



PREVISÃO DE SÉRIES TEMPORAIS DO MERCADO DE AÇÕES UTILIZANDO PROGRAMAÇÃO GENÉTICA COM INSPIRAÇÃO QUÂNTICA

Reinaldo Gomes Moraes

reimorais1986@hotmail.com

Douglas Mota Dias

douglasmota.uerj@gmail.com

Nadia Nedjah

nadia@eng.uerj.br

Luiza Macedo Mourelle

ldmm@eng.uerj.br

Universidade do Estado do Rio de Janeiro

R. São Francisco Xavier, 524 - Cep 20550-900 - Maracanã - Rio de Janeiro - RJ – Brasil

Abstract. *Financial market's time series forecasts still a hard challenge. It is research field that has been explored in various scientific works to demonstrate the precision of many methods of time series forecasting. The genetic programming for time series forecasting gained more visibility, because of its effectiveness. Moreover, in some cases, it provided high quality forecasts in comparison to the results obtained by well-established statistical techniques. The proposal of this study is to test the performance of Quantum-Inspired Linear Genetic Programming for time series forecasting. A set of IBOVESPA Brazilian stocks is used to check if there is an improvement in performance when the principles of quantum mechanics are explored. For this purpose, the statistical method ARIMA is considered for comparing, as well the Tree-based Genetic Programming.*

Keywords: *Time series. forecast, Ibovespa stocks, ARIMA, genetic programming, quantum-inspired.*

INTRODUÇÃO

Um dos maiores desafios relacionados à economia de mercado está em prever os movimentos de ações antecipadamente, realizando uma boa estimação para precificação dos ativos. A completa previsibilidade do seu comportamento não é algo possível devido a muitas variáveis intrínsecas ao seu resultado, como: especulações e opiniões, elementos políticos, realidades sociais, entre outros. No entanto, a existência de um modelo para o mesmo pode ser de grande interesse econômico, mesmo que apresente uma baixa acurácia, pois, dado que se trata de um processo de alta volatilidade, mesmo apresentando baixa acurácia apresenta melhores resultados do que seguir uma linha conservadora de investimento.

Em sistemas complexos desta natureza, são usados modelos estatísticos e de inteligência computacional na tentativa de alcançar um resultado mais competitivo.

Dentre as principais técnicas de inteligência computacional temos a Computação Evolutiva, que é o termo empregado para designar o conjunto de técnicas de resolução de problemas baseados em princípios de evolução biológica.

Uma das principais frentes de pesquisa da Computação Evolutiva é a Programação Genética (PG), concebida por John Koza (1992), possuindo como proposta a solução de problemas através geração automática de programas de computador com seus mecanismos baseados na Seleção Natural de Darwin (1859).

O mecanismo de busca da PG pode ser descrito como um ciclo “criar-testar-modificar”, semelhante a forma com que os humanos desenvolvem seus programas, o que pode ser um processo computacionalmente custoso.

Como uma alternativa para contornar este problema, começou-se a usar processos baseados na mecânica quântica, que explora o paralelismo inerente à superposição de estados quânticos, chegando-se assim ao algoritmo de Programação Genética Linear com Inspiração Quântica (PGLIQ), método desenvolvido por Dias, em 2010.

Apesar do grande potencial da PGLIQ, até o presente momento não havia sido testado seu desempenho para *forecasting*, tão pouco em séries históricas do mercado financeiro, que são um problema com grande grau de complexidade.

Este trabalho tem por objetivo verificar o desempenho da técnica de PGLIQ em relação à PG e o modelo estatístico *Autoregressive Integrated Moving Average* (ARIMA).

Como medida de performance considerou-se a previsão de cada mês do último trimestre de 2014 das séries temporais do preço de fechamento de 2 ações da IBOVESPA, pontuadas por especialistas no site Infomoney (2014) como as possíveis boas oportunidades de investimento no mercado acionário para o ano de 2014: Itaú Unibanco (ITUB4) e Kroton (KROT3).

O artigo está dividido em seis seções adicionais. Na Seção 1 é apresentado um resumo da técnica de previsão ARIMA. Em seguida, na Seção 2, é explicado o conceito do método de programação genética e seu funcionamento. Posteriormente, na Seção 3, é realizada a apresentação sobre o funcionamento da programação genética com inspiração quântica. Na Seção 4, é visto as configurações de cada um dos modelos para o procedimento experimental. Após, na Seção 5, são detalhados os resultados obtidos em cada uma das duas ações com a utilização dos três modelos desenvolvidos seguindo cada uma das técnicas. Por fim, na Seção 6, descreve as conclusões do trabalho.

