

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO**

**ESTIMAÇÃO DO PREÇO DO ABATE BOVINO NO
RIO GRANDE DO SUL RECEBIDO PELO PRODUTOR
POR MEIO DE PREVISÕES**

DISSERTAÇÃO DE MESTRADO

Analice Marchezan

**Santa Maria, RS, Brasil
2009**

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**ESTIMAÇÃO DO PREÇO DO ABATE BOVINO NO RIO
GRANDE DO SUL RECEBIDO PELO PRODUTOR POR MEIO
DE PREVISÕES**

por

Analice Marchezan

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Engenharia de Produção, Área de concentração em Gerência da Produção, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Engenharia de Produção.**

Orientador: Prof. Dr. Adriano Mendonça Souza

**Santa Maria, RS, Brasil
2009**

**Universidade Federal de Santa Maria
Centro de Tecnologia
Programa de Pós-Graduação em Engenharia de Produção**

A Comissão Examinadora, abaixo assinada,
aprova a Dissertação de Mestrado

**ESTIMAÇÃO DOPREÇO DO ABATE BOVINO NO RIO GANDE
DO SUL RECEBIDO PELO PRODUTOR POR MEIO DE PREVISÕES**

elaborada por
Analice Marchezan

Como requisito parcial para obtenção do grau de
Mestre em Engenharia de Produção

Comissão Examinadora:

Adriano Mendonça Souza, Dr. (UFSM)
(Presidente/Orientador)

Paulo Sérgio Ceretta, Dr. (UFSM)

José Vanderlei Prestes De Oliveira, Dr. (UFSM)

Santa Maria, 18 de Setembro de 2009.

AGRADECIMENTO

Nesse momento é difícil conseguir agradecer a todas as pessoas que estiveram envolvidas direta e indiretamente nesta conquista:

a família e aos amigos pela compreensão, força e por entenderem meus momentos de ausência;

aos amigos pela compreensão, apoio, companheirismo,...;

ao Adriano que além orientador é um grande amigo;

aos membros da banca examinadora, pelas sugestões que enriqueceram este trabalho;

à Universidade Federal de Santa Maria e ao PPGEP pelo crescimento proporcionado;

à CAPES, pelo apoio financeiro.

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Engenharia de Produção
Universidade Federal de Santa Maria

ESTIMAÇÃO DO PREÇO DO ABATE BOVINO NO RIO GRANDE DO SUL RECEBIDO PELO PRODUTOR POR MEIO DE PREVISÕES

AUTOR: Analice Marchezan
ORIENTADOR: Adriano Mendonça Souza
Data e Local da Defesa: Santa Maria, 18 de Setembro de 2009.

Este trabalho tem por objetivo realizar a combinação de previsões das técnicas ARIMA, ARFIMA e regressão dinâmica, para prever o preço médio mensal recebido pelo produtor pelo abate bovino no Rio Grande do Sul, para o ano de 2009. A pesquisa analisou o comportamento da variável em estudo no período de Janeiro de 2000 a Dezembro de 2008, coletados na Fundação de Economia e Estatística. Optou-se por combinação de previsões, pois estudos desenvolvidos por diversos pesquisadores mostram que as Combinações de Previsões encontraram resultados mais precisos do que as previsões encontradas por um único modelo. O ponto de partida da análise foi encontrar os modelos representativos de cada técnica, através dos critérios penalizadores: O modelo representativo do ARIMA foi $(0,1,1)(1,0,1)_{12}$; do ARFIMA foi $(1,d,0)$, com uma diferença $d=0,492869 < 0,5$ e $|\phi_1| = 0,766281 < 1$, apresentando um ruído fracionário e o modelo de Regressão Dinâmica, que iniciou a análise com 60 variáveis e após a eliminação das variáveis não significativas, reduziu-se a 17. Das previsões resultantes em cada técnica, aplicou-se a Combinação de Previsão através do índice IPACP, o qual é obtido através do somatório de cada previsão multiplicada pela ponderação da variância explicada por cada técnica de previsão. Para verificar se a Combinação de Previsão apresentou resultados mais eficientes que previsões individuais, calculou-se o MAPE, que no modelo ARIMA foi 1,67%; ARFIMA – 3,35%; Regressão Dinâmica – 2,39% e na Combinação previsão foi de 1,35%; e o EQM, que resultou no modelo ARIMA – 0,0021; ARFIMA – 0,0079; Regressão Dinâmica 0,0070 e a Combinação de Previsão 0,0014. Portanto, os valores do o MAPE e do EQM, mostram que a previsões obtidas através da Combinação de Previsões, considerando a ponderação da variância de cada técnica, encontradas através de componentes principais, resultou em previsões que refletem a realidade da série original.

Palavras Chave: ARIMA; ARFIMA; regressão dinâmica; combinação de previsão; preço; abate bovino

ABSTRACT

Master Dissertation
Post Graduation Program in Production Engineering
Federal University of Santa Maria

ESTEEM OF THE PRICE ABATES IN THE RIO GRANDE DO SUL RECEIVED BY THE PRODUCER BY MEANS OF FORECAST

AUTHOR: Analice Marchezan

ADVISOR: Adriano Mendonça Souza

Date and Place of defence: Santa Maria, September 18th 2009.

This main purpose of this research is to carry out the combination of forecasts using ARIMA, ARFIMA and dynamic regression techniques, to foresee received price by the producer for the abates in the Rio Grande do Sul to 2009. The research analyzed the behavior of the variable in study from January, 2000 to December, 2008. The data were collected in the Foundation of Economy and Statistics in Rio Grande do Sul. The starting point was to find a valid and a representative forecast to each technique, using the penalty criteria. The representative model of the ARIMA was $(0,1,1)(1,0,1)_{12}$, the ARFIMA it was $(1, d, 0)$, with a difference $d=0,492869 < 0,5$ and $|\phi_1| = 0,766281 < 1$, presenting a fractionary noise. The Dynamic Regression model, that was initiated with 60 explained variable, was reduced to it 17 significant variables. From the results forecasted in each technique was applied the Forecast Combination through the index, which is gotten through the sum of each forecast multiplied by the variance explained by each forecast technique. To verify if the Forecast Combination is more efficient than individual forecasts, it was calculated the MAPE for individual models, the ARIMA was 1,67%, the ARFIMA was 3,35%; the dynamic regression was 2,39% and in the Combination forecast was 1,35%, the EQM, that resulted from ARIMA model was 0,0021, the ARFIMA was 0,0079, the dynamic regression was 0,0070 and the forecast combination of 0,0014. Therefore, the values of o MAPE and the EQM, show that the forecasts gotten through the Combination, considering the balance using the variance of each technique found through of main components, resulted in forecasts that reflect the reality of the original series.

Keywords: ARIMA; ARFIMA; Dynamic Regression; Forecast Combination; price; abates bovine

LISTA DE FIGURAS

FIGURA 1 - COMPORTAMENTO DAS FUNÇÕES DE AUTOCORRELAÇÃO E AUTOCORRELAÇÃO PARCIAL DE UM MODELO AR (1), SEGUNDO MORETTIN E TOLOI, 2004.	24
FIGURA 2 - COMPORTAMENTO DAS FUNÇÕES DE AUTOCORRELAÇÃO PARCIAL E AUTOCORRELAÇÃO DE UM MODELO MA (1), SEGUNDO MORETTIN E TOLOI, 2004.	27
FIGURA 3 - SÉRIE ORIGINAL SÉRIE PREÇO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO DE JANEIRO DE 2000 A DEZEMBRO DE 2008.....	58
FIGURA 4 - FUNÇÃO DE AUTOCORRELAÇÃO E FUNÇÃO DE AUTOCORRELAÇÃO PARCIAL DA SÉRIE PREÇO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO DE JANEIRO DE 2000 A DEZEMBRO DE 2008	59
FIGURA 5 - SÉRIE ORIGINAL E SÉRIE DIFERENCIADA.....	60
FIGURA 6 - FUNÇÃO DE AUTOCORRELAÇÃO E FUNÇÃO DE AUTOCORRELAÇÃO PARCIAL DA SÉRIE PREÇO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO DE JANEIRO DE 2000 A DEZEMBRO DE 2008	61
FIGURA 7- PREVISÃO PARA O ANO DE 2009 DO PREÇO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO NO RIO GRANDE DO SUL	62
FIGURA 8 - FUNÇÃO AUTOCORRELAÇÃO PARCIAL DOS RESÍDUOS	64
FIGURA 9 - PREVISÃO PARA O ANO DE 2009 DO PREÇO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO NO RIO GRANDE DO SUL	65
FIGURA 10 - PREVISÃO PARA O ANO DE 2009 DO PREÇO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO NO RIO GRANDE DO SUL, ATRAVÉS DE REGRESSÃO DINÂMICA.	69
FIGURA 11 - PROPORÇÃO DA VARIÂNCIA TOTAL EXPLICADA POR CADA AUTOVALOR.....	71
FIGURA 12 – GRÁFICO DAS PREVISÕES E DA COMBINAÇÃO DE PREVISÃO	73

LISTA DE TABELAS

TABELA 1 - MODELOS ARIMA SIGNIFICATIVOS PARA A SÉRIE DO PREÇO DO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO NO RIO GRANDE DO SUL	60
TABELA 2 - PREVISÃO PARA O ANO 2009, DO PREÇO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO NO RIO GRANDE DO SUL	62
TABELA 3 - MODELOS ARFIMA SIGNIFICATIVOS PARA A SÉRIE DO PREÇO DO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO NO RIO GRANDE DO SUL	64
TABELA 4 - PREVISÃO PARA O PREÇO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO NO RIO GRANDE DO SUL, ATRAVÉS DE MODELOS ARFIMA.....	65
TABELA 5 - PREVISÃO PARA O PREÇO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO NO RIO GRANDE DO SUL, ATRAVÉS DE MODELOS DE REGRESSÃO DINÂMICA.....	68
TABELA 6 - TABELA DOS AUTOVALORES E DO PROPORÇÃO DA VARIÂNCIA TOTAL EXPLICADA .	70
TABELA 7 - PREVISÃO PARA O PREÇO RECEBIDO PELO PRODUTOR PELO ABATE BOVINO NO RIO GRANDE DO SUL, ATRAVÉS DE MODELOS DE REGRESSÃO DINÂMICA.....	72

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Tema da Pesquisa.....	12
1.2	Justificativa e Importância da Pesquisa	12
1.3	Objetivos	13
1.3.1	Objetivo Geral	13
1.3.2	Objetivos específicos	13
1.4	Delimitação da Pesquisa	14
2	REFERENCIAL TEÓRICO	15
2.1	Análise de Séries Temporais	15
2.2	Metodologia Box&Jenkins	16
2.2.1	Modelos Estacionários: Auto-regressivo (AR); Médias Móveis (MA), Auto-regressivo de Médias Móveis (ARMA)	19
2.2.2	Modelos Não-Estacionários	29
2.2.3	Modelos Sazonais	30
2.3	Modelos Fracionários – ARFIMA (p,d,q) – Auto-regressivo Fracionário Integrado de Médias Móveis	32
2.4	Identificação, Estimação, Adequação do Modelo, Previsão: ARIMA,SARIMA e ARFIMA	35
2.4.1	Identificação do modelo	36
2.4.2	Estimação dos parâmetros	38
2.4.3	Diagnóstico ou adequação do modelo	39
2.4.4	Previsão	42
2.5	Regressão Dinâmica.....	42
2.6	Combinação de Previsão	47
3	METODOLOGIA	53
3.1	Banco de Dados	53
3.2	Passos Metodológicos.....	53
4	ANÁLISE DOS RESULTADOS	57
4.1	Análise da Série Original	58
4.2	Modelos Auto-regressivos Integrado de Médias Móveis - ARIMA	59
4.3	Modelos Fracionários Auto-regressivo Integrado de Médias Móveis – ARFIMA.....	63
4.4	Regressão Dinâmica.....	66
4.5	Combinação de Previsões	69
5	CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES	74
6	REFERÊNCIAS	76

1 INTRODUÇÃO

A bovinocultura, no Brasil, sempre desempenhou um importante papel no contexto da sociedade e na economia do país desde o período colonial. Essa atividade além do alimento oferece a força, o couro, e além de suas múltiplas utilidades os animais, muitas vezes, servem de transporte para as zonas agrícolas, chegando aos dias de hoje como um importante segmento do agronegócio brasileiro.

O Brasil, considerando a média dos anos 2004 a 2006, tem o maior rebanho bovino comercial e o segundo maior em efetivos do mundo com 205,9 milhões de cabeças, perdendo apenas para a Índia. O Rio Grande do Sul contribui com o sexto maior rebanho brasileiro, o que representa 6,9%, com 14,3 milhões de cabeças. Apesar de a produção gaúcha apresentar um decaimento, a cadeia produtiva é uma das mais tradicionais do Estado, possuindo grande importância econômica e social, poder de integração regional, possibilidade de aumento de valor agregado de seus produtos finais e de melhoria da pauta de exportações. Na produção animal, os bovinos são o segundo maior contribuinte no Valor Bruto da Produção Agropecuária do Estado com 9,31%, conforme o Censo Agropecuário de 2006.

A maior concentração do rebanho está no oeste e sul do Estado, associado à presença dos campos ou integrado com a produção de arroz. As três regiões que apresentam maior percentual do rebanho são: Fronteira Oeste (24,2%), Sul (12,7%) e Campanha (10,6%). Destacam-se os municípios de Alegrete com 625.113 cabeças, Santana do Livramento com 579.413 cabeças, São Gabriel com 415.405 cabeças e Dom Pedrito com 410.534 cabeças, segundo o Atlas Sócio Econômico do Rio Grande do Sul, 2009.

O desenvolvimento social e econômico do Rio Grande do Sul é marcado pela produção bovina desde 1624, quando surgiram as primeiras estâncias baseadas na atividade pecuária. Mas foi fazendo a ligação dos estados do Rio Grande do Sul e São Paulo, que intensificou-se o comércio de carnes e subprodutos, até que, em 1780 é fundada a primeira charqueada, no município de Pelotas, marco da indústria gaúcha de carnes.

Silva e Alves (2009) destacam o Estado como um dos principais produtores brasileiros de bovino de corte e principal berço desta atividade no século XVIII. O Rio Grande do Sul só deixou de ser o maior produtor, quando ocorreu a expansão agrícola para a região Centro-

Oeste. Tellechea (2001) ressalta que embora não seja mais o principal produtor, o Estado destaca-se como o maior banco genético das raças européias, inclusive exportando reprodutores para novos rebanhos no país, além da mão-de-obra qualificada, responsável pela introdução da agropecuária no Centro-Oeste, especialmente no Mato Grosso do Sul.

Apesar do Rio Grande do Sul não ser mais o principal produtor, Montoya & Parré (2000) apontam a bovinocultura como parte integrante da cadeia produtiva do agronegócio, formada pelo setor montante da produção (o produtor rural e o setor da produção), o produtor permanece entre setores de mercado altamente especializados, mas não possui capacidade de formular os preços finais de seu produto.

Devido a isso a pecuária gaúcha enfrenta inúmeras dificuldades, sejam elas produtivas ou econômicas. Essas dificuldades abrangem a expansão de novos mercados e suas exigências sanitárias, a competição com outras carnes e as dos países do Mercosul, necessidade da produção de animais padronizados e jovens, além de fatores naturais e a concorrência com o mercado interno central do país. Outro fator com forte relevância é que nos últimos anos os preços do quilograma do boi mantiveram-se constante, em contrapartida, os insumos apresentaram elevação, ocasionando a diminuição de renda dos pecuaristas e em muitos casos à necessidade de muitos fazendeiros desfazerem-se de suas propriedades ou, até mesmo, ao abandono do ramo pecuário.

Apesar das dificuldades, é notória a importância da bovinocultura para o Estado em todos os aspectos, sejam eles para a economia, alimentação, trabalho, transporte, reprodução entre tantas outras atividades. Mas precisa-se ter consciência de que esta atividade necessita de organização dos produtores, quanto à alimentação desses animais, pastagem, plantação, mão-de-obra, medicamentos, da heterogeneidade produtiva da atividade, pois essa é gerada pela existência de pecuaristas que exercem a atividade por herança cultural e de forma extensiva, mas também por aqueles que buscam constantemente o emprego de novas tecnologias e que fazem dessa atividade um ramo empresarial.

Nesse sentido, pela impossibilidade de interferir no valor final do seu produto, os pecuaristas precisam de organização, ou seja, prever quais serão seus investimentos. E a previsão do preço recebido pelo produtor, possibilita-o analisar a viabilidade dos investimentos e esforços, pois somente com os resultados financeiros refletem o interesse pela produção. Para dar este suporte pode-se utilizar técnicas de previsão, pois conforme Morettin (1989), a previsão não constitui um fim em si, mas um meio de fornecer informações e subsídios para uma conseqüente tomada de decisão, visando a atingir determinados objetivos.

Portanto, essas previsões podem servir de suporte para a organização dos produtores, cooperativas, empresas e do próprio governo e que, ao detectarem as necessidades do Estado, poderá desenvolver os incentivos necessários para sanar problemas enfrentados pelos produtores. Além de poder auxiliar as organizações privadas que necessitam direcionar seus esforços e recursos, num futuro a curto e a longo prazo.

1.1 Tema da pesquisa

O tema da pesquisa aborda previsão por técnicas univariadas, regressão dinâmica e por combinação de previsão. Serão desenvolvidas previsões do preço médio mensal recebido pelo produtor pelo abate bovino por meio de modelos individuais e a combinação dessas previsões através de componentes principais.

1.2 Justificativa e importância da pesquisa

A combinação de previsão é uma técnica importante para descrever o comportamento da série devido à acurácia oferecida, pois a técnica é resultante da combinação de outras previsões. A importância é, que quando se utiliza uma única técnica apesar de encontrar-se vários modelos representativos, é necessário optar-se por um único modelo, seguindo os critérios penalizadores AIC e BIC, cientes de que ao dispensar os demais modelos perdem-se informações. Com o intuito de utilizar o máximo de informações oferecidas sobre a série é que se utilizarão combinações de previsões. Embora existam inúmeros trabalhos relacionados ao assunto, e os autores não chegaram em consenso sobre qual é a combinação que reflete a melhor previsão, é que propõe-se nesta pesquisa, combinar três modelos previsão, escolhidos pelo próprio pesquisador, por meio da Análise de Componentes Principais.

1.3 Objetivos

Neste item descrevem-se o objetivo geral e objetivos principais que nortearão a a estimação do preço do abate bovino no Rio Grande do Sul recebido pelo produtor por meio de previsões.

1.3.1 Objetivo Geral

Combinar as previsões encontradas através das técnicas ARIMA, ARFIMA e regressão dinâmica, para estimar o preço médio mensal recebido pelo produtor pelo abate bovino no Rio Grande do Sul, para o ano de 2009.

