September 2019

We don't know how to accurately forecast future stock prices. We do know that you will be very, very, very wrong if you try to forecast it using Monte Carlo simulations

---

The viability of using geometric brownian motion to forecast stock prices over various timeframes: evidence from a sample of South African listed companies.

PRECIUM INVESTMENTS

Table of Contents

Table of Contents 2

Abstract 3

Introduction 4

Background 4

Simulating Stock Prices 6

Input Data 6

Hypotheses 7

Reviews of Similar Studies 10

Results 11

Conclusions 12

Abstract

This research report seeks to understand how valid and accurate it would be to use Monte Carlo simulations that make use of geometric brownian motion ("GBM"), to simulate stock price paths and ultimately, future stock prices for various timeframes. GBM was chosen for this report as it is the most commonly used stochastic process[1] in stock price modelling. These simulated stock prices are tested against actual observed stock prices for the same timeframe to assess the accuracy of predicting stock prices using GBM. This is explained in further detail later on in this report.

As the 95% confidence interval[2] is the most commonly used interval size in statistical analysis, we have chosen to also make use of this interval size. The sample for this study consists of 10 large cap stocks listed on the Johannesburg Stock Exchange. Daily stock price data was obtained from the Google Finance database over the period 31 August 2013 to 31 August 2019 (6 years). The reason for using 6 years of historical data is to use the initial 5 years' data to simulate the 6th year's values, which were then compared against actual, observed values. The timeframes predicted and forecasted against were: 1 day, 1 week, 1 month, 6 months and 1 year, in order to assess whether there were differences in the accuracy of the prediction over different time periods. Our instinct, prior to the testing, was that accuracy was likely to break down the further into the future we forecasted. The specific stocks chosen were Naspers(NPN), Anglo American(AGL), Sasol(SOL), FirstRand(FSR) Standard Bank(SBK), MTN(MTN), Nedbank(NED), Amplats(AMS), Sanlam(SLM) and Discovery(DSY).

[1]A statistical process used to model randomness

[2]An interval estimate computed from the statistics of observed historical data, that has the possibility of containing the true value of an unknown parameter

Introduction

Financial markets are prime examples of stochastic systems. This means that the market is composed of several different components that exhibit a high level of uncertainty and randomness. Some examples of these components are the individual, and groups of, investors participating in the market, the current state of the economic cycle, and the competitive and collaborative interactions between companies on the stock market. Other factors that might influence stock prices are inflation rates, unemployment rates and major political events. These factors affect one another and ultimately all may have a different weighted effect on the price movements of stocks that trade on an exchange, at different times.

There is an abundance of public information regarding the pricing of securities, however there is still a lot of debate surrounding which method is most reliable. Many market participants are interested in simulating stock prices in order to make important investment and financing decisions. Testing for the accuracy of these simulations thus becomes extremely important. Not much back-testing has been done with regards to using GBM to predict stock prices for South African and this study aims to shed some light on this particular method's feasibility.

Background

Brownian Motion

Brownian motion dates back to the nineteenth century when botanist Robert Brown, first described the phenomenon, whilst observing pollen particles floating in water under a microscope. He observed that the pollen particles were moving with no external force being applied to them. The first person to describe the mathematics behind Brownian motion was Thorvald N. Thiele in a paper on the method of least squares[3] published in 1880. The french mathematician Louis Jean-Baptiste Alphonse Bachelier, as part of his PhD thesis The Theory of Speculation (1900), went on to be the first person to use this random motion to model the stochastic process. Albert Einstein (1905) and Marian Smoluchowski (1906) later presented Brownian motion as a way to prove the existence of molecules and atoms.