1 MODELOS ARIMA

Os modelos ARIMA foram apresentados por Box e Jenkins no livro *Time Series Analysis: forecasting and control*, em 1970, e são usados para a previsão usando séries temporais, que são registros que exibem sequências de informações periódicas relativas a um item de interesse (Ferreira et al, 2014). São modelos de fácil implementação e robustos do ponto de vista estatístico, que usam as correlações entre as observações em diversos instantes.

Como casos particulares dos modelos ARIMA temos os processos AR (Auto Regressivo), em que o valor do processo é expresso apenas com um agregado linear dos valores anteriores e a adição de distúrbios ε_t e MA (Médias móveis), sendo um processo linear unicamente dependente de um número finito q de ruídos brancos.

Resumidamente, podemos observar que os modelos ARIMA representam as séries temporais como uma ponderação dos próprios valores e/ou erros passados da série.

1.1 ARIMA Não Sazonal e Sazonal

A construção dos modelos ARIMA seguem as etapas iterativas apresentadas na Figura 1:

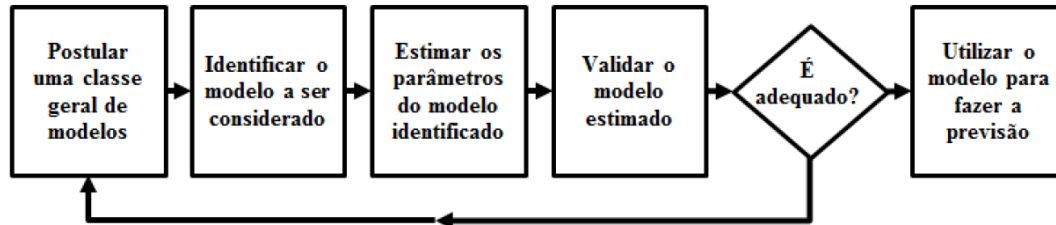


Figura 1 - Fluxograma de construção de modelos ARIMA

A metodologia ARIMA de Box-Jenkins estima modelos de séries temporais da forma:

$$\Phi_p(B)(1-B)^d y_t = a + \Theta_q(B)\varepsilon_t \quad (1)$$

Em alguns casos as séries temporais apresentam um componente sazonal que se repete a cada s observações. Desta forma, para lidar com esta sazonalidade os processos da técnica ARIMA foram generalizados, sendo chamados de SARIMA.

O modelo SARIMA pode ser formulado como:

$$\Phi_P(B)\phi_p(B)(1-B)^D(1-B)^d y_t = a + \Theta_Q(B)\theta_q(B)\varepsilon_t \quad (2)$$

onde temos para estes modelos ARIMA:

p : número de parâmetros relativos à defasagem do fator Auto Regressivo;

d : grau da tendência de variação da série;

q : número de parâmetros relativos à defasagem da Média Móvel;

$\Phi(B)$: operador auto-regressivo e um polinômio de grau p ;

$\theta(B)$: operador da média móvel e um polinômio de grau q ;

P : número de parâmetros relativos à defasagem do fator Auto Regressivo sazonal;

D : grau da tendência de variação sazonal;

Q : número de parâmetros relativos à defasagem da Média Móvel sazonal;

$\phi(B)$: é o operador auto-regressivo e um polinômio de grau P ;

$\theta(B)$: é o operador da média móvel e um polinômio de grau Q ;

B : é o operador de movimentação na série

ε_t : é um erro aleatório com distribuição normal $N(0, \sigma^2)$;

a : é uma constante;

y_t : são os dados da série temporal.

2 PROGRAMAÇÃO GENÉTICA

A Programação Genética é uma extensão dos algoritmos genéticos, entretanto, diferentemente dos algoritmos genéticos, onde cada gene costuma ser representado por um vetor de valores binários, na Programação Genética cada gene costuma ser modelado por uma árvore que, por sua vez, possui toda a informação relevante sobre um determinado indivíduo (Seshadri, 2003). Basicamente, a técnica segue os princípios de seleção natural de Darwin (1859), genética (reprodução, cruzamento e mutação), aprendizado de máquina (busca e heurística) e representação de programas como árvores sintáticas.

