

# Volatility Forecasting II: Stochastic Volatility Models and Empirical Evidence

Rob Reider

October 26, 2009

## Introduction to Stochastic Volatility Models

Assume that returns on an asset are given by  $r_t = \mu + \sigma_t \epsilon_t$  as we did last week. In GARCH-type models, volatility,  $\sigma_t$ , is time varying but not stochastic. In stochastic volatility models, part of the changes in volatility are due to random shocks. To compare the differences, consider an EGARCH(1,1)-type model but with no asymmetries:

$$\ln(\sigma_t) = \alpha_0 + \beta_1 \ln(\sigma_{t-1}) + \alpha_1 (r_{t-1} - \mu)^2$$

and compare this with a typical stochastic volatility (SV) model:

$$\ln(\sigma_t) = \alpha_0 + \beta_1 \ln(\sigma_{t-1}) + v_t$$

where  $v_t$  is an i.i.d. random variable and  $\{v_t\}$  and  $\{\epsilon_t\}$  are stochastically independent.

A more standard way to represent the stochastic volatility process is

$$\ln(\sigma_t) = \alpha + \phi(\ln(\sigma_{t-1}) - \alpha) + \eta_t$$

so that  $\ln(\sigma_t)$  is an AR(1) process, where  $\phi$  is a parameter that represents how quickly volatility gets pulled toward its mean,  $\alpha$ . If  $\eta_t$  is normally distributed with mean 0 and variance  $\sigma_\eta^2$ , then  $\ln(\sigma_t)$  is normally distributed, and  $\sigma_t$  therefore has a lognormal distribution. To get the unconditional mean and variance of  $\ln(\sigma_t)$ ,

$$E[\ln(\sigma_t)] = \alpha + \phi(E[\ln(\sigma_{t-1})] - \alpha) + E[\eta_t]$$

and since  $\ln(\sigma_t)$  is a stationary process,

$$E[\ln(\sigma_t)] = \alpha.$$

For the unconditional variance

$$\begin{aligned} \text{Var}(\ln(\sigma_t)) &= E[(\ln(\sigma_t) - \alpha)^2] \\ &= \phi^2 E[(\ln(\sigma_{t-1}) - \alpha)^2] + E[\eta_t^2] \\ &= \phi^2 \text{Var}(\ln(\sigma_{t-1})) + \text{Var}(\eta_t) \end{aligned}$$

and by stationarity,

$$\text{Var}(\ln(\sigma_t)) = \frac{\sigma_\eta^2}{1 - \phi^2}$$

so  $\ln(\sigma_t) \sim N(\alpha, \beta^2)$  where  $\beta^2 = \sigma_\eta^2/(1 - \phi^2)$ .

Like GARCH models, stochastic volatility models produce returns with kurtosis  $> 3$  and positive autocorrelations between squared excess returns (we will derive the autocorrelations of squared returns and kurtosis of returns in later sections). Unlike GARCH models, however, there are no closed form solutions for the likelihood function. For example, the density function of  $r_t$  is

$$f(r_t) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(r_t - \mu)^2}{2\sigma_t^2}\right) g(\sigma_t|\alpha, \beta^2) d\sigma_t$$

where  $g(\cdot)$  is the lognormal density function. This integral can only be evaluated numerically, so estimating the parameters of a stochastic volatility model are much more complex than that of a GARCH-type model. Taylor describes several methods for estimating the parameters, most of which are beyond the scope of this class. We will briefly describe one technique, the Generalized Method of Moments (GMM), which is often used when maximum likelihood techniques cannot be employed.

## Classical Method of Moments

Before we describe the Generalized Method of Moments, let's digress and review the classical Method of Moments. Suppose we had  $N$  observations of a random variable,  $X, X_1, X_2, \dots, X_N$ . We are interested in estimating a parameter,  $\theta$ , from the distribution of  $X$ . If the parameter is a known function of the moments of  $X$ ,

$$\theta = f(E[X], E[X^2], \dots, E[X^k])$$

then

$$\hat{\theta}_{MM} = f(m_1, m_2, \dots, m_k)$$

is the method of moments estimate of  $\theta$ , where  $m_j = (1/N) \sum_{i=1}^N X_i^j$  is the sample mean of  $X_i^j$ , or the  $j$ th order sample moment. If we are estimating a vector of  $P$  parameters, we would solve  $P$  moment equations simultaneously.