Geometric Brownian Motion

Today, the stochastic process, Geometric Brownian Motion (GBM) is used in various derivative and security pricing techniques. The following formula is widely used in stock price simulations and similarly used in this study:

$$S_{t+1} = S_t . exp[(\mu - \frac{\sigma^2}{2})\Delta t + \sigma\varepsilon\sqrt{\Delta t}] \,..$$

(Equation 1)

Where:

S_t = *the stock price at a specified time*

μ = *the expected rate of return*

σ = *the expected volatility*

ε = *is a randomly drawn number from a* $N(0, 1)$ *distribution*

Δt = *the time that has passed from t to* $t+1$

This formula states that the future stock price S_{t+1} is the current stock price S_t, grown exponentially by two factors:

1. The constant drift that has historically been experienced by the stock (defined as $(\mu - \frac{\sigma^2}{2})\Delta t$) which explains the general direction of the stock's price,
2. A random stochastic component - $\sigma\varepsilon\sqrt{\Delta t}$

The constant drift is the historical expected rate of return experienced by the stock (μ), where this expected return is eroded by the volatility caused by market participants and other external factors. By adding the random stochastic component, the rates of return being used in the simulations are randomised.

The values used for μ in equation 1 can vary. Commonly used values include, but are not limited, to the following:

- The historical periodic returns experienced by the stock over a specified time period
- The risk free rate an investor can otherwise achieve in the market over a specified time period,

[3]This is the standard approach to regression analysis. It aims to minimise the sum of the differences between an observed value and the value fitted by the chosen model.

usually linked to a government bond rate

- The expected return of the stock based on the Capital Asset Pricing Model (CAPM) which is where the $expected\ return = risk\ free\ rate\ + \beta * (market\ return - risk\ free\ rate)$

We have tested each of these 3 values in this study.

The History of Monte Carlo Simulations (MCS)

Monte Carlo simulations refer to a wide range of computational algorithms that utilise randomness in some way or form. These simulations are mostly used to solve problems that might be impossible to solve in a typical algebraic or numerical manner, because they do not claim to give a definitive answer, but rather an approximation in the form of a range of outcomes.

The technique was first developed by Stanislaw Ulam, a mathematician who worked on the Manhattan Project (the research and development undertaking during World War II that produced the first nuclear weapons). After the war, while recovering from brain surgery, Ulam entertained himself by playing countless games of solitaire. He became interested in plotting the outcome of each of these games in order to observe their distribution and determine the probability of winning. After he shared his idea with John Von Neumann, the two collaborated to develop the Monte Carlo simulation process.

In 1946, the physicists at Los Alamos Scientific Laboratory were investigating radiation shielding and the distance that neutrons would likely travel through various materials. Despite having most of the necessary data, such as the average distance a neutron would travel in a substance before it collided with an atomic nucleus, and how much energy the neutron was likely to give off following a collision, the Los Alamos physicists were unable to solve the problem using conventional, deterministic mathematical methods. Ulam's idea of using random experiment gained from playing solitaire was used. Being secret, the work of von Neumann and Ulam required a code name. A colleague of von Neumann and Ulam, Nicholas Metropolis, suggested using the name Monte Carlo, which refers to the Monte Carlo Casino in Monaco where Ulam's uncle would borrow money from relatives to gamble.

Simulating Stock Prices

The idea behind simulating future stock prices and modelling the probability of stock prices is to generate a large number of 'random walks' based on a stochastic stock price model. Each random walk follows a Geometric Brownian Motion (GBM) using $(\mu - \frac{\sigma^2}{2})\Delta t$ as the drift component and $\sigma\varepsilon\sqrt{\Delta t}$ as a volatility shock. μ is the expected return, σ is the expected standard deviation of based on historical returns and Δt is the time step of the simulation. The historical data used for the calculation of these variables ranged from 31 August 2013 to 31 August 2018.

In each step of the MCS the next price is calculated using the formula:

$$S_i = S_{i-1} exp[(\mu - \frac{\sigma^2}{2})\Delta t + \sigma\varepsilon\sqrt{\Delta t}$$

where i = 1, ..., t, S_0 is given by the latest historical price and ε is a random value from the normal distribution with mean 0 and standard deviation 1.

