

WORCESTER POLYTECHNIC INSTITUTE

Exploring the optimal Transformation for Volatility

Alexander Volfson

4/29/2010

APPROVED BY DR. BALGOBIN NANDRAM, PROJECT ADVISOR

ABSTRACT

This paper explores the fit of a stochastic volatility model, in which the Box-Cox transformation of the squared volatility follows an autoregressive Gaussian distribution, to the continuously compounded daily returns of the Australian stock index. Estimation was difficult, and over-fitting likely, because more variables are present than data. We developed a revised model that held a couple of these variables fixed and then, further, a model which reduced the number of variables significantly by grouping trading days. A Metropolis-Hastings algorithm was used to simulate the joint density and derive estimated volatilities. Though autocorrelations were higher with a smaller Box-Cox transformation parameter, the fit of the distribution was much better.

1 INTRODUCTION

1.1 Background

The use of options dates back to the issuance of insurance on the delivery of colonial cargo ships, however, as MacKenzie (2003) argues, “until the 1970s age had not brought them respectability”. As the most basic form of risk transfer, the option remains a building block of modern financial instruments. The purchase of an option is effectively a gamble with a varying payoff. To the purchaser of a European Call, the payoff occurs at an agreed upon date in the future known as maturity and is dependent upon the final value of some asset and the strike price. If the final asset’s price is below the strike, no payoff is made. However, if the asset’s price is above the strike, the difference in price is paid to the purchaser of the option. This particular “gamble” on the price of an asset is called an option because it describes a contract where the purchaser of the option has the *option* of purchasing the agreed upon asset, at maturity, for the strike price K . Of course, such an option would not be exercised in the event that the price of the asset was below the strike. In either case, the payoff, as a function of the stock price, can be visualized as follows:

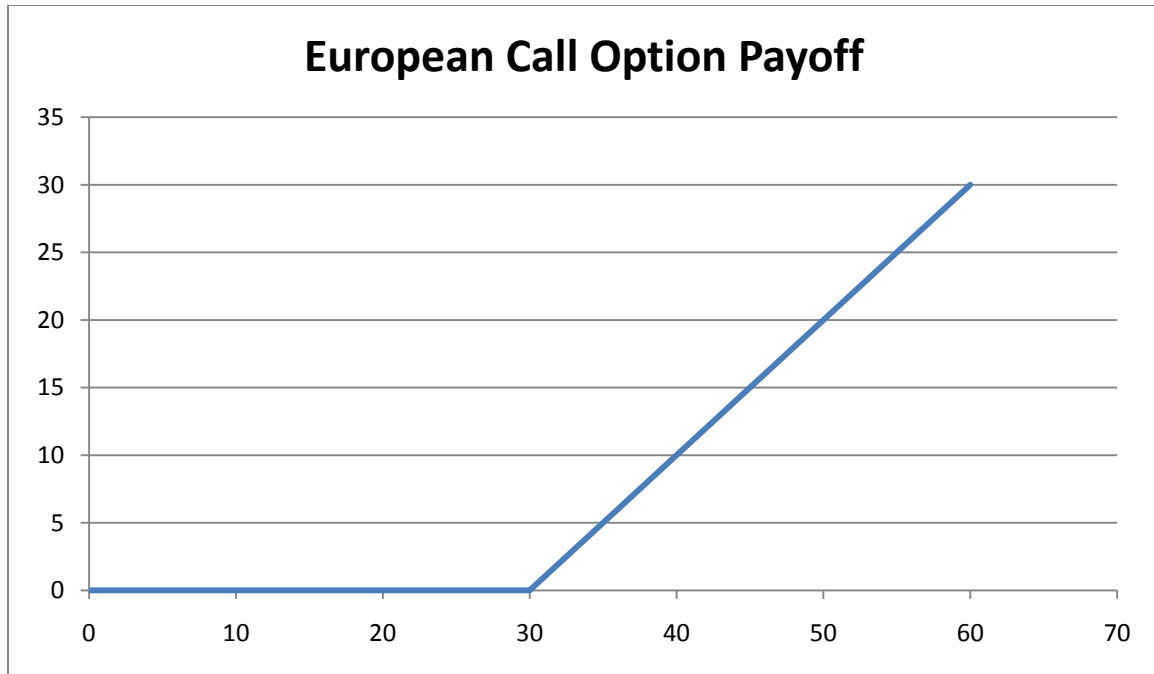


FIGURE 1. THE PAYOFF OF A CALL OPTION WITH STRIKE PRICE OF 30.

While both the asset and the strike price are agreed upon at the time of issue, the strike is a fixed value (typically denoted as K) while the asset's price is variable. Therein lies the obstacle of option pricing; the asset's final price is unknown at the time of issue. Furthermore, option traders may want to exchange options after issue, but before maturity, necessitating an intermediate price. Since the trajectory, or path, of the asset's price is not known in advance, models of the price give traders a framework in which to understand the option prices as well as transact with them.

Apparently, it was Louis Bachelier who first applied probability (and what would later be known as stochastic analysis) to the description of stock price evolution. His description was based on the implicit assumptions that "small fluctuations in price seen over a short time interval should be independent of the current value of the price" and these fluctuations are independent and normally distributed. This leads to a trajectory known as Brownian motion and a rigorous mathematical basis would later be supplied by Wiener based on French physicist, Jean Perrin's experimental work. (Davis 2006)

Brownian motion describes the path of a large particle floating in a fluid. Since the fluid consists of many relatively small particles moving rather quickly in all directions, the large particle is variously impacted and receives a random, non-negligible force at every instant. For a stock price, this can be imagined as a limiting case of a random walk. Imagine a person in possession of a coin and an urge for a workout. Each minute, this person flips the coin and takes a step: stepping to left if the coin lands one way and to the right if it lands the other. In order to make this process more continuous, the person changes his strategy and flips the coin twice per minute while taking steps only one half the size of the previous steps. Next, he flips the coin four times per minute while taking steps only one fourth the size. Continuing until the coin is flipped every instant and the steps are infinitesimally small, we arrive at Brownian motion, a continuous-time random walk.

At least one property of Brownian motion makes it entirely unsuitable for describing long-term stock prices or the price of any asset for the matter: Brownian trajectories can go negative and the probability of this increases over time. Brownian motion remains widely used, however, as a description of a stock's *returns*. In such a model, the stock price itself follows Geometric Brownian motion, described by the stochastic differential equation below:

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

In this equation, S is the price of the stock, μ is the systemic drift, σ is the volatility and W is a Brownian motion (also known as a Wiener process). The left half represents the stock price return and the right consists of a drift term which is constant plus some multiple of the Brownian motion for randomness. This description of stock prices was "in the air" by the late 1950's when Sprenkle as well as Samuelson and Osborne researched. (MacKenzie)

Eventually making use of Sprenkle's work, Black and Scholes were able to develop a pricing formula for options based on the assumption, among others, that stock prices follow Geometric

Brownian motion trajectories. Merton, unhappy with Black and Scholes' assumption of "quadratic utility" as part of the Capital Asset Pricing Model, later, provided an alternative derivation of the same formula under the assumption that continuous trading was possible. Interestingly enough, at the time Black and Scholes went to publish their work, many of their assumptions did not apply: transaction costs were nowhere near negligible, assets could be borrowed in any fraction and loans could not be taken out at the risk free rate. "Wildly unrealistic" is the phrase MacKenzie uses to describe the relevance of Black, Scholes and Merton's assumptions in 1973. MacKenzie argues that it was the very work of Black-Scholes-Merton that led to these assumptions "[becoming], while still not completely realistic, a great deal more so."

The Black-Scholes framework for option pricing relies on the assumption of constant volatility in the log-normal distribution of asset returns. However, since the crash of 1987, many assets have not demonstrated log-normal returns (Rubinstein 1994).

Even before then, however, modelers have found volatility to change over time. Two main approaches have been used to extend the Black-Scholes framework, the autoregressive conditional heteroskedasticity (ARCH) model and the stochastic volatility (SV) model.

1.1.1 Autoregressive Conditional Heteroskedasticity (ARCH)

Articulated by Engle in his 1982 paper, "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom", the ARCH model has been fertile ground for extension. Engle focused on a model which assumes that the processes' volatility on a given day is equal to a weighted average of some number of previous days' returns. Higgins and Bera (1992), writing about their generalization to the ARCH model one decade later, contend "that ARCH models capture some important features of time series data, such as nonlinear dependence, nonnormality and over-dispersion."

Extensions of the ARCH model assume that a particular day's volatility is a function of previous days' returns and volatilities. Additionally, names of these extensions generally end with "ARCH". The generalized autoregressive conditional heteroskedasticity (GARCH, Bollerslev (1986)) model assumes that the processes' volatility is a weighted average of some number of previous days' volatilities as well as some, possibly different, number of previous days' returns. Meanwhile, the Nonlinear ARCH (NARCH) model proposed by Higgins and Bera (1992) introduces a Box-Cox power transformation to the previous days' returns. Between NGARCH, IGARCH, EGARCH, GARCH-M, QGARCH, GJR-GARCH and TGARCH, a medley of constraints, nonlinear terms as well the ability to model asymmetry have resulted in a rich research space. So rich, in fact, that in an effort to capture all of these bountiful features, Hentschel (1995) integrated them into a generalized family of GARCH models (fGARCH) by introducing a few characteristic parameters.

