

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS (CEFET-MG)

KLEBER SILVA LEITE PINTO

**MONITORAMENTO DE CUSTOS COM USO DE MODELAGEM DE
SÉRIES TEMPORAIS E ALGORITMOS**

Belo Horizonte

2010

Livros Grátis

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

Milhares de livros grátis para download.

KLEBER SILVA LEITE PINTO

**MONITORAMENTO DE CUSTOS COM USO DE MODELAGEM DE
SÉRIES TEMPORAIS E ALGORITMOS**

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Modelagem

Linha de Pesquisa: Modelagem, Aperfeiçoamento e Otimização de Processos

Orientadores: Elenice Biazzi e João Francisco de Almeida Vitor

BELO HORIZONTE

2010

Ficha catalográfica: elaborada pela biblioteca da

Será impressa no verso da folha de rosto e não deverá ser contada.

Dissertação intitulada MONITORAMENTO DE CUSTOS COM USO DE MODELAGEM DE SÉRIES TEMPORAIS E ALGORITMOS, de autoria de Kleber Silva Leite Pinto, aprovada pela banca examinadora constituída pelos seguintes professores:

Prof. Dra. Elenice Biazzi (Orientadora)
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. João Francisco de Almeida Vitor (Co-orientador)
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. José Antônio de Sousa Neto
Fundação Dom Cabral

Prof. Dr. Rogério Martins Gomes
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Sérgio Ricardo de Souza
Coordenador do Curso de Mestrado em Modelagem Matemática e Computacional

Data de aprovação: Belo Horizonte, 26 de agosto de 2010.

Coordenação do Programa de Pós-Graduação em Modelagem Matemática e Computacional
Av. Amazonas 7675 - Nova Gameleira - Belo Horizonte - MG - Brasil

DEDICATÓRIA

À minha neta Maria Clara.

AGRADECIMENTOS

Aos professores e alunos do CEFET-MG pelo aprendizado proporcionado neste convívio.

“Não convém a gente levantar escândalo de
começo, só aos poucos é que o escuro é
claro”.

Guimarães Rosa - Grande Sertão Veredas

RESUMO

Este trabalho é o resultado da busca de uma ferramenta de apoio ao controle financeiro para o monitoramento dos custos com mão de obra em instituição dedicada à prestação de serviços, que tem nesse insumo o seu principal item de custo de produção. Essa ferramenta deve propiciar uma identificação dos componentes de custo que são considerados fora de controle, quando estiverem incoerentes com os demais componentes (verbas), ou com a sua própria série histórica. A discrepância entre o custo real apurado e a avaliação das horas consumidas, nas diversas seções administrativas, é inconveniente, uma vez que compromete a eficiência empresarial. Descontinuidades e variações nos custos devem ser prontamente identificadas, em tempo de serem corrigidas ou replanejadas. Todos os custos incorridos são apurados no ciclo contábil mensal e, após consolidados, devem ser transferidos para os projetos que os consumiram. Para isto, utiliza-se o processo de preço de transferência com base no custo padrão. Cada um dos custos é contabilizado segundo a sua natureza, e o valor resultante da soma dos lançamentos mensalmente apurados constitui cada um dos registros das séries temporais em monitoramento. Foram tratadas um total de 6(seis) séries num período de 4 anos, compreendido entre 2006 e 2009, em que se buscou identificar padrões de comportamento que possibilitassem a construção de um modelo de previsão com o objetivo de, no final do período de apuração, permitir o julgamento do último valor contabilizado, isto é, avaliar se tal valor deve ser aceito ou rejeitado. Na hipótese de rejeição, caracteriza-se uma inconformidade que deve ser corrigida ou explicada a sua variação. Para o estabelecimento de critérios de aceitação dos valores a serem aplicados durante o processo de registro contábil, foi utilizado o Modelo Autorregressivo e de Médias Móveis, ARMA, com aplicação para estimação dos seus coeficientes de Algoritmo Genético e Algoritmo de Regressão pelo Método de Mínimos Quadrados. Ambos são processos que têm sido bastante utilizados, tanto na previsão quanto na confirmação de valores de séries temporais. Dentre as técnicas de Análise Multivariada, o modelo ARMA mostrou-se mais adaptável pela sua capacidade de decomposição dos itens de custos em séries temporais independentes. O fato de essas séries serem fortemente influenciadas por perturbações aleatórias e padrões das próprias séries ao longo do tempo encorajou a aplicação desse modelo.

Palavras-chave: Custo. Mão de obra. Rateio de custos. Séries temporais. Modelo ARMA. Algoritmos Genéticos.

ABSTRACT

This project is the result of a quest for a financial control support tool to monitor labor costs at a services institution where such costs make up its main production cost item. This tool must be able to identify the cost components that are thought to be out of control, when they are not coherent with the other components (monies) or with their own historic series. It is inconvenient that there should be discrepancy between real costs and the assessment of hours worked in the various administrative sectors, as it compromises business effectiveness. Discontinuities and cost variations must be promptly identified so they can be either corrected or re-planned. All costs incurred will be identified during the monthly accounting cycle, and after being consolidated they will be transferred to the projects that used them. This is done through the transfer price process based on standard costing. Each cost is registered according to its nature and the monthly end amounts make up each one of the registers of the monitored time series. A total of 6 (six) time series were treated in the 4-year period between 2006 and 2009 as we sought to identify behavior patterns that would make it possible to build a forecasting model so that, at the end of the period, it would be possible to judge if the amount to be accounted should be either accepted or refused. Rejection would characterize noncompliance that must be corrected, or else, its variation be explained. To set acceptance criteria for the values to be applied during the recording accounting process we have used the Autoregressive Moving Average model, ARMA, with Genetic Algorithm (GA) and ordinary least squared (OLS) for the model coefficients estimation. Both processes have been widely used both to forecast and to confirm time series. Among Multivariate Analysis techniques, it was this model that has showed itself to be more adaptable, due to its capability to decompose cost items in independent time series items. The fact that such time series are strongly influenced by random disturbances and patterns from the series themselves throughout time encouraged us to apply this time series model.

Key-words: Cost. Labor. Prorated cost. Time series. ARMA model. Genetic algorithms.

SUMARIO

1	INTRODUÇÃO.....	16
1.1	PROBLEMA	18
1.1.1	Objetivo geral	20
1.2	JUSTIFICATIVA	22
1.3	ESTRUTURA DA DISSERTAÇÃO.....	23
2	CONCEITOS GERAIS E REVISÃO DA LITERATURA.....	25
2.1	CONCEITOS CONTÁBEIS	25
2.2	TRATAMENTO DOS DADOS	29
2.2.1	Dessazonalização	29
2.2.2	Deflação	30
2.2.3	Eliminação de <i>Outliers</i>	30
3	METODOLOGIA.....	31
3.1	SÉRIES TEMPORAIS.....	34
3.1.1	Histórico	34
3.1.2	Definição	34
3.2	MODELO ARIMA	37
3.2.1	Estimação dos coeficientes.....	41
3.2.2	Qualidade dos estimadores.....	45
3.2.3	Fluxograma do Modelo.....	47
3.3	ALGORITMOS GENÉTICOS	49
3.3.1	Histórico	49
3.3.2	Definição	49
4	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS.....	53
4.1	ANÁLISE DE DADOS	53
4.1.1	Tratamento das séries temporais.....	53
4.1.2	Identificação de correlações entre séries.....	57
4.1.3	Tratamento de “ <i>outliers</i> ”	58
4.1.4	Estacionamento das séries	61
4.2	IMPLEMENTAÇÃO DOS ALGORITMOS.....	62
4.2.1	Definição	62
4.2.2	Abordagem de ARMA	63
4.2.3	Abordagem Algoritmos Genéticos.....	64
5	RESULTADOS OBTIDOS.....	67
5.1.1	Descrição dos experimentos.....	67
5.1.2	Avaliação da qualidade dos resultados.....	71
6	CONCLUSÕES E TRABALHOS FUTUROS.....	76
	REFERÊNCIAS.....	79
	ANEXO A – SÉRIES TEMPORAIS.....	83

ANEXO B –PROGRAMA AG PARA SÉRIES TEMPORAIS	84
ANEXO C – EXPERIMENTOS DE PARÂMETROS DO AG.....	94
ANEXO D – DASHBOARD DE ANÁLISE DE AJUSTE E RESÍDUOS	96
ANEXO E – RESULTADOS OBTIDOS PARA R2.....	100

LISTA DE FIGURAS

Figura 1 - Processo contábil de apropriação de custos.....	19
Figura 2 - Discriminação do evento econômico x financeiro.....	21
Figura 3 - Taxonomia das técnicas de otimização.....	32
Figura 4 - Fluxograma do processo de projeção.....	48
Figura 5 - Esquema genérico de um algoritmo evolucionário (Smith, 2003).....	50
Figura 6 (a,b,c) - Tratamento de Deflação e Dessazonalização.....	55
Figura 7 (a,b,c) - Tratamento de estabilização das séries.....	56
Figura 8 (a,b) - Método de tratamento de “ <i>Outliers</i> ”.....	60
Figura 9 - Exemplo de modelos de estacionamento das séries.....	61
Figura 10 - Matriz de Regressão.....	63
Figura 11 - Pseudocódigo ARMA.....	64
Figura 12 - Representação dos coeficientes no cromossomo.....	65
Figura 13 - Análise de sensibilidade à taxa de mutação.....	68
Figura 14 - Exemplo de distribuição de ajuste R^2	70
Figura 15 - Funções de Autocorrelação e Autocorrelação Parcial.....	70
Figura 16 (a,b,c) - Estimativa da série 6 para o período dez/09.....	74
Figura 17 - Esquema básico de AG com elitismos.....	84
Figura 18 - Pseudocódigo básico AG.....	84
Figura 19 - Matriz de autocorrelacionamento.....	85
Figura 20 - Mecanismo de avaliação.....	86
Figura 21 - Mecanismo de seleção.....	86
Figura 22 - Mecanismo de combinação.....	88
Figura 23 - Parametrização de <i>Cross-Over</i>	88
Figura 24 - Exemplo de “Dashboard” que combina série 1 com método 3.....	96
Figura 25 - Coeficientes do ajuste para a série temporal “1”.....	97
Figura 26 - Análise de normalidade dos resíduos.....	98
Figura 27 - Teste de autocorrelação das séries e seus erros de estimação.....	99
Figura 28 - R^2 obtidos nas combinações de p, q e métodos (AG, AR e ARMA).....	100

LISTA DE TABELAS

Tabela 1 - Intervalo τ e critério de seleção <i>Outliers</i>	30
Tabela 2 - Matriz de covariância e correlação das séries.....	58
Tabela 3 - Parametrização do algoritmo genético	66
Tabela 4 - Resultados Ótimos R^2 e indicadores de Adaptação	72
Tabela 5 - Distância do valor contabilizado em relação ao valor estimado.....	75
Tabela 6 - Comparativo dos métodos de estimação	77
Tabela 7 - Série histórica de custos	83
Tabela 8 - Descrição dos parâmetros do programa AG	87
Tabela 9 - Análise de sensibilidade aos parâmetros	94
Tabela 10 - Análise de sensibilidade à taxa de mutação.....	95

LISTA DE EQUAÇÕES

[1] Equação de Regressão.....	39
[2] Equação de Médias Móveis	40
[3] Equação do modelo ARMA.....	40
[4] Modelo Geral de série com tendência determinística.....	40
[5] Equação de máxima verossilhança.....	41
[6] Equação de máxima verossilhança.....	42
[7] Estimação dos coeficiente de regressão.....	42
[8] Coeficiente de Determinação.....	45
[9] Raiz quadrada do erro médio quadrático	45
[10] Equação Geral dos Momentos.....	46
[11] Equação de Assimetria	46
[12] Equação da Curtose	46
[13] Teste Ljung-Box.....	47

LISTA DE TERMOS E ABREVIATURAS

ARIMA - Modelos autorregressivos integrados de média móvel

ARMA - Modelos autorregressivos de média móvel

AR – Autoregressive

MA - Moving Average

AE - Algoritmos Evolutivos

AG - Algoritmos Genéticos

IC – Intervalo de confiança um intervalo estimado de probabilidade de ocorrência

MO – Mão de Obra

NP - "nondeterministic polynomial" - designa problemas sem solução determinística

Outliers – Valor atípico ou discrepante de um conjunto de dados

RMSE - Median Square Error

R^2 - Coeficiente de Determinação

ST – Série Temporal

Verba – Denominação dada a item de custo no contexto de folha de pagamentos

1 INTRODUÇÃO

Uma empresa de prestação de serviços educacionais tem na mão de obra o seu principal insumo, e o tratamento deste como um custo variável é uma necessidade para um adequado acompanhamento/rastreamento, que permita a gestão dos custos mensalmente apurados. É uma necessidade da empresa a montagem de um sistema de controle contábil que planeje e registre os seus eventos financeiros e que possibilite a identificação de desvios que se refiram a erros e omissões, permitindo a reação da administração para a correção dos parâmetros que estejam fora de controle.

O sistema de folha de pagamentos, embora seja de natureza determinística, representa a consolidação dos custos individuais de todos os empregados. Precisa atender às características de sigilo, segurança e complexidade legal. Por essa razão, foi tratado como um modelo que absorve e repassa recursos financeiros de forma não uniforme ao longo do tempo, no nível macro, em que são apresentados à contabilidade.

A modelagem matemática de Séries Temporais, com o uso de técnicas de regressão e médias móveis, foi utilizada para simular o comportamento dos custos ao longo do tempo. A comparação entre os valores estimados por esse modelo e os valores reais apurados na contabilização de um determinado mês é o propósito dessa ferramenta de controle. O objetivo é orientar o analista de custos na tarefa mensal de controle financeiro da organização, no momento em que se comparam os valores apurados, revelados para a contabilidade, e o valor esperado estimado pelo modelo. A técnica de regressão e médias móveis (ARMA) é também conhecida por Box&Jenkins (Box, 1976), que desenvolveu um método para identificação e calibração do modelo.

Para a verificação da eficácia (precisão) dos estimadores obtidos pela modelagem estocástica, a análise foi complementada pelo uso de metodologia heurística com a construção de algoritmo genético, algoritmo da classe dos algoritmos evolutivos, como método alternativo de estimação dos coeficientes. Os resultados obtidos por esses dois métodos foram comparados com a utilização do

coeficiente de determinação (R^2), que é um indicador de qualidade de ajuste entre os valores reais da série temporal e os valores estimados pelo modelo.

Esta pesquisa foi iniciada com o objetivo de identificar uma metodologia ou ferramenta estatística que apoiasse o processo de conciliação contábil dos custos de mão de obra. O processo de conciliação manipula grande número de variáveis: número de verbas de uma folha de pagamentos (salários, INSS, FGTS,...) e variação do número de pessoas, com admissões e demissões.

A primeira fase constitui a verificação da coerência da metodologia de cálculo do custo padrão de MO, adotada pela organização administrada por projetos, o qual se mostrou adequado para a realidade apurada no período de quatro anos (2006 a 2009). Observa-se no Anexo A que o processo encontra-se balanceado, atendendo aos objetivos de equilibrar os custos incorridos com os apropriados aos projetos.

Na segunda fase, são implementados em MATLAB® os modelos estocástico e heurístico para a estimação dos parâmetros do modelo ARMA, que descreve as séries históricas de custo, de forma a otimizar resultados, minimizando os erros de estimativa e que será usado para gerar previsões. O comportamento estatístico dos erros obtidos pela diferença entre os valores estimados por esse modelo e os valores reais das séries temporais será usado para estabelecer intervalo de confiança para o valor projetado.

No curso do desenvolvimento desta pesquisa, foi forte a sugestão da literatura para que se buscassem modelos parcimoniosos, ou seja, aqueles com o menor número de parâmetros possível, mantida a qualidade de estimação. Porém, o que se constatou na prática é que o custo computacional é muito pequeno para se expandir o número de coeficientes do modelo ARMA em simulações.

Uma característica que emergiu com a evolução dos resultados foi a necessidade de tratamentos dos dados com a extração de outliers, sem o que não seria possível a obtenção de suficiente qualidade no ajuste do modelo para possibilitar as inferências estatísticas dos valores a serem estimados pelo modelo.

Para a avaliação da eficiência dos algoritmos de estimação de parâmetros do modelo foi usado o indicador de qualidades do ajuste R^2 -coeficiente de determinação. Pelo critério de número de iterações, o algoritmo genético é o mais complexo, seguido do algoritmo ARMA, que, por sua vez, é seguido pelo AR, modelo de regressão simples. Nas situações em que a série apresenta comportamento estável, ou quando o desvio-padrão da série é pequeno em relação à sua média, as melhorias obtidas no indicador R^2 , ao escolher algoritmos mais sofisticados, não se mostraram vantajosas, sendo nessa situação preferível a seleção do algoritmo mais simples, ou seja, o AR.

1.1 Problema

Em geral os custos de mão de obra de uma empresa apresentam uma uniformidade ao longo dos meses. Tais valores, em muitas empresas, são absorvidos por Departamentos/Centros de custos na proporção de seu consumo. Desta forma, é necessário desenvolver um modelo, conforme esquematizado na figura 1, a seguir, que gerencie essas ações, adquirindo mão de obra quando contrata ou compra horas de profissionais e repassando-a quando oferece, vende, para consumo dos diversos centros de custos. Os gerentes responsáveis pelos centros de custos desejam que os custos que lhes são imputados não excedam o custo real. Essa premissa estará atendida, na medida em que o custo incorrido, o custo real, seja compensado pelo custo transferido. Tanto melhor será a qualidade do custo padrão, quanto menor for a diferença entre o custo real e o custo transferido.

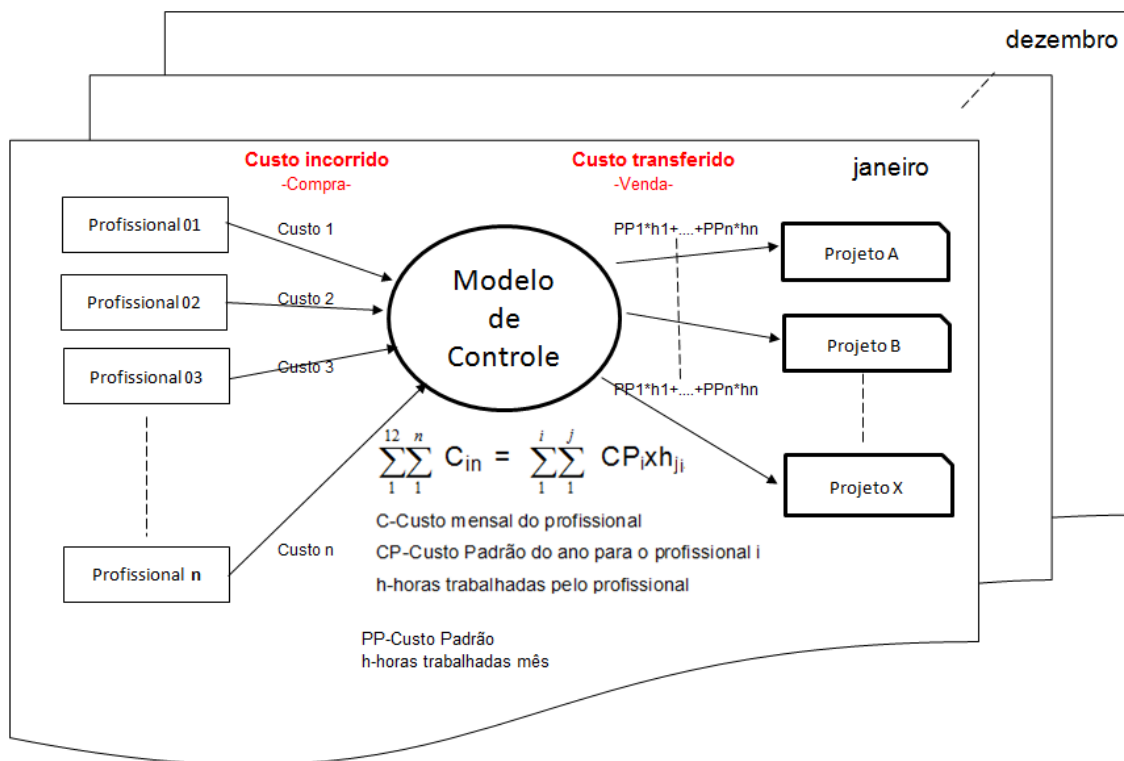


Figura 1 - Processo contábil de apropriação de custos.

A série temporal é representada pelos custos mensais e custos transferidos, os quais precisam ser monitorados. As não conformidades identificadas devem gerar ações corretivas, na hipótese de erro, ou, então, que tenham explicadas as suas diferenças. Havendo desvio/variação, o modelo demonstra domínio sobre a informação, sendo necessária uma intervenção no processo, restabelecendo-se o equilíbrio entre os custos incorridos e os custos transferidos. Para a manutenção desse equilíbrio, o caminho natural é o ajuste do custo padrão, que será definido a seguir na seção 2.1, o que nem sempre é possível no curso de um exercício fiscal. Tal fato agrava a necessidade de controle sobre os custos incorridos. É frequente e natural a ocorrência de flutuações nas contas de despesas com mão de obra, dadas as características de rotatividade, acidentes, afastamentos, gozo de férias etc. Esses fatos tornam o processo de acompanhamento tarefa árdua, tendo em vista o grande número de verbas e a sua variabilidade.

De acordo com a sua natureza, alguns custos, mesmo que não sejam pagos no mês da realização do trabalho, precisam ser contabilizados, pois são obrigações a serem pagas no futuro. É o caso do 13º salário e férias. A esse processo de reconhecimento de custo denomina-se provisão. Essas provisões precisam ser constituídas e baixadas, quando da realização do seu pagamento, e a apuração do seu saldo deve ser contemplada no modelo.

Nesse contexto, faz-se necessária uma ferramenta de apoio que permita ao analista a identificação de potenciais desvios, acelerando o processo de conciliação que possa oferecer diagnóstico de não conformidades de valores atípicos de custos.

1.1.1 Objetivo geral

A determinação do custo padrão unitário do profissional tem as suas regras de negócio bem conhecidas. As incertezas na determinação do custo padrão serão resolvidas pela adoção de premissas que, no acompanhamento, devem ser objeto de verificação, o que pode ser feito por uma análise de resíduos. Como exemplo, para entender a natureza do tipo de incerteza, cita-se o valor de reajuste salarial a ser determinado por dissídios coletivos de diferentes categorias profissionais.

Esses tratamentos de conversão dos eventos financeiros em eventos econômicos (Figura 2) constituem o modelo de precificação de mão de obra que se encontra bem estabelecido e é usado no cálculo do custo a ser transferido para os projetos, custo padrão do profissional. O seu detalhamento é apresentado a seguir.

		Pagamentos											
		Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
	Dias-->	176	144	152	168	160	168	184	160	176	184	160	152
Salário	100,0%	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	
13 Salário	9,1%												100,0
Férias	9,1%												100,0
Adic. Férias	33,3%												33,3
INSS	20,0%	20,0	20,0	20,0	20,0	20,0	20,0	20,0	20,0	20,0	20,0	20,0	
FGTS	8,5%	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	
INSS-13 Salário	1,8%												20,0
INSS-Férias	6,7%												20,0
FGTS-13 Salário	0,8%												8,5
FGTS-Férias	0,8%												8,5
PIS	1,0%	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Transporte	1,0%	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Alimentação	1,0%	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Saúde/Seguro	2,0%	2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0
PLR	200,0%												200,0
Total	163,7%	133,5	133,5	133,5	133,5	133,5	133,5	133,5	133,5	133,5	133,5	133,5	495,3

		Custos											
		Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
	Dias-->	176	144	152	168	160	168	184	160	176	184	160	152
Salário	100,0%	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	
13 Salário	9,1%	9,1	9,1	9,1	9,1	9,1	9,1	9,1	9,1	9,1	9,1	9,1	
Férias	9,1%	9,1	9,1	9,1	9,1	9,1	9,1	9,1	9,1	9,1	9,1	9,1	
Adic. Férias	33,3%	3,0	3,0	3,0	3,0	3,0	3,0	3,0	3,0	3,0	3,0	3,0	
INSS	20,0%	20,0	20,0	20,0	20,0	20,0	20,0	20,0	20,0	20,0	20,0	20,0	
FGTS	8,5%	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	8,5	
INSS-13 Salário	1,8%	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	
INSS-Férias	6,7%	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,8	
FGTS-13 Salário	0,8%	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	
FGTS-Férias	0,8%	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8	
PIS	1,0%	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	
Transporte	1,0%	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	
Alimentação	1,0%	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	
Saúde/Seguro	2,0%	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2	2,2	
PLR	200,0%	18,2	18,2	18,2	18,2	18,2	18,2	18,2	18,2	18,2	18,2	18,2	
Total	163,7%	178,5	178,5	178,5	178,5	178,5	178,5	178,5	178,5	178,5	178,5	178,5	-

Figura 2 - Discriminação do evento econômico x financeiro

A criação de um modelo que propicie a rastreabilidade dos custos e suas apropriações deve contemplar as variáveis que influenciam os saldos contábeis, destacando-se os seguintes pontos:

- Absorção da heterogeneidade do número de dias úteis de cada mês. Os custos da folha são idealmente uniformes.
- O período de gozo das férias é uma opção do profissional.
- Podem ocorrer reajustes de salários imprevistos.
- A legislação sobre a concessão de benefícios financeiros pode mudar.

Esse modelo acompanha os gastos mensais da organização, efetuadas todas as alocações de mão de obra relacionadas aos custos padrões, comparando esse volume com os custos reais pagos registrados na contabilidade. Os desvios observados entre os saldos contábeis das despesas geradas pela folha de pagamento, os custos apropriados pelo produto das horas alocadas e os custos

padrão devem ser explicitados. O propósito desta pesquisa é a modelagem desses processos, realizada com a utilização de técnicas de heurística para a determinação/identificação dos estimadores que apresentem adequada fidelidade, emulando os movimentos financeiros de custo de mão de obra e a utilização de técnicas estatísticas determinando o grau de confiabilidade nos resultados estimados pelo modelo. Além disto, o estudo se completa ao oferecer comparação de valores projetados pelo modelo com os apurados, quando, por critérios de tolerância preestabelecidos, estes serão aceitos ou rejeitados. Sendo rejeitados, aprofunda-se o processo de conciliação contábil detalhando as variações.

