



ISSN 2175-6295 Rio de Janeiro- Brasil, 12 e 13 de agosto de 2010

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

Kleber Silva Leite Pinto

Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)
15054-000, Belo Horizonte, MG
E-mail: kleber@fdc.org.br

Elenice Biazzi

Centro Federal de Educação Tecnológica de Minas Gerais, PPGMMC
Av. Amazonas, 7675, 30.510-000, Belo Horizonte, MG
E-mail: elenice@dppg.cefetmg.br

João Francisco de Almeida Vitor

Centro Federal de Educação Tecnológica de Minas Gerais, PPGMMC
Av. Amazonas, 7675, 30.510-000, Belo Horizonte, MG
E-mail: joaofrancisco@dppg.cefetmg.br

Resumo

Este trabalho tem por objetivo o desenvolvimento de uma ferramenta de apoio ao controle financeiro de uma empresa de serviços educacionais, o qual se dá através do monitoramento dos custos e expurgos dos componentes conhecidos, referentes a variações de volumes físicos, tais como número de horas, número de itens, além dos componentes sazonais e inflacionários. A metodologia proposta busca identificar e extrair dos valores históricos os componentes excepcionais, “*outliers*”, e, com a série resultante, identificar os correlacionamentos de valores passados e suas variações (erro). Desta forma, sintetizar um modelo matemático que melhor simule a série. Obtido o modelo, calcula-se o valor previsto para a série no período $t+1$ e compara-o com o valor a ser contabilizado, que consiste no critério de aceitação ou rejeição. A menor ou maior tolerância dessa diferença dependerá da estatística dos resíduos observados entre a série real e a série modelada. Na construção do modelo foram empregadas técnicas de estimação dos coeficientes do modelo de séries temporais ARMA (Autorregressivo e de Médias Móveis) através da utilização de Algoritmo Genético, processo heurístico que tem sido utilizado na previsão de séries temporais (De Souza, 2006) (Engelbrecht, 2007).

Palavras-chave: Custo, Rateio de custos, Algoritmo genético, Séries temporais, Modelo ARMA.

Abstract:

This paper aims to develop a tool to support financial control in an educational services institution. Such control is carried out by monitoring known component cost and cut-off that refer to volume variation such as the number of hours and items as well as seasonal and inflation components. The methodology used seeks to identify and extract outliers from historical values, and from the resulting series it will identify the correlation of past values and their variation (error). It will thus assemble a mathematical model that will fit the series better. After the model has been obtained, it is used to forecast the amount for the “ $t+1$ ” period and then compared to the actual amount that will be taken into account, that is, the amount that will be either accepted or refused. Lower or higher tolerance for the difference will vary according to the residuals observed between the actual and the modeled series. Genetic Algorithm technics, an heuristic process, has been used to build the ARMA (autoregressive moving average) time series coefficients estimation (De Souza, 2006) (Engelbrecht, 2007).

Key-Words: Costs, Prorated Cost, Genetic Algorithm, Time series, ARMA model.

Exposição do problema

A montagem de um sistema de controle contábil que planeje e registre os seus eventos financeiros tem se mostrado cada vez mais necessária. A identificação de desvios que se refiram a erros, omissões, ou mesmo no caso de registros verdadeiros, possibilita ações por parte da administração para a correção de parâmetros que necessitem correção.

A modelagem matemática de séries temporais, com o uso de técnicas de regressão e médias móveis (ARMA), e o emprego dos Algoritmos Genéticos na estimação dos coeficientes foram utilizados na identificação de anomalias na apuração de custos, através da percepção das diferenças entre os valores estimados e os valores em processo de contabilização. O objetivo é orientar o analista de custos na tarefa mensal de controle financeiro da organização no momento em que são criticados os valores apurados. O critério de aceitação/rejeição dos valores apurados é determinado pela comparação entre o valor esperado e o valor apurado e a tolerância e intervalo de confiança, calculado pelo modelo.

Método

O exercício de previsão busca estabelecer critérios e premissas para a ocorrência de um evento futuro, baseado, pelo menos em parte, no presente ou no passado, assumindo-se que padrões do passado continuarão ocorrendo no futuro. As técnicas de previsão estão classificadas em duas categorias: métodos quantitativos e métodos qualitativos.

