Drifting Along, But Where Am I Going?

Steven Lillywhite *

July 19, 2011

Abstract

Many securities such as stocks exhibit directionality in their prices. This is often modelled as a drift term in diffusion models of the price. We show why it is so hard to estimate this drift.

1 Introduction

Many people feel that stock prices tend to rise over time. Of course, they never rise continuously without any drops. Stock prices go up and down all the time, month after month, day after day, minute by minute. However, many stocks will appear to be going up more on average than going down. This can often be seen by looking at graphs of stock prices covering many years.

One way to attempt to describe this behavior is with stochastic models. One of the simplest models for (non-dividend paying) stock prices is that of geometric Brownian motion (GBM). It looks like the following.

$$\frac{dS}{S} = \mu dt + \sigma dW \tag{1.1}$$

xb Here, $S = S(t)$ represents the stock price, $W = W(t)$ represents a standard Brownian motion, and μ and σ are assumed to be constants. We call μ the *drift* and σ the *volatility*. Intuitively, the equation states that the instantaneous returns from the stock are given by a fixed rate of return μ plus a random amount. If there were no random component, then the return would always be μ and the stock price itself would rise as an exponential function

$$S(t) = S(0)e^{\mu t} \tag{1.2}$$

However, with the random component included, the stock can deviate from a little to a lot from the basic exponential path. (Note that this is not to say that the stock

*__Email__: stevenlillywhite2@yahoo.com

price is mean-reverting to a fixed exponential path). The figures at bottom show some sample paths from the same generating model (1.1). Of course, any particular figure could potentially represent a probabilistically extreme path. But we claim the selection below consists of fairly typical paths, in some sense.

To see why, consider the following. If we make the change of variable $Y = log(S)$ and apply Itô's lemma, we get

$$dY = (\mu - \tfrac{\sigma^2}{2})dt + \sigma dW \tag{1.3}$$

From this we derive the following

$$Y(t) = Y(0) + (\mu - \tfrac{\sigma^2}{2})t + \sigma W(t) \tag{1.4}$$

hence

$$S(t) = S(0)e^{[(\mu - \frac{\sigma^2}{2})t + \sigma W(t)]} \tag{1.5}$$

It follows that $S(t)|S(0)$ is log-normally distributed with

$$E[S(t)|S(0)] = S(0)e^{\mu t} \tag{1.6}$$

and

$$Var[S(t)|S(0)] = S(0)^2 e^{2\mu t}(e^{\sigma^2 t} - 1) \tag{1.7}$$

Hence for the examples below, with $S(0) = 45, \mu = 0.1$ and $\sigma = 0.2$, we find that $E[S(1)|S(0)] = 49.73$ and $Var[S(1)|S(0)] = 100.94$, giving a standard deviation of around 10. For the ten year plots we have $E[S(10)|S(0)] = 122.32$ and $V[S(10)|S(0)] = 7359.09$, or a standard deviation of around 85. So typical ending values would be in the range $40 - 60$ for the one-year plots and $37 - 207$ for the ten-year plots.

We note in passing the pathological behavior that occurs when $\mu > 0$ but $\mu - \frac{\sigma^2}{2} < 0$. In such a case, the expected value of S grows exponentially, while any given path tends toward zero. What happens is that any path will trend towards zero, but from time to time make a large upward excursion.

2 Estimating the Drift

Let us now consider the following inverse problem. We are presented with a time series, such as a history of stock prices, we assume the model (1.1), and we try to then estimate the parameters μ and σ using statistics or other means. We shall assume that we know σ. Not knowing σ adds complications, but does not alter the main point of this note.

Let's look at log returns. If we first take log of the prices, we have the model (1.4). From this it follows that for equal increments of time Δt

$$Y(t + \Delta t) - Y(t) = (\mu - \tfrac{\sigma^2}{2})\Delta t + \sigma[W(t + \Delta t) - W(t)] \tag{2.1}$$

This means that $Y(t+\Delta t) - Y(t)$ is distributed normally with mean $(\mu - \frac{\sigma^2}{2})\Delta t$ and variance $\sigma^2 \Delta t$. Let us now consider a time interval from time zero to T, divided into n equal increments of Δt. So $T = n\Delta t$. Let $Y_i \stackrel{\text{def}}{=} Y(i\Delta t)$ and $Z_i \stackrel{\text{def}}{=} Y_i - Y_{i-1}$. Hence the Z_i are independent draws from $N((\mu - \frac{\sigma^2}{2})\Delta t, \sigma^2 \Delta t)$.