Going back to our stochastic volatility example where  $\ln(\sigma_t) \sim N(\alpha, \beta^2)$  and  $\ln(\sigma_t) = \alpha + \phi(\ln(\sigma_{t-1}) - \alpha) + \eta_t$ , we can use the method of moments to estimate  $\alpha$  and  $\beta$ , but it turns out there is no clear method to estimate  $\phi$ , the rate of mean reversion parameter. Recall that if  $X$  is a lognormal random variable so that  $\ln(X) \sim N(\mu, \sigma^2)$ ,  $E[X] = E[e^{\ln(X)}] = \exp(\mu + \frac{1}{2}\sigma^2)$ . We will compute moments for  $r_t = \mu + \sigma_t \epsilon_t$  (and not  $\sigma_t$ , which is unobservable)

$$\begin{aligned} \text{Var}(r_t) &= E[(r_t - \mu)^2] \\ &= E[\sigma_t^2 \epsilon_t^2] \\ &= E[\sigma_t^2] E[\epsilon_t^2] \\ &= \exp(2\alpha + 2\beta^2) \end{aligned}$$

and

$$\begin{aligned}
\text{kurt}(r_t) &= \frac{E[(r_t - \mu)^4]}{(E[(r_t - \mu)^2])^2} \\
&= \frac{E[\sigma_t^4]E[\epsilon_t^4]}{(E[\sigma_t^2])^2(E[\epsilon_t^2])^2} \\
&= \frac{3 \exp(4\alpha + 8\beta^2)}{\exp(4\alpha + 4\beta^2)} \\
&= 3 \exp(4\beta^2)
\end{aligned}$$

so the method of moment estimate of  $\beta$  is

$$\begin{aligned}
m'_4 &= 3 \exp(4\hat{\beta}_{MM}^2) \\
\hat{\beta}_{MM}^2 &= \frac{1}{4} \ln \left( \frac{m'_4}{3} \right)
\end{aligned}$$

and the method of moment estimate of  $\alpha$  is

$$\begin{aligned}
m'_2 &= \exp(2\hat{\alpha}_{MM} + 2\hat{\beta}_{MM}^2) \\
\hat{\alpha}_{MM} &= \frac{1}{2} \ln(m'_2) - \hat{\beta}_{MM}^2
\end{aligned}$$

where  $m'_j = (1/N) \sum_{i=1}^N (X_i - \bar{X})^j$  ( $j$ th order sample *central* moments). But as we noted above, because of the difficulty of estimating  $\phi$ , we resort to the Generalized Method of Moments.

## Generalized Method of Moments

In this section, we will describe the Generalized Method of Moments (GMM) for estimating the parameters of a distribution for situations when we cannot derive closed form formulas for the likelihood function. Much of this section follows from Hamilton (1994) chapter 14 and Campbell, Lo, and MacKinlay (1997) Appendix A.2-A.3.

In GMM, we derive a set of  $M$  moment conditions,

$$E[\mathbf{f}_t(r_t, \theta_0)] = \mathbf{0}$$

where  $\mathbf{f}_t$  is a vector with  $M$  elements and  $\theta_0$  is the set of true parameters. We compute the corresponding sample averages of  $\mathbf{f}_t$

$$\mathbf{g}(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t(r_t, \hat{\theta})$$

where the number of moments,  $M$ , is larger than the number of parameters,  $P$ . Since we have more equations than parameters, we cannot set all the equations equal to 0, but we can

try to find parameters to minimize the sample moments. In GMM, we actually minimize the scalar

$$Q(\theta) = \mathbf{g}'(\theta) \mathbf{W} \mathbf{g}(\theta)$$

where  $\mathbf{W}$  is an  $M \times M$  weighting matrix that puts less weight on noisy moment conditions.

