

5 Solução por Regressão Simbólica

5.1. Introdução

Neste capítulo é descrito um estudo de caso utilizando-se o modelo proposto no capítulo 4.

Pretende-se com esse estudo de caso, mostrar a viabilidade do modelo, suas vantagens e limitações.

No capítulo 4 é proposta uma metodologia dividida em três blocos. No estudo de caso proposto, será detalhado como cada bloco é aplicado ao caso estudado.

O estudo de caso utiliza o software de “Análise de alternativas de desenvolvimento de um campo sob condições de incerteza de mercado” [25]. Este software modela o problema de investimento em desenvolvimento de campos de petróleo como um problema de opções reais retornando curvas de exercício ótimo e seus valores da opção respectivos. Para o investimento no desenvolvimento de campos de petróleo, há as variáveis utilizadas nos problemas de investimento em opções financeiras (volatilidade, taxa de juros livre de risco, preço de longo prazo, preço inicial, ciclo de vida, etc.) e incertezas relativas a qualidade e quantidade da reserva de petróleo. O software de análise, foi utilizado como Gerador de Amostras para curvas de exercício ótimo.

Utilizando o conjunto de amostras como parâmetro, foi utilizada a biblioteca de PG *OpenBEAGLE* descrita por Gagné em [23]. O *OpenBEAGLE* foi configurado e utilizado para implementar um software que retorna uma função analítica que melhor define este conjunto de amostras. O software implementado é o bloco de Programação de Genética descrito no capítulo 4.

Como parâmetro para bloco de cálculo do valor da opção por SMC, foi utilizada a função analítica da curva de exercício ótimo gerada pelo software de PG. Esta função foi parametrizada no EXCEL, onde foi desenvolvida um macro

em Visual Basic que utiliza a função parametrizada para calcular o valor da opção.

A seguir, são descritos com mais detalhes os procedimentos do estudo de caso.

5.2. Geração de Amostras das Curvas de Exercício Ótimo

O software de “Análise de alternativas de desenvolvimento de um campo sob condições de incerteza de mercado” possui as seguintes entradas de parâmetros já configurados com os respectivos valores padrão:

- Tempo de final do término da opção (em anos) - T : 2.
- Preço atual do Petróleo (US\$/bbl) - P_0 : 20.
- Taxa de conveniência da commodity (*Convenience Yield*) (em % a.a.) - d : 0.08.
- Taxa de desconto exógena ao risco em US\$ (em % a.a.) - r : 0.12
- Preço máximo a ser atingido pelo petróleo (US\$/bbl) - P_{max} : 60.
- Volatilidade do preço do petróleo (em % a.a.) - σ : 0.25.
- Taxa de juros livre de risco em US\$ (em % a.a.) - r : 0.08.
- Quantidade Média de Barris Estimada no Campo (Milhões) - B : 400.
- Velocidade de Reversão à Média - h : 0.3466
- Preço de Longo Prazo do petróleo (US\$/bbl) - \bar{P} : 20.

Além dos parâmetros acima, há o custo de investimento que é análogo ao preço do exercício e a incerteza relativa à qualidade econômica da reserva. Para estas duas variáveis, há três opções, pois o software trabalha na tomada de decisão de investimento entre três alternativas de investimento em um campo de petróleo. Para estes parâmetros, foram designados os seguintes valores:

- Custo de Investimento 1 (US\$/bbl) - $D1$: 1600.
- Custo de Investimento 2(US\$/bbl) - $D2$: 1650.
- Custo de Investimento (US\$/bbl) - $D3$: 2500.
- Qualidade Econômica da Reserva 1 - $q1$: 0.08.

- Qualidade Econômica da Reserva 2 – q_2 : 0.09.
- Qualidade Econômica da Reserva 3 – q_3 : 0.22.

Através destes valores geram-se curvas de exercício ótimo para alternativa 3 (D_3 e q_3) que é utilizada como amostra.

O software ainda possui parametrização do processo estocástico utilizado podendo ser MGB ou PRM de Bhattacharya, descrito pela equação (15). Para o estudo de caso, foi escolhido o PRM de Bhattacharya por ser intrínseco ao tema pesquisado. Para o PRM de Bhattacharya foram escolhidos os intervalos discretos para a trajetória do preço do petróleo ΔP ajustado para o valor 0.1 e para o tempo Δt ajustada para o valor 10^{-6} .

Para este estudo de caso foram geradas 50 curvas de exercício ótimo. As curvas de exercício ótimo utilizaram todos os valores padrão mencionados acima, exceto nos parâmetros r e \mathbf{s} onde houve variação. Para o parâmetro \mathbf{s} , foram selecionados valores aleatórios que variam entre 0.05 e 0.5; para o parâmetro r , foram selecionados valores aleatórios que variam entre 0.01 e 0.17.

5.3.

Função analítica de uma curva de exercício ótimo por PG

Para o implementar o software de PG que evolui funções para retornar a função analítica da curva de exercício ótimo, foi utilizada a biblioteca de PG *OpenBEAGLE*.

Para evoluir uma função analítica para a curva de exercício ótimo, é necessário configurar os parâmetros de controle da PG. Além disso é interessante adicionar terminais relativos à função da curva de exercício ótimo na intenção de facilitar a busca por uma solução ótima. A seguir são descritos os terminais adicionados a evolução PG que tem relação à função para o cálculo da curva de exercício ótimo. Em seguida são descritos os parâmetros de controle da PG.

5.3.1.

