

Inference by Genetic Programming of an analytical expression for the Optimal Exercise Threshold of an asset that follows a Mean Reversion Process

Dan Posternak¹, Marco Aurélio C. Pacheco¹, Juan G. Lazo Lazo¹

¹ICA : Applied Computational Intelligence Laboratory

Department of Electrical Engineering

Pontifical Catholic University of Rio de Janeiro, PUC-Rio

Rua Marquês de S. Vicente 225, Gávea, Rio de Janeiro, CEP 22453-900, RJ

danposternak@terra.com.br

juan@ele.puc-rio.br

marco@ele.puc-rio.br

Marco A. G. Dias²

²Petrobras, E&P-Corp./ENGEP/DPR

Av. Chile 65, sala 1702, CEP 20035-900, RJ

marcodias@petrobras.com.br

Inference by Genetic Programming of an analytical expression for the Optimal Exercise Threshold of an asset that follows a Mean Reversion Process

Abstract

The American option evaluation is a relatively complex and expensive process due to commonly used methodologies as Finite Differences, Dynamic Programming, Monte Carlo Simulation, etc. needs high computational performance. Besides that, the complexity needed to calculate the option value and the optimal threshold increases when the price of underlying asset follows the Mean Reversion Stochastic Process. By this way, is interesting to achieve an analytical approximation in order to make easier to obtain the optimal threshold and the option value.

There are many analytical approximations mentioned in bibliography respecting to American Options about asset prices following a Geometric Brownian Motion [1] [2], but none about it follows Mean Reversion Processes.

This work proposes a model based on Symbolic Regression by Genetic Programming to obtain an analytical approximation for the optimal threshold respecting to an American option which its asset follows a Mean Reversion Process.

The Optimal Threshold that separates during the option life-cycle the decision to exercise the option, is later employed to evaluate the option. The result achieved by the proposed model (Threshold Analytical Function) seemed to be satisfactory.

Genetic Programming

The Genetic Programming [3] [4] is inspired by Darwin ideas related to the survival of the most capable individual, to evolve computer programs and mathematical functions to solve problems. Such programs are generally structured in tree diagrams. The idea of the Symbolic Regression is to evolve Mathematical Functions which in the end return us a solution or an approximation for a peculiar problem. To obtain a mathematical solution, the Symbolic Regression uses a set of functions, mathematical operations and terminals. Terminals are the leaves of the

tree while the functions and mathematical operations are always allocated in the tree vertices as shown in Figure 1

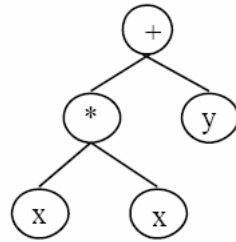


Figure 1 - Tree Structure representing the function $f(x,y)=x^2+y$

This set of functions, mathematical operations and terminals are combined and evolved through mutation and crossover operators (Figure 2 e Figure 3), looking for solving the problem with an equation that is a function of all relevant parameters.

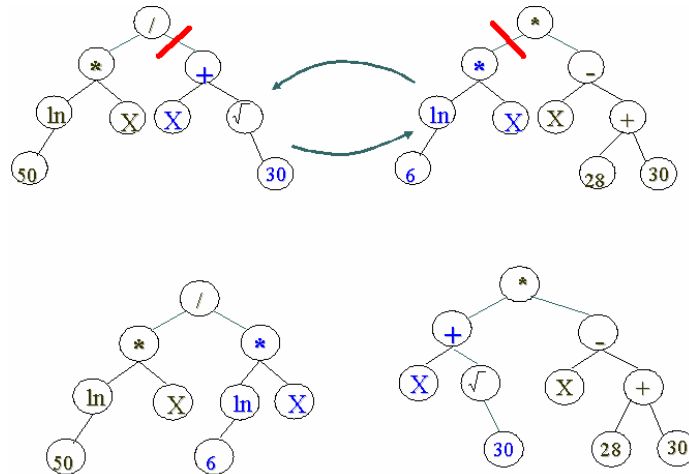


Figure 2 - Crossover between two individuals above generating the descendants right below.

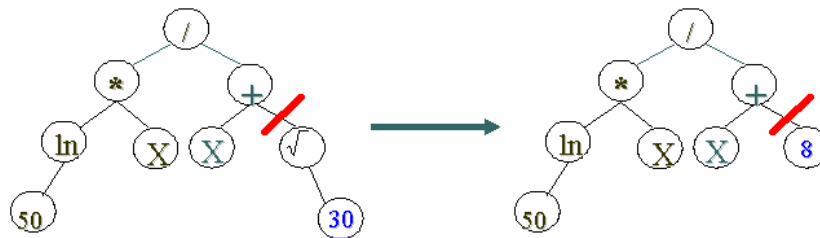


Figure 3 - Mutation. The $\sqrt{30}$ branch was substituted by 8

Model Description

The Symbolic Regression Model uses Genetic Programming and optimal exercise s (optimal thresholds) samples obtained through numerical methods. The system is partitioned into three operational blocks. Sample Generation, Genetic Programming and Option Value calculated by Monte Carlo, Figure 4.

