:rences contract, for

alf of the Organizers
ox and D.V. Hinkley
Aarhus and Oxford

CHAPTER 1

Statistical aspects of ARCH and stochastic volatility

Neil Shephard

1.1 Introduction

Research into time series models of changing variance and covariance, which I will collectively call **volatility models**, has exploded in the last ten years. This activity has been driven by two major factors. First, out of the growing realization that much of modern theoretical finance is related to volatility has emerged the need to develop empirically reasonable models to test, apply and deepen this theoretical work. Second, volatility models provide an excellent testing ground for the development of new nonlinear and non-Gaussian time series techniques.

There is a large literature on volatility models, so this chapter cannot be exhaustive. I hope rather to discuss some of the most important ideas, focusing on the simplest forms of the techniques and models used in the literature, referring the reader elsewhere for generalizations and regularity conditions. To start, I will consider two motivations for volatility models: empirical stylized facts and the pricing of contingent assets. In section 1.4 I will look at multivariate models, which play an important role in analysing the returns on a portfolio.

1.1.1 Empirical stylized facts

In most of this chapter I will work with two sets of financial time series. The first is a bivariate daily exchange rate series of the Japanese yen and the German Deutsche Mark measured against the pound sterling, which runs from 1 January 1986 to 12 April 1994, yielding 2160 observations. The second consists of the bivariate daily FTSE 100 and Nikkei 500 indexes, which are market indexes for the London and Tokyo equity markets. These series run from 2 April 1986 to 6 May 1994, yielding 2113 daily observations.

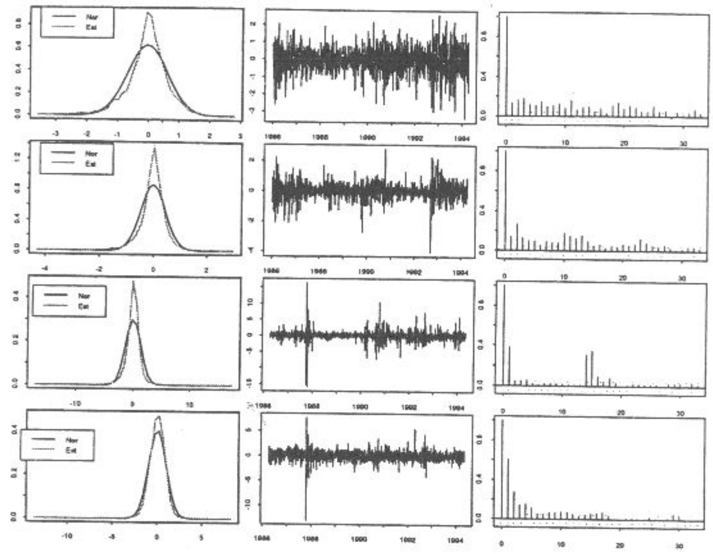
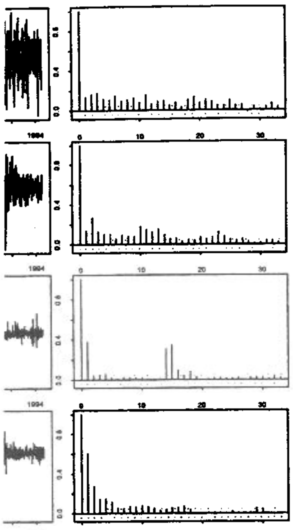


Figure 1.1 *Summaries of the daily returns on four financial assets. From top to bottom: yen, DM, Nikkei 500, FTSE 100. Summaries are: nonparametric density estimate and normal approximation, time series of returns and correlogram of the corresponding squares.*

Throughout I will work with the compounded return on the series $y_t = 100\log(x_t/x_{t-1})$ where x_t is the value of the underlying asset. Figure 1.1 displays some summaries of these two series. It gives a density estimate (using default S-Plus options) of the unconditional distribution of y_t together with the corresponding normal approximation. This suggests that y_t is heavy-tailed. This is confirmed by Table 1.1, which reports an estimate of the standardized fourth moments. In all but the Japanese case they are extremely large.

