LONG MEMORY IN VOLATILITY

How persistent is volatility? In other words, how quickly do financial markets forget large volatility shocks? Figure 1.1, Shephard (attached) shows that daily squared returns on exchange rates and stock indices can have autocorrelations which are significant for many lags. In any stationary *ARCH* or *GARCH* model, memory decays exponentially fast. For example, if $\{\varepsilon_t\}$ are $ARCH(1)$, the $\{\varepsilon_t^2\}$ have autocorrelations $\rho_k = \alpha^k$. Specifically, if $\alpha = .8$ and $k = 20$, we get $\rho_{20} = .012$. This seems an unrealistically fast decay. On the other hand, for any integrated *ARCH* or *GARCH*, $\rho_k = 1$ for all k, so there is no decay at all. This seems unrealistically slow. The progression from $ARCH(1)$ to $ARCH(q)$ to *GARCH* represents an attempt to allow for the strong volatility persistence observed in actual data. Typically, the exponential decay inherent in any stationary *ARCH* or *GARCH* model is too rapid to adequately describe the data (especially high frequency data), forcing the estimated models to be integrated. In reality, however, volatility may not be integrated, and the behavior of the estimated *ARCH* and *GARCH* models may simply be a signal that the memory is decaying relatively slowly compared to the exponential rate. What is needed, then, is a richer class of models allowing for intermediate degrees of volatility persistence.

In stationary long memory models for volatility, the autocorrelations of $\{\varepsilon_t^2\}$ decay slowly to zero as a power law, $\rho_k \sim k^{2d-1}$ where d is between 0 and 1/2. As we will see, typical values of d for financial time series are around 0.4. This provides a volatility series $\{\varepsilon_t^2\}$ with longer memory than the stationary *ARCH* and *GARCH* models, which have $d = 0$, but shorter memory than the integrated models, which have $d = 1$.

ARFIMA: A Long Memory Model for Levels

The most popular long memory model for levels $\{x_t\}$ is the $ARFIMA(p, d, q)$, due to Hosking (1981) and Granger and Joyeux (1980). The *FI* in *ARFIMA* stands for "Fractionally Integrated". In other words, **ARFIMA models are simply ARIMA models in which the d** (the degree of integration) **is allowed to be a fraction of a whole number,** such as 0.4, instead of an integer, such as 0 or 1.

The simplest long memory model is the Gaussian $ARFIMA(0, d, 0)$ with $0 < d < 1/2$. Such a series can be represented in the $MA(\infty)$ form as $x_t = e_t + a_1 e_{t-1} + \cdots$, where the $\{e_t\}$ are Gaussian white noise, and the $\{a_k\}$ coefficients are determined by d and decay as $a_k \sim k^{d-1}$ (slow decay). We can compute the a_k using the *fractional differencing operator* $\Delta^d = (1-B)^d$, as we explain below.

The idea of a fractional difference may seem puzzling at first. It is easy to take the d'th difference when d is 0 1 or 2, but what if $d = 0.4$? A natural definition of fractional differencing was provided by Hosking (1981) and independently by Granger and Joyeux (1980). First, define the backshift operator B by $Bx_t = x_{t-1}$. (B is simply a lag operator, which shifts any time series one time unit into the past). Next, define the differencing operator $\Delta = 1 - B$. The name is appropriate, since $\Delta x_t = (1-B)x_t = x_t - x_{t-1}$, so Δ differences the series. A random walk has $d = 1$ and can be written as $\Delta x_t = e_t$, so that the first difference of $\{x_t\}$ is a white noise. Equivalently, $x_t = \Delta^{-1} e_t$, where $\Delta^{-1} = \dfrac{1}{1-B} = 1 + B + B^2 + \cdots$ is the integration operator. (We used the geometric series for $\dfrac{1}{1-B}$.)