Terminais para a Evolução da Função da Curva de Exercício Ótimo

Para a geração da função analítica da curva de exercício ótimo, utilizam-se os parâmetros citados em 5.2 com os respectivos valores citados. Cada parâmetro

desses, passa a ser para a PG um candidato a terminal em um indivíduo durante a evolução PG.

Além dos terminais já citados, são inseridos fragmentos da fórmula de Black e Scholes como terminais Os fragmentos são:

- DDNOM: $s\sqrt{T-t}$
- $d_1: \frac{\ln((qBP_0)/D) + (r + s^2/2)T}{s\sqrt{T}}$
- $d_2: d_1 - s\sqrt{T}$
- JPS: $r + \frac{s^2(T-t)}{2}$
- JMS: $r - \frac{s^2(T-t)}{2}$

Cada fragmento da fórmula de Black e Scholes foi apelidado para facilitar a codificação da evolução e a procura por uma boa solução dentro da plataforma PG.

5.3.2. Valores para os Parâmetros de Controle da PG

Nesta seção, são mostrados os valores para os parâmetros de controle necessários para a evolução e obtenção de resultados expressivos. Alguns parâmetros foram modificados para cada um dos resultados alcançados, porém alguns permaneceram fixos. Os parâmetros que permaneceram fixos são exibidos nesta seção enquanto os modificados, são exibidos na seção que mostra os resultados.

Os parâmetros que se mantiveram fixos são:

- Probabilidade de ocorrência de Mutação por encolhimento: 0.05. A mutação por encolhimento consiste em trocar um ramo por um de seus sub-ramos.
- Probabilidade de ocorrência de Mutação: 0.05. É a mutação descrita no capítulo 3.

- Probabilidade de ocorrência de Mutação por Troca de Primitivas: 0.05. Esta mutação consiste em trocar funções com o mesmo número de argumentos de dois ramos diferentes.
- Probabilidade de selecionar pontos para Mutação por Troca de Primitivas: 0.5. É a probabilidade de escolher pontos mais externos (terminais) ou mais internos (funções).
- Probabilidade de ocorrência de Permutação: Nulo.
- Freqüência de Edição: Nulo
- Freqüência de Encapsulamento: Nulo
- Probabilidade para Reprodução: 0.05.
- Probabilidade de ocorrência de *crossover*: 0.9.
- Probabilidade de selecionar pontos para *crossover*: 0.9.
- Tamanho máximo permitido de um indivíduo criado por *crossover*: 10.
- Tamanho máximo de árvore permitido para um indivíduo ao ser criado na população inicial ou geração 0: 5.
- Tamanho mínimo de árvore permitido para um indivíduo ao ser criado na população inicial ou geração 0: 2.
- Método de geração da população inicial: *Ramped Half and Half* [13][14][16].
- Método de geração dos pais para *crossover* (primeiro e segundo): *Rank*.
- Uso de aptidão ajustada: sim.

A aptidão ajustada utilizada é função da aptidão padronizada. De acordo com o capítulo 3, a aptidão padronizada deve levar o melhor o mais próximo de zero em casos de minimização do erro. Para o caso estudado, foi utilizada a seguinte aptidão padronizada:

$$s(i, t) = \frac{r(i, t)}{N} \quad (31)$$

Sendo N o número total de indivíduos e $r(i, t)$ a aptidão bruta do indivíduo i , na geração t . Para a avaliação dos indivíduos foi utilizada a aptidão normalizada descrita na equação (28).

Para ajustar os parâmetros das probabilidades de ocorrência de mutação e crossover foi utilizada a taxa adaptativa que aumenta a probabilidade de ocorrência de mutação e diminui a de crossover ao longo da evolução:

$$taxa = \frac{PI - PF}{MAXGEN} \quad (32)$$

Onde *PI* é a Probabilidade Inicial e *PF* a Probabilidade Final do operador genético ambos definidos como parâmetros da evolução. *MAXGEN* é o número máximo de gerações definido como parâmetro da evolução.

Os valores para os parâmetros variáveis são apresentados na seção a seguir junto aos resultados.

5.3.3. Resultados da PG

Foram realizadas diversas evoluções durante a implementação desta pesquisa, porém, vale destacar apenas três das evoluções que retornaram resultados significativos. Destas três funções analíticas para a curva de exercício ótimo, é selecionada a melhor para utilizar-se no próximo item relacionado ao cálculo do valor da opção por SMC. Os resultados são acompanhados dos valores configurados para a evolução da PG.

O primeiro resultado apresentado possui os seguintes parâmetros além dos já citados em 5.3.2:

- Máximo de gerações: 200.
- Máximo de indivíduos por geração: 1000.
- Elitismo: 5 indivíduos.

Além disso, foi utilizada uma máquina com processador de 2.4 GHz e 512Mb de memória RAM tendo a evolução durado três dias. A avaliação obtida através da aptidão normalizada (item 3.4.4) foi 0.66. A função obtida apresenta-se a seguir:

$$P^*(r, \mathbf{s}, t, \dots) = \left\{ \left[\left((\mathbf{h} + \bar{P}) + (\mathbf{d} + d_2) + \bar{P} \right)^{N(d_1)} + \bar{P} + (t \times \bar{P}) + P_0 + (t \times \bar{P}) + d_2 + \bar{P} \right]^{N(s\sqrt{T-t})} \right\}^{N(d_1)}$$

$$+ \left[(d_2 + \bar{P}) + e' + \left((2 \times (t \times \bar{P}))^{\sqrt{s}} + d_1 + \bar{P} \right) \right]^{N(s\sqrt{T-t})} + P_0 + \mathbf{s} \sqrt{T-t}$$

Onde $N(\)$ significa a normal cumulativa do valor entre parênteses.