Os métodos qualitativos podem ser divididos em exploratórios e normativos. Os primeiros utilizam-se do passado e do presente como ponto de partida e se movem em direção ao futuro de uma maneira heurística, investigando todas as possibilidades possíveis. Os métodos normativos começam pela determinação futura de metas e objetivos, movendo-se na direção do passado para verificar se existem restrições de tecnologias e recursos, no presente, para a sua consecução (Makridakis, 1989).

Os métodos quantitativos são divididos em séries temporais e métodos causais, os quais assumem que os objetivos, variáveis dependentes, têm relação de causa e efeito com uma ou mais variáveis independentes. **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 ser extrapolados no futuro (Wong Francis, 2009).

Decomposição da Série:

A análise das séries temporais considera que cada observação resulta da conjunção de vários efeitos, consistindo esse 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. Registre-se que, 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 (Makridakis, 1989):

$$Z = f(T, C, S, \varepsilon) \quad (1)$$

Tratamento dos dados:

Para a extração dos componentes cíclicos e de tendência conhecidos, foram corrigidas as séries com o ajuste dos custos para uma base única de preço. Foram substituídos os valores das séries cujos resultados são influenciados negativamente, por se tratarem de anomalias advindas de erros. Com essa finalidade, utilizou-se a técnica de estatística robusta (Yohai, 2006)(Meckesheimer, 2007) na substituição dos itens que apresentassem distância superior ao limite de confiança de 90% tomado em relação à mediana calculada da população de quatro itens anteriores e quatro itens posteriores ao ponto considerado.

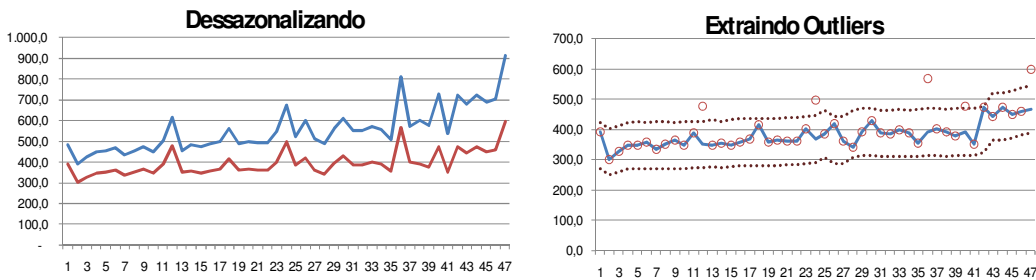


FIGURA 1- Tratamento aplicado à série temporal.

Anterior à aplicação da metodologia proposta, e para que seja possível a aplicação das inferências estatísticas, é necessária a extração da tendência, estacionando-a. Para isso, foi utilizado o modelo de regressão linear no cálculo dos coeficientes determinísticos β_0 e β_1 .

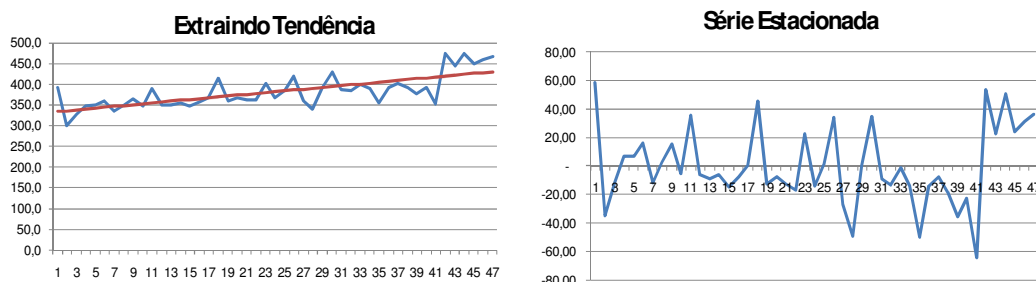


FIGURA 2- Tratamento de estacionamento.

Condição de Validade do Modelo da Série Estacionada:

A aplicação de *inferências* estatísticas em uma série temporal, Y_t , necessita que esta 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(Box, 1976) se:

$$E|Y_t| < \infty$$

$$E(Y_t) = \mu \text{ para todo } t \in Z \text{ e}$$

$$E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma_j$$

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 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 sua variância é constante ao longo do tempo.