The $ARFIMA(0, d, 0)$ is defined by $\Delta^d x_t = e_t$ so that the d'th (fractional) difference of x_t is Gaussian white noise. Equivalently, $x_t = \Delta^{-d} e_t$, where Δ^d and Δ^{-d} are the fractional differencing and fractional integration operators. For example, $\Delta^{-d} = (1-B)^{-d}$, which can be expressed in the infinite (Binomial or Taylor) series $1 + a_1 B + a_2 B^2 + \cdots$, and the a_k are the $MA(\infty)$ weights discussed earlier. The general $ARFIMA(p, d, q)$ model is defined by assuming that $\Delta^d x_t$ is a stationary invertible $ARMA(p, q)$.

There is an interesting connection between the fractional d in long memory models and the fractals studied by Mandelbrot and others. Roughly speaking, a fractal is an object with fractional dimension, and which exhibits self-similarity. (Smaller parts resemble the whole). Since a time series plot is a curve drawn inside a two-dimensional plane, it seems obvious that this curve is one-dimensional. But it is often observed that plots of financial time series at different time scales (e.g., hourly, daily and weekly stock price charts) look similar. These series seem to be self-similar, in some statistical sense. Furthermore, the curves tend to have a very bumpy, craggy appearance, and zooming in on a particular piece of the series reveals even more bumpiness at this higher level of magnification. This suggests that the curves are fractals, of dimension *between* 1 and 2. It turns out that **realizations of long memory**

time series are fractals with dimension that decreases as d increases. The lower the dimension, the smoother the curve will be. So a random walk ($d = 1$) is smoother than a long memory series with $d = 0.4$, which is in turn smoother than a white noise ($d = 0$).

Another very important property of long memory models is that the variance of a sample mean $\bar{x}_n$ based on n observations is $var\ \bar{x}_n \sim n^{2d-1}$. If $d = 0$, we get the familiar $1/n$ rate, but **in the long memory case, d > 0, the variance of $\bar{x}_n$ goes to zero more slowly than $1/$n.** Thus, standard methods (such as the t-test) are invalid for long memory series.

FIGARCH: A Long Memory Model for Volatility

Most financial time series have $d = 1$ for the (raw or log) levels, e.g., log exchange rates, log stock prices. This is consistent with the efficient market theory i.e., the levels are a Martingale and returns are a Martingale Difference. It is the *volatility* (e.g., squared returns) which typically has a fractional value of d. What is needed, then, is a long memory model for the volatility of returns which allows the returns themselves to be a Martingale Difference. The *FIGARCH* (Fractionally Integrated *GARCH*) model of Baillie, Bollerslev and Mikkelsen (1996) is a model for $\varepsilon_t = \log x_t - \log x_{t-1}$.

The definition of *FIGARCH* parallels that of *ARCH*, but allows for long memory in the conditional variance, i.e., $\varepsilon_t \mid \psi_{t-1} \sim N(0, h_t)$, with $h_t = \omega + \sum_{k=1}^{\infty} \alpha_k \varepsilon_{t-k}^2$, where the α_k are the $AR(\infty)$ coefficients of an $ARFIMA(1, d, 0)$ model. Thus, $\{\varepsilon_t\}$ is MD (and therefore white noise), but the volatility series $\{\varepsilon_t^2\}$ has long memory. Specifically, $\{\varepsilon_t^2\}$ is $ARFIMA(1, d, 0)$ and has autocorrelations $\rho_k \sim k^{2d-1}$. A fortunate consequence of this is that the multistep forecasts of volatility will not revert quickly to a constant level, as is the case for stationary *ARCH* and *GARCH* models.

Long Memory Stochastic Volatility: An Alternative to FIGARCH

In the *FIGARCH* (or *ARCH*/*GARCH*) model, the 1-step conditional volatility is directly observable from ε_{t-1}^2, ε_{t-2}^2, $\cdots$, so we refer to these models as *observation driven*. The Stochastic Volatility (*SV*) models, which are not observation driven, provide an alternative to *ARCH*/*GARCH*/*FIGARCH* for modeling volatility clustering. In the *SV* model, the instantaneous volatility (standard deviation) is

$\sigma_t > 0$, an unobserved ("latent") stochastic process. The model is $\varepsilon_t = \sigma_t e_t$, where $\{e_t\}$ are Gaussian white noise, independent of $\{\sigma_t\}$, and $\varepsilon_t = \log x_t - \log x_{t-1}$ are the "returns".