2.1 Maximum Likelihood

Let us consider the maximum likelihood estimator(MLE) of μ. We form the joint probability distribution function of the (independent) Z_i, take its log, and call this function $l(Z, \mu)$. Then

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} Z_i - \frac{\mu T}{\sigma^2} + \frac{T}{2} \tag{2.2}$$

We may show that a maximum for l occurs when this is zero, giving the maximum likelihood estimate $\hat{\mu}$ for μ

$$\hat{\mu} = \frac{1}{T} \sum_{i=1}^{n} Z_i + \frac{\sigma^2}{2} \tag{2.3}$$

Now

$$E[\hat{\mu}] = \frac{1}{T} \sum_{i=1}^{n} E[Z_i] + \frac{\sigma^2}{2} = \frac{1}{T} \sum_{i=1}^{n} (\mu - \frac{\sigma^2}{2})\Delta t + \frac{\sigma^2}{2} = (\mu - \frac{\sigma^2}{2}) + \frac{\sigma^2}{2} = \mu \tag{2.4}$$

so that $\hat{\mu}$ is unbiased. Similarly

$$Var[\hat{\mu}] = \frac{1}{T^2} \sum_{i=1}^{n} Var[Z_i] = \frac{1}{T^2} \sigma^2 T = \frac{\sigma^2}{T} \tag{2.5}$$

With a simple application of the Cauchy-Schwartz inequality one can show that for any unbiased estimator of μ, M, its variance is bounded below by $Var(M) \geq \frac{1}{I(\mu)}$, where $I(\mu)$ is the *Fisher information*. This result is known as the Cramér-Rao Inequality. In our case, we may compute the Fisher information as

$$I(\mu) = -E\left[\frac{\partial^2 l}{\partial \mu^2}\right] = -E[-\frac{T}{\sigma^2}] = \frac{T}{\sigma^2} \tag{2.6}$$

Hence the MLE of μ achieves the Cramér-Rao lower bound for variance and is thus a minimum-variance unbiased estimator. So in this sense, it is a "best" estimator.

Remark 1. *Note that if we change variables back from Z to Y to S, this estimator is*

$$\hat{\mu} = \frac{[log(S(T)) - log(S(0))]}{T} + \frac{\sigma^2}{2} \tag{2.7}$$

In particular, it only depends on the first and last data points.

2.2 Uncertainty in the Estimate

Next we shall consider the uncertainty in the estimate $\hat{\mu}$. From the above calculations of the mean and variance, we have that the distribution of $\hat{\mu}$ is

$$\hat{\mu} \sim N(\mu, \frac{\sigma^2}{T}) \tag{2.8}$$

or equivalently

$$\frac{\sqrt{T}(\hat{\mu} - \mu)}{\sigma} \sim N(0, 1) \tag{2.9}$$

Now choose a real number α between zero and one. We wish to have confidence $1 - \alpha$, which means that typically we shall choose a value for α such as 0.05 or 0.01 in order to have 95%(resp. 99%) confidence. Let $N(x)$ denote the cumulative distribution function for the standard normal distribution $N(0, 1)$. Let $a > 0$ be such that $N(a) = 1 - \alpha/2$. Thus $P(-a < X < a) = 1 - \alpha$ for $X \sim N(0, 1)$. So we have probability of $1 - \alpha$ that

$$-a < \frac{\sqrt{T}(\hat{\mu} - \mu)}{\sigma} < a \tag{2.10}$$

or

$$\frac{-a\sigma}{\sqrt{T}} < \hat{\mu} - \mu < \frac{a\sigma}{\sqrt{T}} \tag{2.11}$$

So, for a given level of probability, α, this gives us bounds on how close our estimate $\hat{\mu}$ is to the actual value μ. Remember that it means that with probability $1-\alpha$, this bound will hold. So if we simulate paths, compute the estimate and use $\alpha = 0.05$, then on average, 95 times out of 100 this bound will hold. On average 5 times out of 100, it won't hold.

