UNIVERSITY OF BIRMINGHAM

Modification of Heston Model and Correlation Estimation

Hanghang Zhang
1957916
Supervisor: Sergey Sergeev

Financial Engineering 2018-2019

1 Introduce

Nowadays, people found the logs return of assets are generally not distributed normally, the distribution of the log returns to a specific asset might be asymmetric, leptokurtic and fat-tailed. This observation against the assumption that used by many financial models. During our research, we found the log returns of most of stocks' prices do follow the normal distribution, but mean and variance of this normal distribution varies over time. The historical values of such stochastic mean and variance can be estimated from the historical data of stocks' price directly, and their behaviour can be described by a slightly modified Heston model. Though the whole paper, we are going to use the historical data of stocks' adjusted price from Bank of American and 3M company as example to generalize a method to modify the Heston model that could fit the data, and then introduce a way to estimate the stochastic correlation between log returns in section 3. Additionally, we developed a general method to test if a time series is a realization of certain stochastic process and use it to test our conclusions.

2 Validation and Modification of Heston Model

The Heston model is a stochastic volatility model that used to describe the process of an asset's price with the assumption that its volatility changes over time. This model can be described by the following stochastic differential equations:

$$dS_t = \mu S_t dt + \sqrt{\nu_t} S_t dW_t^S \tag{2.1}$$

$$d\nu_t = \kappa(\theta - \nu_t)dt + \sqrt{\xi \nu_t} dW_t^{\nu} \tag{2.2}$$

Where W_t^S and W_t^{ν} are two Wiener process with correlation ρ. In the first equation S_t represents the asset's price at time t, μ is the average rate of return and ν_t is the instantaneous volatility. In the second Equation (2.2), parameters κ, θ, ξ are speed that ν_t revert to its long term average, the long term average of ν_t and the volatility of ν_t.

In the Heston model, the historical value of every parameter except S_t does not have a specific definition, that is, when the historical data of S_t is given, there are many different estimators to estimate the historical value of those parameters. For example, possible estimators to measure the historical instantaneous volatility could be moving standard deviation of S_t, exponential moving standard deviation of $\log \frac{S_t}{S_{t-\Delta t}}$, standard deviation of $S_t - S_{t-\Delta t}$. When the estimators for all stochastic parameters like ν_t in the model are selected, estimators for other constant parameters can be chosen with numerical methods like least squares method.

Suppose a set of estimators for each parameter is already given along with the historical data of asset's price in form of time series, the fist problem this section will discuss is: could the Heston model with this set of estimators describe the behaviour of the historical data? We call the process of testing the null hypothesis "H_0 : historical data is a realization of Heston model using given set of estimators." a validation.

To illustrate the idea of validation, let's consider the simple case $\nu_t = \nu$ is a constant for all time t. The equation (2.1) becomes a Geometric Brownian motion, and such stochastic differential equation has solution:

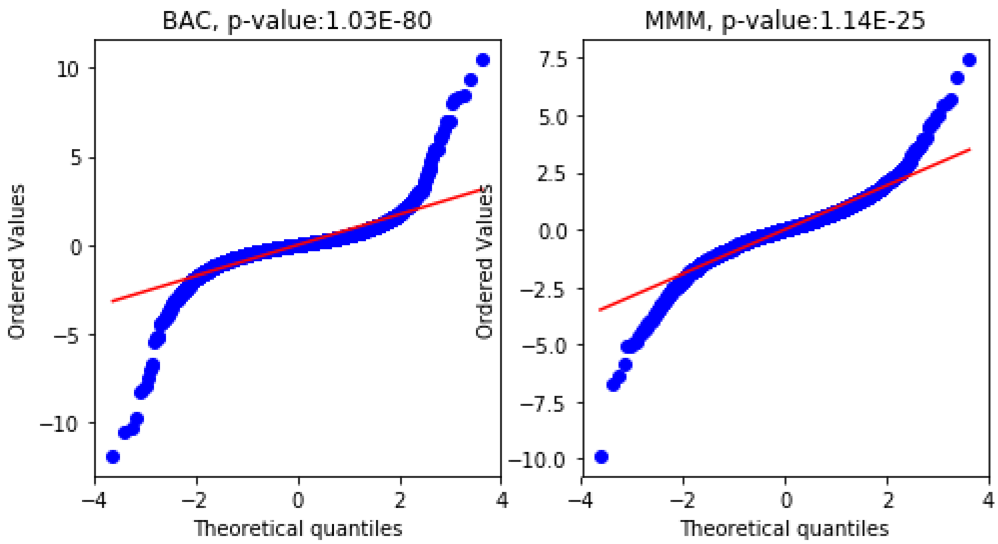
$$S_{t+u} = S_t \exp\Big((\mu - \frac{\nu}{2})t + \sqrt{\nu}(W^S_{t+u} - W^S_t)\Big) \tag{2.3}$$

Since Wiener process has the property of Gaussian increment, $W^S_{t+u} - W^S_t$ is normally distributed with mean 0 and variance u. Let

$$O_t = \frac{\log(\frac{S_t}{S_{t-u}}) - (\mu - \frac{\nu}{2})u}{\sqrt{\nu}} \tag{2.4}$$

If the null hypothesis is true for constant estimators for μ, ν , with significant amount of historical data in form of time series that sampled with a constant frequency, we should be able to observe O_t follows a normal distribution.

