesso em: 12 dez. 2012.

LEMOS, F. O.. **Metodologia para seleção de métodos de previsão de demanda**. 2006. 183 f. Dissertação (Mestrado) - Curso de Engenharia da Produção, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/5949>>. Acesso em: 05 out. 2012.

LESCA, H., FREITAS, H. e JANISSEK-MUNIZ, R. Inteligência Antecipativa para Decisão Estratégica. **Revista Perfil**. Unisc, 2003. http://www.ufrgs.br/gianti/files/artigos/2005/2005_165_rev_eGIANTI.pdf . Acesso em: 05 de fev. 2013.

LII, K.. Transfer Function Model Order And Parameter Estimation. **Journal Of Time Series Analysis**, ., p. 153-169. maio 1985. Disponível em: <<http://dx.doi.org/10.1111/j.1467-9892.1985.tb00406.x>>. Acesso em: 14 dez. 2012.

LIU, L.; HANSSSENS, D. M.. Identification of multiple input transfer function models. **Communications In Statistics**, Los Angeles, n. , 1982, p. 297-314.

LJUNG, G. M. (1993). On Outlier Detection in Time Series. **Journal of the Royal Statistical Society**, v. 55, 1993, p. 559-567, 1993.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R.. **Forecasting: Methods and Applications**. 3. ed. New York: John&Sons, 1998.

MALHOTRA, N. K. **Pesquisa de marketing**: foco na decisão. 3. ed. São Paulo: Pearson, 2010. xx, 491 p.

MARCANO AULAR, Yelitza Josefina; TALAVERA PEREIRA, Rosalba. Minería de Datos como soporte a la toma de decisiones empresariales. **Opcion**, Maracaibo, v. 23, n. 52, jan. 2007 . Disponível em <http://www.scielo.org.ve/scielo.php?script=sci_arttext&pid=S1012-15872007000100008&lng=pt&nrm=iso>. acessos em 01 fev. 2013.

MAYER, R., R. **Administração da Produção**. São Paulo: Atlas, 1990.

MCKINSEY, Global Institute. **Big Data: The next frontier for innovation, competition, and productivity**, 2011.

MENEZES, E. M.. Inteligência competitiva: uma revisão de literatura. **Revista Digital de Biblioteconomia & Ciência da Informação**, Campinas, v. 3, n. 1, jul./dez. 2005, p. 103-130.

MME, Ministério de Minas e Energia. **Anuário Estatístico do Setor Metalúrgico 2011 - ano base 2010**. Disponível em: <<http://www.mme.gov.br/sgm/menu/publicacoes.html>>. Acesso em: 20 jun. 2012.

MONKS, J. G.. **Administração da produção**. São Paulo: McGraw-Hill, 1987. 502 p.

MOREIRA, D. A.. **Administração da produção e operações**. São Paulo: Thomson, 1993. xii, 619 p.

MORETTIN, P. A.; TOLOI, C. M. C.. Modelos para Previsão de Séries Temporais. In : **13º Colóquio Brasileiro de Matemática**. Rio de Janeiro : [s.n.], 1981.

MORETTIN, P. A.; TOLOI, C. M.. **Análise de séries temporais**. 2. ed. São Paulo: Edgard Blucher, 2006. 538 p.

MORETTIN, P. A.; TOLOI, C. M.. **Séries temporais**. 2. ed. São Paulo: Atual, 1987. 136 p.

NONAKA, I.; TAKEUCHI, H.. **Criação de conhecimento na empresa: como as empresas japonesas geram a dinâmica da inovação**. 10. ed. Rio de Janeiro: Campus, 2002. 358 p.

OZAKI, T. On the order determination of ARIMA models. **Applied Statistics**. V. 26, 1977, p. 290-301.

PASSARI, A. F. L. **Exploração de dados atomizados para previsão de vendas no varejo utilizando redes neurais**. 2003. Dissertação (Mestrado em Administração) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2003. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/12/12139/tde-30082004-105146/>>. Acesso em: dez. 2013.