Como exemplo, a Figura 2 apresenta a estrutura da árvore para a expressão matemática $x^3 + 5x + 2$. Semanticamente, esta árvore poderia ser a aproximação de uma função e avaliação da qualidade da expressão matemática criada é medida a partir da diferença entre os resultados da função original e a criada.

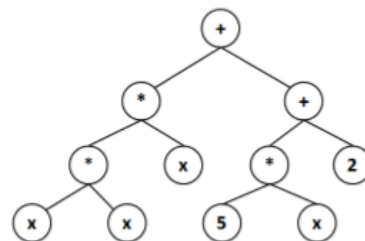


Figura 2 - Fluxograma de construção de modelos programação genética

Segundo Koza (1992) e Li (2000), citado por Prochnowa (2013), programação genética é uma forma de busca no espaço de soluções, e seu algoritmo poderia ser resumido em 3 passos:

1. Definir uma população inicial de indivíduos formados por combinações aleatórias de funções e terminais.
2. Executar as seguintes tarefas até satisfazer o critério de parada (como o número máximo de gerações ou um erro mínimo, por exemplo):
 - a. Avaliar cada membro da população, atribuindo-lhe um valor de aptidão (*fitness*).
 - b. Criar uma nova população a partir da existente aplicando, com diferentes probabilidades, operadores de seleção e reprodução apropriados.
3. Escolher o indivíduo que melhor representa o resultado da execução do algoritmo.

A Figura 3 apresenta os passos citados acima em um fluxograma. Os conceitos apresentados nesses passos serão vistos na seção a seguir.

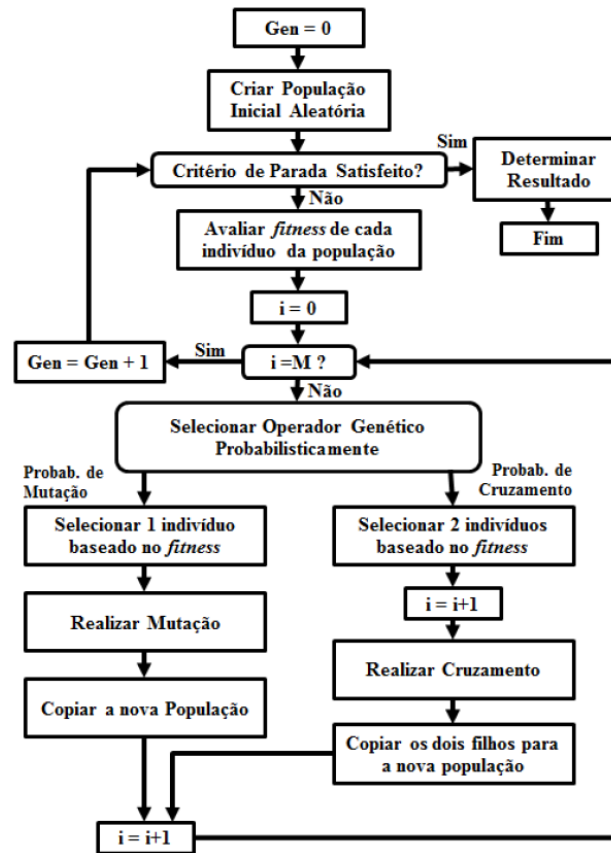


Figura 3 - Fluxograma representando a execução do algoritmo de programação genética (KOZA, 1992)

3 PROGRAMAÇÃO GENÉTICA COM INSPIRAÇÃO QUÂNTICA

A Programação Genética Linear com Inspiração Quântica (PGLIQ) é um novo modelo geral de algoritmos evolucionários inspirados em computação quântica (AEIQ) para a síntese automática de programas de computador, que pode ser aplicado aos mais diversos tipos de problema, como: ajuste de curvas, regressão simbólica, processamento de imagens e sinais, etc. Assim como na PG convencional, são utilizados os operadores de adição, subtração, multiplicação e divisão, além de instruções aritméticas, trigonométricas e de saltos condicionais.

A característica mais representativa dos algoritmos de inspiração quântica é a representação das soluções, pois, diferentemente dos tradicionais algoritmos evolutivos, a população é representada por *qubits* e consiste de distribuição de probabilidades das amostras no espaço de busca, ao invés de pontos exatos no espaço (Moore et al, 1995). No entanto, quando medido, um indivíduo quântico pode indicar um elemento exato do espaço de busca. O diagrama na Figura 4 tem como objetivos ilustrar a estrutura do modelo e auxiliar na introdução ao seu funcionamento básico.

