Proceedings of the 2017 Industrial and Systems Engineering Conference
K. Coperich, E. Cudney, H. Nembhard, eds.

Estimation of Geometric Brownian Motion Parameters for Oil Price Analysis

Abstract ID: 3215

Jakob Croghan and John Jackman
Iowa State University

K. Jo Min
Iowa State University

Abstract

Geometric Brownian motion (GBM), a stochastic differential equation, can be used to model phenomena that are subject to fluctuation and exhibit long-term trends, such as stock prices and the market value of goods. The model uses two parameters, the rate of drift from previous values and volatility, to describe and predict how the continuous-time stochastic process evolves over time. Accurate estimates of the drift rate and volatility are necessary for these models to be useful within quantitative economic decision-making models. Multiple estimation methods have been proposed in previous research. We show how well these methods perform using a GBM with known parameters using different sample sizes. Using a GBM model, we estimated the parameters for historical oil prices and performed statistical analyses to determine how well the oil prices fit a GBM model.

Keywords
Geometric Brownian motion, data analytics, simulation, maximum likelihood

1. Introduction
Many observable phenomena exhibit stochastic, or non-deterministic, behavior over time. Geometric Brownian motion (GBM) is a stochastic differential equation that may be used to model phenomena that are subject to fluctuation and exhibit long-term trends. In particular, [3] has referred to it as "the model for stock prices". [7] indicated that the accuracy of a GBM model for oil prices is yet to be determined. In this paper, we examine estimation methods for GBM model parameters in the context of modelling fossil fuel prices. This is a first step in developing a quantitative decision making model based on an underlying GBM.

GBM differs from the generalized Brownian motion by removing the assumption of a constant drift rate. Instead, the expected rate of return μ is assumed constant [3]. Additionally, for GBM, the drift rate is equal to the current value of x multiplied by expected rate of return μ, or μx. The volatility rate of this function is expressed as the parameter σ. "The stochastic variable $x(t)$ follows a geometric Brownian motion if it satisfies the stochastic differential equation,

$$\mathrm{dx} = \mu\mathrm{x}\,\mathrm{dt} + \sigma\mathrm{x}\,\mathrm{dz} \tag{1}$$

where, dz is the increment of a Wiener process and μ and σ are the parameters to be estimated" [4].

The ratios of $x(t)$ to $x(t\text{-}1)$ have a lognormal distribution [6]; therefore, the $log\left(\frac{x_t}{x_{t-1}}\right)$ values are IID with a normal distribution. In the context of parameter estimation, [5] used Q-Q plots and the Shapiro-Wilk W Test to check the validity of the normality assumption for the log ratios. For the independence assumption, they used a Chi Square Test. However, the role of sample size in the estimation methods was not addressed.

2. Methodology
We evaluated three different methods of estimating the parameters of a GBM by using a dataset generated by a Monte Carlo simulation for a known set of parameters. Based on these results, we determined the appropriate number of data

points to estimate the parameters within 10% of the known values. We applied these estimation methods to real data sets for crude oil and natural gas prices and validated the assumptions for a GBM. The three methods are as follows.

2.1 Parameter Estimation Method 1

Method 1 is based on the difference between successive observations (i.e., $x(t)$ and $x(t\text{-}1)$). The ratios of the difference to $x(t\text{-}1)$ are IID according to a normal distribution [3]. The average of this ratio for the data series corresponds to the drift parameter μ and is given by

$$\hat{\mu} = \sum_{t=1}^{n} \frac{X_t - X_{t-1}}{X_{t-1}}. \tag{2}$$

The sample standard deviation of these ratios is the estimate of the volatility parameter and is given by

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n} \left(\frac{X_t - X_{t-1}}{X_{t-1}} - \hat{\mu} \right)^2} . \tag{3}$$