Remark 2. *We may derive the above results in a simpler fashion. For brevity, we'll consider Brownian motion with drift, although the arguments are the same for GBM. So let X be a Brownian motion with drift.*

$$X(t) = X(0) + \mu t + \sigma W(t) \tag{2.12}$$

We know that $Y \stackrel{def}{=} X(T)|X(0)$ is normally distributed with $E[Y] = E[X(T)|X(0)] = \mu T + X(0)$ and $Var[Y] = Var[X(T)|X(0)] = \sigma^2 T$. Hence, $\frac{Y - X(0) - \mu T}{\sigma\sqrt{T}} \sim N(0, 1)$. Thus for a given level of confidence α as above, choosing a as above so that $N(a) = 1 - \alpha/2$, we have with probability $1 - \alpha$

$$-a < \frac{Y - X(0) - \mu T}{\sigma\sqrt{T}} < a \tag{2.13}$$

and rearranging terms we get

$$\frac{-a\sigma}{\sqrt{T}} < \frac{Y - X(0)}{T} - \mu < \frac{a\sigma}{\sqrt{T}} \tag{2.14}$$

So we define our estimate of the drift $\hat{\mu} \stackrel{def}{=} \frac{Y - X(0)}{T}$, which is the MLE in this case, and we have recovered our results above.

Remark 3. *Note that whether we consider all the data, or only the first and last data points, our estimate of the drift is the same. To estimate the drift, what matters is the maximum time interval of the observations. On the other hand, to estimate the volatility, just observing the endpoints of the data won't help much. The best estimate will use all the data. This is all due to simple properties of Brownian motion.*

2.3 Punchline

Finally, let us plug in some numbers to illustrate how difficult it can be to estimate the drift μ. We shall round all numbers to the second decimal place. For stocks, typical values for μ ought to be in the range -0.2 to 0.2, more or less. Saying the drift of a stock is 0.1 is saying that the stock has an expected return of 10%. Suppose now, for a given level of confidence as expressed by α above, we wish to have a bound on our estimate of length δ, where δ shall be a small number. Specifically, let us impose that with confidence level α, we wish to have

$$-\frac{\delta}{2} < \hat{\mu} - \mu < \frac{\delta}{2} \tag{2.15}$$

which says that the true value μ is within δ of the estimate $\hat{\mu}$. We can solve for this by varying T. That is, let us solve

$$\frac{a\sigma}{\sqrt{T}} = \frac{\delta}{2} \tag{2.16}$$

so that

$$T = \frac{4a^2\sigma^2}{\delta^2} \tag{2.17}$$

Lets run some simple numbers. Let us take confidence level $\alpha = 0.05$, or 95% confidence. This gives a value for a of 1.96. Let us impose a precision of estimation $\delta = 0.01$. A typical volatility for a stock might be $\sigma = 0.2$. Putting this together gives

$$T = \frac{4a^2\sigma^2}{\delta^2} = 6146.33 \tag{2.18}$$

So for fairly reasonable assumptions, we find that it would take over six thousand years of data to estimate the drift within an accuracy of 0.01! If we change the volatility to $\sigma = 0.4$, then it would take over 24,000 years. Even for $\sigma = 0.05$, it would take 384 years.

If we keep the above numbers, and $\sigma = 0.2$, then even to find the drift within a bound of $\delta = 0.1$ requires 61 years of data. That is, 61 years of data just to find an estimate of μ, perhaps something like 0.15, and be able to say at a level of 95% confidence that $0.1 < \mu < 0.2$.

