

Volatility Forecasting I: GARCH Models

Rob Reider

October 19, 2009

Why Forecast Volatility

The three main purposes of forecasting volatility are for risk management, for asset allocation, and for taking bets on future volatility. A large part of risk management is measuring the the potential future losses of a portfolio of assets, and in order to measure these potential losses, estimates must be made of future volatilities and correlations. In asset allocation, the Markowitz approach of minimizing risk for a given level of expected returns has become a standard approach, and of course an estimate of the variance-covariance matrix is required to measure risk. Perhaps the most challenging application of volatility forecasting, however, is to use it for developing a volatility trading strategy. Option traders often develop their own forecast of volatility, and based on this forecast they compare their estimate for the value of an option with the market price of that option. The simplest approach to estimating volatility is to use historical standard deviation, but there is some empirical evidence, which we will discuss later, that this can be improved upon.

In this class, we will start with a simple model for volatility, and gradually build up to more complicated models. Much of these lecture notes are taken from the two textbooks for the class, Tsay (chapter 3) and Taylor (chapters 8-12,15). The texts also provide references for further reading on many of the topics covered in class.

Some Stylized Facts About Volatility

One empirical observation of asset returns is that squared returns are positively autocorrelated. If an asset price like a currency, commodity, stock price, or bond price made a big move yesterday, it is more likely to make a big move today. For example, the autocorrelation of squared returns for the S&P 500 from January 1, 1987 to October 16, 2009 ($N = 5748$) is 0.13. The standard error on this estimate under the i.i.d. hypothesis is $1/\sqrt{N} = 0.013$, so 0.13 is 10 standard deviations away from zero. Similarly, the autocorrelation of squared returns for the \$-yen exchange rate from January 1, 1987 to October 16, 2009 is 0.14, or 11 standard deviations from zero.

Although the stock market crash on October 19, 1987 is an extreme example, we can see anecdotally that large moves in prices lead to more large moves. In Figure 1 below, we show the returns of the S&P 500 around the stock market crash of October 19, 1987. Before the stock market crash, the standard deviation of returns was about 1% per day. On October 19, the S&P 500 was down 20%, which as a 20 standard deviation move, would not be expected to occur in over 4.5 billion years (the age of the earth) if returns were truly

normally distributed. But in 4 of the 5 following days, the market moved over 4%, so it appears that volatility increased after the stock market crash, rather than remaining at 1% per day.

9 Index HCP
 Start document error [0x6BA]
Historical % Changes Page 1 / 2
 SPX S&P 500 INDEX PRICE 1526.75
 Range 9/30/87 to 10/30/87 Period Daily USD

	DATE	PRICE	CHANGE	% Change	4%	CHANGE	% Change
F	10/30/87	251.79	+7.02	+2.87			
T	10/29/87	244.77	+11.49	+4.93			
W	10/28/87	233.28	+0.09	+0.04			
T	10/27/87	233.19	+5.52	+2.42			
M	10/26/87	227.67	-20.55	-8.28			
F	10/23/87	248.22	-.03	-0.01			
T	10/22/87	248.25	-10.13	-3.92			
W	10/21/87	258.38	+21.55	+9.10			
T	10/20/87	236.83	+11.99	+5.33			
M	10/19/87	224.84	-57.86	-20.47			
F	10/16/87	282.70	-15.38	-5.16			
T	10/15/87	298.08	-7.15	-2.34			
W	10/14/87	305.23	-9.29	-2.95			
T	10/13/87	314.52	+5.13	+1.66			
M	10/12/87	309.39	-1.68	-0.54			
F	10/9/87	311.07	-3.09	-0.98			

Australia 61 2 9777 8600 Brazil 5511 3048 4500 Europe 44 20 7330 7500 Germany 49 69 920410
 Hong Kong 852 2377 6000 Japan 81 3 3201 8900 Singapore 65 6212 1000 U.S. 1 212 318 2000 Copyright 2007 Bloomberg L.P.
 6787-196-0 28-Sep-07 17:31:45

Figure 1: Stock returns after the October 19, 1987 stock market crash

Volatility not only spikes up during a crisis, but it eventually drops back to approximately the same level of volatility as before the crisis. Over the decades, there have been periodic spikes in equity volatility due to crises that caused large market drops (rarely are there large sudden up moves in equity markets), such as the Great Depression, Watergate, the 1987 stock market crash, Long Term Capital Management’s collapse in 1998, the September 11 terrorist attacks, and the bankruptcy of WorldCom in 2002. In foreign exchange markets, there was the Mexican Peso crisis in 1994, the East Asian currency crisis in 1997, and the EMS crises in 1992 and 1993. In all these cases, volatility remained high for a while, and then reverted to pre-crisis levels.

During the financial crisis last fall, the VIX index hit a high of 80%, and then gradually reverted over the last year back to a volatility of 21%. This is shown in Figure 2 below.