1.1.2 Stochastic Volatility

Where the original Black-Scholes-Merton model assumes constant volatility in the stochastic differential equation describing the trajectory of a stock price, stochastic volatility (SV) models, assume that the volatility itself follows some stochastic process (possibly Brownian Motion). Thus, the model might be parameterized as:

$$\frac{dS}{S} = \mu dt + \sqrt{v} dW,$$

$$dv = \alpha dt + \beta dB.$$

Here, α and β may be a functions of the current stock price and the time, while B and W are Brownian motions with correlation ρ (possibly zero). Extensions do not stray too far from this framework in order to capture interesting features. The Heston, GARCH(1, 1) and 3/2 models all replace α with $\theta(\omega - v)$, where ω is the value to which the volatility tends to revert and θ represents the rate at which this reversion occurs. This feature is very desirable because volatility

does not tend to wander to arbitrarily large values for very long (nor negative values for that matter). Tangentially, the difference between these three models is in their replacement of β with ζv^δ , where ζ is the volatility of the volatility and δ takes on the values: $\frac{1}{2}$; 1; and $\frac{3}{2}$ respectively.

1.2 Basic Volatility Model

Carlos Blanco and David Soronow (2001) argue that mean reversion for energy prices supports intuition because price shocks (positive and negative) often dissipate and the price goes back to average levels. This paper will explore the basic Box-Cox transformed SV model used by Zhang and King (2008). The use of the Box-Cox transformation is to “allow for skewness in the marginal distribution of the squared volatility” (Zhang). In basic version of the discrete-time model is:

$$y_t = \sqrt{g(\alpha_t, \delta)} \varepsilon_t, t = 1 \dots n$$

$$\alpha_t = \mu + \phi(\alpha_{t-1} - \mu) + \sigma u_t,$$

where $\varepsilon_t \sim N(0,1)$, $u_t \sim N(0,1)$ and

$$g(\alpha_t, \delta) = \begin{cases} (1 + \delta \alpha_t)^{\frac{1}{\delta}}, & \text{if } \delta \neq 0 \\ e^{\alpha_t} & \text{if } \delta = 0. \end{cases}$$

In this model, α_t represents the Box-Cox transformed squared volatility with parameter δ ; y_t represents the continuously compounded return; μ is the value to which the transformed volatility reverts and ϕ is the reversion rate. At each time step, randomness comes into the model via ε_t in a straight-forward way and u_t , which impacts α_t in addition to the mean reversion.

Zhang and King (2008) compare the fit of a variety of their own Box-Cox extension of a log-transformed SV model to: a similar model without the Box-Cox extension; a similar model without the heavy-tailed error assumption; a similar model without the return-volatility correlation; and finally, a t-GARCH(1,1) model. Since the Box-Cox transformation is equivalent to a log transformation when $\delta=0$, they hoped that their Box-Cox transformation model would fit the data

significantly better than the other models and that the δ parameter be significantly far from zero. While they found strong evidence for the latter, two of their three criteria consequentially favored the t-GARCH(1,1) model. The Box-Cox transformation extension with heavy tails and correlation is the best, Zhang and King (2008) argue, because it is favored when comparing the distributions of the fitted residuals.

Since only the y_t variables are observable, the remaining parameters must be estimated to initialize the Metropolis-Hastings algorithm. There is difficulty in estimating μ and δ , however. Consider that:

$$y_t | \alpha_t, \delta \sim N\left(0, (1 + \delta \alpha_t)^{\frac{1}{\delta}}\right);$$

While the variance of this distribution, $(1 + \delta \alpha_t)^{\frac{1}{\delta}}$, can be estimated from the data, identifying the combination of α_t and δ that led to this variance is an underspecified problem. Similarly, consider the conditional density for α_t can be rewritten from:

$$\alpha_t | \mu, \phi, \alpha_{t-1} \sim N(\mu + \phi(\alpha_{t-1} - \mu), \sigma^2)$$

to:

$$\alpha_t - \mu | \phi, \alpha_{t-1} - \mu \sim N(\phi(\alpha_{t-1} - \mu), \sigma^2) \rightarrow \alpha_t^* | \phi, \alpha_{t-1}^* \sim N(\phi \alpha_{t-1}^*, \sigma^2),$$

then, in the log-case ($\delta = 0$):

$$y_t | \alpha_t^*, \delta \sim N(0, e^{\alpha_t^* + \mu}).$$

Thus, while it may be possible to estimate α_t^* , it will not really be possible to discern the contribution from μ .

In this paper, the fit of the basic model (above) as well as a modification of the basic model (below) to the Australian All Ordinaries stock index was explored with the purpose of determining

the optimal δ parameter for the Box-Cox transformation as this is critical to the model and not easily observable.

2 MODEL

The basic model above, assigns one Box-Cox transformation of squared volatility to each time t , known as α_t . The following extension groups these transformed volatilities, assigning one volatility to m trading days. Of course, each trading day retains its individual return, known as y_t . Thus, the model is redefined as:

$$y_t = \sqrt{g(\alpha_s, \delta)} \varepsilon_t, t = 1 \dots n$$

$$\alpha_s = \mu + \phi(\alpha_{s-1} - \mu) + \sigma u_s, \text{ with } s = 1 \dots \frac{n}{m} \text{ and } m|n$$

where m is the number of daily returns per change in volatility, n is the total number of trading days, $s = \left\lceil \frac{t}{m} \right\rceil$, and the remaining parameters are as above.

In order to determine values for the parameters of the model, a random walk Metropolis-Hastings sampler was used with blocking. This algorithm works by sampling each parameter from its conditional posterior distribution iteratively. Proposal densities were used to sample ϕ and α_s .

Letting $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{n/m})$, $y = (y_1, y_2, \dots, y_n)$, and $\theta = (\phi, \mu, \sigma)$, the posterior density of (θ, α) is:

$$\pi(\theta, \alpha | y) = \frac{p(\theta, \alpha, y)}{p(y)} = \frac{p(y | \theta, \alpha) * p(\alpha | \theta) * p(\theta)}{p(y)} \propto p(y | \theta, \alpha) * p(\alpha | \theta) * p(\theta).$$

Following directly from Bayes' Theorem, with $p(\theta)$ representing the prior distribution on θ .

Following fairly directly from the previous definition of the model and variables, we have:

$$y_t \sim N\left(0, (1 + \delta\alpha_s)^{\frac{1}{\delta}}\right)$$

$$\alpha_1 \sim N\left(\mu, \frac{\sigma^2}{1 - \phi^2}\right)$$

$$\alpha_s \sim N(\mu + \phi(\alpha_{s-1} - \mu), \sigma^2).$$

Meanwhile, priors were assumed for the variables of θ :

$$\phi \sim U(-1, 1)$$

$$\mu \sim N(\mu_0, q_0)$$

$$p(\sigma^2) \sim \left(\frac{1}{\sigma^2}\right)^{\frac{\zeta}{2}+1} e^{-\frac{S_t}{2\sigma^2}}.$$

As in Zhang and King (2008), μ_0 , q_0 , S_t , and ζ are hyperparameters. While μ_0 was estimated from the data, the remaining hyperparameters were set to 0.01, 0.001, and 0.001. The original formulation by Zhang and King (2008) contained an underlying Beta distribution for ϕ which was discarded because it necessitated two additional hyperparameters. The original formulation also convolved the distribution for μ with σ . To avoid this dependence and simultaneously simplify the posterior distribution for σ , the variance of the prior for μ simply contains q_0 .

The posterior, distributions for α_s were thus found to be:

$$p(\phi | \alpha, \mu, \sigma^2) \propto \sqrt{1 - \phi^2} p(\psi)$$

$$p(\alpha_{n/m} | z_{n/m}^2, \alpha_{\frac{n}{m}-1}, \mu, \phi, \sigma^2) \propto e^{-\frac{z_{n/m}^2}{2g_{n/m}}} \frac{1}{g_{n/m}^m} * \frac{1}{\sigma} e^{-\frac{(\alpha_{\frac{n}{m}} - \mu - \phi(\alpha_{\frac{n}{m}-1} - \mu))^2}{2\sigma^2}}$$

$$p(\alpha_s | z_s^2, \alpha_{s-1}, \alpha_{s+1}, \mu, \phi, \sigma^2) \propto e^{-\frac{z_s^2}{2g_s}} \frac{1}{g_s^m} * \frac{1}{\sigma^2} e^{-\frac{(\alpha_s - \mu - \phi(\alpha_{s-1} - \mu))^2 + (\alpha_{s+1} - \mu - \phi(\alpha_s - \mu))^2}{2\sigma^2}}$$

$$p(\alpha_1 | z_1^2, \alpha_2, \mu, \phi, \sigma^2) \propto e^{-\frac{z_1^2}{2g_1}} \frac{1}{g_1^m} * \frac{1}{\sigma} e^{-\frac{(\alpha_1 - \mu - \phi(\alpha_2 - \mu))^2}{2\sigma^2}}$$