2.2 Parameter Estimation Method 2

Method 2, a maximum likelihood method, by [4] uses the ratio of $x(t)$ to $x(t\text{-}1)$. The estimate of the drift parameter is given by

$$\hat{\mu} = 1 - \sum_{t=1}^{n} \frac{X_t}{X_{t-1}} . \tag{4}$$

The estimate of the volatility parameter is given by

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(\frac{X_t}{X_{t-1}} + \hat{\mu} - 1 \right)^2} . \tag{5}$$

2.3 Parameter Estimation Method 3

Method 3, similar to Method 2, uses the ratio of $x(t)$ to $x(t\text{-}1)$, which has a lognormal distribution [3]. Therefore, the sample mean of the normal distribution is

$$\bar{X} = \frac{1}{n} \sum_{t=1}^{n} log\left(\frac{X_t}{X_{t-1}} \right) , \tag{6}$$

and the estimate of the drift parameter is

$$\hat{\mu} = \bar{X} + \frac{\hat{\sigma}^2}{2} , \tag{7}$$

where

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n} \left(log\left(\frac{X_t}{X_{t-1}} \right) - \bar{X} \right)^2} . \tag{8}$$

2.4 Evaluation of Parameter Estimation Methods

The three methods for estimating the parameters of a GBM were evaluated using a dataset generated from a Monte Carlo simulation of a GBM with known parameters of 0.05 and 0.15 for μ and σ, respectively. A 95% confidence interval was used to determine if the known parameter value, μ, was included in the confidence interval for different sample sizes.

Once a method was established as accurate and a reasonable value for the sample size n was determined, the parameters of two datasets for crude oil and natural gas prices were estimated. The assumptions for a GBM were tested for the two datasets by checking the normality of the log-ratios with constant mean and variance and the independence from previous data (log ratios independent of their past values). The first assumption was verified using histograms and Q-Q plots. The second assumption was verified using correlograms of the data to compare correlation of the current value to previous values.

Historical datasets on the price of crude oil and natural gas were obtained from the U.S. Energy Information Administration (EIA) for the West Texas Intermediate (WTI) (EIA, 2017b) and the Henry Hub (EIA, 2017b), benchmarks in oil and natural gas pricing, respectively.

3. Results

Given the estimates of the parameter values using the simulated dataset, all three methods are capable of estimating μ for n values as low as 100. At n = 100, the lowest tested value of n, the true value of μ is contained within the calculated confidence interval (see Table 1). However, the confidence intervals are very wide up to n = 1000 as shown in Figure 1. These estimates are also subject to wide variation as shown in Figure 2. From this figure, it can also be seen that for n = 1000 and n = 4000, the estimates of μ generally stay within 10% of the true value. Therefore, n = 1000 was chosen as the basis for evaluating the real data sets.

It should be noted that the three methods produced very similar results (see Table 1) for most values of n. Therefore, Method 2 was used to estimate the parameters for the real datasets because it had the least computational complexity.

Table 1: Parameter estimates for known parameter values

	$\hat{\mu}$ ($\mu = 0.05$)			Confidence Interval Half Width			$\hat{\sigma}$ ($\sigma = 0.15$)		
n	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
100	0.05781	0.05781	0.05662	0.03154	0.03139	0.03072	0.16094	0.16014	0.15672
200	0.03711	0.03711	0.03662	0.02078	0.02073	0.02031	0.14993	0.14955	0.14658
400	0.03787	0.03787	0.03739	0.01555	0.01553	0.01525	0.15870	0.15850	0.15565
600	0.05190	0.05190	0.05089	0.01208	0.01207	0.01185	0.15095	0.15082	0.14803
800	0.05320	0.05320	0.05206	0.01043	0.01042	0.01015	0.15048	0.15039	0.14653
1000	0.04878	0.04878	0.04783	0.00904	0.00903	0.00883	0.14579	0.14571	0.14238
2000	0.05106	0.05106	0.05004	0.00652	0.00652	0.00638	0.14875	0.14872	0.14555
4000	0.04947	0.04947	0.04854	0.00464	0.00464	0.00455	0.14975	0.14973	0.14689
8000	0.05237	0.05237	0.05133	0.00330	0.00330	0.00324	0.15044	0.15043	0.14766
16000	0.05104	0.05104	0.04998	0.00232	0.00232	0.00226	0.14986	0.14986	0.14586

