Other Methods and Conclusions

7.13 Alternative Models

There are as many models proposed for financial data as there are creative people working in the area (i.e. lots). Some find more support in specific communities, and the debate about which are the most appropriate, shows no sign of early resolution. For example the use of Neural Nets have emerged from the field of *artificial intelligence* and provides a locally simple and compelling model originally suggested as model of the brain.

7.13.1 Neural Nets

A basic premise of much of modern research is that many otherwise extremely complex phenomena are much simpler when viewed locally. On this local scale, structures and organizations are substantially simpler. Complex societies of insects, for example, are organized with very simple interactions. Even differential equations like

$$\frac{dy}{dx} = y^2(1-y)$$

describe a simple local structure of a function (its slope is proportional to the square of the distance from zero times the distance from one) but the solution is difficult to write in closed form.

Neural Nets are suggested as devices for processing information as it passes through a network based loosely on the parallel architecture of animal brains. They are a form of multiprocessor computer system, with individual simple processing elements which are interconnected to a high degree. At a given node binary bits b_1, b_2, b_3 enter and then are processed with a very simple processor $g(b_1, b_2, b_3)$ (often a weighted average of the inputs, possibly transformed). This transformed output is then transmitted to one of more nodes.

Thus a particular neural net model consists of a description of the processors (usually simple functions of weighted averages), an architecture describing the routing, and a procedure for estimating the parameters (for example the weights in the weighted average). They have the advantage of generality and

flexibility- they can probably be modified to handle nearly any problem with some success. However, in specific models for which there are statistically motivated alternatives, they usually perform less well than a method designed for a statistical model. Their generality and flexibility makes them a popular research topic in finance see for example Trippi and Turban (1996).

7.13.2 Chaos, Long term dependence and non-linear Dynamics

Another topic, popularized in finance by books by Peters (1996) and Gliek (1987), is *chaos*. Chaotic systems are generally purely deterministic systems that may resemble random or stochastic ones. For example if we define a sequence by a recursion of the form $x_t = f(x_{t-1})$ (this is the same form as the recursion that the linear congruential random number generator satisfies) for some non-linear function f, the resulting system may have many of the apparent properties of a random sequence. Depending on the nature of the function f, the sequence may or may not appear "chaotic". Compare for example the behaviour of the above recursion when $f(x) = ax(1-x), 0 < x < 1, , a \leq 4$ for different initial conditions and different values of a. When $a = 4$, this recursion is extremely sensitive to the initial condition, as Figure 7.20 shows. In the left panel we plot the values of x_n against n for $a = 4$ and $x_0 = 0.4999$ and in the right panel, for $x_0 = 0.5$. This small change in the initial condition makes an enormous difference to the sequence x_n which converges almost immediately to zero when $x_0 = 0.5$ but when $x_0 = 0.4999$ behaves much more like a random sequence except with higher density near 0 and 1. This strong dependence on the distant past is typical of a chaotic system.

Similarly, the recursion

$$x_t = 1 - ax_{t-1}^2 + bx_{t-2}, \quad a = 1.4, \quad b = 0.3$$

describes a *bivariate* chaotic system, which, like an autoregressive process of order 2, requires two "seeds" to determine the subsequent elements of the sequence. In general, a system might define x_t as a non-linear function of n predecessors. Detecting chaos(or lack therof) is equivalent to determining whether the sequences $(x_t, x_{t+1}, \ldots x_{t+n}), t = 1, 2,$ fill $n+1$ dimensional space.

Tests designed to test whether a given sequence of stock returns are independent identically distributed generally result in rejecting this hypothesis but the most plausible explanation of this is not so clear. For example Hsieh (1991) tests for both chaotic behaviour and for arch-garch effects (predictable variance changes) and concludes that the latter is the most likely cause of apparent dependence in the data.