1.2 Justificativa

A pesquisa teve por base a apuração dos custos com mão de obra, que, estatisticamente tratados/sensibilizados, aceleram o processo de conciliação dos custos com mão de obra. Esses custos são utilizados no cálculo anual de custo padrão. As inconsistências apontadas pelo modelo devem conter indicação das possíveis origens e, se possível, trazer o dimensionamento da sua probabilidade.

Esta proposta contempla conceitos e métodos já implementados que se mostram satisfatórios para grande parte das organizações que têm na mão de obra custo relevante dos seus produtos e serviços. Particularmente, a mão de obra é o elemento que requer maior atenção na análise financeira, ao longo dos doze meses de acompanhamento de um exercício fiscal.

A pesquisa busca o desenvolvimento de um modelo que combine variáveis aleatórias de soluções e que explique desvios observados nos custos. O teste realizado utiliza dados históricos e mede a capacidade do modelo para gerar soluções viáveis em tempo compatível. Com o esforço de conciliação contábil, identificam-se as fontes de desvio, ou seja, minimizam-se os erros ao experimentar variáveis de perturbação. Como exemplos de variáveis de perturbação, poderiam ser citadas alterações salariais e férias fora do período programado.

Este trabalho trata da investigação de soluções por intermédio de uma heurística, cuja viabilidade e adequação foram fortemente influenciadas pelos resultados obtidos na aderência das técnicas de estatística e na otimização combinatória de algoritmos genéticos. Assim, buscou-se, através da eliminação de “*outliers*” das séries e da modelagem ARMA das mesmas, a identificação dos erros de estimação, os quais conduzem à definição dos intervalos de confiança para aceitação ou rejeição dos valores futuros das séries.

1.3 Estrutura da dissertação

O foco deste trabalho é a busca de uma ferramenta de modelagem matemática que acelere o processo de conciliação, e que identifique, dentre as diversas verbas, as contabilizações impróprias e os erros, e que ofereça indicações sobre as tendências das séries.

Foram adotadas duas linhas de análise dos dados. A primeira atua sobre as diversas verbas da folha de pagamento identificando inconformidades no comportamento dos custos, os quais são lançados através da identificação de padrões e correlações entre os diversos itens, além do agrupamento de séries que apresentam grau significativo de correlação. A segunda linha de análise focou na criação de um modelo autorregressivo de médias móveis, que utiliza algoritmo genético para identificar os estimadores da equação ARMA e também explora a combinação de parâmetros (p e q) deste modelo econométrico, propiciando uma melhor combinação para cada uma das séries estudadas. Assim, identifica-se o modelo de coeficientes e parâmetros com menor erro de estimação.

O principal instrumento de conciliação é a determinação do intervalo de confiança, com a utilização da hipótese de distribuição gaussiana dos erros obtidos entre os valores reais das séries e o seu valor estimado.

A análise desse conjunto de séries temporais com o emprego de uma ferramenta auxiliar de conciliação dos valores contabilizados constitui o objetivo desta pesquisa. Na verificação da eficiência, foram utilizados os dados coletados em um empresa de serviços educacionais da região metropolitana de Belo Horizonte,

nos meses de janeiro de 2006 a dezembro de 2009. Os 47 primeiros meses foram utilizados para a obtenção dos coeficientes, treinamento do modelo, e o último mês, para teste de eficiência do modelo em prever o valor a ser contabilizado.

2 CONCEITOS GERAIS E REVISÃO DA LITERATURA

Fazem parte da linguagem da contabilidade de custos termos cuja conceituação, apesar de eles serem simples, diretos e associativos à ideia que pretendem transmitir, é necessária à compreensão do texto. Com o objetivo de familiarizar o leitor com esses termos técnicos, neste capítulo, os mesmos são conceituados.

2.1 Conceitos Contábeis

Encontram-se na literatura os conceitos fundamentais utilizados na área de Administração e Finanças, os quais foram utilizados neste trabalho. Como referência citam-se: Introdução à contabilidade gerencial (Horngren, 1985), Contabilidade por atividades (Brimson, 1996) e Contabilidade de Custos (Martins, 2001).

Natureza dos custos

Para a avaliação do resultado empresarial, as organizações estruturam-se com sistemas de controle que melhor reflitam a sua complexidade operacional sem afastar-se das normas contábeis de registros e relatórios. Essas técnicas podem variar substancialmente de empresa para empresa, não sendo, no entanto, variáveis com relação à natureza do tratamento dos custos, que podem ser diretos ou indiretos.

Os custos diretos são identificados quando o seu consumo está associado diretamente à produção de um bem ou serviço. Para o custo indireto essa variação não acontece. Para clareza desse entendimento, cita-se o clássico exemplo de um operário que trabalha dedicado a uma determinada máquina, então, o produto que venha a ser produzido por essa máquina levará como custo direto o custo-hora desse operário na razão direta do tempo necessário à sua produção. Como custo indireto, aproveita-se o exemplo para associar o custo-hora do supervisor do referido operário, que pode estar supervisionando um número indeterminado de operários e máquinas.

Mão de obra Direta

Está relacionada ao pessoal que trabalha diretamente sobre o produto em elaboração, desde que seja possível a mensuração do tempo despendido e a identificação de quem executou o trabalho, sem necessidade de qualquer apropriação indireta ou rateio.

Custo Padrão

Custo padrão é o custo predeterminado para a produção de uma única unidade, ou um número de unidades do produto/serviço, durante um período específico. É o custo planejado de um produto/serviço, segundo condições de operação normais ou ideais de eficiência e volume, especialmente com respeito à despesa indireta. Aplica-se à finalidade de simplificar ou agrupar custos, que possam ser associados a produtos ou serviços, sendo, desta forma, tratados como custos diretos dos produtos/serviços em processo de supervisão sobre o qual se deseja exercer controle.

Projeto

Atividade ou conjunto coordenado de atividades dirigidas para alcançar objetivos explícitos e justificados, segundo uma metodologia definida e empregando recursos humanos e materiais durante certo período de tempo.

Centros de Custos/Resultados

Unidade contábil que segrega os lançamentos, alocação de recursos, particulares de um projeto, com vista à criação de um produto o qual se deseja vender ou item cujo custo deseja-se controlar. A este particionamento/segmento denomina-se centro de resultado, no primeiro caso, ou centro de custo, no segundo caso.

Princípio da competência

Para a adequada apuração do resultado econômico de uma organização, faz-se indispensável a aplicação do princípio contábil da competência, o qual estabelece que o reconhecimento de um custo aplicado a um produto, no caso de custo direto ou absorvido por um departamento, deve ser aplicado no momento do

seu consumo, e não quando do pagamento da sua aquisição. Este último é um evento financeiro e não econômico.

O significado prático desse princípio é que, independentemente do pagamento, se a obrigação foi gerada por consumo, um lançamento (provisão) deve antecipar o registro do custo. Exemplificando: se os salários dos funcionários forem pagos no dia 5 do mês subsequente ao mês trabalhado, deve a obrigação com os funcionários ser registrada no mês correspondente aos serviços prestados. A esse registro denomina-se provisão do custo, que ora então se encontra em processo de pagamento.

Para não alongar mais nos conceitos contábeis, é relevante registrar que se estará ao longo deste trabalho tratando sempre com o conceito econômico do custo no seu consumo, assim como no caso de venda ao cliente interno.

Resultado econômico

O objetivo do controlador de uma organização que trabalhe no conceito de administração por projetos e com a opção por custo padrão na apuração de mão de obra é dimensionar o custo padrão unitário dos profissionais, de modo a equilibrar os custos que se aplicam a esse profissional e o resultado do produto do dito custo padrão, parâmetro de ajuste, e o número de horas trabalhadas no ano. Para atingir esse objetivo, faz-se necessário transferir, caso existam horas ociosas, esses custos a centro de custo próprio, denominado custos de mão de obra não apropriada. Esse procedimento elimina a existência de custos indiretos ou custos a ratear. Estabelecido o procedimento de transferência do custeio de mão de obra aos projetos, podem-se analisar as verbas que compõem o custo e que representam o número de horas disponíveis/trabalhadas em um determinado mês, bastando multiplicá-las pelo custo padrão do profissional que está do lado oposto da equação que avalia o equilíbrio entre o custo incorrido e custo transferido.

Folha de pagamentos

A captação das informações tem lugar, no caso dos custos, na folha de pagamento, que é composta de salários, encargos e benefícios. Os salários e encargos são os itens de maior interdependência e com maior variabilidade pelo

efeito de perturbações sofridas em elementos como opções de férias, adiantamento de 13º salários etc.

Os dados correspondentes ao custo no modelo são captados das informações da folha de pagamento, as quais são transmitidas à contabilidade.

Contabilidade de Custos

Destina-se a produzir informações para diversos níveis gerenciais de uma entidade e funciona como auxílio às funções de determinação de desempenho, de planejamento e controle das operações e de tomada de decisões. Além disto, possibilita a alocação de forma mais criteriosa possível dos custos de produção aos produtos.

Contabilidade Geral

Ciência que tem como objeto de estudo o patrimônio das entidades, seus fenômenos e variações, tanto no aspecto quantitativo quanto no qualitativo, registrando os fatos e atos de natureza econômico-financeira que o afetam e estudando as suas consequências na dinâmica financeira.

Conciliação contábil

A confrontação dos dados de custos com a do mecanismo de transferência destes para os projetos denomina-se conciliação contábil, com a qual se busca explicar as diferenças observadas durante os meses do ano e minimizar a diferença desses dois, no final de cada ano. As incertezas, as variações, as perturbações do ambiente empresarial precisam ser conhecidas com a maior presteza possível.

Na metodologia do cálculo do custo padrão, no domínio da Contabilidade Gerencial, as estimativas dos seus valores não podem produzir discrepância, que no modelo representaria diferenças entre o valor de venda e o custo da folha de pagamento. A conciliação entre o relatório da Contabilidade de Custos e o da Contabilidade Geral é imprescindível, não sendo toleradas diferenças.

Nada pior para uma empresa do que seus executivos receberem relatórios de dois setores da empresa tratando do mesmo assunto, mas apresentando valores diferentes (...) Por esta razão, atenção toda especial deve ser dada ao caso em que a firma necessite dos dois tipos de informação como esses mencionados. E o fundamental para bom andamento do sistema é a completa conciliação entre estas duas informações. São diferentes, é claro, já que prestam a duas finalidades diferentes, mas precisam ser, até o centavo, conciliadas.

(MARTINS, 2001)

2.2 Tratamento dos Dados

Os dados relativos aos custos captados diretamente na contabilidade contêm componentes determinísticos conhecidos e componentes desconhecidos que, por representarem discrepância significativa, permitem ser identificados como anormais. Para a preparação das séries temporais para a aplicação do modelo estocástico, foram tratadas as séries com extração desses componentes.

2.2.1 Dessazonalização

O fator de sazonalidade está relacionado às flutuações periódicas de comprimento constante e intensidade proporcional, que são causadas por fatores conhecidos como temperaturas, chuvas, meses do ano, período de festas e políticas corporativas (Makridakis, 1989).

Para corrigir esse fator conhecido, foi aplicado na eliminação dos efeitos de variações do número de dias úteis em cada mês o tratamento de homogeneização/suavização com aplicação de coeficiente/índice equivalente à razão do número de dias úteis do mês e à média dos dias úteis no ano (Tabela 7 – inclusa no Anexo A).

2.2.2 Deflação

Para a eliminação de “tendência”, advinda dos reajustes salariais e/ou do aumento do número de funcionários, as séries são deflacionadas com a utilização do índice de custo-hora médio, nos quais estão embutidos os aumentos salariais diretamente e a variação do número de colaboradores indiretamente. Após a correção da série com esses índices, pode-se observar a estabilidade gerada pelo significativo estreitamento do intervalo de confiança.

2.2.3 Eliminação de *Outliers*

Com a utilização dos conceitos de estatística robusta, foram extraídos das séries que apresentem desvio superior a 1,28 desvio-padrão tomando-se por referência todos os itens de cada uma das séries. O julgamento de aceitar como sendo ou não um *outlier* foi tomado levando-se em conta o afastamento, em relação à mediana de um conjunto de seis pontos, três pontos anteriores e três posteriores, ser superior a esse desvio. O resultado está mostrado na Tabela 1 abaixo. Este tratamento possibilitou significativa melhora no indicador de qualidade R^2 , assim como estreitamento do intervalo de confiança.

	Venda (a)	Salários e Ordenados P. Admin	Décimo Terceiro Salário	F. G. T. S. e I. N. S. S.	Provisão e Abono de Férias	Outros Indiretos PIS-AVP-FI	Custo (b)	Resultado (c) = (a) - (b)
Média	783	396	31	144	43	199	812	(30)
DesvPadrao	108	60	7	50	18	193	202	201
Outliers	0	5	4	7	6	4	5	5
Percentual de confiança	90,0%	90,0%	90,0%	90,0%	90,0%	90,0%	90,0%	90,0%
Intervalo τ	139	77	9	64	23	248	259	258
Método um lado(1) dois la	2	2	2	2	2	2	2	2

Tabela 1 - Intervalo τ e critério de seleção *Outliers*

3 METODOLOGIA

Previsão significa a expectativa de estabelecer um evento que está para acontecer no futuro baseado, pelo menos em parte, no presente ou no passado. Assim, pode-se definir o princípio da continuidade de um evento quando este possui padrões que se repetem no tempo. As técnicas de previsão podem ser classificadas em duas categorias: métodos quantitativos e métodos qualitativos.

Os métodos qualitativos podem ser divididos em exploratórios e normativos. Os exploratórios utilizam-se do passado e presente como ponto de partida, movendo-se em direção ao futuro, investigando todas as possibilidades possíveis. Os métodos normativos começam pela determinação futura de metas e objetivos, movendo em direção ao passado, para verificar se existem restrições de tecnologias e recursos, no presente, para a sua consecução.

Os métodos quantitativos são divididos em séries temporais e métodos causais, sendo que estes assumem que os objetivos, variáveis dependentes, têm relação de causa e efeito com uma ou mais variáveis independentes, identificados como a sua causa. A projeção de séries temporais é a previsão do futuro baseada nos valores passados dessas mesmas variáveis com o objetivo de identificar padrões na série histórica que permitam a sua exploração no futuro (Wong Francis, 2009).

O resultado de Wold¹(1954) provou que qualquer série temporal pode ser representada por uma estrutura de componentes determinístico e estocástica de médias móveis. Soluções parcimoniosas têm sido obtidas com a metodologia dos pesquisadores George Box e Gwilyn Jenkins (1970). Desde então, encontram-se na literatura várias técnicas de heurística para ajustar os dados. Para a construção desses modelos, novas técnicas da área de Inteligência Artificial vêm sendo aplicadas e têm apresentado um bom desempenho. Dentre essas destacam-se Redes Neurais Artificiais, Algoritmos Evolucionários, Sistemas Nebulosos (De Souza, 2006).

¹ O teorema de Wold estabelece que toda a covariância de uma série estacionária pode ser escrita como um processo de médias móveis infinita. Essa formulação é conhecida como representação de séries temporais de médias móveis ou teorema de decomposição de Wold.

Muitos problemas práticos e teóricos de otimização apresentam grandes espaços de busca multimodais. Dentre eles incluem-se os problemas NP – difíceis de otimização combinatória – cujas estruturas são complexas. Na área de planejamento da produção e logística, esses problemas ocorrem frequentemente. A aplicação de métodos convencionais de pesquisa operacional, como programação dinâmica, método simples, ou técnicas de gradiente falham frequentemente devido ao crescimento exponencial do esforço computacional, tendo em vista a dimensão do problema. Além disso, os métodos heurísticos demandam muito menos esforço computacional, mesmo que não assegurando uma solução ótima global.

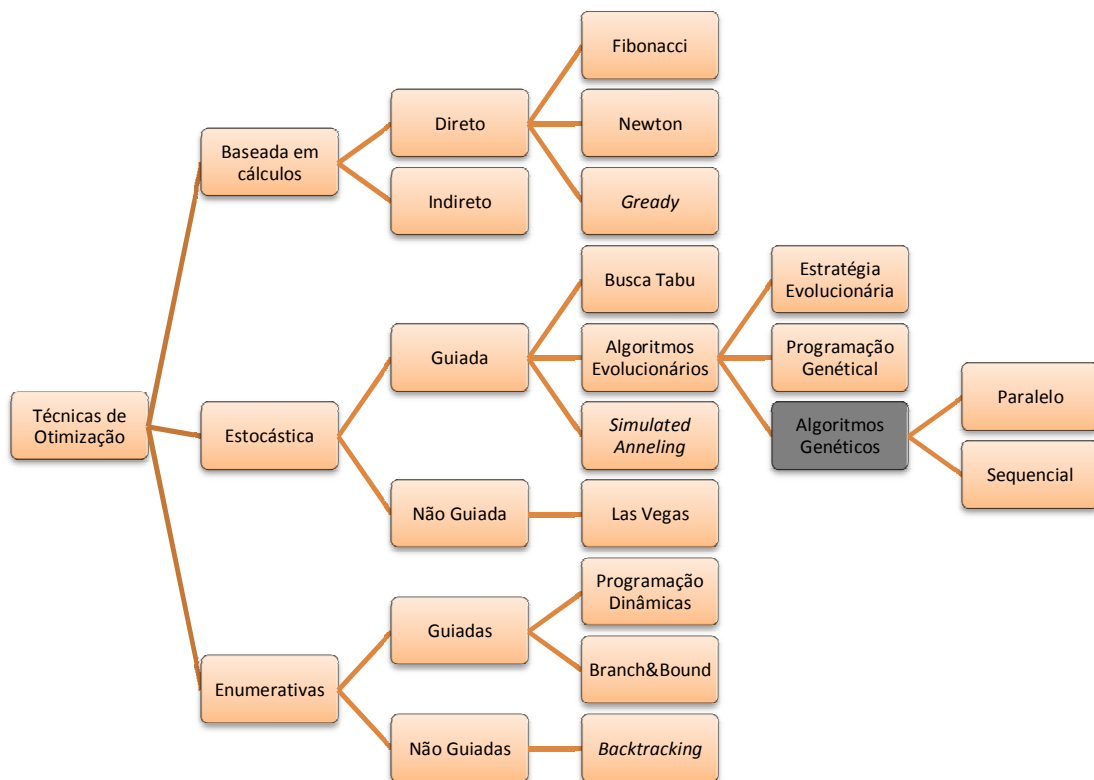


Figura 3 - Taxonomia das técnicas de otimização

Dentre as técnicas heurísticas (Figura 3), os algoritmos genéticos têm sido usados como procedimento de busca de múltipla finalidade, comumente utilizada em otimização. A crescente popularidade dessa metaheurística deve-se à sua adaptabilidade e simplicidade, a qual tem sido aplicada a uma grande variedade de

problemas, com diferentes graus de complexidades. Esse algoritmo combina técnicas de busca e padrões de reconhecimento de dados passados (Mahpol, 2005).

Os algoritmos evolucionários são candidatos ao emprego na estimação de parâmetros quando se implementa uma busca global múltipla, com rápida localização de alta qualidade. É esperado crescimento da sua importância pelas vantagens de gerarem modelos explicitamente representados e adaptativos que evitam estacionar em mínimos locais (Cortez, 2004).

Uma das vantagens em se empregar AG na modelagem de séries temporais reside no fato de ser ele um método estocástico de busca global, que imita os conceitos de seleção natural da sobrevivência do indivíduo mais adaptado. Apesar da possibilidade da sua utilização na identificação do modelo, frequentemente este é usado para a estimação de coeficientes (Chorng, 2005).

Diferentemente de outros métodos, como busca tabu e "*Simulated Annealing*", os algoritmos genéticos exibem um paralelismo implícito, pois consideram muitas soluções a cada iteração, reduzindo as chances de estacionarem em um ótimo local. Apesar da sua capacidade de busca, não existe garantia de que o AG irá alcançar a solução global, embora, frequentemente, obtenha soluções satisfatórias em tempo reduzido (See, 1999).

3.1 SÉRIES TEMPORAIS

3.1.1 Histórico

Em geral, os métodos de previsão partem de valores passados para estimarem valores futuros. Até 1920, a previsão era efetuada através de extrapolação simples de um valor global, ajustado em função do tempo. Só em 1927, Yule criou o modelo Autorregressivo (AR), em que o valor a ser projetado dependia de valores passados. Por duas décadas seguintes, os pesquisadores da área de Estatística acreditaram que o comportamento de uma série dependia dos modelos lineares e do ruído. Com o passar do tempo, porém, percebeu-se que modelos lineares não eram suficientes para a análise de Séries Temporais, já que a maioria das séries reais apresentava fortes tendências de não linearidade.

A partir de 1950, novos estudos foram realizados na área da Previsão de Séries Temporais. Nessa época foi criada a técnica do Alisamento Exponencial, que exigia poucos cálculos e era de fácil utilização. Com o advento do computador nos anos 50, houve uma revolução nos métodos de previsão.

3.1.2 Definição

Uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. A característica mais importante desses tipos de dados reside no fato de que observações vizinhas são dependentes, e o interesse é em analisar e modelar essa dependência. Enquanto em modelos de regressão a ordem das observações é irrelevante, para a análise de séries temporais, a ordem dos dados é

importante. Vale notar também que o tempo pode ser substituído por outras variáveis como espaço, profundidade etc.

Como a maior parte dos procedimentos estatísticos, a técnica das séries temporais foi desenvolvida para analisar observações independentes. Assim, o estudo de séries temporais requer o uso de técnicas específicas. Dados de séries temporais surgem em vários campos do conhecimento, como Economia (preços diários de ações, taxa mensal de desemprego, produção industrial) (Baragonaa, *et al.*, 2001) (Beenstock, *et al.*, 2002) (Chorng, 2005), Medicina (eletrocardiograma, eletroencefalograma), Epidemiologia (número mensal de novos casos de meningite), Meteorologia (Parviz, 2010) (precipitação pluviométrica, temperatura diária, velocidade do vento) etc.

Algumas características são particulares a esse tipo de dados. Observações correlacionadas são mais difíceis de analisar e requerem técnicas específicas. Precisamos levar em conta a ordem temporal das observações. Fatores complicadores como presença de tendências e variação sazonal ou cíclica podem ser difíceis de estimar ou remover. Nesses casos, a seleção de modelos pode ser bastante complicada, e as ferramentas podem ser de difícil interpretação.

Análise

A análise das séries temporais considera que cada observação resulta da conjunção de vários efeitos, consistindo o trabalho na decomposição da série e análise do comportamento de cada uma das suas componentes, com vista à previsão. Normalmente, são consideradas quatro componentes presentes em cada observação:

- a tendência, T, que traduz o nível da sucessão, isto é, o comportamento mais geral da variável sobre o tempo;
- o ciclo, C, que traduz as oscilações em torno do nível da série de prazo superior ao ano, ou seja, os movimentos oscilatórios de longo prazo. Nas séries econômicas, essa componente frequentemente apresenta uma periodicidade pouco definida, o que faz com que, nos estudos de curto prazo, o seu efeito seja integrado na tendência, considerando-se a tendência-ciclo;
- a sazonalidade, S, que se refere às oscilações em torno da tendência de periodicidade inferior ao ano, detectável em séries infra-anuais, com origem

em causas naturais, como o clima, ou em causas sociais e institucionais, como os costumes e disposições;

- a componente errática ou residual, que traduz os movimentos irregulares explicados por causas fortuitas ou desconhecidas, para onde são remetidos os movimentos não explicados pelas componentes anteriores e que se assume ter um comportamento aleatório.

Cada observação, Z , é vista como resultante da combinação daqueles quatro elementos expressando-se na forma genérica

$$O = f (T, C, S, \varepsilon)$$

Para este projeto de pesquisa algumas particularidades do problema em estudo são consideradas e levam à eliminação dos componentes de tendência e sazonalidade pelos seguintes motivos:

- Os salários são reajustados anualmente, sendo proposta a aplicação da deflação dos valores de custo correntes para moeda constante na data atual. As séries são tratadas como séries estacionárias. Esse é o primeiro movimento para o estacionamento da série.
- A sazonalidade que se espera deve estar presente nas séries temporais em estudo e deve ser adequadamente refletida pelos coeficientes de regressão ao atingir o número de doze. A sazonalidade oriunda da variação de dias úteis de cada mês foi tratada em amaciamento da série de Venda.
- Durante a captação dos dados se fez necessária a eliminação de componentes exógenos que foram extraídos como ruídos ou incorporados como parte da parametrização do modelo através do tratamento de estatística robusta, cujo procedimento é abordado na seção 4.1.3 – Tratamento de *Outliers*.

Além desse tratamento, foi necessária a adequação da série, eliminando-se os perturbadores, que eram conhecidos previamente, e aplicando-os na correção da série na forma de deflatores sobre baixas contábeis de provisões, alterações salariais etc.

Pré-requisitos para análise

Para que seja possível a aplicação de inferência estatística de uma série temporal Y_t há necessidade de que a série temporal seja estacionária, o que significa serem a esperança, $E(Y_t)$ e a covariância $E(Y_{t-\mu})(Y_{t-j-\mu})$ independentes de “t”.

Um processo estocástico é denominado fracamente estacionário (Bueno, 2008), se:

- 1) $E|Y_t| < \infty$
- 2) $E(Y_t) = \mu$ para todo $t \in Z$ e
- 3) $E(Y_{t-\mu})(Y_{t-j-\mu}) = \gamma_j$

O significado dessa condição é que (1) o segundo momento da série seja finito, (2) a média seja igualmente independente do período “t” e (3) a covariância seja sempre igual, não dependendo do tempo, mas dependente da distância temporal “j” tomada entre observações. Visualmente observa-se que a estacionaridade de uma série flutua em torno de uma média fixa se a variância da série é constante ao longo do tempo.

Existe ainda o conceito de estacionaridade estrita, que está associada à condição de que a função de distribuição de $\{Y_t\}_{i=1}^k$, seja igual à função de distribuição $\{Y_{t+h}\}_{i=1}^k$, o que significa que os gráficos da função de distribuição da série/processo em quaisquer dois intervalos de tempo exibirão propriedades estatísticas semelhantes. De outra forma, os momentos populacionais, quando existem, são independentes de “t”.