Daily stock price movements for the period 3 September 2018 to 31 August 2019 were simulated. The simulation was run 10,000 for each of the 10 stocks. Important simulation dates tested against actual observed values were:

- 1 day 3 September 2018
- 1 week 7 September 2018
- 1 month 2 October 2018
- 6 months 3 January 2019
- 1 year 30 August 2019

As an example of what the simulations could look like, see figures 1,2,3 and 4 below. It showcases 10 daily simulations of Naspers(NPN), MTN(MTN), Discovery(DSY) and Anglo American(AGL) stocks over the period 2 September 2018 to 31 August 2019 as well as the actual price movement observed over the period 1 September 2017 to 31 August 2019 (Initial blue line).

Input Data

The daily closing stock prices were used for the sample of stocks chosen for this simulation test.

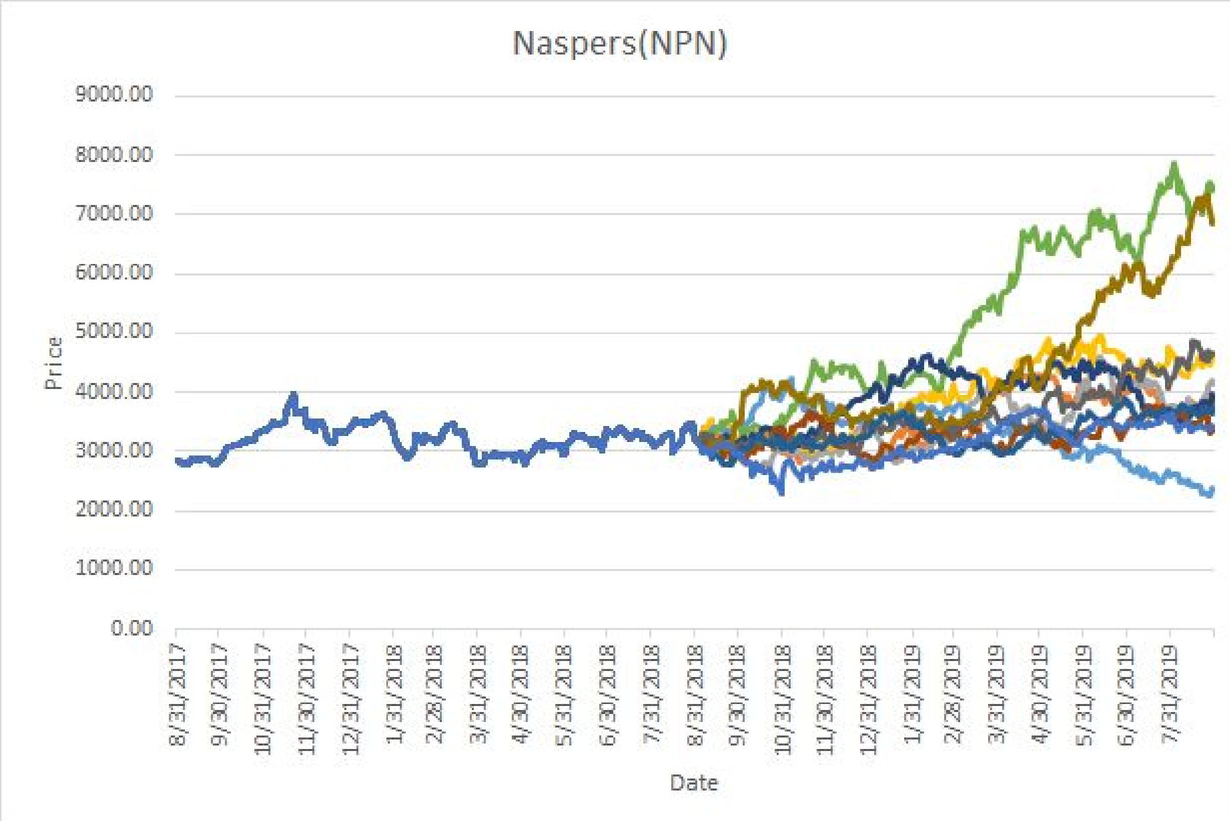
The major inputs into the simulation are the expected returns and volatility of each stock. As mentioned before the expected returns are calculated in 3 ways (the historical return, the risk free rate[4] and the expected return based on CAPM). In addition to this, the historical period used to calculate μ and σ was also changed for each test. The historical periods used for each stock were, 1 month, 6 months, 1 year and 5 years.

