Calibrated Jump Diffusion vs Geometric Brownian Motion: Forecasting Crude Spot Prices

by Alexander Torres, Rutgers University MSMF

April 2018

1 Abstract

Forecasting is an important tool in investment. For this reason, we need well calibrated models to provide accurate bounds on price movement. The sensitivity of commodity prices make jump diffusion models a good choice to forecast crude spots. I find however that the jump frequency λ must be chosen conservatively to avoid what I call the *stunted jump paradox*. This occurs when the distribution of the price process collapses at its starting point. In the conclusion, I provide some possible guidelines to choosing an optimal jump frequency.

2 The General Model

Consider the spot price for crude oil, $O(t)$, over the horizon $t_0 \leq t_n \leq t_N = T$ with drift μ and volatility σ historically over the period $t \in [t_0, t_n)$. Here, t_n is today, t_N is some time in the future, and t_0 is an arbitrary starting time before today. I define $J(u)$ to be a compound Poisson process given by

$$J(u) = \sum_{i=1}^{N(u)} Y(u) \Rightarrow \Delta J(u) = J(u) - J(u-) = Y(u)$$

where $N(u)$ is a Poisson process with intensity λ and $Y(u)$ are i.i.d. random variables. Let $\{W(u)\}$, $t_n \leq u \leq t_N$ be a Brownian motion.

Then I define the jump diffusion model for crude spot prices as,

$$dO(t) = \mu(t)O(t)dt + \sigma(t)O(t)dW(t) + \phi(t, O(t))\Delta J(t), t_n \leq t \leq t_N \tag{1}$$

where the drift and volatility can depend on time. The GBM model is the same, except for the last term which is excluded.

3 Calibration

3.1 Geometric Brownian Motion (GBM)

Consider a general price process $S(t)$, $t_0 \leq t \leq t_n$.For the GBM model there are two parameters $\mu(t) = \mu$ and $\sigma(t) = \sigma$ which are assumed to be constant in the model. To derive the estimator for the parameters, assume a price process moves according the GBM given by

$$dS(t) = \mu S(t)dt + \sigma dW(t)$$

Integrating using Ito calculus we get,

$$S(t) = S(t_0)e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)}$$

Note that Brownian motion has independent increments distributed according to a standard normal random variable with mean 0 and variance t. In other words, $W(t_i)-W(t_{i-1}) \sim N(0, t_i-t_{i-1})$. Suppose $\tau = t_i - t_{i-1}$ for all i. Then by observing log returns we can find an estimator for the drift and volatility of the GBM model.

Proof Let $(\cdot)(t_n) = (\cdot)_n$ for any time dependent function. Then

$$L := ln(\frac{S_n}{S_{n-1}}) = (\mu - \frac{1}{2}\sigma^2)\tau + \sigma(W_n - W_{n-1}) \sim N((\mu - \frac{1}{2}\sigma^2)\tau, \sigma^2\tau)$$

$$\Rightarrow E[L] = (\mu - \frac{1}{2}\sigma^2)\tau, Var(L) = \sigma^2\tau$$

Using sample mean and variance [1] , we obtain unbiased and asymptotically consistent estimators for the drift and the volatility:

$$\hat{\sigma} = \frac{std(L)}{\sqrt{\tau}}, \hat{\mu} = \frac{E[L]}{\tau} + \frac{1}{2}\hat{\sigma}^2 \tag{2}$$

where $std(L) := \sqrt{Var(L)}$ is the standard deviation of the data.

3.2 Merton Jump Diffusion Model

In the jump diffusion model, we need to estimate the impact of the jump, $\phi(t, S(t))$. Particularly, I use a Merton jump diffusion process where the price process takes the form

$$\frac{dS(t)}{S(t)} = (\alpha - \lambda k)dt + \sigma dW(t) + (y(t) - 1)\Delta N(t) \tag{3}$$

where $y_t \sim LogNormal(\alpha, \delta^2)$.

Proof Note that in this model we need to derive estimates for μ, k, and σ. The *sigma* here is not the same as that of the GBM, but its derivation is analogous. To justify equation (3), we follow the proof given by Matsuda [2] which can be referenced for further detail. First assume relative jump size is distributed according to a lognormal distribution with mean μ and variance δ^2. In other words,

$$\frac{dS_t}{S_t} = \frac{y_t S_t - S_t}{S_t} = y_t - 1$$

where $ln(y_t) \sim i.i.dN(\mu, \delta^2)$. This implies that we will be defining $\phi(t, S(t)) = \frac{dS_t}{S_t} = y_t - 1 = \phi(y_t)$. In this way we define $k := E[y_t - 1]$, and conclude $E[(y_t - 1)dN(t)] = k\lambda dt$. We then see that the expected return of (3) is

$$E[\frac{dS_t}{S_t}] = (\alpha - \lambda k)dt + 0 + \lambda k dt = \alpha dt$$

where α is the instantaneous expected return, the desired result.