There is little evidence of any obvious forms of non-symmetry in the unconditional density. A correlogram of y_t shows little activity and so is not given in this figure; Figure 1.1 graphs the raw time series of y_t. Informally this picture suggests that there are periods of volatility clustering: days of large movements are followed by days with the same characteristics. This is confirmed by the use of a correlogram on y_t^2, and the corresponding Box-Ljung statistic reported in Table 1.1, which shows significant correlations which exist at quite extended lag lengths. This suggests that y_t^2 may follow a process close to an ARMA(1,1), for simple AR processes cannot easily combine the persistence in shocks with the



›n four financial assets. From top to
mmaries are: nonparametric density
eries of returns and correlogram of

ɔmpounded return on the series
ıe value of the underlying asset.
of these two series. It gives a
s options) of the unconditional
ːsponding normal approximation.
This is confirmed by Table 1.1,
rdized fourth moments. In all but
rge.
ious forms of non-symmetry in
ım of y_t shows little activity and
.1 graphs the raw time series of
.at there are periods of volatility
e followed by days with the same
e use of a correlogram on y_t^2, and
eported in Table 1.1, which shows
quite extended lag lengths. This
lose to an ARMA(1,1), for simple
he persistence in shocks with the

Table 1.1 *Summary statistics for the daily returns in Figure 1.1. BL denotes the Box–Ljung statistic, computed using the squares, with 30 lags. It should be around 30 if there is no serial dependence in the squares. K denotes the standardized fourth moment of the y_t. K should be around 3 under conditions of normality*

	BL	*K*
Japanese yen	563	5.29
German Deutsche Mark	638	36.2
Nikkei 500 index	828	36.2
FTSE 100 index	1375	28.1

low correlation. Finally, there is some evidence that the exchange rates and equity markets each share periods of high volatility, big movements in one currency being matched by large changes in another. This suggests that multivariate models will be important.

Reasons for changing volatility

It would be convenient to have an explanation for changing levels of volatility. One approach would be to assume that price changes occur as the result of a random number of intra-daily price movements, responding to information arrivals. Hence $y_t = \sum_{i=1}^{n_t} x_{it}$, where x_{it} are independently and identically distributed (i.i.d.) and n_t is some Poisson process. This type of model has a long history, going back to the work of Clark (1973). In this paper the n_t is assumed to be independent over time, which means y_t would inherit this characteristic. It is a trivial matter to allow n_t to be time-dependent, which would lead to volatility clustering, although the resulting econometrics becomes rather involved (see Tauchen and Pitts, 1983).

The more interesting literature in econometric terms is that which ties this information arrival interpretation into a model which also explains volume. The joint models of volume (see also Engle and Russell, 1994) and volatility are the focus of Gallant, Hsieh and Tauchen (1991), who use a reduced-form model, and of Andersen (1995). This is an interesting, but underdeveloped, area.

1.1.2 Pricing contingent assets

Suppose the value of some underlying security, written S, follows a geometric diffusion $dS = \mu S dt + \sigma S dz$, so that $d\log S = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dz$. It is possible to define an asset, c, which is a function of the underlying share price S. Economists call such assets **contingent** or **derivative**. Good introductions to the literature on this topic are given in Ingersoll (1987) and Hull (1993). A primary example of a contingent asset is an **option**, which allows the option owner the ability, but not the obligation, to trade the underlying asset at a given price in the future. The best-known example of this is the European call option whose owner can buy the underlying asset at the fixed price K, at the expiry date $T + v$. An example of K is where it equals $S(T)$, today's price; where the dependence of S on time is now shown explicitly. This special case is called an **at-the-money** option. The value of the general European call option at expiration will be

$$c(T+v) = \max\left\{S(T+v) - K, 0\right\}. \tag{1.1}$$

While equation (1.1) expresses the value of the option at time $T + v$, the option will be purchased at time T, so its purchase value has yet to be determined. A simple approach would be to compute the discounted expected value of the option,

$$\exp(-rv) E_{S(T+v)|S(T)} \max\left\{S(T+v) - K, 0\right\},$$

where r is a riskless interest rate. However, this neglects the fact that traders expect higher returns on risky assets than on riskless assets, a point which will recur in section 1.4. Hence the market will not typically value assets by their expected value. This suggests the introduction of a utility function into the pricing of options, allowing dealers to trade expected gain against risk.