Another observation about returns is they exhibit excess kurtosis (the fourth moment of returns), or fatter tails, relative to a normal distribution. For example, for the S&P 500 from January 1, 1987 to October 16, 2009, the kurtosis is 33.0, whereas for a normal distribution the kurtosis is 3. The standard error on this estimate under the i.i.d. hypothesis is $\sqrt{24/N} = 0.065$. If we exclude the stock market crash of Oct 19, 1987 by looking at Jan 1, 1988 to October 16, 2009, the kurtosis drops to 12.4, which is still significantly larger than 3. For the \$-yen exchange rate, the kurtosis is 8.2.



Figure 2: Implied volatilities after the 2008 financial crisis

The models we look at will attempt to capture the autocorrelation of squared returns, the reversion of volatility to the mean, as well as the excess kurtosis.

Introducing a Simple ARCH(1) Model

The first and simplest model we will look at is an ARCH model, which stands for Autoregressive Conditional Heteroscedasticity. The AR comes from the fact that these models are autoregressive models in squared returns, which we will demonstrate later in this section. The conditional comes from the fact that in these models, next period's volatility is conditional on information this period. Heteroscedasticity means non constant volatility. In a standard linear regression where $y_i = \alpha + \beta x_i + \epsilon_i$, when the variance of the residuals, ϵ_i is constant, we call that homoscedastic and use ordinary least squares to estimate α and β . If, on the other hand, the variance of the residuals is not constant, we call that heteroscedastic and we can use weighted least squares to estimate the regression coefficients.

Let us assume that the return on an asset is

$$r_t = \mu + \sigma_t \epsilon_t$$

where ϵ_t is a sequence of $N(0,1)$ i.i.d. random variables. We will define the residual return at time t , $r_t - \mu$, as

$$a_t = \sigma_t \epsilon_t .$$

In an ARCH(1) model, first developed by Engle (1982),

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2$$

where $\alpha_0 > 0$ and $\alpha_1 \geq 0$ to ensure positive variance and $\alpha_1 < 1$ for stationarity. Under an ARCH(1) model, if the residual return, a_t is large in magnitude, our forecast for next period's conditional volatility, σ_{t+1} will be large. We say that in this model, the returns are conditionally normal (conditional on all information up to time $t - 1$, the one period returns are normally distributed). We will relax that assumption on conditional normality in a later section. Also, note that the returns, r_t , are uncorrelated but are not i.i.d.

We can see right away that a time varying σ_t^2 will lead to fatter tails, relative to a normal distribution, in the unconditional distribution of a_t (see Campbell, Lo, and Mackinlay(1997)). The kurtosis of a_t is defined as

$$\text{kurt}(a_t) = \frac{E[a_t^4]}{(E[a_t^2])^2}.$$

If a_t were normally distributed, it would have a kurtosis of 3. Here,

$$\begin{aligned} \text{kurt}(a_t) &= \frac{E[\sigma_t^4]E[\epsilon_t^4]}{(E[\sigma_t^2])^2(E[\epsilon_t^2])^2} \\ &= \frac{3E[\sigma_t^4]}{(E[\sigma_t^2])^2} \end{aligned}$$

and by Jensen's inequality (for a convex function, $f(x)$, $E[f(x)] > f(E[x])$), $E[\sigma_t^4] > (E[\sigma_t^2])^2$, so $\text{kurt}(a_t) > 3$.

Another intuitive way to see that models with time varying σ_t lead to fat tails is to think of these models as a mixture of normals. I will demonstrate this in class with some figures.

We'll discuss a few properties of an ARCH(1) model in particular. The unconditional variance of a_t is

$$\begin{aligned} \text{Var}(a_t) &= E[a_t^2] - (E[a_t])^2 \\ &= E[a_t^2] \\ &= E[\sigma_t^2 \epsilon_t^2] \\ &= E[\sigma_t^2] \\ &= \alpha_0 + \alpha_1 E[a_{t-1}^2] \end{aligned}$$

and since a_t is a stationary process, the $\text{Var}(a_t) = \text{Var}(a_{t-1}) = E[a_{t-1}^2]$, so

$$\text{Var}(a_t) = \frac{\alpha_0}{1 - \alpha_1}.$$

An ARCH(1) is like an AR(1) model on squared residuals, a_t^2 . To see this, define the conditional forecast error, or the difference between the squared residual return and our conditional expectation of the squared residual return, as

$$\begin{aligned} v_t &\equiv a_t^2 - E[a_t^2 | I_{t-1}] \\ &= a_t^2 - \sigma_t^2 \end{aligned}$$

where I_{t-1} is the information at time $t - 1$. Note that v_t is a zero mean, uncorrelated series. The ARCH(1) equation becomes

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 \\ a_t^2 - v_t &= \alpha_0 + \alpha_1 a_{t-1}^2 \\ a_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + v_t \end{aligned}$$

which is an AR(1) process on squared residuals.

Extension to a GARCH(1,1) model

In an ARCH(1) model, next period's variance only depends on last period's squared residual so a crisis that caused a large residual would not have the sort of persistence that we observe after actual crises. This has led to an extension of the ARCH model to a GARCH, or Generalized ARCH model, first developed by Bollerslev (1986), which is similar in spirit to an ARMA model. In a GARCH(1,1) model,

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

where $\alpha_0 > 0$, $\alpha_1 > 0$, $\beta_1 > 0$, and $\alpha_1 + \beta_1 < 1$, so that our next period forecast of variance is a blend of our last period forecast and last period's squared return.