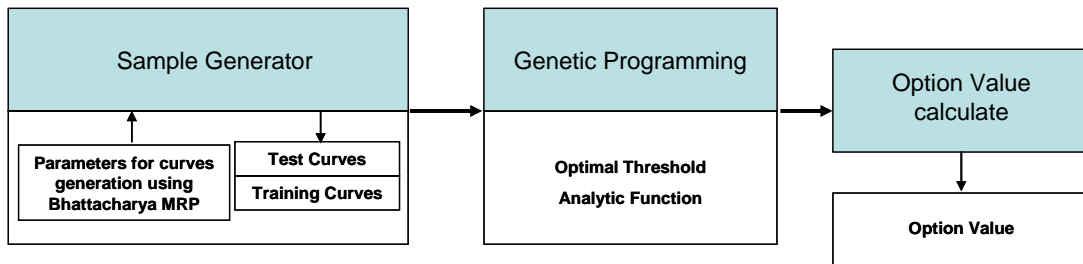


Figure 4 - Block diagram illustrating the proposed model.

Sample Generation: The Generation Sample block refers to optimal thresholds sample generation through numerical methods. To generate those samples, is necessary the input of parameters related to the asset behavior during its life-cycle represented by the Mean Reversion Stochastic Process proposed by Bhattacharya [5] to be used by the numerical method.

The samples were created by proprietary software owned by PETROBRAS to execute analysis of alternatives for oil field development under market uncertainty [6], this software analyzes the investment option in an oil reservoir and calculates the option value using arbitrage theory approach using the solution of partial differential equations through the explicit method. This option to invest is similar to an American call option. The software has generated the sceneries considering the variation of the free risk interest rate and the Petroleum Price Volatility.

The parameters used to generate the option optimal exercise curves are:

- D is the investment cost of the underlying alternative to develop an oil field (US\$/bbl).
- q is technical quality of the respective alternative.
- B is the estimated average size of the reservoir in barrels.
- P_{max} is the maximum price reached by the petroleum during the option life-cycle (US\$ per year).

- P_0 is the petroleum initial price (US\$/bbl).
- \bar{P} is the estimated price reached in long term horizon (US\$/bbl).
- δ is the convenience yield (% per year).
- σ is the petroleum price volatility (% per year).
- r is the free risk interest rate (% per year).
- ρ is the exogenous discount rate adjusted to business' risk in US\$ (% per year).
- T is option life-cycle duration (years).
- η is the Mean Reversion Speed used in Bhattacharya process.

Each parameterization returns an optimal exercise threshold and its option value.

After the samples generation, the thresholds are separated in optimal exercise thresholds for training (50 thresholds) and optimal exercise thresholds for testing (9 thresholds). The testing thresholds have the purpose to evaluate the answer given by the Genetic Programming block.

The training set containing 50 thresholds is responsible for evaluating the mathematical expressions (individuals) during the Genetic Programming evolution.

The test set containing 9 thresholds is used to certify that the analytical expression found by GP works fine also for thresholds out of the training set.

Genetic Programming: The Genetic Programming block is responsible for the Symbolic Regression, i.e., to find an analytical equation to the optimal exercise threshold. This block receives as input the optimal exercise thresholds for training.

The genetic programming uses in its evolutionary process individuals that represents just mathematical functions. This makes easier the adjustments to the functions and terminals sets needed to create individuals randomly.

To implement the Genetic Programming software to evolve functions in order to return the optimal exercise threshold analytical function, there was used the OpenBEAGLE GP framework [7].

To evolve an optimal exercise threshold analytical function, it is necessary set the GP control parameters. Besides, there is interest into add terminals related to the analytical function to cooperate with the search for an optimal solution. In Table 1,

there are parameters used as terminals in the trees from GP with its respective ranges of values and the functions or vertices are described in Table 2.

D (US\$/bbl)	2500
q	0.22
B (MM bbl)	400
P_{max}	60
P_0	20
\bar{P}	20
δ	0.08
ρ	0.12
T	2
η	0.3466
σ (variable)	[0.05; ...;0.5]
r (variable)	[0.01; ...;0.17]

Table 1 – Range of values assigned to terminals

$N()$	Normal Cumulative Density Function
$\sqrt{\quad}$	Squared Root
Pow	Power
exp	Exponential
ln	Natural Logarithm
$+, -, \times, /$	Addition, Subtraction, Product and Division

Table 2 – GP function set.

Besides the terminals mentioned before, there were added encapsulated terminals for helping the individual's evolution. Those terminals are made by other terminals already inserted in search space and the reason for being inserted is due to having relationship to similar problems already solved and could be useful in search for the optimal individual. Those terminals are:

- $JMS = r - [\sigma^2 / 2(T - t)]$

- $JPS = r + [\sigma^2 / 2(T - t)]$
- $DDNOM = \sigma\sqrt{T - t}$
- $d_1 = \frac{\ln(qPB / D) + (r + (\sigma^2 / 2) \times T)}{\sigma\sqrt{T}}$
- $d_2 = d_1 - \sigma\sqrt{T}$

The Genetic Programming process can be described as:

1. An initial population is generated with individuals created randomly by the combination of between functions and terminals related to the problem (analytical function).
2. The following steps are processed until the termination criteria be satisfied:
 - Each Mathematical equation in the population is solved and a fitness is assigned to it representing how precise was this equation to solve the problem.
 - A new population of mathematical equations is created applying genetic operators. The genetic operations (mutation and crossover) are applied to the mathematical functions selected probabilistically by its fitness.
3. The best mathematical equation found so far is returned as result of the GP. This result may be the solution of the problem.