If μ and ν are constant, we can see $\log(\frac{S_t}{S_{t-u}})$ in(2.4) must be distributed normally. So we can simply use the Kolmogorov–Smirnov test to compare the distribution $\log(\frac{S_t}{S_{t-u}})$ against the standard normal distribution to test our null hypothesis. The following graphs shows such distribution comparison and its p-value from KS-test using the stock price data of Bank of American and 3M company:



The p-value suggest to reject the null hypothesis at a significant level of 1%. Thus we know there is no way to describe the behaviour of our data with (2.1) that using constant μ and ν
So now assume that μ_t and $\nu_t$in (2.3) are stochastic, that is

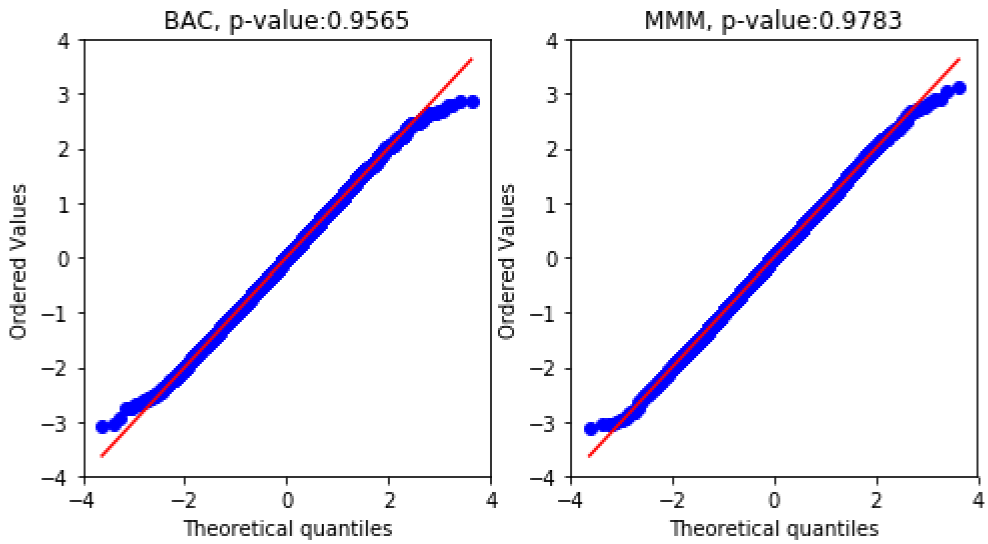
$$S_t = S_{t-\Delta t}\exp((\mu_t - \frac{\nu_t}{2})\Delta t + \sqrt{\nu_t}(W_t^S - W_{t-\Delta t}^S)) \tag{2.5}$$

this time we select the moving average/variance with period m of $\log(\frac{S_t}{S_{t-u}})$ from historical data as estimators for historical values of μ_t and ν_t, namely

$$\mu_t = \frac{1}{m}\sum_{i=t-m-\Delta t}^{t-\Delta t}\log(\frac{S_i}{S_{i-\Delta t}}) = \frac{1}{m}\log(\frac{S_{t-\Delta t}}{S_{t-m-\Delta t}}) \tag{2.6}$$

$$\nu_t = \frac{1}{m-1}\sum_{i=t-m-u}^{t-u}(\log(\frac{S_i}{S_{i-u}}) - \mu_t)^2 \tag{2.7}$$

Substitute μ, ν in (2.4) with above estimators for μ_t, ν_t and set $m = 11$ trading days, we do the KS-test that compares the distribution of normalized O_t against standard normal distribution to get the following results



As we can see, the p-value is above 0.95 for both data from BAC and data from MMM, we can not reject the null hypothesis even at a significant level of 95% this time. Thus we modify the geometric Brownian motion part of the Heston model(2.1) into

$$S_{t+\Delta t} = S_t \exp((\mu_t - \frac{\nu_t}{2})\Delta t + a\sqrt{\nu_t}(\Delta W_t^S) + b\sqrt{\nu_t}) \tag{2.8}$$

Additional constant parameters a, b are added because in the previous step, KS-test was applied on the normalized O_t, a and b are parameters that used to normalize it, they can be obtained from historical data easily.
Now let's validate estimators in the dynamic for instantaneous volatility, (2.2) is called a CIR process and it's closed form solution is:

$$\nu_{t+\Delta t} = \frac{\xi(1 - \exp(-\kappa\Delta t))}{4\kappa} Y_t \tag{2.9}$$

where Y_t follows non central χ^2 distribution

$$Y_t \sim \chi^2(\frac{4\kappa\theta}{\xi}, \frac{4\kappa\nu_t \exp(-\kappa\Delta t)}{\xi(1 - \exp(-\kappa\Delta t))}) \tag{2.10}$$