Definição do Modelo ARMA:

A publicação, em 1970, por Box e Jenkins (Box, 1970), da obra “TIME SERIES ANALYSIS: FORECAST AND CONTROL” instalou uma nova geração de ferramentas de previsão, considerada a mais eficiente e a mais utilizada para previsão de séries temporais(De Souza, 2006), conhecida como metodologia Box&Jenkins ou modelo ARIMA. “AR” refere-se a

autorregressivo, “I” ao fato de ser integrado, na hipótese de se fazer necessário o estacionamento da série tomando-se as diferenças entre seus elementos, forma de eliminação de tendência, e “MA” 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, a variável dependente Y é explicada exclusivamente por suas variáveis independentes X_i ; neste modelo, as variáveis explicativas são os próprios valores da variável dependente defasado no tempo ($X_i = Y_{t-1}$) juntamente com os componentes estocásticos (Gujarati, 2004). Dentro desse conceito, é montado o modelo em que Y_t é representado pela soma dos itens antecessores da série (Y_{t-1}) multiplicados pelo coeficiente Φ_{t-1} e a soma dos erros de estimação antecessores ($a_{i-1} = Y_{t-1} - \hat{Y}_{t-1}$) multiplicados pelo coeficiente θ_{t-i} .

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

Algoritmos Genéticos:

A técnica dos algoritmos genéticos, AG, é uma meta-heurística de busca e otimização de soluções. Pertence à família dos Algoritmos Evolucionários, que buscam inspiração na teoria evolucionária proposta por Darwin¹ e tem seus mecanismos aplicados em modelos computacionais. Consideram uma população inicial de indivíduos, correspondendo, cada um destes, a uma solução candidata à resolução do problema. Os indivíduos, a cada geração, competem entre si, buscando lugar na geração seguinte, diretamente ou através de descendência.

Os indivíduos mais aptos, ou seja, aqueles cuja função de “*fitness*” apresentar melhores valores têm uma maior probabilidade de serem escolhidos para progenitores da geração seguinte. São tomadas de empréstimo à Biologia as noções e definições que se referem à codificação de soluções, 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 dada posição (*locus*) do cromossomo. A maioria das aplicações de algoritmos genéticos utiliza indivíduos haploides. Geralmente os cromossomas têm um número de genes e cada gene tem uma localização fixa dentro do cromossoma. Por esse motivo, a representação em que um cromossoma corresponde a um vetor, sendo os genes os elementos desse vetor, é muitas vezes utilizada.

Duas soluções progenitoras serão combinadas, dando origem a dois descendentes. Outro tipo de operador para a produção de expansão da área de busca é operador de Mutação, que atua introduzindo mudanças aleatórias no cromossoma dos indivíduos.

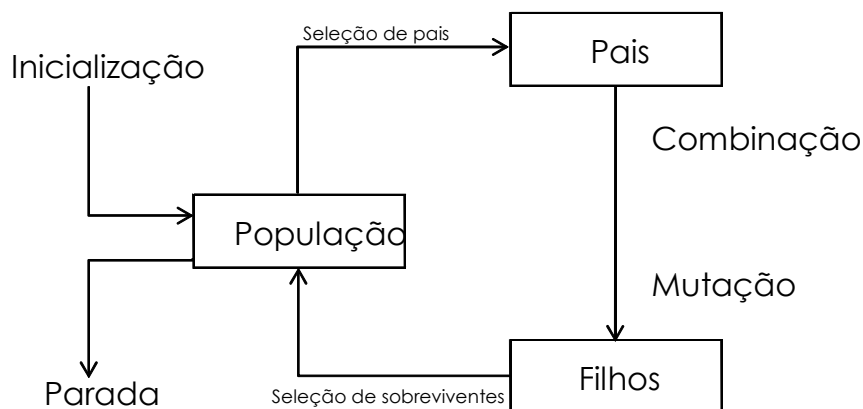


FIGURA 3 - Esquema genérico de um algoritmo evolucionário (Smith, 2003).