The termination criteria used can be the maximum of generations reached by the population.

Option Value Calculated: The Option Value is calculated by Monte Carlo simulation using the optimal exercise threshold analytical function found by Symbolic Regression in the Genetic Programming block.

To calculate the option value, the simulation generates sceneries for the asset price using the Bhattacharya Mean Reversion Stochastic Process in discrete way. The moment each scenery cross the threshold is considered to calculate the difference between the price and the strike price (payoff). Once it is realized, the value is brought to present value and accumulated. In the end, the sum resulted by the accumulation of payoffs is divided by the number of sceneries generated. If the scenery doesn't reach the threshold, the option value for that scenery is zero.

After 10000 sceneries generated the average option value is calculated considering every option values obtained during the simulation.

Results

The model proposed based on Symbolic Regression by Genetic Programming was executed considering a population of 4000 individuals and 200 generations. By each generation, the 300 better individuals were preserved and copied to the following generation. The result achieved at the last generation was:

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

Where P^* represents the critical price or threshold for each instant t , \bar{P} is the long term price, σ is the price volatility, r is the risk free interest rate, T is the expiration time, η is the mean reversion speed, D is the investment cost, q is the reservoir quality, B is the size of the reservoir in barrels, $JPS = r + [\sigma^2 / 2(T - t)]$, $DDNOM = \sigma\sqrt{T-t}$ and $d_1 = \frac{\ln(qPB/D) + (r + (\sigma^2/2) \times T)}{\sigma\sqrt{T}}$.

This analytical expression obtained by GP was tested using a new set for evaluation, the test set, containing nine optimal exercise thresholds. Those thresholds were generated varying the values from the volatility of the price and the free risk interest rate parameters.

In Table 3, the volatility and free risk interest rate parameters used to generate the test and training sets are presented, and the errors between the test thresholds and the thresholds returned by the analytical function discovered by GP. In this table,

it is observed that the analytical function achieves good approximations for the parameters involved.

In the appendix, the figures show the approximation between thresholds returned by the analytical function and the respective sample from test set. It is indicated also the parameters (volatility and risk free interest rate) and the MAPE as well.

<i>r</i>	<i>σ</i>	<i>RMSE</i>	<i>MAPE %</i>
0.04	0.25	0.18	0.48
0.06	0.25	0.14	0.36
0.08	0.25	0.10	0.26
0.1	0.25	0.09	0.23
0.08	0.15	0.50	1.57
0.08	0.20	0.27	0.78
0.08	0.30	0.16	0.40
0.08	0.35	0.22	0.57
0.08	0.40	0.21	0.48

Table 3 - Results according MAPE and RMSE to values variation of the risk free interest rate and volatility in a test set of samples.

In Table 4 the option value calculated by Monte Carlo simulation using thresholds returned by analytical function is compared to the option value obtained by solving partial differential equations through the sample generator software. In this table is also observed that the error between the option values is small. This error can be decreased, increasing the number of sceneries for the price used in the simulation or decreasing the time steps, in this case, 0.01. The step used to generate the thresholds for the test set was 10^{-6} .

<i>Option Value with the GP Threshold</i>	<i>Option Value of the Sample</i>	<i>Option Value Relative Error %</i>
13.09	13.38	2.17
15.79	16.17	2.35
19.34	19.43	0.46
23.27	23.22	0.22
0.68	0.78	12.82
6.26	6.27	0.16
39.27	39.90	1.58
63.13	66.16	4.58
95.07	96.71	1.70

Table 4 – Option value comparison.

Conclusion

This paper has introduced a model based on Symbolic Regression by Genetic Programming to obtain an analytical approximation for an American option optimal exercise threshold which its underlying asset price follows the Mean Reversion Process, considering two varying parameters (volatility and risk free interest rate). The model was tested in a real options problem about investing analysis in petroleum reservoir under market uncertainty conditions. This real option analysis is similar to an American call option.

The proposed model returned an analytical equation that has achieved good approximations for the option optimal exercise threshold.

The Symbolic Regression by Genetic Programming seemed to be a technique able to infer equations based on a sample set producing satisfactory results.

The significance of this work, although its initial simplicity has considered just two varying parameters, lies on the necessity of an analytical approximation that makes easier to calculate an American option optimal exercise threshold which its asset follows a Mean Reversion Process.

Acknowledgment

This project has been sponsored by Petrobras, Petróleo Brasileiro S.A., through PRAVAP-14 program, and coordinated by Cenpes Research Center.

References

- [1] Barone-Adesi, G.; Whalley, Robert E.: Efficient Analytic Approximation of American Option Value. *The Journal of Finance*, Vol. 42, No. 2 (Jun., 1987), 301-320.
- [2] MacMillan, L. W.: Analytic Approximation for the American Put Option. *Advances in Futures and Options Research*, 1 (1986), 119-39.
- [3] Koza, J.: *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Bradford Books (Dec., 1992), 840 pp.
- [4] Koza, J.: *Genetic Programming II: Automatic Discovery of Reusable Programs*. Bradford Books (May, 1994), 746 pp.
- [5] Bhattacharya, S.: Project Valuation with Mean Reverting Cash Flow Streams. *The Journal of Finance*, Vol. 33, No. 5 (Dec., 1978), 1317-1331.