PELLEGRINI, F.R.; FOGLIATTO, F. S.. Passos para implantação de sistemas de previsão de demanda: técnicas e estudo de caso. **Produção**. São Paulo, v. 11, n. 1, June 2001. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-65132001000100004&lng=en&nrm=iso>. Acesso em: 17 nov. 2012.

PEREIRA, R., S. **A Estatística e Suas Aplicações**. Porto Alegre: Edição do Autor, 1979.

PIERCE, D. A.. Residual correlations and diagnostic checking in dynamic-disturbance time series models. **Journal of the American Statistical Association**. 67, 1972, 636-40.

PORTER, M. E. **Competição= On competition : estratégias competitivas essenciais**. ed. rev. e ampl. Rio de Janeiro: Elsevier, 2009.

PORTER, M. E. **Estratégia competitiva: técnicas para análise de indústrias e da concorrência**. 12. ed. Rio de Janeiro: 1997. Campus, 362 p.

PORTER, M. E.; MONTGOMERY, C. A.. **Estratégia: a busca da vantagem competitiva**. Rio de Janeiro: Campus, 1998. 501 p.

PRIESTLEY, M. B.. Fitting relationship between time series. **International Statistical Institute**, Washington D.C., n. , ago. 1971, p.1-29.

RIGGS, J. L.. **Administração da produção: planejamento, análise e controle – uma abordagem sistêmica**. São Paulo: Atlas, 1976. 2 v.

SAMPIERI, R. H.; COLLADO, C. F.; LUCIO, P. B..**Metodología de la investigacion**. 2. ed. México: Mckraw-hill, 2001. 501 p.

SANTOS, C. Estatística Descritiva - Manual de Auto-aprendizagem, Lisboa, Edições Silabo, 2007.

SANTOS, R. S.. **Aplicação de um modelo preditivo de mineração de dados no apoio à decisão de crédito**. 2006. 92 f. Dissertação (Mestrado) - Curso de Ciências da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2006. Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/VALA-6T7R6V/mestrado___roberto_de_souza_santos.pdf?sequence=1>. Acesso em: 18 dez. 2012.

SCHAEFFER, A. G..**Data Mining no Varejo: estudo de caso para loja de materiais de construção**. 2003. 86 f. Dissertação (Mestrado) - Curso de Ciências da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/3430/000400435.pdf?sequence=1>>. Acesso em: 16 nov. 2012.

SCHERRER, C. M.. **A indústria siderúrgica brasileira: um estudo econométrico**. 2006. 89 f. Dissertação (Mestrado) - Curso de Ciências Econômicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006. Disponível em: <<http://hdl.handle.net/10183/8065>>. Acesso em: 05 out. 2012.

SOUZA, R. C.. Modelos Estruturais para Previsão de Séries Temporais: Abordagens Clássica e Bayesiana. In : **17º Colóquio Brasileiro de Matemática**. Rio de Janeiro, 1989.

STAREC, C.; GOMES, E. B. P.; CHAVES, Jorge Bezerra Lopes. **Gestão estratégica da informação e inteligência competitiva**. São Paulo: Saraiva, 2005. xvi, 351 p.

STEWART, T. A.. **Capital intelectual: a nova vantagem competitiva das empresas**. 7. ed. Rio de Janeiro: Campus, 1998.

TARAPANOFF, K.. **Inteligência organizacional e competitiva**. Brasília: UnB, 2001.

TARAPANOFF, K.; ARAUJO JUNIOR, R. H.; CORMIER, P. M. J.. Sociedade da informação e inteligência em unidades de informação. **Ci. Inf.**, Brasília, v. 29, n. 3, dez. 2000 . Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652000000300009&lng=pt&nrm=iso>. acessos em 17 nov. 2012. <http://dx.doi.org/10.1590/S0100-19652000000300009>.