1.3.2 Objetivos específicos

- analisar o comportamento do preço médio mensal recebido pelo produtor pelo abate bovino no Rio Grande do Sul;

- prever o preço médio mensal recebido pelo produtor pelo abate bovino no Rio Grande do Sul, para o ano de 2009, através das técnicas ARIMA, ARFIMA e regressão dinâmica;

- encontrar as ponderações das previsões, obtidas pela proporção da variância total explicada pela técnica de previsão, através de componentes principais;

- criar um índice geral de combinação de previsão, composto pelas previsões de cada técnica e suas respectivas ponderações obtidas através dos autovalores;

- prever o preço médio mensal recebido pelo produtor pelo abate bovino no Rio Grande do Sul para o ano de 2009, através do índice geral de combinação de previsão.

1.4 Delimitação da Pesquisa

A pesquisa analisará o comportamento do preço por quilograma recebido pelo produtor pelo abate de bovinos no Rio Grande do Sul, no período de Janeiro de 2000 a Dezembro de 2008, e realizará previsões para o ano de 2009 por meio da combinação das previsões obtidas pelas metodologias ARIMA, ARFIMA e Regressão Dinâmica.

Optou-se por combinação de previsões pelos estudos desenvolvidos por diversos pesquisadores, onde mostram que as Combinações de Previsões encontram previsões mais precisas do que as previsões encontradas por um único modelo.

2 REFERENCIAL TEÓRICO

Este capítulo abrangerá Séries Temporais, a Metodologia de Box & Jenkins para modelos ARIMA (divididos em Modelos AR, MA, ARMA, ARIMA e SARIMA) e ARFIMA, Modelos de Regressão Dinâmica e Combinação de Previsões, que serão utilizados para estimar o preço médio mensal recebido pelo produtor pelo abate bovino no Estado do Rio Grande do Sul, para o ano de 2009, através da previsão encontrada em cada técnica.

2.1 Análise de Séries Temporais

Uma série temporal é a coleção das observações feitas sequencialmente ao longo do tempo. As características mais importantes dessas observações vizinhas é sua dependência, a qual se tem por interesse analisar e modelar, considerando a ordem dos dados um fator crucial. Vale observar, também, que o tempo pode ser substituído por outra variável, como espaço e profundidade. A técnica de análise de Séries Temporais vem sendo utilizada com frequência em várias áreas de pesquisa e, como a maior parte de procedimentos estatísticos foi desenvolvida para analisar observações independentes, o seu estudo requer o uso de técnicas específicas, pois os dados surgem dos mais diversos campos de conhecimento, como Ciências Sociais e Políticas, Sociologia, História, Meio Ambiente, Economia (preços diários de ações, taxas de desemprego, produção industrial, etc.), Medicina (eletrocardiograma, eletroencefalograma, etc.), Epidemiologia (números mensais de meningite, etc.), Meteorologia (precipitação pluviométrica, temperatura diária, velocidade do vento), entre outros.

Para Barros (2006) a Série Temporal representa um conjunto de observações ordenadas no tempo, que exhibe dependência serial e que pode ser modelado por um processo estocástico, que é a distribuição das probabilidades conjuntas das variáveis aleatórias. A

variável é observada em pontos discretos de tempo, em geral, espaçada igualmente. A análise e a modelagem de Séries Temporais buscam descrever o processo que gera a seqüência da variável determinada, bem como apresentar as suas características mais importantes. Para estimar uma Série Temporal, é necessário representar o comportamento do processo, por meio de um modelo matemático que possa ser estendido ao futuro. Uma vez que o modelo de Séries Temporais é estabelecido, a técnica apropriada de previsão pode ser desenvolvida. A Série Temporal, para ser estimada, precisa estar livre de tendências e componentes sazonais, para assim conseguir capturar a estrutura de dependência da Série observada, a partir da combinação de ruídos brancos. Logo, se a modelagem da Série Temporal estiver bem ajustada, não deve haver correlação serial entre os resíduos, e o comportamento da Série deve ser explicado usando o menor número possível de parâmetros – princípio da parcimônia.

2.2 Metodologia Box & Jenkins

Durante a década de 60, os professores George E. P. Box e Gwilym M. Jenkins escreveram diversos trabalhos sobre a teoria de controle e análise de Séries Temporais. Em 1970, publicaram o livro *Time Series Analysis, forecasting and control*, apresentando uma metodologia para a análise de Séries Temporais e, em 1976, foi lançada a versão revisada desse livro, que normalmente é a mais mencionada. Para Makridakis e Winkler (2008), o grande mérito desse trabalho foi reunir as técnicas existentes em uma metodologia, a fim de construir modelos que descrevessem, com precisão e de forma parcimoniosa, o processo gerador da Série Temporal. Isso proporcionou previsões acuradas de valores futuros. A previsão é uma das principais razões da popularidade desta metodologia e, em muitos casos, as previsões, principalmente de curto prazo, são melhores que as obtidas com base nos modelos econométricos tradicionais.

Esta metodologia é composta pelos Modelos ARIMA, que são modelos estatísticos lineares para análise de Séries Temporais. Esses modelos são mais sofisticados por usarem a correlação entre as observações. No modelo ARIMA, a Série Temporal é gerada por um processo estocástico, cuja natureza pode ser representada através de um modelo. A notação

utilizada para designar o modelo ARIMA é $ARIMA(p, d, q)$, em que p é o número de termos auto-regressivos; d , o número de diferenciações para que a Série torne-se estacionária e q , o número de termos de médias móveis. Os termos p , d e q são todos inteiros maiores ou iguais a zero. São casos particulares: o modelo $ARMA(p, q)$, o modelo auto-regressivo $AR(p)$ e o modelo de médias móveis $MA(q)$, sendo os três modelos utilizados para Séries Temporais estacionárias ($d = 0$).

Uma Série dispõe de operadores que facilitam a manipulação dos modelos. Observe-se, a seguir:

Operador translação para o passado, denotado por B e definido por:

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

Operador translação para o Futuro, denotado por F e definido por:

$$FZ_t = Z_{t+1} \rightarrow F^m Z_t = Z_{t+m} \quad (2)$$

Operador diferença:

$$\Delta Z_t = Z_t - Z_{t-1} \rightarrow \Delta Z_t = (1 - B)Z_t \quad (3)$$

$$\therefore \Delta = 1 - B$$

Operador soma, denotado por S e definido por:

$$SZ_t = \sum_{j=0}^{\infty} Z_{t-j} = Z_t + Z_{t-1} + \dots + (1 + B + B^2 + \dots)Z_t \quad (4)$$

$$SZ_t = (1 - B)^{-1} Z_t = \Delta^{-1} Z_t \quad (5)$$

$$\therefore S = \Delta^{-1} \quad (6)$$

Os modelos e equações apresentados na Metodologia Box & Jenkins é baseada em Morettin e Toloi (2004), Souza & Camargo (1996), Barros (2006) e Vasconcelos e Alves (2000). Todo o capítulo está relacionado aos estudos desses autores.

Os modelos de Séries Temporais possibilitam descrever um processo estocástico, utilizando apenas valores passados da variável dependente e do termo de erro. Assim, dada uma Série, para cada t tem-se uma variável aleatória (v. a.) Z_t , que possui uma média $\mu(t)$ em função Esperança Matemática da Variável Aleatória ($E[Z(t)]$) e uma variância $V(t)$. Por isso, tem-se os parâmetros importantes a serem considerados em um processo estocástico: média e função de autocovariância.

A média é igual ($\mu(t)$):

$$\mu(t) = E[Z(t)] \quad (7)$$

Para a Função de Autocovariância:

$$\gamma(t_1, t_2) = E[Z(t_1)Z(t_2)] - E[Z(t_1)]E[Z(t_2)] \quad (8)$$

No caso de $t_1 = t_2 = t$, a Função de Autocovariância é igual à variância:

$$\gamma(t, t) = V(t) = E[Z^2(t)] - E^2[Z(t)] \quad (9)$$

Portanto, a Função de Autocovariância pode ser representada desta forma pela covariância:

$$\gamma(t_1, t_2) = Cov[Z(t_1)Z(t_2)] \quad (10)$$

É importante o comentário de que a Função de Autocovariância desempenha um papel muito importante na análise dos modelos ARIMA. Esses modelos poderiam simplesmente ser escritos da seguinte forma:

$$Z_t = f(t) + a_t \quad (11)$$

Onde $f(t)$ é chamada de sinal e a_t de ruído.

É partindo dessa função, de acordo com as hipóteses sobre $f(t)$ e a_t , que se chega a vários modelos, sendo um deles os modelos ARIMA.

A importância dos modelos ARIMA encontra-se neste propósito, descrito através de três classes de processos:

Processos lineares estacionários são representados da seguinte forma:

$$Z_t - \mu = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \sum_{k=0}^{\infty} \psi_k a_{t-k} \quad (12)$$

Em que a_t é ruído branco $\mu = E(Z_t)$, ψ_1, ψ_2, \dots é uma seqüência de parâmetros tal que:

$$\sum_{k=0}^{\infty} \psi_k^2 < \infty \quad (13)$$

Existem três casos particulares da Eq.12, que serão descritos a seguir:

- Processo auto-regressivo de ordem p : $AR(p)$;
- Processo de médias móveis de ordem q : $MA(q)$;
- Processo auto-regressivo e de médias móveis de ordem p e q : $ARMA(p, q)$

Processos lineares não-estacionários homogêneos: constituem uma generalidade dos *processos lineares estacionários*, que supõem que o mecanismo gerador da Série sejam não-estacionárias, em nível ou em inclinação. Essas Séries podem tornar-se estacionárias por um meio finito de diferenças (geralmente uma ou duas diferenças são suficientes para estacionarizar a Série).

Processo de memória longa: é um processo estacionário que possui uma longa função de autocorrelação, com decaimento muito longo (hiperbólico) e cuja análise necessitará de uma diferença fracionária ($0 < d < 0,5$).

Esses modelos são bem descritos pelos modelos $ARIMA(p, d, q)$, que aceitam a inclusão de um operador sazonal. Portanto, esta metodologia caracteriza-se por Modelos Estacionários ($AR(p), MA(q), ARMA(p, q)$) e modelos Não-Estacionários ($ARIMA(p, d, q)$)

2.2.1 Modelos Estacionários: Auto-regressivo (AR); Médias Móveis (MA), Auto-regressivo de Médias Móveis (ARMA)

Os Modelos Estacionários interpretam a Série Temporal como uma realização de um vetor aleatório do tamanho das Séries disponíveis, seguindo o princípio da parcimônia e da construção de modelos, através de um ciclo iterativo. Os modelos Box & Jenkins determinam

o processo estocástico adequado para representar uma dada Série Temporal, por meio da passagem de um ruído branco por um filtro linear.

Modelos derivados do filtro linear são capazes de representar estacionariedade e não-estacionariedade de uma Série Temporal. Se uma Série Temporal é estacionária, sua média, variância e autocorrelação (a defasagens diversas) permanecem as mesmas ao longo do tempo, ou seja, constantes.

Formalmente,

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t \quad (14)$$

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots \quad (15)$$

O modelo (15) é denominado Função de Transferência do filtro, e μ é o parâmetro que determina o nível da série.

A Eq.14 é um processo linear. Dessa forma:

$$\begin{aligned} E(a_t) &= 0, \forall t; \\ \text{Var}(a_t) &= \sigma_a^2, \forall t; \\ E(a_t a_s) &= 0, s \neq t \end{aligned}$$

Chamando de $\tilde{Z}_t = Z_t - \mu$ a série centrada na média μ , tem-se:

$$\tilde{Z}_t = \psi(B)a_t \quad (16)$$

Se a seqüência de pesos $\{\psi_i, i \geq 1\}$ for convergente, o filtro é estável e Z_t é estacionária. Nesse caso, μ é a média do processo. Caso contrário, Z_t é não-estacionária e μ não tem significado específico, a não ser como ponto de referência da Série.

Da equação (14), tem-se que:

$$E(Z_t) = \mu + E\left(a_t + \sum_{i=1}^{\infty} \psi_i a_{t-i}\right) \quad (17)$$

E como $E(a_t) = 0, \forall t$, temos que $E(Z_t) = \mu$ se a série $\sum_{i=1}^{\infty} \psi_i$ convergir. Quanto à Função de Autocovariância γ_i de Z_t , ela é dada por:

$$\gamma_i = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+i} \quad (18)$$

Com $\psi_0 = 1$. Um caso particular é para $i = 0$, onde a γ_i é igual à variância de Z_t :

$$\gamma_0 = \text{Var}(Z_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2 \quad (19)$$

A condição de existência das equações (18) e (19) é que $\sum_{i=0}^{\infty} \psi_i^2 < \infty$.

Por conseguinte, percebe-se que a média e a variância de Z_t apresentam-se constantes, e γ_i só depende dos valores de i , sendo, portanto, estacionária.

Pode-se escrever \tilde{Z}_t em uma forma alternativa, como uma soma ponderada dos valores passados e mais um ruído:

$$\tilde{Z}_t = \pi_1 \tilde{Z}_{t-1} + \pi_2 \tilde{Z}_{t-2} + \dots + a_t = \sum_{i=1}^{\infty} \pi_i \tilde{Z}_{t-i} + a_t \quad (20)$$

$$\left(1 - \sum_{i=1}^{\infty} \pi_i B^i\right) \tilde{Z}_t = a_t \quad (21)$$

Ou, simplesmente, é possível representar um modelo estacionário da seguinte forma:

$$\pi(B) \tilde{Z}_t = a_t \quad (22)$$

Em que $\pi(B)$ é o operador do modelo:

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots \quad (23)$$

Unindo as equações (22) e (16), resulta-se na seguinte equação:

$$\pi(B) \psi(B) a_t = a_t \quad (24)$$

Ou, também, representá-la da seguinte maneira:

$$\pi(B) = \psi^{-1}(B) \quad (25)$$

Essa equação é utilizada para obter os valores de π_i , em função ψ_i e vice-versa.

Com base nas particularidades dos processos estacionários estabelecidos acima, serão descritos, abaixo, os modelos que compõem esse processo: $(AR(p), MA(q), ARMA(p, q))$.

2.2.1.1 Modelos Auto-Regressivos de Ordem p – AR (p)

Esse modelo corresponde à defasagem da série Z_t . A versão mais simples de um modelo AR é o de ordem um, sendo aquele em que Z_t depende somente de Z_{t-1} e de a_t , no instante t . O modelo que contém $\phi(B)=1$ é chamado modelo auto-regressivo, sendo denotado por $AR(p)$, em que p indica a ordem do modelo, isto é, o número de defasagens.

Na equação (20), os parâmetros são representados por π_i . Se, nessa mesma equação, o parâmetro assumir o valor de $\pi_i = 0, i > p$, em que p é a ordem do modelo, a equação se reduzirá a um modelo auto-regressivo de ordem p , denotado por $AR(p)$. Então, é possível reescrever a equação auto-regressiva de acordo com a equação usual, e renomear os pesos π_i para ϕ_j . Os parâmetros desse modelo são representados por ϕ :

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t \quad (26)$$

O operador auto-regressivo estacionário de ordem p :

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p. \quad (27)$$

Podendo, então, ser reescrito da seguinte forma:

$$\phi(B) \tilde{Z}_t = a_t. \quad (28)$$

Nas equações (29) e (30), encontram-se os modelos Auto-Regressivos de ordem 1 e 2 respectivamente:

$$AR(1) \quad \tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + a_t \quad (29)$$

$$AR(2) \quad \tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + a_t \quad (30)$$

A identificação do modelo adequado começa pelos valores dos parâmetros, ou seja, o modelo será considerado estacionário se $|\phi_1| < 1$, satisfazendo a condição da estacionaridade. A condição de estacionaridade do $AR(p)$ estabelece que todas as raízes da equação devem cair fora do círculo unitário.

Em contrapartida, o modelo que apresenta ordem finita sempre será inversível. Como $\pi(B) = \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ é finito, não existem restrições sobre os parâmetros para assegurarem a invertibilidade.

Além de $|\phi_1| < 1$, é necessário recorrer às Funções de Autocorrelação. Retornando-se para a equação (26), multiplicam-se todos os membros desta por Z_{t-i} , e tomando-se a esperança, tem-se:

$$E(\tilde{Z}_t \tilde{Z}_{t-i}) = \phi_1 E(\tilde{Z}_{t-1} \tilde{Z}_{t-i}) + \phi_2 E(\tilde{Z}_{t-2} \tilde{Z}_{t-i}) + \dots + \phi_p E(\tilde{Z}_{t-p} \tilde{Z}_{t-i}) + a_t E(a_t \tilde{Z}_{t-i}) \quad (31)$$

Como \tilde{Z}_{t-i} só envolve ruídos até a_{t-i} , não-correlacionados com a_t e $E(a_t \tilde{Z}_{t-i}) = 0, i > 0$, resulta que:

$$\gamma_i = \phi_1 \gamma_{i-1} + \phi_2 \gamma_{i-2} + \dots + \phi_p \gamma_{i-p}, i > 0 \quad (32)$$

Para $i = 0$ em (32):

$$\frac{\gamma_0}{\gamma_0} = \frac{\phi_1 \gamma_{-1} + \phi_2 \gamma_{-2} + \dots + \phi_k \gamma_{-k} + \sigma_a^2}{\gamma_0} \quad (33)$$

como $\gamma_{-i} = \gamma_i$, tem-se que $\gamma_0 = \text{Var}(\tilde{Z}_t) = \text{Var}(Z_t)$:

$$1 = \phi_1 \rho_1 + \phi_2 \rho_2 + \dots + \phi_i \rho_i + \frac{\sigma_a^2}{\gamma_0} \quad (34)$$

Ou seja:

$$\text{Var}(Z_t) = \sigma_z^2 = \frac{\sigma_a^2}{1 - \phi_1 \rho_1 - \dots - \phi_i \rho_i} \quad (35)$$

Logo, dividindo (27) por $\gamma_0 = \text{Var}(\tilde{Z}_t) = \text{Var}(Z_t)$, obtem-se:

$$\rho_i = \phi_1 \rho_{i-1} + \phi_2 \rho_{i-2} + \dots + \phi_p \rho_{i-p}, i > 0 \quad (36)$$

Que também pode ser escrita como:

$$\phi(B)\rho_i = 0 \quad (37)$$

Em que o operador defasagem passa a agir agora em i :

$$B\rho_i = B\rho_{i-1} \dots \quad (38)$$

As Funções de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP) são muito importantes, pois a análise visual permite identificar a ordem do modelo:

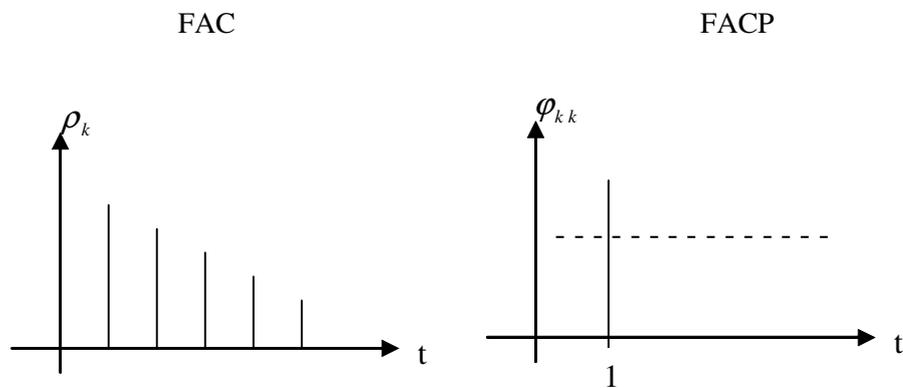


Figura 1 - Comportamento das funções de autocorrelação e autocorrelação parcial de um modelo AR (1), segundo Morettin e Toloí, 2004.