7.13.3 ARCH AND GARCH

There are many failures in the Black-Scholes model for stock returns but two extremely obvious ones common to the application of simple Gaussian time series models to much financial data, evident for at least the past 40 years. The

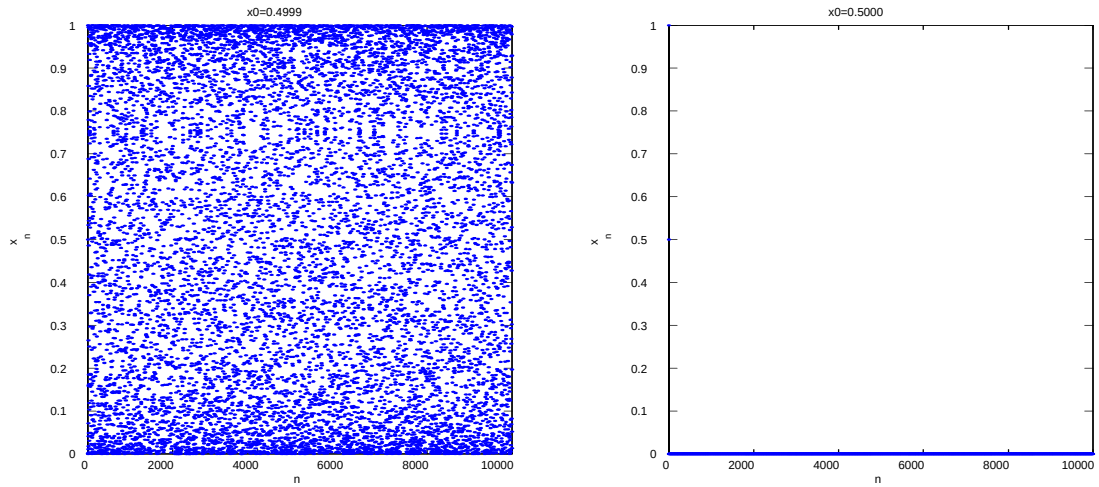

Figure 7.20: The effect of the initial condition on the recursion $x_{n+1} = 4x_n(1 - x_n)$. Left panel, $x_0 = 0.4999$, Right panel, $x_0 = 0.5$.

first is the heavy tails in the distribution of returns. There are many days in which the increase or decrease in a stock price, for example, is well beyond the range of anything reasonable for a normal random variable. Models such as the stable laws or the NIG process have been proposed to ameliorate this problem. However there is another apparent failure in such independent increment models, the failure to adequately represent extended observed periods of high and low volatility. The innovations are supposed in the conventional ARMA models to be independent with 0 mean and constant variance σ^2 and the squared innovations should therefore be approximately independent (uncorrelated) variates but most series show periods when these squared innovations tend to be consistently above the median followed by periods when they are consistently smaller. While there are many ways of addressing this failure in traditional models, one of the most popular is the use of GARCH, or *Generalized Autoregressive Conditional Heteroscedasticity* (see Bollerslev, 1986, Duan, 1995, Engle and Rosenberg, 1995).

Traditional time series models attempt to model the expected value of the series given the past observations, assuming that the conditional variance is constant, but a GARCH model takes this one moment further, allowing the conditional variance to also be modeled by a time series. In particular, suppose that the innovations in a standard time series model, say an ARMA model, are normally distributed given the past

$$a_t \sim N(0, h_t).$$

Assume that h_t, the conditional variance given the past, satisfies some ARMA

relationship, with the squared innovations posing as the new innovations process.

$$\beta(B)h_t = \alpha_0 + \alpha(B)a_t^2$$

where $\beta(B) = 1 - \beta_1 B - \ldots \beta_r B^r$ and $\alpha(B) = \alpha_1 B + \ldots + \alpha_s B^s$ and B is the backwards time shift so that $B^r h_t = h_{t-r}$.

The case $r = 0$ is the original ARCH *Autoregressive Conditional Heteroscedasticity* model, and the most common model takes $r = 0$, $s = 1$ so $h_t = \alpha_0 + \alpha_1 a_{t-1}^2$. For ARCH and GARCH models the parameters must be estimated using both the models for the conditional mean and the conditional variance and diagnostics apply to both models. The advantages of these models are that they provide for some dependence among the observations through volatility rather than through the mean, and that they tend to have heavier tails. As a result, they provide larger estimated prices for deep out-of-the-money options, for example, which are heavily dependent on an accurate model for volatility.