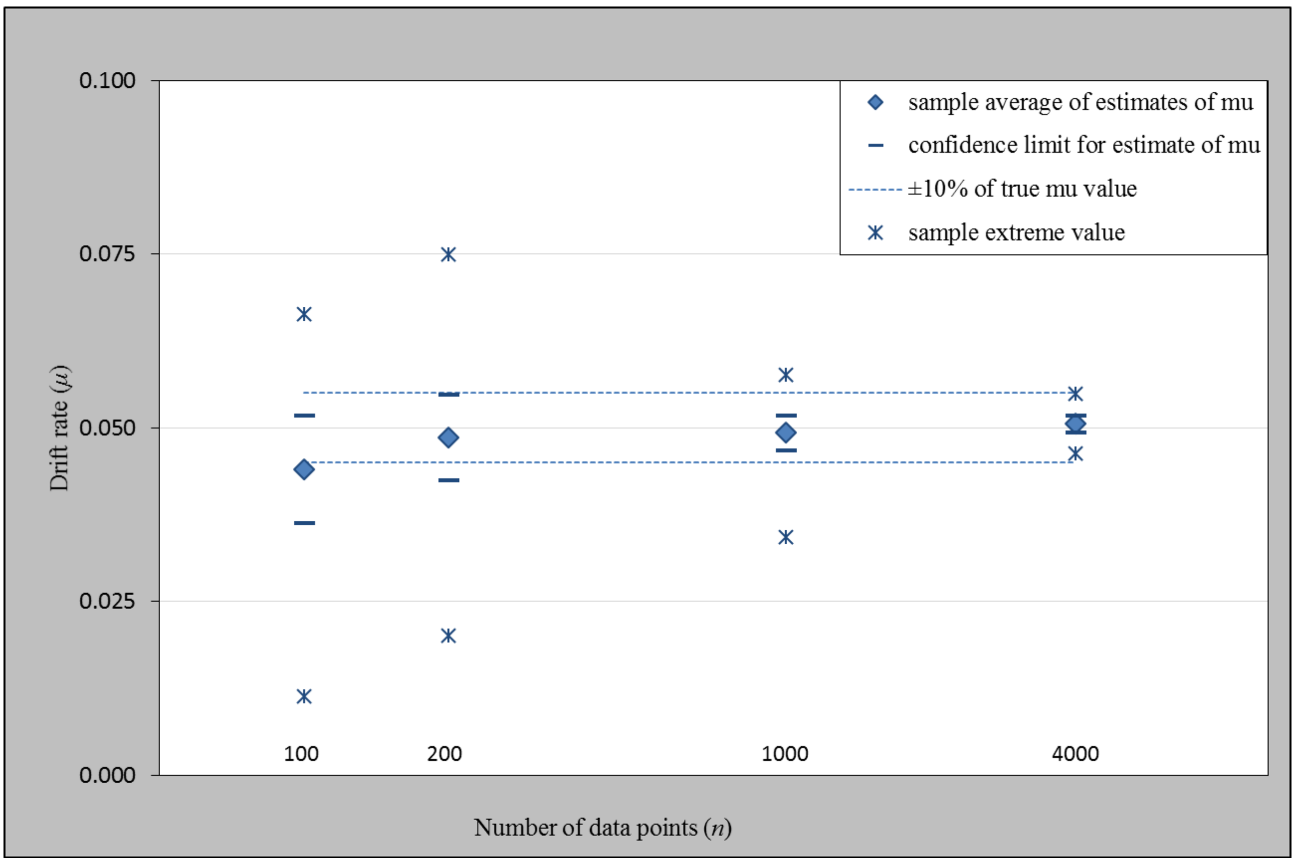


Figure 1: 95% Confidence intervals for μ = 0.05 and σ = 0.15 vs. *log* (n)