O segundo resultado foi gerado pelos seguintes parâmetros utilizados para inicializar a evolução:

- Máximo de gerações: 250.
- Máximo de indivíduos por geração: 4000.
- Elitismo: 50 indivíduos.

A evolução teve duração de 1 semana em uma máquina com processador de 2.4 GHz e 512Mb de memória RAM, e obteve-se, uma avaliação de 0.77 para a seguinte equação:

$$P^*(r, \mathbf{s}, t, \dots) = \left[\frac{N \left\{ e^{(t-\sqrt{T})/(r+t+e^{d_1})} \right\} \times T \times \ln(\bar{P})}{N(JPS)} \right] \left(\frac{DDNOM}{P} \right)^{h(r+s)}$$

$$\times N \left[\left\{ \left(\frac{(\mathbf{r} + D)}{(DDNOM)^r} \right)^{\left(\frac{\mathbf{s}}{P} \right)^{\left(\frac{h}{\sqrt{DDNOM}} \right)^{JPS}}} \right\} \right] \times \sqrt{D} \times \ln(T)$$

Este resultado foi utilizado como semente para uma nova evolução, ou seja, esta função é re-inserida em uma nova evolução como vários indivíduos clonados.

A nova evolução foi realizada com os seguintes parâmetros:

- Máximo de gerações: 250.
- Máximo de indivíduos por geração: 1000.
- Elitismo: 50 indivíduos.
- Número de clones da semente: 100 indivíduos.

A nova evolução teve duração de 4 dias para uma máquina de 1.8 GHz e 512 Mb de RAM, alcançou a avaliação 0.79:

$$P^*(r, \mathbf{s}, t, \dots) = \left[\frac{N \left\{ e^{(t-\sqrt{T})/(t+e^{d_1})} \right\} \times T \times \ln(\bar{P})}{N(JPS)} \right]^{\left(\frac{DDNOM}{\bar{P}} \right)^{h(r+s)}} \\ \times N \left\{ \left[\frac{\frac{t/r + D}{DDNOM^{d_1}} + (\mathbf{s}/\bar{P}) \sqrt{D}^{JPS}}{(\mathbf{s}/\bar{P})^r} \right]^{\left(\frac{\mathbf{s}}{\bar{P}} \right)^{h/\sqrt{DDNOM}^{JPS}}} \right\} \times \sqrt{D} \times \ln(T)$$

Com os resultados anteriores, foram analisados as respectivas conformidades das curvas encontradas com um conjunto de validação. Para análise foram empregadas duas medidas utilizadas em conformidade de séries e comportamento de padrões. As medidas são “Raiz Quadrada do Erro Médio”, “Erro Médio Absoluto Percentual”:

A “Raiz Quadrada do Erro Médio” ou “Root Mean Square Error” (RMSE), é uma medida de distância voltada para o quanto perto a previsão ficou do real. A sua avaliação é dada por:

$$RMSE = \sqrt{\frac{\sum_{k=1}^N (r_k - p_k)^2}{N}} \quad (33)$$

Onde r_k é o valor real conhecido de uma variável no seu estado k e p_k é o valor previsto desta variável no estado k e N é o número de estados utilizados para a variável.

O “Erro Médio Absoluto Percentual” ou “Mean Absolute Percentage Error” (MAPE), como o próprio nome já diz, é uma medida absoluta e não relativa como é o RMSE. Sua equação é dada por:

$$MAPE = \frac{\sum_{k=1}^N \left[\frac{r_k - p_k}{r_k} \right]}{N} \times 100\% \quad (34)$$

Através destas medidas foram comparados os resultados obtidos com a curva de mesmos parâmetros do conjunto de validação. Nas tabelas abaixo encontram-se os valores escolhidos para validação, para os três casos:

Juros	Volatilidade	RMSE	MAPE %
0.04	0.25	0.418024	0.988855
0.06	0.25	0.485543	1.265198
0.08	0.25	0.623191	1.769067
0.1	0.25	0.781657	2.284573
0.08	0.15	0.5752	1.804056
0.08	0.20	0.568114	1.684772
0.08	0.30	0.706612	1.93282
0.08	0.35	0.784555	2.057869
0.08	0.40	0.859297	2.146048

Tabela 3 - Dados para validação conforme métricas de previsão relativos ao indivíduo com avaliação normalizada de 0.66.

Juros	Volatilidade	RMSE	MAPE %
0.04	0.25	0.270168	0.753849
0.06	0.25	0.216488	0.598295
0.08	0.25	0.146439	0.375483
0.1	0.25	0.082379	0.208078
0.08	0.15	0.632053	2.016481
0.08	0.20	0.373024	1.121278
0.08	0.30	0.124712	0.286749
0.08	0.35	0.207905	0.524481
0.08	0.40	0.219712	0.507695

Tabela 4 - Dados para validação conforme métricas de previsão relativos ao indivíduo com avaliação normalizada de 0.77.

Juros	Volatilidade	RMSE	MAPE %
0.04	0.25	0.181625	0.474913
0.06	0.25	0.138446	0.362882
0.08	0.25	0.100954	0.262634
0.1	0.25	0.092577	0.229314
0.08	0.15	0.500863	1.569081
0.08	0.20	0.271866	0.775043
0.08	0.30	0.161623	0.400284
0.08	0.35	0.223979	0.57265
0.08	0.40	0.211906	0.478214

Tabela 5 - Dados para validação conforme métricas de previsão relativos ao indivíduo com avaliação normalizada de 0.79.

Os resultados na Tabela 3, Tabela 4 e Tabela 5, mostram que o MAPE se apresentou satisfatório para todos os casos, apresentando em média, erros inferiores a 1%, entre a curva real e a prevista.