A Função de Autocorrelação de um processo auto-regressivo é constituída de uma mistura de polinômios, exponenciais e senóides amortecidas, e a Função de Autocorrelação Parcial (FACP) apresenta um corte rápido no *lag* significativo, indicando a ordem p do modelo.

Para Morettin e Toloí (2004), a função de autocorrelação decai exponencialmente quando ϕ_1 é positivo; quando ϕ_1 é negativo, a função de autocorrelação também decai exponencialmente, mas apresenta alternância de sinais positivos e negativos.

2.2.1.2 Modelo de Médias Móveis de ordem q – MA (q)

O modelo de Médias Móveis é a combinação do ruído branco a_t , do período atual com os ocorridos anteriormente. A notação do modelo é $MA(q)$, na qual q indica a ordem do modelo. Voltando à equação (14), suponha-se que $\psi_j = 0, j > q$, passa-se, então, a utilizar a notação:

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}; \quad (39)$$

Chamando $\tilde{Z}_t = Z_t - \mu$, resulta que:

$$\tilde{Z}_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t = \theta(B) a_t \quad (40)$$

O operador de médias móveis estacionário de ordem q :

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q. \quad (41)$$

É possível observar como são descritos os modelos de ordem:

$MA(1)$

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} \quad (42)$$

$MA(2)$

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \quad (43)$$

O próximo passo será analisar as condições de estacionaridade e invertibilidade, que precisam ser satisfeitas. Dado que $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$, não existem restrições sobre os parâmetros θ_j , para que o processo seja estacionário. Para isso, usaram-se argumentos completamente similares aos modelos $AR(p)$, a fim de verificar a estacionariedade, as condições de invertibilidade para um modelo $MA(q)$, observando que as raízes da equação característica $\theta(B) = 0$ estejam fora do círculo unitário. Esse modelo é igual a um modelo $AR(p)$, em que p é infinito.

A próxima análise a ser feita no modelo é a da Função de Autocorrelação.

Primeiramente, a Função de Autocovariância desse modelo é:

$$\gamma_i = E[\tilde{Z}_t \tilde{Z}_{t-i}] E \left[\left(a_t - \sum_{k=1}^q \theta_k a_{t-k} \right) \left(a_{t-i} - \sum_{l=1}^q \theta_l a_{t-i-l} \right) \right] \quad (44a)$$

$$= E(a_t a_{t-i}) - \sum_{k=1}^q \theta_k E(a_{t-i} a_{t-k}) - \sum_{l=1}^q \theta_l E(a_t a_{t-i-l}) + \sum_{k=1}^q \sum_{l=1}^q \theta_k \theta_l E(a_{t-k} a_{t-i-l}) \quad (44b)$$

Não esquecendo que:

$$\gamma_a(j) = E(a_t a_{t-i}) = \begin{cases} \sigma_a^2, & i = 0 \\ 0, & i \neq 0 \end{cases} \quad (45a)$$

$$\gamma_0 = \text{Var}(Z_t) = \sigma_Z^2 = (1 + \theta_1^2 + \dots + \theta_q^2) = \sigma_a^2 \quad (45b)$$

Em termos de $\gamma_a(i)$, a função de autocovariância de Z_t fica:

$$\gamma_i = \left(-\theta_i \sum_{l=1}^{q-i} \theta_l \theta_{i+l} \right) \sigma_a^2 \quad (46a)$$

$$\gamma_i = (-\theta_i + \theta_1 \theta_{i+1} + \theta_2 \theta_{i+2} + \dots + \theta_q \theta_{i-q}) \sigma_a^2, i = 1, \dots, q \quad (46b)$$

$$\gamma_i = 0, i > q$$

Unindo as equações (45) e (46), obtém-se a função de autocorrelação do processo:

$$\rho_i = \begin{cases} \frac{-\theta_i + \theta_1 \theta_{i+1} + \theta_2 \theta_{i+2} + \dots + \theta_q \theta_{i-q}}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & i = 1, \dots, q \\ 0, & i > q \end{cases} \quad (47)$$

Pode-se verificar que a função de autocorrelação de um processo $MA(q)$ é igual a zero para *lag's* maiores que q , ao contrário do que acontece em um processo $AR(p)$.

Dessa forma, identificando a condição de estacionaridade, o processo apresenta ordem finita, portanto, será sempre estacionário.

Quanto à condição de invertibilidade, é imposta a restrição de que o parâmetro em módulo é menor que um, ou seja, $|\theta_1| < 1$.

A identificação do modelo é feita através das autocorrelações parciais (, sendo que a função de autocorrelação fornece a ordem do modelo:

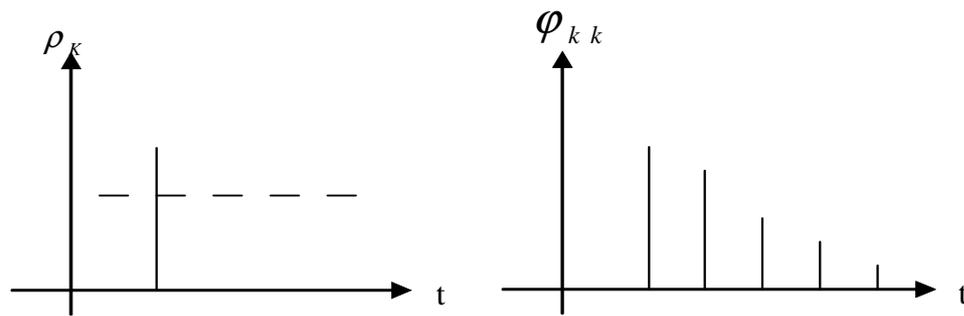


Figura 2 - Comportamento das funções de autocorrelação parcial e autocorrelação de um modelo MA (1), segundo Morettin e Toloí, 2004.

Para Souza (2006), a autocorrelação apresentará um corte rápido no *lag* significativo, indicando a ordem do modelo, enquanto que a autocorrelação parcial decairá exponencialmente, se todos os parâmetros forem positivos, conforme Figura 2. Caso contrário, formará uma senóide amortecida.

2.2.1.3 Modelo Auto-Regressivo de Médias Móveis – ARMA (p, q)

Esse modelo é a união dos modelos vistos anteriormente, os auto-regressivos e o de médias móveis. Ele é utilizado em Séries encontradas na prática, para as quais não se deseja números muito grandes de parâmetros.

Um modelo $ARMA(p, q)$ apresenta-se da seguinte forma:

$$\tilde{Z}_t = \underbrace{\phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p}}_{AR(p)} + a_t - \underbrace{\theta_1 a_{t-1} - \dots - \theta_q a_{t-q}}_{MA(q)} \quad (48)$$

Se $\phi(B)$ e $\theta(B)$ são os operadores auto-regressivos e de médias móveis, respectivamente, pode-se reescrever um modelo $ARMA(p, q)$ através dos seus operadores:

$$\phi(B)\tilde{Z}_t = \theta(B)a_t \quad (49)$$

Um modelo frequentemente usado é o $ARMA(1,1)$, no qual a ordem de ambos os parâmetros é igual a um ($p = q = 1$). Tem-se, também, que $\phi(B) = 1 - \phi B$ e $\theta(B) = 1 - \theta B$, ou seja, reduz-se a equação (48) em:

$$\tilde{Z}_t = \phi \tilde{Z}_{t-1} + a_t - \theta a_{t-1} \quad (50)$$

Facilmente, pode-se escrever (50) na forma de um processo linear ou de médias móveis infinito, apenas substituindo $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots$:

$$\tilde{Z}_t = \psi(B)a_t \quad (51)$$

Em que $\psi_i = \phi^{i-1}(\phi - \theta), i \geq 1$, sendo o processo:

Estacionário: $\sum \psi_i = (\phi - \theta) \sum \phi^{i-1} < \infty$, ou seja, $|\phi| < 1$.

Também podendo ser escrito da seguinte forma:

$$\pi(B)\tilde{Z}_t = a_t \quad (52)$$

Em que os pesos $\pi_i = \theta^{i-1}(\phi - \theta), i \geq 1$, de modo que o processo é irreversível se $\sum \pi_i < \infty$, ou seja, $|\phi| < 1$.

Em seguida, percebe-se que as condições de estacionaridade de um processo $ARMA(1,1)$ são as mesmas que as de um processo $AR(1)$, e as condições de invertibilidade são as mesmas que as de um processo $MA(1)$.

Essas conclusões se generalizam para qualquer ordem de um modelo $ARMA(p, q)$.

Dessa forma, pode-se concluir que o processo é inversível, se todas as raízes de $\phi(B) = 0$ e $\theta(B) = 0$ caírem fora do círculo unitário.

Quanto à Função de Autocorrelação, sua ordem vem da multiplicação da equação (47) por \tilde{Z}_{t-i} e tomando a esperança. Logo:

$$\gamma_i = E\left[\left(\phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}\right) \tilde{Z}_{t-i}\right] \quad (53)$$

$$\gamma_i = \phi_1 \gamma_{t-1} + \dots + \phi_p \gamma_{t-p} + \gamma_{za}(i) - \theta_1 \gamma_{za}(i-1) - \dots - \theta_q \gamma_{za}(i-q) \quad (54)$$

Em que $\gamma_{za}(i)$ é a covariância cruzada entre Z_t e a_t , definida por:

$$\gamma_{za}(i) = E(a_t \tilde{Z}_{t-i}) \quad (55)$$

Como \tilde{Z}_{t-i} só depende dos choques até a_t , ocorridos até o instante $t-i$, obtem-se:

$$\gamma_{za}(i) = \begin{cases} = 0, & i > 0 \\ \neq 0, & i \leq 0 \end{cases} \quad (56)$$

De modo que (54) é reescrita:

$$\gamma_i = \phi_1 \gamma_{i-1} + \dots + \phi_p \gamma_{i-p}, i > q \quad (57)$$

E a Função de Autocorrelação é obtida através de (57):

$$\rho_i = \phi_1 \rho_{i-1} + \dots + \phi_p \rho_{i-p}, i > q \quad (58)$$

Observa-se que os *lag's*, $1, 2, \dots, q$ das autocorrelações são afetados diretamente pelos parâmetros das médias móveis, mas, para $i > q$, as funções de autocorrelação assumem o comportamento de um modelo $AR(p)$.

A FAC possui importante papel na identificação do modelo em estudo. A utilização da Função de Autocorrelação Parcial (FACP), facilita o procedimento de identificação do modelo, o qual se denotará ϕ_{kj} de um modelo $AR(k)$, de tal modo que ϕ_{kk} seja o último coeficiente:

$$\rho_i = \phi_{k1} \rho_{i-1} + \dots + \phi_{kk} \rho_{i-k}, i = 1, \dots, k \quad (59)$$

2.2.2 Modelos Não-Estacionários

Séries não-estacionárias são aquelas que não se desenvolvem ao longo do tempo ao redor de uma média constante. Mas, essas Séries podem tornar-se estacionárias, ao recorrer-se ao processo de diferenciação.

Seja Z_t uma Série não-estacionária:

$$W_t = Z_t - Z_{t-1} = (1 - B)Z_t = \Delta Z_t \quad (60)$$

Com um número finito de diferenças, d , transformou-se Z_t em estacionária. Essas Séries também são chamadas de estacionárias homogêneas ou, ainda, de portadoras de raízes unitárias.

2.2.2.1 Modelo Auto-Regressivo Integrado de Médias Móveis – *ARIMA* (p, d, q)

Dada Z_t , não-estacionária, se $W_t = \Delta^d Z_t$ for estacionária, pode-se representar W_t por um modelo ARMA (p, q), isto é:

$$\phi(B)W_t = \theta(B)a_t. \quad (61)$$

Sendo W_t uma diferença de Z_t , pode-se dizer que Z_t segue um modelo auto-regressivo, integrado de médias móveis, de ordem (p, d, q) , ou seja, *ARIMA*(p, d, q), em que p e q , são as respectivas ordens de $\phi(B)$ e $\theta(B)$.

$$\phi(B)\Delta^d Z_t = \theta(B)a_t. \quad (62)$$

No modelo (61), estão fora do círculo unitário. Portanto, escrever (62) equivale a escrever:

$$\omega(B)Z_t = \theta(B)a_t. \quad (63)$$

Em que $\omega(B)$ é um operador auto-regressivo não-estacionário de ordem $p + q$, com d raízes iguais a um (sobre o círculo unitário), e as restantes p fora desse círculo:

$$\omega(B) = \phi(B)\Delta^d = \phi(B)(1 - B)^d. \quad (64)$$

Observa-se que é indiferente escrever $\omega(B)Z_t$ ou $\omega(B)\tilde{Z}_t$, pois $\Delta^d Z_t = \Delta^d \tilde{Z}_t$, para $d > 1$.

2.2.3 Modelos Sazonais

Esses modelos são uma adição dos modelos *ARIMA*, ou seja, é um modelo *ARIMA* mais a Sazonalidade, chamado de *SARIMA*, podendo ser escrito como: *ARIMA*(p, d, q)(P, D, Q) ou *SARIMA*(p, d, q)(P, D, Q), no qual as letras minúsculas

representam a ordem do modelo e diferenciação ARIMA; e as maiúsculas representam a ordem dos modelos e diferenciação Sazonal.

Para Farias, Rocha; Lima (2000), as Séries Temporais apresentam comumente padrões de comportamento periódico, ou seja, repetem-se a cada s períodos de tempo ($s > 1$). Um dos casos mais comuns de dados periódicos é a série sazonal, sendo aquela em que existe o período de repetição. As Séries Temporais sazonais, em geral, exibem intervalos de tempo de um mês e período anuais ($s = 12$); no entanto, é possível ocorrer casos em que o período é trimestral ($s = 4$). O tratamento das Séries Temporais sazonais é normalmente feito usando-se o componente sazonal como fator de ajustamento. Todavia, alguns dos métodos de ajustamento sazonal não são adequados para fazer previsão com modelos de Séries Temporais, visto que o ajustamento sazonal provoca a perda de informações cruciais para o processo de previsão com os modelos de Séries Temporais.

Para esses autores, os modelos *ARIMA* exploram a autocorrelação entre os valores da Série em instantes sucessivos; mas, quando os dados são observados em períodos inferiores a um ano, a Série também pode apresentar autocorrelação para uma estação de sazonalidade s . Os modelos que contemplam as Séries, os quais apresentam autocorrelação sazonal são conhecidos como *SARIMA*.

Os modelos *SARIMA* contêm uma parte não sazonal, com parâmetros (p, d, q) ; e uma sazonal, com parâmetros (P, D, Q) . O modelo $ARMA(p, q)(P, Q)$ é dado por:

$$\begin{aligned} & (1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \dots - \Phi_{ps} B^{ps}) Z_t \\ & = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^s - \dots - \Theta_{qs} B^{qs}) a_t \end{aligned} \quad (65)$$

Conforme Farias, Rocha; Lima (2000), se a Série em estudo for não-estacionária com sazonalidade, temos um $ARIMA(p, d, q)(P, D, Q)$, no qual d representa a ordem de diferenciação não sazonal; e D , a ordem de diferenciação sazonal. A diferenciação sazonal pode ser realizada da seguinte forma:

$$\Delta^D Z = Z_t - Z_{t-s} \quad (66)$$

A ocorrência de estacionaridade está relacionada a uma queda brusca no valor dos picos sazonais. A FAC também mostrará se há componente sazonal, o que pode ser observado quando esta segue um padrão de picos e depressões ano a ano. Da mesma forma, a FAC mostrará a não-estacionaridade sazonal.

A diferenciação da Série em termos não sazonal e sazonal tem por objetivo torná-la estacionária, o que possibilita a aplicação da metodologia de Box & Jenkins.

2.3 Modelos Fracionários – ARFIMA (p, d, q) - Auto-regressivo Fracionário Integrado de Médias Móveis

Os modelos ARIMA (p, d, q), incluem o parâmetro d , um inteiro, que estabelece o nível de diferenciações necessárias para tornar uma série temporal estacionária, onde normalmente a diferenciação de até 2ª ordem é suficiente para estacionarizar a série. Esses modelos são adequados para a modelagem do comportamento de séries temporais a curto prazo. Granger e Joyeux (1980) e ainda Hosking (1981) propõem uma generalização dessa modelagem em relação ao parâmetro d , podendo esse assumir não só valores inteiros, mas também representar graus de diferenciação fracionários. Esses são Modelos com Memória Longa, por permitirem estudar séries caracterizadas por longas dependências temporais. Esses modelos intitulam-se ARFIMA (p, d, q) - Auto-regressivo Fracionário Integrado de Médias Móveis- onde F significa "fractional". A metodologia ARFIMA é fundamentada em Morettin&Toloi (2004).

Foi aproximadamente na década de 80, que economistas notaram evidências do fenômeno memória longa em dados econômicos e financeiros. Desde então, pesquisadores ligados à área de séries temporais e econometria vêm se dedicando a desenvolver modelos que capturam tal fenômeno. A longa dependência é observada não só na série temporal em nível, mas também pode estar presente na dependência serial da série quadrática, onde nesses casos têm-se processos heteroscedásticos condicionais de memória longa.

O fenômeno de Memória Longa foi notado por Hurst (1951, 1967), Mandelbrot e Wallis (1968) e McLeod e Hipel (1978), em conjunção com problemas na área de Hidrologia. Modelos de Memória Longa (ML) também são de interesse na análise de estudos climáticos, como no estudo da aparente tendência crescente em temperaturas globais devido ao efeito estufa. (Moretin&Toloi, 2004, pág. 460). Tais séries apresentam persistência nas autocorrelações amostrais, isto é, dependência significativa entre observações separadas por um longo intervalo de tempo.

O que diferencia os modelos ARIMA e ARFIMA é que, em um modelo ARIMA, a função de autocorrelação (FAC) representada por ρ_j decresce rapidamente para zero:

$$|\rho_j| \leq Cr^{-j} \quad ,j = 1,2,\dots$$

Com: $C > 0$ (67)

$$0 < r < 1$$

Enquanto que os modelos ARFIMA, responsáveis por capturar e modelar processos com longa dependência serial nos dados, onde essa dependência costuma-se chamar de memória longa, por ser processo estacionário em que a função de autocorrelação decresce hiperbolicamente para zero

$$\rho_j \sim Cr^{2d-1} \quad ,j \rightarrow \infty \quad (68)$$

Com:

$$C > 0$$

$$0 < r < 0,5$$

Para modelar uma série temporal, é possível realizar apenas pelos modelos ARFIMA, pois estes modelos são capazes de descrever, as dinâmicas de memória curta e longa de processos fracionários, onde d deve explicar a estrutura de correlação de ordens altas, enquanto que ϕ e θ explicam a estrutura de correlação de ordens baixas.