It is not hard to show that $\{\varepsilon_t\}$ is a Martingale Difference. The $\{\varepsilon_t^2\}$ will be autocorrelated, so there will be volatility clustering. If we work with the *logs* of ε_t^2 (which seems reasonable from a data analysis point of view anyway) then a simple structure emerges. We have $\log \varepsilon_t^2 = \log \sigma_t^2 + \log e_t^2$, the sum of two independent processes, the second of which is a strict white noise. Thus, the autocorrelations of $\log \varepsilon_t^2$ are identical to those of $\log \sigma_t^2$.

Hull and White (1987), working in continuous time, considered the case where $\log \sigma_t^2$ is a stationary Gaussian $AR(1)$ process, and studied the implications of this SV model on option pricing. Since the autocorrelations of an $AR(1)$ decay exponentially fast, however, this model suffers from the same limitations as an $ARCH(1)$ in capturing actual volatility clustering. A useful generalization, therefore, is to take $\log \sigma_t^2$ to be a Gaussian $ARFIMA(p, d, q)$ series. Then the autocorrelations in $\log \varepsilon_t^2$ will decay as $\rho_k \sim k^{2d-1}$. The overall model is called Long Memory Stochastic Volatility, or $LMSV$. (Breidt, Crato and De Lima 1998, Harvey 1998).

There is hope here for carrying out tractable option pricing since Hull and White (1987) have shown that if $\{\varepsilon_t\}$ obeys any SV model, the fair price of a European option is simply the conditional expectation of the Black-Scholes formula, where the constant volatility σ^2 is replaced by $\overline{\sigma}^2$, the average of σ_t^2 from the current time t to the exercise time T.

Observed Volatility in High Frequency Exchange Rates

Excerpts are attached from "The Distribution of Exchange Rate Volatility", by Andersen, Bollerslev, Diebold and Labys (1999). The complete paper is available as a pdf file from

http://www.ssc.upenn.edu/˜diebold/papers/papers-f.html

The authors used five-minute DM/Dollar and Yen/Dollar returns (actually, changes in log exchange rate), a total of over 1 Million observations, from December 1, 1986 to December 1, 1996. For each series, they summed the squared returns in blocks spanning one trading day, to obtain a daily "observed" volatility. This is treated as if it were the true volatility for that day. The observed volatilities for day t are denoted by $vard_t$ and $vary_t$ for the DM and Yen series, respectively. The square roots of these variances are denoted by $stdy_t$ and $stdd_t$. The logs of these standard deviations are denoted by $lstdd_t$ and $lstdy_t$. The daily "observed" correlation and covariance between the two sets of returns are denoted by $corr_t$ and cov_t.

The third row of Table 3 shows estimated values of d based on each of the eight time series described above. All were significantly greater than zero and less than 1/2, with a typical value of about 0.4. Thus, there seems to be long memory not only in the observed volatilities, but also in the observed correlation between the two series of exchange rate returns. A unit root in volatility is strongly rejected in all cases by the Augmented Dickey Fuller test, also reported in Table 3. This supports the long memory hypothesis, and tends to rule out commonly used models such as integrated $GARCH(1,1)$. (It is noteworthy that Bollerslev, the inventor of $GARCH$, is one of the authors of this paper!)

Further support for long memory is provided in Figure 11, which shows the behavior of h-day partial sums of $lstdd_t$, $lstdy_t$ and $corr_t$. The figure plots the log of the variance of these partial sums against the log of h for $h = 1, \ldots, 30$, where for each value of h the variance is taken over the ten years of daily observations of the partial sums. The plots are strikingly linear, indicating that the variance of the partial sums behaves like a power law in h. The slopes of these lines agree quite well with the scaling rule for long memory (cf. the discussion of sample means earlier in this handout) which dictates that the variance of the partial sums should be proportional to h^{2d+1}.