More specifically, a lognormal distribution has mean $\mu := e^{a+0.5b^2}$ and variance $\delta^2 := e^{2a+b^2}(e^{b^2} - 1)$ for some constant a and b. Thankfully, the estimation of the parameters a and b are not needed here. Instead, we define $\hat{y}_t := \frac{dS_t}{S_t} + 1$, the relative price change, and find

$$\hat{\mu} := E[ln(\hat{y}_t)], \hat{\delta^2} := Var[ln(\hat{y}_t)], \hat{k} := E[\hat{y}_t - 1] \tag{4}$$

[1] Sample mean is unbiased by definition, but sample variance is only asymptotically unbiased. i.e. When we assume variance is constant,

$$E(var_{sample}(Y)) = \frac{1}{(n-1)}\sum_{i=1}^{n} E[(y_i - \bar{y})^2]$$

$$= \frac{n}{n-1}\sum_{i=1}^{n}\sigma_i^2 = \frac{n}{n-1}\sigma^2$$

Although this estimate is biased, it is a good estimator for reasonable finite sample sizes.

Substituting these estimates into equation (3),

$$d\hat{S}(t) := \hat{S}(t)(\hat{\alpha} - \lambda\hat{k})dt + \hat{\sigma}dW(t) + (y_t - 1)\Delta N(t),$$

where $y_t \sim Lognormal(\hat{\mu}, \hat{\delta}^2)$. Taking the exponential of $d[ln(S(t))]$ using Ito calculus, we get the explicit form for the price process:

$$\hat{S}(t_n) := \hat{S}(t_0)e^{(\alpha - 0.5\hat{\sigma}^2 - \lambda\hat{k})\tau + \sigma W(t) + \sum_{i=1}^{N(t_n)} ln(\hat{y}_t)}, \tag{5}$$

Now we may again use log returns to derive estimates for the α and σ paramaters by conditioning on the known data $S(t_0)$,...,$S(t_n)$ as follows:

$$L := ln(\frac{S_n}{S_{n-1}}) = (\alpha - \frac{1}{2}\sigma^2 - \lambda\hat{k})\tau + \sigma(W_n - W_{n-1}) + (N(t_n) - N(t_{n-1}))ln(\hat{y}_t)$$

$$E(L) = (\alpha - \frac{1}{2}\sigma^2 - \lambda\hat{k})\tau + \lambda\hat{\mu}\tau$$

$$Var(L) = \sigma^2\tau + \lambda\hat{\delta}^2\tau$$

$$\Rightarrow \hat{\sigma} := \sqrt{\frac{Var(L) - \lambda\hat{\delta}^2\tau}{\tau}}, \hat{\alpha} := \frac{E(L_t) - \lambda\hat{\mu}\tau}{\tau} + \frac{\hat{\sigma}^2}{2} + \lambda\hat{k}\tau$$

4 Hypothesis and Empirical Analysis

The question I am addressing in this paper is whether or not the jump diffusion model captures the price movement of crude oil better than a geometric Brownian motion. Intuitively, I would expect the jump model to perform better. Oil is a sensitive commodity. The announcement of a closure/opening of a pipeline, geopolitics with big oil producers, supply shocks or booms—all of these things and more cause the spot price to jump. Geometric Brownian motions have continuous paths and cannot capture their effects. Later we will see how this influences parameter estimation for GMB. To formally test which model performs better, I use the difference of mean squared error of each model to form a hypothesis:

First, let v, v_G and v_M be the true data and realizations of the GBM and a MJD process for the crude spot respectively. Then $v_G = \{v_G(t_n), ..., v_G(t_N)\}(\omega)$ is a sample path for the forecasted spot price $N - n$ periods into the future using the GBM model. We interpret v_M the same way for the Merton jump diffusion model. In this way define $MSEG := \{v - v_G(\omega_1), v - v_G(\omega_2), ..., v - v_G(\omega_l)\}$ as the realized mean squared error for l paths of the GBM model. Define $MSEM$ in the same way for the Merton jump diffusion model. I can now state the hypothesis as

$$H_0 : MSEG = MSEM, \forall\omega_i, i = 1, ..., l$$

$$H_1 : MSEG > MSEM, \forall\omega_i, i = 1, ..., l$$

Because testing each individual case would induce sequential test bias, I rely on the sample mean asymptotic distribution and reformulate my hypothesis:

$$H_0 : D := \sum_{i=1}^{l} MSEG(\omega_i) - MSEM(\omega_i) = 0$$

$$H_1 : D := \sum_{i=1}^{l} MSEG(\omega_i) - MSEM(\omega_i) > 0$$

where the sample mean has an asymptotic normal distribution[2] with mean $\mu_{\bar{X}} = \mu$, the population mean, and $\sigma^2_{\bar{X}} = \sigma^2/l$, the scaled population variance. Using $\alpha = 0.05$, we have the test statistic

$$t := \frac{\sqrt{l}(\hat{D} - \mu)}{std(D)}$$

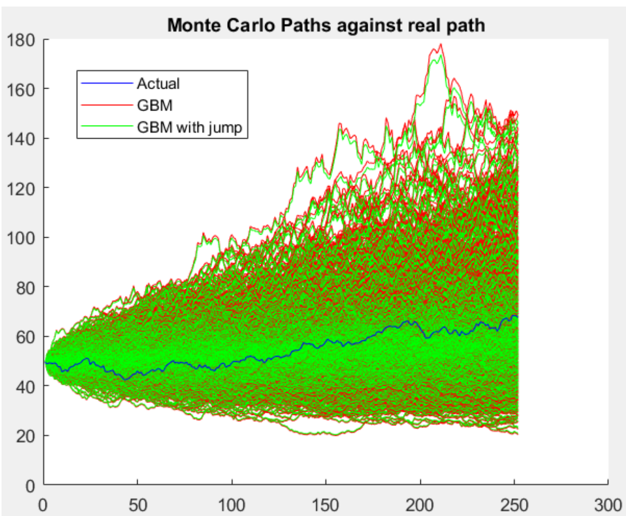
[2] I am using a Monte Carlo simulation with 1000 paths which I think is sufficiently large to use the asymptotic distribution.

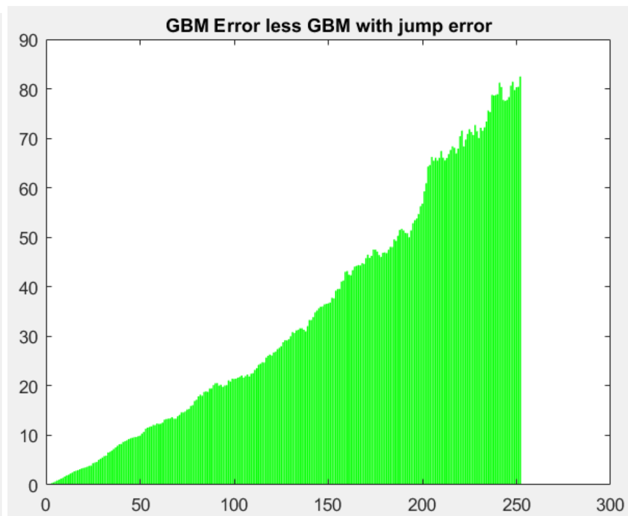
where $\hat{D}$ is the sample average of the difference of mean squared errors and $\mu = 0$ is the hypothesized population mean. Since for a standard normal random variable X, $P(X <= 1.943) = 0.95$, if $t < 1.943$ then we may not reject the null. Otherwise, we reject the null in favor of the jump diffusion model performing better (i.e. having a lower average mean squared error).

5 Results

For my research I used the Cushing WTI spot price (dollars/barrel) [1] from dates t_0 =December 5th, 2016 to t_n =August 4th, 2017 as my testing data. Then, a Monte Carlo simulation is performed to generate paths for the two models and compare the results against the data from t_n to t_N =April 23, 2018. In this way I have chosen to make the time step of $\tau = \frac{1}{252}$ with one year of data used to calibrate the models, and one year of data to test them. The following graphs and tables are the Monte Carlo simulation results over 1000 paths for a series of jump frequencies, and the calculated parameter calibrations.