3.2 Modelo ARIMA

O modelo autorregressivo (AR) foi descrito por Yule (Yule, 1927). Em 1937, Slutsky (Slutsky, 1937) propôs o modelo de médias móveis (MA), e em 1954 Wold (Wold, 1954) descreveu o modelo misto ARMA. Apesar de bem estabelecidas, essas

3 (três) classes de modelos tiveram seu desenvolvimento e aplicações severamente limitados por não estarem, à época da sua proposição, disponíveis recursos computacionais adequados. A expansão da disponibilidade dos computadores tornou possível a utilização dos modelos ARMA (Makridakis, 1989).

A publicação, em 1976, por Box e Jenkins da obra "Time series analysis: forecast and control" instalou uma nova geração de ferramentas de previsão. Por sua autoria, é conhecida como metodologia Box&Jenkins, mas, pela metodologia, é denominado modelo ARIMA, em que AR refere-se a autorregressivo, I a ser integrado, na hipótese de se fazer necessário o estacionamento da série, e MA refere-se a médias móveis. Essas denominações agregam os princípios matemáticos em que se baseiam. A ênfase desse método está na análise probabilística ou estocástica de séries temporais econômicas, em que os dados falam por si, inexistindo análise causal. Da mesma forma que um modelo de regressão, tem a variável dependente Y explicada exclusivamente por suas variáveis independentes Xi. Nesse modelo, as variáveis explicativas são os próprios valores da variável dependente defasados no tempo ($X_t = Y_{t-1}$) juntamente com os componentes estocásticos. Por essa razão, esses modelos são denominados teóricos (Gujarati, 2004).

O processo de modelagem pode ser decomposto em:

1. Transformação da série temporal para uma série estacionária através da tomada das diferenças que a transformam em uma série integrada. A partir desta, calculam-se as Funções de Autocorrelação(FAC) e Autocorrelação Parcial(FACP), que, na forma apresentada pelo gráfico dos valores dessas funções, em relação à defasagem, terão no decaimento para o valor zero indicação da ordem do modelo, ou seja, indicação para escolha do número de parâmetros adequados de p (número máximo de regressores) e q (número máximo de médias móveis) a serem tomadas. O objetivo é obter-se um modelo parcimonioso, isto é, com o menor número possível de parâmetros (coeficientes) conjugado com o menor erro de estimativa. O resultado dessa fase é obtido pela extração da ordem do modelo com a definição dos parâmetros "p" e "q".
2. Estimação dos parâmetros/coeficientes do modelo ARMA através do método de mínimos quadrados para a parte de regressão, que é linear, e

iterativamente para a parte de médias móveis, que não o é. Encontram-se na literatura referências à utilização de algoritmos iterativos para a obtenção de coeficientes a serem aplicados aos componentes MA do modelo.

3. Com o modelo obtido no passo 2 são gerados os resíduos da estimatimação, que, na hipótese de obtenção de sucesso na parametrização do modelo, conterão apenas ruído branco. A verificação nessa fase é a da aderência dos erros da estimatimação à distribuição de probabilidades inferidas, que na maior parte dos casos é a própria distribuição gaussiana. Não confirmada essa hipótese, retorna-se ao passo 2, ampliando os valores atribuídos a “p” e “q” na busca de melhoria na precisão do modelo.

4. Confirmada a aderência dos erros de estimatimação, o modelo está apto para ser empregado em projeções e previsões, observando-se que as previsões terão a sua confiabilidade dependente do tipo de função de distribuição de probabilidade e do grau de confiança requerida pela estimativa.

Modelos Autorregressivos

Os métodos de regressão assumem que as componentes são funções determinísticas, devendo-se estimar regressões com variáveis explicativas.

Chamamos a equação [1] de modelo autorregressivo de ordem “p” ou simplesmente AR(p)

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} \quad [1]$$

onde Y_{t-p} é o elemento da série temporal posicionado p períodos antes do elemento Y_t .

Modelos de Médias Móveis

Os métodos de médias móveis são métodos não causais, não se definindo variáveis explicativas, em que se assume que as componentes são funções alisadas do tempo, com um comportamento estocástico.

Considere a série Y_t , fazendo os Φ 's = $-\theta$'s no processo linear geral, e tomando a série como finita, chamamos de médias móveis de ordem “q” o modelo:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad [2]$$

Essa terminologia vem do fato de que Y_t é obtido aplicando-se os pesos 1, $-\theta_1$, $-\theta_2$, ..., $-\theta_q$, as variáveis a_t , a_{t-1} , a_{t-2} , ..., a_{t-q} e então movendo os mesmos pesos 1 unidade do tempo à frente e aplicando-lhes a a_{t+1} , a_t , a_{t-1} , ..., a_{t-q+1} para obter Y_{t+1} . onde os Y_{t-1} , Y_{t-2} , ..., Y_{t-p} são independentes de a_t . Os valores da série Y_t são uma combinação linear dos p valores passados mais um termo a_t , no qual incorpora coisas na série até o tempo t que não é explicado pelos valores passados.

Modelo Autorregressivo-Médias Móveis ARMA(p,q)

Se considerarmos uma série formada pelas partes autorregressivas e médias móveis, vamos ter um modelo mais geral de séries temporais, ou seja, se

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad [3]$$

diz-se que Y_t é um processo autorregressivo-médias móveis de ordens p e q , respectivamente, e parâmetros Φ s e θ s, ou abreviadamente ARMA(p,q).

Estacionamento da série

Uma série não estacionária não pode ser estimada trivialmente por ser impossível calcular todos os seus momentos, tornando inválidas as inferências estatísticas. O modelo geral de uma série não estacionária tem 3(três) partes:

$$Y_t = \text{Tendência} + \text{componente estacionário} + \text{ruído} \quad [4]$$

$$Y_t = Y_0/\beta_0 + \beta_1 t + \sum_{i=1}^t \varepsilon_i + \varphi(L)\eta_t$$

Esse modelo é chamado de modelo de tendência estacionária, porque o seu valor flutua em torno de uma tendência determinística β_1 . Essa série poderia ser estacionada pela primeira diferença, porém essa transformação introduziria ruído por tomar o erro não inversível. É recomendável neste caso, sendo a série dependente de "t", estacioná-la usando a variável explicativa t (Bueno, 2008).

Para remover a tendência de uma série sendo essa tendência determinística, o procedimento é estimar Y_t contra o tempo e armazenar os resíduos, os quais constituem uma nova série, que será modelada em separado da seguinte forma:

1. Estimando por mínimos quadrados ordinários os parâmetros β_i da equação $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \dots + \beta_n t^n$
2. Estimando o modelo ARMA(p,q) a partir dos resíduos extraídos do modelo determinístico anterior.

Para a determinação da ordem “n” do modelo, devem ser usados os testes estatísticos convencionais t e F na forma de estimar-se um “n” máximo e experimentando se $\beta_{n \text{ máx}}$ é significativo segundo o teste “t”. Não o sendo, retira-se o termo t^n e repete-se o procedimento para t^{n-1} , t^{n-2} ...

Em seguida, deve ser confirmado se os coeficientes em conjunto são todos iguais a zero usando o teste F.

As 6(seis) séries foram testadas para $n=3$ e os testes se mostraram insignificantes para $n>1$.

3.2.1 Estimação dos coeficientes

Ajuste do processo autorregressivo

Para um processo de ordem p com média μ e n observações, os coeficientes são calculados pelos estimadores de máxima verossimilhança, o que coincide com o método dos mínimos quadrados quando a função de distribuição de probabilidade é gaussiana (Ehlers, 2007).

Tomando-se a função $Y_i = X_i \Phi_i + \varepsilon_i$ e a hipótese de normalidade dos erros, i.e., $\varepsilon \sim N(0, \sigma^2 I_{n-p})$, obtém-se uma função de verossimilhança aproximada dada por:

$$L(\Phi, \sigma_\varepsilon^2) \propto (\sigma_\varepsilon^2)^{-\frac{n-p}{2}} \exp \left\{ \frac{-\sigma_\varepsilon^{-2} (y - X\Phi)' (y - X\Phi)}{2} \right\}, \quad [5]$$

$$L(\Phi, \sigma_\varepsilon^2) \propto (\sigma_\varepsilon^2)^{-\frac{n-p}{2}} \exp \left\{ \frac{-\sigma_\varepsilon^{-2} (y - X\Phi)' (y - X\Phi)}{2} \right\},$$

$$\frac{\partial \log(L(\phi, \sigma_\epsilon^2))}{\partial \phi} = \frac{-\sigma_\epsilon^{-2} \partial(y-X\phi)'(y-X\phi)}{2 \partial \phi} \Rightarrow$$

$$\frac{\partial \log(L(\phi, \sigma_\epsilon^2))}{\partial \phi} = \frac{-\sigma_\epsilon^{-2} \partial(-2\phi'X'y + \phi'X'X\phi)}{2 \partial \phi} \Rightarrow$$

$$\frac{\partial \log(L(\phi, \sigma_\epsilon^2))}{\partial \phi} = \frac{-\sigma_\epsilon^{-2}}{2} (-2X'y + 2X'X\phi) \Rightarrow$$

$$\left| \frac{\partial \log(L(\phi, \sigma_\epsilon^2))}{\partial \phi} \right|_{\phi=\hat{\phi}} = \hat{\phi} = (X'X)^{-1} X'y \quad [6]$$

Outra forma de se chegar a esse mesmo resultado é resolvendo o sistema de equações representado pela equação matricial $Y_i = X_i \Phi_i + \epsilon_i$, onde:

$$Y_i = X_i * \Phi_i + \epsilon_i$$

$$\begin{pmatrix} Y_{p+1} \\ Y_{p+2} \\ \cdot \\ \cdot \\ Y_n \end{pmatrix} = \begin{pmatrix} Y_p & Y_{p-1} & \cdot & Y_1 \\ Y_{p+1} & Y_{n-3} & \cdot & Y_2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ Y_{n-1} & Y_{n-2} & \cdot & Y_{n-p} \end{pmatrix} * \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \cdot \\ \cdot \\ \Phi_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_p \end{pmatrix}$$

Executando as operações de álgebra matricial para determinar Φ_i , onde $\epsilon_i=0$ são:

$$X'y = X'X\phi_i \quad \text{e} \quad \phi_i = (X'X)^{-1}X'y \quad [7]$$

Usando-se o valor estimado de Φ_i na equação [1] calculam-se os resíduos:

$$\epsilon_t = Y_t - \sum_{j=1}^p \hat{\phi} X_{t-j} \quad \text{onde } t=p+1, \dots, n$$

A estimativa de mínimos quadrados da variância de erro é $\frac{1}{n-p} \sum_{p+1}^n \varepsilon_t^2$.

Ajuste do processo do processo de médias móveis

O problema de estimação dos parâmetros desse processo é mais complicado por serem os erros ε_t funções não lineares dos parâmetros Φ_1, \dots, Φ_q , não sendo possível obterem-se expressões analíticas para seus estimadores. Métodos computacionais iterativos são utilizados para o objetivo de obter-se a mínima soma dos quadrados residuais.

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Dado o modelo MA(q) o processo consiste em fixar os valores de $a_t, \theta_1, \dots, \theta_q$ e calcular os resíduos ε_t ,

$$\varepsilon_t = Y_t - a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

sequencialmente para $t=1, \dots, n$. Assumindo-se que $\varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{-q+1} = 0$ e substituindo $\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ pelo resíduos calculados:

$$\varepsilon_1 = Y_1 - a_1$$

$$\varepsilon_2 = Y_2 - a_2 - \theta_1 \varepsilon_1 = Y_2 - a_2 - \theta_1 Y_1 - \theta_1 a_1$$

...

Obtidos esses resíduos, calcula-se a soma de quadrados residuais $\sum_1^n \varepsilon_t^2$. Repetindo esse procedimento para $a_1, \theta_1, \dots, \theta_q$ variando em uma grade de pontos, pode-se selecionar a combinação que minimiza a soma de quadrados. Em termos práticos, se o procedimento de otimização leva a muitas iterações para convergir ou não, deve-se desconfiar dos estimadores (Ehlers, 2007).

Ajuste do processo ARMA

Os problemas de estimação para modelos ARMA são similares ao modelo MA, no sentido de que um procedimento iterativo se faz necessário, tendo em vista que os erros ε_t são funções não lineares dos coeficientes Φ_i e θ_i .

O processo em que se utilizam os resíduos como estimativa de variável regressora e se aplicam os estimadores de mínimos quadrados para a obtenção dos coeficientes Φ_i e θ_i , é indicado como eficiente algoritmo que atinge convergência entre 3 a 10 iterações (Aguirre, 2007).

Os passos do algoritmo são:

1. A partir da equação de regressão e dos dados disponíveis, monte a equação matricial $Y_i = X_i \Phi_i + \varepsilon_i$, como no método de mínimos quadrados, e determine $\Phi_i = (X^T X)^{-1} X^T Y$
2. Calcule o vetor de resíduos $\varepsilon_t = Y - X \hat{\phi}$
3. Faça $i=2$ (i indica o número de iterações).
4. Com ε_{t-1} monte a matriz estendida de regressores X^* e estime $\hat{\phi}^* = [X^{*T} X^*]^{-1} X^{*T} Y$
5. Determine o vetor de resíduos $\varepsilon_t = Y - X^* \hat{\phi}^*$
6. Faça $i=i+2$ e volte ao passo 4. Repita até convergir.

Análise dos resíduos

Após o modelo ter sido ajustado, deve-se verificar se ele fornece uma descrição adequada dos dados através do comportamento dos seus resíduos (valor observado – valor ajustado). Será considerado um bom ajuste se as premissas de validação do modelo forem razoavelmente atendidas no caso em que os resíduos se distribuam aleatoriamente em torno de zero, com a sua variância aproximadamente constante, e que não sejam correlacionadas. A ausência de constância observada na variância é denominada volatilidade, pela literatura de séries temporais, e pode ser tratada através da transformação nos dados. Assim como em outros modelos estatísticos, a ideia é que os resíduos denunciem distorções de tratamentos inadequados nos indicadores de qualidade dos ajustes.

3.2.2 Qualidade dos estimadores

A suposição básica de qualquer técnica de previsão de séries temporais é que o valor observado na série fica determinado por um padrão que se repete no tempo e por alguma influência aleatória. Isto significa dizer que, mesmo quando o padrão exato que caracteriza o comportamento da série temporal tenha sido isolado, algum desvio ainda existirá entre os valores da previsão e os valores realmente observados. Essa aleatoriedade não pode ser prevista, entretanto, se isolada, sua magnitude pode ser estimada e usada para determinar a variação ou erro entre as observações e previsões realizadas.

Para a avaliação da qualidade do modelo são propostas as seguintes análises sobre os resíduos:

1) O coeficiente de determinação (R^2) que representa a relação entre as variações explicadas pelo modelo e variação total dos valores da serie temporal relativamente a cada verba estudada pode ser representada por:

$$R^2 = \frac{SQSérie}{SQTtotal} = 1 - \frac{SQErro}{SQTtotal} \quad [8]$$

2) raiz quadrada do erro médio quadrático (RMSE – Root Mean Square Error) onde Y_t são os valores observados e \hat{Y}_t são os valores previstos:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x})^2}{n}} \quad [9]$$

3) Para verificação da normalidade dos resíduos são utilizados teste Jarque-Bera (Bueno, 2008) e os testes Kolmogorov-Smirnov e Lilliefors (Campos, 1979).

Teste Jarque-Bera: Para uma variável aleatória X tal que $E(X)=\mu$ e $\text{Var}(X)=\sigma^2$ definem-se os coeficientes de assimetria e curtose, via a determinação dos seus momentos amostrais

$$m_j = \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X})^j \quad [10]$$

e os estimadores de assimetria e curtose que são dados por, respectivamente:

$$A(x) = E\left(\frac{(X - \mu)^3}{\sigma^3}\right) \quad \therefore \quad \hat{A} = \frac{m_3}{\sqrt{m_2^3}} \quad [11]$$

e

$$K(x) = E\left(\frac{(X - \mu)^4}{\sigma^4}\right) \quad \therefore \quad \hat{K} = \frac{m_4}{\sqrt{m_2^4}} \quad [12]$$

Sob a hipótese de normalidade as variáveis aleatórias $\sqrt{n/6}\hat{A}$ e $\sqrt{n/24}(\hat{K} - 3)$ são independentes e têm distribuição assintótica $N(0,1)$, e assim a estatística

$$\frac{n\hat{A}^2}{6} + \frac{n(\hat{K} - 3)^2}{24}$$

tem distribuição assintótica χ^2 com 2 graus de liberdade e pode ser usada para testar a normalidade de Y .

Teste Lilliefors: Objetiva verificar a adaptação de uma distribuição desconhecida $F_0(x)$ a uma distribuição conhecida $F(x)$, cuja hipótese nula é admitir que uma amostra x_1, x_2, \dots, x_n retirada de uma população desconhecida pertence a uma função de distribuição conhecida. A sua vantagem em relação à aplicação do teste χ^2 (Qui-quadrado), também utilizado para essa finalidade, é a possibilidade de o mesmo poder ser aplicado para pequenas amostras. Os dados são tratados individualmente, o que evita a perda de informações por agrupamentos, fato comum nesse tipo de teste. Ao contrário do teste Kolmogorov-Smirnov que depende do conhecimento da média e variância da distribuição, o teste Lilliefors utiliza-se

diretamente da média e variâncias amostrais para a criação da variável padronizada $Z_i = \frac{x - \bar{x}}{s}$, cujo valor limite para uma “n” amostra maior que 30 e com o nível de significância 20% é $0.738/\sqrt{n}$. O teste consiste em verificar a variação máxima $F(Z_i)$, probabilidade acumulada para os valores da distribuição e $S(Z_i)$, variação linear representado por i/n , onde n é o número de intervalos da amostra.

4) Para a verificação da existência de covariância entre os resíduos, que garantiria serem estes ruídos brancos, atestando que o modelo identificou todos os componentes não aleatórios, é proposto o teste Ljung-Box. Esse teste estabelece como hipótese nula a inexistência de covariância entre os resíduos que estejam dentro do intervalo de defasagem “n” e o valor calculado da distribuição “Q” converge para uma distribuição qui-quadrada com “n” graus de liberdade.

$$Q = T(T + 2) \sum_{j=1}^n \frac{\hat{\rho}_j^2}{T - j} \quad [13]$$

onde T é o tamanho da amostra e n é a defasagem entre itens da série.

Esse teste busca investigar a evidência de que não haja correlações dos resíduos. Em caso de a hipótese nula ser rejeitada, concluiu-se que ainda existem nos resíduos valores passíveis de serem identificados e removidos por modelo ARMA (p,q) de ordem superior. A confirmação da hipótese nula é evidência de que os resíduos sejam ruídos brancos.

3.2.3 Fluxograma do Modelo

O fluxograma da Figura 4 resume todas as etapas do processo de projeção na proposta deste trabalho de desenvolver técnica de apoio ao processo de conciliação contábil de custos com mão de obra.

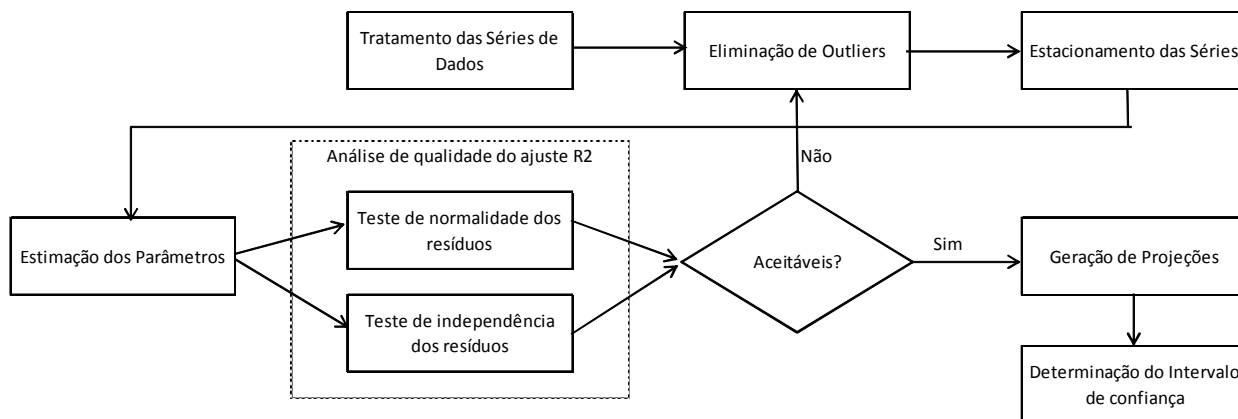


Figura 4 - Fluxograma do processo de projeção

Na etapa de estimação dos parâmetros do modelo ARMA, todos os algoritmos de otimização foram implementados no aplicativo MATLAB, cujos códigos foram incluídos no Anexo B.

3.3 ALGORITMOS GENÉTICOS

3.3.1 Histórico

As técnicas de metaheurísticas evolutivas tiveram o seu primeiro registro efetuado por Cannon (Cannon, 1932). Neste trabalho foi aventada a possibilidade de aplicar o processo baseado na teoria da evolução natural em máquinas de aprendizado - "*learning machines*". As primeiras tentativas de aplicação formal da teoria evolucionária apareceram para solucionar problemas práticos de engenharia na área de controle de processos, máquinas de aprendizado e otimização de funções. Um marco histórico relevante na utilização do mecanismo evolucionário para a resolução de problemas práticos trata-se da metodologia do EVOP – "evolutionary operation", formulada pelo famoso estatístico G.E.P. Box em 1950. Essa metodologia foi aplicada com sucesso na melhoria do processo gerencial de plantas industriais, sem, no entanto, até aquele momento, ter sido implementada em simulação computacional (Fogel, 2006).

Segundo Fogel (Fogel, 2006), os métodos de simulação evolucionária, hoje conhecidos como algoritmos genéticos, aparecem no curso de uma década através de três manifestações independentes: Frazer (Fraser, 1957), Bremermann (Bremermann, 1958) e na publicação de John Holland's da Universidade de Michigan (Holland, 1975).

3.3.2 Definição

A evolução via seleção natural de uma população aleatoriamente escolhida segundo a influência da sua capacidade de adaptação pode ser entendida como mecanismo de busca no espaço compreendido pelos possíveis valores dos cromossomos. Neste sentido, os algoritmos evolutivos (AE) são mecanismos estocásticos de solução ótima para um determinado problema. Os operadores de um algoritmo genético são Inicialização, Avaliação, Seleção, Combinação e Mutação.

Os algoritmos genéticos, AG, são uma metaheurística de busca e otimização de soluções. Pertencem à família dos AE, os quais aplicam técnicas inspiradas na Biologia na resolução de modelos computacionais. Consideram uma população inicial de indivíduos, correspondendo, cada um destes, a um candidato à solução do problema. Os indivíduos de cada geração competem entre si, de forma a ter lugar na geração seguinte, diretamente ou através de um descendente seu (vide Figura 5).

Os indivíduos mais fortes, isto é, cuja função de avaliação de desempenho, *fitness*, apresentar melhores valores, têm maior probabilidade de serem escolhidos para progenitores da geração seguinte. São emprestadas da Biologia noções e definições tais como cromossomos, genes (elementos individuais que constituem um cromossoma) e alelos (os possíveis valores que um determinado gene pode tomar). Cada gene está localizado numa posição fixa (*locus*) do cromossoma. Geralmente, os cromossomos têm um número de genes e cada gene tem uma localização definida dentro do cromossomo. Por esse motivo, os cromossomos são representados por vetores, sendo os genes elementos desses vetores. Nos algoritmos genéticos, o cruzamento (combinação) entre soluções pode ser conduzido utilizando diferentes procedimentos que dependem, por exemplo, da representação do cromossomo escolhido, do número de progenitores, do número de descendentes. Duas soluções progenitoras serão recombinadas, dando origem a dois descendentes.

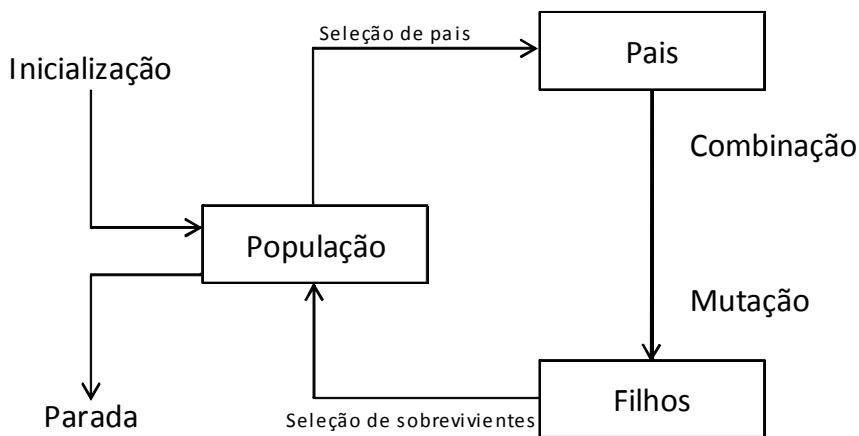


Figura 5 - Esquema genérico de um algoritmo evolucionário (Smith, 2003)

Outro tipo de operador para a produção de expansão da pesquisa na área de busca é operador de Mutação, que atua introduzindo mudanças aleatórias no cromossoma dos indivíduos na expectativa do aparecimento de indivíduos mais aptos.

As variações nas novas populações advindas da mutação podem ter influência negativa ou positiva nas habilidades de sobrevivência e reprodução dos indivíduos. Sem mutação, a população tenderia a convergir para estagnação em que cada indivíduo diferiria muito pouco uns dos outros. Dada uma probabilidade predefinida, os parâmetros de cada indivíduo poderão sofrer alterações, o que poderá traduzir-se na fuga de um candidato a um extremo local da função de avaliação de desempenho ampliando, assim, a pesquisa sobre o espaço de busca. O aumento exagerado da taxa de mutação, entretanto, distorce a estrutura genética, que pode implicar a perda de bom material genético de indivíduos mais aptos. Entretanto, pesquisas mostraram que uma taxa de mutação inicialmente elevada que decresça exponencialmente com o aumento do número de mutações pode vir a favorecer a velocidade de convergência e a acurácia do algoritmo (Engelbrecht, 2007).