Como exemplo da discrepância entre as métricas de RMSE e MAPE, pode-se utilizar o par $r=0.08$ e $s=0.15$ na Tabela 3 onde o RMSE de 0.50 parece baixo, porém o MAPE, apesar de baixo, apresenta uma diferença exagerada dos MAPEs encontrados para as demais parametrizações, o que não é refletido de mesmo modo pelo RMSE em relação aos demais RMSEs.

Os resultados exibidos na Tabela 3, Tabela 4 e Tabela 5 estão ilustrados a seguir através das curvas:

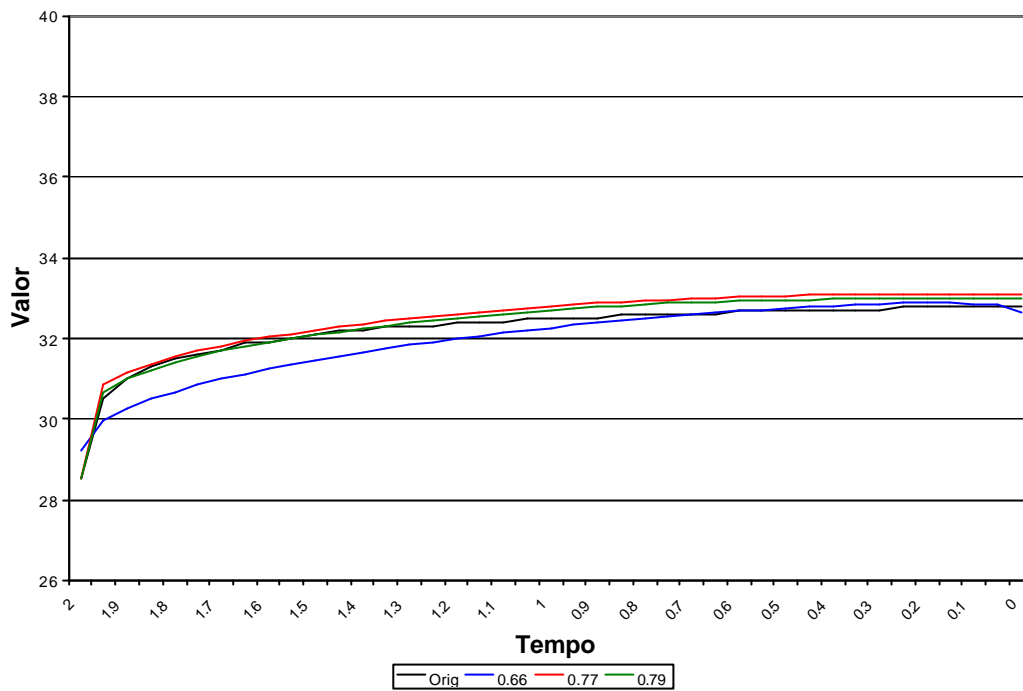


Figura 13 - GP X Curva Original, juros com 4% e volatilidade de 0.25.

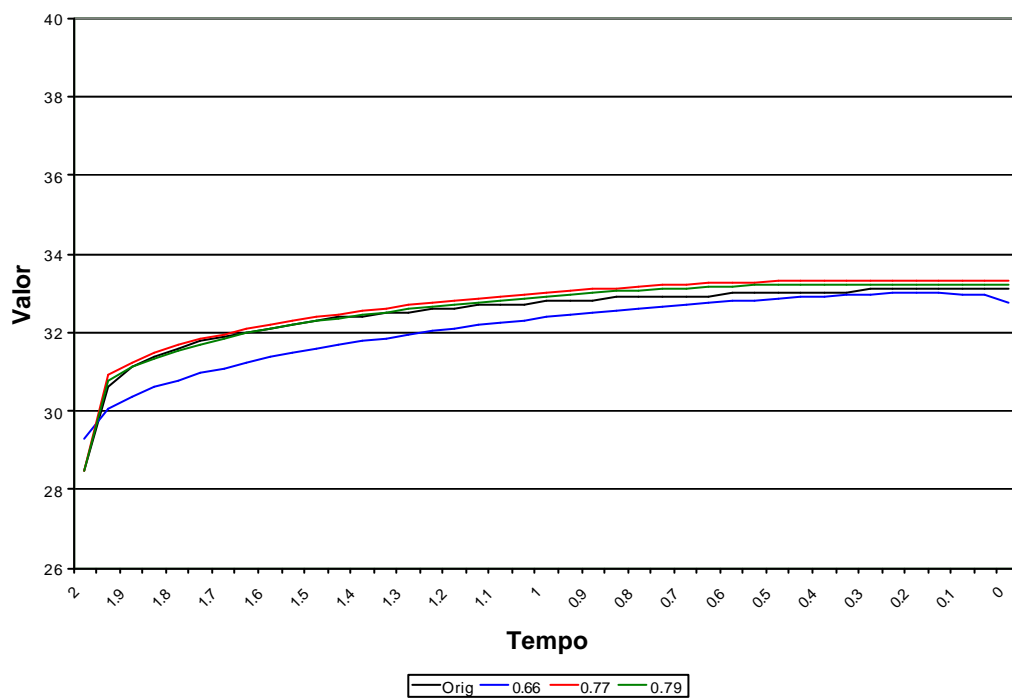


Figura 14 - GP X Curva Original, juros com 6% e volatilidade de 0.25.