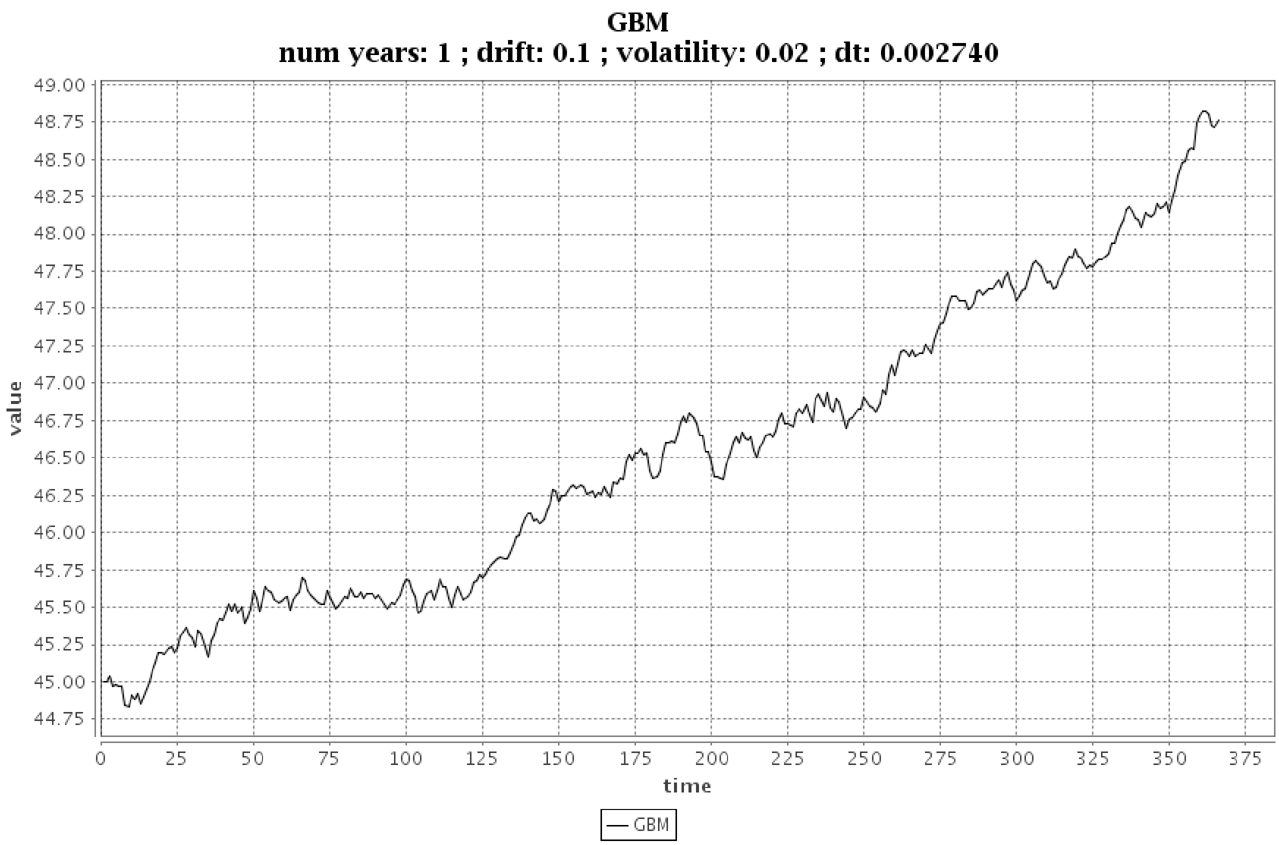
For a stock with ten years of history, we get $\delta = 0.25$. For most of these stocks, you can essentially say nothing about the drift with any good level of confidence. For a stock with 20 years of history, $\delta = 0.18$. With only a 90% level of confidence, $\delta = 0.15$. Same story. If we estimate the drift to be 0.1 for a stock with 20 years of history, then at a 90% confidence level, we can only assert that it is likely to be between 0.03 and 0.18, which isn't likely to impress or help anyone.