Um processo de memória longa é um processo estacionário em que a função de autocorrelação decresce hiperbolicamente para zero (MORETTIN, 2006). Formalmente, suponha que Z_t tenha auto-correlação ρ_j , diz-se que Z_t possui memória longa se em uma análise gráfica da função de autocorrelação amostral possuir um decaimento lento. As séries com memória longa, a Função de Autocorrelação da série original parece ser não estacionária, mas ao realizar a primeira diferença, a série parece ser super diferenciada, por isso a necessidade da diferenciação ser fracionária.

Portanto, os modelos ARFIMA (p, d, q) são uma generalização dos modelos ARIMA (p, d, q) , ou seja, os modelos ARFIMA são modelos ARIMA com integração fracionária ($d \in \mathfrak{R}$), modelos esses propostos inicialmente por Granger e Joyeux (1980) e Hosking (1981).

Para $\forall \{d \in \mathfrak{R} / d - 1\}$, o operador diferença fracionária:

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \quad (69)$$

$$(1-B)^d = 1 - dB + \frac{1}{2!}d(d-1)B^2 - \frac{1}{3!}d(d-1)(d-2)B^3 + \dots \quad (70)$$

Diz-se que Z_t é um processo auto-regressivo fracionário integrado de médias móveis, ou ARFIMA (p, d, q) com $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, se $\{Z_t\}$ for estacionário e satisfizer a equação:

$$\phi(B)(1-B)^d Z_t = \theta(B)a_t \quad (71)$$

Onde:

- a_t é ruído branco ;
- e $\phi(B)$ e $\theta(B)$ são polinômios em B de graus p e q, respectivamente. O efeito do parâmetro d decai hiperbolicamente conforme a distância aumenta e ϕ e θ decaem exponencialmente, portanto:

- d explica a estrutura de correlação de ordem alta da série,

- ϕ e θ explicam a correlação de ordem baixa da série.

Condições de Estacionaridade e Invertibilidade

O processo ARFIMA (p,d,q) , é:

- **estacionário** se $d < \frac{1}{2}$ e todas as raízes de $\phi(B) = 0$ estiverem fora do círculo unitário.

- **invertível** se $d > -\frac{1}{2}$ e todas as raízes de $\theta(B) = 0$ estiverem fora do círculo unitário.

- **Estimação de "d"**

Na literatura existem vários métodos para estimar o parâmetro d do modelo ARIMA(p,d,q), sendo os métodos paramétricos os mais usuais. Reisen (1995) destaca três métodos:

- o método da regressão utilizando o periodograma;
- o método da regressão utilizando o periodograma suavizado;
- o método utilizando o coeficiente de Hurst.

Dentre os métodos destacados, dois são baseados na utilização do periodograma, pois este é estimador da função espectral de uma série temporal.

Nesta pesquisa a estimação, como será encontrada por software, será através de sucessivas tentativas.

2.4 Identificação, Estimação, Adequação do Modelo e Previsão: ARIMA, SARIMA e ARFIMA

É importante destacar que os modelos ARIMA, SARIMA e ARFIMA não apresentam especificidade gráfica, somente plotando a Série Temporal, não sendo possível identificar qual o modelo que a representa. Para que se possa determinar qual o modelo dessa família que melhor representa a Série, e que poderá ser utilizado para fazer as previsões, a sua construção será baseada no ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados.

Caso o modelo não seja adequado, o ciclo é repetido, voltando à fase de identificação. Um procedimento que muitas vezes é utilizado é identificar não um só modelo, os quais serão então estimados e verificados. Se o propósito é previsão, escolher-se-á, entre os modelos ajustados, o melhor; por exemplo, no sentido de fornecer o menor erro quadrático médio de previsão.

Para os autores, a fase crítica do procedimento é a identificação, pois é possível que vários pesquisadores identifiquem diferentes modelos para uma única Série em estudo. Isso é possível, inclusive, pelo critério de decisão adotado para escolher o modelo adequado. Em geral, os modelos postulados são parcimoniosos, por terem um pequeno número de parâmetros e pelas previsões obtidas serem bastante precisas.

A estratégia utilizada para a construção de modelos, pela metodologia Box & Jenkins, é baseada no ciclo iterativo: identificação do modelo, estimação dos parâmetros, diagnóstico ou adequação do modelo e previsão.

2.4.1 Identificação do modelo

Para Barros (2006), no passo de identificação de um modelo ARIMA, busca-se encontrar procedimentos que permitem identificar os parâmetros do modelo para a Série Temporal em estudo. Um ponto importante na identificação é a parcimônia, pois de dois modelos com ajustes igualmente bons, escolhe-se aquele com menor número de parâmetros. Isso deve-se à simplicidade da estrutura, permitindo uma melhor compreensão do processo, enquanto que os modelos superparametrizados se ajustam bem ao que acontece dentro da amostra, mas não generalizam o comportamento da Série. O primeiro passo para a identificação é realizar a ordem da diferenciação da Série, que tende a introduzir correlações negativas. Por isso, é importante ter o cuidado de não diferenciar a Série original mais do que o necessário para estacionarizá-la.

Nesse sentido, considera-se essa a fase mais crítica do processo. A escolha do melhor modelo ajustado é feita principalmente com base nas Funções de Autocorrelação (FAC) e Funções de Autocorrelação Parcial (FACP) estimadas.

A FAC ρ_i é estimada por:

$$r_i = \frac{c_i}{c_0}, i = 0, 1, \dots, N - 1. \quad (72)$$

Em que c_i é estimada pela Função de Autocovariância (FACV) γ_i :

$$c_i = \frac{1}{N} \sum_{t=1}^{N-i} [(Z_t - \bar{Z})(Z_{t+i} - \bar{Z})], i = 0, 1, \dots, N - 1. \quad (73)$$

Em que $\bar{Z} = \frac{1}{N} \sum_{t=1}^N Z_t$ é a média amostral. É importante lembrar que $r_{-i} = r_i$:

A variância aproximada de r_i :

$$\text{Var}(r_i) \cong \frac{1}{N} \sum_{v=-\infty}^{\infty} [\rho_v^2 + \rho_{v+i}\rho_{v-i} - 4\rho_i\rho_v\rho_{v-i} + 2\rho_v^2\rho_i^2]. \quad (74)$$

Em um processo no qual as autocorrelações são nulas para $v > q$, todos os termos do lado direito de (74) anulam-se para $i > q$, exceto o primeiro, obtendo-se:

$$\text{Var}(r_i) \cong \frac{1}{N} \left[1 + 2 \sum_{v=1}^q \rho_v^2 \right], i > q. \quad (75)$$

Como as autocorrelações são desconhecidas, substitui-se ρ_v por r_v , obtendo-se as estimativas para (75):

$$\sigma^2(r_i) \cong \frac{1}{N} \left[1 + 2 \sum_{v=1}^q r_v^2 \right], i > q. \quad (75)$$

Para N suficientemente grande e sob a hipótese de $\rho_i = 0$, para $i > q$, será criado um intervalo de confiança aproximado para as autocorrelações, pois a distribuição da estimativa aproxima-se de uma distribuição normal, com média igual a zero e variância encontrada através da equação (75). Logo:

$$r_i \pm t_\gamma \cdot \hat{\sigma}(r_i). \quad (76)$$

Em que t_γ é o valor de t de *Student* com $N-1$ graus de liberdade, tal que $P(-t_\gamma < t < t_\gamma) = \gamma$. Na prática, é utilizado $t_\gamma = 2$, sendo assim, possível considerar ρ_i como sendo significativamente diferente de zero, se:

$$|r_i| \pm 2 \cdot \hat{\sigma}(r_i), i > q. \quad (77)$$

A FACP é analisada sob a hipótese de um modelo $AR(p)$, em que:

$$\text{Var}(\hat{\phi}_i) \cong \frac{1}{N}, i > p \quad (78)$$

Como as FAC são analisadas através de estimativas:

$$\hat{\sigma}(\hat{\phi}_{ii}) \cong \frac{1}{\sqrt{N}}, i > p \quad (79)$$

No caso de N apresentar-se muito grande e sob hipótese de o modelo ser $AR(p)$, $\hat{\phi}_{ij}$ terá uma distribuição aproximadamente normal, com média zero e variância encontrada através da equação (78), sendo possível considerar $\hat{\phi}_{ii}$ significativamente diferente de zero, se:

$$|\hat{\phi}_{ii}| > \frac{1}{\sqrt{N}}, j > p \quad (80)$$

O objetivo do processo de identificação é encontrar a ordem dos parâmetros do modelo $ARIMA(p, d, q)$. Essa fase do modelo consiste em três partes:

- Verificar a necessidade de transformação da Série original, com o objetivo de estabilizar a variância, que pode ser identificada através de uma análise visual do gráfico da Série;
- Tomar diferenças da Série até torná-la estacionária. O número de diferenças necessárias é alcançado quando a FAC da Série decresce rapidamente para zero;
- Encontrar o processo através das FAC e FACP estimadas, no qual essas encontram seus valores dentro dos limites de confiabilidade estipulado.

Nesse processo de identificação, busca-se determinar a ordem dos modelos, tomando-se por base o comportamento das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP), além dos seus respectivos correlogramas (exposição gráfica das FAC e FACP contra as defasagens temporais).

2.4.2 Estimação dos parâmetros

Nesta fase da metodologia, cada um dos modelos sugeridos na fase de identificação é ajustado, e os vários parâmetros são examinados. Os modelos ajustados são comparados usando os critérios de parcimônia e os de estacionaridade e invertibilidade.

A estimação pode ser por mínimos quadrados e por máxima verossimilhança. Para a estimação da máxima verossimilhança, é necessário admitir, inicialmente, que os ruídos brancos têm a distribuição Normal.

Experimentos com simulação pelo método de Monte Carlo, realizados por vários autores, indicam que o método de máxima verossimilhança é superior ao método de mínimos quadrados, quando o tamanho da Série é pequeno e, sobretudo, quando os valores dos parâmetros aproximam-se dos limites da condição de invertibilidade. Para mais detalhes, ver Vasconcelos e Alves (2000, p.69-82).

2.4.3 Diagnóstico ou adequação do modelo

Como já comentado anteriormente, existe a possibilidade de serem encontrados vários modelos, candidatos à representação em estudo. Dentre esses, somente um deve ser escolhido. Portanto, recorre-se aos seguintes critérios de decisão:

- Parcimoniosidade nos parâmetros. Deve-se escolher, preferencialmente, entre os modelos de melhor ajuste, aqueles com um número menor de parâmetros;
- Invertibilidade e estacionaridade dos coeficientes dos modelos estimados;
- Ajustamento do modelo aos dados da Série Temporal em estudo, com base nos critérios *Akaike Information Criterion*, AIC, e *Schwartz Bayesian Criterion*, BIC.

Esses critérios permitem a comparação de diferentes modelos para a Série e, por conseguinte, a escolha do modelo com melhor ajustamento. É nesse critério que se minimiza a variância dos resíduos e que se estabelece a conseqüente penalidade pela inclusão de mais parâmetros. Isso porque se busca modelos parcimoniosos e, geralmente, a inclusão de mais parâmetros leva à melhora do ajuste do modelo, mas peca pelo excesso de parametrização.

Criado por Akaike, em 1974, o critério *AIC* (*Akaike Information Criterion*) estabelece que o modelo ótimo para uma Série é aquele que minimiza o *AIC*, e é representado por:

$$AIC = \ln \hat{\sigma}_\varepsilon^2 + \frac{2(p+q)}{n} \quad (81)$$

Segundo Barros (2006), esse critério tende a superestimar os modelos autorregressivos. Com o intuito de corrigir esse problema é que se desenvolveu o próximo critério, *Schwartz Bayesian Criterion*, BIC.

$$BIC = \ln \hat{\sigma}_\varepsilon^2 + \frac{(p+q)\ln(n)}{n} \quad (82)$$

Em que:

$\hat{\sigma}_\varepsilon^2$ = soma do quadrado dos resíduos;

p, q = número de parâmetros estimados;

n = número de observações utilizadas.

Criado por Schwartz, em 1978, esse critério vem a ser um refinamento do critério AIC, desenvolvido justamente para resolver o problema encontrado nele.

Os critérios AIC e BIC podem ser negativos. Isso ocorre quando a soma do quadrado dos resíduos é menor que 1 e $\ln \hat{\sigma}_\varepsilon^2 > 2(p+q)$ ou $\ln \hat{\sigma}_\varepsilon^2 > (p+q)\ln(n)$

Um fator a ser observado é quanto aos resíduos do modelo estimado, se são estimativas do ruído branco a_t . Assim sendo, devem comportar-se como um ruído branco, se o modelo estiver adequadamente especificado. Em particular, seus coeficientes de autocorrelação devem ser estatisticamente iguais a zero. Os resíduos se distribuem como uma Normal de média zero e variância $1/n$.

Nos modelos ARFIMA utilizam apenas o AIC como critério penalizador para decidir qual o modelo mais parcimonioso. Em vez de estabelecer p e q precisamente, estimam-se os modelos correspondentes a vários pares (p, q) e escolhe-se aquela especificação que apresentar o menor valor para AIC.

A presença de p e q nas fórmulas dos critérios AIC e tem por objetivo “penalizar” os modelos com muitos parâmetros, tendo em vista que modelos mais parcimoniosos devem ser privilegiados por apresentarem menor número de parâmetros a ser estimado. A crítica que se faz a esses critérios é que eles podem conduzir a modelos super especificados, ou seja, modelos com valores de p e/ou q maiores do que o correto. Assim sendo, o critério AIC devem ser usados como procedimento complementar e não alternativo àquele baseado na FAC e na FACP.

O processo ARFIMA (p,d,q) estacionário e invertível dado por

$$\phi(B)(1-B)^d Z_t = \theta(B)a_t, \text{ com } d \in \left(-\frac{1}{2}, \frac{1}{2}\right) \quad (83)$$

Pode-se reescrever o processo na forma de choques aleatórios

$$Z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad (84)$$

e na forma invertida

$$\sum_{j=0}^{\infty} \pi_j Z_{t-j} = a_t \quad (85)$$

Onde

$$\sum_{j=0}^{\infty} \psi_j B^j = \theta(B)\phi^{-1}(B)(1-B)^{-d} \quad (86)$$

e

$$\sum_{j=0}^{\infty} \pi_j B^j = \phi(B)\theta^{-1}(B)(1-B)^d \quad (87)$$

Utilizando as equações 86 e 87, podem-se realizar as previsões dos valores futuros do processo Z_t com h passos à frente. Outra forma de se reescrever um processo ARFIMA é pela equação da diferença:

$$\varphi(B)Z_t = \theta(B)a_t, \quad (88)$$

Onde $\varphi(B) = \phi(B)(1-B)^d = \phi(B)D(B)$ e $D(B) = 1 - d_1B - d_2B^2 - \dots$ é um polinômio em B de ordem infinita, pois quando tem-se uma série com n observações, utiliza-se somente os L primeiros termos desse polinômio, com $L < n$.

2.4.4 Previsão

Essa é a última etapa do processo dos modelos ARIMA, seguindo o Ciclo Iterativo. Os passos construídos até aqui se devem ao interesse de utilizar o modelo identificado e estimado para se realizar as previsões. Tais Previsões são representadas por Z_{t+h} , $h > 1$, sendo que se tem uma Série original Z_t é até o instante t ; e a previsão de origem t e *horizonte* h será denotada por $\hat{Z}_t(h)$. Também, será assumido que $W_t = (1 - B)^d Z_t$ é estacionário e invertível, e os parâmetros são conhecidos.

Já que o objetivo agora é a previsão, é possível reescrever os modelos ARIMA(p,d,q) nas três formas básicas:

Forma de Equação de Diferença:

$$Z_{t+h} = \omega_1 Z_{t+h-1} + \dots + \omega_{p+d} Z_{t+h-p-d} - \theta_1 a_{t+h-1} - \dots - \theta_q a_{t+h-q} + a_{t+h} \quad (89)$$

Forma de Choques Aleatórios:

$$Z_{t+h} = \sum_{i=-\infty}^{t+h} \psi_{t+h-i} a_i = \sum_{i=0}^{t+h} \psi_j a_{t+h-i} \quad (90)$$

Em que $\psi_0 = 1$.

Forma Invertida:

$$Z_{t+h} = \sum_{i=1}^{\infty} \pi_i Z_{t+h-i} + a_{t+h} \quad (91)$$

2.5 Regressão Dinâmica

O termo regressão foi introduzido por Francis Galton, em um famoso ensaio, onde verificou que, embora houvesse uma tendência de pais mais altos terem filhos altos e de pais baixos terem filhos baixos, a altura média dos filhos de pais de uma dada altura tendia a se

deslocar ou “regredir” até a altura média da população. A lei de regressão universal de Galton foi confirmada por Karl Pearson, que coletou mais de mil registros das alturas dos membros de grupos das famílias. Ele verificou que a altura média dos filhos de um dos grupos de pais altos era inferior á altura de seus pais, e que a altura média dos filhos de um grupo de pais baixos era superior á altura de seus pais. Os dois autores estavam interessados em provar a validade de uma suposta ‘Lei de Regressão Universal’, que estabelecia que as características de uma pessoa eram transmitidas de pai para filho de forma amortecida (Gujarati,2000; Vasconcellos & Alves, 2000).

Ao se realizar uma análise de regressão entre uma variável dependente e no mínimo uma variável independente, tem-se por objetivo encontrar a relação existente, através de um modelo matemático. Quando fala-se em variáveis dependentes e independentes é devido a relação matemática entre elas, não necessariamente uma relação de causa e efeito.

Embora as variáveis independentes possam explicar a variação na variável dependente, isto não implica necessariamente uma causação. O emprego das expressões variável dependente ou de critério, variável independente ou prognosticadora ou preditora na análise de regressão decorre da relação matemática entre as variáveis. Esses termos não implicam que a variável de critério seja dependente das variáveis independentes em um sentido causal. O objetivo da análise de regressão é a natureza e o grau de associação entre variáveis, não implica nem supõe qualquer causalidade (Malhorta, 2001).

Para Barbetta (2004) os modelos dinâmicos, são essencialmente um modelo econométrico de regressão múltipla. Portanto, na sua análise, objetiva-se encontrar um modelo estatístico-matemático para descrever a relação existente entre variáveis independentes e dependentes e conhecer a influência de cada variável independente, bem como prever a variável dependente em função das independentes.

Os modelos ditos dinâmicos, porém, não consideram as variáveis apenas em seus valores correntes, como nas análises “puras” de regressão múltipla, mas também seus valores defasados. Esta característica, em função de permitir que sejam descritas as relações entre as variáveis consideradas ao longo do tempo, é que lhes dão o adjetivo de dinâmicos, pois “retrata o caminho temporal da variável dependente em relação ao(s) seu(s) valor(es) passado(s)” (GUJARATI, 2000, p.590).