Next, we will describe how to estimate the weighting matrix,  $\mathbf{W}$ . Let  $\mathbf{S}$  be the asymptotic covariance of the moment conditions

$$\begin{aligned} \mathbf{S} &= \lim_{T \rightarrow \infty} T \text{Var}(\mathbf{g}(\theta_0)) \\ &= \lim_{T \rightarrow \infty} T E[\mathbf{g}(\theta_0) \mathbf{g}'(\theta_0)] \\ &= \lim_{T \rightarrow \infty} (1/T) E \left[ \sum_{t=1}^T \mathbf{f}_t(r_t, \theta_0) \sum_{t=1}^T \mathbf{f}_t(r_t, \theta_0)' \right]. \end{aligned}$$

The optimal weighting matrix turns out to be the inverse of the covariance matrix,  $\mathbf{W} = \mathbf{S}^{-1}$ . Intuitively, if a moment condition is noisier, less weight is placed on it relative to a less noisy moment condition.

If  $\mathbf{f}_t$  is serially uncorrelated, we could estimate the covariance matrix,  $\hat{\mathbf{S}}$ , as

$$\hat{\mathbf{S}} = \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t(r_t, \theta_0) \mathbf{f}_t(r_t, \theta_0)'$$

but since this requires us to know the true parameters,  $\theta_0$ , an iterative approach is often taken where we start out with equal weights,  $\mathbf{W}^{(1)} = \mathbf{I}$  to get an initial estimate of  $\theta_0$ , call it  $\hat{\theta}^{(1)}$ . From that initial estimate, we reestimate the covariance matrix  $\mathbf{S}$  and take the inverse to get an updated weighting matrix,  $\mathbf{W}^{(2)}$ . From this updated weighting matrix, we update our estimate of the parameters,  $\hat{\theta}^{(2)}$ , and we can continue to iterate until the changes in the parameter estimates from iteration to iteration are less than a certain tolerance.

If  $\mathbf{f}_t$  is serially correlated, the estimate of  $\mathbf{S}$  is more complex. It would be tempting to estimate  $\hat{\mathbf{S}}$  with sample covariances

$$\begin{aligned} \hat{\mathbf{S}} &= \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t(r_t, \theta_0) \mathbf{f}_t(r_t, \theta_0)' + \frac{1}{T} \sum_{t=2}^T \mathbf{f}_t(r_t, \theta_0) \mathbf{f}_{t-1}(r_{t-1}, \theta_0)' + \frac{1}{T} \sum_{t=2}^T \mathbf{f}_{t-1}(r_{t-1}, \theta_0) \mathbf{f}_t(r_t, \theta_0)' \\ &\quad + \cdots + \frac{1}{T} \sum_{t=T}^T \mathbf{f}_t(r_t, \theta_0) \mathbf{f}_{t-T+1}(r_{t-T+1}, \theta_0)' + \frac{1}{T} \sum_{t=T}^T \mathbf{f}_{t-T+1}(r_{t-T+1}, \theta_0) \mathbf{f}_t(r_t, \theta_0)' \end{aligned}$$

which, if we define the sample covariances as

$$\mathbf{\Gamma}_{j,T} \equiv \frac{1}{T} \sum_{t=j+1}^T \mathbf{f}_t(r_t, \theta_0) \mathbf{f}_{t-j}(r_{t-j}, \theta_0)'$$

can be written as the sum of sample covariances

$$\hat{\mathbf{S}} = \mathbf{\Gamma}_{0,T} + \sum_{j=1}^T (\mathbf{\Gamma}_{j,T} + \mathbf{\Gamma}_{j,T}').$$

However, the two problems with this estimate of  $\mathbf{S}$  is that there is no guarantee that  $\mathbf{S}$  will be positive definite in small samples. The second problem is that with a finite sample size and estimating a large number of covariance matrices, the estimator of  $\mathbf{S}$  may not be consistent.

The estimator of Newey and West (1987) is often used to deal with these problems by attenuating and cutting off the number of covariances to be estimated

$$\hat{\mathbf{S}} = \mathbf{\Gamma}_{0,T} + \sum_{j=1}^q \frac{q-j}{q} (\mathbf{\Gamma}_{j,T} + \mathbf{\Gamma}'_{j,T})$$

which guarantees both consistent estimates and positive definite matrices.