- [6] Dias, M. A. G.: Selection of Alternatives of Investment in Information for Oilfield Development Using Evolutionary Real Options Approach with Monte Carlo Simulation (Jan., 2001), 29 pp, presented at the 5th Annual International Conference on Real Options, UCLA, Los Angeles, July 2001.
- [7] Gagné C., Parizeau, M., Open BEAGLE Manual em http://easynews.dl.sourceforge.net/sourceforge/beagle/beagle-manual-2_1_3-r1.pdf 80pp. April, 2004. Último acesso em 28 de julho de 2004.

Appendix: Threshold Curves returned by the GP Analytical Function X Test Curves.

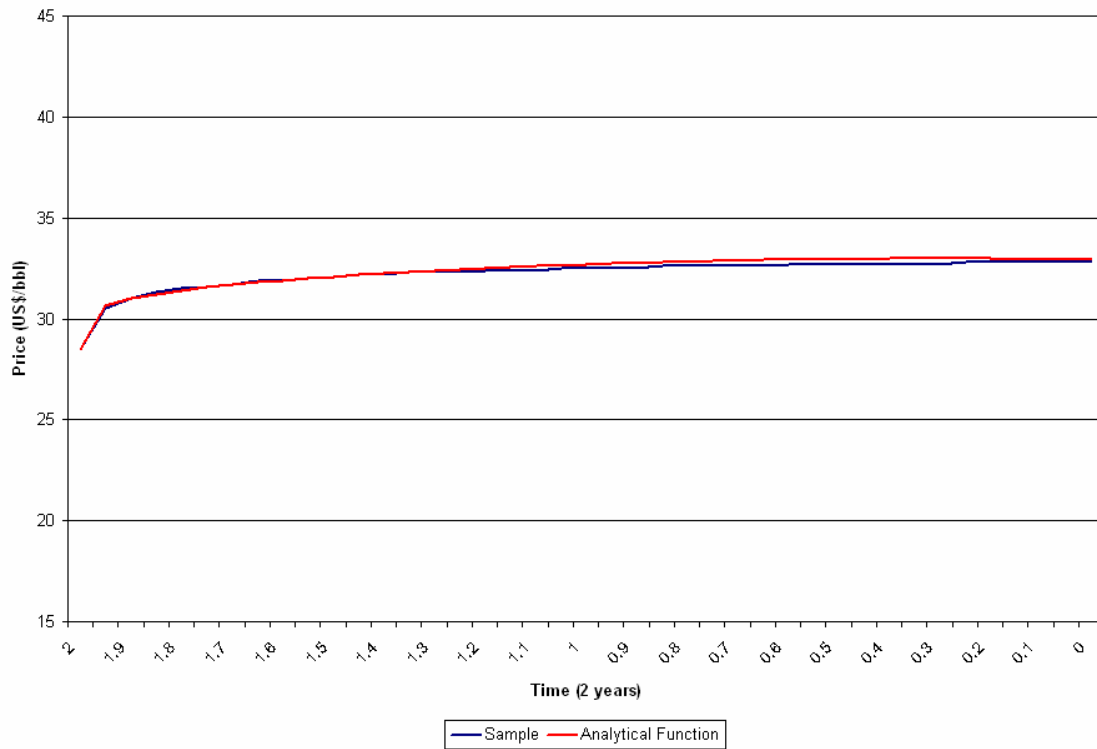


Figure 5 – $r=0.04$, $\sigma=0.25$, MAPE=0.48

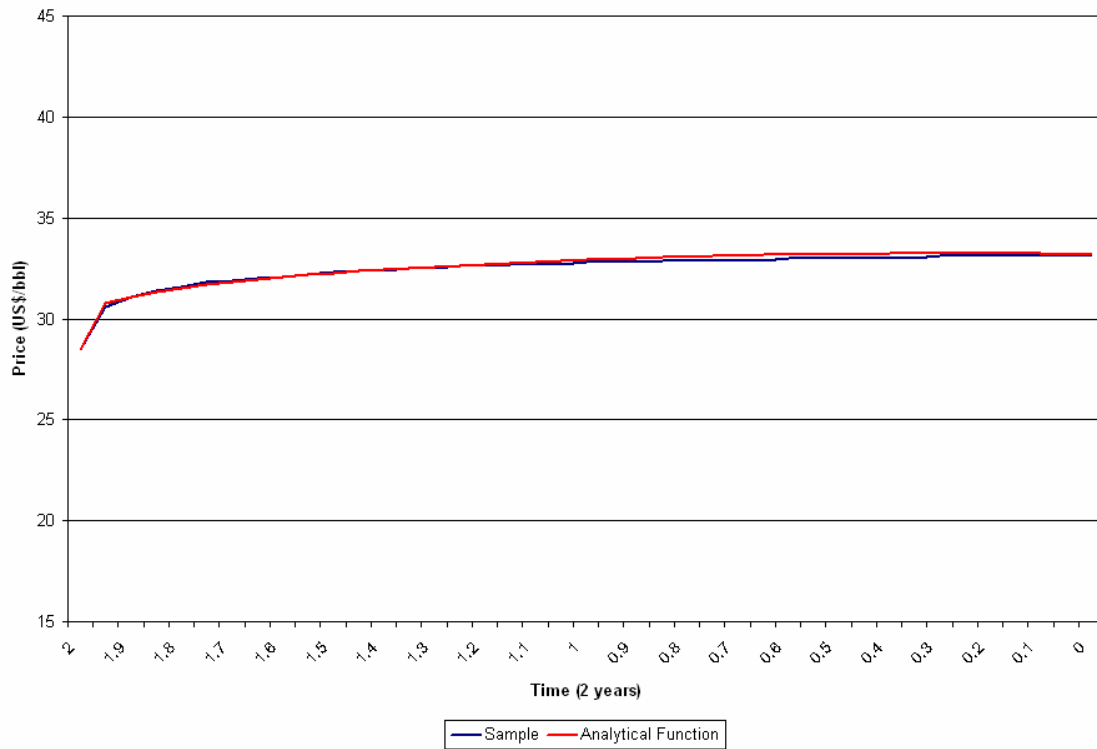


Figure 6 - $r=0.06$, $\sigma=0.25$, MAPE=0.36

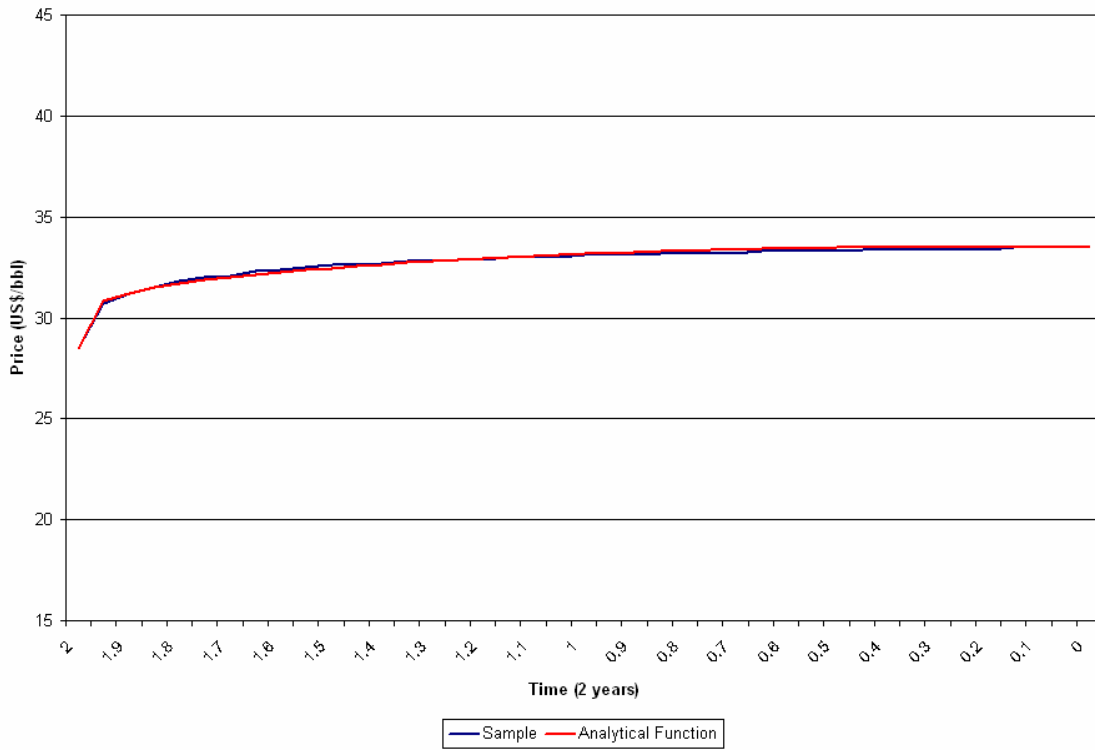


Figure 7 – $r=0.08$, $\sigma=0.25$, MAPE=0.26