Esses operadores influenciam nas habilidades de sobrevivência e reprodução dos indivíduos. Sem o emprego da mutação, a probabilidade de estagnação da população com a subsequente

¹ Charles Robert Darwin (* 12 de fevereiro de 1809 — † 19 de abril de 1882). Cientista inglês que elaborou a moderna teoria da evolução, a qual se baseia em um processo denominado seleção natural.

convergência prematura do algoritmo seria muito alta. Dada uma probabilidade predefinida de mutação, os parâmetros de cada indivíduo poderão sofrer alterações, possibilitando a sua fuga de um ótimo local da função de avaliação, ampliando, assim, o espaço de busca. O aumento exagerado da taxa de mutação distorce a estrutura genética, o que pode implicar a perda de bom material genético de indivíduos mais aptos. Entretanto, pesquisas mostraram que uma taxa maior de mutação que decresça exponencialmente com o aumento do número de mutações favorece a velocidade de convergência e a acurácia do algoritmo (Engelbrecht, 2007)

Cada sequência de operações de seleção dos mais fortes, cruzamento e mutação dá origem a uma nova geração (*off-spring*), seguida de nova avaliação de desempenho dos indivíduos que a constituem. O ciclo de evolução é interrompido quando uma condição predefinida de parada é satisfeita, a qual poderá ocorrer se o algoritmo tiver obtido a solução ótima do problema pelos desempenhos dos vários indivíduos, geração após geração, e convergir para um determinado valor, por terem ocorrido: 1) um número suficiente de gerações, ou 2) por outra razão estabelecida, que são definidos como critérios de parada do algoritmo.

Estimação dos coeficientes:

Representação e inicialização

Respeitando os limites do espaço de busca e as defasagens máximas de 12 períodos para a regressão e para a média móvel, a representação genética do indivíduo foi estabelecida como representação real com 24 genes, um 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.

TABELA 1

Representação dos coeficientes no cromossomo

| | | | | | | | | | | | | | | | | | |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------|------------|-----------|-----------|-----------|-----------|-----------|-----------|------|------------|
| 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | | 12 | 13 | 14 | 15 | 16 | 17 | 18 | | 24 |
| Φ1 | Φ2 | Φ3 | Φ4 | Φ5 | Φ6 | Φ7 | Φ8 | | Φ12 | 01 | 02 | 03 | 04 | 05 | 06 | | 012 |

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 que possuem maior aptidão possuem uma probabilidade maior de serem selecionados para a próxima geração, e este pode ser considerado, então, 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 esse indivíduo propague as suas características para a população seguinte. Para a seleção dos indivíduos que participam da combinação 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 na geração de novos indivíduos.

Função de aptidão

Foi utilizado o indicador de qualidade R², que representa a relação da soma quadrada dos erros de estimação e a soma dos quadrados das diferenças dos valores da série em relação a 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 R² se aproxime de 1(um). O significado dessa proximidade é a capacidade do modelo de 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 quando dois pais gerassem dois filhos. Na situação inversa, descartado, quando o material genético na operação de troca (*cross-over*) fosse para geração de único filho (*offspring*). Em outro operador de combinação foi introduzida a possibilidade de selecionar-se 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, variando 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.

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).

TABELA 2
Parâmetros do Algoritmo

| Módulo de População | Operadores | Melhores Resultados |
|------------------------------------|--|----------------------------|
| Técnica de Representação | <i>Real</i> | |
| Técnica Inicialização da População | <i>Progressiva</i> | |
| Técnica Eliminação da População | <i>Elitismo 5-10%</i> | <i>Elitismo 20%</i> |
| Técnica de Reprodução | <i>Troca da geração</i> | |
| Técnica de Seleção de Genitores | <i>Proporcional/Roleta</i> | <i>Proporcional/Roleta</i> |
| Técnica de Aptidão | <i>R² coeficiente de determinação</i> | |
| Tamanho da População | <i>100</i> | |
| Número de Gerações (parada) | <i>50-300</i> | <i>300</i> |
| Operadores | <i>Combinação linear</i> | |
| Taxa Mutação | <i>1-90%(Gene=0,04-3,45%)</i> | <i>90%(Gene=3,45%)</i> |
| Taxa <i>Crossover</i> | <i>50-100%</i> | <i>Indiferente</i> |

Avaliação de qualidade:

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, quando o padrão que caracteriza o comportamento da série temporal tiver sido isolado, o desvio remanescente representará uma aleatoriedade, ruído branco, cuja 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, foram utilizadas as seguintes análises sobre os resíduos:

- 1) O coeficiente de determinação (R²), que representa a relação entre as variações explicadas pelo modelo e variação total dos valores da série temporal relativamente a cada verba estudada na forma abaixo, em que SQErro é a soma de quadrado dos erros de estimação e SQTotal a variância da série estimada.