## Using GMM to Estimate Stochastic Volatility Parameters

We will now finally apply GMM to estimating the parameters of the stochastic volatility model. Andersen and Sorensen (1996), based on a Monte Carlo simulations study, address several issues related to using GMM to estimate the parameters of this particular stochastic volatility model. One issue is how many moment conditions to use. You may think that by increasing the number of moment conditions, you are using additional information, which, if weighted appropriately, cannot make the parameter estimates worse. But the weighting matrix,  $\mathbf{W}$ , must itself be estimated, and with  $M$  moment conditions, we need to estimate  $M(M+1)/2$  elements of  $\mathbf{W}$ , and a larger number of moment conditions could lead to poorer estimates of  $\mathbf{W}$  and worse estimates of the parameters. Another issue in general with GMM is that there is not much guidance on which moment conditions to use. For the stochastic volatility model, moment conditions can be based on infinitely many functions of returns (see Melino and Turnbull (1990)), such as

$$\begin{aligned} E[|r_t|^i] & i = 1, 2, 3, \dots \\ E[|r_t||r_{t-i}|] & i = 1, 2, 3, \dots \\ E[r_t^2 r_{t-i}^2] & i = 1, 2, 3, \dots \end{aligned}$$

For this particular model, however, Andersen and Sorensen suggest, based on their Monte Carlo analysis, the following 24 moment conditions:

$$\begin{aligned} g_i(\hat{\theta}) &= \frac{1}{T} \sum_{t=1}^T |r_t - \bar{r}|^i - E[\sigma_t^i] E[|\epsilon_t^i|] & i = 1, \dots, 4 \\ g_{i+4}(\hat{\theta}) &= \frac{1}{T} \sum_{t=1}^T |r_t - \bar{r}| |r_{t-i} - \bar{r}| - E[\sigma_t \sigma_{t-i}] E[|\epsilon_t|] E[|\epsilon_{t-i}|] & i = 1, \dots, 10 \\ g_{i+14}(\hat{\theta}) &= \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2 (r_{t-i} - \bar{r})^2 - E[\sigma_t^2 \sigma_{t-i}^2] E[\epsilon_t^2] E[\epsilon_{t-i}^2] & i = 1, \dots, 10 \end{aligned}$$

where we can derive analytical solutions for  $E[\sigma_t^i]$ ,  $E[|\epsilon_t|^i]$ ,  $E[\sigma_t \sigma_{t-i}]$  and  $E[\sigma_t^2 \sigma_{t-i}^2]$ .

Since  $\ln(\sigma_t) \sim N(\alpha, \beta^2)$  and  $\ln(\sigma_t^i) = i \ln(\sigma_t)$ , then  $\ln(\sigma_t^i) \sim N(i\alpha, i^2\beta^2)$  and

$$E[\sigma_t^i] = \exp(i\alpha + \frac{1}{2}i^2\beta^2).$$

For a standard normal random variable,  $\epsilon_t$ , it can be shown that

$$E[|\epsilon_t|^i] = \frac{2^{i/2}}{\sqrt{\pi}} \Gamma\left(\frac{i+1}{2}\right)$$

so for  $i = 1$ , for example,

$$E[|\epsilon_t|] = \sqrt{2/\pi}.$$

To derive  $E[\sigma_t \sigma_{t-i}]$ , note that  $\ln(\sigma_t) \sim N(\alpha, \beta^2)$  and  $\ln(\sigma_{t-i}) \sim N(\alpha, \beta^2)$  and since  $\ln(\sigma_t)$  is an AR(1) process,  $\text{corr}(\ln(\sigma_t), \ln(\sigma_{t-i})) = \phi^i$ . Therefore,  $E[\ln(\sigma_t) + \ln(\sigma_{t-i})] = 2\alpha$  and  $\text{Var}(\ln(\sigma_t) + \ln(\sigma_{t-i})) = \beta^2 + \beta^2 + 2\beta^2(1 + \phi^i)$  so  $\ln(\sigma_t) + \ln(\sigma_{t-i}) = \ln(\sigma_t \sigma_{t-i}) \sim N(2\alpha, 2\beta^2(1 + \phi^i))$  and