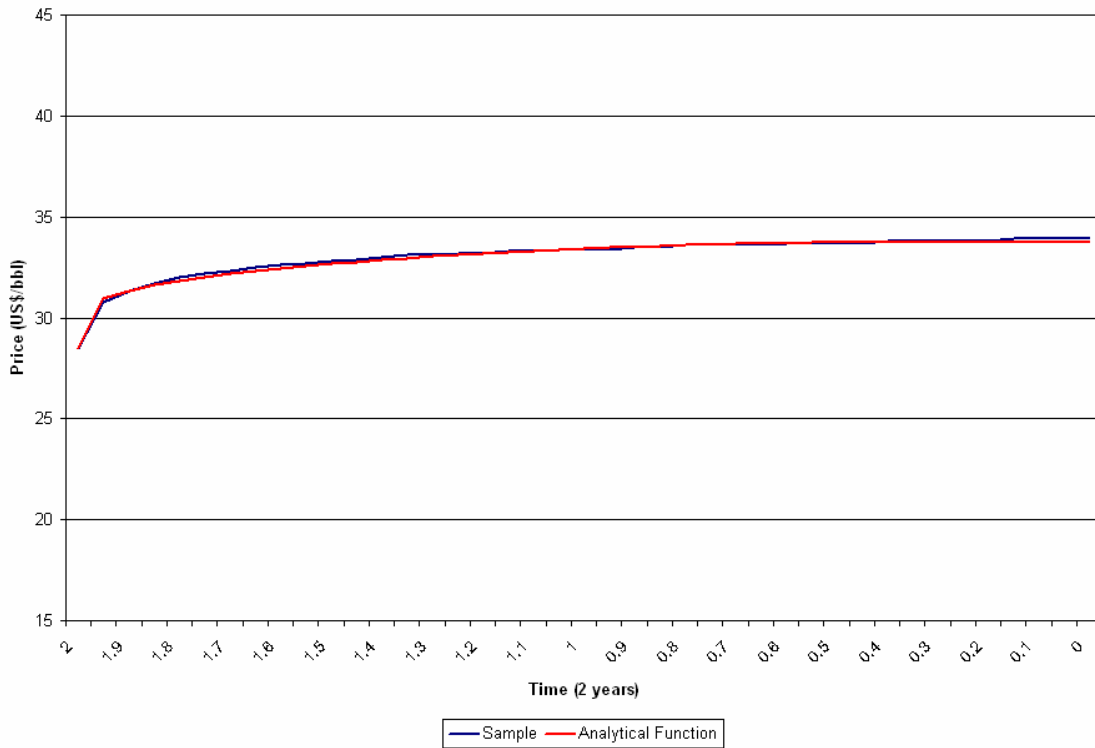


Figure 8 - $r=0.10$, $\sigma=0.25$, MAPE=0.23

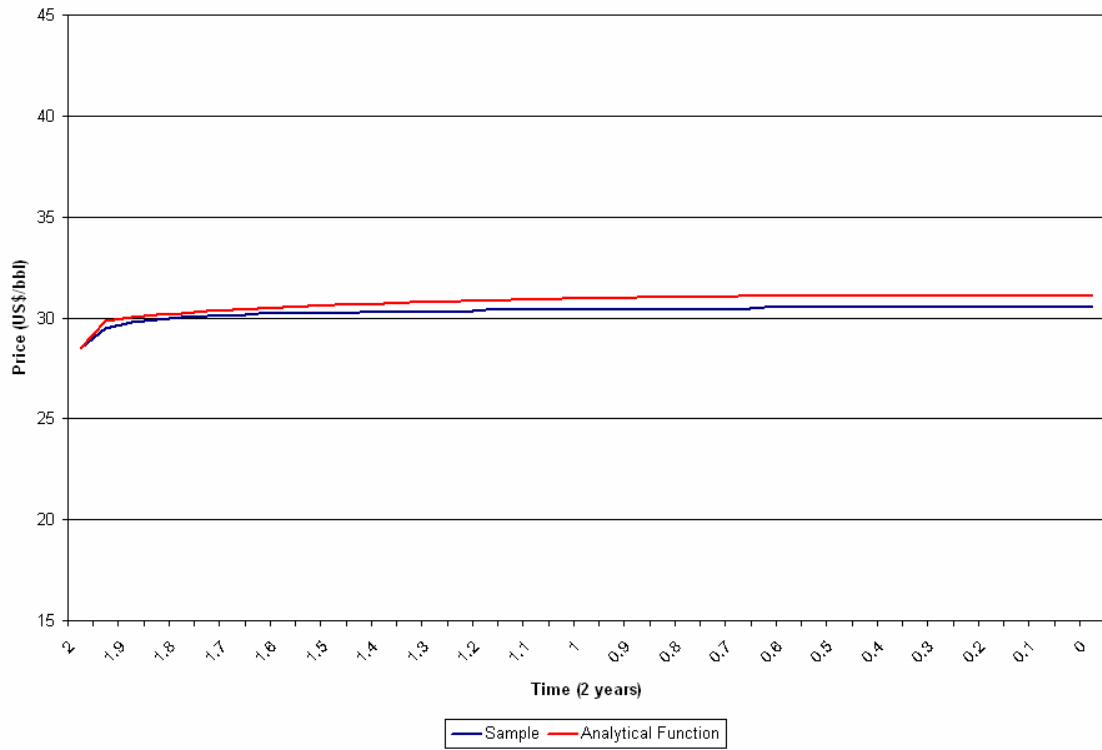


Figure 9 - $r=0.08$, $\sigma=0.15$, MAPE=1.57

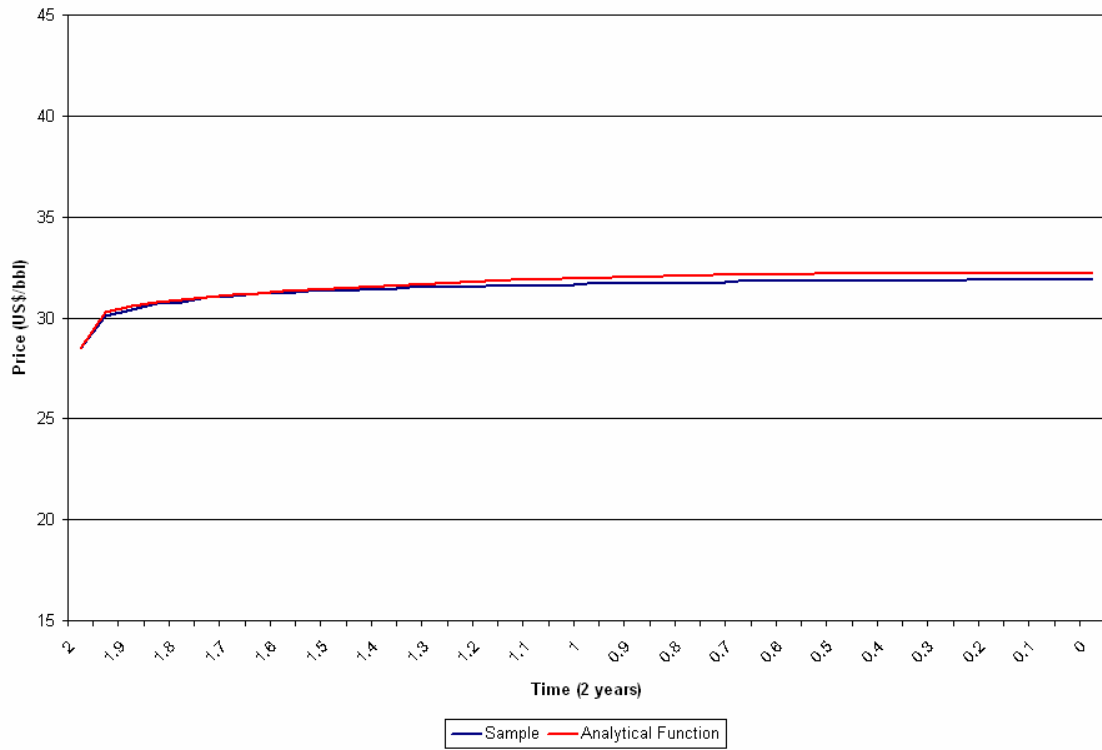


Figure 10 - $r=0.08$, $\sigma=0.20$, MAPE=0.78