We can see that just as an ARCH(1) model is an AR(1) model on squared residuals, an ARCH(1,1) model is an ARMA(1,1) model on squared residuals by making the same substitutions as before, $v_t = a_t^2 - \sigma_t^2$

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ a_t^2 - v_t &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 (a_{t-1}^2 - v_{t-1}) \\ a_t^2 &= \alpha_0 + (\alpha_1 + \beta_1) a_{t-1}^2 + v_t - \beta_1 v_{t-1} \end{aligned}$$

which is an ARMA(1,1) on the squared residuals.

The unconditional variance of a_t is

$$\begin{aligned} \text{Var}(a_t) &= E[a_t^2] - (E[a_t])^2 \\ &= E[a_t^2] \\ &= E[\sigma_t^2 \epsilon_t^2] \\ &= E[\sigma_t^2] \\ &= \alpha_0 + \alpha_1 E[a_{t-1}^2] + \beta_1 \sigma_{t-1}^2 \\ &= \alpha_0 + (\alpha_1 + \beta_1) E[a_{t-1}^2] \end{aligned}$$

and since a_t is a stationary process,

$$\text{Var}(a_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

and since $a_t = \sigma_t \epsilon_t$, the unconditional variance of returns, $E[\sigma_t^2] = E[a_t^2]$, is also $\alpha_0/(1 - \alpha_1 - \beta_1)$.

Just as an ARMA(1,1) can be written as an AR(∞), a GARCH(1,1) can be written as an ARCH(∞),

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 (\alpha_0 + \alpha_1 a_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \\ &= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_0 \beta_1 + \alpha_1 \beta_1 a_{t-2}^2 + \beta_1^2 \sigma_{t-2}^2 \\ &= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_0 \beta_1 + \alpha_1 \beta_1 a_{t-2}^2 + \beta_1^2 (\alpha_0 + \alpha_1 a_{t-3}^2 + \beta_1 \sigma_{t-3}^2) \\ &\vdots \\ &= \frac{\alpha_0}{1 - \beta_1} + \alpha_1 \sum_{i=0}^{\infty} a_{t-1-i}^2 \beta_1^i \end{aligned}$$

so that the conditional variance at time t is the weighted sum of past squared residuals and the weights decrease as you go further back in time.

Since the unconditional variance of returns is $E[\sigma^2] = \alpha_0/(1 - \alpha_1 - \beta_1)$, we can write the GARCH(1,1) equation yet another way

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= (1 - \alpha_1 - \beta_1) E[\sigma^2] + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

Written this way, it is easy to see that next period's conditional variance is a weighted combination of the unconditional variance of returns, $E[\sigma^2]$, last period's squared residuals, a_{t-1}^2 , and last period's conditional variance, σ_{t-1}^2 , with weights $(1 - \alpha_1 - \beta_1)$, α_1 , β_1 which sum to one.

It is often useful not only to forecast next period's variance of returns, but also to make an l -step ahead forecast, especially if our goal is to price an option with l steps to expiration using our volatility model. Again starting from the GARCH(1,1) equation for σ_t^2 , we can

derive our forecast for next period's variance, $\hat{\sigma}_{t+1}^2$

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
\hat{\sigma}_{t+1}^2 &= \alpha_0 + \alpha_1 E[a_t^2 | I_{t-1}] + \beta_1 \sigma_t^2 \\
&= \alpha_0 + \alpha_1 \sigma_t^2 + \beta_1 \sigma_t^2 \\
&= \alpha_0 + (\alpha_1 + \beta_1) \sigma_t^2 \\
&= \sigma^2 + (\alpha_1 + \beta_1) (\sigma_t^2 - \sigma^2) \\
\hat{\sigma}_{t+2}^2 &= \alpha_0 + \alpha_1 E[a_{t+1}^2 | I_{t-1}] + \beta_1 E[\sigma_{t+1}^2 | I_{t-1}] \\
&= \alpha_0 + (\alpha_1 + \beta_1) \hat{\sigma}_{t+1}^2 \\
&= \sigma^2 + (\alpha_1 + \beta_1) (\hat{\sigma}_{t+1}^2 - \sigma^2) \\
&= \sigma^2 + (\alpha_1 + \beta_1)^2 (\sigma_t^2 - \sigma^2) \\
&\vdots \\
\hat{\sigma}_{t+l}^2 &= \alpha_0 + (\alpha_1 + \beta_1) \hat{\sigma}_{t+l-1}^2 \\
&= \sigma^2 + (\alpha_1 + \beta_1) (\hat{\sigma}_{t+l-1}^2 - \sigma^2) \\
&= \sigma^2 + (\alpha_1 + \beta_1)^l (\sigma_t^2 - \sigma^2)
\end{aligned}$$

where we have substituted for the unconditional variance, $\sigma^2 = \alpha_0 / (1 - \alpha_1 - \beta_1)$.