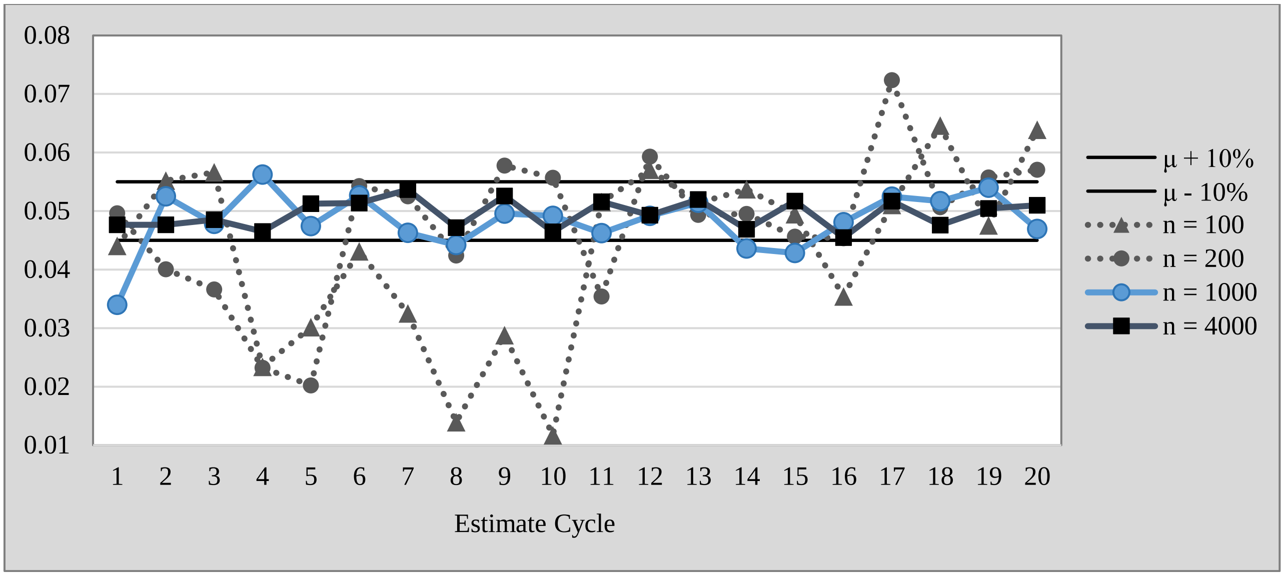


Figure 2: Sample estimates of μ = 0.05

The datasets for the daily WTI spot prices of crude oil and the Henry Hub values for natural gas futures were modeled as GBM processes. The graphs of $x(t)$ vs. time are shown in Figures 3 and 4.