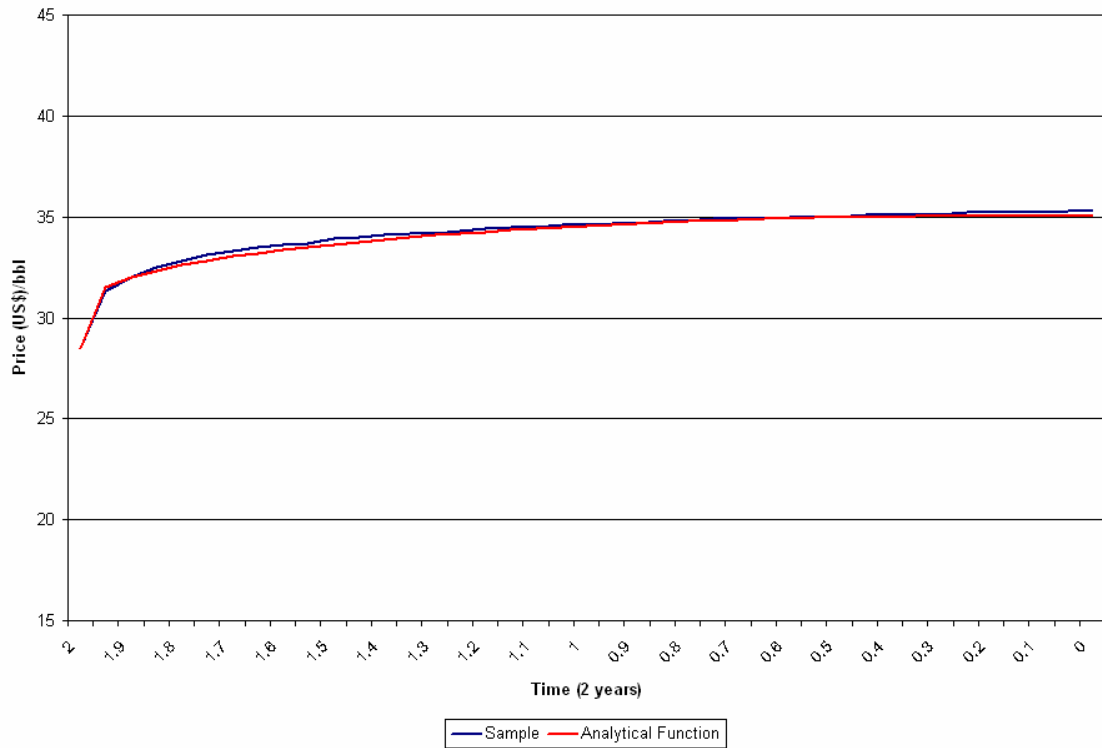


Figure 11 - $r=0.08$, $\sigma=0.30$, MAPE=0.40

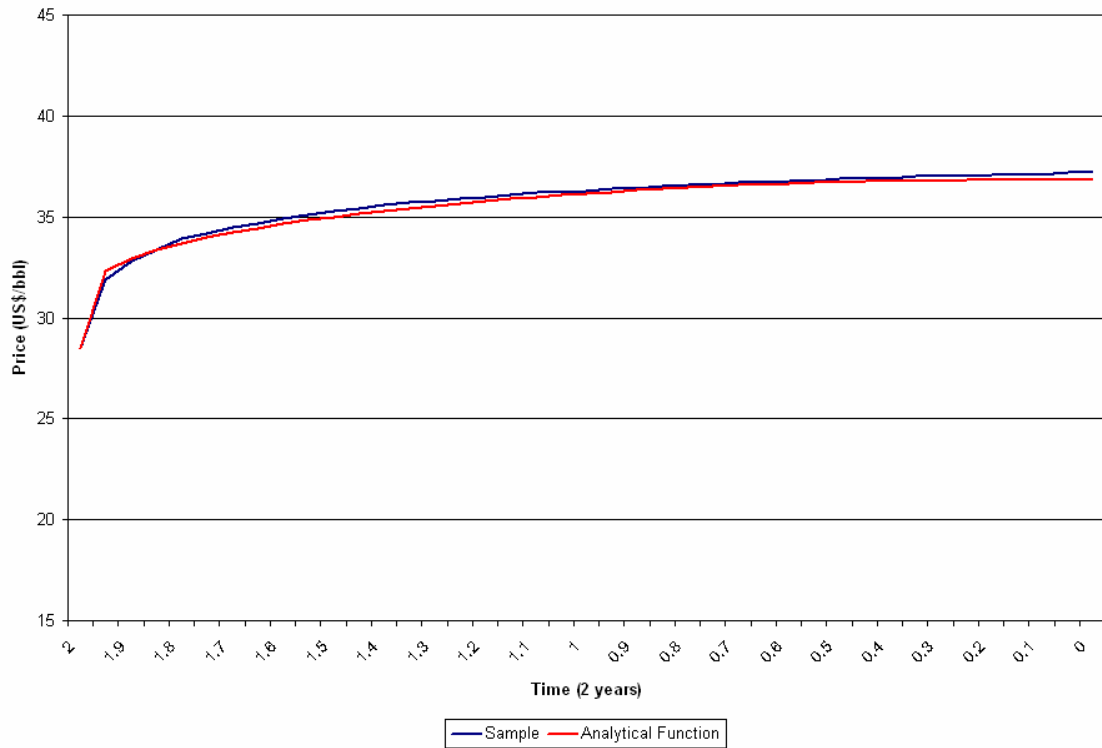


Figure 12 - $r=0.08$, $\sigma=0.35$, MAPE=0.57

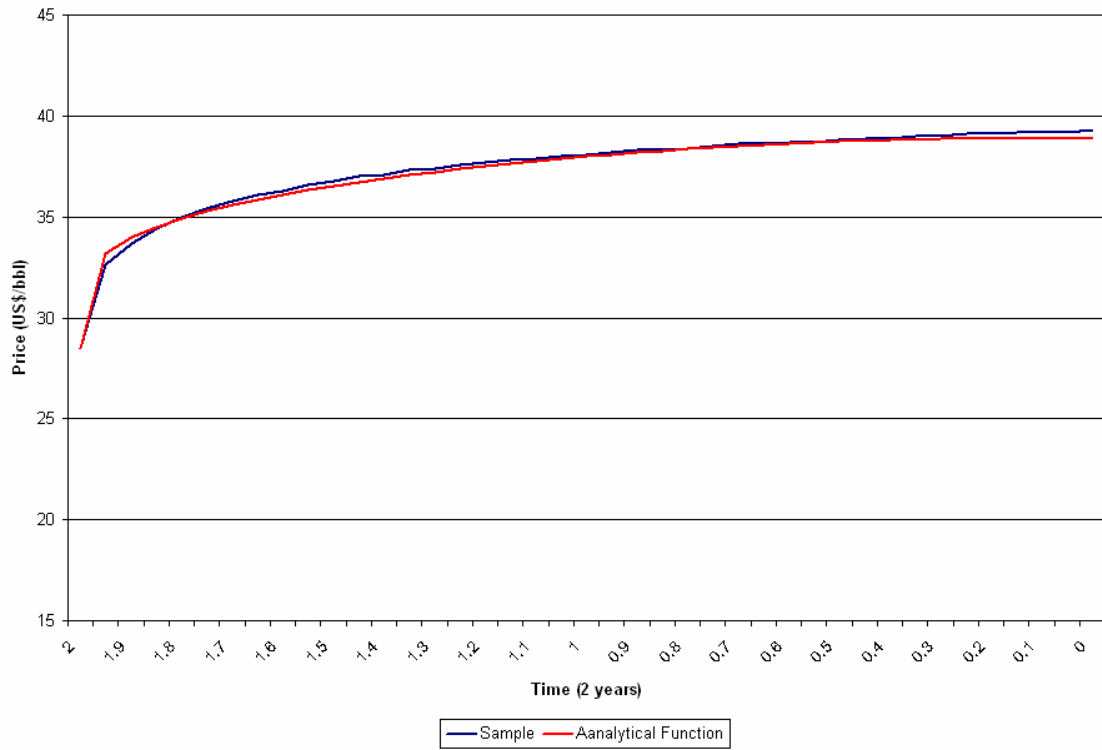


Figure 13 - $r=0.08$, $\sigma=0.40$, MAPE=0.48