The validation of the CIR part of Heston model is not that straightforward as the validation of the geometric Brownian motion part, because a non central χ^2 distribution $\chi^2(\delta, \lambda)$ can not be obtained by linearly transform another

non central χ^2 distribution.
Suppose there a stochastic process X_t, we know that at time t, X_t follows certain distribution whose cumulative density function F_t is known. For a given time series data D_t, we can test the hypothesis that D_t is realized by stochastic process X_t with the following mythology:
Let D_t be a realization of X_t, then for given $p \in [0,1]$, we assert

$$\lim_{N\to\infty} \frac{\sum_{t=t_0}^{t_0+N\Delta t} H[F_t^{-1}(p) - D_t]}{N} = p \tag{2.11}$$

Where H is the Heaviside function defined as

$$H(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases} \tag{2.12}$$

Thus

$$\lim_{N\to\infty} \int_0^1 \left(\frac{\sum_{t=t_0}^{t_0+N\Delta t} H[F_t^{-1}(p) - D_t]}{N} - p\right)^2 dp = 0 \tag{2.13}$$

For given $\Delta p, \Delta t, N$, let

$$B = \sum_{i=0}^{\lfloor \frac{1}{\Delta p} \rfloor} \left(\frac{1}{N} \sum_{t=t_0}^{t_0+N\Delta t} H[F_t^{-1}(i\Delta p) - D_t] - i\Delta p\right)^2 \tag{2.14}$$

Be the statistic, select a huge number M, use $\Delta t, \Delta p$ to simulate N steps of the stochastic process X_t for M times to get M realizations $R_1, R_2, ..., R_M$ and their B-statistics $\hat{B}_1, \hat{B}_2, ...\hat{B_M}$.
Sort $B, \hat{B}_1, \hat{B}_2..., \hat{B_M}$ in decreasing order, then find the rank r of B, $\frac{r}{M} \times 100\%$ will be the empirical p-value for this test.
With this method, we test if the Y_t we observed from historical data follows the time dependent distribution(2.10) to do the validation.
Since restriction $2\kappa\theta > \xi$ is needed to ensure positive ν_t, we rewrite κ as $\kappa = k\frac{\xi}{2\theta}$ or $\kappa = \frac{\xi}{2\theta} + \epsilon$ with restriction $\epsilon > 0$ or $k > 1$, usually, a κ that closes to $\frac{\xi}{2\theta}$ will result in a smaller B. As θ means the long term average of ν_t, we can simply set θ to be the average of all observed historical ν_t, Through our experiments, we noticed that how ξ is estimated from historical data effects the result of B-statistic most significantly. For example, for some assets, select ξ to be the variance of historical ν_t will minimize the B-statistic while other assets' B-statistic might be minimized with ξ equals the variance of

$\log(\frac{\nu_t}{\nu_{t-\Delta t}})$ or the standard deviation of ν_t. Use a unique estimator for all assets damages the performance of model badly.
To find the best value of ξ_t, we can use

$$L(\xi|p,t) = (\frac{1}{N}\sum_{t=t_0}^{t_0+N\Delta t} H[F_t^{-1}(\xi|p) - D_t] - p)^2 \tag{2.15}$$

as loss function and apply gradient descent method to minimize the B-statistic $B(\xi) = \sum_p L(\xi|p)$. Or more aggressively, we can apply gradient descent method to minimize ξ, θ and k or ϵ at the same time.
To apply gradient descent, we need to figure out the derivative of B respect to ξ.

$$\frac{\partial}{\partial \xi}B = \sum_p \frac{\partial}{\partial \xi}L(\xi|p,t) \tag{2.16}$$

$$= \sum_p \Big[2(\frac{1}{N}\sum_t H[F_t^{-1}(\xi|p,t) - D_i] - p)\frac{1}{N}\sum_t \frac{\partial}{\partial \xi}(H[F_t^{-1}(\xi|p,t) - D_t])\Big]$$

$$= \sum_p \Big[2(\frac{1}{N}\sum_t H[F_t^{-1}(\xi|p) - D_i] - p)\frac{1}{N}\sum_t(\delta[F_t^{-1}(\xi|p) - D_t]\frac{\partial}{\partial \xi}(F_t^{-1}(\xi|p|)))\Big]$$

$\delta(x)$ here is the Dirac delta function, in application, we can replace δ with $\frac{\exp(-(x/a)^2)}{|a|\sqrt{\pi}}$ for a very small a. Since in our case, F_t is the cumulative density function of non central χ^2 distribution, by choosing $\kappa = k\frac{\xi}{2\theta}$, the degree of freedom of such distribution at any time t is $2k$, thus the cumulative density function is

$$F_t(\xi|x) = 1 - Q_k(\sqrt{\frac{4\kappa\nu_t\exp(-\kappa\Delta t)}{\xi(1-\exp(-\kappa\Delta t))}}, \sqrt{x}) \tag{2.17}$$

$$= 1 - Q_k(\xi^{-\frac{1}{1}}C_t, \sqrt{x})$$

where Q_k is the Marcum Q function and $C_t = \sqrt{\frac{4\kappa\nu_t\exp(-\kappa\Delta t)}{(1-\exp(-\kappa\Delta t))}}$. The derivative of $Q_m(a,b)$ respect to a is