It turns out that the added complexity of a utility function can be avoided by using some properties of diffusions and by assuming continuous and costless trading. This can be seen by constructing a portfolio worth π made up of owning θ of the underlying shares and by borrowing a single contingent asset c. Then the value of the portfolio evolves as

$$\begin{aligned} d\pi &= \theta dS - dc \\ &= \theta\left(\mu S dt + \sigma S dz\right) - (c_s \mu S + c_t + \tfrac{1}{2} c_{ss} \sigma^2 S^2) dt - c_s \sigma S dz \\ &= (\theta - c_s)(\mu S dt + \sigma S dz) - (c_t + \tfrac{1}{2} c_{ss} \sigma^2 S^2) dt, \end{aligned}$$

by using Itô's lemma, where $c_t = \partial c / \partial t$ and $c_s = \partial c / \partial S$. The investor, by selecting $\theta = c_s$ at each time period, can ensure $d\pi$ is instantaneously riskless by eliminating any dependence on the random dz. This result, of making $d\pi$ a deterministic function of time, is due to Black and Scholes (1973). As time passes, the portfolio will have continually to adjust to maintain risklessness – hence the need for continuous costless trading.

As this portfolio is riskless, its return must be the riskless interest rate

r, for otherwise traders will take this arbitrage opportunity and make instant riskless profits. The riskless interest rate can be taken as the return on a very short-duration government bond. Consequently the riskless portfolio follows

$$\begin{aligned} d\pi &= r\pi dt = r(c_s S - c)dt, \text{ as } \pi = c_s S - c \text{ to achieve risklessness} \\ &= -(c_t + \tfrac{1}{2}c_{ss}\sigma^2 S^2)dt, \end{aligned}$$

implying that the contingent asset follows the stochastic differential equation

$$c_t + \frac{1}{2}c_{ss}\sigma^2 S^2 + rc_s S = rc, \quad \text{with end condition } c = \max(S - K, 0).$$

This equation is remarkably simple. In particular, it does not depend on μ or the risk preference of the traders. Hence we can evaluate it as if the world was risk-neutral, in which case we can assume that the share price follows a new diffusion, with mean rS^* and variance $\sigma^2 S^{*2}$: $dS^* = rS^*\,dt + \sigma S^* dz$. This is the **risk-neutral** process; see Hull (1993, pp. 221–222). Using the log-normality of the diffusion we have

$$\log S^*(T+v)|\log S(T) \sim N\{\log S(T) + (r - \sigma^2/2)v, \sigma^2 v\}.$$

Straightforward log-normal results give us the Black–Scholes valuation of the option v periods ahead, using an instantaneous variance of σ^2, of

$$bs_v(\sigma^2) = \exp(-rv)E\left[\max\{S^*(T+v) - K, 0\}\,|S(T)\right]$$

which is

$$bs_v(\sigma^2) = S(T)\Phi(d) - K\exp(-rv)\Phi(d - \sigma\sqrt{v}), \qquad (1.2)$$

where

$$d = \frac{\log\{S(T)/K\} + (r + \sigma^2/2)v}{\sigma\sqrt{v}}. \qquad (1.3)$$

Note that v and K are given by institutional norms, $S(T)$ and r are observed, leaving only σ^2 as unknown. In a real sense, option prices are valuing volatility. As with much of finance, it is the volatility which plays the crucial role, rather than the mean effect.

Empirically there are two straightforward ways of using (1.2). The first is to estimate σ^2 and then work out the resulting option price. The second is to use the observed option prices to back out a value for σ^2. This second method is called an **implied volatility estimate**; see Xu and Taylor (1994) for a modern treatment of this.

A difficulty with all of this analysis is the basic underlying assumption of the process, that stock returns follow a geometric diffusion. Figure 1.1 indicates that this is a poor assumption, in turn suggesting that (1.2) may give a poor rule on which to base option pricing. This realization has prompted theoretical work into option pricing theory under various changing volatility regimes. The leading paper in this field is by Hull and White (1987). I will return to this later.