where

$$p(\psi | \alpha, \mu, \sigma^2) \sim N\left(\frac{\sum_{s=2}^n (\alpha_s - \mu)(\alpha_{s-1} - \mu)}{\sum_{s=2}^{n-1} (\alpha_s - \mu)}, \frac{\sigma^2}{\sum_{s=2}^{n-1} (\alpha_s - \mu)}\right)$$

$$z_i^2 = \sum_{t=(i-1)m+1}^{im} y_t^2.$$

Obtaining the conditional posterior (not simply the posterior) allows the algorithm proceeds faster and more efficiently. For α_s and ϕ , this was not possible, however for μ and σ it was:

$$\sigma^2 | \alpha, \mu, \phi \sim IG\left(\frac{n + \zeta + 2}{2}, \frac{k}{2}\right)$$

$$\mu | \alpha, \phi, \sigma^2 \sim N\left(\frac{A}{B}, \frac{\sigma^2}{B}\right),$$

where

$$k = (1 - \phi^2)(\alpha_1 - \mu^2) + \sum_{s=1}^{\frac{n}{m}-1} (\alpha_{s+1} - \mu - \phi(\alpha_s - \mu))^2 + S_t$$

$$A = (n - 1)(1 - \phi)^2 + (1 - \phi^2) + \frac{\sigma^2}{q_0}$$

$$B = (1 - \phi^2)\alpha_1 + (1 - \phi) \sum_{s=2}^{\frac{n}{m}} (\alpha_s - \phi\alpha_{s-1}) + \mu_0 \frac{\sigma^2}{q_0}$$

The sampling of (θ, α) was done in this order:

- 1) Sample ϕ from its posterior via random walk Metropolis step
- 2) Sample σ^2 and μ directly based on their conditional posteriors

3) Sample α using the random walk Metropolis step with Gaussian proposal densities

3 METHODOLOGY

Using the distributions from above, we have a mixture model with parameters α, y , and $\theta=(\phi, \delta, \mu, \sigma)$. Conditional distributions for each parameter facilitated the Bayesian, empirical Bayesian, analysis. Starting with the index prices, the follow transformation resulted in a dataset of mean-corrected and variance scaled continuously compounded returns:

$$r_t = \ln\left(\frac{p_t}{p_{t-1}}\right); y_t = \frac{r_t - \bar{r}}{s}, \text{ for } t = 1 \dots n$$

where p_t is the price process, r_t is the return, \bar{r} is the mean return, and s is the standard deviation of the returns.

Starting with $\{y_t\}$, the Markov Chain Monte Carlo (MCMC) method was used to simulate the joint distribution of the parameters. Because of the difficulty of sampling δ , it was fixed. Thus, δ was fixed at various values to see which one works the best.

In order to kick-off the algorithm, parameter estimates were derived from the dataset. Since the returns were mean-corrected, the squared returns are representative of that day's volatility. Thus, the Box-Cox transformation of the squared returns was used as the initial α_t values. Because the α_t variables effectively form a series of equations with σu_t taking the role of a Gaussian error term, a simple linear regression was used to derive values for ϕ and μ . The residuals of the regression were then used to generate an initial value for σ .

The following are plots of the closing price path of the index from January 2nd, 2000 through December 30th, 2005 as well as the mean-corrected and variance scaled returns for the same time period:

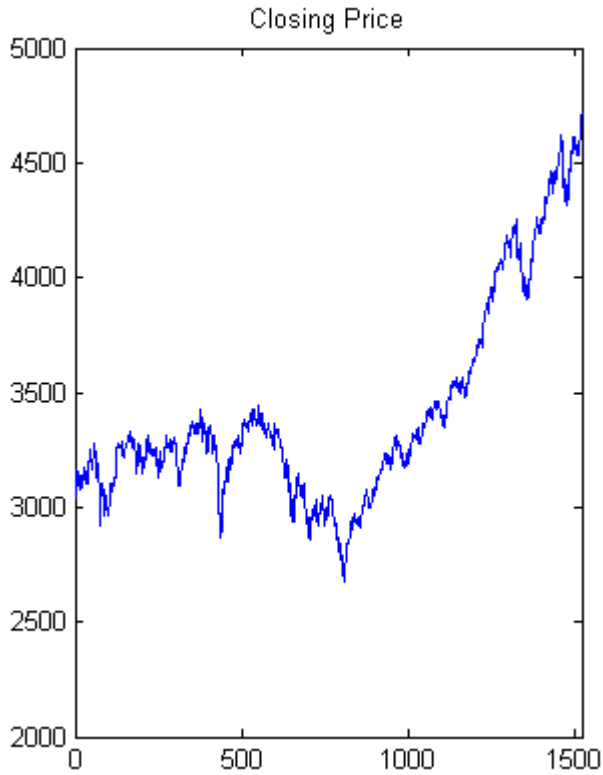


FIGURE 2.

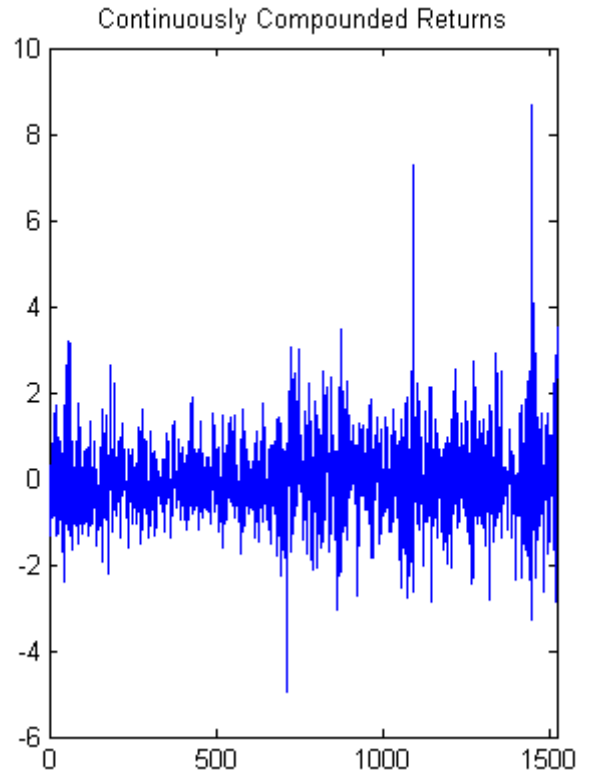


FIGURE 3.

3.1 Fixing μ and δ

The MCMC method seeks to create a joint distribution for the variables of the model. Since Metropolis-Hastings steps are used, the possibility of autocorrelation is high yet undesirable. The first 500 iterations were used as a burn-in period to ensure convergence, and 50% thinning was used to minimize the autocorrelations. To determine how much thinning is best, the autocorrelations for ϕ and σ are shown below in a run of 20,000 iterates holding μ constant (at an estimated value) and $\delta=-0.4$:

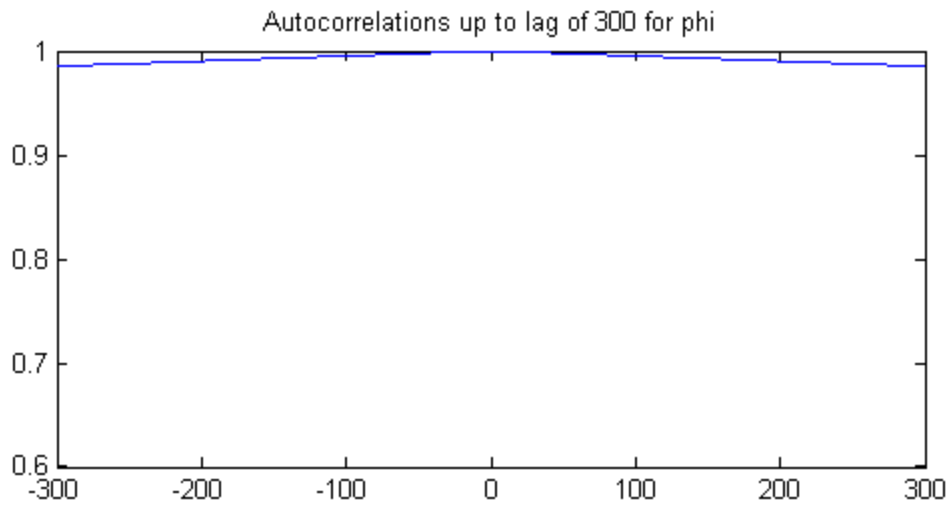


FIGURE 4. THE AUTOCORRELATIONS FOR PHI ARE ABOUT 1.0 EVEN OUT TO LAG 300

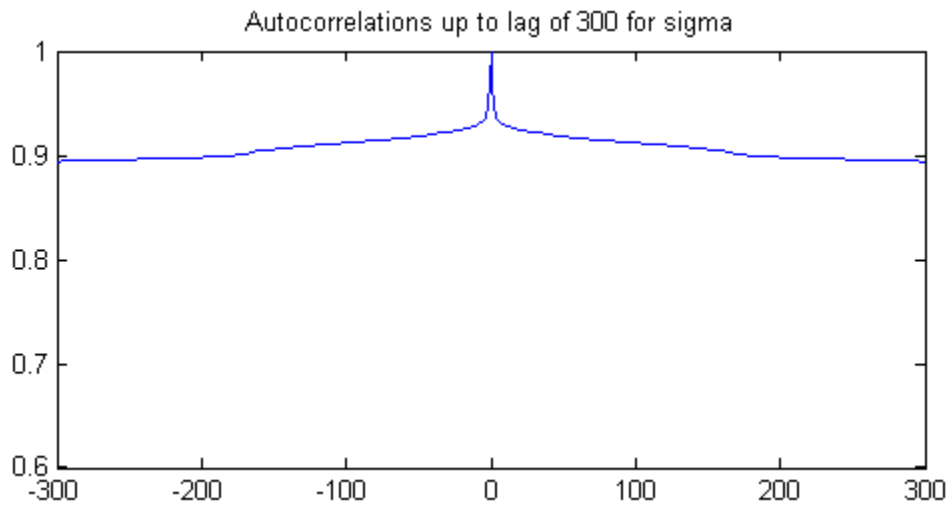


FIGURE 5. THE AUTOCORRELATIONS FOR SIGMA DROP TO 0.9 QUICKLY

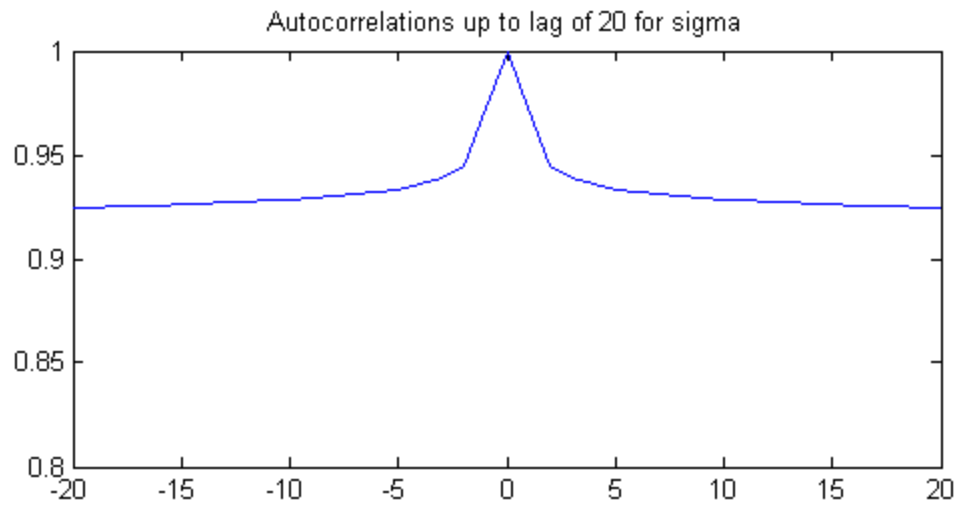


FIGURE 6. AUTOCORRELATION FOR SIGMA DROPS TO 0.93 AT LAG OF 5

The high autocorrelations of ϕ and σ are disconcerting. With 50% thinning, the trajectories look reasonable since ϕ and σ are not expected to move too much. The trajectories show movement, but not wide swings.

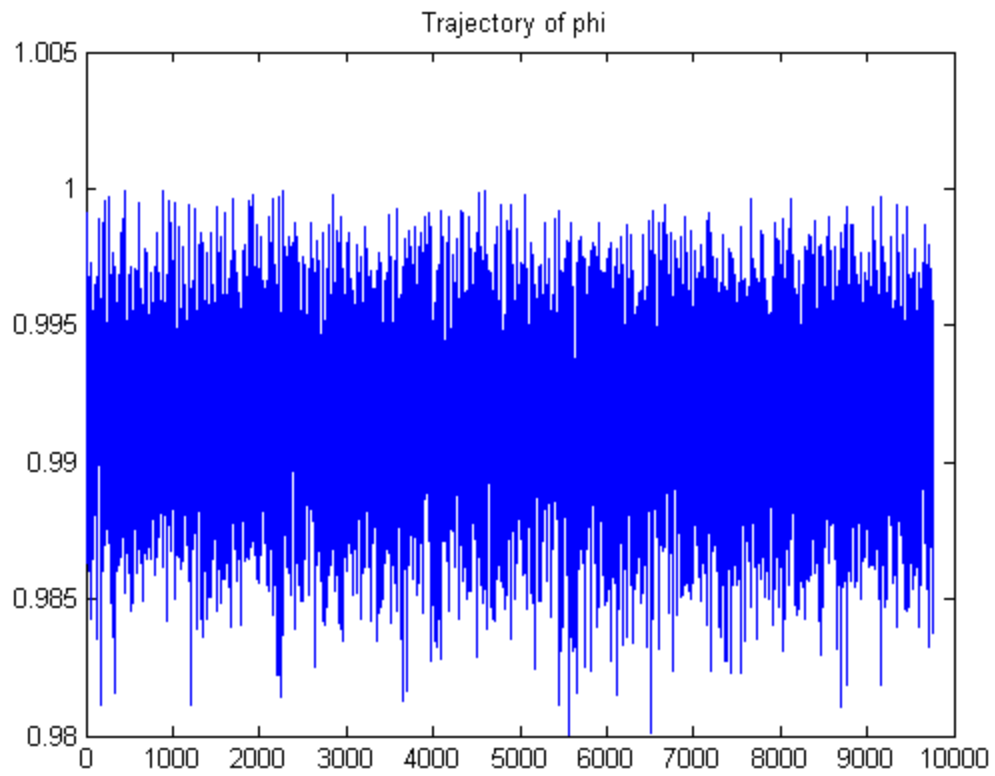


FIGURE 7.

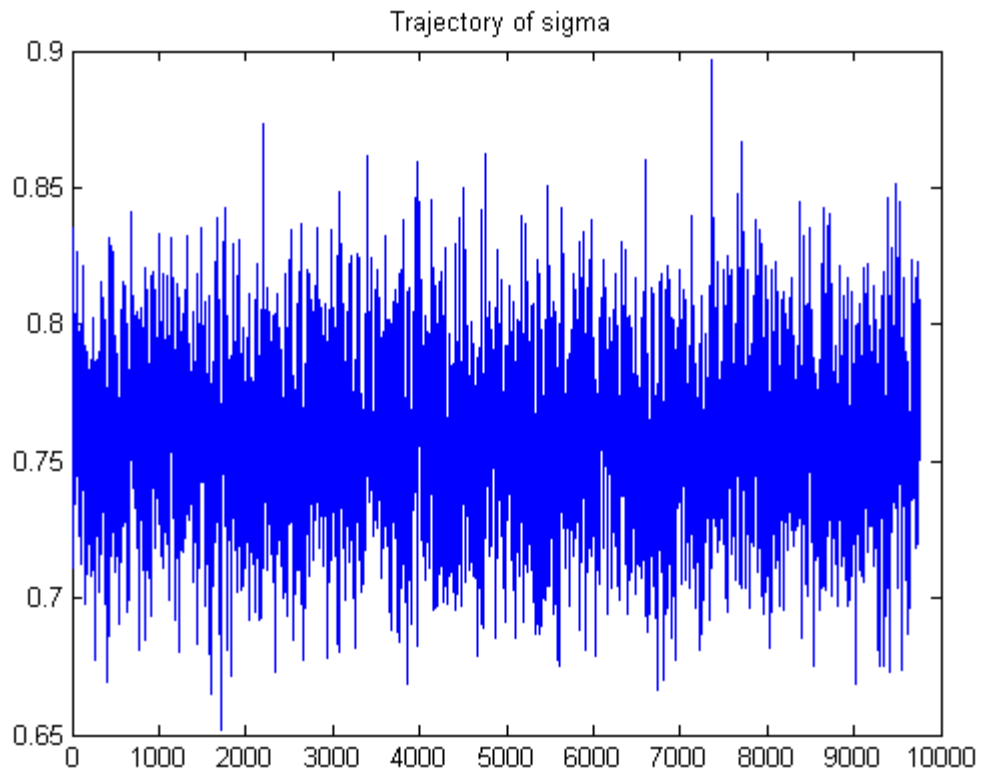


FIGURE 8.