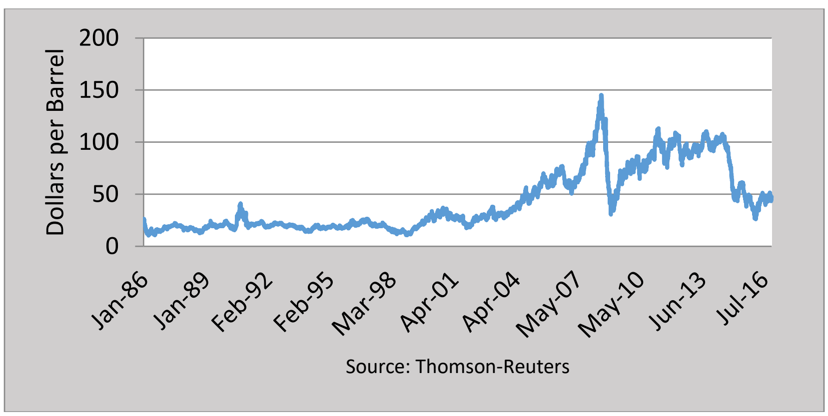


Figure 3: Cushing, OK WTI spot price FOB, daily

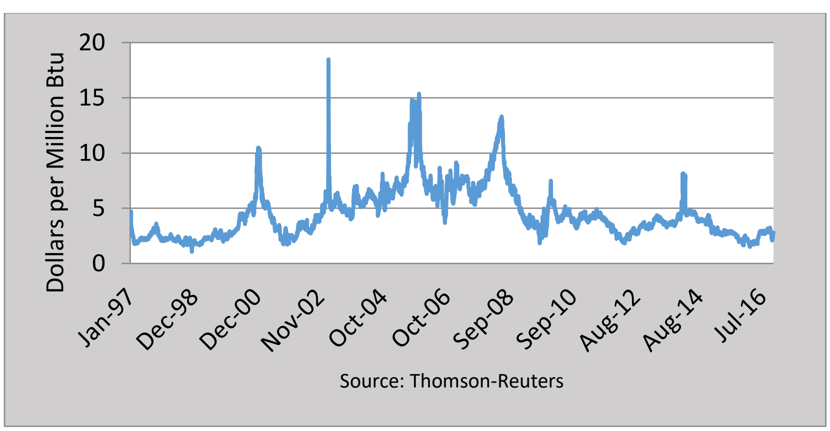


Figure 4: Henry Hub natural gas spot price, daily

Using the statistical package JMP, Q-Q plots and histograms for $log\left(\frac{x_t}{x_{t-1}}\right)$ were generated and a normal distribution was fitted to the data. The results for the two datasets can be seen in Figure 5. Ideally a Q-Q plot for a normal distribution would follow a straight line along the diagonal of the plot.