From the above equation we can see that $\hat{\sigma}_{t+l}^2 \rightarrow \sigma^2$ as $l \rightarrow \infty$ so as the forecast horizon goes to infinity, the variance forecast approaches the unconditional variance of a_t . From the l -step ahead variance forecast, we can see that $(\alpha_1 + \beta_1)$ determines how quickly the variance forecast converges to the unconditional variance. If the variance spikes up during a crisis, the number of periods, K , until it is halfway between the first forecast and the unconditional variance is $(\alpha_1 + \beta_1)^K = 0.5$, so the half life is given by $K = \ln(0.5) / \ln(\alpha_1 + \beta_1)$. For example, if $(\alpha_1 + \beta_1) = 0.97$ and steps are measured in days, the half life is approximately 23 days.

Maximum Likelihood Estimation of Parameters

In general, to estimate the parameters using maximum likelihood, we form a likelihood function, which is essentially a joint probability density function but instead of thinking of it as a function of the data given the set of parameters, $f(x_1, x_2, \dots, x_n | \Theta)$, we think of the likelihood function as a function of the parameters given the data, $L(\Theta | x_1, x_2, \dots, x_n)$, and we maximize the likelihood function with respect to the parameters, which is essentially finding the mode of the distribution.

If the residual returns were independent of each other, we could write the joint density function as the product of the marginal densities, but in the GARCH model, returns are not, of course, independent. However, we can still write the joint probability density function as the product of conditional density functions

$$\begin{aligned}
f(r_1, r_2, \dots, r_T) &= f(r_T | r_1, r_2, \dots, r_{T-1}) f(r_1, r_2, \dots, r_{T-1}) \\
&= f(r_T | r_1, r_2, \dots, r_{T-1}) f(r_{T-1} | r_1, r_2, \dots, r_{T-2}) f(r_1, r_2, \dots, r_{T-2}) \\
&\vdots \\
&= f(r_T | r_1, r_2, \dots, r_{T-1}) f(r_{T-1} | r_1, r_2, \dots, r_{T-2}) \cdots f(r_1).
\end{aligned}$$

For a GARCH(1,1) model with Normal conditional returns, the likelihood function is

$$\begin{aligned}
L(\alpha_0, \alpha_1, \beta_1, \mu | r_1, r_2, \dots, r_T) &= \\
\frac{1}{\sqrt{2\pi\sigma_T^2}} \exp\left(-\frac{(r_T - \mu)^2}{2\sigma_T^2}\right) &\frac{1}{\sqrt{2\pi\sigma_{T-1}^2}} \exp\left(-\frac{(r_{T-1} - \mu)^2}{2\sigma_{T-1}^2}\right) \cdots \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(r_1 - \mu)^2}{2\sigma_1^2}\right).
\end{aligned}$$

Since the $\ln L$ function is monotonically increasing function of L , we can maximize the log of the likelihood function

$$\ln L(\alpha_0, \alpha_1, \beta_1, \mu | r_1, r_2, \dots, r_T) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^T \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^T \left(\frac{(r_i - \mu)^2}{\sigma_i^2} \right)$$

and for a GARCH(1,1), we can substitute $\sigma_i^2 = \alpha_0 + \alpha_1 a_{i-1}^2 + \beta_1 \sigma_{i-1}^2$ into the above equation, and the likelihood function is only a function of the returns, r_i and the parameters.

Notice that besides estimating the parameters, α_0 , α_1 , β_1 , and μ , we must also estimate the initial volatility, σ_1 . If the time series is long enough, the estimate for σ_1 will be unimportant.

In class, we will demonstrate how to create and maximize the likelihood function using only a spreadsheet.

Nonnormal Conditional Returns

In order to better model the excess kurtosis we observe with asset prices, we can relax the assumption that the conditional returns are normally distributed. For example, we can assume returns follow a student's t-distribution or a Generalized Error Distribution (GED), both of which can have fat tails. For example, the density function for the GED is given by

$$f(x) = \frac{\nu \exp\left(-\frac{1}{2} \left| \frac{x}{\lambda \sigma} \right|^\nu\right) \frac{1}{\sigma}}{\lambda 2^{(1+1/\nu)} \Gamma(1/\nu)}$$

where $\nu > 0$ and

$$\lambda = [2^{-(2/\nu)} \Gamma(1/\nu) / \Gamma(3/\nu)]^{1/2}$$

and where $\Gamma(\cdot)$ is the Gamma function, defined as

$$\Gamma(x) = \int_0^{\infty} y^{x-1} e^{-y} dy$$

so that $\Gamma(x) = (x - 1)!$ if x is an integer. The parameter, ν , is a measure of fatness of the tails. The Gaussian distribution is a special case of the GED when $\nu = 2$, and when $\nu < 2$, the distribution has fatter tails than a Gaussian distribution.