5.1 10 jumps





Historical sample statistics

Z	EZ	VarZ
"log Returns (LR)"	0.000604908792943367	0.000450440853713653
"Relative Returns + 1 (y)"	0.000604908792943365	0.000450440853713654

Merton Jump Diffusion calibration, jumps = 10

lambda	mu_y	delta_y	k
0.0396825396825397	0.000604908792943367	0.0212235919135676	0.000829946695876774

alpha	sigma
0.20092420558434	0.330161606790832

Geometric Brownian Motion calibration

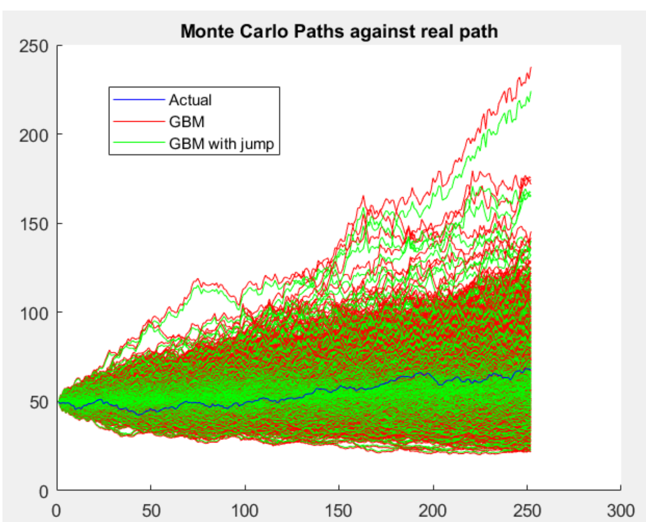
mu_LR	sigma_LR
0.209192563389648	0.336914076784929

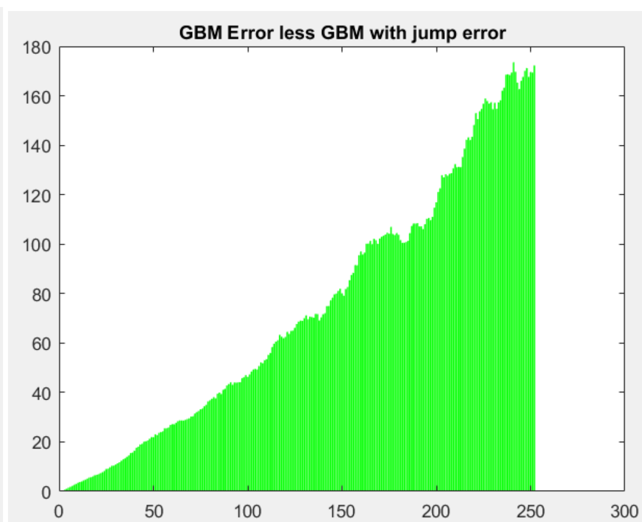
Parameter values based of estimates

model	drift	volatility	mean_MSE	var_MSE
"MJD"	0.146387927892294	0.330161606790832	814.718791287319	284199.191829536
"GBM"	0.152437015821728	0.336914076784929	848.601153428976	310346.77551664

paths	mean_MSED	std_MSED	t_stat	p_value
1000	8.5383552596976	16.7184994445965	16.1501635848363	9.18071374824098e-58

5.2 20 jumps





Historical sample statistics

Z	EZ	VarZ
"log Returns (LR)"	0.000604908792943367	0.000450440853713653
"Relative Returns + 1 (y)"	0.000604908792943365	0.000450440853713654

Merton Jump Diffusion calibration, jumps = 20

lambda	mu_y	delta_y	k
0.0793650793650794	0.000604908792943367	0.0212235919135676	0.000829946695876774

alpha	sigma
0.192655847779032	0.323268121010358

Geometric Brownian Motion calibration

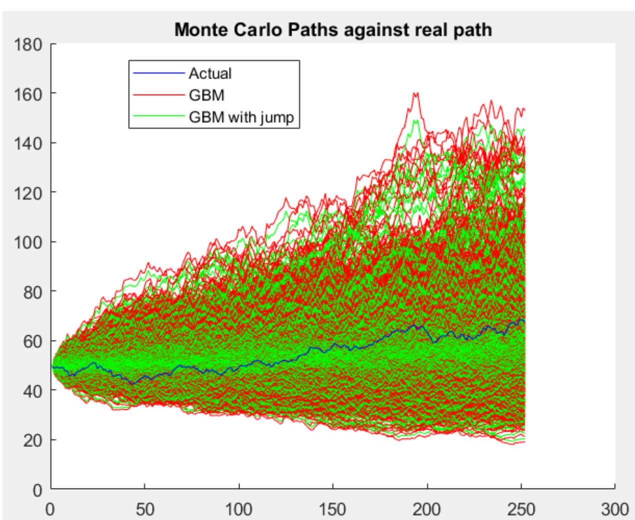
mu_LR	sigma_LR
0.209192563389648	0.336914076784929