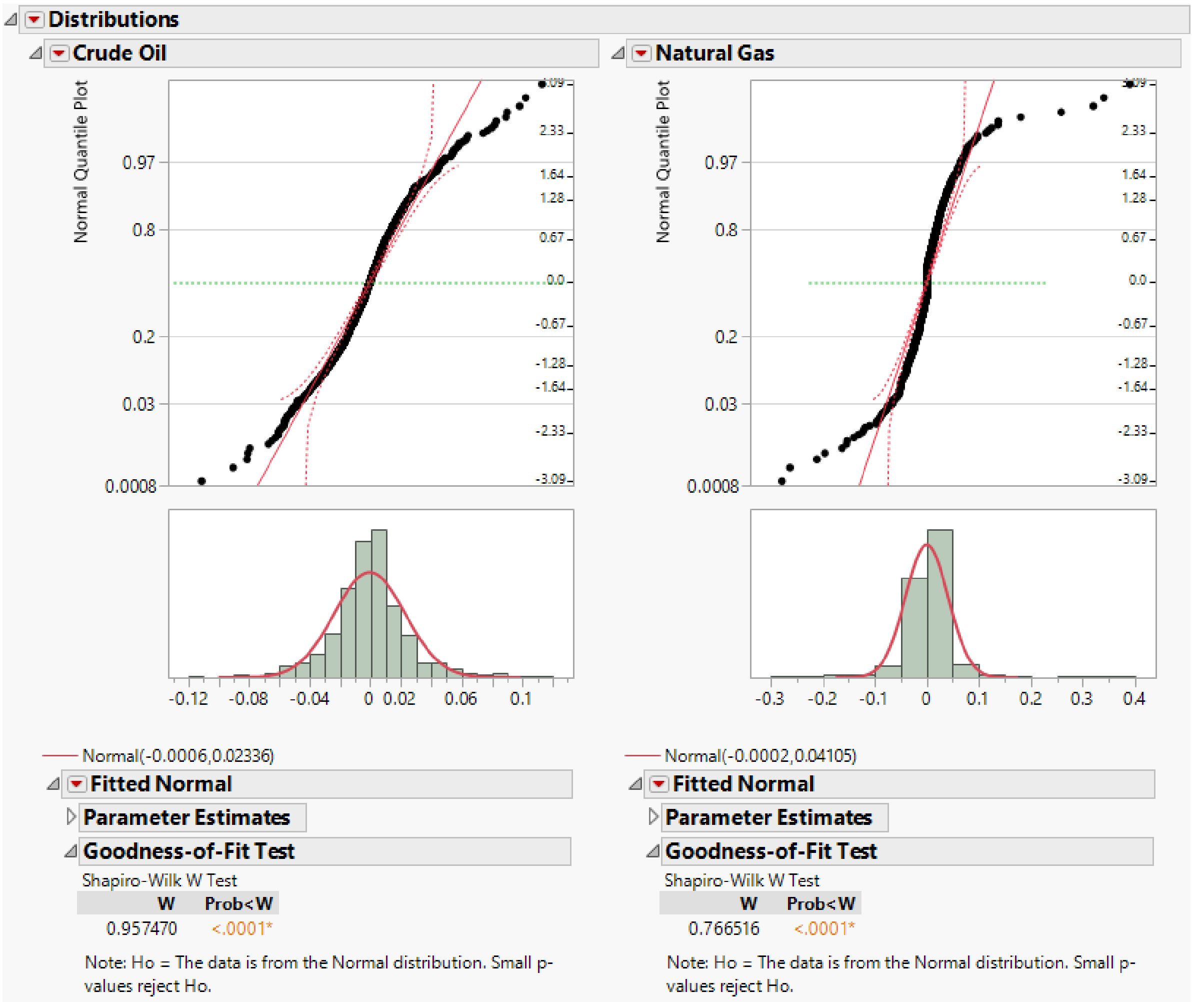
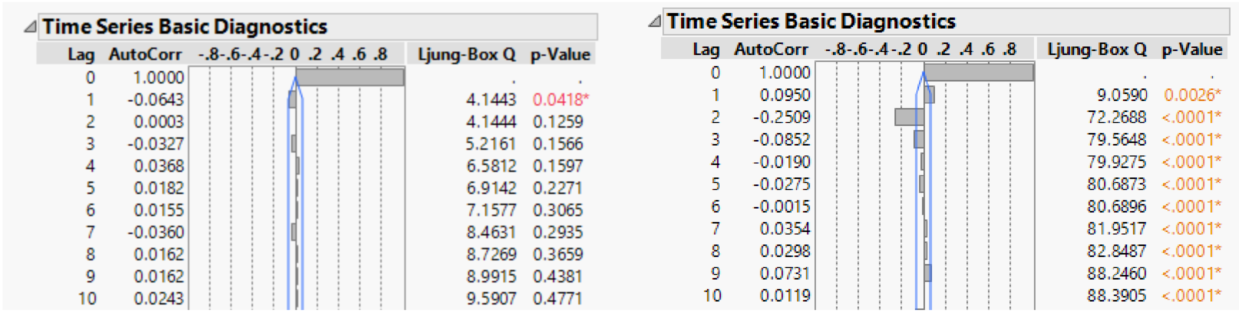


Figure 5: Q-Q plots of the log ratios against the normal distribution and histograms (n = 1000) generated using JMP

The Shapiro-Wilk W test was used to evaluate the goodness of fit. In this test, the null hypothesis states the distribution is normal. The test yields a W value and a corresponding p-value for α = 0.05, the latter of which is compared to the significance level of the test. If the p-value is greater than α, it fails to reject the null hypothesis. If the p-value is less than α, it rejects the null hypothesis. The p-value for both our data sets is less than α. This causes a rejection of the null hypothesis. Additionally, both histograms show a slightly skewed distribution to the right hand side. However, the log ratios of the crude oil prices follow a nearly normal distribution with a slight deviation from normality. These deviations can be affected by phenomena that were not taken into account with our model such as the seasonality of demand for oil. The log ratios of natural gas prices fail to exhibit the expected normality, as the tails of the distribution deviate significantly from the straight line.