Cada sequência de operações formada pela seleção, cruzamento e mutação, dá origem a uma nova geração, seguida de sua avaliação de desempenho. O ciclo de evolução é interrompido quando uma das condições de parada é satisfeita. Os critérios de parada geralmente empregados são (a) número de gerações sem melhoras, (b) erro relativo ou absoluto entre gerações e (c) número máximo de iterações.

O processo matemático de seleção para reprodução, tomando por base a sua aptidão relativa, pode não garantir uma convergência assintótica para o ótimo global. O melhor cromossomo de uma população tem probabilidade de ser perdido na geração seguinte, não assegurando que o ganho obtido em uma geração seja mantido nas gerações futuras. Para evitar esse problema, é possível aplicar heurística denominada Elitismo, em que sempre são conservados para combinação os melhores indivíduos de uma população. Esse procedimento garante a convergência assintótica no processo de busca (Fogel, 2006).

Segundo (Engelbrecht, 2007), os algoritmos genéticos (AG) têm sido aplicados a uma vasta gama de problemas:

- Planejamento- rotinas de otimização e agendamento.
- Projetos de filtros, arquitetura de redes neurais e otimização estrutural.
- Controladores de turbinas a gás, sistema de direcionamento visual de robôs.
- Classificação e agrupamento.
- Funções de aproximação e modelagem de séries-temporais .
- Regressões.
- Mineração de dados.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Neste capítulo são apresentadas as ações desenvolvidas na análise dos custos, no intuito de identificar e corrigir os elementos determinísticos, além da análise estatística de correlações entre essas séries. Nas situações em que foram identificadas significativas correlações, foi procedido o agrupamento das séries com a consequente economia de processamento computacional.

4.1 Análise de Dados

Foram aplicados no processo de eliminação dos componentes determinísticos a extração de “*outliers*”, a dessazonalização e o deflacionamento das dez séries temporais representadas por Salários e Ordenados, Décimo Terceiro, FGTS, INSS, PIS, Abono de Férias, Férias, Aviso Prévio Indenizado, Férias e Outros Indiretos, além da série de custos transferidos.

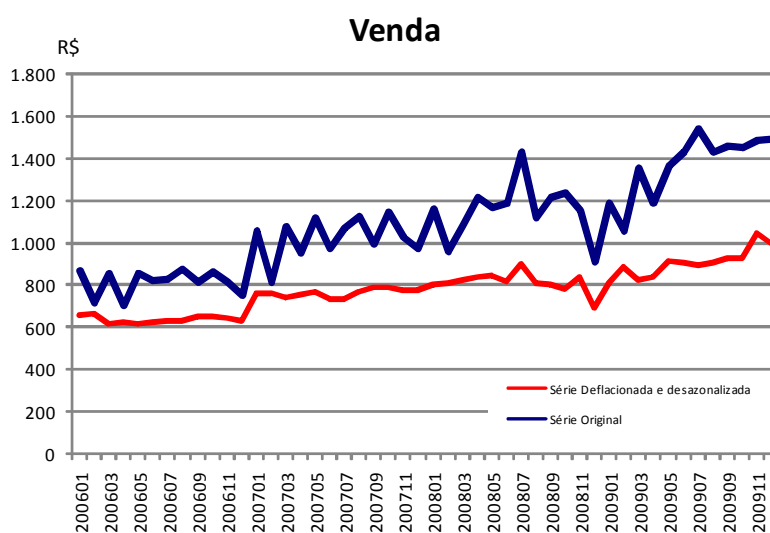
4.1.1 Tratamento das séries temporais

Seguindo o modelo de controle proposto na Figura 1, as séries temporais em observação foram divididas em dois grupos. O primeiro trata dos custos transferidos pelo modelo para os projetos que consumiram as horas dos profissionais. O segundo trata os custos incorridos pelo modelo.

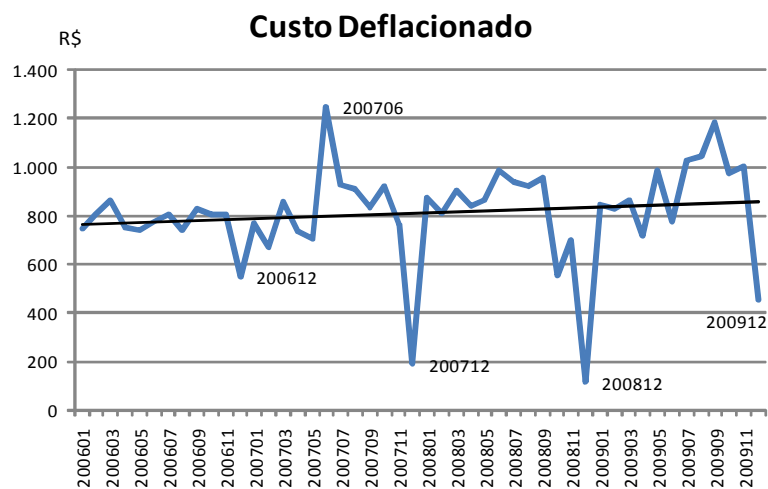
O primeiro grupo, metaforicamente denominado Venda, refere-se aos custos a serem transferidos aos projetos, os quais são definidos pelo produto das horas, (h_{ji}), trabalhadas por um profissional em um projeto, e o custo padrão (CP_i), que é fixado anualmente para cada um dos recursos humanos. O h_{ji} é o número de horas que esse profissional “i” contribuiu para um determinado projeto “j”, ou seja, o número de horas que o projeto venha a utilizar desse recurso.

O segundo grupo, metaforicamente denominado Compra, refere-se ao efetivo custo incorrido, que é composto por salário, encargos e benefícios despendidos pela empresa para a sua manutenção. Um sistema de controle ideal de gestão de custos de mão de obra deve dimensionar o custo/hora dos profissionais de forma a absorver todos os custos despendidos, equilibrando-os, no final do exercício fiscal, com os valores transferidos para os projetos que se utilizaram desses recursos, em observância do princípio da competência contábil.

Para a análise comparativa do comportamento dos custos ao longo do tempo são tomados valores reais na formação das séries temporais. Essas séries são a seguir tratadas por um índice de inflação, aqui denominado deflator, de forma a atualizar os seus valores na data-base de fevereiro de 2009. Esse índice de reajuste de preço, diretamente proporcional à atualização dos salários, corresponde aos reajustes obtidos nas negociações de acordo coletivo de trabalho da categoria, negociados nos meses de fevereiro de 2006 a fevereiro de 2009, relativos aos respectivos anos (vide Anexo A). Para eliminação do efeito da heterogeneidade no número de dias dos meses, dentro de um mesmo ano, as séries de venda são corrigidas com a aplicação do índice de dessazonalização, que corresponde à relação entre o número de dias do mês em referência e a média de dias do referido ano. O resultado desses tratamentos encontra-se na Figura 6 (a,b,c) abaixo:



(a)



(b)



(c)

Figura 6 (a,b,c) - Tratamento de Deflação e Dessazonalização

O resultado desse tratamento produziu a diminuição da declividade da série, quando aplicada a deflação, pois os preços fixados na data-base de fevereiro de 2009 e o efeito do tratamento de dessazonalização podem ser visualmente percebidos pelo amaciamento, diminuição do número de picos apresentados nas séries (Figura 7).

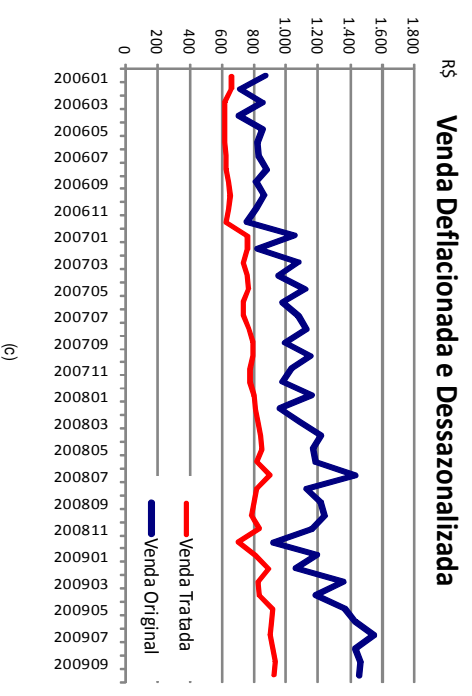
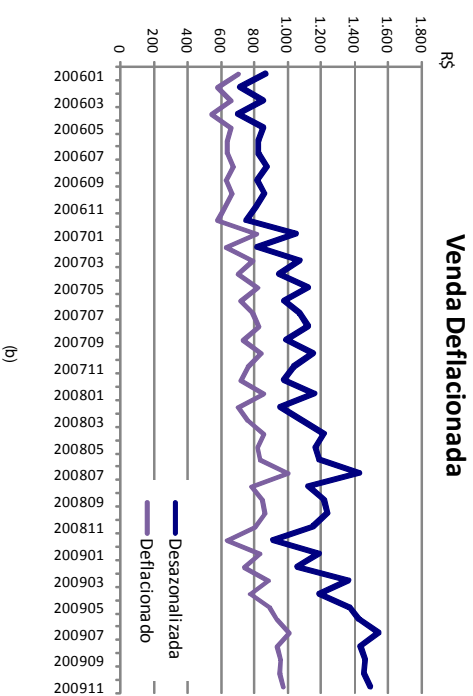
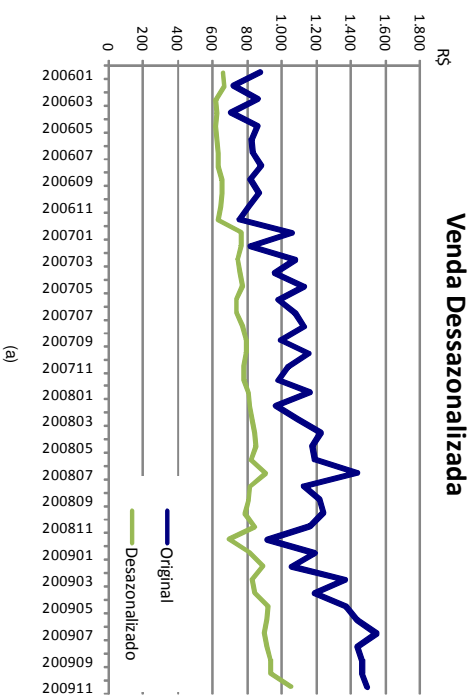


Figura 7 (a,b,c) - Tratamento de estabilização das séries

Quanto à análise dos custos, foram identificados visualmente grandes variações nos meses de janeiro de 2006, julho de 2006, dezembro de 2007 e dezembro de 2008. Foi expurgado o lançamento de ajuste de janeiro de 2006 por referir-se a custos do exercício de 2005. O gráfico (Figura 6c) de resultado, que resume a diferença entre custos e vendas (custos transferidos), apresenta um desequilíbrio evidenciado pela existência de uma componente de tendência, em que os custos incorridos comportaram-se ligeiramente inferiores aos custos transferidos. Para a decomposição da série, fez-se necessária a eliminação do componente de crescimento antes do processo de estimação de parâmetros. Dois são os tratamentos recomendados: o estacionamento da série com a utilização da diferenciação da série, calculando-se nova série a partir das diferenças, ou a incorporação do elemento de crescimento, tendência, na equação do modelo que passa a ter o componente determinístico dependente do tempo (β_1). A adoção da primeira alternativa apresenta o inconveniente de introduzir ruído por tornar os erros não invertíveis (Bueno, 2008) e a dificuldade na leitura do fenômeno físico observado, custo mensal calculado na base de preço de fevereiro de 2009.

Foram calculados os coeficientes de tendência para cada uma das séries no período de fevereiro de 2006 e dezembro de 2009 deduzindo do valor das séries o componente determinístico, que é o custo explicado, componente de tendência de cada série. O objetivo desse tratamento é o estacionamento da série para que sejam promovidas as inferências estatísticas.

4.1.2 Identificação de correlações entre séries

Análise de correlacionamentos

Para a análise do inter-relacionamento das nove séries foram calculadas as matrizes de covariância e de correlação, sendo possível identificar a irrelevância de algumas das verbas para a variação total do custo, assim como a alta correlação observada entre algumas verbas em valores superiores a 70%. Essa análise levou à decisão de agruparem-se as verbas de P. I. S., Provisão e Abono de Férias, Aviso

Prévio Indenizado à verba de outros custos indiretos, que contêm custo de natureza aleatória para pagamento de ajuda alimentação, despesas com saúde etc. As três verbas citadas representam menos de 1% da média de custos e menos de 3% da variância observada. Foi possível observar, ainda, na alta correlação entre os outros custos indiretos e o custo total a representatividade da sua variância e variância total, levando a acreditar que a eficiência do modelo em equilibrar custo e venda é influenciada fortemente por esse item de custo indireto. Representando na média 26% do custo com mão de obra, a variância observada nos outros custos indiretos é de 77% da variância observada no total do custo.

Após os ajustes de agrupamento referidos, a matriz de covariância e de correlação é apresentada na Tabela 2, a seguir:

	Matriz de Covariância								Matriz de Correlação							
	Venda (a)	Salários e Ordenados P. Admin.	Décimo Terceiro Salário	F. G. T. S. e I. N. S. S.	Provisão e Abono de Férias	Outros Indiretos PIS-AVP-FI	Custo (b)	Resultado (c)= (a)- (b)	Venda (a)	Salários e Ordenados P. Admin.	Décimo Terceiro Salário	F. G. T. S. e I. N. S. S.	Provisão e Abono de Férias	Outros Indiretos PIS-AVP-FI	Custo (b)	Resultado (c)= (a)- (b)
Venda (a)	27448	9342	434	539	1192	9142	20649	6799	1,00	0,63	0,24	0,04	0,26	0,21	0,41	0,15
Salários e Ordenados P. Admin.	9342	8113	264	-379	234	947	9178	164	0,63	1,00	0,26	(0,06)	0,09	0,04	0,34	0,01
Décimo Terceiro Salário	434	264	124	-152	31	-1414	-1147	1581	0,24	0,26	1,00	(0,18)	0,10	(0,48)	(0,34)	0,51
F. G. T. S. e I. N. S. S.	539	-379	-152	5797	399	2991	8655	-8116	0,04	(0,06)	(0,18)	1,00	0,19	0,15	0,38	(0,38)
Provisão e Abono de Férias	1192	234	31	399	765	178	1607	-415	0,26	0,09	0,10	0,19	1,00	0,02	0,19	(0,05)
Outros Indiretos PIS-AVP-FI	9142	947	-1414	2991	178	70516	73218	-64076	0,21	0,04	(0,48)	0,15	0,02	1,00	0,91	(0,87)
Custo (b)	20649	9178	-1147	8655	1607	73218	91512	-70863	0,41	0,34	(0,34)	0,38	0,19	0,91	1,00	(0,84)
Resultado (c)= (a)- (b)	6799	164	1581	-8116	-415	-64076	-70863	77662	0,15	0,01	0,51	(0,38)	(0,05)	(0,87)	(0,84)	1,00

Tabela 2 - Matriz de covariância e correlação das séries

4.1.3 Tratamento de “outliers”

Um *outlier* pode ser definido como um ponto fora da série temporal que seja significativamente diferente dos demais. Ele é uma observação não usual que afeta a análise dos dados e, portanto, precisa ser tratado.

Estimativas utilizando estatística clássica como médias amostrais, variâncias, covariâncias e correlações, ou ajustes utilizando modelos de regressão de mínimos

quadrados podem ser influenciados de forma adversa por *outliers*, mesmo que seja apenas um, e falhar para prover bons ajustes para amostras (Yohai, 2006).

Ocorrem dois tipos de *outliers* nas séries temporais: os que têm origem em erros de medição causados por fatores externos ao sistema e são denominados *outliers* aditivos e os que têm origem nas mudanças intrínsecas do processo, cuja denominação é *outlier* inovativo. A diferença entre estes é que o segundo indica início da mudança de tendência do processo e pode eventualmente retornar ao estado original anterior. Essas alterações levam a outras na média do processo e passam a determinar medidas que o desestabilizam, deixando até mesmo de ser um processo estacionário. Não se pode ignorar, também, a possibilidade de esses dois tipos ocorrerem ao mesmo tempo, tornando a tarefa de detecção mais difícil.

Detectados e classificados os *outliers*, o problema seguinte passa a ser decidir o tratamento a ser empregado nos mesmos.

O principal problema na estratégia tradicional de detecção de *outliers* é que estes são dependentes do modelo. Uma vez conhecido o tipo de *outliers*, o modelo que o incorpora também pode ser estimado. Isto nem sempre é prático quando se lida com grandes séries de dados cujo processo de leitura está sujeito a frequentes mudanças. Neste caso, as séries temporais resultantes podem ser altamente instáveis, impossibilitando a aplicação do modelo de dependência. Para a eliminação desse problema nas séries estudadas nesta pesquisa, o componente determinístico de estabilização será aplicado após a extração dos *outliers*.

O modelo para a limpeza da série significa a substituição dos valores identificados por valores razoáveis que representem valores esperados. O método utilizado para ambas as aplicações, identificação e substituição, foi o uso de valor da mediana na vizinhança dos pontos em processo de teste de tolerância ao critério de aceitação. Em caso de não aceitação, procedeu-se à substituição do valor deste pelo valor da mediana no intervalo de “k” pontos antes e depois do ponto em apreciação $\eta_t^{(k)} = \{Y_{t-k}, \dots, Y_t, \dots, Y_{t+k}\}$. Esse método é denominado “*two-side median*” (Meckesheimer, 2007) por utilizar no cálculo da mediana k pontos à direita e k pontos à esquerda. O outro parâmetro do processo de limpeza é o critério de

rejeição ou aceitação do que se define pelo afastamento absoluto “ τ ” em relação à mediana na forma mostrada no Figura 8a, a seguir. Para o tratamento das séries foram utilizados como critério de aceitação valores com afastamento máximo de 1,28 desvios-padrão no pressuposto de ser uma distribuição normal, significando o intervalo conter 90% dos pontos das séries.

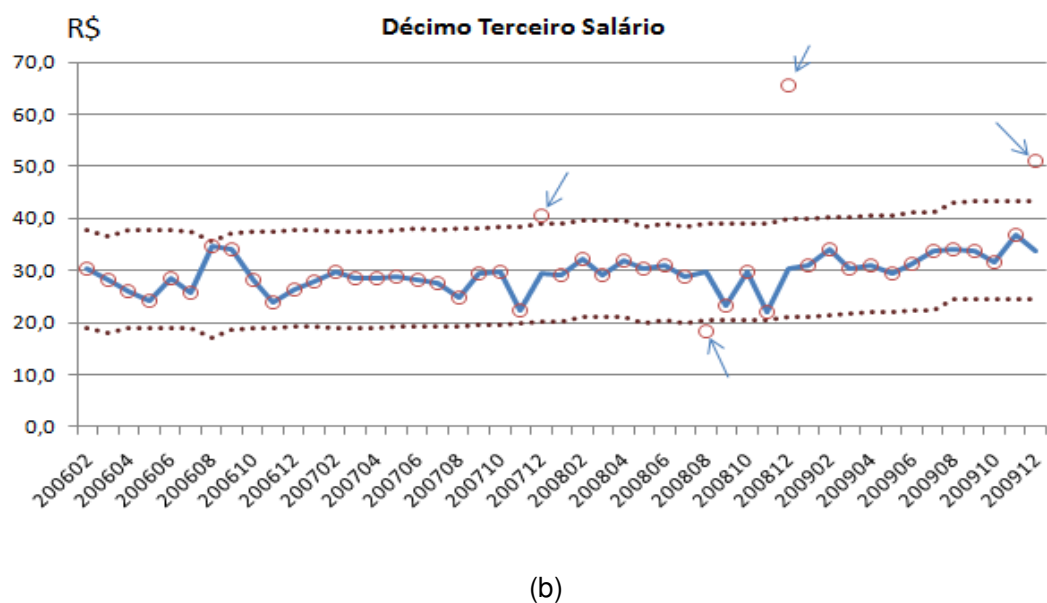
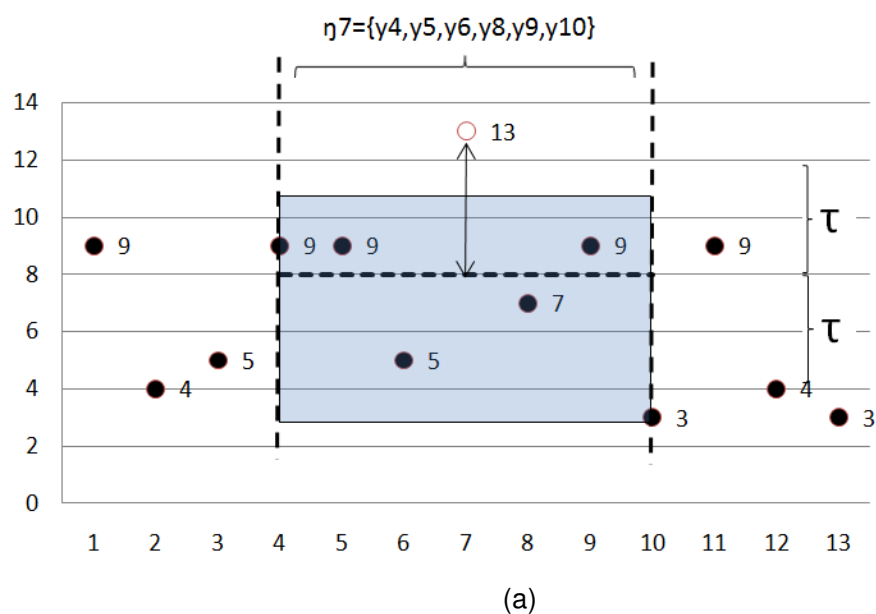


Figura 8 (a,b) - Método de tratamento de “Outliers”

4.1.4 Estacionamento das séries

O estacionamento das seis séries foi efetuado utilizando função de primeiro, segundo e terceiro grau. O teste de significância t-student entre as diferentes séries estimadas por essas funções resultaram em 100% de compatibilidade, confirmando a adequação do modelo linear para a extração de tendência das séries (Figura 9).

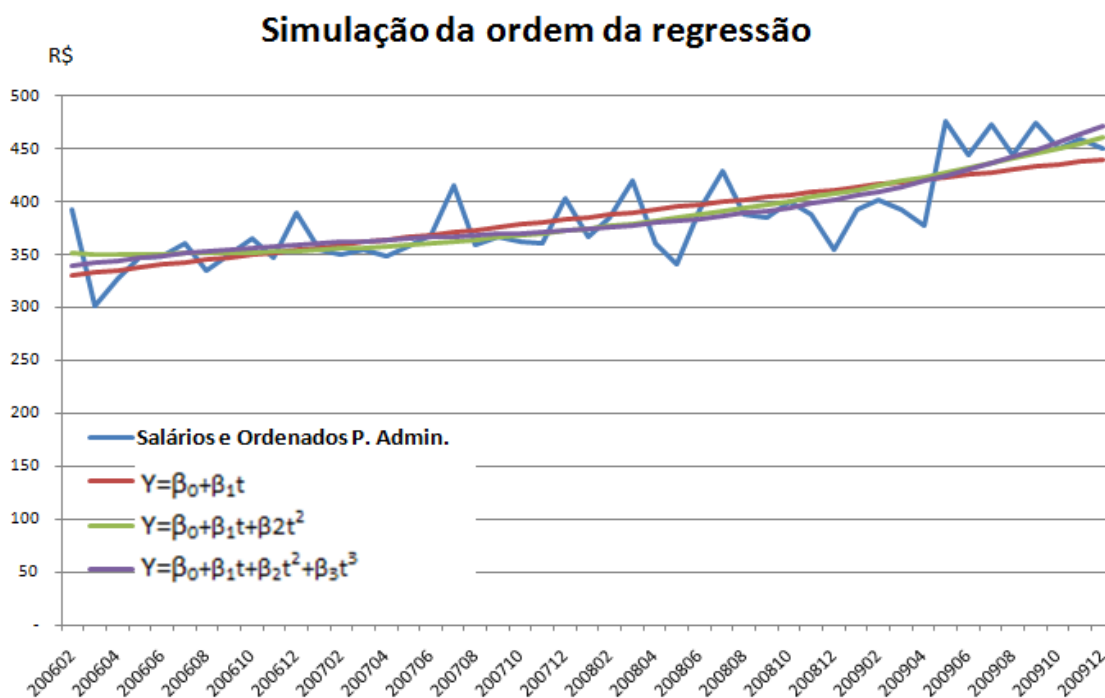


Figura 9 - Exemplo de modelos de estacionamento das séries

4.2 Implementação dos Algoritmos

Nesta seção são detalhados os procedimentos utilizados na transferência dos dados tratados para os modelos de previsão, mostrando explicitamente as premissas utilizadas na obtenção dos coeficientes Θ_i e Φ_i do modelo.

4.2.1 Definição

O objetivo do programa é, primeiramente, identificar o modelo que melhor se ajusta à descrição do fenômeno, série temporal e seus componentes, que permita, da melhor forma possível, representá-lo, isto é, minimizando erros e mantendo as propriedades estatísticas dos resíduos, para vir a suportar inferências nas etapas de projeções das séries, que é o objetivo final do processo de conciliação. O modelo proposto é um modelo linear que se utiliza dos valores passados e suas diferenças para as médias. Para a determinação dos coeficientes, foram utilizadas as abordagens de máxima verossimilhança da função de distribuição de probabilidade como sendo normal, que coincide com o processo de mínimos quadrados ordinários. Na busca de estimadores do modelo de séries temporais, foram desenvolvidos 3 (três) algoritmos distintos para que explorassem soluções para cada uma das séries analisadas e os seus resultados fossem comparados entre si. O algoritmo genético foi desenvolvido de forma parametrizada, para permitir a análise dos resultados e a velocidade de convergência com diferentes combinações de parâmetros.