$$\frac{\partial}{\partial a}Q_m(a,b) = a(Q_{m+1}(a,b) - Q_m(a,b)) \tag{2.18}$$

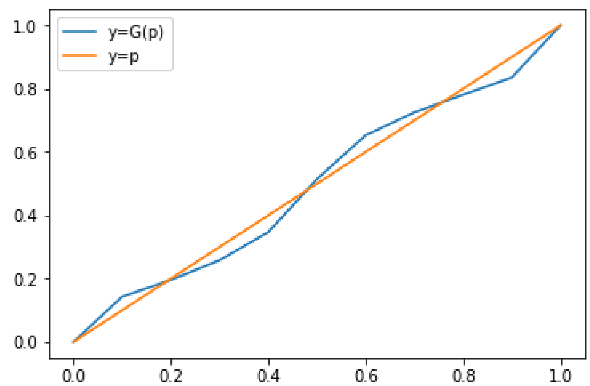
Therefore we have

$$\frac{\partial}{\partial \xi}F(\xi|x) = \frac{1}{2}C_t^2\xi^{-2}(Q_{k+1}(C\xi^{-\frac{1}{2}}, \sqrt{x}) - Q_k(C\xi^{-\frac{1}{2}}, \sqrt{x})) \tag{2.19}$$

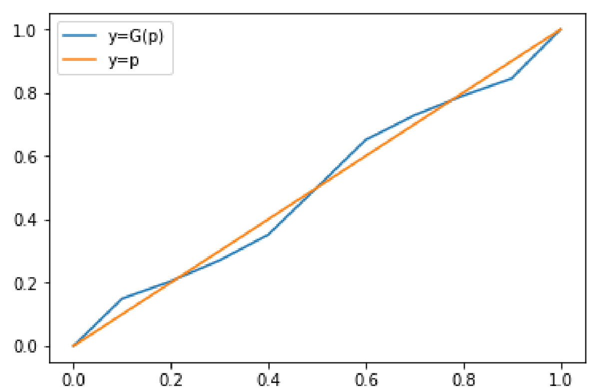
By the derivative rule of inverse functions, we have

$$\frac{\partial}{\partial \xi} F_t^{-1}(\xi|p) = \frac{1}{\frac{1}{2} C_t^2 \xi^{-2} (Q_{k+1}(C\xi^{-\frac{1}{2}}, \sqrt{F_t^{-1}(p|\xi)}) - Q_k(C\xi^{-\frac{1}{2}}, \sqrt{F_t^{-1}(p|\xi)}))} \tag{2.20}$$

Put (2.20) back to (2.16) will give the expression for the gradient of B. Similar approach could be used to do gradient decent on all parameters ξ, θ, k. Gradient descent method using such cost function can not guarantee a global optimal solution, but it can help us to find a relatively good one. With this method, choosing $\Delta p = 0.1$, we found $\xi = 3.773 \times 10^{-6}, \theta = 8.614 \times 10^{-4}$ and $k = 1.01$ will give a locally minimized $B = 1.459 \times 10^{-2}$ from historical data of Bank of American. The following graph shows the plot of $G(p) = \frac{1}{N} \sum_{t=t_0}^{t_0+N\Delta t} H[F_t^{-1}(p) - D_t]$ using above parameters:



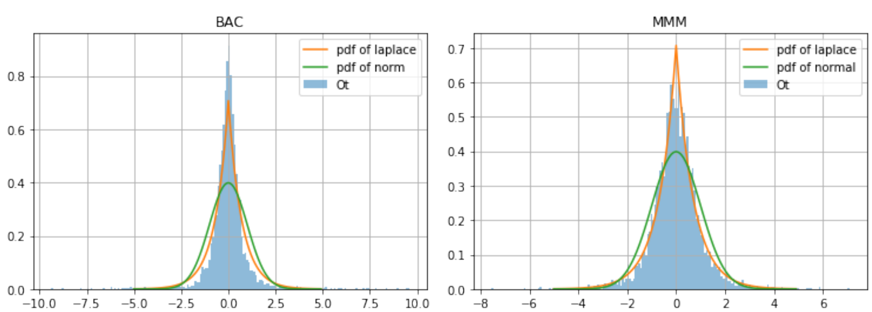
Use the same Δp and k, $\theta = 2.0126 \times 10^{-4}$ was observed from the historical data of 3M company, gradient decent gives $\xi = 1.754 \times 10^{-6}$ and the result B-statistic is 1.22×10^{-2}. The $G(p)$ plot for 3M company is shown below:



The B-statistics for constant θ, κ, ξ that estimated from historical data of BAC and 3M companies are still too large to achieve a significant level of 1% using the empirical p-value table. (To achieve a significant level of 1%, the B-statistics for both data from BAC and 3M should be less than 1.6×10^{-3}) As we didn't found any obvious stochastic estimators for θ, ξ that could improve the B-statistic significantly, we have to accept the hypothesis that these parameters are constant for now. As we can tell from above graphs, $G(p)$ is very close to p intuitively.
As mentioned earlier, the moving average of $\log \frac{S_t}{S_{t-\Delta t}}$ with a window size of m which is equivalent to $\frac{1}{m} \log \frac{S_t}{S_{t-m\Delta t}}$ is used as the estimator of μ_t, a stochastic process that describes the behaviour of μ_t is needed for further analysis. When $m = 1$, the estimator for μ_t becomes $\log \frac{S_t}{S_{t-\Delta t}}$ which is already modeled by (2.8), thus it makes sense to guess the stochastic process for μ_t is