7.13.4 ARCH(1)

The basic model investigated by Engle(1982) was the simplest case in which the process has zero conditional mean (it is reasonable to expect that arbitrageurs in the market have removed a large part of any predictability in the mean) but that the squares are significantly autocorrelated. Most financial data exhibits this property to some degree. Engle's ARCH(1) model is:

$$\begin{aligned} x_t &\sim N(0, h_t) \text{ and} \\ h_t &= \alpha_0 + \alpha_1 x_{t-1}^2. \end{aligned} \tag{7.109}$$

An ARCH regression model allows the conditional mean of x_t in (7.109) to depend on some observed predictors. The GARCH-IN-MEAN process fit by French et. al.(1987) allow the mean of x_t to be a function of its variance so that $x_t \sim N(a + bh_t^{p/2}, h_t)$. This would allow testing the hypotheses of relative risk aversion, for example. However, there is little evidence that b may be non-zero, and even less evidence to determine whether the linear relation should be between mean and standard deviation ($p = 1$) or between mean and variance ($p = 2$).

7.13.5 Estimating Parameters

The conditional log likelihood to be maximized with respect to the parameters α_i, β_j is:

$$\ln(L) = -\frac{1}{2}\sum_t [\ln h_t + \frac{\hat{a}_t^2}{h_t}]$$

Various modifications of the above GARCH model are possible and have been tried, but the spirit of the models as well as most of the methodology remains basically the same. There is also a system of Yule-Walker equations that can

be solved for the coefficients β_i in an ARCH model. If γ_i is the autocovariance function of the *innovations squared* a_i^2 process, then

$$\gamma_n = \sum_{i=1}^{s} \alpha_i \gamma_{n-i} + \sum_{i=1}^{r} \beta_i \gamma_{n-i}$$

for $n \geq r+1$. These provide the usual *Partial Autocorrelation Function* for identification of the suitable order r of the autoregressive part.

7.13.6 Akaike's Information Criterion

A model which leads to small estimated variances for the innovations is obviously preferable to one with highly variable innovations if everything else is the same. In other words when we select a model, we are inclined to minimize the estimated residual variance $\frac{1}{N-k}\sum \hat{a}_i^2$ (or equivalently its logarithm) over the parameters themselves and k, the number of autoregressive+moving average parameters in the model. Unfortunately each additional parameter results in what may be only a marginal improvement in the residual variance so minimizing $\frac{1}{N-k}\sum \hat{a}_i^2$ would encourage the addition of parameters which do not improve the ability of the model to forecast or fit new observations. A better criterion, the Akaike's Information Criterion, penalizes the model for each additional parameter:

$$AIC = log[\frac{1}{N-k}\sum \hat{a}_i^2] + \frac{2k}{N}. \tag{7.110}$$

The AIC criterion chooses that model and number of parameters k which minimizes this quantity. In some software such as in *R* and *Splus*, the AIC differs from (7.110) approximately by a multiple of N, for example $AIC_2 = -2log(L) + 2\times k$ is approximately N times the value in (7.110). The advantage in multiplying by N is that differences operate on a more natural scale. When nested models are compared (i.e. one model is a special case of the other), differences between values of the statistic $-2log(L)$ have a distribution which is Chi-squared with degrees of freedom the difference in the number of parameters in the two models under the null hypothesis that the simpler model holds.

7.13.7 Estimation and testing ARCH Effects.

The function *ugarch* in Matlab estimates the parameters in a Garch model. In particular, if a is the vector of innovations from a time series for which we wish to fit a GARCH model, the command $[Alpha0, Alpha, Beta] = ugarch(a, p, q)$ fits a GARCH(p, q) model.

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + ... + \alpha_p h_{t-p} + \beta_1 a_{t-1}^2 + .. + \beta_q a_{t-q}^2$$

For example if we fit the Garch(1,1) model to the mean adjusted daily returns for the S&P 500 index over the period 1997-2002. The estimated model is

$$h_t = 0.000006 + 0.8671 h_{t-1} + 0.0974 a_{t-1}^2.$$

The large coefficient $\widehat{\alpha_1} = 0.8671$ on h_{t-1} indicates a strong tendency for the variance to remain near its previous value. In R there is a similar function in the package *tseries*, (see for example *http://pbil.univ-lyon1.fr/library/tseries/html/00Index.html*) run with a command like

garch(a,order=$c(p,q)$,coef=NULL, itmax=200,eps=NULL,grad=c("numerical"),series=NULL,trace=TR

where the NULL parameters indicates that the default values are used.