$$R^2 = 1 - \frac{SQErro}{SQTotal} \quad (3)$$

- 2) Para verificação da normalidade dos resíduos, são utilizados os testes de *Lilliefors*, que se utilizam da métrica de verificação do desvio máximo entre a distribuição acumulada dos resíduos estudados e a densidade esperada por uma distribuição Gaussiana. Esse teste apresenta a vantagem de ser adequado para pequenas amostras e é desconhecida a variância da sua população (Campos, 1979).
- 3) Para a verificação da existência de covariância entre os resíduos, que garantiria serem esses ruídos brancos, atestando que o modelo identificou todos os componentes não aleatórios, é proposto o teste Ljung-Box, que 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 (Bueno, 2008).

$$Q = T(T + 2) \sum_{j=1}^n \frac{\hat{\rho}_j^2}{T - j} \quad (4)$$

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

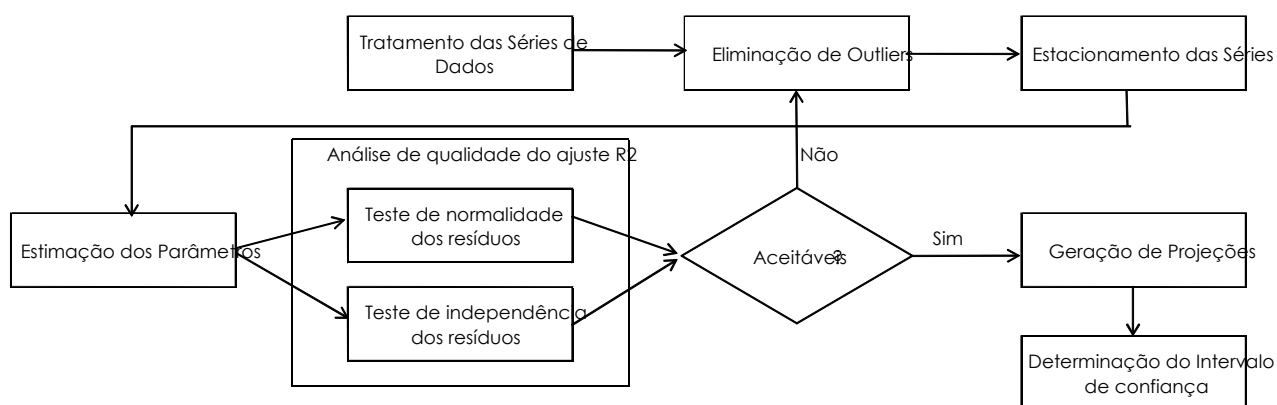


FIGURA 4 - Fluxograma do processo de projeção.

Resultados

Teste de robustez:

Para avaliação da eficiência do algoritmo genético foram gerados para cada uma das 6 (seis) séries temporais 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 deste método.

TABELA 2

Análise de sensibilidades aos parâmetros do algoritmo

| Média de R2 | Iterações--> | | | | | |
|-------------|--------------|------|------|------|------|------|
| Mutação | 50 | 100 | 150 | 200 | 250 | 300 |
| 1% | -0,20 | 0,11 | 0,26 | 0,30 | 0,35 | 0,39 |
| 11% | 0,40 | 0,46 | 0,48 | 0,49 | 0,49 | 0,49 |
| 21% | 0,44 | 0,48 | 0,49 | 0,49 | 0,49 | 0,49 |
| 31% | 0,45 | 0,48 | 0,49 | 0,49 | 0,49 | 0,49 |
| 41% | 0,46 | 0,49 | 0,49 | 0,49 | 0,49 | 0,49 |
| 51% | 0,46 | 0,49 | 0,49 | 0,49 | 0,49 | 0,49 |
| 61% | 0,47 | 0,49 | 0,49 | 0,49 | 0,49 | 0,49 |
| 71% | 0,47 | 0,49 | 0,49 | 0,49 | 0,49 | 0,49 |
| 81% | 0,47 | 0,49 | 0,49 | 0,49 | 0,49 | 0,49 |
| 91% | 0,47 | 0,49 | 0,49 | 0,49 | 0,49 | 0,49 |
| Total | 0,39 | 0,45 | 0,47 | 0,47 | 0,48 | 0,48 |