With a GED, the log-likelihood function becomes

$$\ln L(\alpha_0, \alpha_1, \beta_1, \mu, \nu | r_1, r_2, \dots, r_T) = T(\ln(\nu) - \ln(\lambda) - (1 + 1/\nu) \ln(2) - \ln(\Gamma(1/\nu))) - \frac{1}{2} \sum_{i=1}^T \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^T \left(\frac{(r_i - \mu)^2}{\lambda^2 \sigma_i^2} \right)^{\nu/2}$$

We can substitute the GARCH(1,1) equation $\sigma_i^2 = \alpha_0 + \alpha_1 a_{i-1}^2 + \beta_1 \sigma_{i-1}^2$ just as before with normal conditional returns, except we have an extra parameter, ν , to estimate.

Instead of using a three parameter distribution for conditional returns that captures kurtosis, one could use a four parameter distribution for conditional returns that captures both skewness and kurtosis. The Johnson Distribution (Johnson (1949)) is one such family of distributions. It is an easy distribution to work with because it is a transformation of normal random variables. In the unbounded version, the upper and lower tails go to infinity which is an important property for financial returns.

Asymmetric Variations of GARCH Models

One empirical observation is that in many markets, the impact of negative price moves on future volatility is different from that of positive price moves. This is particularly true in equity markets. Perhaps there is no better example of this than the changes in volatility during the financial crisis in 2008. Consider Figures 3 and 4 below, which show, for the end of September and beginning of October of 2008, how implied volatility reacted asymmetrically to up and down stock market moves. Figure 3 shows the returns for the S&P 500 during this period, and Figure 4 shows the VIX Index, which measures the weighted average of the implied volatility of short term S&P 500 options. We use the VIX as a proxy for market expectations about future volatility. As expected, when the market declined sharply, like on 9/29, 10/2, 10/6, 10/9, and 10/15, when the market dropped 8.8%, 4.0%, 5.7%, 7.6%, and 9.0%, the VIX rose on each of those days. However, on 10/13, when the market rose 11.6%, which was the largest percentage rise in the stock market since 1933, the VIX actually fell 21.4%, from 69.95 to 54.99. And even on small down days in the market, like 10/1, 10/8, 10/14, and 10/17, when the market dropped by less than 1.2% on each day, the VIX actually rose each time.

It is clear that not only does the magnitude of a_t^2 affect future volatility, but the sign of a_t also affects future volatility, at least for equities. It is not clear why volatility should increase more when the level of stock prices drop compared to a stock price rise. On one

Index HCP

Historical % Changes Page 1 / 1

SPX S&P 500 INDEX PRICE 940.55

Range 9/29/08 to 10/17/08 Period 1 Daily USD

	DATE	PRICE	CHANGE	% Change	VOLUME	CHANGE	% Change
F	10/17/08	940.55	-5.88	-0.62	1.6BLN	-272.6MLN	-14.59
T	10/16/08	946.43	+38.59	+4.25	1.9BLN	+306.6MLN	+19.63
W	10/15/08	907.84	-90.17	-9.03	1.6BLN	-154.6MLN	-9.01
T	10/14/08	998.01	-5.34	-0.53	1.7BLN	+118.4MLN	+7.41
M	10/13/08	1003.35	+104.13	+11.58	1.6BLN	-1.1BLN	-40.86
F	10/10/08	899.22	-10.70	-1.18	2.7BLN	+837.0MLN	+44.88
T	10/ 9/08	909.92	-75.02	-7.62	1.9BLN	-64.9MLN	-3.36
W	10/ 8/08	984.94	-11.29	-1.13	1.9BLN	+348.7MLN	+22.05
T	10/ 7/08	996.23	-60.66	-5.74	1.6BLN	-143.2MLN	-8.30
M	10/ 6/08	1056.89	-42.34	-3.85	1.7BLN	+439.0MLN	+34.16
F	10/ 3/08	1099.23	-15.05	-1.35	1.3BLN	-51.2MLN	-3.83
T	10/ 2/08	1114.28	-46.78	-4.03	1.3BLN	+126.9MLN	+10.49
W	10/ 1/08	1161.06	-5.30	-0.45	1.2BLN	-221.9MLN	-15.50
T	9/30/08	1166.36	+59.97	+5.42	1.4BLN	-405.9MLN	-22.09
M	9/29/08	1106.39	-106.62	-8.79	1.8BLN	+778.3MLN	+73.49

Australia 61 2 9777 8600 Brazil 5511 3048 4500 Europe 44 20 7330 7500 Germany 49 69 9204 1210 Hong Kong 852 2977 6000
Japan 81 3 3201 8900 Singapore 65 6212 1000 U.S. 1 212 318 2000 Copyright 2008 Bloomberg Finance L.P.
H252-689-0 19-Oct-08 11:36:26