Most of the tests for the adequacy of a given time series model are inherited from regression, although in some cases the autocorrellation of the series induces a different limiting distribution. For example, if there is an ARCH or GARCH effect, then there should be a significant regression of $\hat{a}_t^2$ on its predecessors $\hat{a}_{t-1}^2, \hat{a}_{t-2}^2, \hat{a}_{t-3}^2 \ldots$. Suppose we are able to obtain residuals $\hat{a}_l, \hat{a}_{l+1}, \ldots \hat{a}_N$ from an ARMA model for the original series. We might test for ARCH effect by regressing the vector $(\hat{a}_{l+s}^2, \ldots \hat{a}_N^2)$ on a constant as well as the s "predictors"

$$(\hat{a}_{l+s-1}^2, , \hat{a}_{N-1}^2), (\hat{a}_{l+s-2}^2, , \hat{a}_{N-2}^2) \ldots . (\hat{a}_l^2, , \hat{a}_{N-s}^2)$$

and obtaining the usual coefficient of determination or *squared multiple correlation coefficient* R^2. Standardized, $(N-l)R^2$ has an approximate chi-squared distribution with s degrees of freedom under the null hypothesis of homoscedasticity so values above the 95'th chi-squared percentile would lead to rejecting the homoscedasticity null hypothesis and concluding arch-like effects. One can also fit a GARCH model and compare the values of the coefficient estimators with their standard errors to see if the model can be further simplified. Finally, it is easy to simulate an ARCH or a GARCH model (see for example the function $[a, h]$ = *ugarchsim(Alpha0 , Alpha , Beta , NumSamples)* in *Matlab*). Any test statistic which is sensitive to persistence in the volatility can be adapted to test for a GARCH model by using simulations to determine the distribution of this test statistic, where we fix the parameter values in the simulation at their estimated values.

7.13.8 Example. Canadian dollar, US dollar Exchange.

As an example we downloaded the US/Canadian dollar exchange rate close for a 10 year period from October 7, 1994 to October 8, 2004 from the Bank of Canada website *http://www.bankofcanada.ca*. There are approximately 2514 daily observations of the value of the US dollar priced in Canadian dollars. Suppose we first fit an autoregressive moving average model to this returns data of order $(1,1)$ using the systems identification toolbox in *Matlab*. The command *armax(data,[p, q])* fits an autoregressive moving average model in general, with autoregressive order p and moving average order 1. We fit an $AR(2)$ model to the returns from this series resulting in the model x_t $+0.03657x_{t-1}$ $-0.02497x_{t-2} = a_t$ with innovations process a_t and then we fit a GARCH(1,1)

model to the innovations a_t with the following estimated model for the variance h_t of a_t:

$$h_t = 0.9524h_{t-1} + 0.0474a_{t-1}^2.$$

Once again the large coefficient 0.954 on h_{t-1} indicates a high degree of persistence in the volatility.

Diebold and Nerlove (1989), confirm the ARCH effect on the exchange rate for a number of different currencies, observing ARCH effects at lag 12 weeks or more.

7.13.9 Conclusions

Research and modeling is a dynamic task, but in no discipline more than in finance. In physics theories change over time but at least the target is often a physical law which is, at least in terms of our meagre lifespans, relatively constant. Not so in the modeling of a financial time series. First order autocorrelation in the Dow Jones average was once quite strong, but with increased liquidity and statistical literacy of arbitrageurs, it has largely disappeared. Tools which permit "trading" volatility, interest rates may alter other features of the market as well. The standard approach to derivative pricing which we take here is to assume a model for an asset price, in which case the derivative, a function of the asset price, has a price functionally related to the asset price. Which is the asset and which is the derivative is semantic (since the derivative may be more heavily traded than the underlying); highly dependent and liquid assets will result in a near functional relationship between the corresponding asset prices, and will tend to "tie down" each to a functional relationship. Each new liquid financial instrument or asset in a market can substantially effect the model for related assets. Models and their parameters are not only subject to constantly changing economic conditions, they are effected by every related product that enters and leaves the market, by the information and the technology base of traders, by political events and moods.

Today's financial model is almost certainly inadequate tomorrow, and the model parameters are evidently in constant flux. As the complexity of the model changes, and as the explosion of new types of instruments in the market continues to constrain current asset prices in new ways, the need for new statistical and computational tools, often dependent on computer simulation, can only continue to grow. Evolution, from which we have ourselves developed, is a remarkably efficient stochastic optimization (in this case in a high dimensional space). The diversity of models, tools, and approaches to financial analysis that can be accommodated by simulation ensure that our ability to reflect the real processes will continue to improve. I am conscious that only a fraction of these tools and models are discussed here, a tribute to the wealth of research in this important area.