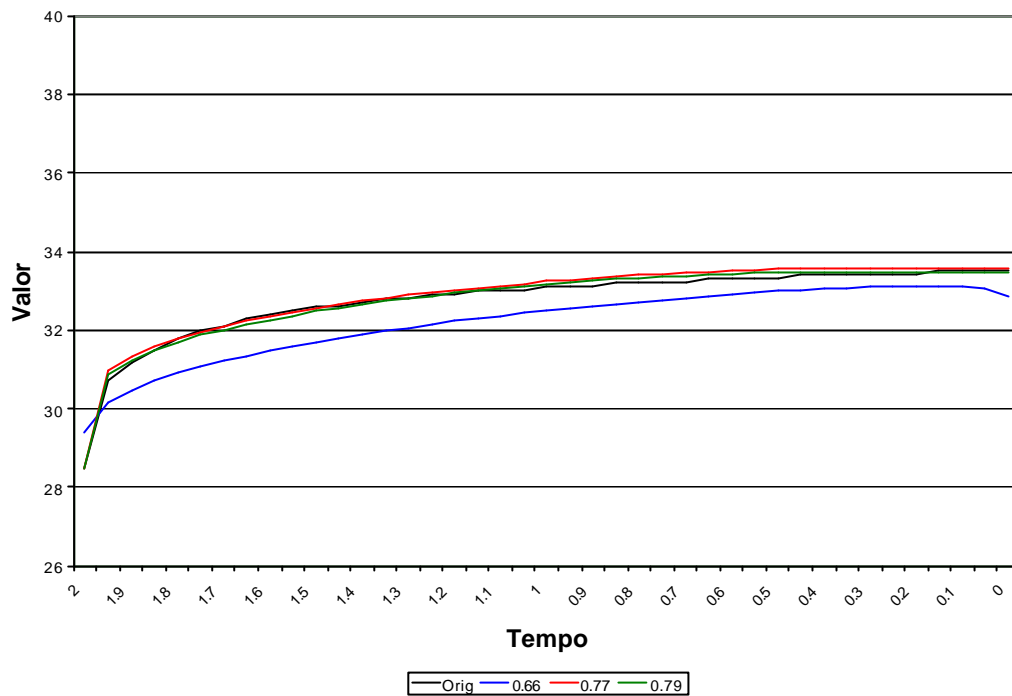


Figura 15 - GP X Curva Original, juros com 8% e volatilidade de 0.25.

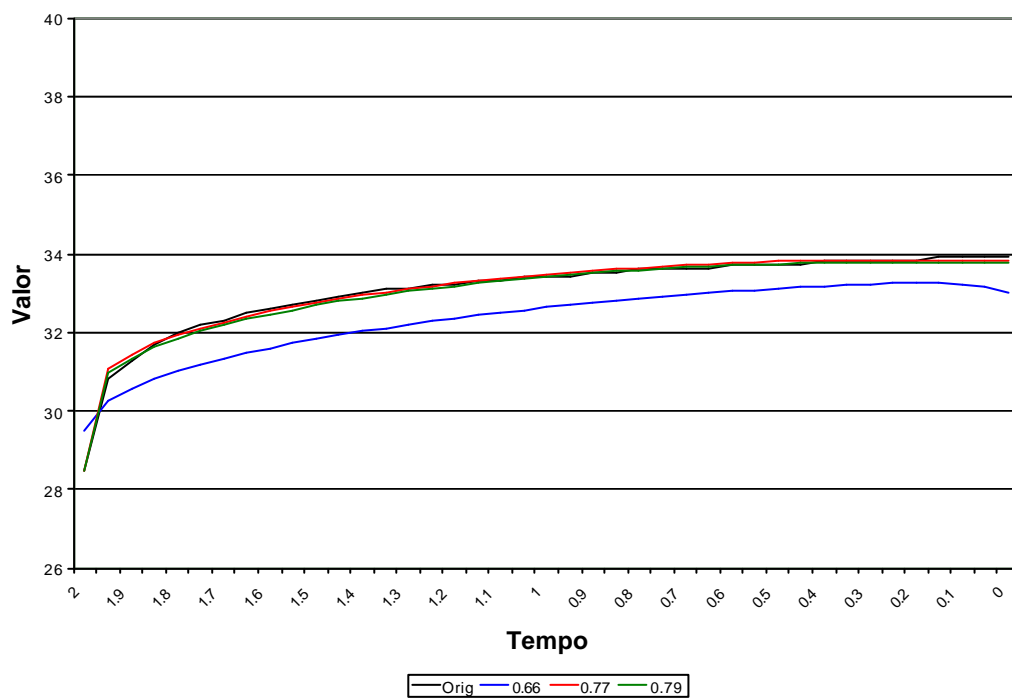


Figura 16 - GP X Curva Original, juros com 10% e volatilidade de 0.25.

