Towards a Bayesian Method to Estimate Future Realized Volatility

Abstract

Estimation of ex-ante or future realized volatility is crucial to the financial industry. In this paper we explore the use of a Bayesian method to estimate the implied volatility on stock options which in turn will allow us to estimate the future realized volatility on the underlying. We find that this method is more accurate in estimating the implied volatility than estimates using historical volatility. Thus this method might be helpful in estimating ex-ante volatility for the underlying stock and will therefore be useful in pricing derivatives which does not have any implied volatility data on them.

1 Background and Significance

The derivatives market has grown dramatically over the past 20 years. In 1998 the notional value of outstanding OTC derivatives were estimated to be around \$60 trillion and at the end of June 2018 the same value was estimated to be around \$595 trillion[1].

Most research in the field has been based on the classical paradigm, although some Bayesian analyses have been done. Most notably Boyle and Ananthanarayanan (1977) who created credible intervals for the Black-Scholes option price and Karolyi (1993) who showed how the use of Bayesian methods can improve the precision of price estimations of options.

This paper seeks to further develop a method to better estimate the the price of derivatives where the implied volatility cannot be observed through using a Bayesian method to better estimate the implied volatility data for the underlying asset in question.

At the foundation of this paper lies the notion that implied volatility used in the appropriate manner will provide a better estimation of future volatility of the stocks. This notion is supported by many academics in the financial field, most notably Latané and Rendleman (1976), Chiras and Manaster (1978), and Beckers (1981).

2 Methods

2.1 Data

The underlying asset we picked is the SPDR SP 500 ETF(SPY) which is an ETF that is designed to track the S&P 500 index. We will be focusing on the 30-day historical volatility. The implied volatility data comes from the VIX, the CBOE Volatility Index, which is a measure of anticipated movements in the S&P 500, derived from the current traded prices of S&P 500 options.

The time frame for the data is from November 1st 2017 to April 10th 2019.

2.2 Choice of Likelihood Distribution

We assume that the daily returns are log-normally distributed and centered at zero, which is in accordance with Karolyi (1993) and Ho et al. (2011). Thus our likelihood distribution for the log returns looks as follows:

$$y_i \sim \text{Normal}(0, \sigma_i^2)$$
$$L(\sigma_i^2|y_1, ..., y_n) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-y_i^2}{2\sigma_i^2}\right)$$

Where y_i is the log-return on day i and σ_i^2 is the 30-day realized volatility for the same day.

2.3 Prior Distributions

We used three different prior distributions to see which would produce values most similar to the actual implied volatility: an Inverse Gamma, a Gamma, and a truncated Normal prior. For all of our priors we used weakly informative priors to let the data drive the posterior.

2.3.1 Inverse Gamma Prior

The Inverse Gamma prior distribution is conjugate with the Normal sampling distribution and has thus been commonly used in estimating the volatility in asset returns, most notably by Karolyi (1993).

$$\sigma_i^2 \sim \text{InverseGamma}(\alpha, \beta)$$
$$\pi(\sigma_i^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma_i^2)^{(-\alpha-1)} \exp\left(\frac{-\beta}{\sigma_i^2}\right)$$

2.3.2 Gamma Prior

The Gamma distribution although not conjugate with the Normal distribution was used by Ho et al. (2011) and was shown to be a better fit for

[1] https://www.bis.org/statistics/derstats.htm

modeling asset return volatility than the Inverse Gamma prior Ho et al. (2011).

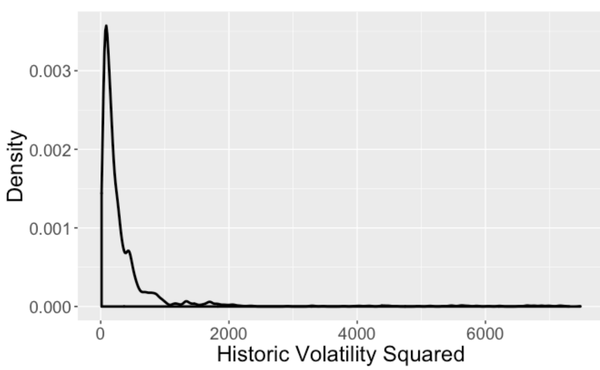
$$\sigma_i^2 \sim \text{Gamma}(\alpha, \beta)$$
$$\pi(\sigma_i^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma_i^2)^{(\alpha-1)} \exp\left(\frac{-\beta}{\sigma_i^2}\right)$$