TENA, M., J., & Comai, A. 2005. Análisis e Interpretación de la Información de Inteligencia. PUZZLE - **Revista Hispana de la Inteligencia Competitiva**,4: 11-17.

THE ECONOMIST, **A Special Report on managing information: Data, data everywhere**. fev. 2010. Disponível em: <<http://www.economist.com/node/15557443>> Acesso em 05 ago.2012.

TIAO, G. C.; BOX, G. E. P.; HAMMING, W. J. Analysis of Los Angeles photochemical smog data: a statistical overview. **Journal of the Air Pollution Control Association**, [S.l.], v. 25, n. 3, p. 260-268, mar. 1975.

TSAY, R. S. Outliers, Level Shifts, and Variance Changes in Time Series. **Journal of Forecasting**, v. 7, 1-20. 1988.

TSAY, R. S. Time Series Model Specification in the Presence of Outliers. **Journal of the American Statistical Association**, v. 81, n. 393, 132-141, 1986.

VANDAELE, W.. **Applied Time series and Box-Jenkins models**. New York: Academic Press, 1983. 417 p

WERNER, L.; RIBEIRO, J. L. D.. Previsão de demanda: uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais. **Gest. Prod.**, São Carlos, v. 10, n. 1, abr. 2003 . Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-530X2003000100005&lng=pt&nrm=iso>. Acesso em 17 nov. 2012.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 4. ed. Porto Alegre: Bookman, 2010. xviii, 248 p.

APÊNDICE A- ANÁLISE DAS CORRELAÇÕES CRUZADAS DOS RESÍDUOS

THE RESIDUAL CROSS-CORRELATION ANALYSIS

CROSS-CORRELATION ANALYSIS

THE CROSS-CORRELATIONS

(A DIAGNOSTIC CHECK FOR TRANSFER MODEL SUFFICIENCY)

INPUT SERIES : PREWHITENED VAR1065.DTA
 OUTPUT SERIES : THE ESTIMATED RESIDUALS FROM THE TRANSFER FUNCTION MODEL

MEAN OF THE INPUT SERIES : -.29735E-03
 STANDARD DEVIATION : .61712E-01
 NUMBER OF OBSERVATIONS : 48

THE CROSS-CORRELATIONS

LAGS 0- 7	.048	-.036	-.243	-.062	.003	-.013	.057	-.094
STANDARD ERROR	(.147)	(.149)	(.151)	(.152)	(.154)	(.156)	(.158)	(.160)
LAGS 8- 15	-.095	-.163	.141	-.064	.110	.155	.007	.176
STANDARD ERROR	(.162)	(.164)	(.167)	(.169)	(.171)	(.174)	(.177)	(.180)
LAGS 16- 23	.054	.044	.082	.161	.048	-.090	-.129	-.103
STANDARD ERROR	(.183)	(.186)	(.189)	(.192)	(.196)	(.200)	(.204)	(.209)
LAG 24	.153							
STANDARD ERROR	(.213)							

----- section 28 -----

THE CROSS-CORRELATIONS

(THE DIAGNOSTIC CHECK FOR UNIDIRECTIONAL CAUSALITY)
(I.E. LACK OF FEEDBACK)

INPUT SERIES : THE ESTIMATED RESIDUALS FROM THE TRANSFER FUNCTION MODEL
OUTPUT SERIES : PREWHITENED VAR1065.DTA

THE CROSS-CORRELATIONS

LAGS 0- 7	.048	-.120	.120	.066	-.154	-.135	-.298	.030
STANDARD ERROR	(.147)	(.149)	(.151)	(.152)	(.154)	(.156)	(.158)	(.160)
LAGS 8- 15	.146	.068	-.046	.023	.164	.094	.006	-.095
STANDARD ERROR	(.162)	(.164)	(.167)	(.169)	(.171)	(.174)	(.177)	(.180)
LAGS 16- 23	-.044	.100	.052	.086	.014	.048	.001	.005
STANDARD ERROR	(.183)	(.186)	(.189)	(.192)	(.196)	(.200)	(.204)	(.209)
LAG 24	-.114							
STANDARD ERROR	(.213)							