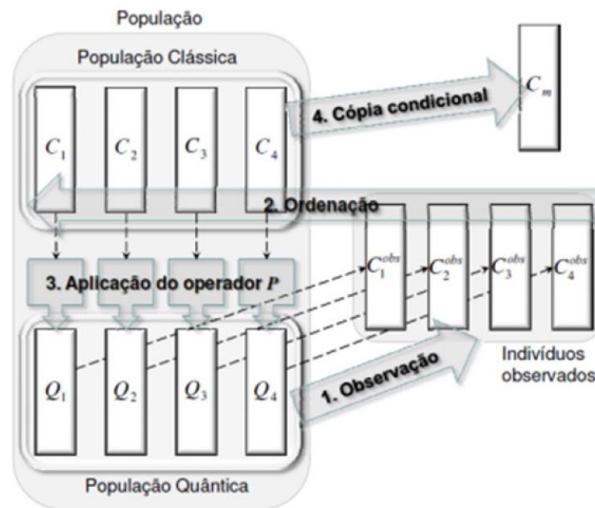


Figura 4 - Diagrama descritivo básico do modelo PGLIQ (DIAS, 2010)

Conforme Dias (2010), com relação à estrutura, a PGLIQ possui uma população (híbrida) que, por sua vez, é composta por duas populações, uma quântica e outra clássica, ambas possuindo o mesmo número M de indivíduos. Também possui M indivíduos clássicos auxiliares C_i^{obs} , que resultam das observações dos indivíduos quânticos Q_i , onde $1 \leq i \leq M$. No caso exemplificado pelo diagrama da Figura 4, $M = 4$.

Além da estrutura, o mesmo diagrama apresenta, enumeradas, as 4 etapas básicas que caracterizam uma “geração” do modelo, as quais são descritas a seguir (Dias, 2010):

1. Cada um dos M indivíduos quânticos é observado uma vez, resultando nos M indivíduos clássicos C_i^{obs} .
2. É efetuada a ordenação conjunta dos indivíduos da população clássica e dos indivíduos observados quanto às suas avaliações. Como resultado, os M melhores indivíduos clássicos dentre os $2M$ avaliados são mantidos na população clássica, ordenados do melhor para o pior, de C_1 para C_M , os demais M indivíduos clássicos permanecem armazenados e ordenados, do melhor para o pior, de C_1^{obs} para C_M^{obs} .
3. O operador P , que representa a funcionalidade de uma porta quântica, é aplicado a cada indivíduo da população quântica, tomando como referência seus indivíduos correspondentes na população clássica. É nesta etapa que ocorre efetivamente a evolução, uma vez que, a cada nova geração, a aplicação do operador aumenta a probabilidade de que a observação dos indivíduos quânticos gerem indivíduos clássicos mais parecidos com os melhores encontrados até então.
4. Caso algum dos indivíduos da população clássica, avaliados na geração atual, seja melhor que o melhor indivíduo clássico avaliado até então (em relação às gerações anteriores), uma cópia do mesmo é armazenada em C_M , o melhor indivíduo clássico encontrado pelo algoritmo até então.

4 PROCEDIMENTO EXPERIMENTAL

Neste trabalho utilizou-se o Projeto R (RPROJECT, 2012) juntamente com a ferramenta R Studio (RSTUDIO, 2012) para realizar as previsões dos métodos ARIMA e Programação Genética, além de gerar os gráficos referentes a cada uma das execuções. Para a execução do modelo PGLIQ foi usado o algoritmo concebido por Dias (2010), em C e C#, utilizando-se o ambiente integrado de desenvolvimento *Microsoft Visual Studio*.

Para este artigo, foram escolhidas duas séries históricas do mercado de ações, de forma que, a partir dos testes pudessem apresentar a diferença de acurácia das técnicas escolhidas. Do Mercado de Ações foram obtidas as informações de fechamento ajustado para dividendos e desdobramentos, medidas diariamente, referentes às ações das empresas Itaú Unibanco (ITUB4) e Kroton (KROT3), disponibilizadas no site Yahoo Finanças, em 2015. Foram utilizados os dados entre 02 de janeiro de 2012 e 30 de dezembro de 2014 para as duas séries.