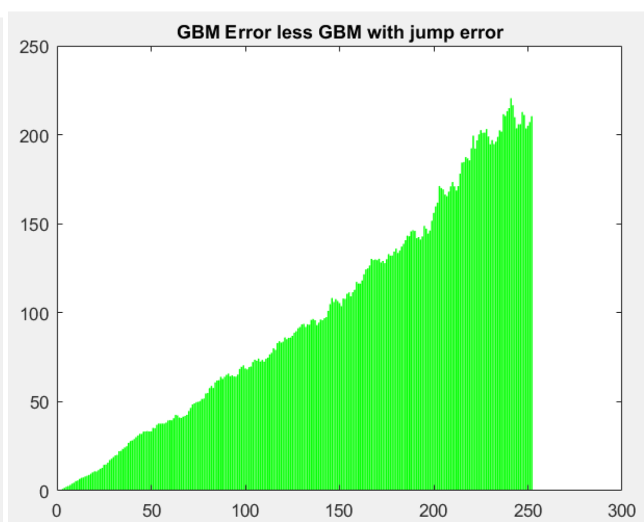
Parameter values based of estimates

model	drift	volatility	mean_MSE	var_MSE
"MJD"	0.140338839962861	0.323268121010358	820.237816223512	276820.249724178
"GBM"	0.152437015821728	0.336914076784929	893.313673725595	332743.125959084

paths	mean_MSED	std_MSED	t_stat	p_value
1000	18.4151160905248	48.3143145891318	12.0530966273034	1.13339243378161e-32

5.3 30 jumps





Historical sample statistics

Z	EZ	VarZ
"log Returns (LR)"	0.000604908792943367	0.000450440853713653
"Relative Returns + 1 (y)"	0.000604908792943365	0.000450440853713654

Merton Jump Diffusion calibration, jumps = 30

lambda	mu_y	delta_y	k
0.119047619047619	0.000604908792943367	0.0212235919135676	0.000829946695876774

alpha	sigma
0.184387489973723	0.316224397421248

Geometric Brownian Motion calibration

mu_LR	sigma_LR
0.209192563389648	0.336914076784929

Parameter values based of estimates

model	drift	volatility	mean_MSE	var_MSE
"MJD"	0.134289752033427	0.316224397421248	732.391902548881	202904.935942591
"GBM"	0.152437015821728	0.336914076784929	829.008931035027	262885.850160757

paths	mean_MSED	std_MSED	t_stat	p_value
1000	24.3474911785086	47.8068049341502	16.1051397475729	1.89771137950212e-57

6 Analysis

It is clear that the Merton Jump Diffusion model is superior to a Geometric Brownian Motion model by observing the p-values. For all jump frequencies, we strongly reject the null that the GBM is at least as good as the MJD model. This can be seen upon inspection of the graphics which show a concentration of MJD paths (denoted as 'GBM with jump' in the key) around the true blue path. When looking at the GBM paths, we can see they are more frequently higher than the MJD paths. As I said before, the calibration of the drift and volatility parameters is affected by not included a jump. To make up for the missing information, these parameters are scaled up (see tables) for the GBM model, causing it to yield exaggerated bounds on the spot price. The stronger drift will dominate the Brownian motion and cause the price to be inaccurate. The inclusion of the jump term mitigates this effect, with a higher jump intensity λ causing a weaker drift and volatility.

The question now is how to choose the optimal number of jumps per period (1 year in my case or 252 time steps). In short, my opinion would be to choose the number of jumps conservatively. In this case I would suggest 10 to be a good option. The reason for this is two-fold: Looking at the sample paths, we notice that the more jumps that occur, the more concentrated the MJD paths (green) become around the actual (blue) path of the crude spot. This suggests that the optimal jump frequency would be higher, say 30 paths. However, we must also take into consideration the bar chart which represent the average difference in squared error at each point in time between the GBM and MJD models. Although the distance between the two grows in proportion to jump frequency, the shape of the bar chart remains intact across the board. This suggests that forecasts are increasingly bad the further out in time you go. With this in mind, the forecast is not mean to provide a tight bound on the spot price for far away projections. Therefore, I would choose

10 to be an appropriate amount of jumps.