1.1.3 Classifying models of changing volatility

There are numerous models of changing variance and covariance. A useful conceptual division of the models, following Cox (1981), is into **observation-driven** and **parameter-driven** models. For convenience I will discuss these two approaches within the confines of a tightly defined parametric framework which allows

$$y_t \mid z_t \sim N(\mu_t, \sigma_t^2).$$

For compactness of exposition μ_t will often be set to zero as I do not intend to focus on that feature of the model. Observation-driven models put z_t as a function of lagged values of y_t. The simplest example of this was introduced by Engle (1982) in his paper on autoregressive conditional heteroscedasticity (ARCH). This allows the variance to be a linear function of the squares of past observations

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \ldots + \alpha_p y_{t-p}^2,$$

and so the model becomes one for the one-step-ahead forecast density:

$$y_t \mid Y_{t-1} \sim N(0, \sigma_t^2),$$

where Y_{t-1} is the set of observations up to time $t-1$. This allows today's variance to depend on the variability of recent observations.

Models built out of explicit one-step-ahead forecast densities are compelling for at least three reasons. First, from a statistical viewpoint, combining these densities delivers the likelihood via a prediction decomposition. This means estimation and testing are straightforward, at least in principle. Second, and more importantly from an economic viewpoint, finance theory is often specified using one-step-ahead moments, although it is defined with respect to the economic agents' information set not the econometricians'. The third reason for using observation-driven models is that they parallel the very successful autoregressive and moving average models which are used so widely for models of changing means. Consequently some of the techniques which have been constructed for these models can be used for the new models. ARCH type models have attracted a large amount of attention in the econometrics literature. Surveys of this work are given in the papers by Bollerslev, Chou and Kroner (1992), Bollerslev, Engle and

Nelson (1995), Bera and Higgins (1995) and Diebold and Lopez (1995). Finally, Engle (1995) is an extensive reprint collection of ARCH papers.

Parameter-driven or state-space models allow z_t to be a function of some unobserved or latent component. A simple example of this is the log-normal stochastic variance or volatility (SV) model, due to Taylor (1986):

$$y_t \mid h_t \sim N\{0, \exp(h_t)\}, \quad h_{t+1} = \gamma_0 + \gamma_1 h_t + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2),$$

where *NID* denotes normally and independently distributed. Here the log-volatility h_t is unobserved (at least by the econometrician) but can be estimated using the observations. These models parallel the Gaussian state-space models of means dealt with by Kalman (1960) and highlighted by Harrison and Stevens (1976) and West and Harrison (1989). In econometrics this type of models is associated with the work of Harvey (1989).

Unfortunately, unlike the models of the mean which fit into the Gaussian state-space form, almost all parameter-driven volatility models lack analytic one-step-ahead forecast densities $y_t \mid Y_{t-1}$. As a result, in order to deal with these models, either approximations have to be made or numerically intensive methods used. There seems to be only one constrained exception to this: stochastic volatility models which possess analytic filtering algorithms. Shephard (1994a) suggests setting h_{t+1} to be a random walk with $\exp(\eta_t)$ using a highly contrived scaled beta distribution, following some earlier work on some different non-Gaussian models by Smith and Miller (1986) and Harvey and Fernandes (1989). This delivers a one-step-ahead prediction distribution which has some similarities to the ARCH model. It has been generalized to the multivariate case by Uhlig (1992), who uses it to allow the covariance matrix of the innovations of a vector autoregression to change in a highly parsimonious way. Unfortunately, it does not seem possible to move away from h_{t+1} being a random walk without losing conjugacy. This inflexibility is worrisome and suggests this approach may be a dead end.

Although SV models are harder to handle statistically than the corresponding observation-driven models, there are some good reasons for still investigating them. We will see that their properties are easier to find, understand, manipulate and generalize to the multivariate case. They also have simpler analogous continuous-time representations, which is important given that much of modern finance employs diffusions. An example of this is the work by Hull and White (1987) which uses a log-normal SV model, replacing the discrete-time AR(1) for h_{t+1} with an Ornstein–Uhlenbeck process. A survey of some of the early work on SV models is given in Taylor (1994).