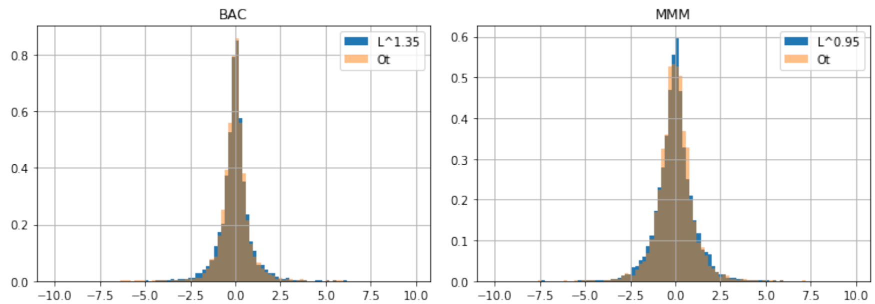
$$\mu_{t+\Delta t} = (r\mu_t - \frac{s^2}{2})\Delta t + s\Delta W_t^{\mu} \tag{2.21}$$

With r be the average of μ_t, s be the standard deviation of μ_t and $\Delta t = 1$ trading day, $m = 11$. Validating the normality of historical value of $O_t = \frac{1}{s}[\mu_{t+\Delta t} - (r\mu_t - \frac{s^2}{2})\Delta t]$ failed but the following distributions are observed:



It seems like the distribution of O_t is more likely to be a Laplace distribution. KS-tests for normality gives p-values 1.05×10^{-75}(BAC) and 3.48×10^{-17}(MMM) while testing if they are Laplace distribution gives p-values 4.58×10^{-22}(BAC) and 1.66×10^{-2}(MMM). KS test tells us that they are not Laplace distributions neither, but by their histogram, it is clear that they are related to Laplace distribution some how.
Let L_1, L_2 be two independent and identically distributed random variables that follows $L(0, \frac{\sqrt{2}}{2})$, the Laplace distribution with mean 0 and variance 1. For a real number ω, we denote the distribution of random variable $L_1(|L_2|)^{\omega}$

as $L^{1+\omega}$ distribution. Then we compare the distribution of O_t against distribution of samples from $L^{1+\omega}$ distribution:



In this example, ω used for BAC is 0.35 and for MMM is -0.05. KS 2-sample test gives p-value 0.102 and 8.313×10^{-2} respectively. Thus we can not reject the hypothesis that O_t follows a $L^{1+\omega}$ distribution at a significant level of 5%. Therefore, we modify (2.21) into

$$\mu_{t+\Delta t} = (r\mu_t - \frac{s^2}{2})\Delta t + s\Delta\mathcal{L}_t^{\omega} \tag{2.22}$$

Where $\mathcal{L}_t^{\omega}$ is a process with independent increment property and $\Delta\mathcal{L}_t^{\omega}$ follows a $L^{1+\omega}$ distribution.
So the Heston model is finally modified as

$$S_{t+\Delta t} = S_t \exp((\mu_t - \frac{\nu_t}{2})\Delta t + a\sqrt{\nu_t}(\Delta W_t) + b\sqrt{\nu_t}) \tag{2.23}$$

$$\nu_{t+\Delta t} = \frac{\xi(1-\exp(-\kappa\Delta t))}{4\kappa} Y_t \tag{2.24}$$

$$Y_t \sim \chi^2(\frac{4\kappa\theta}{\xi}, \frac{4\kappa\nu_t \exp(-\kappa\Delta t)}{\xi(1-\exp(-\kappa\Delta t))}) \tag{2.25}$$

$$\mu_{t+\Delta t} = (r\mu_t - \frac{s^2}{2})\Delta t + s\Delta\mathcal{L}_t^{\omega} \tag{2.26}$$

And the path simulation can be done with these equations directly.

3 Instantaneous Correlation

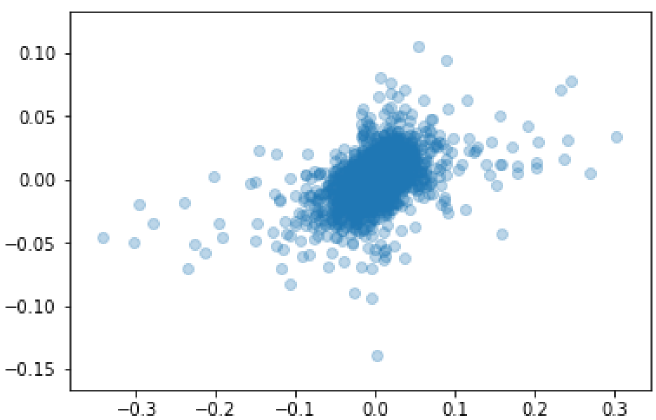
Suppose there are two assets $S^{(1)}, S^{(2)}$ whose historical price data can be described by Heston model using suitable estimators, in this section, we will focus on finding a estimator to estimate the historical value of correlation, next, validate if this estimator we established consists with modified Heston model.