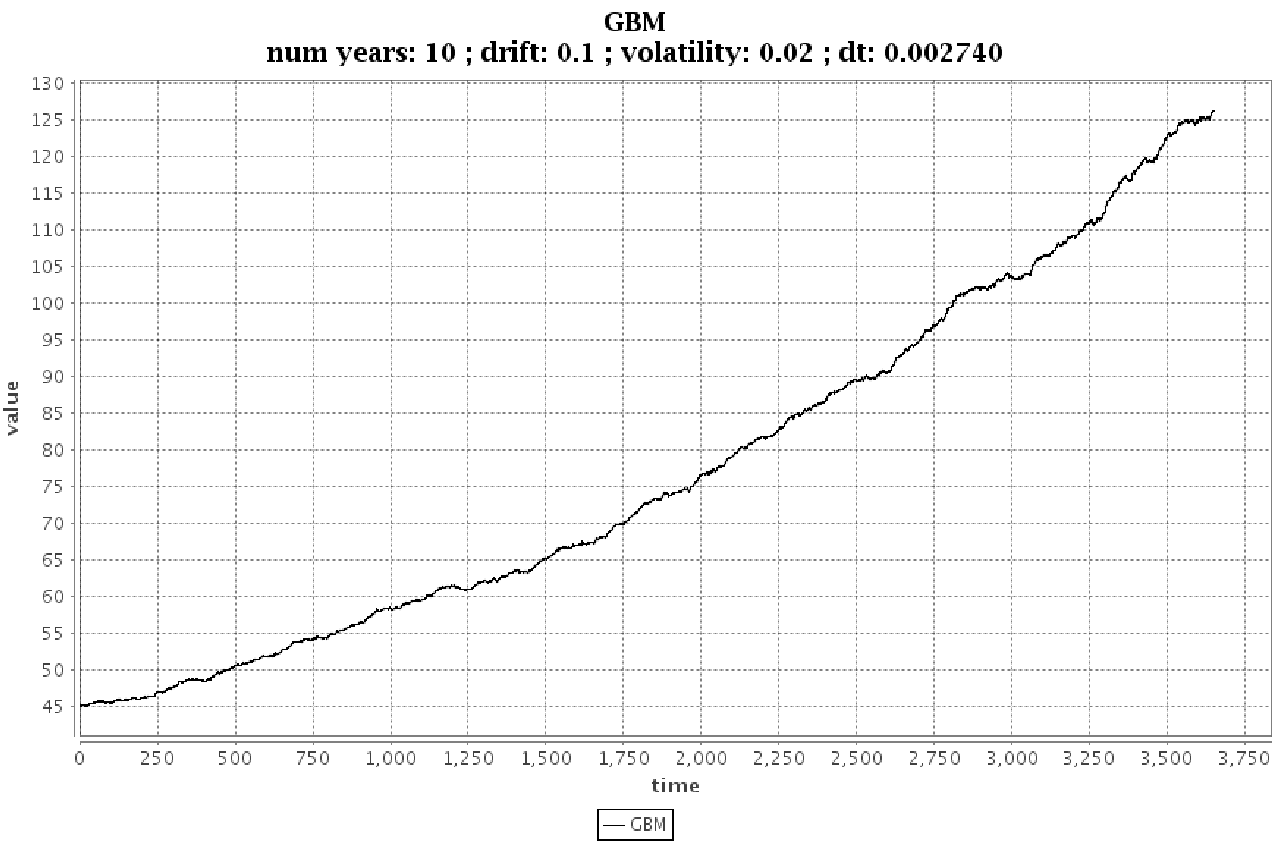
3 Summary

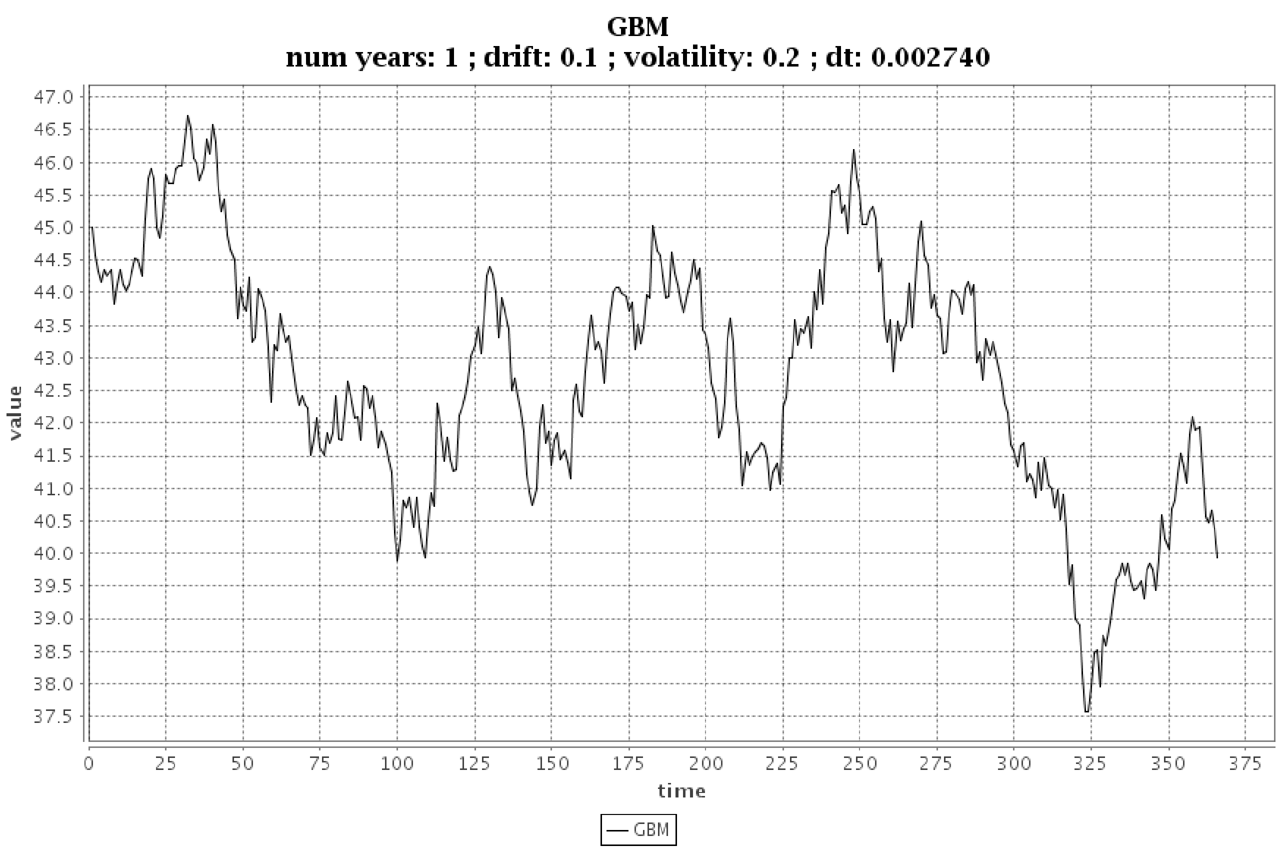
We have shown how estimates of the drift in a GBM model using a "best" estimator are quite imprecise over time scales commonly used in economic and financial applications. The same arguments apply as well for ordinary Brownian motion with drift. We should thus be suspicious whenever anyone attempts to estimate drifts in such models.