In order to get a sense of the autocorrelations for α_t the smallest autocorrelation (considering all lags up to 300) was taken and plotted in the following histogram:

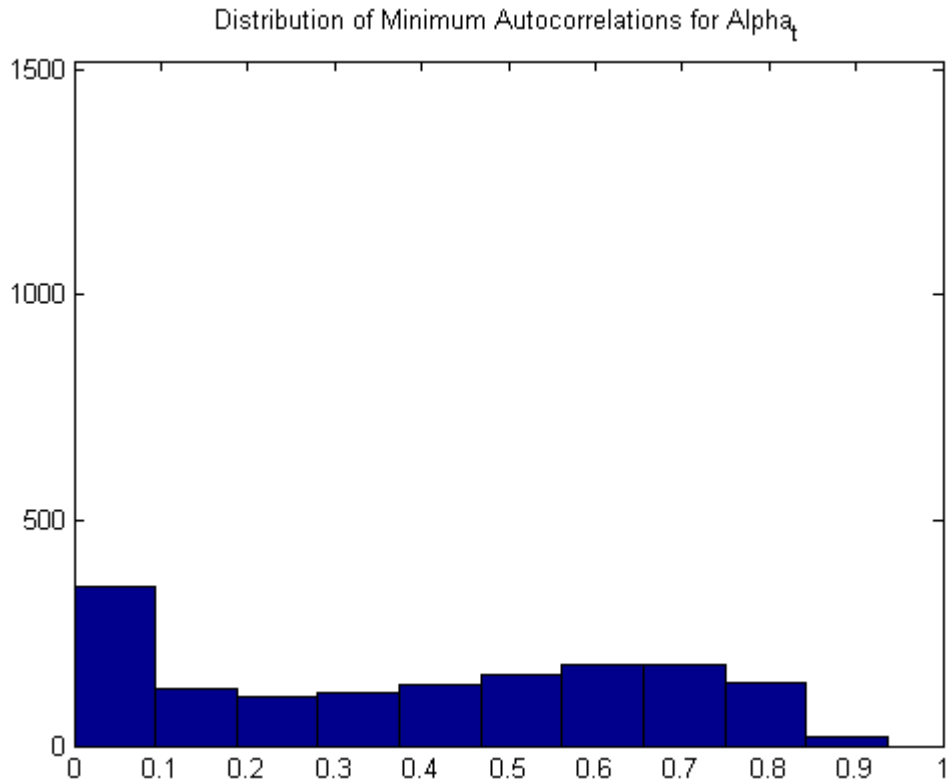


FIGURE 9.

The tallest bar on the left represents about 25% of the α_t 's which is desirable. However, the remaining α_t 's are distributed all the way out to 0.5 and 0.8 autocorrelations. This is fairly surprising because a look at a histogram of the jumping frequencies for the α_t 's is quite promising:

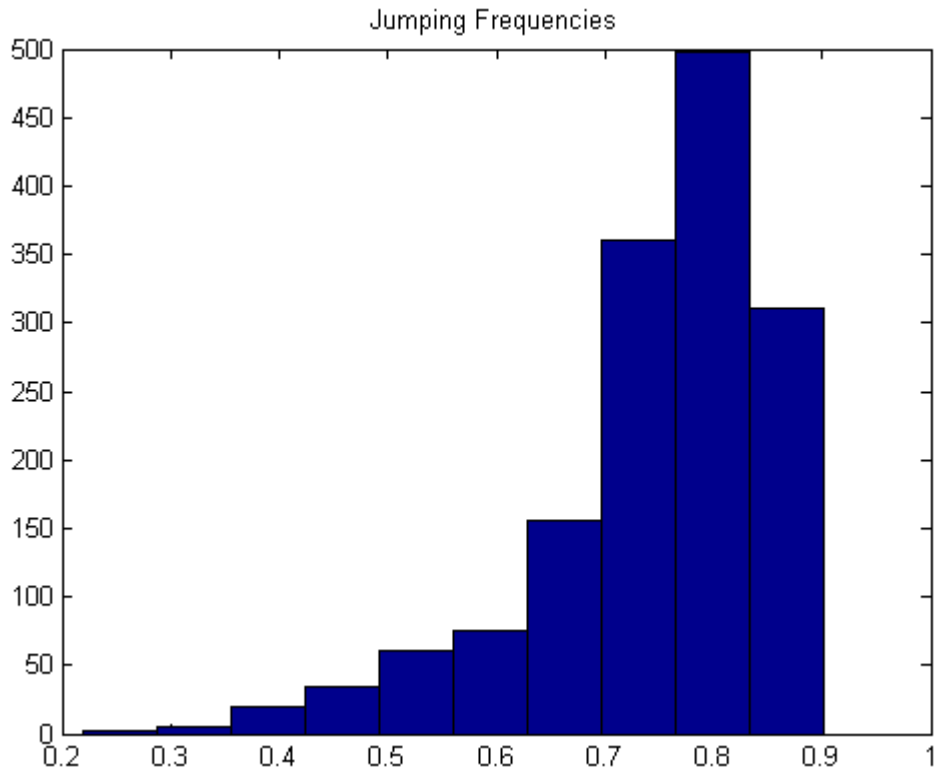


FIGURE 10.

Frequencies of 25% to 50% are to be expected with the Metropolis-Hastings algorithm, and in this run, almost all α_t 's are within or above that range.

The trajectories of the α_t 's themselves also look very stable:

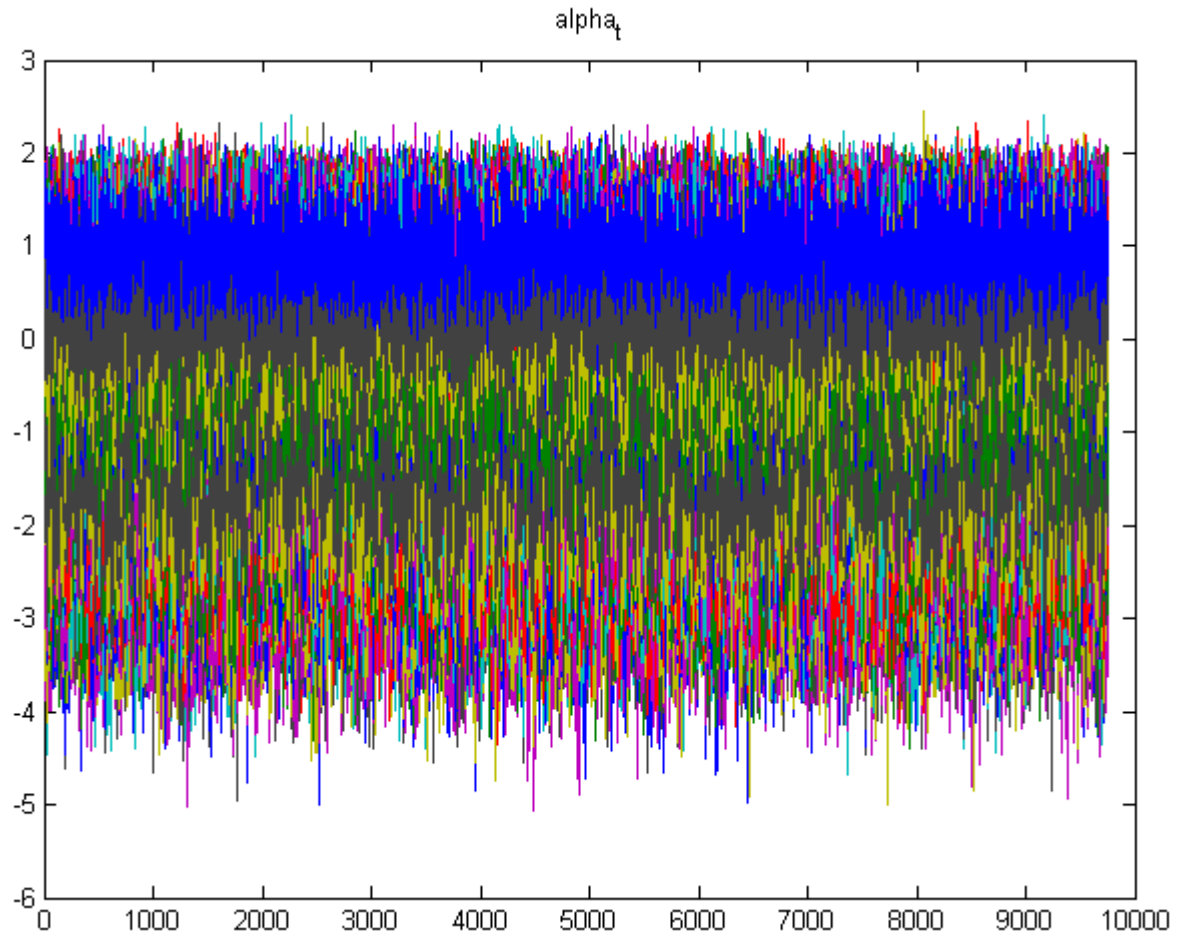


FIGURE 11. ALPHA TRAJECTORIES

To get a better sense of the estimates for each α_t , the mean, along with a 90% band is plotted below.