| Média de R2 | Iterações--> | | | | | |
|-------------|--------------|------|------|------|------|------|
| Seleção | 50 | 100 | 150 | 200 | 250 | 300 |
| 50% | 0,39 | 0,44 | 0,47 | 0,46 | 0,48 | 0,48 |
| 55% | 0,40 | 0,44 | 0,45 | 0,47 | 0,47 | 0,48 |
| 60% | 0,40 | 0,46 | 0,46 | 0,47 | 0,48 | 0,48 |
| 65% | 0,35 | 0,46 | 0,47 | 0,47 | 0,48 | 0,48 |
| 70% | 0,40 | 0,44 | 0,46 | 0,47 | 0,47 | 0,48 |
| 75% | 0,38 | 0,45 | 0,48 | 0,47 | 0,48 | 0,48 |
| 80% | 0,39 | 0,45 | 0,47 | 0,48 | 0,48 | 0,49 |
| 85% | 0,39 | 0,44 | 0,47 | 0,47 | 0,48 | 0,49 |
| 90% | 0,40 | 0,43 | 0,46 | 0,47 | 0,48 | 0,49 |
| 95% | 0,40 | 0,45 | 0,47 | 0,47 | 0,47 | 0,48 |
| 100% | 0,39 | 0,45 | 0,47 | 0,48 | 0,47 | 0,48 |
| Total | 0,39 | 0,45 | 0,47 | 0,47 | 0,48 | 0,48 |

| Média de R | Iterações--> | | | | | |
|------------|--------------|------|------|------|------|------|
| Elitismo | 50 | 100 | 150 | 200 | 250 | 300 |
| 2% | 0,38 | 0,44 | 0,47 | 0,47 | 0,48 | 0,48 |
| 4% | 0,39 | 0,45 | 0,47 | 0,47 | 0,48 | 0,48 |
| 6% | 0,38 | 0,46 | 0,46 | 0,47 | 0,48 | 0,48 |
| 8% | 0,40 | 0,44 | 0,47 | 0,47 | 0,48 | 0,48 |
| 10% | 0,39 | 0,44 | 0,46 | 0,48 | 0,48 | 0,48 |
| Total | 0,39 | 0,45 | 0,47 | 0,47 | 0,48 | 0,48 |

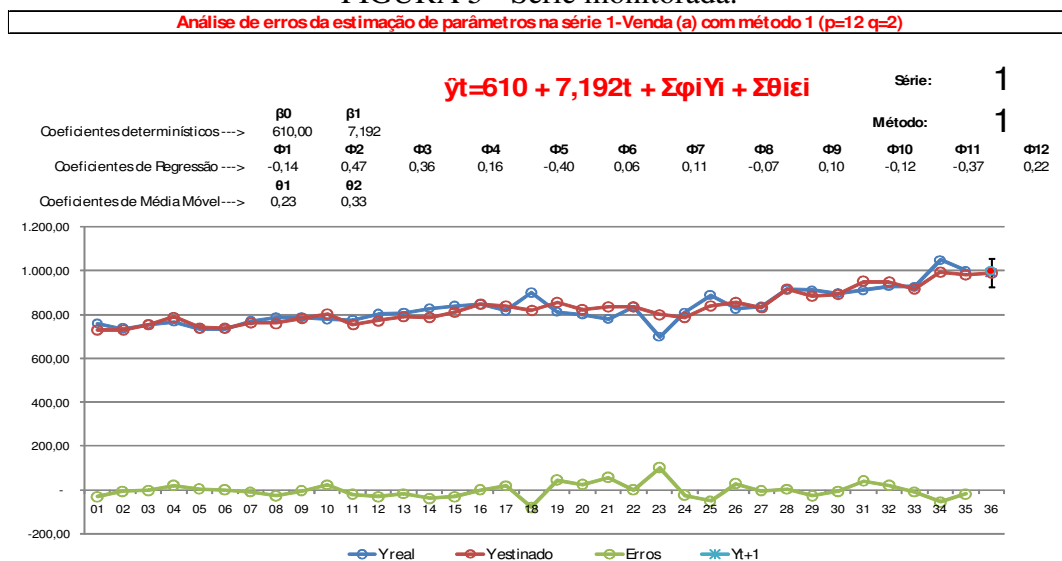
O resultado da série “1” mostra que é possível observar a rápida convergência do algoritmo para atingir os valores ótimos da regressão (igual a 0,49) com um pequeno número de iterações. Em menos de 150 iterações, o resultado já se posicionou bastante próximo dos valores máximos obtidos quando se utilizava maior número de iterações das séries (300 iterações). Nota-se também a pequena sensibilidade aos parâmetros de elitismos e aos parâmetros de seleção pela pequena variação ao variar esse parâmetro.