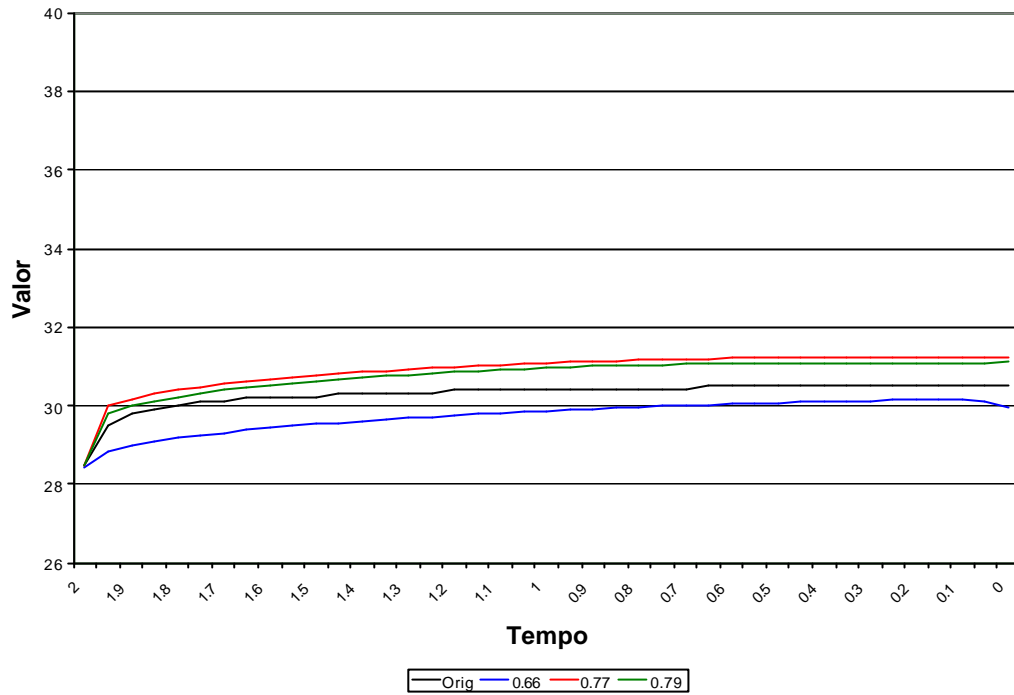


Figura 17 - GP X Curva Original, juros com 8% e volatilidade de 0.15.

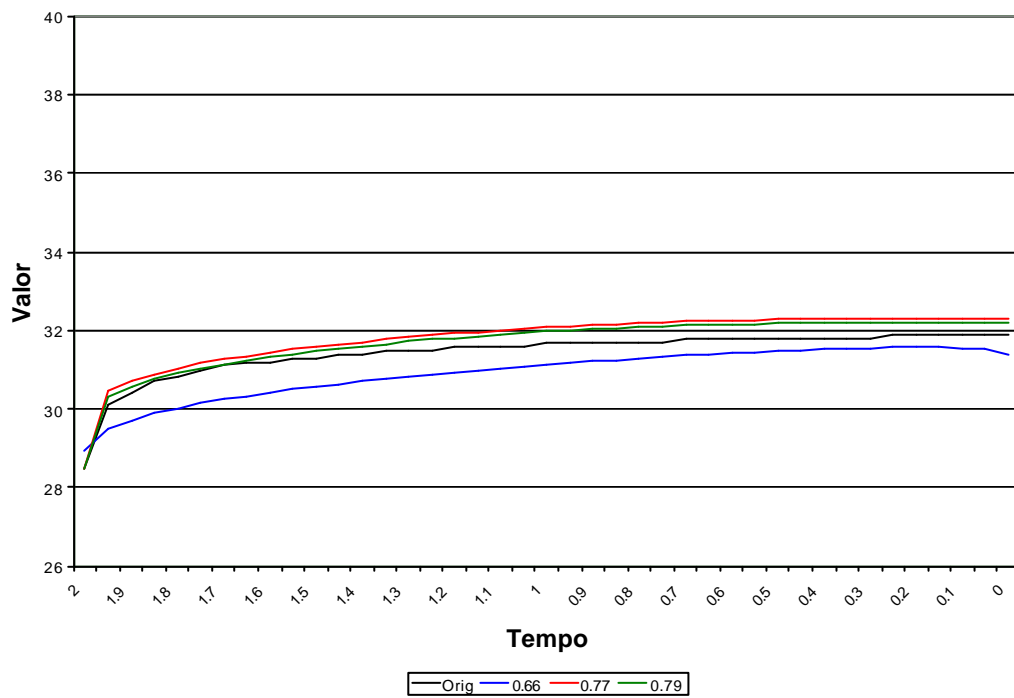


Figura 18 - GP X Curva Original, juros com 8% e volatilidade de 0.20.

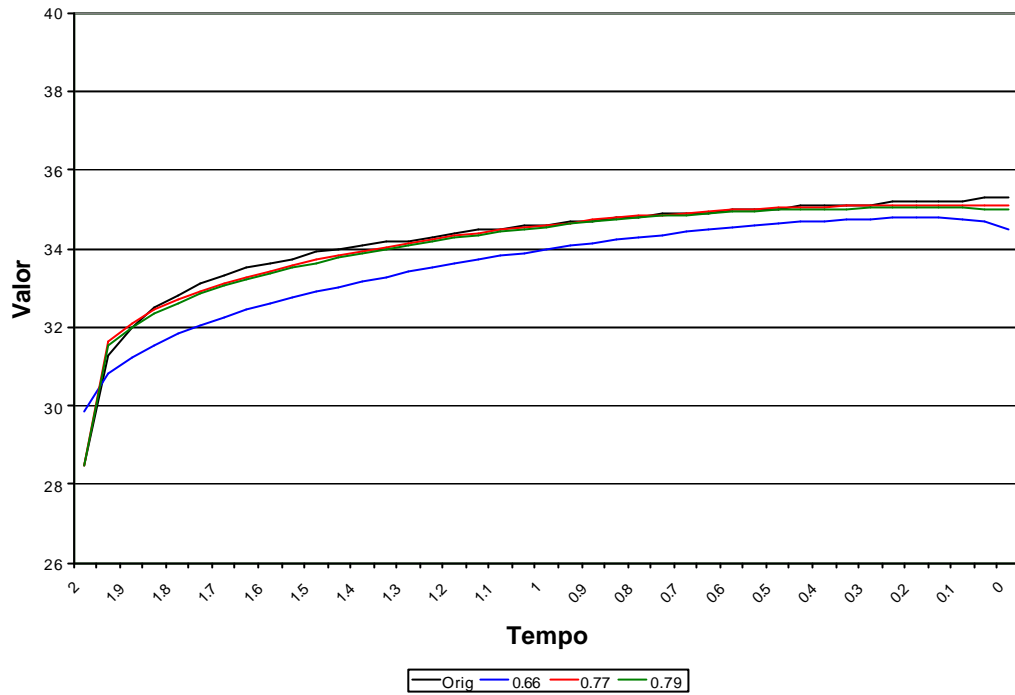


Figura 19 - GP X Curva Original, juros com 8% e volatilidade de 0.30.