Basic statistical background

To understand the properties of volatility models it is important to have careful definitions of some of the most basic time series concepts for, unusually in time series modelling, small differences in these definitions can have substantial impact. The most commonly used will be:

- White noise (WN). This means $E(y_t) = \mu$, $\text{var}(y_t) = \sigma^2$ and $\text{cov}(y_t, y_{t+s}) = 0$, for all $s \neq 0$. Often μ will be taken to be zero. These unconditional moment conditions are sometimes strengthened to include y_t being independent, rather than uncorrelated, over time. This will be called strong WN, a special case of which is i.i.d.
- Martingale difference (MD). A related concept, y_t being MD stipulates that $E|y_t| < \infty$ and that $E(y_t|Y_{t-1}) = 0$. All MDs have zero means and are uncorrelated over time. If the unconditional variance of the MD is constant over time, then the series is also white noise.
- Covariance stationarity. This generalizes WN to allow autocovariance of the form $\text{cov}(y_t, y_{t+s}) = \gamma(s)$ for all t : the degree of covariance among the observations depends only on the time gap between them. The notation $\text{corr}(y_t, y_{t+s}) = \rho(s) = \gamma(s)/\sigma^2$ denotes the autocorrelation function.
- Strict stationarity. For some models moments will not exist, even in cases where the corresponding unconditional distributions are perfectly well behaved. As a result strict stationarity, where $F(y_{t+h}, y_{t+h+1}, \ldots, y_{t+h+p}) = F(y_t, y_{t+1}, \ldots, y_{t+p})$ for all p and h, will play a particularly prominent role.

1.2 ARCH

The simplest linear ARCH model, ARCH(1), puts:

$$y_t = \varepsilon_t \sigma_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2, \quad t = 1, \ldots, T, \tag{1.4}$$

where $\varepsilon_t \sim NID(0,1)$. The parameter α_1 has to be non-negative to ensure that $\sigma_t^2 \geq 0$ for all t. Crucially $y_t \mid Y_{t-1} \sim N(0, \sigma_t^2)$, which means y_t is a MD and under strict stationarity has a symmetric unconditional density. To show it is zero-mean white noise, we need to find its variance. Clearly the model can be written as a non-Gaussian autoregression:

$$y_t^2 = \sigma_t^2 + (y_t^2 - \sigma_t^2) = \alpha_0 + \alpha_1 y_{t-1}^2 + v_t, \tag{1.5}$$

where $v_t = \sigma_t^2(\varepsilon_t^2 - 1)$ and the sign of y_t is randomized. As v_t is a martingale difference, then if $\alpha_1 \in [0,1)$, $E(y_t^2) = \alpha_0/(1-\alpha_1)$. After some effort, it can be seen that

$$E(y_t^4)/\{E(y_t^2)\}^2 = 3(1-\alpha_1^2)/(1-3\alpha_1^2),$$

if $3\alpha_1^2 < 1$, which when it exists is greater than 3. Under this tight condition, y_t^2 is covariance stationary, its autocorrelation function is $\rho_{y_t^2}(s) = \alpha_1^s$, and y_t has leptokurtosis (fat tails). Notice that $\rho_{y_t^2}(s) \geq 0$ for all s, a result which is common to all linear ARCH models.

These are interesting results. If $\alpha_1 < 1$, y_t is white noise while y_t^2 follows an autoregressive process, yielding volatility clustering. However, y_t^2 is not necessarily covariance stationary for its variance will be finite only if $3\alpha_1^2 < 1$.

The conflicting conditions for covariance stationarity for y_t and y_t^2 prompt the interesting question as to the condition needed on α_0 and α_1 to ensure strict stationarity for y_t. This can be found as a special case of the results of Nelson (1990a), who proved that α_1 had to satisfy $E\{\log(\alpha_1\epsilon_t^2) < 0\}$, which in Gaussian models implies that $\alpha_1 < 3.5622$.

1.2.1 Estimation

At first sight it is tempting to use the autoregressive representation (1.5) to estimate the parameters of the model (this was used by Poterba and Summers, 1986). If v_t is white noise this can be carried out by least squares; in effect this estimate will be reported as the first spike of the correlogram for y_t^2. Although a best linear unbiased estimator, this estimate would be inefficient.