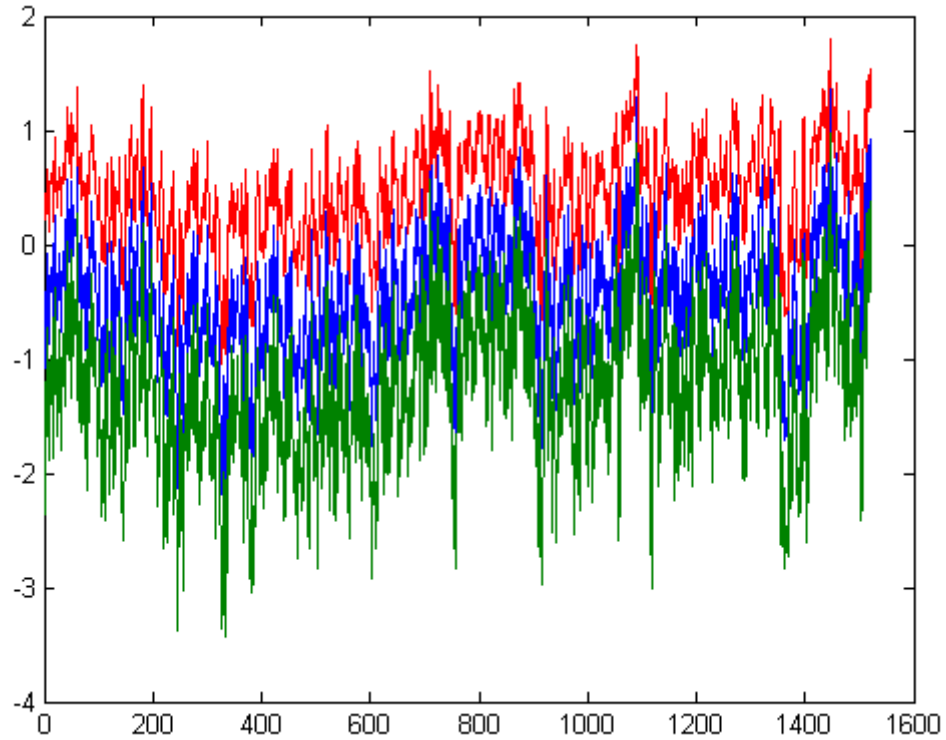


FIGURE 12. FINAL ALPHA ESTIMATES (BLUE) ALONG WITH 90% BANDS (RED AND GREEN)

The variability of the α_t 's (above) is not too great, suggesting that it is reasonable to derive estimated volatilities from them. Thus, this simulation yields estimated volatilities of:

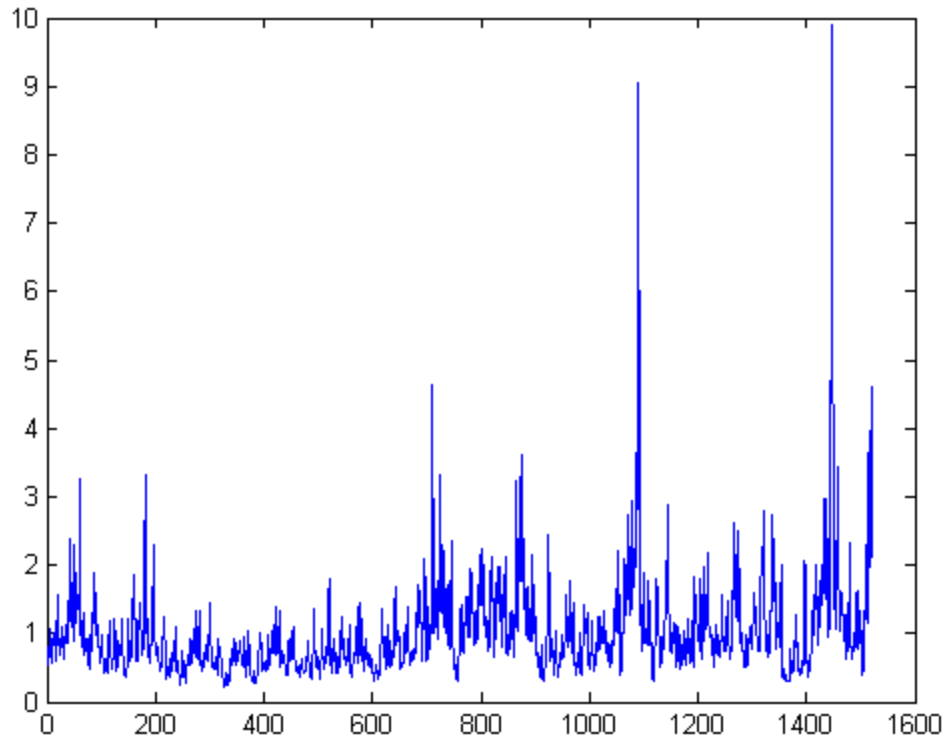


FIGURE 13. ESTIMATED VOLATILITIES FOR FIXED MU AND DELTA=-0.4

3.2 Grouping

One of the pitfalls of the model as presented above is that there are more variables than data. For each stock price, there is one variable representing volatility, α_t , in addition to a handful of other variables. Furthermore, μ and δ are not derivable from the data. These issues may be the underlying reason for the autocorrelations noted earlier because the presence of too many free variables might result in an over-fit model and, thus, Metropolis-Hastings steps that leave all of the variables very close to their previous values.

Thus, the model was redefined so that each α_t is not only the transformed volatility for y_t , but is the transformed volatility for a consecutive group of m such y_t 's. On this basis, runs were made for a variety of δ values between negative two and positive two in order to determine which provides the optimal fit.

The autocorrelations look similar across different values of δ , but some differences are apparent. The general pattern is demonstrated with $\delta = -0.54$:

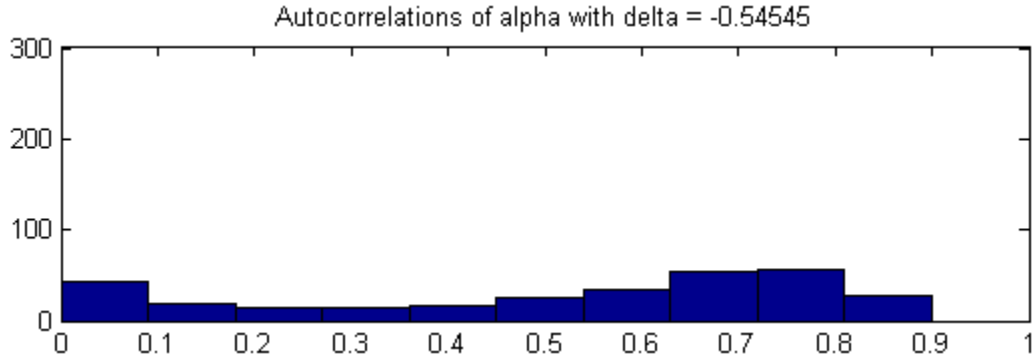


FIGURE 14.

The following table summarizes the distributions of the autocorrelations of the α_t parameters for each δ .

Percentage of Autocorrelation Distribution

Delta	Below 0.1	Above 0.3	Above 0.8
-2	13.16	76.64	3.29
-1.6364	13.49	77.96	4.61
-1.2727	14.47	76.32	9.54
-0.9091	14.80	75.33	11.18
-0.5455	15.46	72.37	10.53
-0.1818	16.12	72.37	10.20
0.1818	17.76	72.37	11.51
0.5455	19.41	70.72	10.20
0.9091	20.72	66.45	8.88
1.2727	22.70	62.83	8.22
1.6364	23.03	59.87	7.57
2	29.93	48.68	5.26

TABLE 1. SUMMARY OF THE DISTRIBUTION OF THE AUTOCORRELATIONS OF THE TRANSFORMED VARIANCE PARAMETERS

On the basis of these autocorrelations, it seems that larger (more positive) δ yields lower autocorrelations. When considering the autocorrelations of σ , however, a slightly different picture

emerges. For δ between -1.25 and +1.25, the autocorrelation is very high. The farther away δ goes from this range, however, the lower the autocorrelations. For $\delta = -2$, for example, the autocorrelations drop to 0.1 around a lag of 5.

To determine which delta fits the data best, it may be worthwhile to look at the variability present in the trajectories of the distributions of the volatility of each grouping of days. Smaller coefficients of variance suggest that the trajectories vary little compared to their average values, while a larger value suggests more variance. The distribution of the coefficient of variation (standard deviation divided by mean) of the variance trajectories is plotted below for all α_s trajectories for each δ .

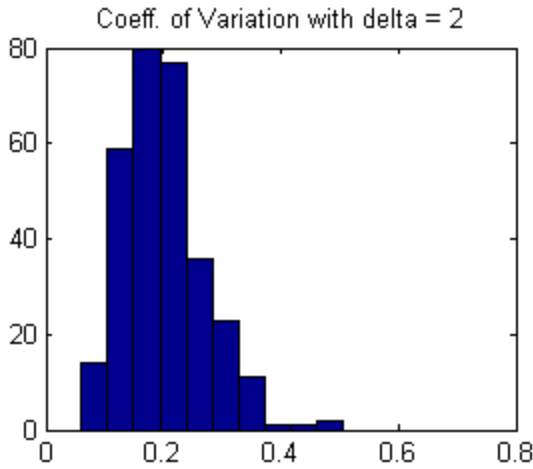


FIGURE 15.

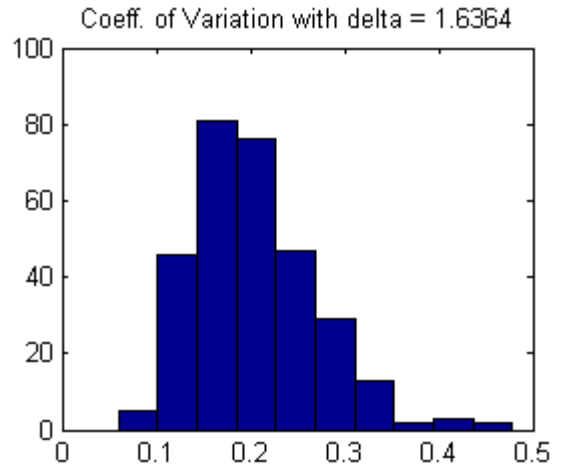


FIGURE 16.