Let $X_t = \begin{bmatrix} X_t^{(1)} \\ X_t^{(2)} \end{bmatrix}$ where $X_t^{(i)} = \log \frac{S_t^{(i)}}{S_{t-\Delta t}^{(i)}}$, from (2.23), we know that $X_t^{(i)} \sim N(a_t^{(i)}, b_t^{(i)})$ where

$$a_t^{(i)} = (\mu_t^{(i)} - \frac{\nu_t^{(i)}}{2})\Delta t + b^{(i)}\sqrt{\nu_t^{(i)}} \tag{3.1}$$

$$b_t^{(i)} = a^{(i)}\sqrt{\nu_t^{(i)}}\Delta t \tag{3.2}$$

Notice that $a^{(i)}, b^{(i)}$ comes from (2.23), do not confuse them with $a_t^{(i)}, b_t^{(i)}$. we have X_t follows a bi-variate normal distribution with an unknown correlation ρ_t. First, let us assume $\rho = \rho_t$ is a constant that does not dependent on time t, as X_t follow bi-variant distributions with different mean and variance at different time t, we can not use the sample correlation from X_t as estimator for ρ directly, the historical values of X_t is shown below:

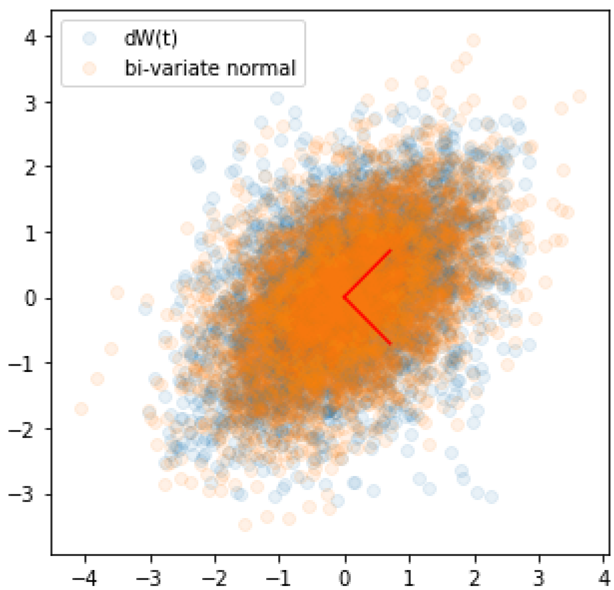


(3.3)

We already known that the distribution of historical realizations of $\Delta W_t^{(i)}$

$$\Delta W_t^{(i)} = \frac{\log \frac{S_t^{(i)}}{S_{t-\Delta t}^{(i)}} - a_t^{(i)}}{b_t^{(i)}} \tag{3.4}$$

follows a normal distribution that independent to time t, let ρ^W the correlation between $\Delta W_t^{(1)}$ and $\Delta W_t^{(2)}$. Use the the sample correlation from historical values of $\Delta W_t = \begin{bmatrix} \Delta W_t^{(1)} \\ \Delta W_t^{(2)} \end{bmatrix}$ as the estimator for ρ, the following graph shows the historical data of ΔW_t, compares with samples from bi-variate normal distribution with mean $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and correlation matrix $\begin{bmatrix} 1 & 0.447 \\ 0.447 & 1 \end{bmatrix}$



$$(3.5)$$

The next step is to validate $\rho = \rho^W$ represents the correlation between $X_t^{(1)}, X_t^{(2)}$. At time t, we have the co-variance matrix

$$M_t = \text{Cov}[X_t^{(1)}, X_t^{(2)}] = \begin{bmatrix} (b_t^{(1)})^2 & \rho^W b_t^{(1)} b_t^{(2)} \\ \rho^W b_t^{(1)} b_t^{(2)} & (b_t^{(2)})^2 \end{bmatrix} \tag{3.6}$$

Apply Eigen decomposition on M_t to get matrices Q_t, λ_t such that $M_t = Q_t^{-1} \Lambda_t Q^t$ with

$$\Lambda_t = \begin{bmatrix} \lambda_t^{(1)} & 0 \\ 0 & \lambda_t^{(2)} \end{bmatrix}, Q_t = \begin{bmatrix} Q_t^c & -Q_t^s \\ Q_t^s & Q_t^c \end{bmatrix} \tag{3.7}$$

If the ρ^W is close to the true correlation between $X_t^{(1)}$ and $X_t^{(2)}$, then $Y_t^{(1)}, Y_t^{(2)}$ should be uncorrelated normal distributions where

$$Y_t = \begin{bmatrix} Y_t^{(1)} \\ Y_t^{(2)} \end{bmatrix} = Q_t \begin{bmatrix} X_t^{(1)} \\ X_t^{(2)} \end{bmatrix} \tag{3.8}$$

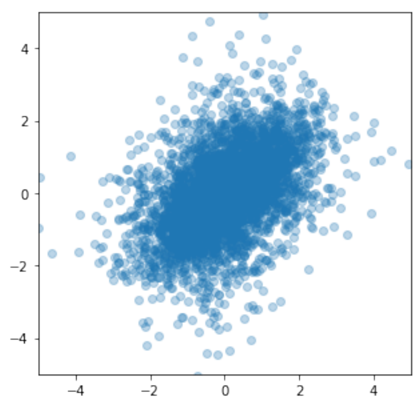
the expected value and variance of $Y_t^{(i)}$ are:

$$\mathbb{E}[Y_t] = Q_t \begin{bmatrix} a_t^{(1)} \\ a_t^{(2)} \end{bmatrix} \tag{3.9}$$

$$Var[Y_t^{(1)}] = (Q_t^c b_t^{(1)})^2 + (Q_t^s b_t^{(2)})^2 - 2Q_t^s Q_t^c \rho_t b_t^{(1)} b_t^{(2)} \tag{3.10}$$

$$Var[Y_t^{(2)}] = (Q_t^s b_t^{(1)})^2 + (Q_t^c b_t^{(2)})^2 + 2Q_t^s Q_t^c \rho_t b_t^{(1)} b_t^{(2)} \tag{3.11}$$