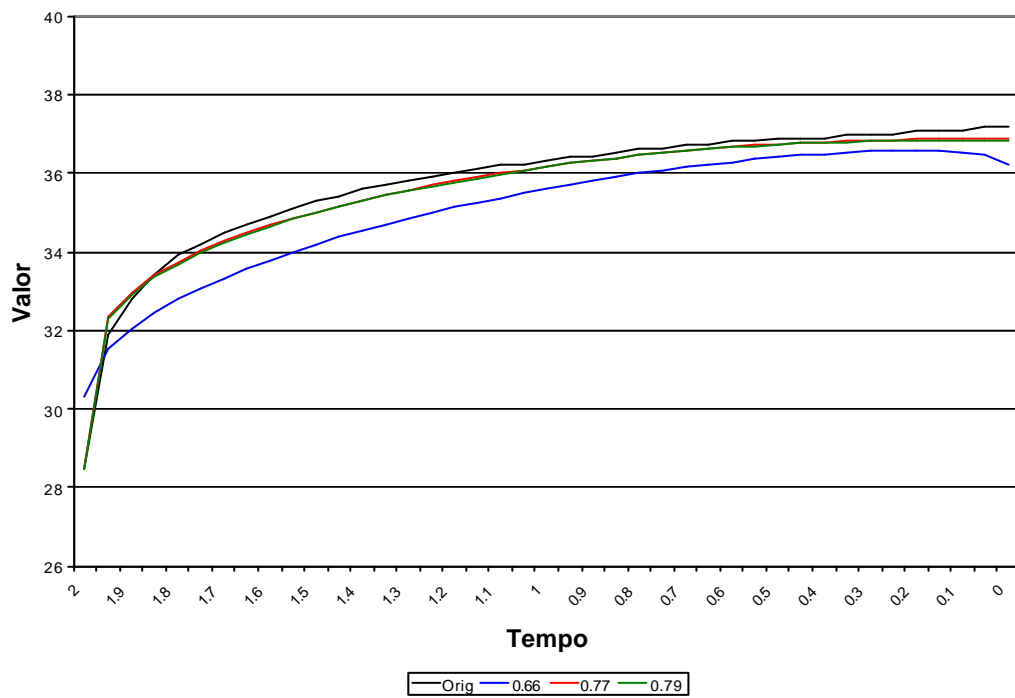


Figura 20 - GP X Curva Original, juros com 8% e volatilidade de 0.35.

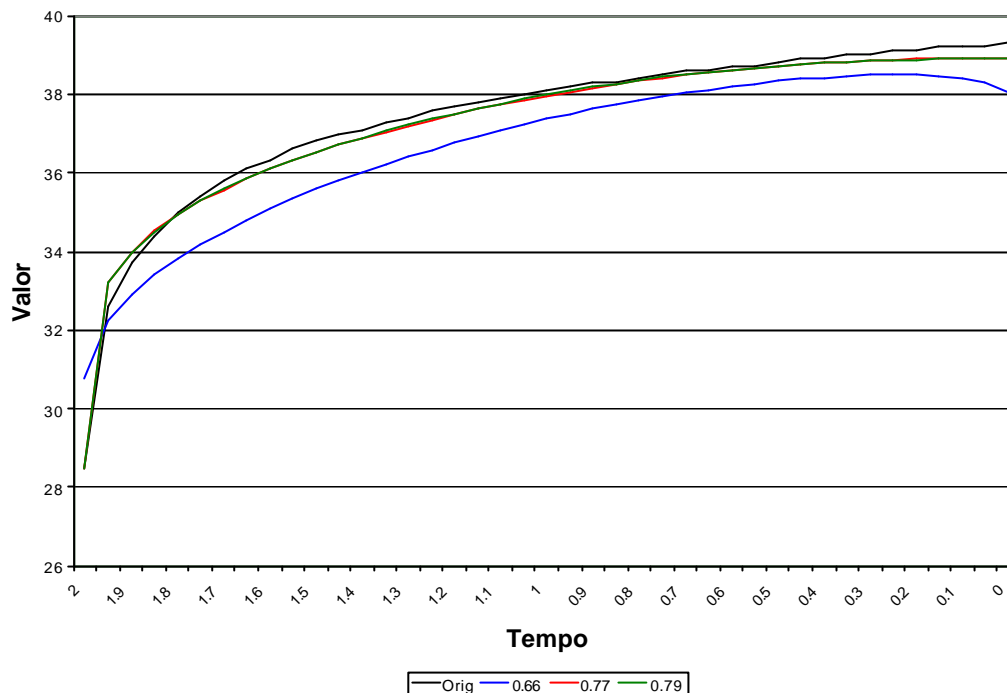


Figura 21 - GP X Curva Original, juros com 8% e volatilidade de 0.40.

Tendo em vista os resultados obtidos pelas métricas estabelecidas (MAPE e RMSE), é considerada apenas a curva com aptidão normalizada de 0.79, por ter os MAPEs no conjunto de validação mais baixos, como retorno para o bloco de Avaliação da Opção por SMC.