It is obvious that in a world of finite resources, nothing can grow forever at a positive exponential rate. So the use of GBM to model most phenomena would be interpreted as applying over the short term. For stocks, in particular, we would expect regime changes in the prices long before we would ever have enough data to estimate precisely the drift under a GBM model. Note that we are not claiming that GBM is a good or appropriate model to describe stock prices. We are saying that *even if it were*, no one would ever be able to estimate the expected rate of returns with any confidence for just about any stock you care to name.

Where the GBM model for stock prices often gets used is in the pricing of options. One of the most widely employed techniques is that of the risk-neutral framework, where the physical drift is not needed to compute option prices. The inherent difficulties in estimating the real-world drift of asset prices make such risk-neutral theories all the more appealing.


GBM
num years: 1 ; drift: 0.1 ; volatility: 0.02 ; dt: 0.002740
value
time
GBM


GBM
num years: 10 ; drift: 0.1 ; volatility: 0.02 ; dt: 0.002740
value
130
125
120
115
110
105
100
95
90
85
80
75
70
65
60
55
50
45
0
250
500
750
1,000
1,250
1,500
1,750
2,000
2,250
2,500
2,750
3,000
3,250
3,500
3,750
time
GBM


GBM
num years: 1 ; drift: 0.1 ; volatility: 0.2 ; dt: 0.002740
value
time
47.0
46.5
46.0
45.5
45.0
44.5
44.0
43.5
43.0
42.5
42.0
41.5
41.0
40.5
40.0
39.5
39.0
38.5
38.0
37.5
0
25
50
75
100
125
150
175
200
225
250
275
300
325
350
375
GBM


GBM
num years: 1 ; drift: 0.1 ; volatility: 0.2 ; dt: 0.002740
value
time
GBM


GBM
num years: 10 ; drift: 0.1 ; volatility: 0.2 ; dt: 0.002740
value
time
135
130
125
120
115
110
105
100
95
90
85
80
75
70
65
60
55
50
45
40
35
0
250
500
750
1,000
1,250
1,500
1,750
2,000
2,250
2,500
2,750
3,000
3,250
3,500
3,750
GBM


GBM
num years: 10 ; drift: 0.1 ; volatility: 0.2 ; dt: 0.002740
value
120
115
110
105
100
95
90
85
80
75
70
65
60
55
50
45
40
35
0
250
500
750
1,000
1,250
1,500
1,750
2,000
2,250
2,500
2,750
3,000
3,250
3,500
3,750
time
GBM


GBM
num years: 10 ; drift: 0.1 ; volatility: 0.2 ; dt: 0.002740
value
60.0
57.5
55.0
52.5
50.0
47.5
45.0
42.5
40.0
37.5
35.0
32.5
30.0
0
250
500
750
1,000
1,250
1,500
1,750
2,000
2,250
2,500
2,750
3,000
3,250
3,500
3,750
time
GBM


GBM
num years: 10 ; drift: 0.1 ; volatility: 0.2 ; dt: 0.002740
value
100
95
90
85
80
75
70
65
60
55
50
45
40
35
0
250
500
750
1,000
1,250
1,500
1,750
2,000
2,250
2,500
2,750
3,000
3,250
3,500
3,750
time
GBM