A avaliação dos resultados das técnicas leva em conta as previsões de cada mês do último trimestre de 2014, sendo assim, os testes consideram 3 subconjuntos para cada série histórica para processamento, ou seja, foram previstos cada mês considerando os demais meses anteriores para treino, somando assim 6 previsões realizadas no total. A Tabela 1 mostra quantos dias foram previstos em cada mês.

Tabela 1. Dias previstos em cada mês testado

Mês	Out/14	Nov/14	Dez/14
Qtde de dias	23	19	20

Na realização das previsões utilizando o modelo ARIMA, foi usada a primeira diferença (ou seja, a diferença entre um dia d e o dia anterior $d - 1$) do log das séries. Esta transformação foi feita para diminuir a variância das séries. A título de exemplo da separação das base de treino no modelo ARIMA, considerando a ação ITUB4, para a previsão do mês de outubro/14, foi usado para treino o log da série iniciando em 02 de janeiro de 2012 até 30 de setembro 2014.

Nas previsões usando programação genética, foi usada a primeira diferença do log das séries, sendo o log considerado na criação das séries devido a este operador não pertencer ao conjunto de operadores possíveis usado na evolução das funções da programação genética. Para a predição foram usados valores de até 23 dias antes para prever os valores futuros. A título de exemplo, considerando a ação ITUB4, para a previsão do mês de março/14, foi usada para treino a série iniciando em 02 de janeiro de 2012 até 03 de janeiro de 2014 como variável explicativa e como variável resposta 02 de fevereiro de 2012 até 31 de janeiro de 2014. Como função *fitness* foi usado o *MAPE* (*Mean Absolut Percentual Error*), formulada como:

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{Valor Real_t - Valor Predito_t}{Valor Real_t} \right| \quad (3)$$

onde N é a quantidade de valores a serem preditos.

A previsão usando a programação genética com inspiração quântica segue os critérios de criação da variável explicativa e variável resposta usados na programação genética convencional mostrada acima, bem como o *MAPE* para função *fitness*.

4.1 Configuração dos Modelos ARIMA

O método ARIMA depende da definição de parâmetros para ser executado e, após uma série de testes com diferentes combinações de valores, foram definidos os valores como segue na Tabela 2. A ação KROT3 apresentou a configuração de um modelo ARIMA, enquanto que a ação ITUB4, devido a existência da componente sazonal Q , usou um modelo SARIMA.

Tabela 2. Configuração dos parâmetros do ARIMA

Ação	Mês Previsto	p	d	q	P	D	Q
KROT3	out/14	0	1	0	0	0	0
	nov/14	0	1	0	0	0	0
	dez/14	0	1	0	0	0	0
ITUB4	out/14	0	1	0	0	0	1
	nov/14	0	1	0	0	0	1
	dez/14	0	1	0	0	0	1

4.2 Configuração dos modelos PG

Foram realizadas várias combinações diferentes de parâmetros na tentativa de alcançar uma melhor acurácia. Os testes utilizaram as séries selecionadas contendo no mínimo 471 pontos cada uma, e com a configuração para a Programação Genética apresentada na Tabela 3.

Tabela 3. Configuração para programação genética

Parâmetros	Valor
Conjuntos de funções	+, -, *, Seno, Coseno
Conjuntos de terminais	Série defasada*, constantes de 0,025 a 1
Número de gerações	10000
Tamanho da população	100
Função de <i>fitness</i>	<i>MAPE</i>
Número de selecionados por elitismo	10

*Séries históricas anterior ao período a ser previsto, com seu final defasado em até 23 dias, conforme definido anteriormente.

4.3 Configuração dos modelos PGLIQ

A configuração dos parâmetros do algoritmo PGLIQ é mostrada pela Tabela 4. Note que gsm_{max} é o número máximo de gerações executadas, M é o tamanho da população, gsm_{max} é o número máximo de gerações sem melhoria, $p_{0,0}$ é a probabilidade de NOPs (nenhuma instrução), máximo do programa, s_T é o tamanho do passo de reinício, P_s é o tamanho do passo do operador, L é o tamanho máximo do programa, e finalmente R é o número máximo de registros usados. Os demais parâmetros que são comuns aos PG são configurados conforme Tabela 3.