5.4. Cálculo do Valor da Opção por Simulação de Monte Carlo (SMC)

Neste item, é detalhado o cálculo do valor da opção por SMC.

No item anterior foi descrito o experimento que obteve a função analítica da curva de exercício ótimo. Na SMC, esta função foi parametrizada através de duas variáveis: volatilidade (σ) e taxa de juros livre de risco (r). A parametrização da função é realizada desta forma de modo a ficar de acordo com a geração das amostras conforme descrito no item 5.2.

O caso estudado, por se tratar da análise de alternativas para investimento em desenvolvimento de campo de petróleo, utiliza para o cálculo do valor da opção, características intrínsecas à *commodity* do petróleo. Uma dessas características são as incertezas q (qualidade da reserva) e B (quantidade em milhões de barris da reserva) já mencionadas anteriormente. Estas duas incertezas

são utilizadas no cálculo do valor da reserva de petróleo de acordo com [5][12]. O valor da reserva é análogo ao preço da ação no estudo de opções financeiras e é calculado como:

$$V = qPB \quad (35)$$

Onde P é o preço de um barril da reserva de petróleo, q é a qualidade da reserva e B , o tamanho da reserva. O valor da opção é medido por:

$$VO = V - D \quad (36)$$

Onde D é o investimento inicial e funciona análogo ao preço do exercício em Opções Financeiras.

A variável P obedece o PRM de Bhattacharya. Para a SMC, utiliza-se este processo com intervalos discretos descontando-se o prêmio de risco da opção. A simulação considera o valor da opção para situações em que o investidor não quer sofrer riscos [7][24]:

$$\Delta P = h \left(\left[\bar{P} - \frac{r-r}{h} \right] - P \right) \Delta t + sP e \sqrt{\Delta t} \quad (37)$$

Onde $e \sim N(0,1)$.

A SMC utilizou a equação (37) iterativamente do seguinte modo:

$$P_{t+1} = P_t \left(1 - (h + r - r) \Delta t + s e \sqrt{\Delta t} \right) + h \bar{P} \Delta t \quad (38)$$

Para simular os caminhos do Preço do Petróleo ao longo do tempo, utiliza-se a variável e na equação (38) a cada intervalo de tempo Δt . A geração de números aleatórios e é feita de acordo com a equação (30).

A cada intervalo discreto Δt , em que o preço do petróleo P se desloca, é gerado um novo e até o tempo de expiração do projeto T ou até que a trajetória do

ativo ultrapasse a curva de exercício ótimo compondo um ciclo. O intervalo Δt para cada ciclo utilizado é 0.01, que equivale a 7.3 dias dado um tempo de expiração T de 2 anos. Com 10.000 ciclos ou cenários, para cada parametrização, são rodadas 10 séries, isto é 10 SMCs.

Após as 10 séries rodadas, é calculada uma média aritmética dos valores da opção retornados para cada série, definindo-se o valor da opção para cada parametrização. Este resultado é comparado ao valor da opção obtido pela mesma parametrização na geração das amostras do conjunto de validação. Uma ilustração deste experimento é exibida na Tabela 6 para o cálculo do valor da opção com curva de exercício ótimo parametrizada por taxa de juros livre de risco de 0.04 e volatilidade de 0.25 e demais parâmetros fixos:

Tempo	Valor SMC para 10.000 cenários
0:10:19	13.03
0:10:17	12.99
0:10:16	13.23
0:10:18	12.96
0:10:19	13.04
0:10:18	13.2
0:10:16	13.07
0:10:18	13.05
0:10:18	13.41
0:10:17	12.95
Valor Médio das SMCs	Variância
13.09	0.02

Tabela 6 - Simulação de Monte Carlo para Juros 0.04 e Volatilidade 0.25. Tempo aproximado de 10 minutos.

A Tabela 6 mostra o cálculo do valor da opção para uma parametrização escolhida para validar a função analítica da curva de exercício ótimo retornada pelo bloco de PG. O tempo aproximado para cada SMC foi de 10 minutos. Após realizar este cálculo para as demais parametrizações escolhidas, produz-se o resultado da Tabela 7:

Juros	Volatilidade	Valor SMC médio (US\$MM) em 10 rodadas	Valor da amostra (US\$MM)	Erro Relativo %
0.04	0.25	13.09	13.38	2.17
0.06	0.25	15.79	16.17	2.35
0.08	0.25	19.34	19.43	0.46
0.1	0.25	23.27	23.22	0.22
0.08	0.15	0.68	0.78	12.82
0.08	0.20	6.26	6.27	0.16
0.08	0.30	39.27	39.90	1.58
0.08	0.35	63.13	66.16	4.58
0.08	0.40	95.07	96.71	1.70

Tabela 7 - Comparação entre o Valor da Opção das amostras e o Valor da Opção calculado para os mesmos parâmetros através de Simulação de Monte-Carlo.

A Tabela 7 exhibe os resultados comparando o cálculo do valor da opção para cada parametrização com o valor da amostra gerada pelo software “Análise de alternativas de desenvolvimento de um campo sob condições de incerteza de mercado”. Na comparação também é exibida a variância do experimento com 10 SMCs para cada parametrização. Os resultados apresentados são satisfatórios considerando que o intervalo Δt de 0.01 utilizado nas SMCs é bem maior do que o intervalo utilizado na geração das amostras (10^{-6}). O passo de 0.01 foi utilizado para acelerar o tempo da simulação que é bem mais lenta do que a técnica de obtenção das amostras.