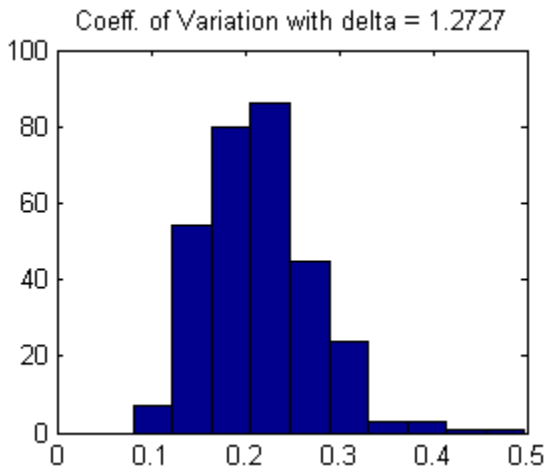


FIGURE 17.

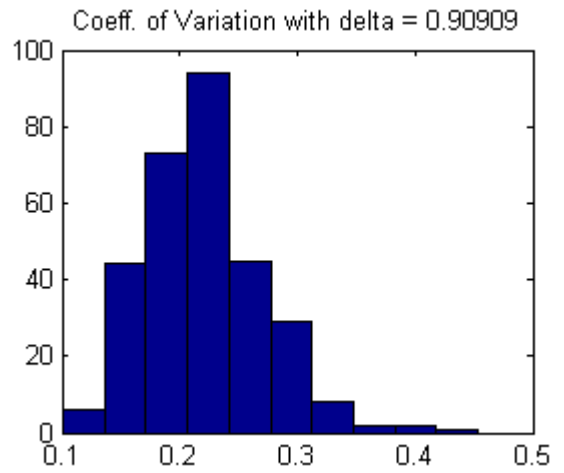


FIGURE 18.

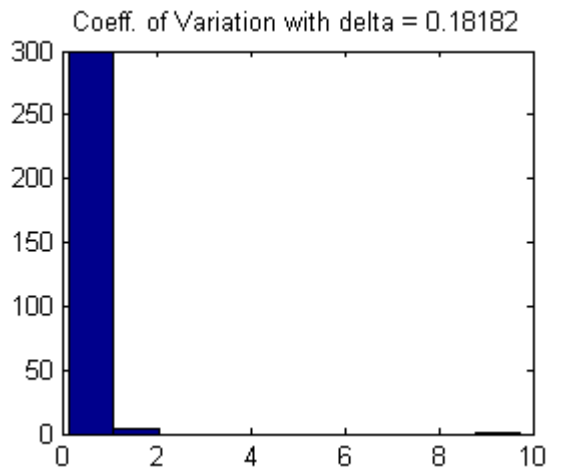
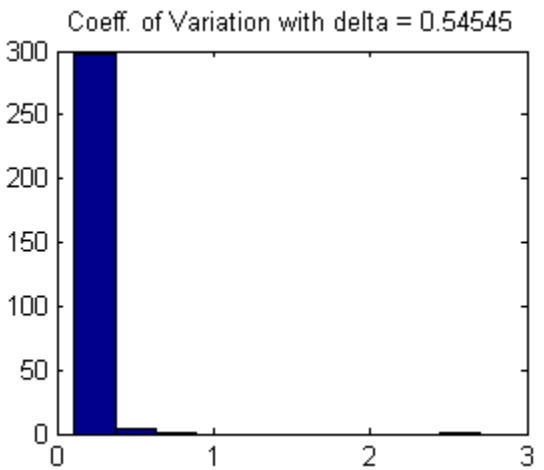


FIGURE 19.

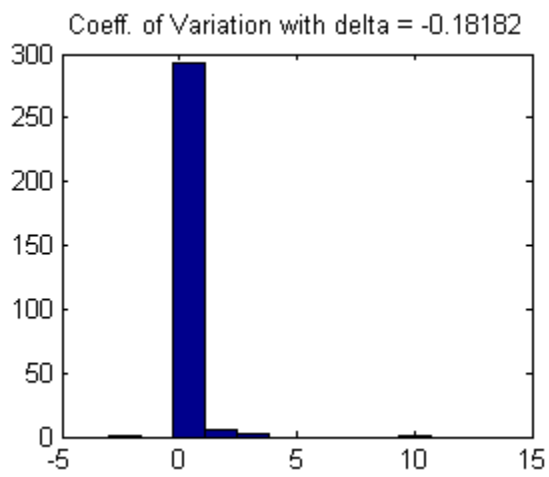


FIGURE 20.

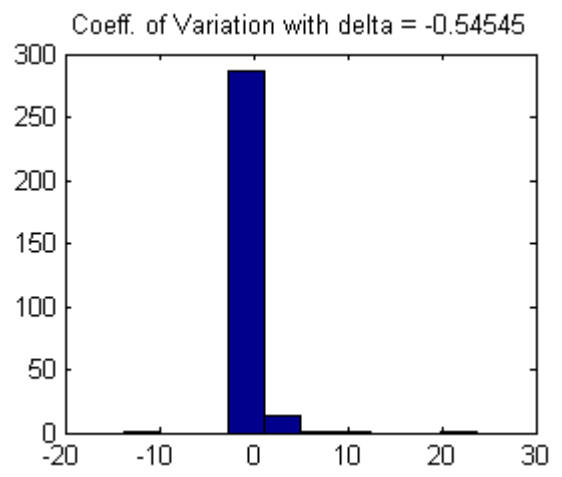


FIGURE 21.

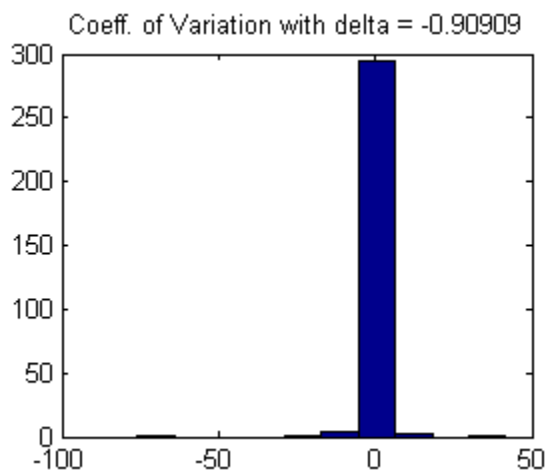


FIGURE 22.

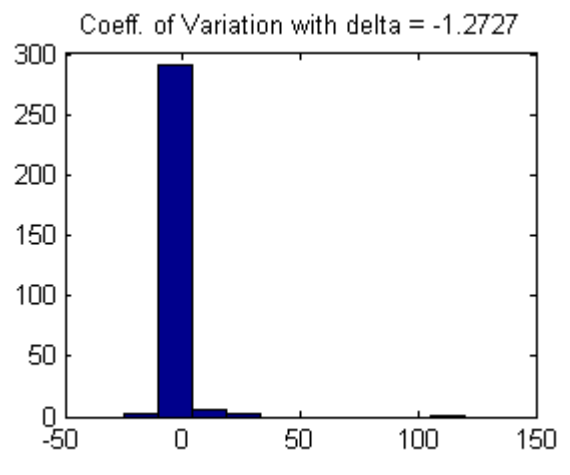


FIGURE 23.

FIGURE 24.

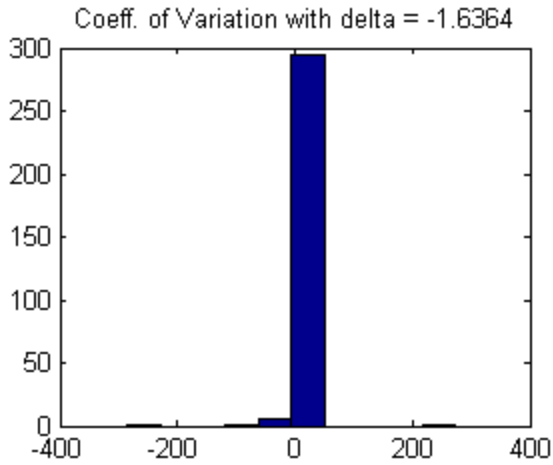


FIGURE 25.

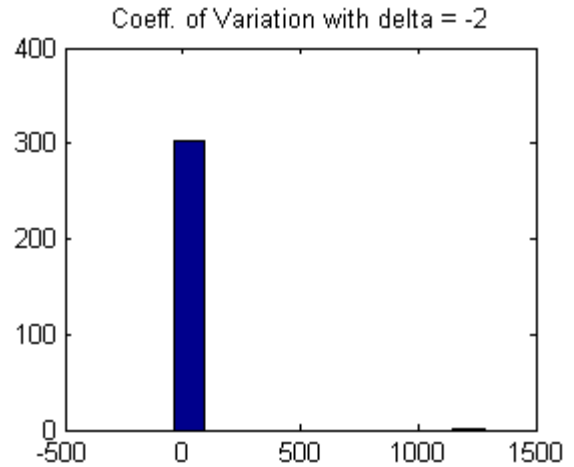


FIGURE 26.