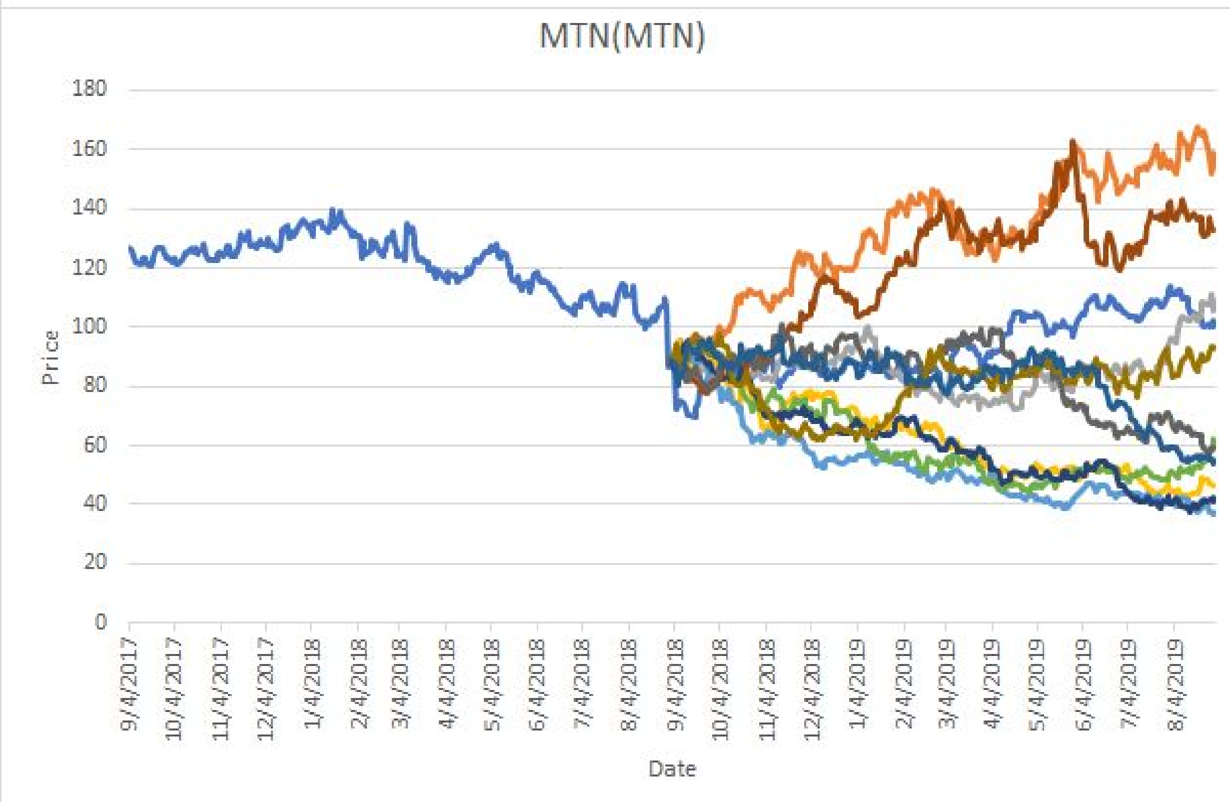
Hypotheses

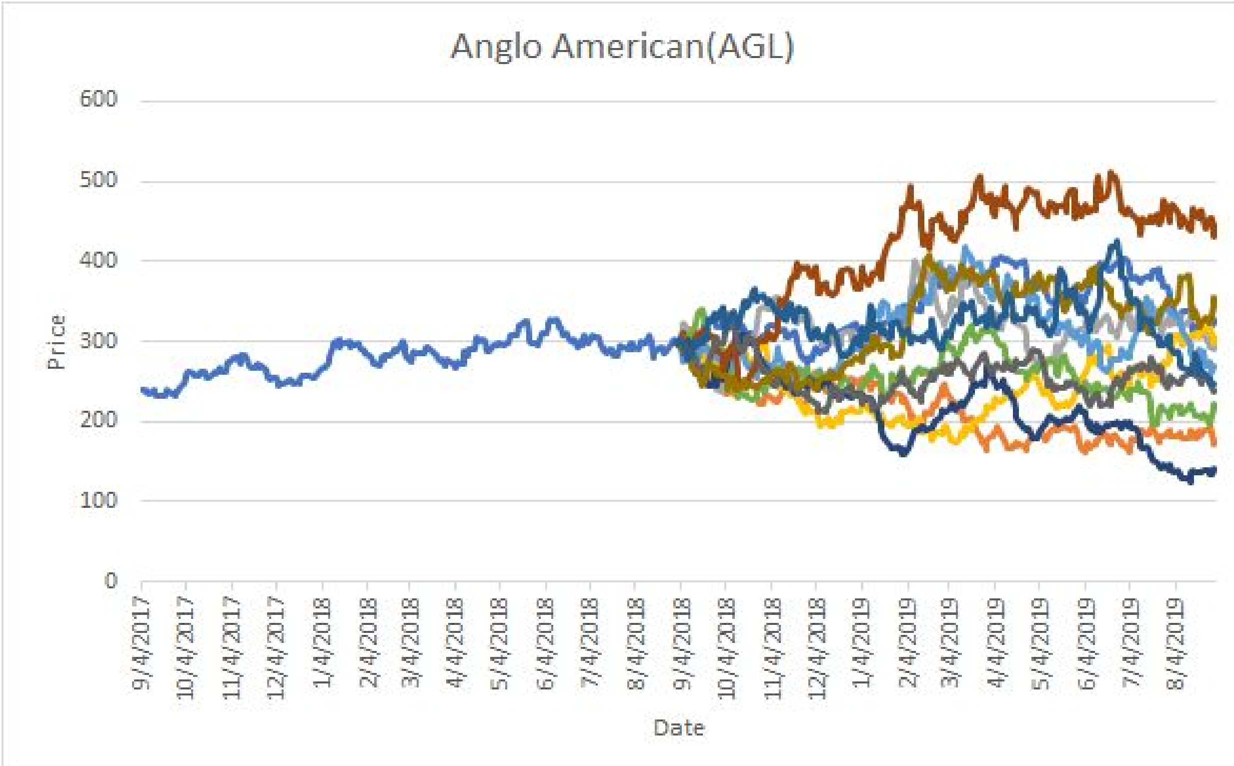
Based on the nature of the GBM being used to simulate the daily stock movements and sample simulation graphs such as figures 1,2,3 and 4 below we make the following hypotheses:

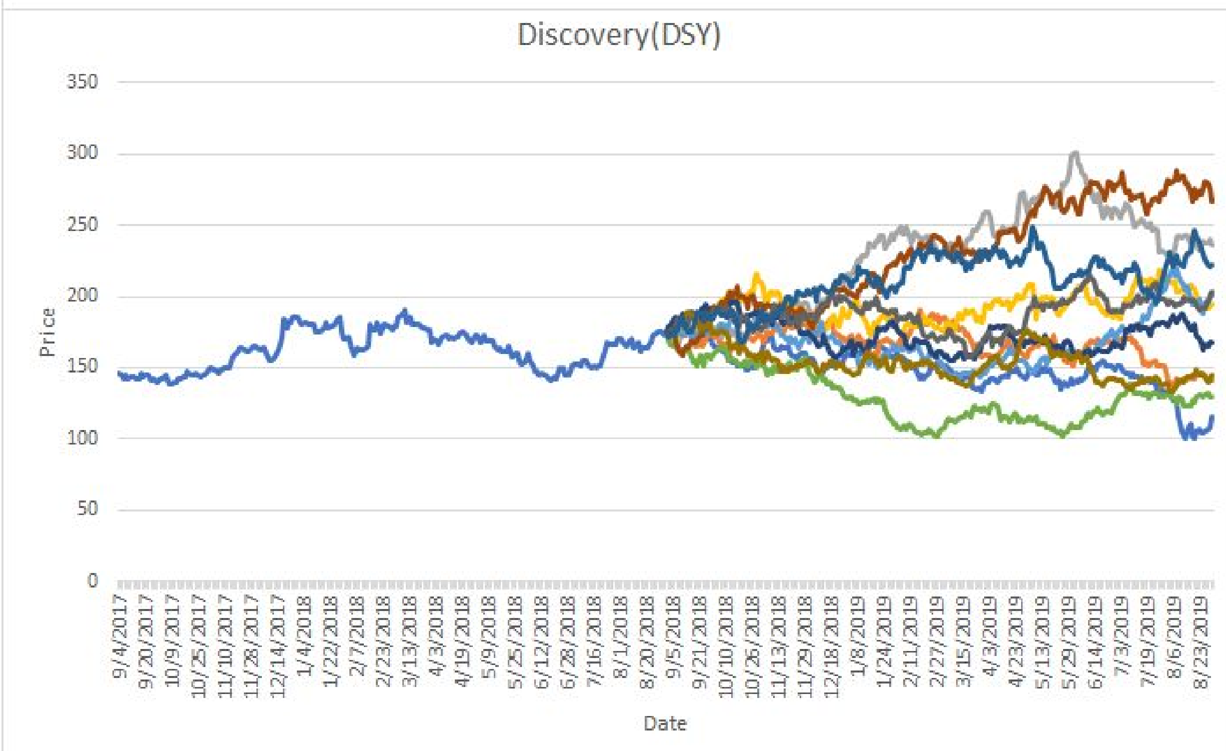
1. Geometric brownian motion, even over shorter time periods is not a viable prediction method for stock prices and should not be extensively used to make investment decisions.
2. The accuracy of the stock prices simulated using geometric brownian motion deteriorates as the time horizon for the simulation is increased.
3. Using longer historical time periods (i.e 5 years historical data vs. 1 years historical data) lead to less accurate forecasting results

[4]The R186 yield was used for the risk free rate


Naspers(NPN)
9000.00
8000.00
7000.00
6000.00
5000.00
4000.00
3000.00
2000.00
1000.00
0.00
Price
8/31/2017
9/30/2017
10/31/2017
11/30/2017
12/31/2017
1/31/2018
2/28/2018
3/31/2018
4/30/2018
5/31/2018
6/30/2018
7/31/2018
8/31/2018
9/30/2018
10/31/2018
11/30/2018
12/31/2018
1/31/2019
2/28/2019
3/31/2019
4/30/2019
5/31/2019
6/30/2019
7/31/2019
Date


MTN(MTN)
180
160
140
120
100
80
60
40
20
0
Price
9/4/2017
10/4/2017
11/4/2017
12/4/2017
1/4/2018
2/4/2018
3/4/2018
4/4/2018
5/4/2018
6/4/2018
7/4/2018
8/4/2018
9/4/2018
10/4/2018
11/4/2018
12/4/2018
1/4/2019
2/4/2019
3/4/2019
4/4/2019
5/4/2019
6/4/2019
7/4/2019
8/4/2019
Date