Gujarati (2000) observa que na análise de modelos de regressão dinâmica envolvendo dados de séries temporais, apresenta a duas características:

- se apresentar tanto na forma de defasagem distribuída, que utilizada como variáveis explicativas os valores defasados de X 's, representado por:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \varepsilon_t \quad (92)$$

- ou na forma auto-regressiva explicativa dos valores defasados de Y . Se o modelo incluir no mínimo um valor defasado da variável dependente entre suas variáveis explicativas, será conhecido como modelo auto-regressivo, também conhecido como modelo dinâmico, por retratar o caminho temporal da variável dependente em relação aos seus valores passados. O modelo dinâmico é descrito pela seguinte equação:

$$Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + \varepsilon_t \quad (93)$$

Observa-se que β é o peso da defasagem distribuída de X enquanto que γ é a influência auto-regressiva de Y no modelo de previsão.

Generalizando, temos

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_k X_{t-k} + \varepsilon_t \quad (94)$$

Um modelo de defasagem distribuída com uma defasagem finita de k períodos de tempo. O coeficiente β_0 é o multiplicador a curto prazo, por alterar o valor médio de Y devido a alteração de uma unidade em X no mesmo período de tempo¹. Se a alteração em X é mantida no mesmo nível, então:

- $(\beta_0 + \beta_1)$ Altera, em valor médio, Y no próximo período;
- $(\beta_0 + \beta_1 + \beta_2)$ altera no seguinte período e assim sucessivamente

Estimativas dos Modelos

Segundo Gujarati (2000), são vários os modelos a seguir, dentre eles:

- *Koyck*:

$$Y_t = \alpha(1 - \lambda) + \beta_0 X_t + \lambda Y_{t-1} + (\varepsilon_t - \lambda \varepsilon_{t-1}) \quad (95)$$

- *Estimativa adaptativa*:

¹ Tecnicamente, β_0 é a derivada parcial de Y em relação a X_t ; β_1 em relação a X_{t-1} ; β_2 em relação a X_{t-2} e assim sucessivamente. Portanto, resumindo: $\frac{\partial Y_t}{\partial X_{t-k}} = \beta_k$

$$Y_t = \gamma\beta_1 X_t + (1 - \gamma)Y_{t-1} + [\varepsilon_t - (1 - \gamma)\varepsilon_{t-1}] \quad (96)$$

- Ajustamento Parcial

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1 - \delta)Y_{t-1} + \delta\varepsilon_t \quad (97)$$

Observando os modelos anteriores, nota-se que apresentam uma forma em comum:

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 X_{t-1} + \varepsilon_t \quad (98)$$

Os modelos de regressão dinâmica é uma teoria de previsão, que busca modelos ótimos de previsão baseados em casualidade das variáveis exógenas para as endógenas a serem estimadas. Propõe-se o modelo linear dinâmico, o qual considera o passado das variáveis defasadas no tempo.

Tem-se ainda, em Gujarati (2000), a seleção do modelo segundo a abordagem desenvolvida na London School of Economics- LSE, na década de 60, ou metodologia de Hendry, popularmente conhecida como modelagem “**geral para específico**”, por se iniciar com um modelo mais geral, com diversos regressores, e em seguida vá se reduzindo gradualmente esse mesmo modelo, conservando apenas as variáveis importantes, ou seja, significativas. Esta pesquisa está fundamentada nesse modelo.

Tal redução, feita através da eliminação dos regressores que mostram-se estatisticamente insignificantes ao modelo, é embasada nos resultados do teste t - Gosset (que mensura a significância da variável de forma individualizada no modelo) e do teste F - Snedecor (que mensura a significância da variável em conjunto com as demais), simultaneamente, até o momento em que se acredite ter encontrado o modelo específico.

Gujarati (2000) mostra que a metodologia LSE foi desenvolvida para lidar com dados em forma de séries temporais, sendo a relação a longo prazo, onde a relação é sintetizada por

$$Y_t = \alpha X_t \quad (99)$$

Esse modelo apresenta uma relação a longo prazo, por levar tempo a se concretizar, portanto essa metodologia propõe o modelo dinâmico reescrito da seguinte forma:

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_m X_{t-m} + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \dots + \delta_m Y_{t-m} + \varepsilon_t \quad (100)$$

Observa-se que Y é a variável dependente no tempo t , regredida sobre os regressores X nos tempos $t, (t - 1), (t - 2), \dots, (t - m)$, como também seus valores defasados nos instantes $(t - 1), (t - 2), \dots, (t - m)$. Este é um modelo de regressão dinâmica por ser auto-regressivo de defasagem distribuída.

Como o modelo encontrado é através da eliminação dos valores menos significativos, fazendo parte do modelo final somente as variáveis e suas defasagens que apresentarem $p < 0,05$, e a regularidade do modelo se dará através das saídas do software utilizado, *Eviews*. Portanto será necessário analisar os testes de saída que o modelo significativo apresentar, satisfazendo os seguintes critérios, segundo Gujarati (2000):

- Ser admissível quanto aos dados, o que significa dizer que as previsões geradas a partir do modelo devem assumir valores aceitáveis tendo-se em vista a amostra utilizada;
- *Ser compatível com os dados*: o modelo deve apresentar previsões logicamente possíveis;
- *Ser compatível com a teoria*: deve fazer sentido econômico;
- *Ter regressores fracamente exógenos*: que os regressores não tenham correlação com o erro;
- *Exibir constância nos parâmetros*: estabilidade dos parâmetros, possibilitando a sua utilização para realizar previsões;
- *Exibir coerência nos dados*: que consiste basicamente em os resíduos do modelo ser ruído branco com variância constante (homocedásticos) e seguirem uma distribuição normal;
- *Ser abrangente*: o modelo escolhido deve abranger todos os outros modelos.

Encontrado o modelo, testar se:

- *os erros ou ruído branco* apresentam variância constante, ou seja, se os erros são homocedásticos. Nesse caso é o ARCH quem testa a heterocedasticidade do modelo auto-regressivo condicionado, ao analisar se a constante que multiplica o erro nos períodos defasados é igual a zero;
- *Autocorrelação residual do modelo*, através do teste AR, que verifica a existência de correlação entre os resíduos atuais e os defasados;
- *Normality test* que verifica se todos os resíduos seguem uma distribuição normal;
- *Hetro test*, que testa se a variância encontrada está ligada aos regressores;
- *Reset test* verifica se o modelo não apresenta problemas na sua fórmula funcional.

Em todos estes testes são testadas as hipóteses H_0 e H_1 , ao nível de significância estipulado pelo pesquisador. Esta significância é analisada através do p valor. Por exemplo, testando a normalidade:

H_0 : os resíduos seguem uma distribuição normal, ou seja, $p > \alpha$.

H_1 : os resíduos não seguem uma distribuição normal, ou seja, $p < \alpha$.

2.6 Combinação de Previsão

Os precursores da combinação de previsão foram Bates & Granger em 1969 e ela tem sido objeto de estudo de vários autores, que observam que as combinações de previsões podem chegar a resultados superiores dos que gerados por técnicas individuais. Apesar da literatura sobre essa técnica ser vasta, autores ainda não chegaram a um consenso de qual combinação apresenta melhores resultados.

Clement (1989), considera a combinação de previsão uma técnica atraente devido aos diversos fatores que influenciam as previsões, então cada técnica contribui capturando informações características, que os demais modelos não conseguiram capturar. Apesar de não serem definidas as melhores condições para se realizar combinações, nem qual o método mais eficiente em determinada situação, o autor enfatiza que são necessárias previsões individuais eficientes para que se atinjam resultados significativos com a combinação. Portanto, cada técnica precisa ser criteriosa na escolha de seus modelos.

De acordo com Zou e Yang (2004 apud SOUZA, 2005), nas últimas décadas de estudos sobre combinação de previsões, vários métodos foram propostos e, apesar de não se ter chegado a um consenso com relação a qual a melhor forma de se combinar previsões, pode-se concluir que “combinar tende a reduzir o erro de predição”. Ao se realizar essas combinações é possível reduzir a variabilidade das previsões em relação ao valor verificado, o que podem torná-las mais exatas e precisas, sob o critério de avaliação empregado.

Makridakis&Winkler (1983) enfatizam que a combinação de previsões aumenta a acurácia conforme o número de técnicas utilizadas, embora o grau de saturação tenha sido encontrado com a combinação de quatro ou cinco técnicas, enquanto Libby e Blashfield (1978) destacam que as melhores previsões e acurácias resultam da combinação de duas a três previsões.

Rausser e Oliveira (1976), lembram ser muito comum encontrarmos mais de um modelo representativo da série em estudo em cada técnica, e comumente aceitarmos o melhor

modelo, segundo algum critério. Quando é realizado o descarte dos demais modelos, geralmente algumas informações úteis também são descartadas. Logo, é aconselhável considerar várias técnicas e combiná-las em vez de considerar uma única técnica para a previsão.

Para Mahmoud (1989 apud WERNER, 2004), a combinação de previsão é incentivada por três razões: melhora a acurácia e diminui a variância do erro da previsão, pesquisadores com pouca experiência sobre o assunto podem utilizar métodos simples de combinação, que também resulta em melhoras à acurácia e a combinação pode ter pouco ou nenhum custo.

Armstrong (1984 apud WERNER), realizou um levantamento de estudos publicados que comparava modelos de previsão e encontrou 39 estudos, dos quais 13 envolviam modelos Box & Jenkins. Desses 13 estudos, 4 foram mais eficientes que técnicas simples, em 8 não apresentaram diferença significativa com relação as técnicas simples e em apenas um estudo as técnicas simples foram mais acuradas que os modelos Box & Jenkins. Com esse levantamento, conclui-se que os modelos Box & Jenkins comportam informações mais úteis para aumentar a acurácia da combinação de previsões.

Para Makridakis (1989), a metodologia Box & Jenkins analisa séries temporais univariadas, não sendo possível a inclusão de outras informações no modelo, como temperatura, fatores naturais ou outras variáveis que podem contribuir com o aumento da acurácia. Ainda para o autor, modelos econométricos são ferramentas valiosas para aumentar o entendimento da série em estudo, sendo muito interessante a combinação de modelos de regressão com modelos Box & Jenkins.

Newbold e Granger (1974 apud WERNER, 2004) combinaram três previsões obtidas de modelos Box & Jenkins, *Holt-Winters* e auto-regressão de *Stepwise*, e concluíram que a combinação ponderada trouxe melhorias para a acurácia da previsão combinada. Os autores concluíram que a combinação de três técnicas de previsão oferecem melhores combinações, pois capturam diferentes informações.

Libby e Blashfield, *Newbold e Granger e Makridakis*, apresentaram contribuições sobre as melhores previsões serem resultado da combinação de três modelos.

Paiva (2006, pág 99), desenvolveu um índice Geral Multivariado (IGM) para realizar a combinação de previsão através de um índice multivariado, que será encontrado considerando as previsões de cada modelo como componentes principais.

$$IGM = \sum_{i=1}^m [\lambda_i(PCs_i)] \quad (101)$$

Onde: m = número de componentes principais significativos, λ_i = i -ésimo autovalor e PCs_i = i -ésimo escore de componentes principais.

Buscou-se adaptar esse índice para o Índice de Ponderação de Autovalores para Combinação de Previsões, o qual levará em consideração o percentual de explicação de cada autovalor, ou seja:

$$IPACP = \sum_{i=1}^n \left[\frac{\lambda_i}{\sum_{i=1}^n \lambda_i} (P_i) \right] \quad (102)$$

O objetivo da análise de componentes principais é reduzir as dimensões, de modo que sejam utilizadas apenas as informações mais importantes, ou seja, os m -primeiro componentes principais. Para Johnson e Wichern (2007), para se fazer uma interpretação correta de quais componentes utilizar no estudo, basta selecionar as primeiras componentes que acumulem uma percentagem de variância explicada, igual ou superior a 70%.

O percentual da variância explicada é dado pelo percentual que cada autovalor está explicando a componente principal, ou seja, nesta pesquisa, na está se buscando as componentes principais, pois a previsão encontrada em cada técnica é uma componente principal. O que precisa-se descobrir através dos autovalores, qual é o percentual de contribuição que cada técnica para a combinação da previsão. Portanto, encontra-se os autovalores, através de componentes principais e na seqüência o percentual de explicação de cada autovalor e aplica-se fórmula do IPACP (102).

Para encontrar os autovalores, parte-se da matriz de variância e covariância, expressa pela ligação entre as n variáveis, encontradas pela sua covariância. Considerando as variáveis P_1, P_2, \dots, P_n , denota-se a matriz da covariância por:

$$S = \begin{bmatrix} Var(P_1) & Cov(P_1, P_2) & \dots & Cov(P_1, P_n) \\ Cov(P_1, P_2) & Var(P_2) & \dots & Cov(P_2, P_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(P_1, P_n) & Cov(P_2, P_n) & \dots & Var(P_n) \end{bmatrix}$$

Os termos situado na intersecção i -ésima linha com j -ésima coluna com $i \neq j$, são os termos da covariância e os termos situados na diagonal principal $i = j$, são os termos da variância.

$$\text{Variância} \Rightarrow \text{Var}(P_j) = \frac{1}{n-1} \left[\sum_{i=1}^n P_{ij}^2 \frac{\sum_{i=1}^n (P_{ij})^2}{n} \right] \quad (103)$$

$$\text{Covariância} \Rightarrow \text{Cov}(P_i, P_j) = \frac{1}{n-1} \left[\sum_{i=1}^n P_{ij} P_{ij} - \frac{\sum_{i=1}^n (P_{ij})^2}{n} \right] \quad (104)$$

Quando existir a necessidade de padronização dos dados, com o objetivo de se evitar problemas de magnitude, é utilizada a matriz de correlação. Considerando P_1, P_2, \dots, P_n , a matriz de correlação denotada por r :

$$r = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix}$$

Na qual,

$$r_{ij} = r(P_i, P_j) = \frac{\text{cov}(P_i, P_j)}{\sqrt{\text{var}(P_i) \cdot \text{var}(P_j)}} \quad (105)$$

Encontrada a matriz S , da variância e covariância, quadradas de ordem $p \times p$ e sendo I a matriz identidade $p \times p$, então os escalares $\lambda_1, \lambda_2, \dots, \lambda_p$, satisfazem:

$$|S - \lambda I| = 0 \quad (106)$$

Da equação (106), encontraremos p autovalores onde $\lambda_1 > \lambda_2 > \dots > \lambda_p$.

Acrescentando um vetor (X) não nulo ($X \neq 0$), tal que

$$S\vec{X} = \lambda\vec{X} \quad (107)$$

Onde \vec{X} é uma matriz $p \times p$ de todos os autovetores e λ uma matriz $p \times p$ de todos os autovalores, então \vec{x} será o autovetor da matriz S , associado a um autovalor λ .

Encontrados os autovalores é possível encontrar os autovetores. Como não estamos em busca de encontrar as componentes principais, pois as mesmas serão as próprias previsões de cada técnica, torna-se desnecessário calcular os autovetores. Mas a fins de conhecimento, conforme Ragazzi (2001), tomando os elementos do vetor \vec{x} , assim determinados como coeficiente de Y_i , tem-se que o i -ésimo componente principal é dado por:

$$Y_i = x_{i1}X_1 + x_{i2}X_2 \cdots + x_{ip}X_p \quad (108)$$

Onde, Y_i é componente principal, \vec{x} é um autovetor e \vec{X} é a matriz dos autovetores.

Tem-se ainda que:

- $Var(Y_i) = \lambda_i$, logo a $Var(Y_1) > Var(Y_2) > \cdots > Var(Y_p)$;
- $\sum Var(X_i) = \sum \lambda_i = \sum varY_i$;
- $Cov(Y_i, Y_j) = 0 \Leftrightarrow \sum_{j=i}^p x_{ij} x_{kj} = 0$

Dessas propriedades observa-se que a variância da i -ésima componente principal é igual ao autovalor correspondente a está componente principal e que a componente principal é medida em termos da variância, logo pode-se representar por percentagem.

$$\frac{Var(Y_i)}{\sum_{i=1}^p Var(Y_i)} \times 100 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \times 100 \quad (109)$$

A equação (109) pode ser encontrada através da equação (106), sendo que essa expressão representa a proporção da variância total explicada pelo seu componente principal, no caso desta pesquisa, proporção da variância total explicada por cada técnica de previsão. Será através do peso proporcional cada modelo contribui, para a previsão, que será utilizada para realizar a previsão, através de combinação de previsão.

Encontradas as previsões representativas de cada modelo, os autovalores e através desse, a proporção da variância total explicada por cada técnica de previsão, realizar a Combinação das Previsões.

A Combinação de Previsão é proposta com o objetivo de apresentar uma combinação que inclua todas as informações dos modelos, e que apresente previsões melhores do que as propostas por modelos individuais. Para certificar-se que realmente as previsões combinadas apresentam melhores acurácias, será utilizado o MAPE – Erro Percentual Absoluto Médio, e o EQM - Erro Quadrado Médio, o qual compara os modelos através dos desvios em relação aos valores esperados.

O MAPE será calculado através de:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Z_t - P_t}{Z_t} \right| \quad (110)$$

E o EQM

$$EQM = \frac{\sum_{t=1}^n (Z_t - P_t)^2}{n} \quad (111)$$

Sendo Z_t o valor real, P_t o valor previsto e n o tamanho da amostra.

Para a combinação de previsão ser mais importante que as previsões analisadas individualmente ela precisa refletir a realidade, o que pode ser comprovado se essa técnica apresentar o MAPE e o EQM, menor que os outros modelos.

3 METODOLOGIA

3.1 Banco de dados

Os dados utilizados para análise referem-se ao preço médio mensal recebido pelo produtor pelo quilograma vivo do boi para abate, no período de janeiro de 2000 a dezembro de 2008, no estado do Rio Grande do Sul.

Os dados foram coletados junto a FEE – DADOS – Fundação de Economia e Estatística, disponibilizados pela EMATER/RS.

Para a análise dos dados será utilizados os Software

- *Statistic 7: para analisar os modelos ARIMA e ARFIMA;*
- *Ewies: para modelar os modelos ARFIMA e de Regressão Dinâmica.*

3.2 Passos Metodológicos

Os passos metodológicos utilizados para o cumprimento dos objetivos são:

➤ Análise do preço médio mensal recebido pelo produtor pelo quilograma vivo do boi para abate por meio:

✓ **Metodologia de Box & Jenkins: Modelos Auto-Regressivo Integrado de Médias Móveis - ARIMA:** Essa metodologia segue um ciclo iterativo onde, primeiramente é desenvolvido o gráfico da série original para identificar visualmente o comportamento da série, se a mesma é estacionária ou apresenta tendência crescente ou decrescente. É a funções

de autocorrelação que auxiliará na identificação da estacionaridade, ao apresentar um decaimento rápido ou a não estacionaridade através de um decréscimo muito lento.

Ao analisar o gráfico da série original e identificar-se que a mesma apresenta um comportamento não estacionário ou tendencioso será realizada sucessivas diferenças até obter-se uma série estacionária. Prosseguindo a análise, será o momento de estimar os modelos significativos, pois é possível encontrar vários modelos significativos para uma única série. Devido a isso, recorre-se aos critérios penalizadores, para optar qual é o modelo que melhor explica a série em estudo através dessa metodologia. Para decidir qual o melhor modelo será utilizado o critério penalizador AIC e BIC, que quanto menores seus valores, mais parcimonioso é o modelo.