Figure 3: The S&P 500 around the October 2008 financial crisis

Index HCP

Historical % Changes Page 1 / 1

VIX CBOE SPX VOLATILITY INDX PRICE 70.33

Range 9/29/08 to 10/17/08 Period 1 Daily USD

	DATE	PRICE	CHANGE	% Change		CHANGE	% Change
F	10/17/08	70.33	+2.72	+4.02			
T	10/16/08	67.61	-1.64	-2.37			
W	10/15/08	69.25	+14.12	+25.61			
T	10/14/08	55.13	+1.14	+0.25			
M	10/13/08	54.99	-14.96	-21.39			
F	10/10/08	69.95	+6.03	+9.43			
T	10/ 9/08	63.92	+6.39	+11.11			
W	10/ 8/08	57.53	+3.85	+7.17			
T	10/ 7/08	53.68	+1.63	+3.13			
M	10/ 6/08	52.05	+6.91	+15.31			
F	10/ 3/08	45.14	-.12	-0.27			
T	10/ 2/08	45.26	+5.45	+13.69			
W	10/ 1/08	39.81	+1.42	+1.07			
T	9/30/08	39.39	-7.33	-15.69			
M	9/29/08	46.72	+11.98	+34.48			

Australia 61 2 9777 8600 Brazil 5511 3048 4500 Europe 44 20 7330 7500 Germany 49 69 9204 1210 Hong Kong 852 2977 6000
Japan 81 3 3201 8900 Singapore 65 6212 1000 U.S. 1 212 318 2000 Copyright 2008 Bloomberg Finance L.P.
H252-689-0 19-Oct-08 11:37:01

Figure 4: The VIX implied volatility index around the October 2008 financial crisis

hand, as stocks drop, the debt/equity ratios increase and stocks become more volatile with higher leverage ratios. But the changes in volatility associated with stock market drops are much larger than that which could be explained by leverage alone.

One model to account for this asymmetry is the Threshold GARCH (TGARCH) model, also known as the GJR-GARCH model following the work of Glosten, Jagannathan, and Runkle (1993):

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \gamma_1 S_{t-1} a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

where

$$S_{t-1} = \begin{cases} 1, & \text{if } a_{t-1} < 0; \\ 0, & \text{if } a_{t-1} \geq 0. \end{cases}$$

Now there is an additional parameter, γ_1 to estimate using the maximum likelihood techniques described earlier.

Another variant on GARCH to account for the asymmetry between up and down moves described above is the EGARCH model of Nelson (1991). In an EGARCH(1,1),

$$\ln(\sigma_t^2) = \alpha_0 + \frac{\alpha_1 a_{t-1} + \gamma_1 |a_{t-1}|}{\sigma_{t-1}} + \beta_1 \ln(\sigma_{t-1}^2).$$

Notice that there is an asymmetric effect between positive and negative returns. Also, to avoid the possibility of a negative variance, the model is an AR(1) on $\ln(\sigma_t^2)$ rather than σ_t^2 .

An alternative representation of an EGARCH model, which is often found in the literature, is to write the EGARCH as an AR(1) process on $\ln(\sigma_t^2)$ with zero mean, i.i.d. residuals, $g(\epsilon_t)$,

$$\ln(\sigma_t^2) = \alpha + \beta(\ln(\sigma_{t-1}^2) - \alpha) + g(\epsilon_{t-1})$$

where $g(\epsilon_t) = \theta \epsilon_t + \gamma(|\epsilon_t| - E[|\epsilon_t|])$ and if $\epsilon_t \sim N(0, 1)$, $E[|\epsilon_t|] = \sqrt{2/\pi}$, and $\alpha = E[\ln(\sigma_t^2)]$.

There are many other models in the literature that try to capture the asymmetry between up and down moves on future volatility. For example, a slight variation on a GARCH(1,1) is

$$\sigma_t^2 = \alpha_0 + \alpha_1 (a_{t-1} - x)^2 + \beta_1 \sigma_{t-1}^2$$

so that the next period's variance is not necessarily minimized when the squared residuals are zero. There are a host of other variations of asymmetric GARCH models which can be found in the literature.

The Integrated GARCH Model

In the case where $\alpha_1 + \beta_1 = 1$, the GARCH(1,1) model becomes

$$\sigma_t^2 = \alpha_0 + (1 - \beta_1) a_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

This model, first developed by Engle and Bollerslev (1986), is referred to as an Integrated GARCH model, or an IGARCH model. Squared shocks are persistent, so the variance follows

a random walk with a drift α_0 . Since we generally do not observe a drift in variance, we will assume $\alpha_0 = 0$. Just as a GARCH model is analogous to an ARMA model, the IGARCH model where the variance process has a unit root is analogous to an ARIMA model.

When $\alpha_1 + \beta_1 = 1$ and $\alpha_0 = 0$, the l -step ahead forecast that we derived for a GARCH(1,1) model becomes

$$\begin{aligned}\hat{\sigma}_{t+l}^2 &= \alpha_0 + (\alpha_1 + \beta_1)\hat{\sigma}_{t+l-1}^2 \\ &= \hat{\sigma}_{t+l-1}^2 \\ &= \sigma_t^2\end{aligned}$$

so the forecast for future variance is the current variance, just as in a random walk.