Montagem da Matriz de Regressão

O modelo ARMA define uma combinação de “p” períodos de regressão, ou seja, influenciam o valor de Y_i os “p” valores antecessores deste, os quais são tratados como as variáveis independentes de um modelo de regressão múltipla e cuja determinação dos fatores/coeficientes é obtida com o auxílio de algoritmos genéticos (AG) e pelo método dos mínimos quadrados com o uso do cálculo matricial $\beta = (X'X)^{-1}(X'Y)$.

	p=3			q=3		
Y_{i-6}					* $\mathcal{E}_{i-1}=Y_{i-1}-\hat{Y}_{i-1}$	
Y_{i-5}	Y_{i-5}			* \mathcal{E}_{i-5}		
Y_{i-4}	Y_{i-5}	Y_{i-6}		\mathcal{E}_{i-5}	\mathcal{E}_{i-6}	
Y_{i-3}	Y_{i-4}	Y_{i-5}	Y_{i-6}	\mathcal{E}_{i-4}	\mathcal{E}_{i-5}	\mathcal{E}_{i-6}
Y_{i-2}	Y_{i-3}	Y_{i-4}	Y_{i-5}	\mathcal{E}_{i-3}	\mathcal{E}_{i-4}	\mathcal{E}_{i-5}
Y_{i-1}	Y_{i-2}	Y_{i-3}	Y_{i-4}	\mathcal{E}_{i-2}	\mathcal{E}_{i-3}	\mathcal{E}_{i-4}
Y_i	Y_{i-1}	Y_{i-2}	Y_{i-3}	\mathcal{E}_{i-1}	\mathcal{E}_{i-2}	\mathcal{E}_{i-3}

X_p
 X_q

Figura 10 - Matriz de Regressão

A inclusão das diferenças entre Y_i e a média com a defasagem de “q” períodos melhora a qualidade do modelo e a estimação dos seus coeficientes também segue os dois métodos citados.

O modelo Box&Jenkins tem por orientação a observância do princípio da parcimônia, na busca de coeficientes limitando o seu número ao que realmente traga melhorias/progresso no processo de estimação. Quando o acréscimo no número de coeficientes não apresentar melhora no resultado dos erros, termina o processo de busca.

Neste trabalho, tendo em vista a disponibilidade de recursos computacionais, foi executada, para cada série, a cobertura de todas as combinações possíveis de “p” e “q”, limitados a 12 períodos de defasagem cada. A razão para o estabelecimento desse limite está na periodicidade do exercício fiscal, em que os eventos estarão encontrando forte correlação, o que se confirma na análise gráfica das séries.

4.2.2 Abordagem de ARMA

Foram testados para a estimação dos parâmetros de regressão (θ) e de médias móveis (Φ) dois algoritmos distintos, sendo o primeiro, denominado ARMA, iterativo, que se utiliza do erro entre o valor estimado pelo período t-1 para o

cálculo de \hat{y}_t . O segundo modelo utilizou-se da matriz de regressão com erros de média móvel (Figura 10) para obter os coeficientes do modelo regredindo a referida matriz. A vantagem do primeiro modelo sobre o segundo é a melhor qualidade de estimação, perdendo-se, no entanto, por ser um processo iterativo, a possibilidade de explicitar ou descrever os valores de \hat{y} em única fórmula matemática.

```

Faça enquanto  $\sum \varepsilon_t, \sum \varepsilon_{t-1} < \text{Precisão}$ 
  -Extraia a Matriz  $X(n-\text{máx}(p,q), p+q)$ 
  -Calcule os coeficientes  $\alpha$  e  $\beta = (X'X)^{-1}(X'Y)$ 
  - Se primeiro loop Salve solução R2ARME
  - Salve solução R2ARME e R2ARMA
Fim
- Salve solução R2ARMA e R2ARME

```

Figura 11 - Pseudocódigo ARMA

O algoritmo ARME executa os cálculos dos estimadores para o modelo estático ARMA na primeira execução do laço, enquanto o algoritmo ARME executa o mesmo laço até que o processo não evolua em melhoria de precisão, o que significa o crescimento de R^2 durante a execução do algoritmo. Os resultados obtidos pelo processo iterativo foram significativamente superiores ao modelo de estático e, mesmo com a desvantagem da não representação do modelo por fórmula definida por Y_i e ε_i , justificaram a sua utilização.

4.2.3 Abordagem Algoritmos Genéticos

REPRESENTAÇÃO E INICIALIZAÇÃO

O espaço de busca foi estabelecido através da fixação dos limites entre -1 e 1, para cada um dos coeficientes, tanto para os coeficientes de regressão quanto para os coeficientes de médias móveis. A representação genética do indivíduo foi estabelecida como representação "Real" com 24 genes, um gene para cada coeficiente. A população inicial dos algoritmos genéticos foi gerada randomicamente respeitando o domínio de busca e as defasagens de, no máximo, 12 períodos para a regressão e para as médias móveis.

01	02	03	04	05	06	07	08	12	13	14	15	16	17	18	24
$\Phi 1$	$\Phi 2$	$\Phi 3$	$\Phi 4$	$\Phi 5$	$\Phi 6$	$\Phi 7$	$\Phi 8$	$\Phi 12$	$\theta 1$	$\theta 2$	$\theta 3$	$\theta 4$	$\theta 5$	$\theta 6$	$\theta 12$

Figura 12 - Representação dos coeficientes no cromossomo

SELEÇÃO

Foi adotada a seleção proporcional, que é um método de seleção aplicado aos Algoritmos Evolutivos e especifica a probabilidade de que cada indivíduo seja selecionado para a próxima geração. Os indivíduos com maior aptidão possuem uma probabilidade maior de serem selecionados para a próxima geração, e este pode ser considerado, então, como um problema de maximização.

No processo de seleção foi adotado o procedimento de transferir o melhor indivíduo da população para a população seguinte. A essa escolha dá-se o nome de elitismo, que tem por objetivo privilegiar a melhor solução, de forma que este indivíduo propague as suas características para a população seguinte. Para seleção dos indivíduos que participam da recombinação genética, foi utilizado o método de seleção SUS – Sistema Universal de Seleção –, também conhecido por roleta, em que os indivíduos mais aptos, com maior valor de “*Fitness*”, têm maior chance de serem selecionados para combinarem-se na geração de novos indivíduos.

FUNÇÃO DE APTIDÃO

Foi utilizado o indicador de qualidade R^2 , que representa a relação da soma quadrada dos erros de estimação e a soma dos quadrados das diferenças dos valores da série em relação à sua média. Esse indicador é denominado na literatura coeficiente de determinação, e o seu valor deve estar situado sempre entre 0(zero) e 1(um). O modelo será tanto melhor quanto R^2 aproxime de 1(um). O significado dessa proximidade é a capacidade do modelo em explicar as variações observadas nas séries.

COMBINAÇÃO

O programa foi parametrizado para permitir que a troca de material genético entre dois pais fosse totalmente aproveitado, na hipótese de que dois pais gerassem dois filhos, ou descartado, na situação do material genético na operação de troca (*cross-over*) para geração de único filho (*offspring*). Em outro operador de recombinação parametrizado foi introduzida a possibilidade de selecionar um percentual dos indivíduos para reprodução, que se transferirá para as gerações futuras sem alteração do seu material genético.

MUTAÇÃO

O operador de mutação foi parametrizado para variar durante o processo de convergência, sendo desejáveis índices maiores no início da execução do programa para o aproveitamento da característica de aceleração da convergência, estando o espaço de busca ampliado. Com a aproximação do critério de parada, número determinado de iterações, decresceu o índice de mutação para melhorar a varredura de ótimos locais sem dispersão, garantindo as conquistas obtidas. Também foi previsto no programa a possibilidade de parametrizar o algoritmo com taxa de mutação fixa. Os resultados obtidos por experimentação na combinação dos parâmetros são reportados no Anexo C, para os quais foram utilizados os seguintes parâmetros e definições para o algoritmo (Tabela 3):

Módulo de População	Operadores
Técnica de Representação	<i>Real</i>
Técnica Inicialização da População	<i>Progressiva</i>
Técnica Eliminação da População	<i>Elitismo 5-40%</i>
Técnica de Reprodução	<i>Troca da geração</i>
Técnica de Seleção de Genitores	<i>Proporcional/Roleta</i>
Técnica de Aptidão	R^2
Tamanho da população	<i>100</i>
Número de Gerações(parada)	<i>50-300</i>
Operadores	<i>Recombinação linear</i>
Taxa Mutação	<i>5-40%</i>
Taxa Crossover	<i>50-100%</i>

Tabela 3 - Parametrização do algoritmo genético

5 RESULTADOS OBTIDOS

Neste capítulo são apresentados os resultados obtidos pelos três diferentes métodos de estimação: AG, AR e ARMA. Os resultados do modelo AG, por sua característica não determinística, demandaram a realização de experimentos estatísticos para a obtenção de parâmetros de calibração do algoritmo que conduzissem a resultados mais assertivos. A comparação dos resultados obtidos pelos três algoritmos foi realizada com a utilização de indicadores de ajuste R^2 , testes de normalidade dos resíduos e teste de eliminação de autocorrelacionamento nas séries estimadas.

5.1.1 Descrição dos experimentos

Método Algoritmo Genético (AG)

Para avaliação da eficiência do algoritmo genético foram geradas para cada uma das 6(seis) séries experimentos com parametrização de regressão para a situação mais desfavorável para o AG, ou seja, com 12 coeficientes ($p=12$) combinando os parâmetros de mutação(mPM), seleção(mPS) e elitismo(mPE). A razão dessa escolha é que, sendo o método AR um processo determinístico, o seu resultado é sempre ótimo e conhecido, consistindo dessa forma referência para a avaliação de desempenho desse método. Os resultados estão mostrados no Anexo C, no qual é possível observar a rápida convergência do algoritmo para atingir os valores ótimos de regressão com um pequeno número de iterações. Em menos de 150 iterações, o resultado já se posicionou bastante próximo dos valores máximos obtidos. Nota-se, também, a pequena sensibilidade aos parâmetros de elitismos e de seleção. O comportamento do algoritmo para ocorrências superiores a 300 iterações se mostrou pouco compensadora, já que o ganho de qualidade no crescimento de R^2 resultou muito pequeno, significando pequena melhoria no ajuste do modelo. No entanto, o parâmetro de mutação do modelo apresentou-se bastante influente, notando-se a aceleração da convergência com a variação desse parâmetro.

Os resultados observados na análise de sensibilidade da taxa de mutação no AG recomendaram o aprofundamento da verificação da influência desse parâmetro no desempenho do algoritmo. Para essa finalidade foi executado ensaio sobre a série “1” parametrizando-se o algoritmo com taxas fixas de elitismo em 10%, taxa de combinação em 50% e variando a taxa de mutação entre 0,1% até 95% em intervalos de 1% conforme Tabela 9 (Anexo C). Os resultados, apresentados na Figura 13, indicaram que a eficiência do algoritmo fica comprometida quando se utilizam níveis muito baixos da taxa de mutação, valores inferiores a 1%, assim como quando se utilizam níveis muito altos desse parâmetro, valores superiores a 40%. Também é possível observar na mesma figura que a variação dos resultados de R^2 acima desse limite tende a aumentar, indicando que, além da perda de eficiência, o excesso mutação, mecanismo utilizado para escapar de ótimos locais, provoca desvios que não permitem a convergência do algoritmo. O algoritmo genético na determinação dos coeficientes do modelo é bastante robusto, e os testes apresentados na Tabela 10 atenderam a 78% dos resultados para os experimentos com 50, 100,...,300 iterações, admitindo uma tolerância de 5% de erro em relação ao valor ótimo de $R^2=0,4825$.

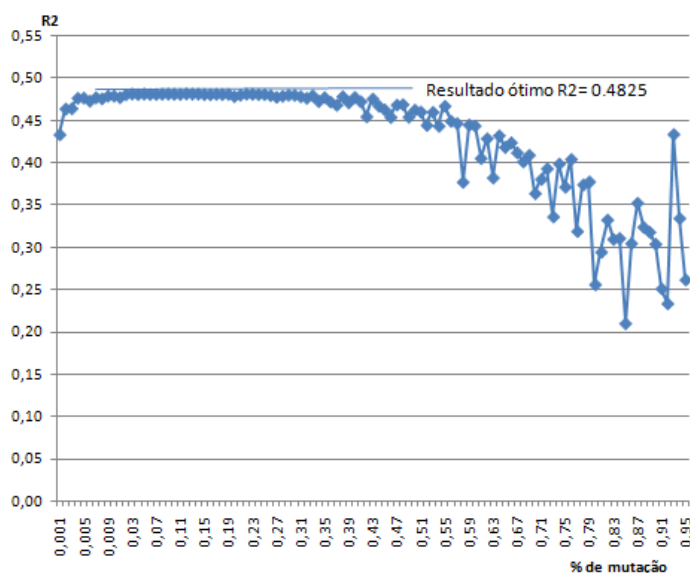


Figura 13 - Análise de sensibilidade à taxa de mutação

Método Autorregressivo(AR)

Os resultados obtidos com regressão de mínimos quadrados, os quais independem do erro de estimação, é uma particularização do modelo ARMA, no qual $q=0$. Por ser o processo mais simples, necessitando apenas da estimação dos coeficientes Φ_i , demanda menor número de cálculos e pode ser estimado diretamente de uma planilha, dispensando a necessidade de processos iterativos. Outra vantagem da simplicidade desse método é o menor descarte de dados quando se aumenta o número de ordem do modelo ($q=0$). A relevância dessa característica é mais sensível quando se manuseiam séries curtas, em pequeno número de períodos (Levine, et al., 2008). O método AR é preferível aos processos mais sofisticados quando se efetua análise preliminar da série temporal, para uso em resultados provisórios de caracterização dos dados, por exemplo, tratamento de “*outliers*”.

Em todos os tratamentos confirmou-se o fato esperado de que a sofisticação do modelo com aumento do número de parâmetros “ q ” produz melhorias do resultado que podem ser verificadas através do coeficiente R^2 (Anexo E).

Método Autorregressivo e Médias Móveis(ARMA)

As referências bibliográficas encontradas sobre o modelo ARMA (BOX; JENKINS, 2006) despendem um volume relevante de páginas com o objetivo de identificar modelos parcimoniosos que apresentem boa relação de custo-benefício entre o número de parâmetros e a qualidade da estimação (mínimos erros). Demonstrou-se nesta pesquisa que o esforço computacional para a obtenção dos resultados não é significativo em comparação com as análises das FAC- Função de Autocorrelação e FACP- Função de Autocorrelação Parcial, que, além de trabalhosas, não conduzem a conclusões seguras na identificação do modelo mais parcimonioso.

Série 5 - Método ARMA												
	1	2	3	4	5	6	7	8	9	10	11	12
1	9	31	18	20	21	21	25	28	29	44	40	50
2	21	21	19	20	22	21	30	31	31	48	35	50
3	23	28	35	20	22	21	29	33	36	52	36	49
4	19	22	23	29	30	33	37	34	34	49	36	42
5	21	22	23	27	31	29	39	34	34	48	42	46
6	25	27	26	30	31	31	40	35	36	50	43	56
7	27	24	25	33	34	34	39	47	39	51	42	57
8	22	23	34	31	36	35	40	37	48	53	42	53
9	26	39	40	39	37	40	41	39	46	53	54	61
10	30	38	38	37	42	41	42	46	50	50	74	51
11	33	37	41	43	49	41	50	48	78	58	62	86
12	32	42	45	50	46	45	51	49	61	87	76	81

Figura 14 - Exemplo de distribuição de ajuste R^2

Como exemplo, observe-se a comparação dos resultados obtidos pelo método ARMA para a Série 5, Figura 14, e cujos gráficos de FAC e FACP são apresentados na Figura 15. É difícil afirmar que a análise conduísse pela escolha do modelo $p=12$ e $q=10$, que foi o melhor resultado obtido para o coeficiente de correlação R^2 .

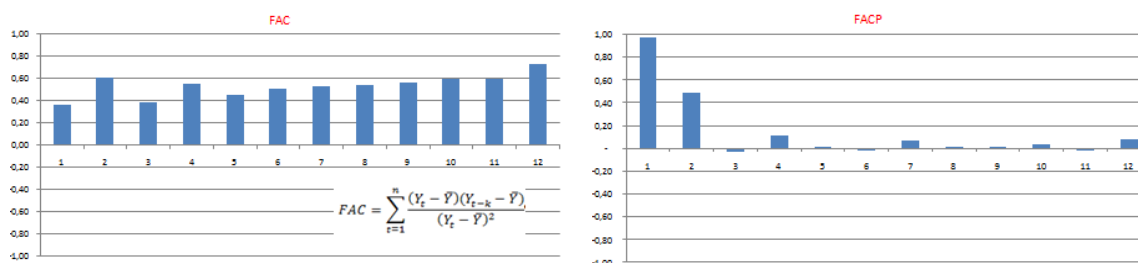


Figura 15 - Funções de Autocorrelação e Autocorrelação Parcial

Recomenda-se, então, partir diretamente para as experimentações e simulações e avaliar a qualidade do ajuste diretamente.

O modelo ARMA, apesar de ser mais eficiente na identificação de coeficientes que produzam resíduos menores, por serem obtidos por processo mais sofisticado gerados iterativamente, apresentam dois inconvenientes: 1) podem retornar erro quando a matriz de regressão “ Xp ” não for reversível; 2) não permitem a obtenção

de valores projetados sem a reconstrução de todas as iterações que lhes deram origem. Isto significa que, para a projeção do item da série Y_{t+1} , faz-se necessário recálculo de Y_0 até Y_t para a geração dos erros necessários à obtenção da matriz “Xq”, sobre a qual se aplicam os coeficientes θ_i . Esse fato constitui uma desvantagem do método quando se projetam valores futuros, já que se faz necessária a reconstrução do erro $(Y_i - \hat{Y}_i)$, que foi obtido de forma iterativa.

5.1.2 Avaliação da qualidade dos resultados

Para a análise da qualidade dos resultados obtidos reuniu-se, em um painel, o sumário das informações dos resíduos com indicador da qualidade do ajuste R^2 , soma de quadrado dos erros, histograma, gráfico, indicador da assimetria e curtose da função de distribuição de probabilidade dos erros de estimação, probabilidade de normalidade desses erros, além do teste de Ljung-Box, para verificar se persistem nos resíduos indicadores de correlacionamentos. O modelo de painel é apresentado no Anexo E.

Obtidos os modelos que melhor se ajustam às séries e apresentando os resíduos distribuição de probabilidade gaussiana, é possível inferir que o valor projetado para a série no período “t+1” terá probabilidade de ocorrência em intervalo calculado para 95%, cujo valor vem a ser 1,96 vez o desvio-padrão obtido para a série de resíduos da diferença entre o valor real e o valor estimado. É importante registrar que, ao estacionarem as séries, foram utilizados os coeficientes determinísticos β_0 – intersepto e β_1 coeficiente angular, que são incorporados à série, além do fator de sazonalidade correspondente ao número de dias úteis aplicado quando se analisa especificamente a Série 1- Venda de Serviço.

O objetivo principal desta pesquisa é a obtenção de uma ferramenta auxiliar de análise para o exercício da atividade de conciliação contábil dos custos com mão de obra. A base da metodologia é que os valores apurados de custo para um determinado período contábil, mês de apuração de resultado, se mantenham dentro do intervalo de confiança, sendo este o pressuposto do controle exercido sobre a

variável de custo. Na hipótese de não se manter neste intervalo IC, conclui-se pela quebra da confiabilidade do dado, indicando ao analista a necessidade de investigar as razões que contribuíram para o seu desvio. A consequência da investigação tem duas ações corretivas possíveis: 1) identificação e correção do erro gerador do desvio observado; 2) se confirmado o comportamento anômalo nos custos, poderá a sua incorporação à série ter consequência de, no futuro, ser o mesmo classificado como “*outlier*”. Um exemplo desse tipo de custo fora de padrão é a ocorrência de baixas/estorno de provisões que podem ser observadas em encerramento de exercícios fiscais, cuja ocorrência é frequentemente apresentada nos meses de dezembro e janeiro de cada ano, cuja característica é de serem meses de baixa atividade, e quando os valores são severamente afetados.

Observa-se na Tabela 4 que os indicadores de ajuste R^2 das séries, produzidos pelo algoritmo ARMA, são os que melhor ajuste de R^2 produziram. Entretanto, este algoritmo tem pior desempenho na eliminação de autocorrelacionamentos dos resíduos pelo índice de Ljung-Box. Todas as séries atendem ao critério Kolmogorov – Smirnov, com probabilidade maior do que 20% de risco de rejeição dos resíduos por não apresentarem uma distribuição gaussiana. Esse resultado valida os valores estimados para os intervalos de confiança, critério de aceitação dos valores em processo de contabilização.

Método	1	2	3	4	5	6
R²						
AG	47,88%	26,82%	17,40%	42,42%	31,60%	62,08%
AR	48,26%	26,65%	16,20%	41,62%	31,87%	62,38%
ARMA	88,57%	73,96%	72,70%	77,66%	87,32%	95,98%
Prob. Normalidade%						
AG	49,44%	15,54%	1,04%	94,28%	46,00%	0,00%
AR	70,00%	6,77%	0,57%	98,17%	47,40%	0,00%
ARMA	79,58%	19,65%	38,15%	70,86%	73,72%	26,89%
Prob. KS%						
AG	20,00%	20,00%	5,00%	20,00%	20,00%	20,00%
AR	20,00%	20,00%	5,00%	20,00%	20,00%	20,00%
ARMA	20,00%	20,00%	20,00%	20,00%	20,00%	20,00%
Lilliefors%						
AG	20,00%	20,00%	1,00%	20,00%	20,00%	5,00%
AR	20,00%	20,00%	1,00%	20,00%	20,00%	20,00%
ARMA	20,00%	20,00%	20,00%	20,00%	20,00%	20,00%
Q - Ljung-Box						
AG	76,24%	98,55%	97,47%	67,38%	88,48%	72,23%
AR	78,49%	97,86%	96,49%	78,31%	91,16%	79,43%
ARMA	0,92%	55,08%	18,85%	49,12%	92,63%	54,73%

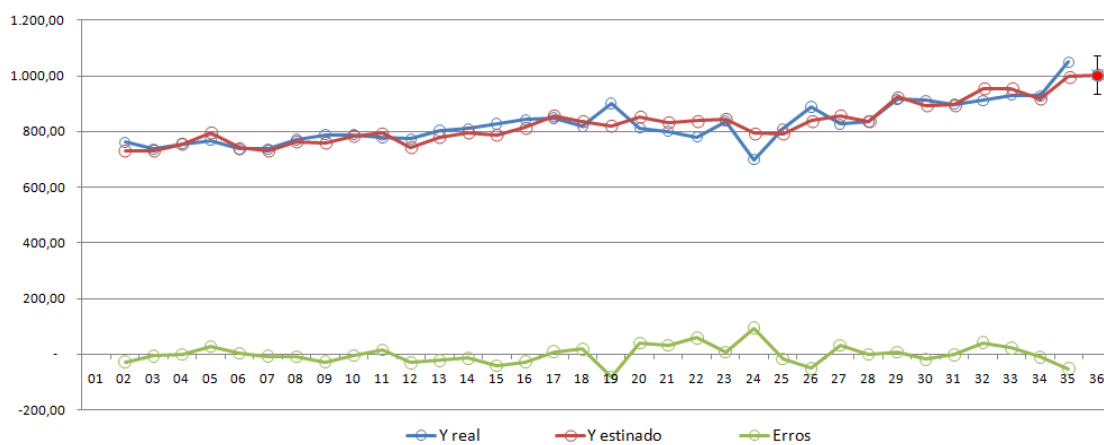
Tabela 4 - Resultados Ótimos R^2 e indicadores de Adaptação

O valor observado nas médias de Q- Ljung Box que melhor resultado obteve foi de 98,85% pelo método AG. Ainda com a finalidade de testar a normalidade da distribuição dos resíduos, o teste de Lilliefors obteve um resultado menos favorável, tendo em dois casos o nível mínimo de significância atingido, apenas 1%.

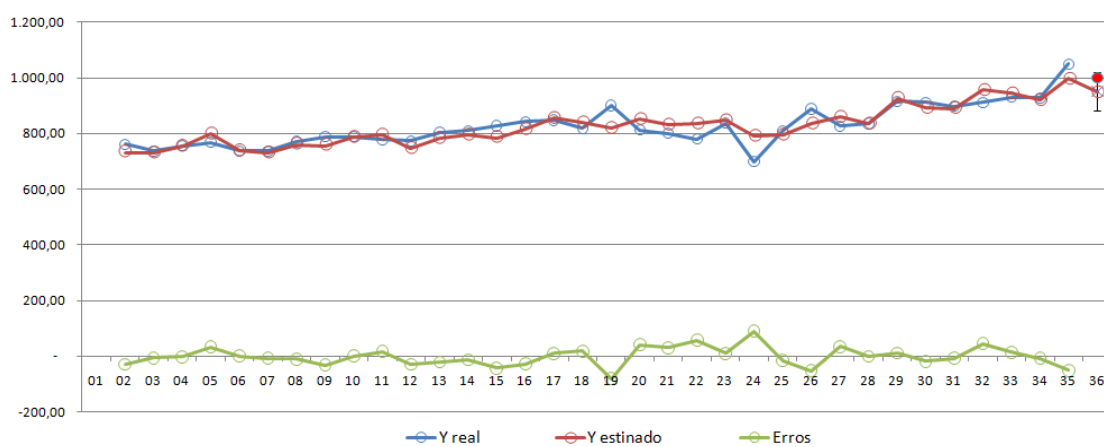
O critério de exclusão de *outliers* (90%) e os modelos de estimação AG, AR e ARMA definem a amplitude do intervalo de aceitação (intervalo de confiança). O primeiro produz o estreitamento desse intervalo ao se eliminarem itens considerados *outliers*, e o segundo também causa a diminuição do intervalo de confiança ao melhorar a qualidade do ajuste advindo do aumento do coeficiente de determinação R^2 . Comparando-se o ajuste dos modelos pelos diferentes métodos, a indicação é de que o melhor resultado de R^2 obtido pelo modelo ARMA produz um estreitamento acentuado do Intervalo de aceitação, além de não eliminar resíduos de autocorrelacionamento. O estabelecimento de padrões de qualidade muito altos conduz a uma diminuição do referido intervalo com a conseqüente rejeição dos valores em processo de contabilização (Figura 16).