The datasets were evaluated for independence by estimating the autocorrelation of $log\left(\frac{x_t}{x_{t-1}}\right)$ and plotting the correlograms. As there are negligible autocorrelation values for lags greater than zero (see Figure 6), the log ratio of crude oil prices and gas futures appear to be independent.



Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value
0	1.0000	.	.
1	-0.0643	4.1443	0.0418*
2	0.0003	4.1444	0.1259
3	-0.0327	5.2161	0.1566
4	0.0368	6.5812	0.1597
5	0.0182	6.9142	0.2271
6	0.0155	7.1577	0.3065
7	-0.0360	8.4631	0.2935
8	0.0162	8.7269	0.3659
9	0.0162	8.9915	0.4381
10	0.0243	9.5907	0.4771

Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value
0	1.0000	.	.
1	0.0950	9.0590	0.0026*
2	-0.2509	72.2688	<.0001*
3	-0.0852	79.5648	<.0001*
4	-0.0190	79.9275	<.0001*
5	-0.0275	80.6873	<.0001*
6	-0.0015	80.6896	<.0001*
7	0.0354	81.9517	<.0001*
8	0.0298	82.8487	<.0001*
9	0.0731	88.2460	<.0001*
10	0.0119	88.3905	<.0001*

Figure 6: Correlograms of crude oil and natural gas log ratios (n = 1000)

4. Conclusions

Using a GBM with known parameters, we generated a dataset using Monte Carlo simulation to study the performance of different parameter estimation methods. We found that all three methods had similar performance and that a sample size of 1000 or more afforded accurate estimates of the parameters. In validating a GBM model, it is important to verify that the assumptions for a GBM are satisfied for a given dataset, including normality and independence of ratios. We sought to establish two important items – the ability of the parameters estimation methods to provide good estimates of the parameters and to establish the validity of using a GBM to model the phenomena under examination. Using actual datasets for crude oil prices and natural gas futures, we found that the log ratios did not have a normal distribution. Therefore, a GBM would not be an accurate representation of the price. Estimating the drift and volatility parameters for a GBM can provide insights on future behavior and provide a basis for informed decision-making. These are the first important steps in making a quantitative decision making model that can be used to decide when to purchase crude oil. Future research will examine other stochastic processes' ability to model similar phenomena.

Acknowledgements and Disclaimer

The authors would like to thank the two anonymous referees for constructive and developmental feedback and other comments. This material is based upon work supported by the National Science Foundation under Grant No. 1504912 and support from the Department of Industrial and Manufacturing Systems Engineering (IMSE) Undergraduate Research Assistantship (URA) program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the IMSE department.

References

1. EIA, United States Energy Information Administration. (2017a). Cushing, OK WTI Spot Price FOB [Data file]. Retrieved from http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RWTC&f=D.
2. EIA, United States Energy Information Administration. (2017b). Henry Hub Spot Prices [Data file]. Retrieved from http://www.eia.gov/dnav/ng/hist/rngwhhdd.htm.
3. Hull, John C. *Options, Futures, and Other Derivatives*. Upper Saddle River, New Jersey: Prentice Hall, 2008.
4. Hurn, A. S., K. A. Lindsay, and V. L. Martin. "On the Efficiency of Simulated Maximum Likelihood for Estimating the Parameters of Stochastic Differential Equations." *Journal of Time Series Analysis* 24, no. 1 (2003): 45-63.
5. Marathe, Rahul Ratnakar, and Sarah M. Ryan. "On the Validity of the Geometric Brownian Motion Assumption." *Engineering Economist* 50, no. 2 (2005): 159-192.
6. Phillips, Peter C. B., and Jun Yu. "Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance." *Handbook of Financial Time Series*, 2009: 497-530.
7. Postali, Fernando A.S., and Paulo Picchetti. "Geometric Brownian Motion and structural breaks in oil prices: A quantitative analysis." *Energy Economics* 28 (2006): 506-522.