Also, if we now write the model as an ARCH(∞) as we did before with a GARCH(1,1) model, after repeated substitutions we get

$$\begin{aligned}\sigma_t^2 &= \frac{\alpha_0}{1 - \beta_1} + \alpha_1 \sum_{i=0}^{\infty} a_{t-1-i}^2 \beta_1^i \\ &= (1 - \beta_1) \sum_{i=0}^{\infty} a_{t-1-i}^2 \beta_1^i \\ &= (1 - \beta_1)(a_{t-1}^2 + \beta_1 a_{t-2}^2 + \beta_1^2 a_{t-3}^2 + \dots)\end{aligned}$$

which is an exponential smoothing of past squared residuals. The “weights” on the squared residuals, $(1 - \beta_1), \beta_1(1 - \beta_1), \beta_1^2(1 - \beta_1), \dots$ sum to one, so exponential weighting can be used as an alternative to historical variance, $\sigma_t^2 = (1/N) \sum_{i=1}^N (r_{t-i} - \bar{r})^2$, which estimates historical variance by placing equal weights on the last N data points, where N is often chosen arbitrarily.

Fractionally Integrated GARCH Models

In GARCH models, the estimate of the persistence parameter, $\phi = \alpha_1 + \beta_1$, is often very close to 1, but even so, these models often do not capture the persistence of volatility shocks. When we look at the data, we notice that volatility shocks decay at a much slower rate than they do under GARCH models. One way to increase the persistence of shocks is to introduce a “long memory”, or “fractionally integrated” model. In these models, the autocorrelations of squared returns decay at a much slower rate.

Recall the EGARCH model

$$(1 - \phi B)(\ln(\sigma_t^2) - \mu) = g(\epsilon_{t-1})$$

where B is the back-shift operator, $Bx_t = x_{t-1}$, and $g(\epsilon_{t-1})$ was defined earlier. In a Fractionally Integrated EGARCH Model, known as FIEGARCH(1, d ,1),

$$(1 - \phi B)(1 - B)^d(\ln(\sigma_t^2) - \mu) = g(\epsilon_{t-1}).$$

The binomial expansion of $(1 - B)^d$ is given by

$$\begin{aligned} (1 - B)^d &= 1 - dB + \frac{d(d-1)}{2!}B^2 - \frac{d(d-1)(d-2)}{3!}B^3 + \dots \\ &= \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \end{aligned}$$

and

$$\begin{aligned} (1 - \phi B)(1 - B)^d &= 1 - dB + \frac{d(d-1)}{2!}B^2 - \frac{d(d-1)(d-2)}{3!}B^3 + \dots \\ &\quad - \phi B + d\phi B^2 - \frac{d(d-1)}{2!}\phi B^3 + \dots \\ &= 1 + \sum_{k=1}^{\infty} \left[\binom{d}{k} + \phi \binom{d}{k-1} \right] (-B)^k \\ &= 1 + \sum_{k=1}^{\infty} b_k B^k \end{aligned}$$

where

$$b_k = \left[\binom{d}{k} + \phi \binom{d}{k-1} \right] (-1)^k$$

and if we truncate the infinite series at some K , the FIEGARCH(1,d,1) becomes

$$\ln(\sigma_t^2) = \mu + \sum_{k=1}^K b_k (\ln(\sigma_{t-k}^2) - \mu) + g(\epsilon_{t-1}).$$

With a fractionally integrated model, it turns out that the autocorrelation of squared returns asymptotically has a polynomial decay, $\rho_\tau \propto \tau^{2d-1}$, in contrast to a GARCH model, where the autocorrelations have a much faster exponential decay, $\rho_\tau \propto \phi^\tau$.

The GARCH-M Model

Another variation of a GARCH model tests whether variance can impact the mean of future returns. These models are referred to as GARCH in the mean, or GARCH-M models. A GARCH(1,1)-M is represented as

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ r_t &= \mu + \sigma_t \epsilon_t + \lambda \sigma_t^2. \end{aligned}$$

In some specifications, the volatility, rather than the variance, affects returns as $r_t = \mu + \sigma_t \epsilon_t + \lambda \sigma_t$.

If $\lambda \neq 0$, these models imply a serial correlation of returns since variance is serially correlated and the returns depend on the variance. Many studies have tried to determine whether λ is significantly different from zero, usually with mixed conclusions. In fact, of the studies that find that $\lambda \neq 0$, they do not even agree on the sign of λ .

Other Variations of GARCH Models

There have been dozens of papers published on variations of the basic GARCH models. Theoretically, you could add any number of variables on the right hand side that are known at time $t - 1$ for forecasting time t volatility. For example, you could use implied volatility as an explanatory variable. However, if you wanted to forecast more than one period ahead, these variables may not work.

Some papers have included a dummy variable, D_t , which equals 1 if t is a Monday, if you believe the Friday-to-Monday volatility is higher than other day's volatility.