Foram definidos como parâmetros de calibração: 1) para a eliminação de itens *outliers*, a aceitação de valores com até 90% de probabilidade ($\tau=1,64$ DP), descarte de 5% nas caudas, na presunção de distribuição normal e 2) para o critério de aceitação do valor no momento de contabilização, como conforme, o valor em contabilização deve estar dentro de intervalo de probabilidade de 95%, descarte de 2,5% nas caudas. Apresentando o valor de custo em processo de análise uma diferença em relação ao valor estimado pelo modelo superior a 1.96 desvios-padrão, este será rejeitado decidindo-se pelo processo de conciliação convencional. Isto significará o exame de todos os valores analíticos que o compõem. Por serem processos não causais, em que os valores futuros não estão fundamentados em fatores passados, o aumento da ordem do modelo e a melhoria do processo de estimação podem conduzir ao estabelecimento de intervalos de aceitação muito estreitos. Esse efeito contraria o objetivo de diminuição do esforço de conciliação com a verificação de muitos itens.

(a) AG – IC=68,3



(b) AR – IC=68,1



(c) ARMA – IC=31,8

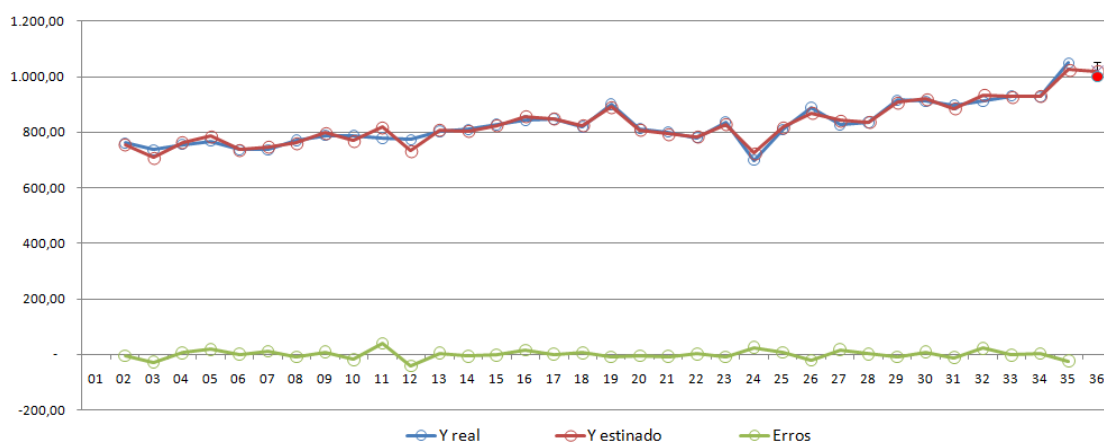


Figura 16 (a,b,c) - Estimativa da série 6 para o período dez/09

O período de encerramento fiscal, mês de dezembro, que é muito influenciado pelos ajustes oriundos de decisões estratégicas e gerenciais, mostrou-se menos adequado para a aplicação desse método de apoio na tomada de decisão. Nos meses regulares, intermediários, no entanto, o modelo proporciona grande ganho de tempo em análises irrelevantes e inconclusivas. Na Tabela 5, são apresentados os resultados das seis séries temporais na avaliação dos resultados financeiros do período de dezembro de 2009 a junho de 2010, portanto, 42 séries temporais. O método utilizado foi o AR, e os resultados recomendaram a aceitação de 23 valores desses valores (55%). Os indicados em fundo azul são aceitos com a dispensa verificações mais rigorosas e onerosas.

% IC	Séries					
Período	1	2	3	4	5	6
200912	72%	368%	408%	50%	120%	477%
201001	600%	464%	85%	283%	639%	55%
201002	14%	47%	39%	205%	337%	10%
201003	72%	102%	16%	161%	57%	67%
201004	12%	12%	38%	177%	22%	18%
201005	45%	25%	362%	146%	555%	49%
201006	70%	130%	383%	82%	323%	39%
Número de $\Delta < IC$	6	3	4	2	2	6
Total Periodos	7	7	7	7	7	7
% de aceitação	86%	43%	57%	29%	29%	86%
				Número de $\Delta < IC$	23	
				Total Periodos	42	
				% de aceitação	55%	

Tabela 5 - Distância do valor contabilizado em relação ao valor estimado

No que se refere à verificação do modelo quanto à calibração do custo padrão, que busca transferir todos os custos incorridos para os projetos (vide Figura 1), não foram identificadas grandes variações entre os custos incorridos (Compra) e os custos transferidos (Venda) (vide Figura 1). As diferenças apuradas nesse balanço, em 2009, foram de 1.73% antes dos tratamentos e de 1.76% após os tratamentos. Esse resultado é considerado satisfatório para o acompanhamento orçamentário e as variações são toleráveis tomando-se em consideração outros itens de custo, como variações de preço de material de consumo, passagens etc.

6 CONCLUSÕES E TRABALHOS FUTUROS

Esta pesquisa foi motivada pela necessidade de identificar uma ferramenta auxiliar para a solução de problema de natureza determinística, tendo em vista a limitação de tempo e recursos para executar a verificação de todos os dados, mensalmente. Apesar do grande número de eventos/lançamentos contábeis, estes precisam ter uma adequada investigação em curto espaço de tempo, no máximo duas semanas. Seria de grande ajuda a diminuição do esforço de investigação com a identificação de potenciais zonas de problema, reduzindo a grupos menores, melhorando o foco de análise para itens estatisticamente identificados como suspeitos de descontrole e erros.

Na busca de ferramenta para essa aplicação, foram consideradas técnicas estatísticas de análise multivariada, dentre elas Dados em Painel, Teoria dos Jogos, Simulação de Monte Carlo, Vetores Autorregressivos e Análise de Séries Temporais. Dentre estas, a última mostrou-se mais adequada, devido à forte independência entre as séries temporais e a robustez do modelo ARMA. Este não é um modelo causal, ou seja, não analisa a causa de um evento. Está fundamentado unicamente na existência de um padrão embutido/implícito nos valores passados e que influenciam o seu resultado futuro. Esse componente é incorporado na forma de pesos/coeficientes com o objetivo de minimizar os erros entre os valores reais e os modelados.

O resultado da pesquisa, que na fase de planejamento tinha por objetivo final a construção de um software para a realização da tarefa de diagnóstico de processo, rotina financeira nas empresas, pode ser considerado como tendo sido atingido, quando se analisa o grande número de informações disponíveis e a praticidade oferecida pelo DASHBOARD - Painel de análise das séries temporais (Figura 24).

Outro resultado parcial do projeto foi o desenvolvimento e a implementação do algoritmo genético, que mostrou ser uma ferramenta robusta na estimação de parâmetros do modelo ARMA. Apresenta vantagem por ser um modelo simples e

não depender da reversibilidade da matriz autorregressiva X_p (vide Tabela 6) ou da matriz de erros X_q (Figura 10).

	Vantagens	Desvantagens
AG	<ul style="list-style-type: none"> • Simplicidade • Reporta equação diretamente 	<ul style="list-style-type: none"> • Menos preciso • Não é repetitivo
AR	<ul style="list-style-type: none"> • Simplicidade • Determinístico • Reporta equação diretamente 	<ul style="list-style-type: none"> • Depende da reversibilidade da matriz X_p
ARMA	<ul style="list-style-type: none"> • Mais eficiente • Determinístico 	<ul style="list-style-type: none"> • Depende da reversibilidade da matriz X_p e X_q • Depende do recálculo de toda a série para gerar valores projetados

Tabela 6 - Comparativo dos métodos de estimação

O algoritmo proposto por Box e Jenkins (1970) para a obtenção dos parâmetros “p” e “q” de modelos parcimoniosos de representação de séries temporais não foi utilizado nesta pesquisa. A evolução tecnológica dos processos e os custos computacionais atuais possibilitam a varredura de todas as combinações possíveis desses parâmetros, melhorando a confiabilidade aos valores projetados. Em nenhuma das seis séries observou-se piora do erro entre a série real e a série estimada, quando se crescia a ordem do modelo, com o crescimento de “p” e “q”. Essa constatação é evidenciada na visualização do Anexo D.

A busca de ferramenta de apoio à atividade de análise financeira tem obtido grande avanço nos últimos anos, proporcionado pela tecnologia da informação, com o acesso a grande massa de dados com os sistemas integrados de gestão – ERP – e ferramentas de inteligência de negócios – BI. No entanto, nada foi oferecido para a melhoria do processo de conciliação contábil. Os objetivos propostos podem ser considerados atingidos pela redução do custo no trabalho investigativo, que no experimento descrito concluiu conforme 50% das séries analisadas, dispensando exame pormenorizado. Na realidade das empresas não são raras as situações em que o custo do controle excede o custo do risco exposto pelo não controle. A

substituição do método convencional de exame de todos os lançamentos contábeis pelo método de análise das movimentações das contas sintéticas da contabilidade representa grande economia de recursos e tem sido aplicada de forma ampla com auditores e contadores. Critérios estatísticos que aumentem a assertividade do processo de conciliação contábil foi o esforço deste trabalho. Utilizando-se de método de decomposição das séries temporais de custo de mão de obra, seguido da identificação das tendências e dos autocorrelacionamentos, obteve-se maior transparência. Esse fato transmitiu segurança na decisão de aceitar ou não um valor dessas séries para a apuração do resultado econômico da organização.

Os resultados atingidos com o uso de heurística para a identificação de não conformidades no processo de contabilização de custos com mão de obra encorajam a sua aplicação na detecção de outros custos diretos. Para essa finalidade está planejado o aperfeiçoamento do modelo no sentido de atender a múltiplos objetivos. Além do refinamento conceitual do software, é desejada a integração das informações disponíveis na contabilidade para que diretamente possibilite a sua análise crítica. Para esse último objetivo está planejada a construção um de software, desenvolvido no conceito de web, que seja amigável e que ofereça interatividade de requisição de mais de um objetivo simultaneamente.

REFERÊNCIAS

Aguirre, Luiz Antonio. 2007. *Introdução à identificação de sistemas*. Belo Horizonte : UFMG, 2007.

Baragonaa, Roberto, Francesco, Battagliab e Calzina, Claudio. 2001. Genetic algorithms for the identification of additive and innovation outliers in time series. *Computational Statistics & Data Analysis 37 (2001) 1–12*. 37, 2001, Vols. 1-12.

Beenstock, Michael e George, Szpiro. 2002. Specification search in nonlinear time-series models using the genetic algorithm. *Journal of Economic Dynamics & Control*. 26, 2002, Vols. 811-835, 26.

Box, G.E.P. and Jenkins, G.M. 1970. *Time series analysis: Forecast and control*. San Francisco : Holden Day, 1970.

Box, George E.P. 1976. *Time Series Analysis: forecasting and control*. Oakland - USA : Holden-Day, 1976.

Bremermann, H.J. 1958. The evolution of intelligence. The nervous system as a model of its environment. *Technical Report - Univ. Washington - Seattle*. 477, 1958, Vol. 17.

Brimson, A. James. 1996. *Contabilidade por Atividades*. São Paulo : Editora Atlas, 1996.

Bueno, Rodrigo de Losso da Silveira. 2008. *Econometria de séries temporais*. São Paulo : Cengage Learning, 2008.

Campos, H. 1979. *Estatística Experimental Não-Paramétrica*. Piracicaba : USP, 1979.

Cannon, W.D. 1932. *The Wisdom of the Body*. New York : Norton and Company, 1932.

Chorng, Shyong Ong, Jih-Jeng Huang ,Gwo-Hshiung Tzeng. 2005. *Model identification of ARIMA family using genetic algorithms.* Taipei - Taiwan : National Taiwan University, 2005.

Cortez, Paulo e Rocha,Miguel. 2004. *Evolving Time Series Forecasting ARMA.* Braga - Portugal : Kluwer Academic Publishers, 2004.

De Souza, Luzia Vidal. 2006. *Tese Doutorado - Programação Genética e Combinação de Preditores para Previsão de Séries Temporais.* Curitiba : Universidade Federal do Paraná, 2006.

Ehlers, S. Ricardo. 2007. *Análise de Séries Temporais.* São Paulo : Ricardo Sandes Ehlers, 2007.

Engelbrecht, Andries P. 2007. *Computational Intelligence.* West Sussex, England : John Wiley & Sons, 2007.

Fogel, David B. 2006. *Evolucionary Computation.* New Jersey : IEEE Press, 2006.

Fraser, A. S. 1957. Simulation of genetic systems by automatic digital computers. I. Introduction. *Aust. J. Biol. Sci.* 1957, Vol. 10.

Gujarati, Damodar N. 2004. *Basic Econometrics.* New York : McGraw-Hill, 2004.

Holland, J.H. 1975. *Adaptation in Natural and Arti"cial Systems - 2nd Edition.* Cambridge : MIT Press, 1975.

Horngren, Charles T. 1985. *Introdução à Contabilidade Gerencial.* Rio de Janeiro : Prentice-Hall do Brasil Ltda, 1985.

Hung, Jui-Chung. 2008. A genetic algorithm approach to the spectral estimation of time series. *Department of Information Technology, Ling Tung University.* 1, 2008, Vol. 178.

Levine, David M., et al. 2008. *Estatística Teoria e Aplicações*. Rio de Janeiro : LTC, 2008.

Mahpol, Khairil Asmani b. 2005. *Box-Jenkins and Genetic Algorithm Hybrid Modeling Electricity Forecasting System*. Kuala Lumpur : Universiti Teknologi Malaysia, 2005.

Makridakis, Spyros G. 1989. *Forecasting Methods for Management*. New York : John Wiley & Sons, 1989.

Martins, Eliseu. 2001. *Contabilidade de custos*. São Paulo : Atlas, 2001.

Meckesheimer, S. Basu and M. 2007. Automatic outlier detection for time séries: an application to sensor data. *Knowledge and Information Systems*. 11, 2007, Vol. 2.

Mendes, L.P. Vale. 2008. *Dissertação de Mestrado - Algoritmo genéticos aplicados a séries temporais em mercados cambias*. Coimbra : Faculdade de Economia da Universidade de Coimbra, 2008.

Michael Affenzeller, Stephan Winkler and Stefan Wagner. 2008. *Evolutionary Systems Identification:New Algorithmic Concepts and Applications*. Hagenberg - Austria : Upper Austrian University of Applied Sciences, 2008.

Núñez-Letamendia, Laura. 2007. Fitting the control parameters of a Genetic Algorithm: An application to technical trading system design. *European Journal of Operational Research*. 2007, pp. 179(3):847-868 por Cortez, Paulo e Rocha.

Parviz, L. , Kholghi,M. and A. Hoorfar. 2010. A Comparison of the Efficiency of Parameter Estimation Methods in the Context of Streamflow Forecasting. *J. Agr. Sci. Tech*. Parviz et al., 2010, Vol. 12.

See, Linda & Openhaw Stan. 1999. Applying soft computing approaches to river level forecasting. *Hydrological Sciences-Journal-des Sciences Hydrologiques*. 44, 1999, Vol. 5.

Slutsky, E. 1937. The summation of Random Causes as Source of Cyclic Processes. *Econometrica* . 5, 1937.

Smith, A.E. Eiben and J.E.Smith. 2003. *Introduction to Evolutionary Computing*. Berlin : Springer, 2003.

Wong Francis, Pan Yong Tan. 2009. Neural Networks and Genetic Algorithm for Economic Forecasting. *AI in Economics and Business Administration*. 2009.

Yohai, V.J., Marona A.R. e Martin R.D. 2006. *Robust Statistics*. West Sussex-England : John Wiley & Sons, 2006.

Yule, G.U. On the method of investigating periodicities in disturbed series, with Special Reference to Wolfer's Sunspot Numbers. *Philosophical Transactions A*. 89,1927 Citado por Makridakis, Spyros G. 1989. *Forecasting Methods for Management*. New York : John Wiley & Sons, 1989.

Tabela 7 - Série histórica de custos

Período	Mês	Venda (e)	Salários e Ordenados P. A admn.	Décimo Terceiro Salário	F. G. T. S.	I. N. S. S.	P. I. S.	Provisão e Abono de Férias	Férias	Av. Próvio Inden. Férias Inden	Outros Indiretos	Custo (b)	Resultado (c)= (a)- (b)
01	200601	656	503	-292	348	77	3	-70	0	0	177	744	-88
02	200602	661	392	30	35	117	4	57	0	0	172	807	-146
03	200603	616	301	28	69	211	9	37	0	5	200	861	-245
04	200604	618	327	26	36	105	3	44	0	0	208	750	-132
05	200605	614	347	24	31	95	3	21	0	0	220	741	-127
06	200606	620	349	29	33	95	4	39	1	0	227	776	-156
07	200607	625	360	26	33	100	3	70	0	2	212	806	-181
08	200608	630	335	35	30	90	3	2	0	0	248	743	-113
09	200609	646	351	34	36	108	3	49	0	0	245	825	-179
10	200610	651	366	28	33	100	3	38	0	0	236	805	-154
11	200611	642	347	24	43	100	3	39	0	0	249	805	-163
12	200612	625	390	26	25	20	9	34	0	4	42	550	75
13	200701	759	476	28	35	107	3	38	0	5	79	770	-12
14	200702	761	350	30	33	104	4	49	0	4	98	671	90
15	200703	737	355	29	82	261	10	45	0	0	77	859	-121
16	200704	756	348	29	33	111	4	36	0	0	174	735	21
17	200705	768	358	29	33	104	3	40	0	0	139	706	62
18	200706	737	368	28	30	96	4	40	0	0	683	1.249	-512
19	200707	736	415	27	31	97	5	40	0	3	307	926	-190
20	200708	771	359	25	32	101	4	38	0	0	350	908	-137
21	200709	789	366	30	33	103	2	37	0	6	259	835	-47
22	200710	788	363	30	34	107	4	42	0	4	336	920	-132
23	200711	779	361	22	46	84	4	33	0	4	209	762	17
24	200712	774	403	40	25	24	9	44	0	1	-354	191	583
25	200801	803	496	29	46	102	4	33	0	0	164	874	-71
26	200802	810	385	32	49	102	4	52	0	0	186	810	-1
27	200803	828	421	29	35	111	3	32	0	4	271	906	-78
28	200804	842	361	32	37	116	4	41	0	4	243	839	3
29	200805	847	341	30	36	115	4	44	0	5	289	864	-16
30	200806	819	393	31	36	115	4	41	0	8	359	987	-168
31	200807	900	430	29	33	105	4	51	0	9	279	940	-40
32	200808	813	388	18	32	93	4	37	0	2	349	923	-110
33	200809	801	385	23	31	104	4	41	0	6	361	957	-156
34	200810	780	400	30	34	108	4	28	15	0	-64	555	225
35	200811	836	389	22	32	101	4	44	0	0	109	701	136
36	200812	699	355	66	16	53	9	34	0	0	-418	115	585
37	200901	810	568	31	36	117	3	15	0	5	71	845	-35
38	200902	889	401	34	44	140	4	126	0	4	77	830	59
39	200903	826	393	30	63	203	7	48	0	1	119	866	-39
40	200904	835	378	31	40	109	4	44	0	0	109	715	120
41	200905	915	476	30	4	119	4	52	0	1	300	986	-71
42	200906	911	353	31	33	18	4	35	0	4	298	777	134
43	200907	898	473	34	36	109	4	53	0	2	318	1.029	-131
44	200908	912	444	34	37	122	4	46	0	0	353	1.042	-130
45	200909	930	474	34	40	132	5	83	0	2	417	1.187	-257
46	200910	928	450	32	35	115	4	11	0	1	327	975	-47
47	200911	1.049	459	37	41	132	5	46	0	0	284	1.005	44
48	200912	1.000	598	51	39	83	11	56	0	7	-389	456	544

- Tendência das séries
- Itens substituídos
- Critério de extração de outliers

Série Original

Série Agrupadas

Venda (e)	Salários e Ordenados P. A admn.	Décimo Terceiro Salário	F. G. T. S. e I. N. S. S.	Provisão e Abono de Férias	Outros Indiretos PIS-AVP-FI
661	392	30	151	57	176
616	301	28	281	37	214
618	327	26	141	44	212
614	347	24	126	21	223
620	349	29	128	39	231
625	360	26	133	70	217
630	335	35	120	2	251
646	351	34	143	49	248
651	366	28	133	38	239
642	347	24	142	39	252
625	390	26	45	34	55
759	476	28	141	38	86
761	350	30	137	49	105
737	355	29	343	45	87
756	348	29	143	36	178
768	358	29	137	40	143
737	368	28	126	40	886
736	415	27	128	40	316
771	359	25	132	38	354
789	366	30	135	37	267
788	363	30	141	42	344
779	361	22	129	33	216
774	403	40	49	44	-344
803	496	29	148	33	168
810	385	32	151	52	190
828	421	29	146	32	278
842	361	32	154	41	252
847	341	30	151	44	298
819	393	31	151	41	371
900	430	29	139	51	292
813	388	18	124	37	356
801	385	23	136	41	372
780	400	30	142	28	-45
836	389	22	133	44	113
699	355	66	69	34	-409
810	568	31	153	15	79
889	401	34	184	126	85
826	393	30	266	49	128
835	378	31	149	44	114
915	476	30	124	52	305
911	353	31	51	35	307
898	473	34	144	53	324
912	444	34	159	46	358
930	474	34	173	83	423
928	450	32	150	11	332
1.049	459	37	173	46	289
1.000	598	51	122	56	-372

Série Sem Outliers

Venda (e)	Salários e Ordenados P. A admn.	Décimo Terceiro Salário	F. G. T. S. e I. N. S. S.	Provisão e Abono de Férias	Outros Indiretos PIS-AVP-FI
661	392	30	151	57	176
616	301	28	137	37	214
618	327	26	141	44	212
614	347	24	126	21	223
620	349	29	128	39	231
625	360	26	133	39	217
630	335	35	120	39	251
646	351	34	143	49	248
651	366	28	133	38	239
642	347	24	142	39	252
625	390	26	141	34	55
759	355	28	141	38	86
761	350	30	137	49	105
737	355	29	137	45	87
756	348	29	143	36	178
768	358	29	137	40	143
737	368	28	126	40	267
736	415	27	128	40	316
771	359	25	132	38	354
789	366	30	135	37	267
788	363	30	141	42	344
779	361	22	129	33	216
774	403	40	141	44	252
803	366	29	148	33	168
810	385	32	151	52	190
828	421	29	146	32	278
842	361	32	154	41	252
847	341	30	151	44	298
819	393	31	151	41	371
900	430	29	139	51	292
813	388	30	124	37	356
801	385	23	136	41	372
780	400	30	142	28	-45
836	389	22	133	44	113
699	355	30	142	34	113
810	393	31	153	44	79
889	401	34	184	44	85
826	393	30	144	49	128
835	378	31	149	44	114
915	476	30	124	52	305
911	444	31	150	35	307
898	473	34	144	53	324
912	444	34	159	46	358
930	474	34	173	49	423
928	450	32	150	46	332
1.049	459	37	173	46	289
1.000	450	34	122	56	289

Série estacionada

Venda (e)	Salários e Ordenados P. A admn.	Décimo Terceiro Salário	F. G. T. S. e I. N. S. S.	Provisão e Abono de Férias	Outros Indiretos PIS-AVP-FI
44	61	3	19	19	-14
-8	-32	1	4	-1	22
-14	-9	-1	8	5	18
-25	9	-3	-8	-18	28
-26	9	1	-7	1	35
-28	18	-2	-2	0	18
-31	-10	7	-15	0	51
-21	3	6	8	10	46
-24	16	0	-3	-1	36
-40	-5	-4	6	0	47
-64	35	-2	5	-5	-152
62	-2	0	4	-1	-122
57	-9	1	0	9	-105
26	-7	0	-1	5	-125
38	-16	0	5	-4	-35
43	-9	0	2	0	-72
4	-1	-1	-13	0	50
-3	44	-1	-12	-1	98
24	-15	-4	-7	-3	134
35	-10	1	-5	-4	46
27	-15	1	0	1	120
11	-19	-7	-12	-9	-8
-2	19	0	2	0	25
20	-19	0	6	-9	-60
20	-3	3	9	10	-40
31	30	-1	4	-10	47
38	-32	2	11	-1	18
36	-54	1	7	-1	63

ANEXO B –PROGRAMA AG PARA SÉRIES TEMPORAIS

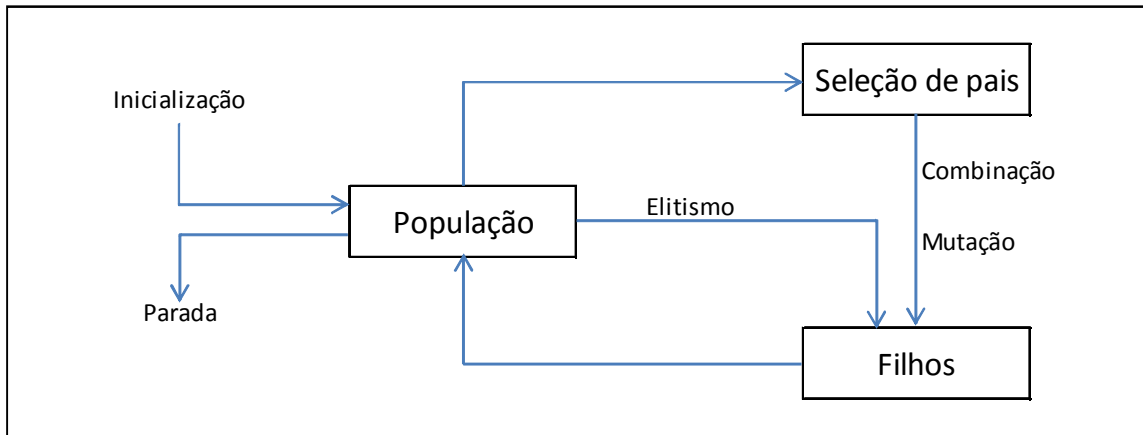


Figura 17 - Esquema básico de AG com elitismos

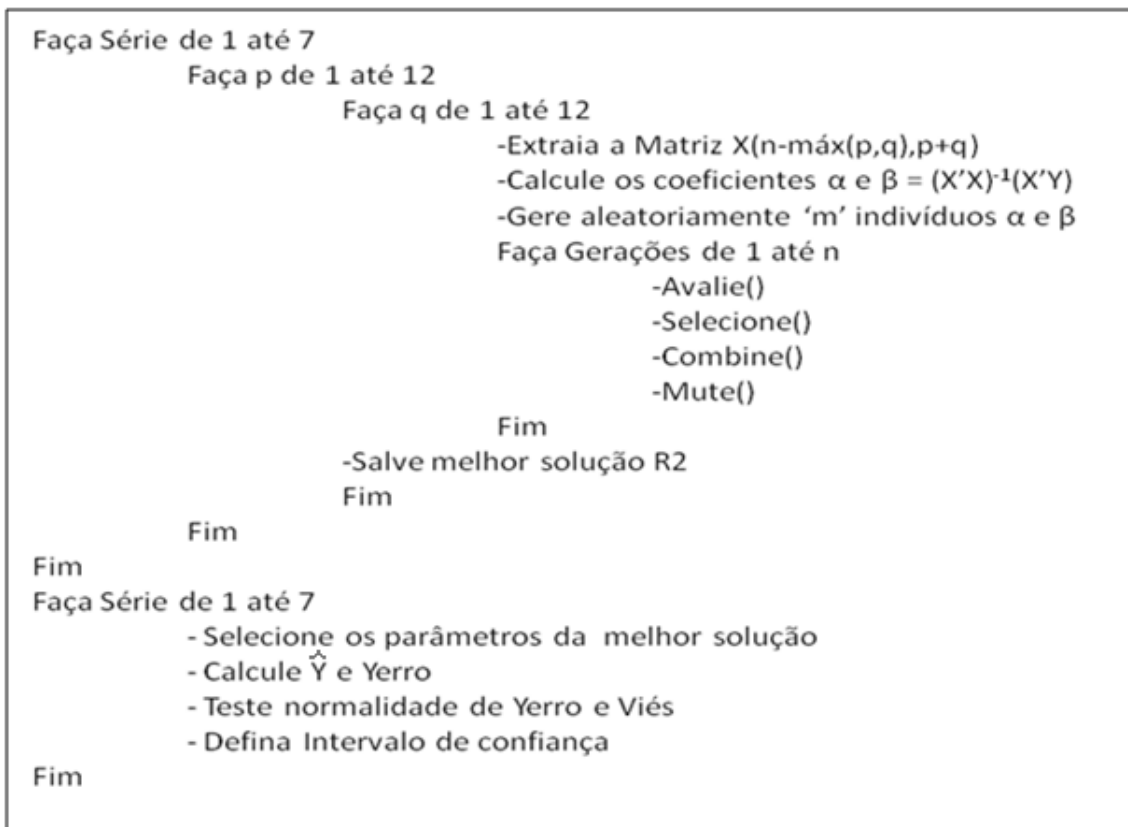


Figura 18 - Pseudocódigo básico AG

Descrição das rotinas do programa com identificação de suas variáveis e esquemas

Inicializa()

Gera, a partir do cálculo das defasagens temporais das séries e dos erros médios, as matrizes X_r e X_m com o número de linhas equivalente ao tamanho da população e as colunas representando p e q , que são, respectivamente, os coeficientes de regressão (p) e os coeficientes de média móvel (q), gerando aleatoriamente a matriz de coeficientes $A(j,p+q)$, estando esses coeficientes iniciais entre 0 e 1.