$$E[\sigma_t \sigma_{t-i}] = \exp(2\alpha + \beta^2(1 + \phi^i)).$$

Along the same lines,  $\ln(\sigma_t^2 \sigma_{t-i}^2) \sim N(4\alpha, 8\beta^2(1 + \phi^i))$ , so

$$E[\sigma_t^2 \sigma_{t-i}^2] = \exp(4\alpha + 4\beta^2(1 + \phi^i)).$$

## Asymmetric Stochastic Volatility Models

In the last class, we went over the evidence that there may be an asymmetric impact of returns on volatility. In equity markets, for example, large down moves in stock prices have a bigger impact on future volatility than large up moves. This effect can be modeled with EGARCH models, as well as other variations of GARCH models.

With stochastic volatility models, we can model the asymmetry by adding a correlation between today's stock return shock,  $\epsilon_t$ , and tomorrow's shock to the log of volatility,  $\eta_{t+1}$ . Thus, we can treat  $\epsilon_t, \eta_{t+1}$  as bivariate normal random variables with a correlation  $\delta$ , so that

$$\begin{pmatrix} \epsilon_t \\ \eta_{t+1} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \delta\sigma_\eta \\ \delta\sigma_\eta & \sigma_\eta^2 \end{pmatrix}\right).$$

When we add a correlation, the moment conditions we derived earlier will change. Consider the 15th moment condition

$$g_{15}(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2 (r_{t-1} - \bar{r})^2 - E[\sigma_t^2 \sigma_{t-1}^2 \epsilon_t^2 \epsilon_{t-1}^2]$$

The  $\epsilon_{t-1}^2$  is no longer independent of  $\sigma_t^2$ , so

$$\begin{aligned} E[\sigma_t^2 \sigma_{t-1}^2 \epsilon_t^2 \epsilon_{t-1}^2] &= E[\sigma_t^2 \epsilon_{t-1}^2 \sigma_{t-1}^2] E[\epsilon_t^2] \\ &= E[\sigma_t^2 \epsilon_{t-1}^2 \sigma_{t-1}^2] \\ &= E[\exp(\ln(\sigma_t^2) + \ln(\sigma_{t-1}^2)) \epsilon_{t-1}^2] \\ &= E[\exp(X) Y^2] \end{aligned}$$

where

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left( \begin{pmatrix} 4\alpha \\ 0 \end{pmatrix}, \begin{pmatrix} 8\beta^2(1+\phi) & 2\delta\sigma_\eta \\ 2\delta\sigma_\eta & 1 \end{pmatrix} \right).$$

In order to derive the expectation above, we can use the Moment Generating Function for a bivariate normal random variable,  $(X, Y)$ , which is

$$\begin{aligned} M(t_1, t_2) &\equiv E[\exp(t_1 X + t_2 Y)] \\ &= \exp \left( \mu_1 t_1 + \mu_2 t_2 + \frac{\sigma_1^2 t_1^2 + \sigma_2^2 t_2^2 + 2\rho\sigma_1\sigma_2 t_1 t_2}{2} \right) \end{aligned}$$

and

$$\begin{aligned} \left. \frac{\partial^2 M}{\partial t_2^2} \right|_{\substack{t_1=1 \\ t_2=0}} &= E[\exp(X)Y^2] \\ &= (\sigma_2^2 + (\mu_2 + \rho\sigma_1\sigma_2)^2) \exp(\mu_1 + \frac{\sigma_1^2}{2}) \\ &= (1 + 4\delta^2\sigma_\eta^2) \exp(4\alpha + 4\beta^2(1+\phi)). \end{aligned}$$