ARCH models, like all observation-driven models, are designed to allow the likelihood to be found easily. Using a prediction decomposition (and ignoring constants):

$$\begin{aligned}\log f(y_1,\dots,y_T \mid y_0;\theta) &= \sum_{t=1}^{T} \log f(y_t \mid Y_{t-1};\theta) \\ &= -\frac{1}{2}\sum_{t=1}^{T}\log\sigma_t^2 - \frac{1}{2}\sum_{t=1}^{T} y_t^2/\sigma_t^2 \quad (1.6)\end{aligned}$$

where θ will denote the parameters which index the model, in this case $(\alpha_0, \alpha_1)'$.

Notice that this likelihood conditions on some prior observations (or in real problems, the first few observations). This is convenient since the analytic form for the unconditional distribution of an ARCH model is unknown. The consequence is that the likelihood does not impose $\alpha_1 < 1$.

Table 1.2 *Aspects of the distribution of the ML estimator of ARCH(1) and GARCH(1,1). RMSE denotes root mean square error. ARCH true values of* $\alpha_0 = 0.2$ *and* $\alpha_1 = 0.9$. *GARCH true values of* $\alpha_0 = \alpha_1 = 0.2$ *and* $\beta_1 = 0.7$. *Based on 1000 replications. Top table – ARCH, bottom table – GARCH*

ARCH

T	$E(\hat{\alpha}_1)$	RMSE($\hat{\alpha}_1$)	Pr($\hat{\alpha}_1 \geq 1$)
100	0.85221	0.25742	0.266
250	0.88386	0.16355	0.239
500	0.89266	0.10659	0.152
1000	0.89804	0.08143	0.100

GARCH

T	$E(\hat{\alpha}_1 + \hat{\beta}_1)$	RMSE($\hat{\alpha}_1 + \hat{\beta}_1$)	Pr($\hat{\alpha}_1 + \hat{\beta}_1 \geq 1$)
100	0.87869	0.14673	0.206
250	0.88680	0.10246	0.143
500	0.89680	0.06581	0.060
1000	0.89913	0.04893	0.019

It is possible to find the scores for the model:

$$\frac{\partial \log f}{\partial \theta} = \frac{1}{2} \sum_{t=1}^{T} \frac{\partial \sigma_t^2}{\partial \theta} \frac{1}{\sigma_t^2} \left(\frac{y_t^2}{\sigma_t^2} - 1 \right), \tag{1.7}$$

where $\partial \sigma_t^2 / \partial \theta = (1, y_{t-1}^2)'$. Typically, even for such a simple model, the likelihood tends to be rather flat unless T is quite large. This means that the resulting maximum likelihood (ML) estimates of α_0 and α_1 are quite imprecise. Table 1.2 gives an example of this, reporting the mean and root mean squared error for the ML estimate of α_1. Notice the substantial probability that the estimated model is not covariance stationary, even when $T = 1000$.

The asymptotic behaviour of the ML estimation of the ARCH model has been studied by Weiss (1986) who showed normality if y_t has a bounded fourth moment. Unfortunately this rules out most interesting ARCH models. More recently Lumsdaine (1991) and Lee and Hansen (1994) have relaxed this condition substantially. Further, both papers look at the consequences of the possible failure of the normality assumption on ε_t (see Bollerslev and Wooldridge, 1992). By relaxing this assumption they treat (1.6) as a quasi-likelihood, which still ensures consistent estimation, but requires the use of the robust sandwich variance estimator (for the calculation of the variance of the estimators, see White, 1982; 1994).

Lee and Hansen (1994) state the following two main sufficient conditions for the consistency of the quasi-likelihood estimator using (1.6):

1. $E(\varepsilon_t \mid Y_{t-1}) = 0, \quad E(\varepsilon_t^2 \mid Y_{t-1}) = 1;$
2. $E\{\log(\alpha_1 \varepsilon_t^2) \mid Y_{t-1}\} < 0.$

The first ensures that the ARCH model correctly specifies the first two moments, the second that y_t is strictly stationary. Asymptotic normality additionally requires that $E(\varepsilon_t^4 \mid Y_{t-1})$ is bounded and that $\alpha_0, \alpha_1 > 0$.