Portanto, analisa-se o gráfico da série original, realiza-se a diferenciação necessária, estima-se os modelos e encontrado mais de um modelo, opta-se pelo modelo que apresentar os menores valores para AIC e BIC.

Definido o melhor modelo, chega-se a última etapa do Ciclo Iterativo: a previsão para o ano de 2009. Realizada a previsão, será calculado o MAPE e o EQM, com o intuito de certifica-se que a previsão aproxima-se dos valores reais.

✓ **Modelo Auto-repressivo Fracionário de Médias Móveis - ARFIMA:** esse modelo apresenta as mesmas características dos modelos ARIMA, o que os diferencia é que, muitas vezes, uma diferença está superdiferenciando a série, e a análise será prejudicada devido à perda de informações, portanto, uma diferença fracionária, $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, é suficiente para estacionarizar a série, sem perda de informações.

Ao analisa-se o comportamento da série original e observar a necessidade de estacionarizar a série para então encontrar o modelo que a represente, realizar diferenças, com $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. A análise prosseguirá com a estimação dos modelos que representarão a série e o critério que penalizador utilizado nos modelos ARFIMA é o AIC. O modelo que apresentar o menor AIC será o modelo utilizado para realizar a previsão do preço recebido pelo produtor pelo abate bovino no Rio Grande do Sul. Após, realizada a previsão, calculou-se o MAPE e EQM, com a finalidade de comparar os valores previstos dos valores reais.

✓ **Regressão Dinâmica:** na Regressão Dinâmica, buscou-se analisar se o preço do quilograma recebido pelo produtor é influenciado pelo preço de outros 14 animais da pecuária. Primeiramente, observou-se que as séries em estudo não são estacionárias, e por serem valores expressos em unidade monetária, a regressão dinâmica espera que esses valores sejam estacionários. Portanto, iniciou-se a análise diferenciando as 15 variáveis em estudo.

Na seqüência, a variável dependente, abate bovino, e as demais variáveis foram analisadas com 3 defasagens, totalizando 60 parâmetros no modelo inicial.

O modelo foi fornecido através do software *Ewies*, onde os parâmetros não significativos, $p > \alpha$, foram sendo eliminados em ordem decrescente, um a um. O modelo final resultou em 17 parâmetros com todos os seus valores significativos, $p < \alpha$, e satisfazendo as especificidades do teste, fornecidos pelo aplicativo do Software analisado, o *Ewies*. No modelo, o nível de significância adotado para análise dos testes é de 5%. Os testes analisados foram:

- AR, o qual testa a existência de autocorrelação serial entre os resíduos atuais e defasados;
- ARCH, para testar a heterocedasticidade do modelo auto-regressivo condicionado, ou seja, se o ruído branco apresenta variância constante, ou se os erros são homocedáticos;
- A normalidade dos resíduos, ou seja, os resíduos seguem uma distribuição normal;
- A heterocedasticidade dos dados, mostrando se a variância encontrada está ligada aos regressores;
- O teste *Reset*, que testa se o modelo apresenta problemas na sua fórmula funcional, ou seja, não apresenta problema de má especificação na equação do modelo e portanto as previsões podem ser utilizadas.

✓ **Combinação de Previsão:** Para realizar a combinação de previsão, desenvolveu-se um Índice de Ponderação de Autovalores para Combinação de Previsões (IPACP), equação 102, composto pelo somatório das previsões multiplicadas pela proporção total da variância explicada por cada previsão, ou seja, através das ponderações das previsões.

Esse índice foi uma adaptação da equação 101 - Índice Geral Multivariado (IGM), proposto por Paiva (2006, pág 99).

Para a Combinação de Previsões, utilizaram-se as previsões encontradas pelos modelos ARIMA, ARFIMA e Regressão Dinâmica. Com as previsões desenvolveu-se Componentes Principais para encontrar a proporção da variância total explicada por cada técnica de previsão, ou seja, que é o percentual de explicação de cada autovalor. A previsão fornecida por cada modelo são as componentes principais.

Portanto, com as previsões encontradas em cada técnica e a proporção da variância total explicada por cada técnica foi aplicado-se as informações na fórmula:

$$IPACP = \sum_{i=1}^n \left[\frac{\lambda_i}{\sum_{i=1}^n \lambda_i} (P_i) \right]$$

Realizada a combinação de previsão, utilizou-se o Erro Quadrado Médio e o MAPE para verificar se a previsão encontrada através da Combinação de previsão mostrou-se mais eficiente que as previsões individualmente encontradas por cada técnica analisada.

4 ANÁLISE DOS RESULTADOS

O Rio Grande do Sul caracteriza-se pelas condições climáticas, topográficas e hidrográficas favoráveis para a criação de gado, de maneira organizada e sistemática. Mas nos primórdios desta atividade, o principal objetivo era a produção de couro em grande quantidade, pois a população ainda era insuficiente para um grande consumo de carne.

Foi no final do século XVIII, com a decadência da pecuária nordestina, que desenvolveu-se o interesse pela criação da indústria de charques e o fornecimento de carne pra as regiões brasileiras, ultrapassando o século XIX com exportação para outros países. Segundo Gomes, (2007 apud SCHLESINGER), um dos maiores pecuaristas da região, foi José Antônio dos Anjos, que abatia 50.000 cabeças de gado por ano. O Rio Grande do Sul produzia trigo e gado, usado na fabricação de charque, mantas de couro, sebo e chifre.

A atividade bovina é muito importante e acompanha a história do Rio Grande do Sul por sua produção ser a segunda maior contribuinte no Valor Bruto da Produção Agropecuária do Estado. Portanto, será analisado o comportamento da série do preço médio mensal do quilograma do abate bovino recebido pelo produtor no período de janeiro de 2000 a dezembro de 2008 e fundamentada na análise desta série, realizar a previsão para o ano de 2009, através da combinação de previsão.

A previsão do preço recebido pelo produtor pelo abate bovino, através de combinação de previsão, se dará em cinco momentos:

- análise da série original;
- previsão através dos modelos ARIMA,
- previsão através dos modelos ARFIMA;
- previsão através dos modelos de Regressão Dinâmica;
- a proporção da variância total explicada por cada técnica de previsão
- Combinação das previsões.

4.1 Análise da Série Original

A análise da série original será utilizada para todas as técnicas de previsão por ser a única série em estudo, da qual será encontrada a previsão para o ano de 2009, por três diferentes técnicas de previsão e pela combinação dessas técnicas.

Analisando a série original, Figura 3, é possível observar que o preço recebido pelo produtor pelo abate bovino apresenta um comportamento tendencioso crescente ao longo dos meses. Através de uma análise visual identifica-se que a série é não estacionária e que é necessário realizar diferenças para estacionarizar a série.



Figura 3 - Série Original Série preço recebido pelo produtor pelo abate bovino de janeiro de 2000 a dezembro de 2008

Na Figura 4 encontra-se a função de autocorrelação da série original, é possível verificar que a não estacionaridade e, portanto o próximo procedimento é identificar o número de diferenças necessárias para a obter sua estacionaridade.

Observando ainda a Figura 4, a FACP mostra que a série começa a declinar em 1, pois apresenta apenas um lag fora dos limites de confiabilidade, indicando que uma única diferença é suficiente para estacionarizar a série.

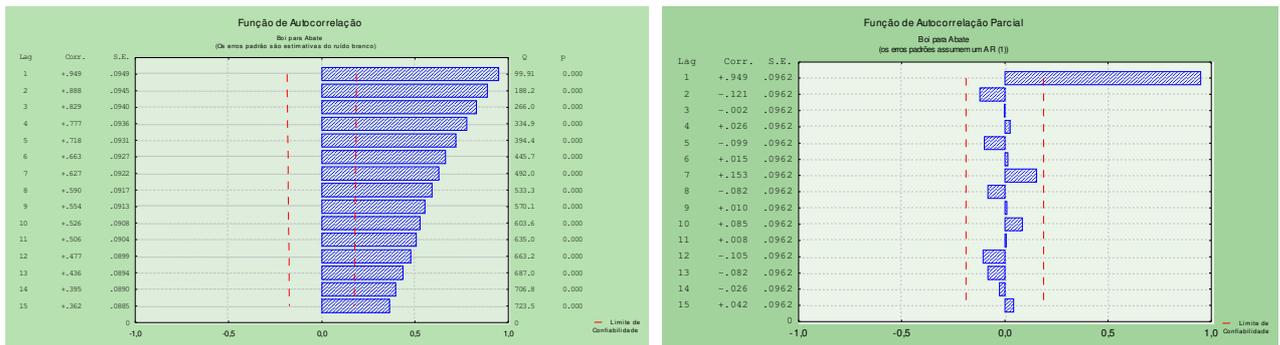


Figura 4 - Função de Autocorrelação e Função de Autocorrelação Parcial da Série preço recebido pelo produtor pelo abate bovino de janeiro de 2000 a dezembro de 2008

4.2 Modelos Auto-regressivos Integrados de Médias Móveis – ARIMA

Nos modelos ARIMA, necessitam que a série seja estacionária para assim poder encontrar modelos representativos da série. Normalmente utilizam-se até duas diferenças sucessivas e não mais que isso. Na série do preço recebido pelo produtor pelo abate bovino, na Figura 5, observa-se que uma única diferença foi suficiente para estacionarizar a série e assim encontrar os modelos representativos da série, não havendo a necessidade de realizar mais diferenças para não sofrer superdiferenciações.

Estacionarizada a série, próximo procedimento é encontrar os modelos significativos que a representam. A metodologia de modelos ARIMA encontrou sete modelos que representam a série, conforme Tabela 1. Observa-se que os sete modelos representativos apresentaram os critérios penalizadores semelhantes. Mas o mais significativo é o que apresentar os menores AIC e BIC, sendo, portanto o modelo $SARIMA(0,1,1)(0,1,0)_{12}$ por apresentar o AIC e BIC mais próximos a zero o modelo com características mais representativa da série original e ser o modelo mais parcimonioso.

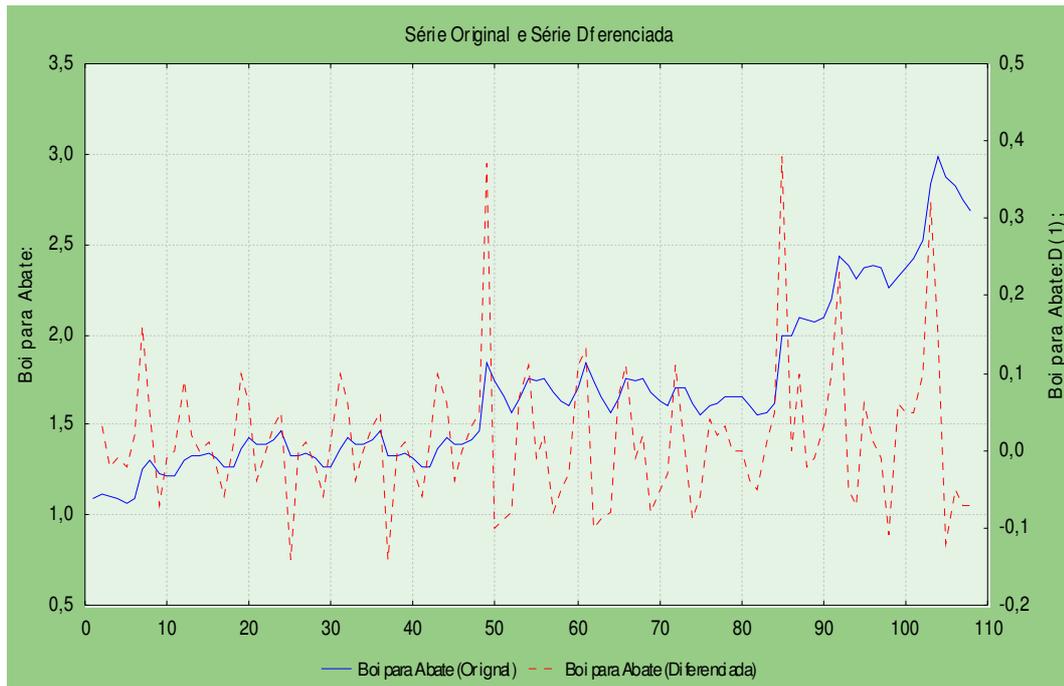


Figura 5 - Série Original e Série Diferenciada

Tabela 1 - Modelos ARIMA significativos para a Série do preço do recebido pelo produtor pelo abate bovino no Rio Grande do Sul

Modelos	Estimativas dos parâmetros do modelo	AIC	BIC
SARIMA (0,1,1)(1,0,1)₁₂	$\theta_1 = -0,2107$ $\Phi_1 = -0,4062$ $\Theta_1 = -0,9079$	-5,14	-5,07
<i>SARIMA (1,1,2)(0,1,0)₁₂</i>	$\theta_1 = 0,891$ $\theta_1 = 0,7824$ $\theta_2 = 0,2173$	-4,54	-4,46
SARIMA (2,1,1)(0,0,1) ₁₂	$\phi_1 = 0,83672$ $\phi_2 = -0,2248$ $\theta_1 = 0,70156$ $\Theta_1 = -0,6973$	-4,95	-4,86
SARIMA (0,1,1)(0,0,2) ₁₂	$\theta_1 = -0,2030$ $\Theta_1 = -0,5620$ $\Theta_2 = 0,22415$	-4,98	-4,91
SARIMA (2,1,1)(0,0,2) ₁₂	$\phi_1 = 0,78435$ $\phi_2 = -0,2800$ $\theta_1 = 0,61385$ $\Theta_1 = -0,5574$ $\Theta_2 = 0,25671$	-5,01	-4,88
SARIMA (1,1,1)(1,0,1) ₁₂	$\phi_1 = -0,9093$ $\theta_1 = -0,9955$ $\Phi_1 = -0,3388$ $\Theta_1 = -0,8888$	-4,99	-4,89

Modelos	Estimativas dos parâmetros do modelo	AIC	BIC
SARIMA (2,1,1)(1,0,1) ₁₂	$\phi_1 = 0,7786$ $\phi_2 = -0,2801$ $\theta_1 = 0,6003$ $\Phi_1 = 0,4549$ $\Theta_1 = -0,9280$	-5,16	-4,95

O próximo procedimento é analisar a FAC e FACP residuais, na Figura 6, na qual é possível verificar que o modelo ajustado apresenta todos seus *lag* dentro do limite de confiabilidade, demonstrando a adequação do modelo estimado.

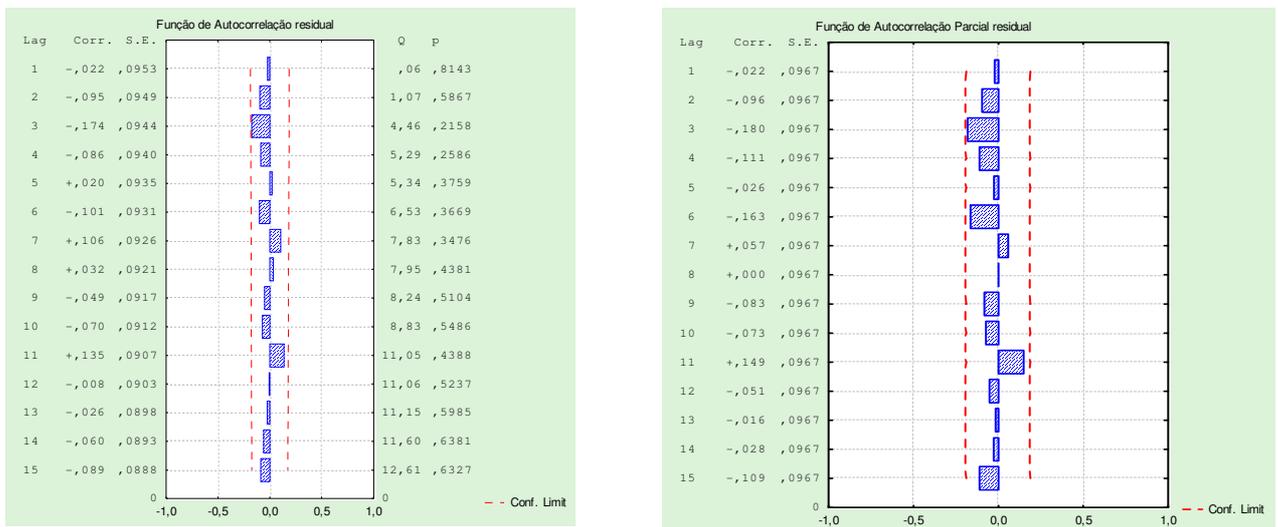


Figura 6 - Função de Autocorrelação e Função de Autocorrelação Parcial da Série preço recebido pelo produtor pelo abate bovino de janeiro de 2000 a dezembro de 2008

Após a análise da série original, realizada as diferenças necessárias, encontrado os modelos significativos e seus respectivos parâmetros, definido o modelo significativo representativo da série através dos critérios penalizadores AIC e BIC, analisado os resíduos identificando a adequação do modelo, chega-se a etapa final do ciclo iterativo: realizar as previsões.

Os valores resultante do modelo representativo encontra-se na Tabela 2, bem como a a representação gráfica encontra-se na Figura 7.

Tabela 2 - Previsão para o ano 2009, do preço recebido pelo produtor pelo abate bovino no Rio Grande do Sul

Meses	Previsões	Valores Reais	Inferior 95%	Superior 95%	Erro Padrão
Jan/09	2,63	2,67	2,41	2,73	0,08
Fev/09	2,60	2,66	2,24	2,74	0,13
Mar/09	2,59	2,62	2,13	2,76	0,16
Abr/09	2,56	2,54	2,13	2,87	0,19
Mai/09	2,55	2,50	2,13	2,97	0,21
Jun/09	2,60	2,54	2,13	3,06	0,23
Jul/09	2,72		2,22	3,22	0,25
Ago/09	2,69		2,15	3,22	0,27
Set/09	2,63		2,06	3,20	0,29
Out/09	2,62		2,01	3,22	0,30
Nov/09	2,59		1,92	3,18	0,32
Dez/07	2,57		1,87	3,20	0,33



Figura 7- Previsão para o ano de 2009 do preço recebido pelo produtor pelo abate bovino no Rio Grande do Sul

Logo, para os modelos ARIMA, o modelo representativo foi o **SARIMA** $(0,1,1)(0,1,0)_{12}$ o que melhor representa a série em estudo e oferece as melhores previsões para o preço recebido pelo produtor pelo abate bovino no Rio Grande do Sul para o ano de 2009.

4.3 Modelos Fracionários Auto-Regressivos Integrados de Médias Móveis – ARFIMA

Nos modelos ARIMA, encontrou-se mais de um modelo significativo, sendo necessária a utilização dos critérios penalizadores, AIC e BIC, para escolher o modelo mais significativo e mesmo assim, as FAC e FACP apresentaram um *lag* fora dos limites de confiabilidade. A Figura 3 mostra que a série do preço recebido pelo produtor pelo abate bovino é não estacionária, e a Figura 4, na Função de Autocorrelação da Série Original, vem confirmar esta não estacionariedade. A Figura 5, mostra que apenas uma diferença foi suficiente para a série diferenciada assumir um comportamento estacionário, mas devido a FAC e a FACP apresentarem um *lag* fora dos limites de confiabilidade, talvez uma diferença possa ter tornado a série super diferenciada, apesar de o modelo encontrado ser o mais significativo para a série.