Notice that the correlation parameter,  $\delta$ , only appears as  $\delta^2$  in the moment condition. If we are to estimate the sign of  $\delta$ , we need to add at least one more moment condition where  $\delta$  appears unsquared. Consider an additional condition

$$g_{25}(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T |r_t - \bar{r}| (r_{t-1} - \bar{r}) - E[\sigma_t \sigma_{t-1} | \epsilon_t | \epsilon_{t-1}]$$

where

$$\begin{aligned} E[\sigma_t \sigma_{t-1} | \epsilon_t | \epsilon_{t-1}] &= E[\sigma_t \sigma_{t-1} \epsilon_{t-1}] E[|\epsilon_t|] \\ &= \sqrt{2/\pi} E[\exp(\ln(\sigma_t) + \ln(\sigma_{t-1})) \epsilon_{t-1}] \\ &= \sqrt{2/\pi} E[\exp(X)Y] \end{aligned}$$

where

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left( \begin{pmatrix} 2\alpha \\ 0 \end{pmatrix}, \begin{pmatrix} 2\beta^2(1+\phi) & \delta\sigma_\eta \\ \delta\sigma_\eta & 1 \end{pmatrix} \right).$$

We again use the Moment Generating Function for a bivariate normal random variable,  $(X, Y)$ , which is

$$\begin{aligned} \left. \frac{\partial M}{\partial t_2} \right|_{\substack{t_1=1 \\ t_2=0}} &= E[\exp(X)Y] \\ &= (\mu_2 + \rho\sigma_1\sigma_2) \exp(\mu_1 + \frac{\sigma_1^2}{2}) \\ &= (\delta\sigma_\eta) \exp(2\alpha + \beta^2(1+\phi)) \end{aligned}$$

so this moment condition (as well as other lags) can be used to determine  $\delta$ .

As you can see, when maximum likelihood techniques cannot be used, estimating the parameters becomes much more complex. Several other methods besides GMM have been proposed in the literature for estimating the stochastic volatility parameters, including Quasi Maximum Likelihood estimation, simulated method of moments, and Markov Chain Monte Carlo Methods. See Taylor for more information and references on these methods. One reason I chose to focus on GMM versus these other techniques is that GMM can be used in a variety of other time series applications, so it is a good tool to understand.

## Estimating Volatility Using High-Frequency Returns

Another approach to forecasting volatility is to use high-frequency, intraday data to estimate the daily "realized volatility". The intraday variance is defined as

$$INTRA_t^2 = \sum_{j=1}^n r_{t,j}^2$$

where  $r_{t,i}$  is the return on day  $t$  over interval  $j$  and  $n$  is the number of intervals in a day. The time series of intraday variances can then be used in a traditional time series model, like an ARFIMA model for forecasting variances with long memory, or it can be used to augment a traditional GARCH model. For example, in Blair, Poon, and Taylor (2001), they estimate a model that is roughly

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_t^2 + \gamma_1 S_{t-1} \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \alpha_2 INTRA_{t-1}^2 + \alpha_3 VIX_{t-1}^2$$

which combines the TGARCH model from the last class that uses daily squared excess returns,  $a_t^2$ , with the intraday volatility defined above and the VIX index of implied volatility. Using intraday volatility shows promising results, both in-sample and out-of-sample. In-sample tests based on standardizing returns by normalizing them with volatility estimates,  $\hat{\epsilon}_t = (r_t - \hat{\mu})/\hat{\sigma}_t$ , appear to look more like standard normal variables when  $\hat{\sigma}_t$  is estimated using intraday volatility than they do with GARCH volatility. Since high frequency models are relatively recent, there are only a few studies that have examined out-of-sample properties of these models, but there is some evidence that intraday volatility improves forecasts.

Using intraday volatility presents additional difficulties, however. In particular, market microstructure effects like the bid/ask bounce can bias the intraday volatility estimates. Some authors try to mitigate the bid/ask effects by filtering high frequency returns through an MA(1) model. Several authors suggest that a good tradeoff between sampling more frequently to get more data points and sampling less frequently to reduce bid/ask effects is to use a sampling interval of around five minutes. Thus, for foreign exchange, which trades 24 hours a day, the intraday volatility estimate would be based on 288 daily observations, and for equities, the intraday variance would be the sum of the squared return from the 4:00 pm close to the 9:30 am open plus the sum of the 78 five-minute squared returns from the 9:30 am open to the 4:00 pm close. Even with a five-minute sampling interval, Blair et al. (2001) find that the annualized intraday volatility is 13.4% but the volatility calculated



over the same period using close-to-close returns is 17.6%, so they estimate an additional proportionality factor that is used to adjust the intraday volatility.