----- section 29 -----

THE CROSS-CORRELATIONS

(A DIAGNOSTIC CHECK FOR TRANSFER MODEL SUFFICIENCY)

INPUT SERIES : PREWHITENED VAR1386.DTA
OUTPUT SERIES : THE ESTIMATED RESIDUALS FROM THE TRANSFER FUNCTION MODEL

MEAN OF THE INPUT SERIES	:	.10031E+00
STANDARD DEVIATION	:	.85142E+00
NUMBER OF OBSERVATIONS	:	48

THE CROSS-CORRELATIONS

LAGS 0- 7	-.151	-.040	.075	-.049	-.054	.136	.008	-.063
STANDARD ERROR	(.147)	(.149)	(.151)	(.152)	(.154)	(.156)	(.158)	(.160)
LAGS 8- 15	-.088	-.093	.170	-.052	.274	.136	.023	.073
STANDARD ERROR	(.162)	(.164)	(.167)	(.169)	(.171)	(.174)	(.177)	(.180)
LAGS 16- 23	.042	.095	.064	.096	-.114	-.069	-.105	-.100
STANDARD ERROR	(.183)	(.186)	(.189)	(.192)	(.196)	(.200)	(.204)	(.209)
LAG 24	.287							
STANDARD ERROR	(.213)							

----- section 30 -----

THE CROSS-CORRELATIONS

(THE DIAGNOSTIC CHECK FOR UNIDIRECTIONAL CAUSALITY)
(I. E. LACK OF FEEDBACK)

INPUT SERIES : THE ESTIMATED RESIDUALS FROM THE TRANSFER FUNCTION MODEL
OUTPUT SERIES : PREWHITENED VAR1386.DTA

THE CROSS-CORRELATIONS

LAGS 0- 7	-.151	.077	-.204	.137	.017	-.066	-.225	-.023
STANDARD ERROR	(.147)	(.149)	(.151)	(.152)	(.154)	(.156)	(.158)	(.160)
LAGS 8- 15	.138	-.113	.182	.008	.068	-.057	-.156	-.047
STANDARD ERROR	(.162)	(.164)	(.167)	(.169)	(.171)	(.174)	(.177)	(.180)
LAGS 16- 23	.110	-.051	.066	-.019	.038	-.024	-.002	-.022
STANDARD ERROR	(.183)	(.186)	(.189)	(.192)	(.196)	(.200)	(.204)	(.209)
LAG 24	-.087							
STANDARD ERROR	(.213)							

----- Section 31 -----

THE CROSS-CORRELATIONS

(A DIAGNOSTIC CHECK FOR TRANSFER MODEL SUFFICIENCY)

INPUT SERIES : PREWHITENED V11066.DTA
OUTPUT SERIES : THE ESTIMATED RESIDUALS FROM THE TRANSFER FUNCTION MODEL

MEAN OF THE INPUT SERIES	:	-.10207E-02
STANDARD DEVIATION	:	.49025E-01
NUMBER OF OBSERVATIONS	:	48

THE CROSS-CORRELATIONS

LAGS 0- 7	.016	.015	-.131	-.220	.019	-.078	.096	-.042
STANDARD ERROR	(.147)	(.149)	(.151)	(.152)	(.154)	(.156)	(.158)	(.160)
LAGS 8- 15	-.055	-.266	.092	-.055	.006	.193	.154	-.074
STANDARD ERROR	(.162)	(.164)	(.167)	(.169)	(.171)	(.174)	(.177)	(.180)
LAGS 16- 23	.070	.085	.012	.260	.018	-.114	-.050	-.260
STANDARD ERROR	(.183)	(.186)	(.189)	(.192)	(.196)	(.200)	(.204)	(.209)
LAG 24	.267							
STANDARD ERROR	(.213)							