We will have $Z_t^{(i)} = \frac{Y_t^{(i)} - \mathbb{E}[Y_t^{(i)}]}{\sqrt{Var[Y_t^{(i)}]}}$ follows the standard normal distribution $N(0,1)$ and $Z_t^{(1)}, Z_t^{(2)}$ are independent. But the following graph shows that $Z_t^{(1)}, Z_t^{(2)}$ are still correlated with correlation 0.3517



(3.12)

Thus the validation failed, the estimator we used for correlation can not be close to the true correlation. Now, let the constant correlation between $Z_t^{(1)}, Z_t^{(2)}$ be $\rho^Z = 0.3517$. Since ρ^W could not reflect the true correlation, we rename $Var[Y_t^{(i)}]$ calculated using ρ^W as $VY_t^{(i)}$, namely,

$$VY_t^{(1)} = (Q_t^c b_t^{(1)})^2 + (Q_t^s b_t^{(2)})^2 - 2Q_t^s Q_t^c \rho^W b_t^{(1)} b_t^{(2)} \tag{3.13}$$

$$VY_t^{(2)} = (Q_t^s b_t^{(1)})^2 + (Q_t^c b_t^{(2)})^2 + 2Q_t^s Q_t^c \rho^W b_t^{(1)} b_t^{(2)} \tag{3.14}$$

we have

$$\rho^Z = \mathrm{Corr}[Z_t^{(1)}, Z_t^{(2)}] = \frac{VY_t^{(1)}\, VY_t^{(2)} \mathrm{Cov}[Z_t^{(1)}, Z_t^{(2)}]}{\sqrt{Var[Y_t^{(1)}]\, Var[Y_t^{(2)}]}} \tag{3.15}$$

$$\mathrm{Cov}[Z_t^{(1)}, Z_t^{(2)}] = \mathrm{Cov}[\frac{Q_t^c X_t^{(1)} - Q_t^s X_t^{(2)}}{\sqrt{VY_t^{(1)}}}, \frac{Q_t^s X_t^{(1)} + Q_t^c X_t^{(2)}}{\sqrt{VY_t^{(2)}}}] \tag{3.16}$$

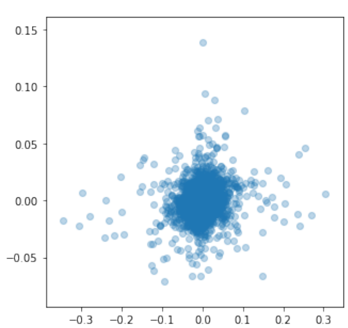
$$= \frac{Q_t^c Q_t^s}{\sqrt{VY_t^{(1)}\, VY_t^{(2)}}}(b_t^{(1)})^2 + \frac{(Q_t^c)^2}{\sqrt{VY_t^{(1)}\, VY_t^{(2)}}}\mathrm{Cov}[X_t^{(1)}, X_t^{(2)}]$$

$$- \frac{(Q_t^s)^2}{\sqrt{VY_t^{(1)}\, VY_t^{(2)}}}\mathrm{Cov}[X_t^{(1)}, X_t^{(2)}] - \frac{Q_t^c Q_t^s}{\sqrt{VY_t^{(1)}\, VY_t^{(2)}}}(b_t^{(2)})^2$$

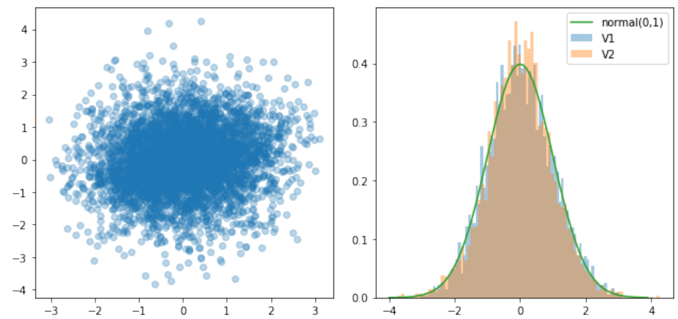
Therefore

$$\frac{\rho^Z}{\sqrt{VY_t^{(1)}\, VY_t^{(2)}}} = \frac{Q_t^c Q_t^s((b_t^{(1)})^2 - (b_t^{(2)})^2) + ((Q_t^c)^2 - (Q_t^s)^2) b_t^{(1)} b_t^{(2)} \rho_t}{\sqrt{Var[Y_t^{(1)}]\, Var[Y_t^{(2)}]}} \tag{3.17}$$