## Empirical Evidence

In this section, we will look at how well these volatility forecasting techniques perform, relative to naive forecasts using only the historical standard deviation, relative to implied volatility, and relative to other forecasting methodologies.

In finance, it is rare to find a method that is consistently superior for forecasting the price of financial assets, and empirical studies are often inconclusive. To gain an appreciation for how difficult it is to evaluate the forecasting ability of various models, consider that the results can vary depending on the time period, the frequency of data, the asset class, the country within an asset class, and the forecast horizon. This has clearly been the case for papers on volatility forecasting, for there have been over a hundred academic papers that have attempted to measure and compare the performance of various volatility forecasting techniques. Many of these papers use different methodologies for merely measuring the accuracy of the forecasts. For example, authors use mean squared error, mean absolute error, root mean square error, and all of these applied to estimating  $\sigma$  and  $\sigma^2$ . Recently, some authors have started to use next period's realized variance instead of next period's squared returns as the target to which volatility forecasts are compared. And what is the best way to measure realized volatility? Is  $(1/T) \sum_{i=1}^T r_i^2$  better than  $(1/T) \sum_{i=1}^T (r_i - \bar{r})^2$ ? And if  $\hat{\sigma}^2 = (1/T) \sum_{i=1}^T (r_i - \bar{r})^2$  is an unbiased estimator of  $\sigma^2$ ,  $\sqrt{\hat{\sigma}^2}$  is a biased estimate of  $\sigma$  by Jensen's inequality. There is even contradictory evidence on whether GARCH models outperform some simple volatility forecasts, like using the historical volatility or an exponentially weighted version of historical volatility.

Many of the papers that do attempt to compare various volatility forecasting models cannot be used for drawing conclusions about option trading strategies. First of all, many of the tests in the literature are in-sample tests, like the test we described earlier for looking at the standardized residuals from a GARCH model to test for normality, or the Portmanteau tests on correlations of squared residuals. These tests fit the parameters with a set of data, and then use the same data to perform tests. An out-of-sample test involves estimating the parameters of a model using data up to time  $t$ , and then forecasting the variance at  $t + 1$ , or  $t + m$  for an  $m$ -step ahead forecast. Then the parameters would be reestimated using data up to time  $t + 1$ . This is called a rolling estimate. Many of the out-of-sample tests compare the forecast ability of implied volatility to that of a GARCH model, which is a useful test if you want to choose a source of volatility for risk management purposes, but is not a useful test for examining option trading strategies.

Keeping in mind all these warnings, Poon and Granger (2003) provide a review and summary of 93 other papers that examine the performance of volatility forecasting methodologies. Figure 1 below, which is reproduced from Poon and Granger, summarizes these studies. In the figure, HISVOL are methodologies that use historical volatility, including exponentially weighted historical volatility, GARCH includes ARCH, GARCH, EGARCH,

or any other GARCH variant, ISD are implied volatility forecasts, and SV are stochastic volatility forecasts.

	Number of Studies	Studies Percentage
HISVOL > GARCH	22	56%
GARCH > HISVOL	17	44%
HISVOL > ISD	8	24%
ISD > HISVOL	26	76%
GARCH > ISD	1	6
ISD > GARCH	17	94
SV > HISVOL	3	
SV > GARCH	3	
GARCH > SV	1	
ISD > SV	1	

Figure 1: A Comparison of Volatility Forecasting Methods from Poon and Granger (2003)

Notice from the Figure how poorly GARCH does compared with implied volatility. This does not imply that you cannot make money by trading GARCH vs. implied volatility. For example, suppose ISD=10%, GARCH=13%, and realized volatility=10.5%. ISD is more accurate than GARCH, but GARCH would lead you to buy volatility, and presumably, since realized volatility > implied volatility, you would make money if you bought an option and delta hedged it. Very few academic studies actually test the profitability of an options trading strategy. They often look at which model is a better predictor of tomorrow's volatility.