O comportamento do algoritmo para ocorrências superiores a 300 iterações se mostrou pouco compensadora, já que o ganho de qualidade no crescimento de R2 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 o crescimento desse parâmetro. Esse fato sugeriu a introdução de modulação, mecanismo de variação desse parâmetro durante a execução do programa. Buscou-se a aceleração da mutação nos primeiras gerações, decrescendo o parâmetro mPM (percentual de mutação) com a aproximação do critério de parada. Dessa forma, apresentou-se na fase inicial uma maior dinâmica para exploração do espaço de busca e, na final, maior precisão. É quando se buscava refinar os resultados diminuindo o risco de se afastar de conquistas obtidas nos ótimos locais. Os resultados alcançados pelo algoritmo, quando expandido para utilizar os parâmetros de calibração de médias móveis, apresentaram-se satisfatórios, mostrando-se superiores aos alcançados com a regressão simples, ainda que inferiores aos obtidos com o processo de estimação ARMA.

Critério de Aceitação:

A capacidade preditiva do modelo para de cada uma das seis séries foi verificada tomando-se por base o encerramento do exercício contábil de dezembro de 2009, com a utilização dos dados de quatro anos, retroativos a esta data.

FIGURA 5 - Série monitorada.



Como o interesse desta aplicação é verificar se o valor do item de mão de obra em processo de apuração em determinado mês no exercício do registro contábil é aceitável, os desvios em relação ao valor esperado são tolerados, desde que não ultrapassem a estatística definida no

representarem erros de coleta de dados (possibilitando a sua pronta correção), neste último caso justificando a sua imediata correção.

As ferramentas de apoio à atividade de análise financeira têm se beneficiado de grande avanço nos últimos anos, proporcionado pelas melhorias na tecnologia da informação, que permitem 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 trabalho investigativo que, no experimento descrito, levou à conclusão de que 50% das séries analisadas eram conformes, dispensando, conseqüentemente, exames mais pormenorizados, com grau de confiabilidade de 90% de segurança. Na realidade das empresas não são raras as situações em que o custo do controle excede o custo do risco exposto pelo não controle.

Referências

- Box G.E.P. and Jenkins, G.M.** Time series analysis: Forecast and Control [Book]. - San Francisco : Holden-Day, 1970.
- Box George E.P** Time Series Analysis: forecasting and control [Book]. - Oakland - USA : Holden-Day, 1976.
- Bueno Rodrigo de Losso da Silveira** Econometria de séries temporais [Book]. - São Paulo : Cengage Learning, 2008.
- Campos H** Estatística Experimental Não-Paramétrica [Book]. - Piracicaba : USP, 1979.
- De Souza Luzia Vidal** Tese Doutorado - Programação Genética e Combinação de Preditores para Previsão de Séries Temporais. - Curitiba : Universidade Federal do Paraná, 2006.
- Engelbrecht Andries P.** Computational Intelligence [Book]. - West Sussex, England : John Wiley & Sons, 2007.
- Gujarati Damodar N.** Basic Econometrics. - New York : McGraw-Hill, 2004.
- Makridakis Spyros G.** Forecasting Methods for Management [Book]. - New York : John Wiley & Sons, 1989.
- Meckesheimer S. Basu and M.** Automatic outlier detection for time séries: an application to sensor data [Article] // Knowledge and Information Systems. - 2007. - 11. - Vol. 2.
- Núñez-Letamendia Laura e M. de Molina** Fitting the control parameters of a Genetic Algorithm to optimise technical trade rules [Article] // European Journal of Operational Research. - 2007. - pp. 179(3):847-868.
- Smith A.E. Eiben and J.E.Smith** Introduction to Evolutionary Computing [Book]. - Berlin : Springer, 2003.
- Wong Francis Pan Yong Tan** Neural Networks and Genetic Algorithm for Economic Forecasting [Article] // AI in Economics and Business Administration. - 2009.
- Yohai V.J., Marona A.R. e Martin R.D.** Robust Statistics [Book]. - West Sussex-England : John Wiley & Sons, 2006.