2.3.3 Truncated Normal Prior

Our motivation for using this distribution as a prior is because it closely resembles the data of the historical volatility data, as can be seen in the following graph. Another reason is because volatility cannot be negative.

$$\sigma_i^2 \sim \text{Normal}(\mu, \tau)\ T(0,)$$
$$\pi(\sigma_i^2) = \frac{\sqrt{2}}{\sqrt{\pi}\tau}\left(erf\left(\frac{\mu}{\sqrt{2}\tau}\right) - 1\right)\exp\left(-\frac{(\sigma_i^2 - \mu)^2}{2\tau^2}\right)$$

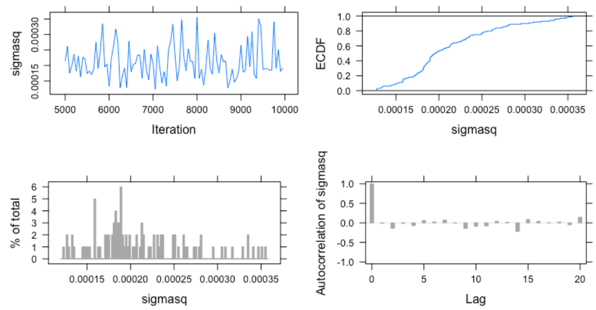
Where erf is the Guass error function.



2.4 Posterior Distributions

We used Just Another Gibbs Sampler (JAGS)[2] to get draws from our posterior distributions. To get estimates of the volatility for each day we ran the MCMC with the previous 30 days' log returns and took a sample of 100 draws and annualized them. From these draws we recorded the mean and median values.

[2] http://mcmc-jags.sourceforge.net/



By examining a longer chain we see that it has already converged at this point and it clearly has low auto-correlation. We saw similar MCMC diagnostics for our other prior choices.

2.5 Calculation of Estimation Error

Since we are investigating if a Bayesian estimate can better approximate the implied volatility than the historical volatility. Thus, the benchmark for the estimation error is the implied volatility data.

We used the following equations to estimate the estimation error where σ_i^B is the Bayesian estimated volatility for day i, σ_i^H is the historical volatility for day i, and σ_i^I is the implied volatility for day i.

$$\text{Bayesian Error} = \left(\frac{\sigma_i^B - \sigma_i^I}{\sigma_i^I}\right)^2$$
$$\text{Historical Error} = \left(\frac{\sigma_i^H - \sigma_i^I}{\sigma_i^I}\right)^2$$

The Bayesian error is the error between the Bayesian estimated volatility and the implied volatility and the historical error is the error between the historical volatility and the implied volatility.

The lower the error the better the performance of the estimator in question.

3 Results

The results of the average error for each estimating method are displayed in the table below. As seen in the table, the Inverse Gamma prior performed significantly worse compared to the historical volatility data, whereas the Gamma prior performed better than the historical data, which is in line with what was observed by Ho et al. (2011). Notably, the truncated Normal prior turned out to perform as well, up to three decimal places, as the Gamma prior.

Average Data Error

Inverse Gamma	Gamma	Normal	Historical Error
625	0.097	0.097	0.104

A more comprehensive table is found in the appendix.

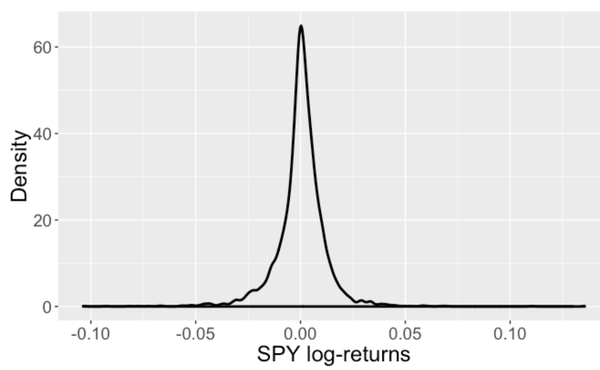
The significance of the improvement in estimating implied volatility compared to the historical volatility is 6.29% for both the Gamma and truncated Normal priors, as computed by the formulas for estimation error shown in Section 2.5.