Figures 1,2,3 and 4 - *Ten Simulations 12 months into the future.*

Reviews of Similar Studies

In their paper: *"Simulating Stock Prices Using Geometric Brownian Motion: Evidence from Australian Companies"* (2016), Krishna Reddy and Vaughan Clinton (University of Wollongong, Australia) make use of GBM to simulate stock price movements and test if these simulated stock prices align with actual stock prices observed. They focussed on 50 large, listed, Australian companies, running simulations for various forecast periods (1 week, 2 weeks, 1 month, 6 months and 1 year). By making use of mean absolute percentage errors (MAPE), amongst other testing methods, they came to the conclusion that using this simulation technique is quite accurate for predicting stock prices. They did however find that the longer the time horizon for the prediction was the less accurate the prediction became. They do not make any mention of how many simulations were run for each individual stock. From the way that they present their findings, it seems that they only ran one simulation for each individual stock, which causes us to doubt the validity of the results produced in their paper. If one runs only one simulation, one could cherry pick the simulation that supports one's hypothesis. The probability of a stock following a single random simulated path is effectively zero.

In their paper: *"Monte Carlo Simulations of Stock Prices: Modelling the probability of future stock returns"* (2018), Tobias Brodd and Adrian Djerf (KTH Royal Institute of Technology: Sweden) make use of GBM to model the probability of future stock prices. They focussed on 10 Swedish large-cap stocks, running 100 000 simulations over various time periods (1 month, 3 months, 6
months, 9 months and 1 year) for each individual stock. They found that this method was not statistically significant. However, they found that if they used weighted values for mean returns and standard deviations, this method became statistically significant for only a 1 month prediction period. This led them to the conclusion that by varying the historical data used and using different weights, this method has the potential to become more accurate. This conclusion is fairly obvious as varying the inputs to the simulation model will naturally have a different impact on the simulated prices. It is for this reason that our simulations use a multitude of different inputs.

In his masters dissertation, *"Stock price modelling: Theory and Practice"* (2006), Abdelmoula Dmouj (Vrije Universiteit: Netherlands) tested the accuracy of modelling stock prices by making use of GBM. He makes use of only one stock, Hewlett-Packard, to do this. He simulated 1 year stock prices 1000 times (for the years 1997 - 2005) and finds that GBM is less accurate over shorter forecast periods and becomes more accurate over longer periods of time. This contradicts the findings of the other papers, including our own research. His method for testing consisted of checking if actual stock prices fell within the 95% confidence interval produced by the simulated stock prices. Because his sample size was so small (1000) his confidence intervals were very large, one example of this would be his 1 year 95% confidence level for 2001, which ranged from 35.23 to 260.09. This presents a very large target to aim at and equates to false positive results.

Results

In order to test the viability of the model we made use of 95% confidence intervals. Due to the fact that we ran 10,000 simulations for each case being investigated our intervals were quite narrow. We went on to check if the actual prices observed fell within these simulated confidence intervals over the various testing horizons. Table 1 is an example of the output obtained. All figures are quoted in South African Rand. Column 1 indicates which historical data type was used in the simulation. Columns 2,5,8,11 and 14 show the lower bound (LCL) of the 95% confidence interval based on the historical data type used and the forecast period. Columns 3,6,9,12 and 15 show the upper bound (UCL) of the 95% confidence interval based on the historical data type used and the forecast period. Columns 4,7,10,13 and 16 gives the actual share price observed for the various forecast periods. We found that on actual share prices fell within the 95% confidence interval once for 1 day forecasts, once for 1 week forecasts, zero times for 1 month forecasts, once for 6 month forecasts and 2 times for 1 year forecasts. Thus, these various iterations showed that this method led to accurate predictions less than 2% of the time.