Tabela 4. Configuração para o PGLIQ

Parâmetros	Valor
$gsm_{max}; p_{0,0}$	20.000; 0,9
$s_T; P_s$	1; 0,004
$L; R$	128; 8

5 RESULTADOS DOS EXPERIMENTOS

Nesta seção, são apresentados os resultados para cada uma das ações, onde será apresentada a série original, a tabela com o *MAPE* e *MAE* (*Mean Absolut Error*) de cada mês e os gráficos dos resultados das previsões para cada mês de 2014 separadamente. O *MAE* segue a mesma fórmula do *MAPE*, sem a divisão pelo valor real, e foi usado somente como indicador adicional para comparação entre os modelos,

A Figura 5 mostra o gráfico da série histórica da KROT3 com o período utilizado para o experimento. Os resultados obtidos com a utilização das técnicas aplicados à série KROT3 são apresentados na Figura 6.

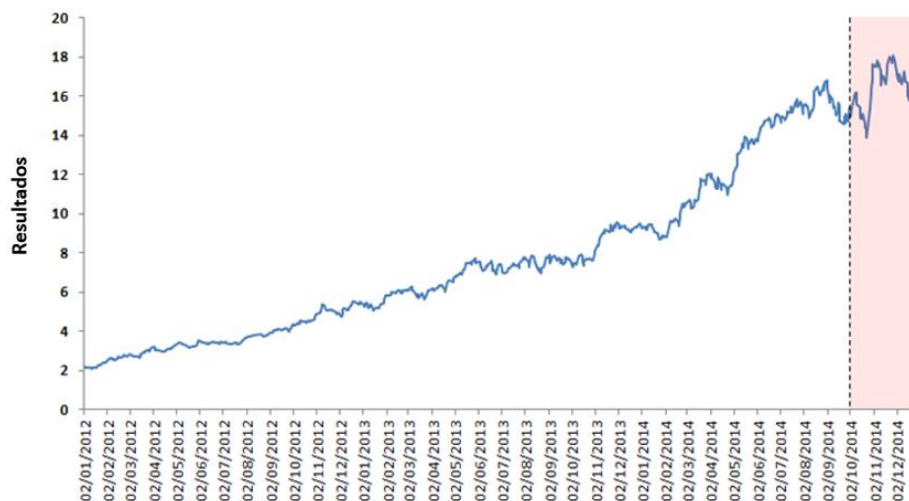


Figura 5 - Série Histórica KROT3 de jan/12 a dez/14

A Figura 6 mostra os gráficos dos períodos previstos e suas respectivas previsões para cada técnica, com o ARIMA apresentando o melhor desempenho no período avaliado. O PG foi a técnica que apresentou os piores resultados em novembro.

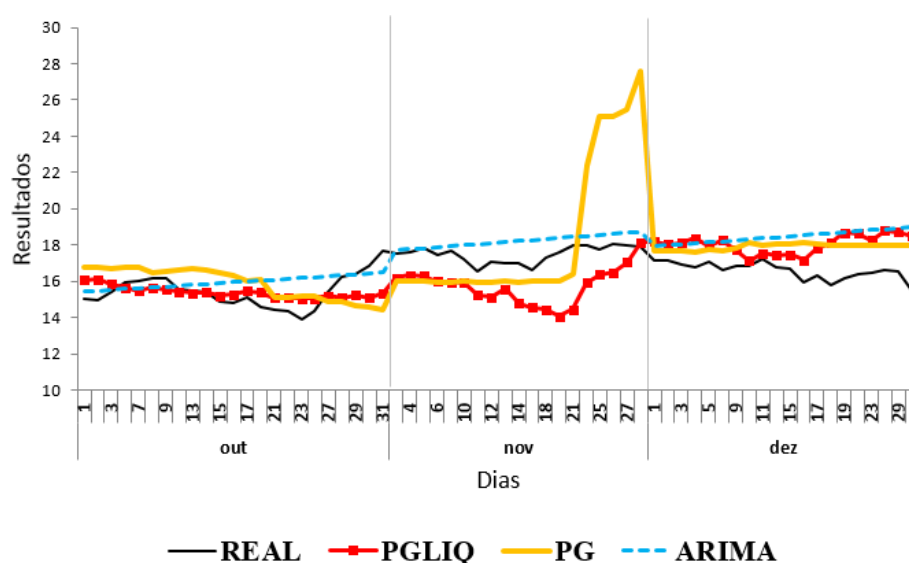


Figura 6 - Predição para os meses do último trimestre de 2014 da ação KROT3

A Figura 7 mostra o gráfico da série histórica da ITUB4 com o período utilizado para o experiment e a Figura 8 mostra os gráficos dos períodos previstos e suas respectivas previsões para cada técnica.