4 Discussion

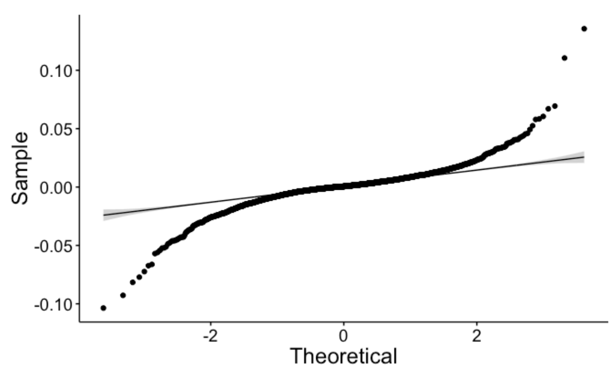
The results in this paper show that Bayesian volatility estimates with a Gamma or a truncated Normal prior provide around a 6.29% more accurate estimation of the implied volatility compared to using historical volatility estimates. Thus, given the previous research concluding that implied volatility can be used to provide better estimates of ex-ante stock price volatility, using these methods may provide more accurate pricing of derivatives where there is no data on implied volatility available. An example of such a derivative might be employee stock option schemes for many US and offshore companies that does not have any publicly traded options.

4.1 Future Research

At the base of this paper lies the assumption that stock returns are Gaussian. This notion has been widely disproven, most notably by Lisa Borland, who goes on to show that a more accurate distribution to use would be a Tsallis distribution rather than a Normal distribution for the likelihood Borland (2002). Below is the log-returns of the SPDR SP 500 ETF:



Next is a graph showing how this data fails the Shapiro-Wilk test, which further shows how the data is non-Normal in nature:



Thus an area for future research is implementing a Tsallis distribution as the distribution for the likelihood. Further areas of future research include using a mixture model for the prior distribution.

References

Beckers, S. (1981). Standard deviations implied in option prices as predictors of future stock price variability. *Journal of Banking Finance*, 5(3):363–381.

Borland, L. (2002). Option pricing formulas based on a non-gaussian stock price model. *Physical review letters*, 89(9):098701.

Boyle, P. P. and Ananthanarayanan, A. L. (1977). The impact of variance estimation in option valuation models. *Journal of Financial Economics*, 5(3):375–387.

Chiras, D. P. and Manaster, S. (1978). The information content of option prices and a test of market efficiency. *Journal of Financial Economics*, 6(2-3):213–234.

Ho, S., Lee, A., and Marsden, A. (2011). Use of bayesian estimates to determine the volatility parameter input in the black-scholes and binomial option pricing models. *Journal of Risk and Financial Management*, 4(1):74–96.

Karolyi, G. A. (1993). A bayesian approach to modeling stock return volatility for option valuation. *The Journal of Financial and Quantitative Analysis*, 28(4):579.

Latané, H. A. and Rendleman, R. J. (1976). Standard deviations of stock price ratios implied in option prices. *The Journal of Finance*, 31(2):369–381.

5 Appendix

5.1 Table for Section 3

Date	InverseGamma	Gamma	Normal	Historical Error
Nov 1, 2017	1493	0.099	0.094	0.143
Dec 2, 2017	554	0.160	0.127	0.135
Jan 2, 2018	298	0.000	0.000	0.002
Feb 2, 2018	834	0.010	0.018	0.019
Mar 5, 2018	569	0.142	0.158	0.179
Apr 5, 2018	690	0.181	0.174	0.189
May 6, 2018	889	0.194	0.188	0.244
Jun 6, 2018	499	0.089	0.080	0.070
Jul 7, 2018	98	0.113	0.103	0.152
Aug 7, 2018	534	0.205	0.197	0.072
Data Average	625	0.097	0.097	0.104