----- Section 32 -----

THE CROSS-CORRELATIONS

(THE DIAGNOSTIC CHECK FOR UNIDIRECTIONAL CAUSALITY)
(I.E. LACK OF FEEDBACK)

INPUT SERIES : THE ESTIMATED RESIDUALS FROM THE TRANSFER FUNCTION MODEL
OUTPUT SERIES : PREWHITENED V11066.DTA

THE CROSS-CORRELATIONS

LAGS 0- 7	.016	.038	-.118	.190	-.064	-.102	-.333	.172
STANDARD ERROR	(.147)	(.149)	(.151)	(.152)	(.154)	(.156)	(.158)	(.160)
LAGS 8- 15	.072	.051	.122	-.059	.050	.026	-.015	-.048
STANDARD ERROR	(.162)	(.164)	(.167)	(.169)	(.171)	(.174)	(.177)	(.180)
LAGS 16- 23	.125	.100	.029	.012	-.051	.122	.022	-.011
STANDARD ERROR	(.183)	(.186)	(.189)	(.192)	(.196)	(.200)	(.204)	(.209)
LAG 24	-.117							
STANDARD ERROR	(.213)							

----- Section 33 -----

THE CROSS-CORRELATIONS

(A DIAGNOSTIC CHECK FOR TRANSFER MODEL SUFFICIENCY)

INPUT SERIES : PREWHITENED VAR20105.DTA
OUTPUT SERIES : THE ESTIMATED RESIDUALS FROM THE TRANSFER FUNCTION MODEL

MEAN OF THE INPUT SERIES	:	.77270E-02
STANDARD DEVIATION	:	.42173E-01
NUMBER OF OBSERVATIONS	:	48

THE CROSS-CORRELATIONS

LAGS 0- 7	-.044	-.039	-.093	-.156	.054	.074	.190	-.077
STANDARD ERROR	(.147)	(.149)	(.151)	(.152)	(.154)	(.156)	(.158)	(.160)
LAGS 8- 15	-.021	-.085	.122	.070	.095	.197	.126	.007
STANDARD ERROR	(.162)	(.164)	(.167)	(.169)	(.171)	(.174)	(.177)	(.180)
LAGS 16- 23	.143	-.021	.063	.204	-.089	-.087	-.102	-.122
STANDARD ERROR	(.183)	(.186)	(.189)	(.192)	(.196)	(.200)	(.204)	(.209)
LAG 24	.092							
STANDARD ERROR	(.213)							

----- Section 34 -----

THE CROSS-CORRELATIONS

 (THE DIAGNOSTIC CHECK FOR UNIDIRECTIONAL CAUSALITY)
 (I. E. LACK OF FEEDBACK)

INPUT SERIES : THE ESTIMATED RESIDUALS FROM THE TRANSFER FUNCTION MODEL
 OUTPUT SERIES : PREWHITENED VAR20105.DTA

THE CROSS-CORRELATIONS

LAGS 0- 7	-.044	.122	-.165	.077	-.126	-.161	-.275	-.055
STANDARD ERROR	(.147)	(.149)	(.151)	(.152)	(.154)	(.156)	(.158)	(.160)
LAGS 8- 15	.232	-.050	.087	-.190	-.018	-.016	-.013	-.069
STANDARD ERROR	(.162)	(.164)	(.167)	(.169)	(.171)	(.174)	(.177)	(.180)
LAGS 16- 23	.102	.107	-.047	.004	-.135	.157	.080	.050
STANDARD ERROR	(.183)	(.186)	(.189)	(.192)	(.196)	(.200)	(.204)	(.209)
LAG 24	-.245							
STANDARD ERROR	(.213)							

----- section 35 -----

Fonte: Elaborado pelo pesquisador