Here is an example of how you can create your own variant of a GARCH model. In foreign exchange markets, people do not talk about asymmetries between up and down moves and future volatility as they do in equity markets. Indeed, an up move in the yen against the dollar is a down move of the dollar against the yen. People do talk about how volatility goes up when an exchange rate approaches the edge of a trading range, and how volatility goes down when the exchange rate moves back to the middle of its historical range. So you could run a GARCH variant, call it RR-GARCH for Rob Reider's GARCH, where

$$\ln(\sigma_t^2) = \alpha_0 + \alpha_1 \frac{|a_{t-1}|}{\sigma_{t-1}} + \beta_1 \ln(\sigma_{t-1}^2) + \gamma_1 |I_{t-1}|$$

where

$$I_t = \left[2 \frac{S_t - \min_{1 \leq \tau \leq 60} (S_{t-\tau})}{\max_{1 \leq \tau \leq 60} (S_{t-\tau}) - \min_{1 \leq \tau \leq 60} (S_{t-\tau})} - 1 \right]^\lambda$$

and S_t is the exchange rate on day t . I_t increases as the exchange rate moves closer to the edge of a range.

Once you have the parameters estimated for a model, we have not discussed how you would price an option under the estimated volatility model so that it can be compared with the market price of an option. For some models, like the EGARCH model, there are closed form solutions for option prices. For more complicated models, like the RR-GARCH model above, you would have to use Monte Carlo simulations to price options. Starting with an initial stock price, S_0 , you would generate a ϵ_1 from $N(0, 1)$ and multiply it by σ_1 to get r_1 (we would add r_f instead of μ to price an option), and $S_1 = S_0 \exp(r_1)$. From r_1 , we could compute σ_2^2 from our model, and start the process over again by generating ϵ_2 and r_2 .

Choosing a Model

Suppose you had two models. How do you know which is a better fit to the data? For example does an EGARCH(1,1) fit better than a GARCH(1,1)? Or how about a GARCH(1,1) compared with a GARCH(1,1) with nonnormal conditional returns? If the two models have the same number of parameters, you could simply compare the maximum value of their likelihood functions. What if the models have a different number of parameters? In that case, you can use the Akaike Information Criterion (AIC), which makes adjustments to the

likelihood function to account for the number of parameters. If the number of parameters in the model is P , the AIC is given by

$$\text{AIC}(P) = 2 \ln(\text{maximum likelihood}) - 2P$$

Some have found that this leads to models chosen with too many parameters, and a second criterion attributed to Schwartz, the Schwartz Bayesian Criterion (SBC), is given by

$$\text{SBC}(P) = 2 \ln(\text{maximum likelihood}) - P \ln(T)$$

where T is the number of observations. So AIC gives a penalty of 2 for an extra parameter and SBC gives a penalty of $\ln(T)$ for an additional parameter, which will be larger than 2 for any reasonable number of observations.

Another diagnostic check of a model (not, as before, for comparing two models but to see whether a model is a good fit) is to compute the residuals, $\hat{\epsilon}_t = (r_t - \hat{\mu})/\hat{\sigma}_t$ of a GARCH-type model, and test whether these residuals are i.i.d., as they should be (but probably won't be) if the model were properly specified. A common test performed on $\hat{\epsilon}_t$ is the Portmanteau test, which jointly tests whether several squared autocorrelations of $\hat{\epsilon}_t$ are zero.

The test statistic proposed by Box and Pierce (1970), for the null hypothesis $H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0$ is

$$Q^*(m) = T \sum_{l=1}^m \hat{\rho}_l^2$$

and a modified version for smaller sample sizes proposed by Ljung and Box (1978) is

$$Q(m) = T(T+2) \sum_{l=1}^m \hat{\rho}_l^2 / (T-l)$$

where m is the number of autocorrelations, T is the number of observations. When we are looking at GARCH models, it is more relevant to test the autocorrelations of squared residuals of the model, so $\hat{\rho}_l$ is the autocorrelation of $\hat{\epsilon}_t^2$ with $\hat{\epsilon}_{t-l}^2$ rather than the autocorrelation of $\hat{\epsilon}_t$ with $\hat{\epsilon}_{t-l}$, which might be the relevant autocorrelation if we were looking at the residuals from an ARMA model, for example. Under the assumption that the process is i.i.d., the asymptotic distribution of Q and Q^* is chi-squared with m degrees of freedom. This has shown to be conservative, and Q and Q^* are often compared to a chi-squared with $m - P$ degrees of freedom, where P is the number of parameters being estimated. The number of correlations, m , is often chosen so that $m \approx \ln(T)$.

Note that the tests in this section are all within sample. In other words, the same data that is used to fit the model is also used to evaluate it. Later, we talk about out-of-sample tests, where one part of the data is used to fit the model and subsequent data is used to evaluate it.

References

- Bollerslev, T. 1986. Generalized Autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31:307-327.
- Campbell, J.Y., A.W. Lo, and A.C. MacKinlay. 1997. *The econometrics of financial markets*. Princeton University Press.
- Engle, R.F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987-1007.
- Engle, R.F. and T. Bollerslev. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* 5:1-50.
- Glosten, L.R., R. Jagannathan, and D. Runkle. 1993. Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48:1779-1801.
- Hamilton, J.D. 1994. *Times series analysis*. Princeton University Press.
- Johnson, N.L. 1949. Systems of frequency curves generated by methods of translation. *Biometrika* 36:149-176.
- Nelson, D.B. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59:347-370.