1.2.2 Non-normal conditional densities

The Gaussian assumption on ε_t is arbitrary, indicating that we should explore other distributions. Although ARCH models can display fat tails, the evidence of the very fat-tailed unconditional distributions found for financial data (Mandelbrot, 1963; Fama, 1965) suggests that it may be useful to use models based on distributions with fatter tails than the normal distribution. Obvious candidate distributions include the Student t, favoured by Bollerslev (1987), and the generalized error distribution, used by Nelson (1991); see Evans, Hastings and Peacock (1993) for details of this error distribution. Notice that in both cases it is important to define the new ε_t so that it has unit variance.

Finally, there has recently been considerable interest in the development of estimation procedures which either estimate semi-parametrically the density of ε_t (Engle and Gonzalez-Rivera, 1991) or adaptively estimate the parameters of ARCH models in the presence of a non-normal ε_t (Steigerwald, 1991; Linton 1993). These seem promising areas of research; however, given that parametric estimation of ARCH models requires such large data sets, their effectiveness for real data sets seems questionable.

1.2.3 Testing for ARCH

Using the score (1.7) and corresponding Hessian it is possible to construct a score test of the hypothesis that $\alpha_1 = 0$, i.e. there is no volatility clustering in the series. It turns out to be the natural analogue of the portmanteau score test for AR(1) or MA(1), but in the squares. A generalization to more complicated ARCH models results in the analogue of the Box–Pierce statistic, which uses serial correlation coefficients for the squares

$$r_j = \sum \left(y_t^2 - \overline{y^2}\right)\left(y_{t-j}^2 - \overline{y^2}\right) / \sum \left(y_t^2 - \overline{y^2}\right)^2$$

rather than for the levels. It is studied in Engle, Hendry and Trumble (1985).

More recently, in the econometric literature, some concern has been expressed about the fact that these types of test do not exploit the full information about the model. In particular, $\sigma_t^2 \geq 0$ and so $\alpha_0, \alpha_1 \geq 0$, so that tests of the null hypothesis and more complicated variants of it have to be one-sided. Papers which address this issue include Lee and King (1993) and Demos and Sentana (1994).

1.2.4 Forecasting

One of the aims of building time series models is to be able to forecast. In ARCH models attention focuses not on $E(y_{T+s} \mid Y_T)$ as this is zero, but rather on $E(y_{T+s}^2 \mid Y_T)$ or more usefully, in my opinion, the whole distribution of $y_{T+s} \mid Y_T$.

In ARCH models it is easy to evaluate the forecast moments of $E(y_{T+s} \mid Y_T)$ (see Engle and Bollerslev, 1986; Baillie and Bollerslev, 1992). In the ARCH(1) case

$$E(y_{T+s}^2 \mid Y_T) = \alpha_0(1 + \alpha_1 + \ldots + \alpha_1^{s-1}) + \alpha_1^s y_T^2.$$

In non-covariance stationary cases, such as when $\alpha_1 = 1$, this forecast continually trends upwards, going to infinity with s. This may be somewhat unsatisfactory for some purposes, although if there is not much persistence in the process a normal approximation based on $y_{T+s} \mid Y_T \sim N\{0, E(y_{T+s}^2 \mid Y_T)\}$ may not be too unsatisfactory. That is the conclusion of Baillie and Bollerslev (1992).

In more complicated models it seems sensible to have simple methods to estimate informatively and report the distribution of $y_{T+s} \mid Y_T$. This is studied, using simulation, by Geweke (1989), who repeatedly simulates (1.4) into the future M times, and summarizes the results. A useful graphical representation of the simulation results is the plot of various estimated quantiles of the distribution against s. The results of Koenker, Ng and Portnoy (1994) are useful in reducing the required amount of simulation through smoothing quantile techniques.

1.2.5 Extensions of ARCH

The basic univariate ARCH model has been extended in a number of directions, some dictated by economic insight, others by broadly statistical ideas. The most important of these is the extension to include moving average parts, namely the generalized ARCH (GARCH) model. Its simplest example is GARCH(1,1) which puts

$$y_t = \varepsilon_t \, \sigma_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

This model is usually attributed to Bollerslev (1986), although it was formulated simultaneously by Taylor (1986). It has been tremendously successful in empirical work and is regarded as the benchmark model by many econometricans.