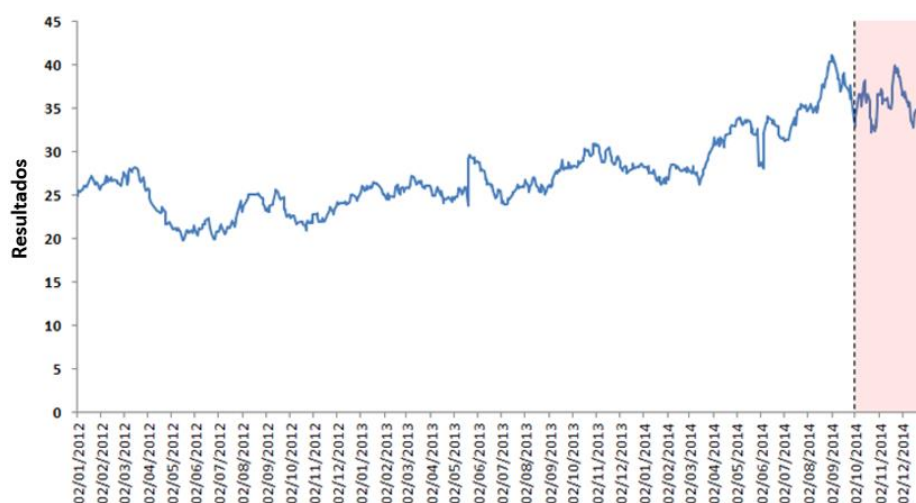


Figura 7 - Predição para os meses do último trimestre de 2014 da ação ITUB4

A partir da Figura 8 é possível observar que nenhuma das três técnicas conseguiu explicar bem o comportamento da série KROT3 entre outubro e dezembro. O PG apresentou os piores resultados, com destaque ao mês de out/14.

A Tabela 5 apresenta os resultados encontrados por cada técnica em cada uma das duas ações. A partir da primeira parte da Tabela 5 referente à ação KROT3, pode-se observar que o modelo PGLIQ apresentou os melhores resultados que o PG em dois meses avaliados,

entretanto, foi superior ao ARIMA apenas em outubro/14. Já na segunda parte da tabela referente à ação ITUB4, observa-se que o modelo PGLIQ apresentou um resultado de qualidade superior ao PG nos três meses avaliados e possuiu um resultado superior ao ARIMA apenas no mês de dezembro/14.