Y	X					
	XR			XM		
477,00	01
481,00	02	477	4
477,00	03	481	477	-4	-2
490,00	04	477	481	477	13	11
460,00	05	490	477	481	-30	-24
449,00	06	460	490	477	-11	-26
463,00	07	449	460	490	14	9
480,00	07	463	449	460	17	24
	j

	A**					
	p			q		
01	0,13	0,79	0,31	0,25	0,26	0,13
02	0,32	0,77	0,03	0,98	0,31	0,20
03	0,36	0,22	0,94	0,00	0,32	0,61
04	0,40	0,25	0,54	0,09	0,62	0,80
05	0,51	0,45	0,36	0,94	0,98	0,57
n

Figura 19 - Matriz de autocorrelacionamento

Fitness()

É a rotina mais simples do algoritmo que procura avaliar a soma das diferenças entre as séries estimadas \hat{Y} , que é calculada pelo produto da matriz X ($n \times \max(p,q)$; $p+q$) e dos coeficientes gerados pelo algoritmo A (n , $p+q$) e a série real. Essas diferenças têm seus valores elevados ao quadrado e somados apurando-se o valor de SQE (soma de quadrado dos erros) da estimativa de cada solução (indivíduo). A soma dos quadrados das diferenças dos valores da série e a média da série, cuja soma dos quadrados é denominada SQT (soma de quadrados totais), variância, é calculada para apuração do coeficiente de determinação R^2 cuja formula é $1-SQE/SQT$. Esse indicador será usado para a avaliação da qualidade das soluções obtidas. O "fitness" de cada indivíduo da matriz A é apurado na matriz coluna SQM(j), com j linhas, cujo conteúdo é a diferença da soma de quadrados dos erros para o maior erro encontrado, ou seja, o melhor indivíduo é o que apresenta a maior diferença, sendo um problema de maximização. Para utilização da rotina de seleção é gerada a coluna SSQM(j), que é a acumulação dos valores de "fitness" SQM(j). O objetivo é criar para o mecanismo de roleta indexador que amplie as chances de seleção dos indivíduos que apresentam maior aptidão (Figura 17).

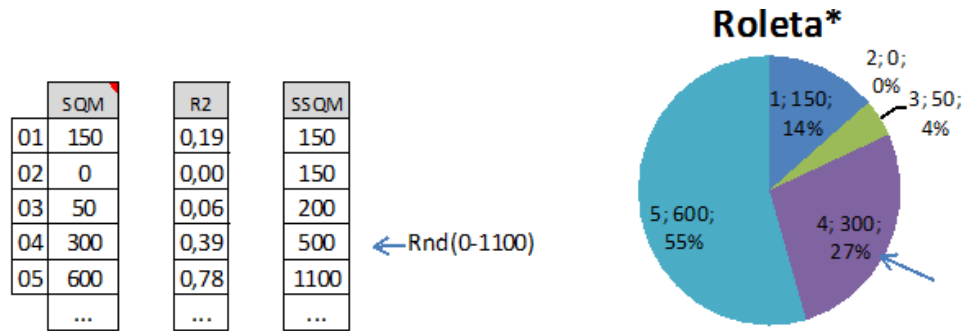


Figura 20 - Mecanismo de avaliação

Seleção()

Essa rotina define o processo de seleção dos indivíduos que serão usados para combinar e gerar novos indivíduos. Os indivíduos são selecionados pelos seus índices na matriz A, que corresponde aos mesmos índices na matriz SQM e cuja manipulação é processada na matriz **IndSQM**. O algoritmo prevê a seleção de um percentual do total da população de melhores indivíduos mPE . Sendo atribuído a essa variável o valor zero, ocorrerá exclusão da estratégia de elitismo. Nessa etapa são selecionados $j \cdot mPE$ índices (número da linha) dos melhores indivíduos e armazenados na matriz **IndSQM**.

A matriz **BEST()** recebe o melhor resultado de R^2 obtida em cada geração e serve apenas para o acompanhamento da dinâmica de convergência.

Após a seleção randômica dos elementos a serem transferidos para as gerações futuras, em número de elite ($mPE \cdot J$), são adicionados mais ($j - mPE$) indivíduos através da utilização de ponderação da variável SSQM, "fitness". Esses índices são armazenados na matriz coluna "roleta". Localiza-se o índice do indivíduo que apresenta aptidão imediatamente superior ao valor sorteado e é registrado o seu índice na matriz **IndSQM**.

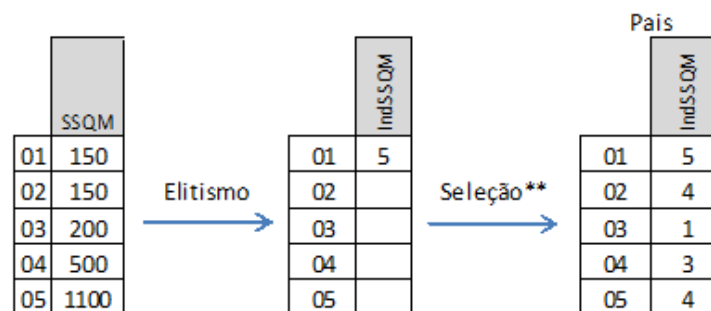


Figura 21 - Mecanismo de seleção

A estratégia do algoritmo pode ser ajustada atuando nos parâmetros **mPE** e **mPS** conforme a Tabela 8 seguir:

O parâmetro mPE é o percentual de elitismo. Atribuir zero a este parâmetro equivale no caso de ausência de elitismos	mPE=0
Desconsideração dos elementos selecionados no elitismo como candidatos a pais. Este parâmetro transfere a geração da matriz SSQM após seleção de elitismo. Os elementos selecionados no elitismo têm seus “ <i>fitness</i> ” alterados para zero, sendo que a sua probabilidade de seleção é nula.	oPE=1
Alteração do critério de seleção mPS define o percentual da população que sofrerá cruzamento(<i>cross-over</i>). O valor de 1-mPS são pais que se transferem para a geração futura.	mPS=0
Esta opção de <i>cross-over</i> indica que onde 2 pais gerem dois filhos cada um, o parâmetro oPS deve ser ajustado para 2. Em caso contrário, o algoritmo estará gerando para 2 pais selecionados para cruzamento, 1 filho. Sendo necessário gerar, para completar a nova população, o dobro de operações de combinação (Figura 23).	oPS=1 ou 2
Este parâmetro, sendo igual a 1, passa a instrução ao programa para aplicar percentual de mutação decrescente conforme gerações aproximem do critério de parada.	oPM=0 ou 1

Tabela 8 - Descrição dos parâmetros do programa AG

Combinação()

Com os pais selecionados na matriz coluna **IndSQM**, passa-se à montagem da matriz de coeficientes **Anew**, transferindo inicialmente os elementos selecionados por elitismo (Figura 22). A combinação é feita na geração da matriz PPC, matriz de 3 colunas que randomicamente têm selecionados na primeira e na segunda parte os pais que cederam material genético (coeficientes), e na terceira coluna é escolhido randomicamente o ponto onde ocorrerá o corte para a geração de 1 ou 2 indivíduos. Caso a opção seja pelo aproveitamento do material genético dos dois pais combinados, multiplicado pelo coeficiente “ $(1-\text{floor}(\text{rand}(\text{nprole},1)/\text{mPS}))$ ”, que avalia se será ou não efetuada a troca de material genético. Foi introduzido, também, o parâmetro oPS para definir o tipo de mecanismo de *cross-over*.

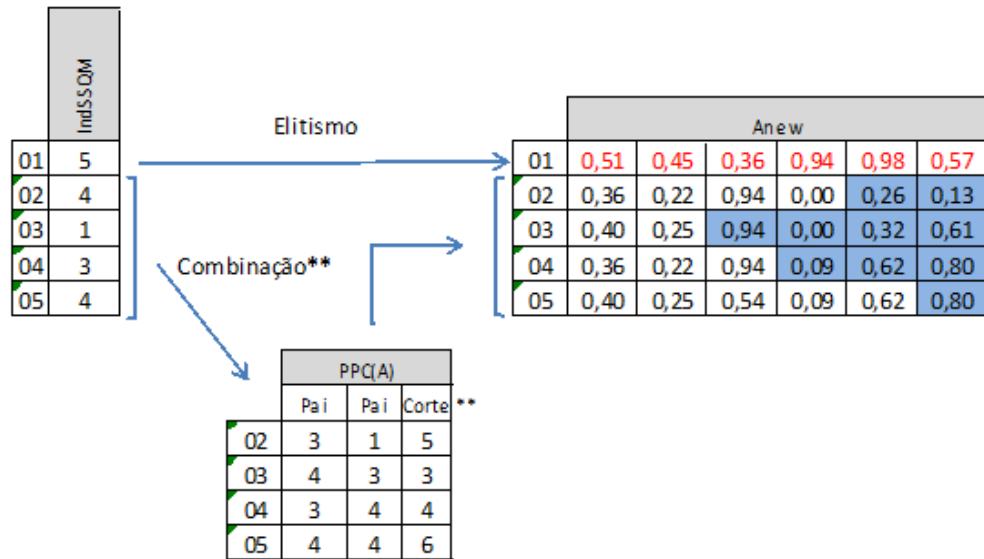
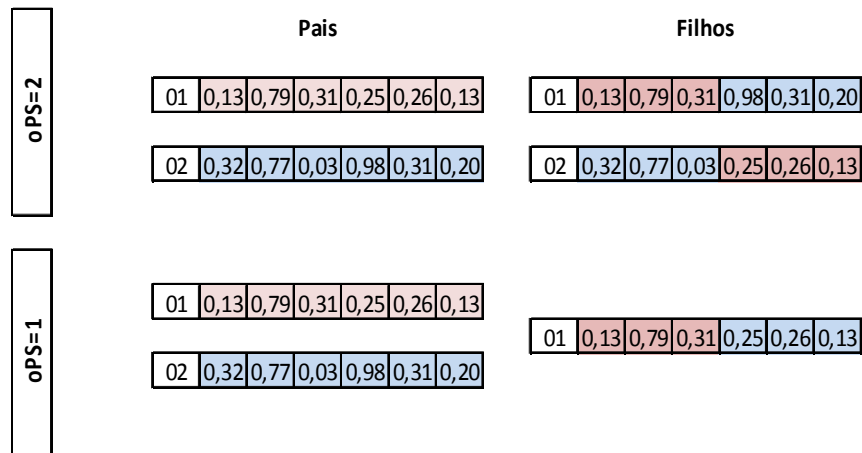


Figura 22 - Mecanismo de combinação

A propriedade do algoritmo de cada par de pais gerar 1 ou 2 filhos é aceita por SMITH(2003), sendo essa variação importante por ser a recombinação o principal operador de um AG. É aceito, também, por este autor a possibilidade de combinação de 3 pais, não existindo equivalência na natureza, mas que pode produzir pela regra de combinação bons efeitos para o algoritmo.

Figura 23 - Parametrização de *Cross-Over*

Mutação()

Em geral, os autores recomendam baixos índices de probabilidade de mutação, sendo que, em alguns casos, foram obtidos bons resultados nas velocidades de

convergência e de precisão, quando se adotam índices altos no início e que decaíam com o aumento do número de gerações (Engelbrecht, 2007).

Os melhores resultados, em termos de robustez de um GA aplicado a séries temporais de dados financeiros, foram obtidos com a combinação de probabilidades altas de cruzamentos com probabilidades baixas de mutação, ou com probabilidades baixas de cruzamentos e probabilidades altas de mutação, o que faz sobressair uma complementaridade entre as operações de cruzamento e de mutação (Núñez-Letamendia, 2007) (Mendes, 2008).

Código Principal

```

%-----CMDMain-----
% Algoritmo Genético para calculo de series temporais
% Autor : Kleber Silva Leite Pinto
% Escola: Cefet - MG
% Data : 13/03/10
% Parametros:
%-----Introdução da serie analisada-----
Yn=xlsread('C:\U\Input_DadosCustoMO','Estacionada');
%-----
pmax=12; qmax=12;
if l=1
    nserie=size(Yn,2);
else
    nserie=5;
end
%-----Cálculo das soluções p q de cada série(serie)-----
Sol_A=zeros(1,4+5+24); % gera a primeira linha da tabela de soluções
Sol_Y=[zeros(1,4+5+size(Yn,1)+1)]; % gera a primeira linha da tabela de soluções

for serie=1:nserie % número da série em análise
    Y=Yn(:,serie); % extrai para Y a série em análise
    m=size(Y,1); % m - tamanho da série size(Y,1)
    %Sol_A=zeros(1,4+5+24); % gera a primeira linha da tabela de soluções
    Sol_Y=[Sol_Y;zeros(1,4),serie,zeros(1,3),-100,Y',zeros(1,1)]; % gera a linha da serie
    Xt=zeros(m,m-1); % inicializar matriz máxima para m-1 regressores
    for i=2:m
        Xt(i:m,i-1)=Y(1:m-i+1); % gera as série defazadas dos lags 2 à m
    end
    for p=1:pmax % loop para gerar os p regressores
        for q=1:qmax % loop para gerar as q médias
            %-----Parametros dos algoritmo genético-----
            ml=min(m-p,m-q); %ml - número de itens projetados/ajustados
            j=100; % tamanho da população
            n=100; % número de iterações
            mPM=.99; % percentual de mutações
            mPS=.50; % percentual de seleção de pais
            mPE=.20; % percentual de seleção por elitismo
            % Opções do algoritmo
            oPE=1; % (1) Elite participa do processo de seleção pais (0) não participa
            oPS=1; % (1) dois pais i filho 2x (2) dois pais dois filhos 1x
            oPM=0; % (0) Percentual de mutação fixo (1) percentual de mutação decrescente
            maxF=1; % Coeficiente máximo-range
            Yr=Y(m-ml+1:m); %Série regressível
            AG(); % Estima série usando modelo AG(p,q)
            AR(); % Estima série usando modelo AR(p)
            ARMA(); % Estima série usando modelo ARMA(p,q)
            ARME(); % Estima série usando modelo ARME(p,q) erros iterativos
            Sol_A=[Sol_A;[n,mPM,mPS,mPE,serie,p,q,1,R2(1,1),A(1,1:p),A(1,p+1:p+q),zeros(1,24-p-q)]];
            Sol_A=[Sol_A;[n,mPM,mPS,mPE,serie,p,q,2,R2ar(1,1),Aar',zeros(1,24-p)]]; % tabela de soluções
            Sol_A=[Sol_A;[n,mPM,mPS,mPE,serie,p,q,3,R2Arme(1,1),AArme(1:p)',AArme(p+1:p+q)',zeros(1,24-p-
q)]]; % tabela de soluções
            Sol_A=[Sol_A;[n,mPM,mPS,mPE,serie,p,q,4,R2arma(1,1),Aarma(1:p)',Aarma(p+1:p+q)',zeros(1,24-p-
q)]]; % tabela de soluções
            %-----
            Forecast();
            %-----
            Sol_Y=[Sol_Y;[n,mPM,mPS,mPE,serie,p,q,1,R2(1,1),zeros(1,m-ml)],Yhat1']; % tabela de soluções
            Sol_Y=[Sol_Y;[n,mPM,mPS,mPE,serie,p,q,2,R2ar(1,1),zeros(1,m-ml)],Yhatar']; % tabela de soluções
            Sol_Y=[Sol_Y;[n,mPM,mPS,mPE,serie,p,q,3,R2Arme(1,1),zeros(1,m-ml)],YhatArme']; % tabela de
soluções
            Sol_Y=[Sol_Y;[n,mPM,mPS,mPE,serie,p,q,4,R2arma(1,1),zeros(1,m-ml)],Yhatarma'];
            %[size(Yhat1,1),size(Yhatar,1),size(YhatArme,1),size(Yhatarma,1)]
            if l==0
                figure(2)
                GrafSerie_3_1()
            end
        end
    end
end
end
Analise();
SolClear=zeros(size(xlsread('C:\Users\Kleber\Documents\KleberPai\Cefet\Material
Dissertação\XLS\Output_SoluçõesE.xls','Sol_Y')));
xlswrite('C:\Users\Kleber\Documents\KleberPai\Cefet\Material
Dissertação\XLS\Output_SoluçõesE.xls',SolClear,'Sol_A');
xlswrite('C:\Users\Kleber\Documents\KleberPai\Cefet\Material
Dissertação\XLS\Output_SoluçõesE.xls',SolClear,'Sol_Y');
xlswrite('C:\Users\Kleber\Documents\KleberPai\Cefet\Material
Dissertação\XLS\Output_SoluçõesE.xls',Sol_A,'Sol_A');
xlswrite('C:\Users\Kleber\Documents\KleberPai\Cefet\Material
Dissertação\XLS\Output_SoluçõesE.xls',Sol_Y,'Sol_Y');

```

```

%-----Inicializa-----
% Algoritmo Genético para calculo de series temporais
% Autor : Kleber Silva Leite Pinto
% Escola: Cefet - MG
% Data : 28/09/09
% Parametros:
% Inicialização de série temporal
% Parametros de regressão
% matriz de regressão Xr
% matriz de erro Xm
%-----
if p>0
    Xr=Xt(m-ml+1:m,1:p); % gera a matriz Xr utilizando os últimos m-max(p,q) itens da série
end
% Parametros de erro
if q>0
    Xm=Xt(m-ml+1:m,1)-Xt(m-ml+1:m,2); % gera a matriz Xm utilizando os últimos m-max(p,q) itens da série
    for i=3:q+1
        Xm=[Xm,Y(m-ml:m-1)-sum(Xt(m-ml+1:m,2:i)')/(i-1)]; %variação em relação à média fixa da serie
    end
end
if q==0
    X=Xr; % monta a matriz para modelo AR
else
    if p==0
        X=Xm; % monta a matriz para modelo MA
    else
        X=[Xr,Xm]; % monta a matriz para modelo ARMA
    end
end
%Xproj=[Yr(ml),Xr(ml,1:p-1),Xm]; % dados para projeção de i+1
Xproj=X(ml,:);
% População Inicial
A=rand(j,q+p)*maxF; % estima os p+q coeficientes dos j individuos
%-----Fitness-----
% Algoritmo Genético para calculo de series temporais
% Autor : Kleber Silva Leite Pinto
% Escola: Cefet - MG
% Data : 28/09/09
% Parametros:
%-----
Yhat=X*A'; % Calcula série ajustada
SQE=(Yhat-rep(Yr,[1 j]))'*(Yhat-rep(Yr,[1 j])); %Calcula o erro quadrado
SQE=diag(SQE);
SQT=(Yr-mean(Yr))'*(Yr-mean(Yr)); % Calcula soma de quadrados total
R2=1-SQE/SQT;
SQM=max(SQE)-SQE;
%-----Geração de roleta-----
SSQM=zeros(j,1);
SSQM(1)=SQM(1);
for i=2:j
    SSQM(i)=SSQM(i-1)+SQM(i);
end
%-----Selecao-----
% Algoritmo Genético para calculo de series temporais
% Autor : Kleber Silva Leite Pinto
% Escola: Cefet - MG
% Data : 28/09/09
% Parametros:
%-----
% Algoritmo Genético seleção de indivíduos
% Inicio de elitismo
IndSQM=max(find(max(SQM)==SQM,1)); %IndSQM é o indice da matriz de coeficientes
%SQM(IndSQM)=0;
if k==1
    Best=[k ,R2(IndSQM)];
else
    Best=[Best;k ,R2(IndSQM)];
end
if mPE==0 % sem elitismo
    nelite=0;
else
    nelite=round(mPE*j); % função da população(j) e percentual da elite
    for i=1:nelite
        IndSQM=[IndSQM;find(max(SQM)==SQM,1)];
        SQM(IndSQM)=0;
    end
end
end
if oPE==0 % exclusão da elite dos candidatos a pais SQM é nulo

```

```

%-----Geração de roleta novamente elite fora--
SSQM=zeros(j,1);
SSQM(1)=SQM(1);
for i=2:j
    SSQM(i)=SSQM(i-1)+SQM(i);
end
%-----
end
% Início de seleção de randômica dos pais
%roleta=(rand(floor(mPS*j),1)*max(SSQM));
roleta=(rand(floor(j-nelite),1)*max(SSQM)); % sorteio dos valores de SSQM;
%for i=1:mPS*j número de pais limitados a mPS*j
for i=1:j-nelite % número de pais correspondentes a elite+seleção randômica
    mSelecao=max(find(roleta(i)>=SSQM,1));
    IndSQM=[IndSQM;mSelecao];
end
% os pais estão selecionados no índice IndSQM
%-----Combinacao-----
% Algoritmo Genético para calculo de series temporais
% Autor : Kleber Silva Leite Pinto
% Escola: Cefet - MG
% Data : 28/09/09
% Parametros:
% PPC - indice de [individuo 1, individuo 2, local crossover]
%-----
% Algoritmo Genético combinação de indivíduos
Pais=A(IndSQM,:);
npais=size(Pais,1); % número de pais
nprole=j-size(Anew,1); % Transferencia de elite

Anew=A([IndSQM(1:mPE*j)],:); % Anew=Nova matriz de coeficientes transferencia de elite
% Combinação contatenação dos pais cortados elite+seleção aleatória

nprole=j-size(Anew,1); % número de filhos

PPC=[fix(rand(nprole,1)*(npais))+1,fix(rand(nprole,1)*(npais))+1,fix(rand(nprole,1)*(p+q)).*(1-
floor(rand(nprole,1)/mPS))]; %paipaicorte
% (1-floor(rand(nprole,1)/mPS)) assume valor=0 para probabilidade > mPS não trocando material

if oPS==1 % (1) dois pais i filho 2x (2) dois pais dois filhos 1x
    for i=1:nprole
        Anew=[Anew;[Pais(PPC(i,1),1:PPC(i,3)),Pais(PPC(i,2),PPC(i,3)+1:(p+q))]];
    end
else
    for i=1:nprole/2+2
        Anew=[Anew;[Pais(PPC(i,1),1:PPC(i,3)),Pais(PPC(i,2),PPC(i,3)+1:(p+q))]];
        Anew=[Anew;[Pais(PPC(i,2),1:(p+q)-PPC(i,3)),Pais(PPC(i,1),(p+q)-PPC(i,3)+1:(p+q))]];
    end
end
Anew=Anew(1:j,:); %Ajusta matriz para fitness

%-----Mutacao-----
% Algoritmo Genético para calculo de series temporais
% Autor : Kleber Silva Leite Pinto
% Escola: Cefet - MG
% Data : 28/09/09
% Parametros:
%-----
% Algoritmo Genético combinação de indivíduos
Anew;
% Introdução de mutação dentro da faixa de fatores
for i=1:j*mPM
    Anew(ceil(rand()* (j-nelite))+nelite,ceil(rand()* (p+q)))=randn()*maxF;
end
A=Anew; % substitui as gerações
%-----Regressão usando a função %ar()-----
% Inicialização de série temporal
% Parametros de regressão
% matriz de regressão Xr
%-----Cálculo de Regressão -----
% [Yr,Xr];
% Xar=[ones(mL,1),Xr];
% Xar=Xr;
% Aar=inv(Xar'*Xar)*Xar'*Y(m-mL+1:m);
% Yhatar=Xar*Aar;
%-----Avaliação da Regressão-----
% SQEar=(Yhatar-Y(m-mL+1:m))* (Yhatar-Y(m-mL+1:m)); %Calcula o erro quadrado
% SQTar=(Y(m-mL+1:m)-mean(Y(m-mL+1:m)))* (Y(m-mL+1:m)-mean(Y(m-mL+1:m)));
% SQEar=(Yhatar-Yr)* (Yhatar-Yr); %Calcula o erro quadrado
% SQTar=(Yr-mean(Yr))* (Yr-mean(Yr));
% R2ar=1-SQEar/SQTar;