One can also argue this is a reasonable caveat because the inclusion of a large number of jumps tends to eliminate the drift and driving Brownian motion. As a result, the probability density starts to collapse around the starting point. Since we include jumps to induce sporatic movement in the price process, I term refer to this phenomena as the *stunted jump paradox*.

Theorem *Stunted Jump Paradox* For a Merton Jump Diffusion model,

$$lim_{\lambda\to\tau^{-2}} P(S(t_n) = S(t_0); \lambda) = 1$$

Proof Take the MJD price process given by equation (5). Notice that the drift component becomes

$$drift := \hat{\alpha} - 0.5\hat{\sigma}^2 - \lambda\hat{k} = \frac{E(L_t) - \lambda\hat{\mu}\tau}{\tau}$$
$$= \frac{1}{\tau}\{E(ln(\frac{S_t}{S_{t-1}})) - \lambda E(ln(\frac{dS_t}{S_t} + 1))\tau\}$$
$$\leq \frac{1}{\tau}\{E(ln(\frac{S_t}{S_{t-1}})) - \lambda E(ln(e^{\frac{dS_t}{S_t}}))\tau\}$$
$$= \frac{1}{\tau}\{E(ln(\frac{S_t}{S_{t-1}})) - \lambda\alpha\tau^2\}$$

Since α is the instantaneous rate of return [3] , the last line may be written as

$$E(ln(\frac{S_t}{S_{t-1}}))(\frac{1}{\tau} - \lambda\tau)$$

This implies

$$lim_{\lambda\to\tau^{-2}} drift \leq lim_{\lambda\to\tau^{-2}} E(ln(\frac{S_t}{S_{t-1}}))(\frac{1}{\tau} - \lambda\tau) \to 0$$

Since we are $ln(x)$ is bounded a bounded function, we know $ln(S_t/S_{t-1})$ is bounded and the limit makes sense. Similarly, for the volatility estimate, we have

$$\hat{\sigma^2} = \frac{1}{\tau}(Var(L) - \lambda\hat{\delta}^2\tau)$$
$$\leq Var(ln(\frac{S_t}{S_{t-1}}))(\frac{1}{\tau} - \lambda\tau)$$

This also tends to zero as $\lambda \to \tau^{-2}$. Finally, because the drift and volatility are eliminated as λ approaches τ^{-2}, the driving source of randomness is the compound poisson process. However, $\sum_i = 1^{N(t_n)} ln(y_{t_n}) = 0$ [2]. Then we have

$$lim_{\lambda\to\tau^{-2}} \hat{S}(t_n) := S(t_0)e^{(\alpha-0.5\hat{\sigma}^2-\lambda\hat{k})\tau+\sigma W(t)+\sum_{i=1}^{N(t_n)} ln(\hat{y}_t)}$$
$$= S(t_0)exp(0dt + 0dW(t) + 0) = S(t_0)$$

QED

In our case, $\tau^{-2} = 63504$, an unrealistically large intensity for something that is supposed to represent 'sudden' shocks to the price process. Notice also that the GBM drift is greater than 0. As we include jumps, the drift is pushed towards the asymptotic upper bound of 0. If we had started with a negative drift, the drift would also be pushed to 0 as λ increases. As such, the process becomes more and more akin to white noise as the jump becomes the driving random process, which holds no value for forecasting.

7 Conclusion

The Merton Jump Diffusion model outperforms the Geometric Brownian motion model for predicted crude oil spot prices when correctly calibrated. The optimal amount of jumps per period should be chosen carefully, taking into consideration market expectations and bearing in mind the stunted jump paradox. Too many jumps will lead to a loss of information as the distribution of the paths converge to a single point. Too little jumps will not capture the effects of discontinuities in the price process. One way to gauge λ based off the graphs is to pick a reasonable frequency, simulate the paths, and see if the actual data falls within an area of high probability. If so, that is a good λ to start testing with.

8 Work Cited

[1] Eia.gov. (2018). Cushing, OK WTI Spot Price FOB (Dollars per Barrel). [online] Available at: https://www.eia.gov/dnav/pet/hist/RWTCD.htm [Accessed 23 Apr. 2018].
[2] Matsuda, Kazuhisa. "Introduction to Merton Jump Diffusion Model", *Department of Economics, The City University of New York*. Dec 2004
[3] Wikipedia contributors. "Continuously compounded nominal and real returns." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 15 Mar. 2017. Web. 30 Apr. 2018.