To test whether the GARCH-type models can be used as a volatility trading strategy, it must be compared with implied volatilities. But there may be a market price on volatility risk, which complicates any comparison between forecasted volatility and implied volatility. Just as stocks are risky and investors earn positive expected returns for holding stocks, volatility is a risk that cannot be hedged and may be biased compared with realized volatility. Moreover, implied volatility may be consistently higher than realized volatility because the market is pricing in the possibility of an important low probability event that has not yet occurred. This phenomenon is sometimes referred to as the peso problem. Between 1955 and 1975, the Mexican peso was fixed by the Mexican government at a constant exchange rate against the U.S. dollar, yet Mexican interest rates were consistently higher than U.S. rates. The rate differential was explained by the market's anticipation of a major devaluation of the Mexican peso, which eventually occurred in 1976. The same issue arises with implied volatility. Deep out-of-the-money puts have a consistently higher implied volatility than just about any volatility forecast, but perhaps the market is pricing in the small possibility of a stock market crash due to some unlikely event.

## Forecasting Implied Volatility

For trading strategies centered around shorter-dated options, we are interested in forecasting realized volatility over the life of the option. If, for example, our forecast for realized volatility is higher than the implied volatility, we could buy the option and delta hedge it to capture this difference. If we are correct and the realized volatility turns out to be higher than the implied volatility, our profits from gamma trading the option should exceed the costs of the option.

For longer-dated options, most of the gains or losses come from changes in implied volatility and not from gamma trading. Long-dated options, relative to short-date ones, have more vega (sensitivity of option price to changes in implied volatility) and less gamma (sensitivity of delta to changes in asset price). Thus, if you bought a one year option, for example, and held it for a few days, almost all the gains and losses would come from changes in implied volatility, which could be partly due to changes in the underlying asset but could also be due to other factors, like the rate at which implied volatility is mean reverting.

Although there are many papers that use implied volatility to try to forecast future realized volatility, very little has been written on modeling and forecasting implied volatility itself. A relatively recent working paper by Ahoniemi (2006) is an attempt to examine this issue. It tries to forecast changes in the VIX, which is a weighted average of put and call volatility for several strikes and two near-dated expirations. No attempt was made to forecast implied volatilities of longer-dated options. She uses an ARMA(1,1) model to forecast changes in the VIX, but augments it with GARCH forecasts as well as several exogenous variables.

## Problems With GARCH-type Models

Besides the poor forecasting ability of GARCH-type models, there are other problems with these models. First of all, parameters of GARCH models using different time scales give inconsistent results, which would not be the case if the model were properly specified. For example, if one used a GARCH model to forecast foreign exchange volatilities, the 2-day ahead forecast from a model estimated with daily data would give very different forecasts than a 48-hour ahead forecast from a model estimated with hourly data. Second, outliers and extremely large moves, like the 1987 stock market crash, reduce the performance. Also, when forecasting single stocks, GARCH models do not distinguish between large moves associated with earnings announcements and those associated with other news. Moreover, the empirical evidence is not clear on how well these models actually do in forecasting volatility. A vast majority of studies show that implied volatility is a better forecaster of future realized volatility than GARCH-type models. There is little or no evidence that these models can be profitably used for making bets on volatility through options.

Despite all these problems, think of this week and last week's lecture as an introduction to a class of models that can be improved upon. There is still much work to be done in generating volatility forecasts that lead to profitable trading strategies.

## References

- Ahoniemi, K. 2006. Modeling and forecasting implied volatility - an econometric analysis of the VIX index. Helsinki School of Economics.
- Andersen, T.G. and B.E. Sorensen. 1996. GMM estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Business and Economic Statistics* 14:328-352.
- Campbell, J.Y., A.W. Lo, and A.C. MacKinlay. 1997. *The econometrics of financial markets*. Princeton University Press.
- Blair, B.J., S.-H. Poon, and S.J. Taylor, 2001. Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high frequency index returns, *Journal of Econometrics* 105:5-26.
- Hamilton, J.D. 1994. *Times series analysis*. Princeton University Press.
- Melino, A. and S.M. Turnbull. 1990. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45:239-265.
- Newey, W.K. and K.D. West 1987. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703-708.
- Poon, S.-H. and C.W.J. Granger 2003. Forecasting volatility in financial markets. *Journal of Economic Literature* 41:478-539.