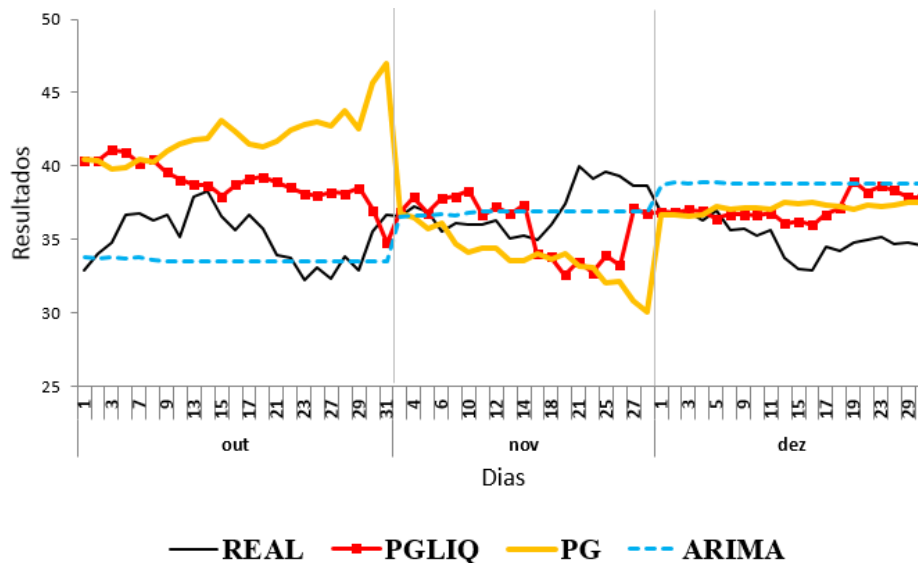


Figura 8 - Predição para os meses do último trimestre de 2014 da ação ITUB3

Tabela 5. Resultados por mês de predição das ações KROT3 e ITUB4

Ação	Mês	Valor Médio	MAE			MAPE		
			PGLIQ	PG	ARIMA	PGLIQ	PG	ARIMA
KROT3	out/14	R\$ 15,44	R\$ 0,73	R\$ 1,25	R\$ 0,80	4,67%	8,05%	5,35%
	nov/14	R\$ 17,48	R\$ 1,77	R\$ 2,85	R\$ 0,73	10,16%	16,05%	4,26%
	dez/14	R\$ 16,59	R\$ 1,46	R\$ 1,32	R\$ 1,87	8,89%	8,03%	11,37%
ITUB4	out/14	R\$ 35,16	R\$ 3,92	R\$ 6,91	R\$ 1,97	11,39%	19,97%	5,44%
	nov/14	R\$ 37,10	R\$ 2,57	R\$ 3,34	R\$ 1,35	6,77%	8,75%	3,59%
	dez/14	R\$ 35,17	R\$ 2,05	R\$ 2,05	R\$ 3,66	5,93%	5,97%	10,53%

6 CONCLUSÃO

Os resultados alcançados com a técnica de programação genética com inspiração quântica sugerem o forte potencial desta técnica no uso para previsão de séries temporais, conforme resumido na Tabela 6.

Para constatar a eficácia da técnica, foi realizada uma série de experimentos comparativos utilizando ações e períodos diferentes, onde chegou-se à conclusão que os modelos de PGLIQ

com apenas uma defasagem como variável explicativa e uma quantidade menor de indivíduos avaliados pode apresentar erros de previsão menores que a técnica PG tradicional.

Na comparação com os resultados obtidos usando o modelo ARIMA, consegue-se notar que o ARIMA leva vantagem, muito por conta de considerar mais de uma variável explicativa para a previsão de alguns casos, enquanto o PG e o PGLIQ consideram apenas uma defasagem de até 23 dias. Como um possível trabalho futuro, pretende-se investigar uma implementação de processamento paralelo do PGLIQ visando acelerar as previsões, bem como o uso de mais defasagens na previsão.

Tabela 6. MAPE no último trimestre de 2014.

Ação	PGLIQ	PG	ARIMA
KROT3	7,71%	10,49%	6,96%
ITUB4	8,21%	12,01%	6,52%

REFERÊNCIAS

Box, George; Jenkins, Gwilym. *Time series analysis: Forecasting and control*, San Francisco: Holden-Day, 1970.

Darwin, Charles. *On the Origin of Species by Means of Natural Selection*. Londres, 1859.

Dias, Douglas Mota. *Programação genética linear com inspiração quântica*. Departamento de Engenharia Elétrica, PUC-Rio, RJ, 2010.

Ferreira, M.; Silva, L. C.; Martins, N. H.; Campagnac, V.; Pinto, A. S.; Oliveira, J. B. P.; Dirk, R. C. *51 Ferramentas para análise criminal*, 2014.

Infomoney. *As melhores e piores ações para 2014, segundo um dos maiores gestores do mundo*. Disponível em: <<http://www.infomoney.com.br/onde-investir/acoes/noticia/3103161/melhores-piores-acoes-para-2014-segundo-dos-maiores-gestores-mundo>>. Acesso em: fevereiro/2015.

Koza, John R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. 1ª ed. Cambridge: MIT Press, 1992.

Moore, M.; Narayanan, A. *Quantum-inspired computing*. Dept. Comput. Sci., Univ. Exeter, Citeseer, Exeter, UK, 1995.

Prochnow, Fabio Alberto. *Programação Genética para Predição de Séries Temporais Aplicados a Mercados Financeiros*. Instituto de Informática, UFRGS, 2013.

RProject. *The R Project for Statistical Computing*. Disponível em: <http://www.rproject.org/>>. Acesso em: fevereiro/2015.

Seshadri, Mukund. *Comprehensibility, Overfitting and Co-Evolution in Genetic Programming for Technical Trading Rules*. Worchester Polytechnic Institute, 2003.

Yahoo Finanças. *Cotações - Bolsas de Valores*. Disponível em: <https://br.financas.yahoo.com/q>>. Acesso em fevereiro/2015.