	Day			Week			Month			6Month			Year		
DataUsed	LCL	UCL	Actual	LCL	UCL	Actual	LCL	UCL	Actual	LCL	UCL	Actual	LCL	UCL	Actual
1mCAPM	3173.96	3177.65	3107.59	3209.76	3218.19	3055.54	3361.20	3379.49	2908.82	4532.82	4593.97	3045.39	6480.73	6608.03	3453.80
1mHist	3168.93	3172.60		3184.80	3193.22		3245.85	3263.59		3670.26	3719.72		4242.49	4325.00	
1mRFR	3168.84	3172.55		3180.08	3188.33		3241.46	3259.02		3659.64	3708.60		4266.24	4349.65	
6mCAPM	3166.50	3169.46		3168.01	3174.55		3186.43	3200.17		3303.88	3338.77		3471.38	3524.29	
6mHist	3167.00	3169.96		3170.52	3177.07		3197.03	3210.82		3369.33	3404.91		3610.28	3665.30	
6mRFR	3168.19	3171.12		3179.82	3186.51		3224.17	3238.19		3527.64	3565.12		3922.85	3982.39	
1yCAPM	3166.97	3169.63		3173.22	3179.30		3196.02	3208.66		3349.44	3381.73		3537.54	3586.10	
1yHist	3168.27	3170.93		3179.73	3185.83		3223.67	3236.42		3524.20	3558.18		3916.31	3970.08	
1yRFR	3167.90	3170.60		3174.91	3180.88		3214.41	3227.02		3477.40	3510.73		3843.33	3896.28	
5yCAPM	3167.48	3170.10		3172.73	3178.53		3204.77	3217.00		3415.87	3447.69		3707.91	3757.52	
5yHist	3170.22	3172.84		3186.48	3192.31		3263.52	3275.97		3802.68	3838.11		4595.22	4656.69	
5yRFR	3167.93	3170.52		3177.91	3183.84		3215.89	3228.26		3474.21	3506.76		3806.28	3857.02	

Table 1: 95% confidence level results - Naspers

Conclusions

This study explored the geometric Brownian motion model for simulating stock price paths, and tests the validity of the method. Returning to our hypotheses the following observations can be made:

1. We maintain that GBM is not a viable prediction method for South African listed stocks.
2. The accuracy of the simulated stock prices vs the actual stock prices observed does deteriorate as the simulation time horizon is increased.
3. Regardless of the time frame used for historical data, the results all proved to be inadequate and thus not usable for future stock price prediction.

Since the objective behind monte carlo simulations is to state with a relatively high degree of confidence that a stock price will lie within a certain interval and not to give a specific value, we tested the probability of the actual stock price falling within the 95% confidence level predicted by the simulations. This study found that this was the case for less than 2% of the different iterations of historical data and forecast periods that we looked at.

If one were to average out the simulated stock prices in order to get to a more specific estimate, the randomness of the simulation would be negated as the number of simulations increases. This would ultimately lead to the simulated stock continuing on the path of the chosen drift component. We believe that the drift component is a fundamental flaw of the GBM model, regardless of which value is used for μ. The chosen drift ultimately leads to the largest amounts of simulations following the direction and magnitude of the chosen drift, this boils down to stating that a stock will continue on the path that it has followed up until now. However, this is not the case in an open market and a more dynamic method should be found to continuously update the drift component or even make certain assumptions regarding future growth expectations for the stock.

We do not completely discount the use of GBM or monte carlo simulations, we just believe that it would be better suited for other forecasting needs that have a higher tendency to follow historical trends and that are less volatile. We would therefore not recommend making any investment decisions regarding listed South African stocks by making use of this method.

We have not tested other different methods for stock price prediction/estimation such as multiple regression analysis, correlation analysis, or financial statement analysis and forecasting to assess their accuracy, and so cannot comment on how accurate or inaccurate they may be.