```

```

%-----Arme-----
% Algoritmo Génético para calculo de series temporais
% Autor : Kleber Silva Leite Pinto
% Escola: Cefet - MG
% Data : 13/03/10
%-----Regressão usando a função %Arme()-----
% Inicialização de série temporal
% Parametros de regressão
% matriz de regressão Xr
% matriz de erro Xe
%-----Cálculo de Regressão Simples-----
[Yr,Xr];
Xar=Xr;
Aar=inv(Xar'*Xar)*Xar'*Y(m-ml+1:m);
Yhatar=Xar*Aar;
%-----Avaliação da Regressão-----
SQEar=(Yhatar-Y(m-ml+1:m))*(Yhatar-Y(m-ml+1:m)); %Calcula o erro quadrado
SQTar=(Y(m-ml+1:m)-mean(Y(m-ml+1:m)))*(Y(m-ml+1:m)-mean(Y(m-ml+1:m)));
R2ar=1-SQEar/SQTar; % calculo do coeficiente de determinação de regressão
%-----Cálculo de Médias de Erro-----
Ye=Yr-Yhatar; % Calculo do erro utilizando estimativa de regressão
Xe=zeros(ml,q);

for i=q+1:ml
    for j=1:q
        Xe(i,q-j+1)=Ye(i-q+j-1);
    end
end
XArme=[Xr,Xe];
AArme=inv(XArme'*XArme)*XArme'*Yr;
YhatArme=XArme*AArme;
%-----Avaliação -----
SQEArme=(YhatArme-Yr)*(YhatArme-Yr); %Calcula o erro quadrado
SQTArme=(Yr-mean(Yr))*(Yr-mean(Yr));
R2Arme=1-SQEArme/SQTArme; % calculo do coeficiente de determinação de regressão e erro
%-----Convergência -----
w=1;
Delta=0;
YhatArmePre=YhatArme;
while w=1 || Delta>.001
    Ye=Yr-YhatArmePre; % Calculo do erro utilizando estimativa de regressão
    Xe=zeros(ml,q);
    for i=q+1:ml; % Monta a matriz de erros
        for j=1:q;
            Xe(i,q-j+1)=Ye(i-q+j-1);
        end
    end
    XArmePre=[Xr,Xe];
    AArmePre=inv(XArmePre'*XArmePre)*XArmePre'*Yr;
    YhatArmePre=XArmePre*AArmePre;
    SQEArmePre=(YhatArmePre-Yr)*(YhatArmePre-Yr); %Calcula o erro quadrado
    SQTArmePre=(Yr-mean(Yr))*(Yr-mean(Yr));
    %-----critério de parada-----
    w=w+1;
    Delta=1-SQEArmePre/SQTArmePre-R2Arme;
    R2ArmePre=1-SQEArmePre/SQTArmePre; % calculo do coeficiente de determinação de regressão e erro
    if Delta>0 % caso a serie esteja convergindo
        YhatArme=YhatArmePre;
        XArme=XArmePre;
        AArme=AArmePre;
        YhatArme=YhatArmePre;
        SQEArme=SQEArmePre;
        SQTArme=SQTArmePre;
        R2Arme=R2ArmePre
    end
end
%-----
end

%-----Arma-----
% Algoritmo Génético para calculo de series temporais
% Autor : Kleber Silva Leite Pinto
% Escola: Cefet - MG
% Data : 28/09/09
%-----Cálculo de Regressão Simples-----

Aarma=inv(X'*X)*X'*Y(m-ml+1:m);
Yhatarma=X*Aarma;

%-----Avaliação da Regressão-----
SQEarma=(Yhatarma-Y(m-ml+1:m))*(Yhatarma-Y(m-ml+1:m)); %Calcula o erro quadrado
SQTarma=(Y(m-ml+1:m)-mean(Y(m-ml+1:m)))*(Y(m-ml+1:m)-mean(Y(m-ml+1:m)));
R2arma=1-SQEarma/SQTarma; % calculo do coeficiente de determinação de regressão e erro

```

ANEXO C – EXPERIMENTOS DE PARÂMETROS DO AG

Média de R2	<---Iterações--->					
Mutação	50	100	150	200	250	300
0%	-1,37	-2,40	-2,31	-1,91	-2,19	-2,89
5%	0,44	0,48	0,48	0,48	0,48	0,48
10%	0,47	0,47	0,48	0,48	0,48	0,48
15%	0,43	0,48	0,48	0,48	0,48	0,48
20%	0,45	0,47	0,48	0,48	0,48	0,48
25%	0,47	0,48	0,48	0,48	0,48	0,48
30%	0,44	0,48	0,48	0,48	0,48	0,48
35%	0,45	0,46	0,47	0,47	0,48	0,47
40%	0,37	0,44	0,47	0,46	0,46	0,48
45%	0,28	0,42	0,44	0,47	0,47	0,47
50%	0,40	0,46	0,44	0,46	0,47	0,46
55%	0,30	0,42	0,43	0,43	0,46	0,46
60%	0,35	0,42	0,35	0,40	0,45	0,45
65%	0,31	0,46	0,44	0,40	0,44	0,43
70%	0,15	0,33	0,33	0,37	0,35	0,45
75%	0,01	0,41	0,32	0,33	0,34	0,40
80%	0,19	0,36	0,36	0,39	0,40	0,34
85%	-0,18	0,29	0,33	0,36	0,22	0,30
90%	0,17	0,26	0,17	0,27	0,27	0,37

(a) Variação da taxa de mutação e fixando elitismo em 2% e taxa de seleção em 50%. Taxa de mutação fixa durante evolução

Média de R2	<---Iterações--->					
Seleção	50	100	150	200	250	300
50%	0,43	0,48	0,48	0,48	0,48	0,48
55%	0,47	0,48	0,48	0,48	0,48	0,48
60%	0,46	0,48	0,48	0,48	0,48	0,48
65%	0,44	0,48	0,48	0,48	0,48	0,48
70%	0,45	0,48	0,48	0,48	0,48	0,48
75%	0,47	0,48	0,48	0,48	0,48	0,48
80%	0,46	0,48	0,48	0,48	0,48	0,48
85%	0,46	0,47	0,48	0,47	0,48	0,48
90%	0,44	0,48	0,48	0,48	0,48	0,48
95%	0,45	0,48	0,48	0,48	0,48	0,48
100%	0,47	0,47	0,48	0,48	0,48	0,48

(b) Variação da taxa de seleção fixando taxa de mutação em 15% e taxa de elitismo em 2%. Taxa de mutação fixa durante evolução

Média de R2	<---Iterações--->					
Elitismo	50	100	150	200	250	300
2%	0,43	0,48	0,48	0,48	0,48	0,48
4%	0,46	0,48	0,48	0,48	0,48	0,48
6%	0,47	0,47	0,48	0,48	0,48	0,48
8%	0,44	0,47	0,48	0,48	0,48	0,48
10%	0,44	0,47	0,48	0,48	0,48	0,48

(c) Variação da taxa de elitismo fixando taxa de mutação em 15% e taxa de seleção em 50%. Taxa de mutação fixa durante evolução

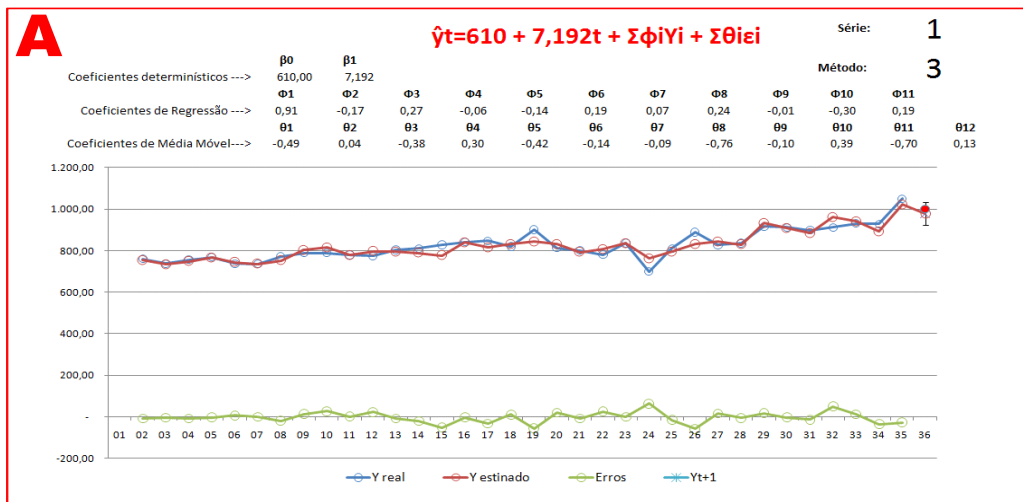
Tabela 9 - Análise de sensibilidade aos parâmetros

Média de R2	<---Iterações--->					
	Mutação	50	100	150	200	250
0%	0,084	0,055	0,321	0,274	0,400	0,434
0%	0,210	0,277	0,418	0,398	0,466	0,464
0%	0,124	0,427	0,405	0,452	0,466	0,465
0%	0,383	0,400	0,462	0,470	0,451	0,477
1%	0,283	0,448	0,434	0,476	0,468	0,477
1%	0,235	0,436	0,469	0,452	0,476	0,474
1%	0,337	0,462	0,471	0,468	0,478	0,478
1%	0,433	0,448	0,473	0,470	0,478	0,477
1%	0,395	0,435	0,467	0,477	0,478	0,480
1%	0,391	0,450	0,475	0,478	0,478	0,480
2%	0,434	0,454	0,478	0,476	0,480	0,481
3%	0,444	0,477	0,481	0,480	0,482	0,482
4%	0,458	0,477	0,476	0,480	0,482	0,482
5%	0,458	0,473	0,480	0,481	0,481	0,482
6%	0,409	0,471	0,477	0,480	0,481	0,482
7%	0,466	0,476	0,481	0,478	0,481	0,482
8%	0,450	0,480	0,481	0,481	0,481	0,482
9%	0,465	0,477	0,481	0,481	0,481	0,482
10%	0,467	0,479	0,482	0,482	0,481	0,482
11%	0,465	0,475	0,480	0,481	0,481	0,482
12%	0,471	0,479	0,478	0,482	0,481	0,482
13%	0,444	0,480	0,481	0,482	0,482	0,482
14%	0,474	0,475	0,481	0,478	0,482	0,482
15%	0,458	0,466	0,481	0,482	0,482	0,481
16%	0,463	0,464	0,478	0,480	0,482	0,481
17%	0,458	0,480	0,476	0,481	0,481	0,482
18%	0,460	0,476	0,481	0,479	0,480	0,482
19%	0,348	0,466	0,480	0,482	0,482	0,482
20%	0,441	0,472	0,474	0,476	0,481	0,479
21%	0,464	0,476	0,480	0,480	0,481	0,481
22%	0,415	0,468	0,478	0,479	0,479	0,482
23%	0,448	0,461	0,481	0,481	0,481	0,482
24%	0,463	0,464	0,480	0,479	0,478	0,481
25%	0,396	0,478	0,474	0,479	0,478	0,481
26%	0,452	0,458	0,480	0,476	0,477	0,480
27%	0,429	0,456	0,480	0,476	0,480	0,478
28%	0,408	0,458	0,469	0,479	0,481	0,479
29%	0,438	0,459	0,476	0,478	0,475	0,481
30%	0,443	0,461	0,471	0,476	0,478	0,481
31%	0,450	0,466	0,450	0,477	0,474	0,479
32%	0,442	0,472	0,476	0,474	0,469	0,477
33%	0,463	0,450	0,473	0,476	0,479	0,480
34%	0,462	0,474	0,464	0,468	0,473	0,473
35%	0,368	0,461	0,458	0,466	0,474	0,478
36%	0,378	0,464	0,458	0,476	0,448	0,473
37%	0,330	0,471	0,477	0,463	0,477	0,469
38%	0,395	0,455	0,459	0,477	0,465	0,479
39%	0,381	0,447	0,430	0,470	0,476	0,472
40%	0,434	0,414	0,439	0,470	0,446	0,478

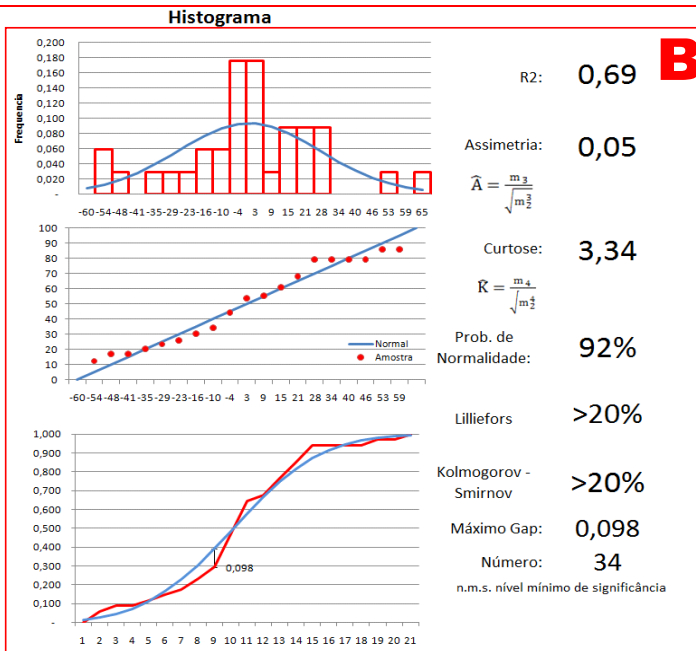
(a) Variação da taxa de mutação e fixando elitismo em 10% e taxa de seleção em 50%. Taxa de mutação fixa durante evolução. O gradiente de cores utilizado para auxiliar na visualização. Valor mínimo vermelho e valor máximo verde.

Tabela 10 - Análise de sensibilidade à taxa de mutação

ANEXO D – DASHBOARD DE ANÁLISE DE AJUSTE E RESÍDUOS



n	Y	\hat{Y}	Erro
13	760,70	753,04	-7,66
14	737,06	732,49	-4,57
15	755,92	748,05	-7,87
16	768,14	765,38	-2,75
17	736,67	743,89	7,22
18	736,09	735,59	-0,50
19	770,87	751,96	-18,91
20	788,56	802,04	13,48
21	787,92	815,40	27,48
22	778,76	779,56	0,80
23	773,85	796,92	23,07
24	802,79	794,35	-8,44
25	809,58	789,11	-20,47
26	828,00	776,91	-51,09
27	841,75	839,02	-2,73
28	847,12	814,50	-32,62
29	819,24	830,51	11,27
30	900,23	844,86	-55,37
31	812,58	832,52	19,94
32	800,81	792,46	-8,35
33	779,98	806,23	26,25
34	836,40	836,72	0,32
35	699,26	763,49	64,23
36	809,98	794,68	-15,30
37	889,02	830,79	-58,23
38	826,26	842,07	15,81
39	835,40	829,37	-6,02
40	915,44	933,10	17,66
41	911,37	908,77	-2,60
42	897,86	884,22	-13,64
43	911,88	961,63	49,76
44	930,25	942,02	11,77
45	927,78	892,52	-35,26
46	1.049,02	1.022,20	-26,82
47	1.000,00	976,57	0,00



C

Teste de Autocorrelação da Série Original												
Defasagem	01	02	03	04	05	06	07	08	09	10	11	12
FAC	0,950	0,992	1,029	1,040	1,060	1,120	1,122	1,120	1,143	1,151	1,158	1,195
Q	34,41	73,01	115,83	161,00	209,49	265,51	323,76	383,96	449,01	517,66	590,00	670,35
Prob	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Teste de Autocorrelação dos Erros de Estimação												
Defasagem	01	02	03	04	05	06	07	08	09	10	11	12
FAC	-0,082	0,014	-0,165	0,073	-0,180	-0,115	-0,175	0,095	0,012	-0,093	0,107	0,116
Q	0,00	0,26	1,37	1,59	2,98	3,57	4,98	5,42	5,43	5,88	6,49	7,26
Prob	1,00	0,88	0,71	0,81	0,70	0,73	0,66	0,71	0,80	0,83	0,84	0,84

FAC- Função de autocorrelação
Q- Teste de Ljung-Box (Hipótese nula é que não existe correlacionamento => pi=0)

Figura 24 - Exemplo de “Dashboard” que combina série 1 com método 3

A

Série: 1

$\hat{y}_t = 610 + 7,192t + \Sigma\phi_i y_i + \Sigma\theta_i e_i$

Método: 3

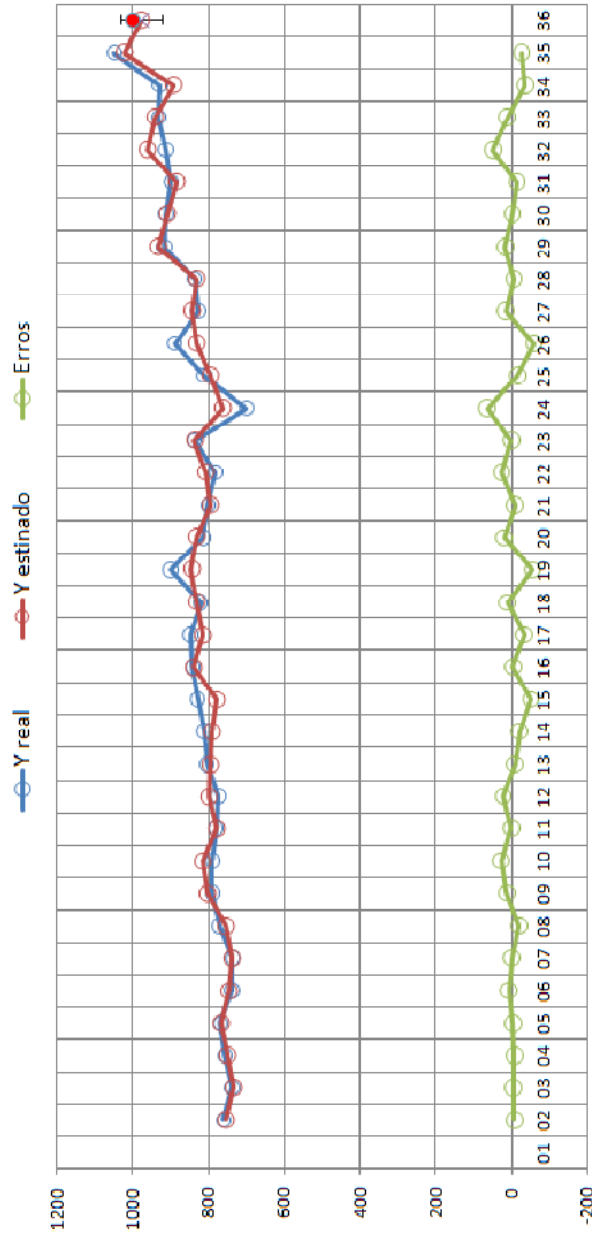
Coefficientes determinísticos ---->

β_0	610,00	β_1	7,192
-----------	--------	-----------	-------

Coefficientes de Regressão ---->

ϕ_1	-0,17	ϕ_2	0,19	ϕ_{10}	-0,30	ϕ_{11}	0,19																
θ_1	0,04	θ_2	-0,42	θ_3	-0,38	θ_4	0,30	θ_5	-0,14	θ_6	-0,14	θ_7	-0,09	θ_8	-0,76	θ_9	-0,10	θ_{10}	0,39	θ_{11}	-0,70	θ_{12}	0,13

Coefficientes de Média Móvel ---->



t	Y real	Y estimado	Erros
01	0,00	732,13	-28,56
02	0,00	760,70	-2,19
03	734,87	737,06	-1,16
04	754,77	755,92	30,26
05	798,40	768,14	0,50
06	737,17	736,67	-1,81
07	734,28	736,09	-8,70
08	762,17	770,87	-31,38
09	757,18	788,56	0,41
10	789,33	787,92	23,95
11	802,71	778,76	-18,89
12	754,97	773,85	-28,95
13	773,84	802,79	-28,95
14	793,68	809,58	-15,90
15	787,15	828,00	-40,85
16	812,49	841,75	-29,27
17	852,10	847,12	4,99
18	838,98	819,24	19,74
19	821,73	900,23	-78,50
20	857,70	812,58	45,11
21	825,06	800,81	24,25
22	832,80	779,98	52,82
23	841,15	836,40	4,76
24	794,02	699,26	94,76
25	789,55	809,98	-20,43
26	837,90	889,02	-51,11
27	863,98	826,26	37,72
28	831,36	835,40	-4,04
29	927,89	915,44	12,45
30	891,23	911,37	-20,13
31	887,22	897,86	-10,64
32	960,83	911,88	48,95
33	947,04	930,25	16,80
34	921,84	927,78	-5,94
35	998,22	1049,02	-50,80
36	981,97	1000,00	0,00

Figura 25 - Coeficientes do ajuste para a série temporal "1"

B

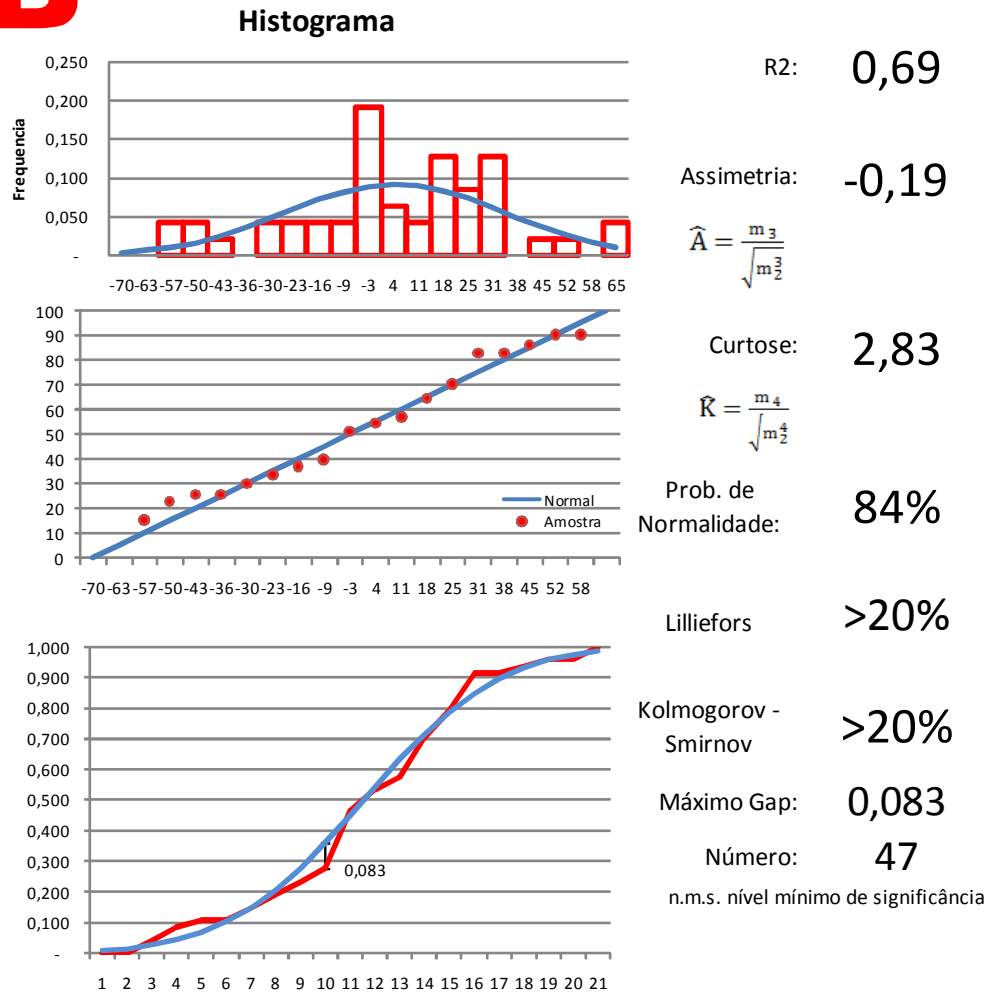


Figura 26 - Análise de normalidade dos resíduos



Teste de Autocorrelação da Série Original												
	01	02	03	04	05	06	07	08	09	10	11	12
Defasagem												
FAC	0,726	0,624	0,685	0,649	0,529	0,623	0,603	0,529	0,601	0,655	0,562	0,641
Q	20,08	35,35	54,32	71,90	83,95	101,31	118,15	131,57	149,58	171,79	188,83	211,99
Prob	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Teste de Autocorrelação dos Erros de Estimação												
	01	02	03	04	05	06	07	08	09	10	11	12
Defasagem												
FAC	-0,048	0,035	-0,172	0,007	-0,206	-0,103	-0,134	0,014	-0,065	0,048	0,123	0,194
Q	0,00	0,13	1,34	1,34	3,17	3,64	4,48	4,49	4,69	4,81	5,63	7,74
Prob	1,00	0,94	0,72	0,85	0,67	0,72	0,72	0,81	0,86	0,90	0,90	0,80

FAC- Função de autocorrelação

Q - Teste de Ljung-Box (Hipótese nula é que não existe correlacionamento => pi=0)

$$Q = T(T + 2) \sum_{j=1}^n \frac{\hat{\rho}_j^2}{T - j}$$

$$FAC = \sum_{k=1}^n \frac{(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{(Y_t - \bar{Y})^2}$$

Figura 27 - Teste de autocorrelação das séries e seus erros de estimação

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](#)