Como o foco principal desta pesquisa é conseguir encontrar uma combinação de previsões que expliquem ao máximo os componente da série em estudo é que o próximo procedimento será estimar as diferenças fracionárias, através dos modelos ARFIMA. Como a série original é analisada através de três técnicas, não convém relembrar o comportamento da série original em cada técnica analisada para não tornar a análise repetitiva.

Já observou-se que a série é não estacionária e aplicou-se uma diferenciação com $d \in \mathbb{Z}_+$, nos modelos ARIMA, agora será aplicada uma diferenciação fracionária com $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ e encontrar os modelos representativos desta técnica. Os modelos significativos encontram-se na Tabela 4, na qual encontram-se quatro modelos representativos e será o critério penalizador Critério de Informação de Akaike – AIC, que definirá o melhor modelo. Analisando por este critério, o melhor modelo é o que apresentar o menor AIC. Como o menor AIC é -1,8389 logo o modelo que representa a série em análise, nesta metodologia é dado por ARFIMA $(1,d,0)$, com uma diferença $d=0,492869 < 0,5$ e $|\phi_1| = 0,766281 < 1$, apresentando um ruído fracionário.

Tabela 3 - Modelos ARFIMA significativos para a Série do preço do recebido pelo produtor pelo abate bovino no Rio Grande do Sul

Modelo	Parâmetros	Erro Padrão	t-valor	t-prob	AIC
ARFIMA (1,d,0)	d=0,492869 $\phi_1=0,766281$	0,0109 0,0673	45,2 11,4	0,000 0,000	-1,8389
ARFIMA (1,d,1)	d=0,493115 $\phi_1=0,756717$ $\theta_1=0,0220227$	0,01062 0,09664 0,1553	46,4 7,83 0,142	0,000 0,000 0,887	-1,8206
ARFIMA (0,d,1)	d=0,498858 $\theta_1=0,630725$	0,001601 0,06380	312,0 9,89	0,000 0,000	-1,5407
ARFIMA (2,d,1)	d=0,493123 $\phi_1=0,316631$ $\phi_2=0,332073$ $\theta_1=0,472233$	0,01052 1,266 0,9804 1,256	46,9 0,250 0,339 0,376	0,000 0,803 0,736 0,708	-1,8030

Encontrado o melhor modelo, analisam-se os resíduos através da Função de Autocorrelação e Autocorrelação Parcial, na Figura 8. Observa-se que os resíduos encontram-se entre os limites de confiabilidade, refletindo na adequação do modelo.

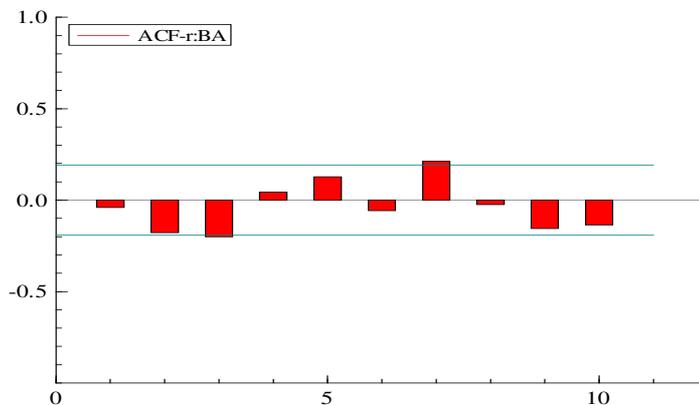


Figura 8 - Função Autocorrelação Parcial dos Resíduos

Encontrado o melhor modelo e verificada a sua adequação, pode-se realizar a previsões para o ano de 2009, para preço recebido pelo produtor pelo abate bovino no Estado. Na Tabela 4, encontra-se os valores das previsões obtidas pelo modelo ARFIMA e na Figura 9, encontra a representação gráfica das previsões. Logo esta metodologia é eficaz para encontrar modelos de previsão para o preço do abate bovino no Estado.

Tabela 4 - Previsão para o preço recebido pelo produtor pelo abate bovino no Rio Grande do Sul, através de Modelos ARFIMA

Meses	Previsões	Valores Reais	Inferior 95%	Superior 95%	Erro Padrão
Jan/09	2,62	2,67	2,41	2,70	0,06
Fev/09	2,57	2,66	2,24	2,71	0,10
Mar/09	2,52	2,62	2,13	2,76	0,15
Abr/09	2,48	2,54	2,13	2,80	0,09
Mai/09	2,44	2,50	2,13	2,91	0,10
Jun/09	2,4	2,54	2,13	2,99	0,23
Jul/09	2,37		2,11	3,12	0,21
Ago/09	2,34		2,10	3,02	0,18
Set/09	2,31		2,04	2,90	0,21
Out/09	2,29		2,01	2,87	0,22
Nov/09	2,27		1,95	2,76	0,24
Dez/07	2,25		1,82	2,55	0,23

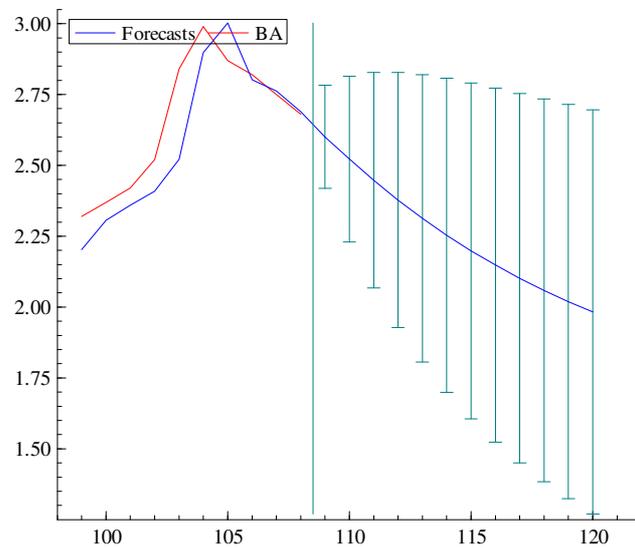


Figura 9 - Previsão para o ano de 2009 do preço recebido pelo produtor pelo abate bovino no Rio Grande do Sul

4.4 Regressão Dinâmica

Com o intuito de desenvolver uma previsão com resultados bem acurados a Regressão Dinâmica também tem muito a contribuir, pois ela leva em consideração valores defasados, tanto da variável dependente, quanto das variáveis explicativas. Essa metodologia mostra que o comportamento da série do preço recebido pelo produtor pelo abate bovino é influenciado pelo preço de outras atividades pecuárias também.

Nesta análise consideraram-se as variáveis com três defasagens e buscou-se explicar o preço recebido pelo produtor, explicado pelas variáveis:

Dependente: o preço do quilograma recebido pelo produtor gaúcho pelo boi para o abate (BA);

Independentes: o preço recebido pelo produtor pelo preço da vaca para abate (VA); touro de 3 anos (T3); Novilho 4/+ anos (N4); Novilho 3 anos (N3); Novilho 2 anos (N2); Novilho 1 anos (N1); Vaca Solteira (VS); Vaca com cria ao pé (VCP); Vaca de Invernar (VI); Vaquilhona de 3 anos (VQ3); Vaquilhona de 2 anos (VQ2); Vaquilhona de 1 ano (VQ1); Terneiro(a) (T) e Vaca Leiteira Comum (VLC). Optou-se por incluir estas variáveis na análise, por pertencerem a pecuária e estarem disponíveis pela Fundação de Economia e Estatística.

O modelo inicial da regressão dinâmica começou a ser rodado com 60 parâmetros, resultado de 15 variáveis com 3 defasagem mensais em cada variável, a fim de captar os efeitos provenientes da combinação de dados defasados no tempo. O modelo adequado foi encontrado através da eliminação dos parâmetros não significativos em ordem decrescente, com $p < \alpha$. Após a eliminação das variáveis estatisticamente não-significativas, o modelo reduziu-se a 17 parâmetros

Logo, o modelo significativo resultante é:

$$\begin{aligned}
BA_t = & - 0.935130 BA - 0.469415BA_{t-2} + 1.12761VA + 1.00052VA_{t-1} + 0.522166 VA_{t-2} + \\
& 9.65608e^{-005} N3_{t-1} + 9.55365e^{-005} N3_{t-2} + 0.000100255 N3_{t-3} + 0.000385023NI_{t-2} - \\
& 0.00104321VS_{t-1} - 0.000697758 VS_{t-2} - 0.000116299VCP_{t-3} + 0.000573516 VI_{t-1} - \\
& 0.000607678 VQ2 + 0.00110725 VQ1 + 0.000364047 T_{t-2} + 0.000181362 VLC_{t-1}
\end{aligned}$$

Os testes de especificação do modelo são apresentados pelo próprio programa, *Ewies*. Em cada modelo gerado, ao se eliminar as variáveis explicativas não significativas, o programa fornece os testes de especificação do modelo. As especificações foram aceitas conjuntamente quando o modelo gerado apresentou todas as variáveis significativas, ou seja, $p < \alpha$, que está expresso pelo modelo descrito acima.

Tabela 5 – Especificidades do modelo encontrado através da Regressão Dinâmica

Teste	Finalidade	p – valor
AR	Avaliar a existência de autocorrelação entre os resíduos	0.3980
ARCH	Heterocedasticidade dos resíduos	0.8705
Normalidade	Os dados seguem uma distribuição normal	0.7042
Heterocedasticidade	Variância encontrada ligada aos regressores	0,3744
Reset	Especificação na equação do modelo	0.0869

No modelo, o nível de significância adotado para análise dos testes é de 5%. A análise da especificidade, dá-se através dos valores para cada teste encontrados na Tabela 5. O teste AR, o qual testa a existência de autocorrelação serial entre os resíduos atuais e defasados. Como o teste apresentou um p-valor de 0.3980, conclui-se portanto que os resíduos não são autocorrelacionados.

Para testar a heterocedasticidade do modelo auto-regressivo condicionado, ou seja, se o ruído branco apresenta variância constante, ou se os erros são homocedásticos, será analisado o teste ARCH. Com um p-valor de 0.8705 o modelo mostra a não heterocedasticidade dos resíduos, ou seja, os resíduos não apresentam variabilidade na variância.

O teste da normalidade dos resíduos apresentou um p - valor de 0.7042, mostrando que os resíduos apresenta-se normalmente distribuídos, ou seja, os dados são normalmente distribuídos. Avaliando a heterocedasticidade, com p-valor de 0,3744, mostra que a variância

encontrada está ligada aos regressores. O Software ainda apresenta o teste *Reset*, com p-valor de 0.0869 afirma que o modelo não apresenta problemas na sua fórmula funcional, ou seja, não apresenta problema de má especificação na equação do modelo e, portanto as previsões podem ser utilizadas.

Atendida as condições de especificidade do modelo, pode-se utilizar o modelo para realizar previsões, as quais encontram-se na Tabela 6 e Figura 10.

Tabela 6 - Previsão para o preço recebido pelo produtor pelo abate bovino no Rio Grande do Sul, através de Modelos de Regressão Dinâmica

Meses	Previsões	Valores Reais	Inferior 95%	Superior 95%	Erro Padrão
Jan/09	2,67	2,67	2,52	2,82	0,02
Fev/09	2,62	2,66	2,50	2,74	0,04
Mar/09	2,68	2,62	2,51	2,84	0,09
Abr/09	2,67	2,54	2,52	2,87	0,13
Mai/09	2,64	2,50	2,58	2,86	0,16
Jun/09	2,53	2,54	2,43	2,63	0,02
Jul/09	2,87		2,69	3,05	0,20
Ago/09	2,98		2,81	3,15	0,18
Set/09	2,92		2,76	3,08	0,17
Out/09	2,85		2,69	3,01	0,19
Nov/09	2,78		2,65	2,91	0,22
Dez/07	2,75		2,62	2,88	0,23

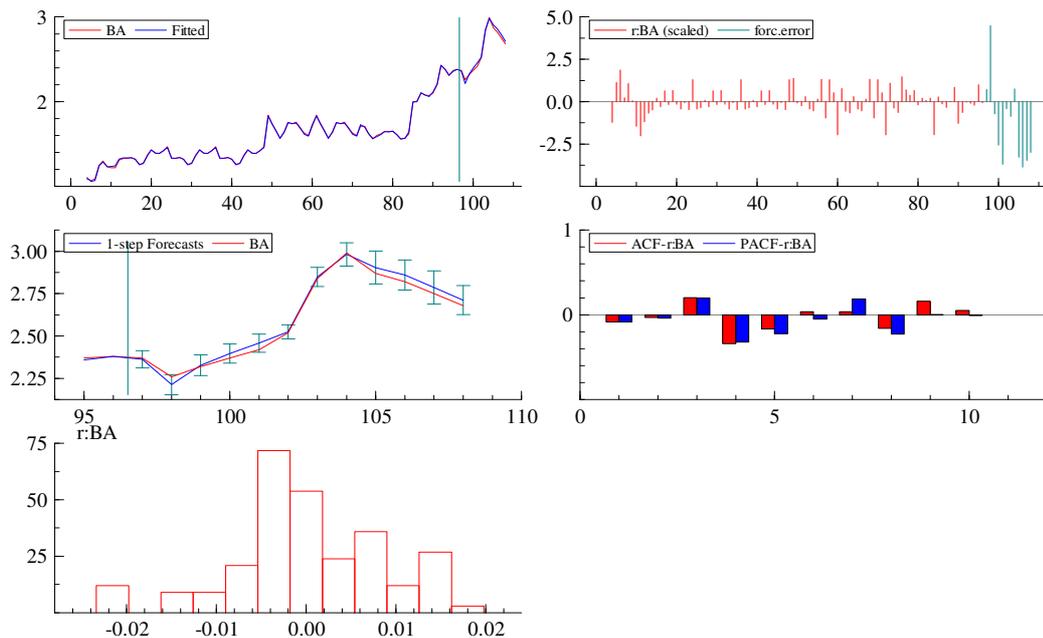


Figura 10 - Previsão para o ano de 2009 do preço recebido pelo produtor pelo abate bovino no Rio Grande do Sul, através de Regressão Dinâmica.

Na Figura 10, observa-se a Série Original, a FAC e FACP, FAC e FACP residual e a previsão da série em pesquisa proporcionada pelo modelo encontrado pela metodologia.

4.5 Combinação de Previsões

A combinação de previsão também é técnica para encontrar valores futuros de uma série temporal. Com a combinação de previsão deseja-se obter resultados mais satisfatórios que as previsões analisadas separadamente por cada técnica. Isso, porque na combinação de previsões, como o próprio nome já diz, combina as melhores previsões de cada técnica de previsão. A combinação das previsões se dá através do percentual da variância explicada de cada previsão, através de componentes principais.

Como ainda não existem estudos que comprovem quais as melhores previsões a se combinar, optou-se por previsões da metodologia Box&Jenkins, ARIMA e ARFIMA e por Regressão Dinâmica.

As previsões de cada técnica encontram-se na Tabela 8, podendo se comparar que não apresentam os mesmos valores futuros, apresentando certa discrepância. Observando a previsão fornecida pelo modelo ARIMA e pela Regressão Dinâmica, ambos apresentam um comportamento semelhante e que os mesmos apresentam inicialmente uma tendência decrescente e em seguida crescente, enquanto o modelo ARFIMA já apresenta comportamento decrescente em toda sua previsão.

Analisando as medidas descritivas, observa-se que a média do preço encontrada em cada modelo é representativa da previsão do preço recebido pelo produtor no ano de 2009, pois o ARIMA apresentou coeficiente de variação de 2%, o ARFIMA de 5% e a Regressão Dinâmica de 5%, ou seja, a média é representativa de cada modelo, pois a variabilidade encontrada é pequena. Embora a média as médias de cada previsão sejam diferentes, cabe ressaltar que a média da combinação de previsões (R\$2,59) foi igual à média dos valores reais (R\$2,59) e com desvio padrão menor. Portanto, em uma análise prévia, mas não conclusiva, pode-se observar que a combinação das previsões, através da variância explicada, apresentou mais condizentes com a realidade.

Devido à capacidade que cada modelo apresenta de captar informações características e específicas é que deseja-se realizar a combinação de previsões, buscando unir a contribuição que cada modelo é capaz de oferecer. A combinação da previsão se dará através da equação 102.

Encontradas as previsões, deve-se encontrar o quanto cada previsão contribui para a combinação de previsão. Essa contribuição se dará através da análise de componentes principais, onde cada previsão representa uma componente e deseja-se encontrar o seu percentual de explicação. Esses percentuais são os autovalores ou a proporção da variância total explicada por cada modelo de previsão, que irão compor a Combinação das Previsões.

Na Tabela 7 e na Figura 11 encontram-se respectivamente os autovalores e proporção da variância total explicada por cada modelo de previsão, respectivamente.

Tabela 7 - Tabela dos autovalores e da proporção da variância total explicada

	Autovalores	% variância total explicada	Cumulativo Autovalores	Cumulativo do percentual
ARIMA	2,139670	71,32232	2,139670	71,3223
ARFIMA	0,558096	18,60319	2,697766	89,9255
Regressão Dinâmica	0,302234	10,07448	3,000000	100,0000

Observa-se na Tabela 7 o percentual da variância explicada pelas previsões obtidas em cada técnica analisada, e a combinação de previsão se dará pelo por esse percentual, ou seja, a combinação das previsões se dará da seguinte maneira:

- 71,32% da combinação previsão é representada pelo modelo ARIMA;
- 18,60% da combinação previsão é representada pelo modelo ARFIMA, e
- 10,07% da combinação previsão é representada pelo modelo de Regressão Dinâmica.