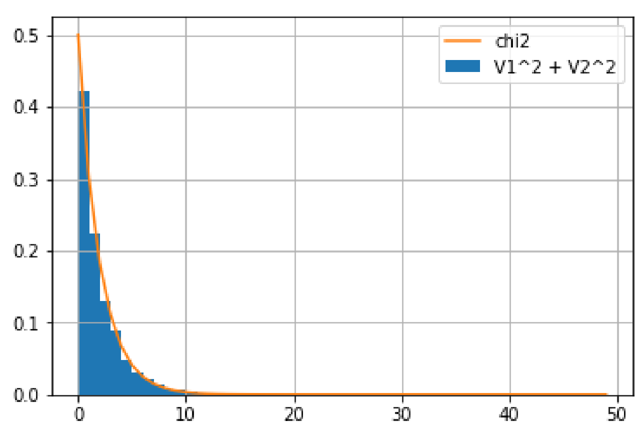
Notice that $Var[Y_t^{(i)}]$ is an expression of ρ_t from (3.10) and (3.11), solve above equation for ρ_t at time t will give an estimator for stochastic correlation ρ_t automatically.
Using the new stochastic ρ_t from (3.17), recalculate the co-variance matrix $\mathcal{M}_t$ for X_t(3.6)we have eigen decomposition $\mathcal{M}_t = P_t \Lambda_t P_t^{-1}$. The data points of $U_t = P_t X_t$ looks like



After normalization with stochastic mean $\bar{U}_t = P_t \begin{bmatrix} a_t^{(1)} \\ a_t^{(2)} \end{bmatrix}$ and stochastic variance (3.10),(3.11), it becomes



Let V_t denotes the normalized U_t, the sample correlation between $V_t^{(1)}$ and $V_t^{(2)}$ is 0.10992. So far, we can conclude that ρ_t estimated this way is close to the true correlation between $X_t^{(1)}, X_t^{(2)}$. We can repeat the procedure with ρ^Z replaced by ρ_t we found here and Q_t replaced by P_t to find a better estimator to eliminate the residue correlation 0.10992 from V_t which will only complicated the computation.
A quantitative method to validate estimator ρ_t consists with modified Heston model is: use KS-test to compare the sample distribution of $(V_t^{(1)})^2 + (V_t^{(2)})^2$ with a χ^2 distribution with degree of freedom 2. Because if we have the ρ_t estimated correctly, $V_t^{(1)}$ and $V_t^{(2)}$ should follow the standard normal distribution with correlation 0, which means $(V_t^{(1)})^2 + (V_t^{(2)})^2$ gives a chi-squared distribution with degree of freedom equals 2. In this case, KS-test gives p-value: 7.933×10^{-5}



The KS-test failed because we still have an uncaptured correlation of 0.10992 in V_t. As mentioned above, we are able to reduce the uncaptured correlation, but we will use the ρ_t we have so far as the estimator to avoid over complicated calculation.

A general method to find a good estimator for correlation is summarized as follow:

When we have a series of data $X_i = \begin{bmatrix} X_i^{(1)} \\ X_i^{(2)} \end{bmatrix}$indexed with integer i, the distribution of $X_i^{(j)}$ is known as $N(\mu_i^{(j)}, \sigma_i^{(j)})$ with unknown correlation $\rho_i = \frac{\mathrm{Cov}[X_i^{(1)}, X_i^{(2)}]}{\sigma_t^{(1)}\sigma_t^{(2)}}$. Denote the $\mathcal{N}$ as the normalization operator, e.g. $\mathcal{N}[X_i] = \begin{bmatrix} \mathcal{N}[X_t^{(1)}] \\ \mathcal{N}[X_t^{(2)}] \end{bmatrix} = \begin{bmatrix} \frac{X_t^{(1)} - \mu_t^{(1)}}{\sigma_t^{(1)}} \\ \frac{X_t^{(2)} - \mu_t^{(2)}}{\sigma_t^{(2)}} \end{bmatrix}$, denote $\mathcal{R}_\theta$ as the rotation operator that rotate an vector for an angle θ, e.g. $\mathcal{R}_\theta \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. For an arbitrarily given θ, our task is to find θ_i that makes the following diagram communicate:

$$\begin{tikzcd} X_i \arrow[r, "\mathcal{R}_{\theta_i}"] \arrow[d, "\mathcal{N}"'] & U_i \arrow[d, "\mathcal{N}"] \\ Y_i \arrow[r, "\mathcal{R}_\theta"] & V_i \end{tikzcd}$$

Then using eigen decompostion to find the correlation ρ_i that correspondence to θ_i. This is true based on the assumption that after normalization, the correlation becomes a constant and it can be reflected by the sample correlation. This assumption was made by comparing graphs (3.12) and (3.3)

References

[1] Broadie, M., & Kaya, O., *Exact simulation of stochastic volatility and other affine jump diffusion processes*

[2] Baldeaux, J *Exact Simulation of the 3/2 Model*

[3] Bohrnstedt, G. W., & Goldberger, A. S. *On the Exact Covariance of Products of Random Variables*

[4] Heston, Steven L. *A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options*

[5] Deelstra, G& Parker, G *A Coveriance Equivalent Discretisation of the CIR Model*

[6] Wilson A. G. & Ghahramani Z. *Generalised Wishart Processes*

[7] van Emmerich C. *Modelling correlation as a stochastic process.*