Overall the majority of coefficients of variance are rather small. The spread of these variances, however, is quite different. For larger (and more positive) δ , the distribution of these variances tends to have very little variation and simultaneously, tends to have very small values. On this basis, δ above 0.90 seems to be the optimal range where the coefficient of variation of trajectories is minimized. Negative values of δ have the opposite effect with a few trajectories displaying very large variance. Most likely, these outliers are the result of a small estimated volatility.

Finally, comparing the empirical variance with the volatility estimates derived via Metropolis-Hastings can be used to determine how well the estimates fit. By totaling the squares of these differences, the following table was generated:

Delta	Volatility Error
-2	38.19
-1.6364	37.56
-1.2727	37.58
-0.9091	37.64
-0.5455	39.75
-0.1818	41.70
0.1818	44.29

Delta	Volatility Error
0.5455	46.49
0.9091	48.25
1.2727	49.86
1.6364	51.26
2	52.01

TABLE 2. MEAN SQUARE VOLATILITY ERROR FOR EACH FITTED MODEL

Clearly, more negative values of δ result in estimated volatilities that fit the empirical distribution better. The causal relationship here is questionable, however. The empirical distribution of volatility is skewed far-right and extends into the teens, as seen below:

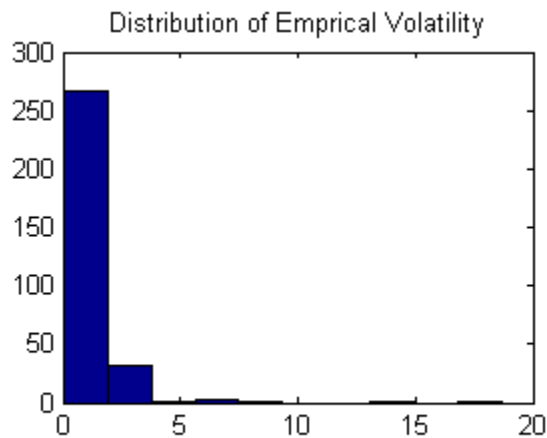


FIGURE 27.

Looking at a few representative iterative distributions with various values for δ , it's clear that there is a large difference:

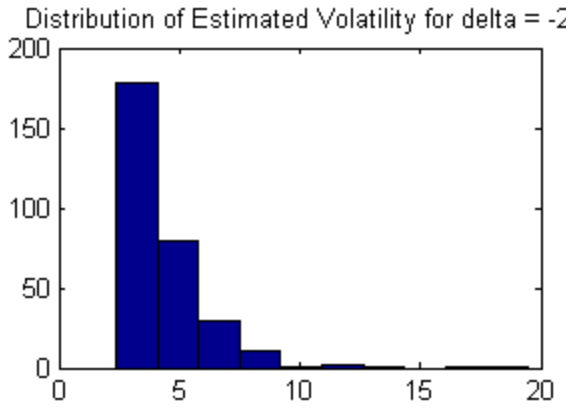


FIGURE 28.

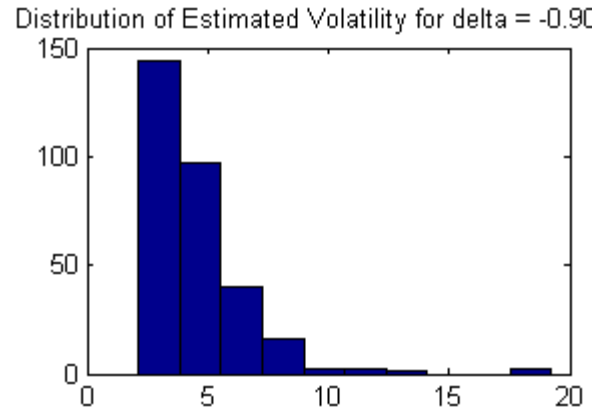


FIGURE 29.

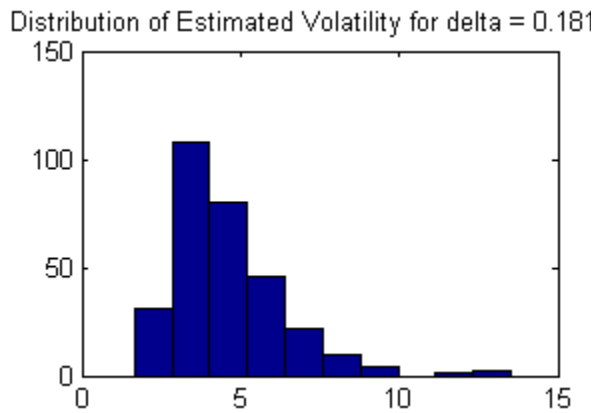


FIGURE 30.

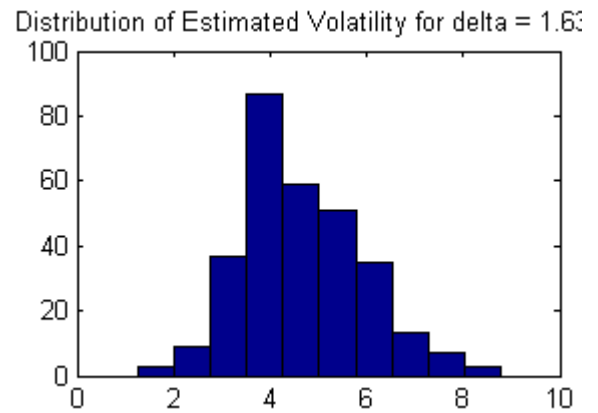


FIGURE 31.

The following chart quantifies the differences between these distributions and the empirical

one:

	Mean	Standard Deviation	Skewness	Kurtosis
Empirical	1.00	1.59	7.05	68.26
Delta				
-2	0.87	0.39	3.39	21.00
-1.6364	0.88	0.40	3.49	22.33
-1.2727	0.89	0.41	3.27	20.24
-0.9091	0.90	0.41	3.17	19.44
-0.5455	0.91	0.39	2.58	14.21
-0.1818	0.92	0.37	2.11	10.59
0.1818	0.92	0.34	1.64	7.49

0.5455	0.92	0.31	1.26	5.55
0.9091	0.92	0.29	0.97	4.46
1.2727	0.93	0.27	0.70	3.75
1.6364	0.93	0.25	0.50	3.29
2	0.94	0.23	0.43	3.24

TABLE 3. A COMPARISON OF KEY QUANTITIES FOR THE EMPIRICAL AND FITTED VOLATILITY DISTRIBUTIONS

The difference in skewness and kurtosis is especially wide. Because these values really address the underlying purpose of the modeling, which is to fit the empirical distribution, there is strong evidence to support the selection of a large negative value for δ , possibly more negative than -2.

4 CONCLUSION

An attempt was made to fit a stochastic volatility model with a Box-Cox transformation applied to the squared volatilities. First the one day grouping model and then the five day grouping model was used to explore the applicability of the model. The Metropolis-Hastings algorithm enabled sampling from joint posterior distribution successfully and with large jumping frequencies. In the final assessment, the distribution of the estimated volatilities supplied by the model was compared to the empirical volatilities leading to a recommendation of a large, negative value for δ .

The inconsistency between the shapes of the empirical distribution and the estimated volatilities suggests a systemic issue in the model. Despite seemingly immediate stabilization of the α trajectories, it is possible that more iteration would have significantly altered the sampled distributions and autocorrelations. Extending the model to include the correlation and heavy-tailed features of Zhang and King (2008) as well could also be fruitful.

5 REFERENCES

Box, G. E. P. and D. R. Cox, (1964) 'An Analysis of Transformations'. *Journal of Royal Statistical Society B*, 26: 211-243.

Blanco, Carlos & David Soronow (2001). 'Mean Reverting Processes - Energy Price Processes Used For Derivatives Pricing & Risk Management'. *Commodities Now*. June.

Bollerslev, T. (1986). 'Generalized autoregressive conditional heteroscedasticity'. *Journal of Econometrics* 31: 307-327.

Davis, Mark and Etheridge, Alison 'Louis Bachelier's Theory of speculation: The origins of modern finance', Princeton University Press, Princeton, NJ (2006).

Hentschel, Ludger (1995). 'All in the family: Nesting symmetric and asymmetric GARCH models'. *Journal of Financial Economics* 39: 71-104.

M.L. Higgins and A. Bera (1992). 'A class of nonlinear ARCH models'. *International Economic Review* 62: 137-158.

MacKenzie, D. (2003). 'An Equation and its Worlds: Bricolage, Exemplars, Disunity and Performativity in Financial Economics'. *Social Studies of Science* 33(6): 831-68.

Rubinstein, Mark (1994). 'Implied Binomial Trees'. *Journal of Finance* 49: 771-818.

Zhang, Xibin & King, Maxwell L. (2008). 'Box-Cox stochastic volatility models with heavy-tails and correlated errors'. *Journal of Empirical Finance*, Elsevier, vol. 15(3), pages 549-566, June.