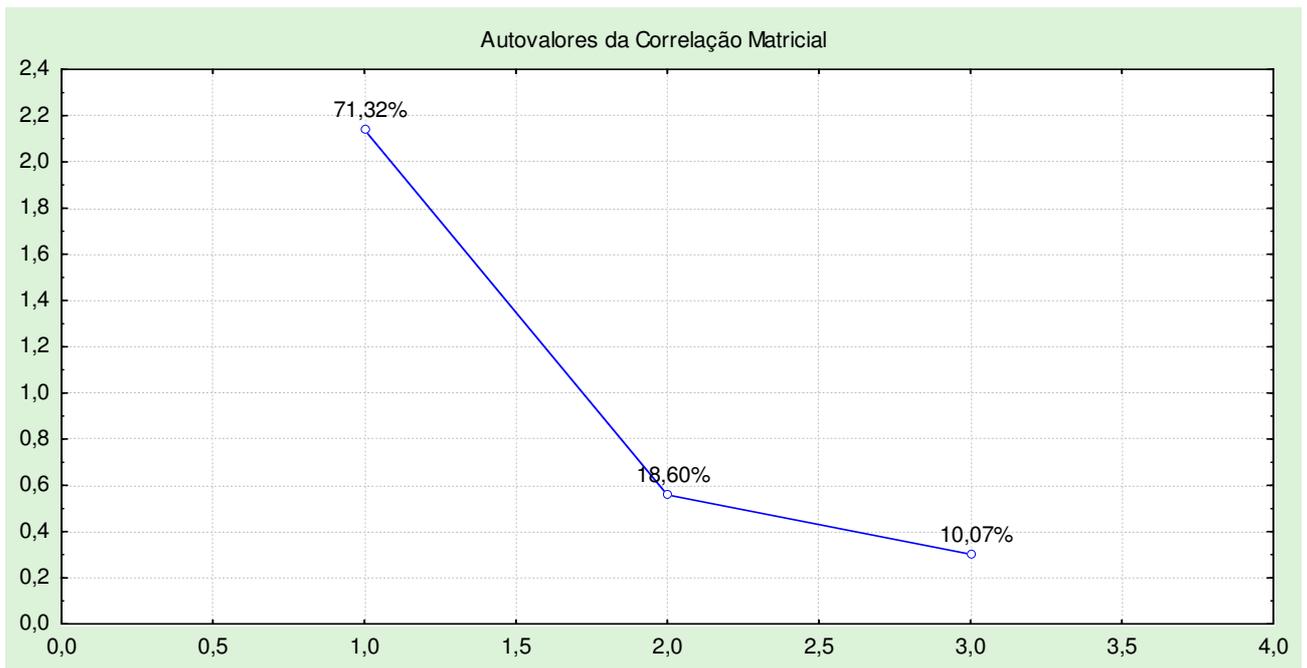


Figura 11 - Proporção da variância total explicada por cada autovalor

Encontrados os autovalores e, conseqüentemente, a proporção total da variância explicada por cada autovalor, pode-se encontrar a Combinação de previsão, através do IPACP:

$$\text{IPACP} = \sum_{i=1}^n \left[\frac{\lambda_i}{\sum_{i=1}^n \lambda_i} (P_i) \right]$$

Portanto, a equação do IPACP é o somatório da proporção total da variância explicada por cada técnica de previsão multiplicada por suas respectivas previsões. Então, tem-se que:

- Para obter o primeiro valor da combinação de previsão:

$$i = 1 \rightarrow \text{IPACP}_1 = \frac{2,14}{3} * 2,62 + \frac{0,56}{3} * 2,62 + \frac{0,30}{3} * 2,67 = 2,63$$

- Para obter o segundo valor da combinação de previsão:

$$i = 2 \rightarrow \text{IPACP}_2 = \frac{2,14}{3} * 2,60 + \frac{0,56}{3} * 2,57 + \frac{0,30}{3} * 2,62 = 2,60$$

⋮

- E assim sucessivamente até a décima segunda previsão:

- $i = 12 \rightarrow \text{IPACP}_{12} = \frac{2,14}{3} * 2,57 + \frac{0,56}{3} * 2,25 + \frac{0,30}{3} * 2,75 = 2,55$

Tabela 8 - Previsão para o preço recebido pelo produtor pelo abate bovino no Rio Grande do Sul, através de Modelos de Regressão Dinâmica

Meses	Previsão				Valor Real	Erro			
	ARIMA	ARFIMA	Regressão Dinâmica	Combinação de Previsão		ARIMA	ARFIMA	Regressão Dinâmica	Com. Prev
Jan/09	2,63	2,62	2,67	2,63	2,67	0,04	0,05	0,00	0,04
Fev/09	2,60	2,57	2,62	2,60	2,66	0,06	0,09	0,04	0,06
Mar/09	2,59	2,52	2,68	2,59	2,62	0,03	0,10	-0,06	0,03
Abr/09	2,56	2,48	2,67	2,56	2,54	-0,02	0,06	-0,13	-0,02
Mai/09	2,55	2,44	2,64	2,54	2,50	-0,05	0,06	-0,14	-0,04
Jun/09	2,60	2,4	2,53	2,56	2,54	-0,06	0,14	0,01	-0,02
Jul/09	2,72	2,37	2,87	2,67		Preço do recebido pelo produtor pelo abate bovino de jan/2000 a dez/2008			
Ago/09	2,69	2,34	2,98	2,66					
Set/09	2,63	2,31	2,92	2,60					
Out/09	2,62	2,29	2,85	2,58					
Nov/09	2,59	2,27	2,78	2,55					
Dez/09	2,57	2,25	2,75	2,53					
Média	2,61	2,41	2,75	2,59					
Desvio Padrão	0,05	0,12	0,14	0,05	0,071	0,45			
Coef. de Variação	2%	5%	5%	2%	3%	27%			
MAPE	0,0167	0,0335	0,0239	0,0135					
EQM	0,0021	0,0079	0,0070	0,0014					

Na Tabela 8, encontra-se a previsão do modelo ARIMA, ARFIMA, Regressão Dinâmica, a Combinação das Previsões e os valores reais do preço recebido pelo produtor de janeiro a junho de 2009. Observando os valores, é possível identificar que cada modelo tem

informações que lhe é particular. O objetivo de realizar a combinação de previsão foi justamente combinar as informações fornecidas por cada previsão. Por tal motivo a utilização do percentual da variância de cada modelo.

Como visualmente não pode-se afirmar que a combinação de previsão realmente apresenta melhores resultados, utilizou-se do MAPE, que é erro percentual absoluto médio, e do EQM, Erro Quadrado Médio, com o intuito de verificar qual previsão mais se aproxima da realidade.

Analisando a Tabela 8, constata-se que a Combinação das previsões apresentaram os valores do MAPE = 0,0135 e Erro Quadrado Médio = 0,0014, valores esses menores que as técnicas de previsão individuais, ainda sendo importante enfatiza que a média da combinação de previsões e dos valores reais são iguais. Portanto, a técnica de combinação de previsão de modelos ARIMA, ARFIMA e Regressão Dinâmica apresentaram valores mais eficiente, semelhantes a realidade. Na Figura 12, encontram-se todas as previsões e a Combinação de Previsão para o preço médio mensal recebido pelo produtor pelo abate bovino.

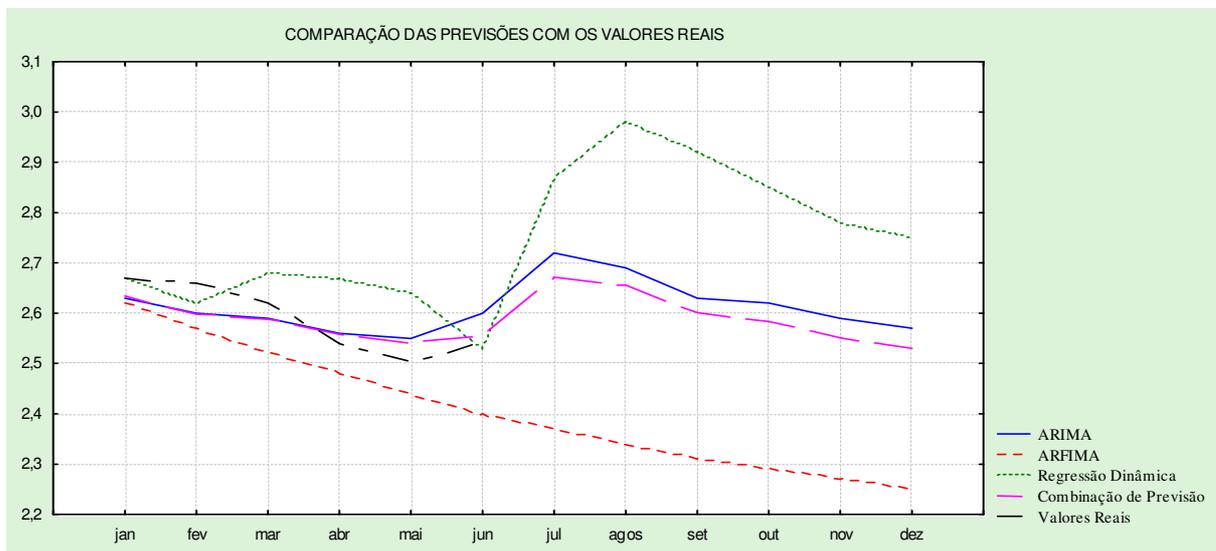


Figura 12 – Gráfico das previsões e da Combinação de Previsão

5 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES

Desenvolveu-se esta pesquisa com o objetivo de analisar o comportamento da Série do preço médio do quilograma recebido pelo produtor pelo abate bovino, através de técnicas consagradas pela literatura, com o intuito de se obter resultados mais precisos que os métodos individuais. Essa técnica é a combinação das previsões, com a qual se buscou contemplar informações e características fornecidas por todos os modelos analisados e não considerar em específico.

A Combinação de Previsão, como o próprio nome já diz, é a combinação das previsões resultante de técnicas diferentes, através do percentual da variância explicada, obtida através de Componentes Principais, onde cada previsão representa uma componente principal. Para chegar a essa combinação, desenvolveu-se o Índice de Ponderação de Autovalores para Combinação de Previsões - IPACP, o qual é composto pela proporção da variância total explicada por cada previsão.

As previsões individuais desenvolvidas para compor o IPACP, são provenientes da Metodologia Box&Jenkins, modelos ARIMA e ARFIMA, e Regressão Dinâmica. A seguir encontram-se os modelos significativos para cada técnica de previsão:

- **ARIMA:** resultou em modelo *SARIMA* $(0,1,)(1,0,1)_{12}$;
- **ARFIMA:** $(1,d,0)$, com uma diferença $d=0,492869 < 0,5$ e $|\phi_1| = 0,766281 < 1$, apresentando um ruído fracionário;
- **Regressão Dinâmica:** que inicialmente tinha 60 variáveis e com a eliminação das variáveis não significativas, resultou em um modelo com 17 parâmetros.

$$BA_t = - 0.935130 BA - 0.469415BA_{t-2} + 1.12761VA + 1.00052VA_{t-1} + 0.522166 VA_{t-2} + 9.65608e^{-005} N3_{t-1} + 9.55365e^{-005} N3_{t-2} + 0.000100255 N3_{t-3} + 0.000385023NI_{t-2} - 0.00104321VS_{t-1} - 0.000697758 VS_{t-2} - 0.000116299VCP_{t-3} + 0.000573516 VI_{t-1} - 0.000607678 VQ2 + 0.00110725 VQ1 + 0.000364047 T_{t-2} + 0.000181362 VLC_{t-1}$$

Através de Componentes Principais foi possível encontrar a variância explicada de cada modelo, sendo o ARIMA 71,32%; o ARFIMA 18,60 e a Regressão Dinâmica 10,07%. Esses três modelos combinados apresentaram previsões melhores que as previsões específicas

de cada modelo. Como a buscou-se combinar previsões e mostrar que a combinação apresenta previsões mais eficientes utilizou-se o MAPE – Erro Percentual Absoluto Médio e o EQM - Erro Quadrado Médio. O MAPE encontrado para o modelo ARIMA foi 1,67%, ARFIMA – 3,35%; Regressão Dinâmica – 2,39% a Combinação das previsões foi de 1,35%, mostrando ser a combinação de previsões uma técnica mais precisa nas previsões, o que também é comprovado através do EQM, mostrou a Combinação com resultados mais condizentes com a realidade. No modelo ARIMA o EQM encontrado foi 0,0021; no ARFIMA foi 0,0079; Regressão Dinâmica 0,0070 e a Combinação de Previsão 0,0014.

Portanto a Combinação de Previsão, combinando modelos ARIMA, ARFIMA e Regressão Dinâmica apresentam previsões melhores para o preço médio por quilograma recebido pelo produtor pelo abate bovino no Rio Grande do Sul. Por ter apresentado valores condizentes com a realidade, é uma técnica que precisa ser sugerida aos tomadores de previsões para que transfiram esses conhecimentos aos produtores.

Conclui-se esta pesquisa, com objetivo atendido, conhecimentos e resultados satisfatórios, mas novas idéias acompanharam a pesquisa e que merece o prosseguimento destes estudos. Sugere-se para estudos futuros outras combinações de previsão, incluindo outros métodos de previsão e fazer a comparação de qual combinação resulta em previsões mais satisfatórias, visto que ainda não se tem um estudo que comprove qual é a melhor combinação existente para a realização de previsões. É uma proposta muito importante, visto que ainda não existe um estudo que realizou todas as possíveis combinações e concluiu qual a melhor. Com certeza este novo estudo vem a contribuir com muitas pesquisas, as quais apresentarão erros de previsão cada vez melhores e mais representativos da realidade. Outra sugestão refere-se a combinar previsões dos modelos representativos em cada técnica, como por exemplo, foram encontrados 7 modelos significativos dos modelos ARIMA, desenvolver a combinação de previsões desses modelos, através de componentes principais.

Espera-se que a idealização de realizar as n possíveis combinações e apresentar qual é o melhor método não seja ilusório e que possa ser aprimorado em pesquisas de um futuro bem próximo, não somente pelos conhecimentos científicos, mas pelos benefícios que pode trazer a sociedade em geral.

6 REFERÊNCIAS

Atlas Sócio Econômico do Rio Grande do Sul. Disponível em <http://www.seplag.rs.gov.br/atlas/default.asp> Acesso em: 25 jul. 2009.

BARBETTA, P., A. **Estatística Aplicada às Ciências Sociais.** 5 ed. Florianópolis: Editora da UFSC, 2004.

BARROS, M. Consultoria Ltda: **Capítulo 5: Introdução às Séries Temporais e aos Modelos ARIMA, 2006.** Disponível em: www.mbarros.com/sitebuildercontent/sitebuilderfiles/LivroProcessosCap5parte.pdf. Acesso em: 30 out. 2006.

BOX, G. E. P., JENKINS G. M., REINSEL G. C. **Time Series Analysis: Forecasting and Control.** Editora: Wiley. Pg 768. 4ª ed, 2008.

CENSO AGROPECUÁRIO 2006. **Instituto Brasileiro de Geografia e Estatística – IBGE.** Disponível em: <http://www.sidra.ibge.gov.br> Acesso em: 11 ago. 2009.

CLEMEN, R.T. **Combining Forecasts: A Review and Annotated Bibliography.** International Journal of Forecasting. V.5, 1989

FARIAS, E. R.; ROCHA, F. J. S; LIMA, R. C. **Critérios de seleção de modelos sazonais de séries temporais: uma aplicação usando a taxa de desemprego da região metropolitana de Recife.** III Encontro Regional de Estudos do Trabalho – ABET, 22 a 24 de novembro de 2000 – Recife, PE. Disponível em: <http://www.race.nuca.ie.ufrj.br/abet/3reg/39.DOC>. Acesso em: 20 nov. 2005.

FEE – **Fundação de Economia e Estatística.** Disponível em <http://www.fee.rs.gov.br>. Acesso em: 10 mai. 2009.

GRANGER, C. W. J. e Joyeux, R. (1980). **An introduction to long-memory time series models and fractional differencing.** Journal of Time Series Analysis, 1, 15-29.

GUJARATI, D.N. **Econometria Básica.** 3 ed. São Paulo: MAKRON Books, 2000.

HOSKING, J. (1982). **Some models of persistence in time series analysis: theory and practice 1.** North Holland Publishing Company, 641-653.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. Ed. Prentice Hall, 800 pg., 6ª, 2007.

LIBBLY, R. e BLASHFIELD, R.K **Performance of a Composite as a Function of Number of Judges**. *Organizational Behavior and human Performance*, v 21, 1978.

MAKRIDAKIS, S. **Why Combining Woks?** *International Journal of forecasting*. v.5,1989.

MAKRIDAKIS, S.G.; WINKLER, R.L. **Averages of Forecasts: Some Empirical Results**. *Maganament Science*. v.29,n.9, 1983.

MALHORTA, N. K. **Pesquisa de Marketing: uma orientação aplicada. 3 ed. Bookmann**. Porto Alegre. 2001.

MONTOYA, M.A.; PARRÉ, J.L. **O agronegócio brasileiro no final do século XX: estrutura produtiva, arquitetura organizacional e tendências**. 2v. Passo Fundo: UPF, 2000. 337p.

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

MORETTIN, P. A. & TOLOI, C. M. C., **Modelos de Função de Transferência**. Rio de Janeiro, 3ª Escola e séries Temporais Econometria, 1989.

MORETTIN, P. A.; TOLOI, C. M. **Análise de Séries Temporais**. 1ª ed. São Paulo: E. Blücher, 2004.

PAIVA, A.P. **Metodologia de Superfície de Resposta Multivariada: uma Proposta de Otimização para Processos de Soldagem com Múltiplas Respostas Correlacionadas Baseada na Análise de Componentes Principais**. Tese de Doutorado do Instituto de Engenharia Mecânica da Universidade Federal de Itajubá, 2006.

RAUSSER, G.C.; OLIVEIRA, R.A. **An Econometric Analysis of Wilerness Area Use**. *Journal of the American Statical Association*, v.71, n.354, 1976.

REGAZZI, A. J. **Análise multivariada**. Viçosa: Universidade Federal de Viçosa, Centro de Ciências Exatas e Tecnológicas, Departamento de Informática, 2001.

REISEN, V. A. **ARFIMA – O modelo ARIMA para o d fracionário. 6ª Escola de Séries Temporais.** 1995. Vitória, ES.

SCHLESINGER, S. **O gado bovino no Brasil.** Disponível em <http://www.boell-latinoamerica.org/downloads/Texto_Gado_Boll_2009-4.pdf> Acesso em: 23 jul. 2009.

SILVA, C.F; ALVES, T.W. **Perfil da Pecuária de Corte do Rio Grande do Sul.** Disponível em <www.apec.unesc.net/III%20EEC> Acesso em: 01 ago. 2009.

SOUZA, F. M. **Modelos Box & Jenkins Aplicados a Demanda de Leitos Hospitalares** (Especialização em Estatística e Modelagem Quantitativa) – Universidade Federal de Santa Maria, Santa Maria, 2006.

SOUZA, G.P. **Previsão do Consumo Industrial de Energia Elétrica no Estado de Santa Catarina: uma Aplicação da Combinação de Previsões entre Modelos Univariados e de Regressão Dinâmica.** Dissertação de Mestrado, Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina. 2005

SOUZA, R. C.; CAMARGO, M. E. **Análise e previsão de séries temporais: Os modelos ARIMA.** Ijuí, RS: Sedigraf, 1996.

TELLECHEA, F. **Análise dos custos de transação no setor industrial da cadeia produtiva de carne bovina no Rio Grande do Sul.** Curso de Pós-Graduação em economia rural da Universidade Federal do Rio Grande do Sul, 2001.

VASCONCELOS, M. A. S.; ALVES, D. **Manual de Econometria.** São Paulo: Atlas, 2000.

VIANA, J.G.A.; SILVEIRA, V.C.P.; OLIVEIRA, C.B **A relação entre o preço pago pelo consumidor de carne bovina em Santa Maria e o recebido pelo produtor de gado de corte no Rio Grande do Sul, Brasil** - Revista Ciência Rural, 2008 - ISSN 0103-8478 - doi: 10.1590/S0103-84782007000400033 Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782007000400033&lng=e&nrm=iso&tlng=e> Acesso em: 30 jul. 2009.

WERNER, L. **Um modelo Composto para realizar previsão de demanda através da integração da Combinação de previsões e do ajuste baseado na opinião.** Tese de Doutorado do Curso de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul, 2004.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)