GARCH

The GARCH model can be written as a non-Gaussian linear ARMA model in the squares:

$$y_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + v_t = \alpha_0 + (\alpha_1 + \beta_1) y_{t-1}^2 + v_t - \beta_1 v_{t-1}, \tag{1.8}$$

following (1.5). The original series y_t is covariance stationary if $\alpha_1 + \beta_1 < 1$. In practice the fourth moment of y_t will not usually exist (see the conditions needed in Bollerslev, 1986), but y_t will be strictly stationary if $E \log(\beta_1 + \alpha_1 \varepsilon_t^2) < 1$ and $\alpha_0 > 0$. Nelson (1990a) graphs the combinations of α_1 and β_1 that this allows: importantly, it does include $\alpha_1 + \beta_1 \leq 1$.

The case of $\alpha_1 + \beta_1 = 1$ has itself received considerable attention. It is called integrated GARCH (IGARCH) (see Bollerslev and Engle, 1993). We will see later that for many empirical studies $\alpha_1 + \beta_1$ is estimated to be close to one, indicating that volatility has quite persistent shocks.

In ARCH models the likelihood can be constructed by conditioning on initial observations. In the GARCH(1,1) model both σ_{t-1}^2 and y_{t-1}^2 are required. A standard approach to this problem is to use an initial stretch of 20 observations, say, to calculate σ_{21}^2 by using a simple global variance estimate and computing $\log f(y_{21}, \ldots, y_T | \sigma_{21}^2, y_{20}^2; \theta)$. This is somewhat unsatisfactory, although for large n the initial conditions will not have a substantial impact. Standard normal asymptotics have been proved so long as y_t is strictly stationary (see Lee and Hansen, 1994). Interestingly asymptotic normality does hold for the unit root case, $\alpha_1 + \beta_1 = 1$, unlike for the corresponding Gaussian AR models studied in, for example, Phillips and Durlauf (1986).

To glean some idea of the sampling behaviour of the ML estimator for this model, I repeat the ARCH(1) simulation experiment but now with $\alpha_0 = \alpha_1 = 0.2$ and $\beta_1 = 0.7$. Table 1.2 reports the properties of $\widehat{\alpha}_1 + \widehat{\beta}_1$, as this is the most meaningful parametrization. It inherits most of the properties we found for the ARCH(1) model. Again there is a substantial probability of estimating this persistence parameter as being greater than one.

This model can be generalized by allowing p lags of y_t^2 and q lags of σ_t^2 to enter σ_t^2. This GARCH(p, q) is also strictly stationary in the integrated case, an extension of the GARCH(1,1) case proved by Bougerol and Picard (1992). This suggests normal asymptotics for the ML estimator can also be used in this more complicated situation.

Log GARCH

To statisticians ARCH models may appear somewhat odd. After all $y_t^2 = \varepsilon_t^2 \sigma_t^2$ is a scaled χ_1^2 or gamma variable. Usually when we model the changing mean of a gamma distribution, a log link is used in the generalized linear model (see, for example, McCullagh and Nelder, 1989, Chapter 8). Consequently for many readers a natural alternative to this model might be

$$y_t^2 = \varepsilon_t^2 \exp(h_t), \quad h_t = \gamma_0 + \gamma_1 \log y_{t-1}^2.$$

This suggestion has been made by Geweke (1986) but has attracted little support. A major reason for this is that y_t is often close to zero (or quite often exactly zero). In a rather different context Zeger and Qaqish (1988) have proposed a simple solution to this problem by replacing h_t by

$$h_t = \gamma_0 + \gamma_1 \log\{\max(y_{t-1}^2, c)\}, \quad c > 0.$$

The constant c is a nuisance parameter which can be estimated from the data.

Exponential GARCH

Although the log GARCH models have not had very much impact, another log-based model has, but for rather different reasons. Nelson (1991) introduced an exponential GARCH (EGARCH) model for h_t which in its simplest form is

$$h_t = \gamma_0 + \gamma_1 h_{t-1} + g(\varepsilon_{t-1}), \quad \text{where } g(x) = wx + \lambda(| x | - E | x |). \tag{1.9}$$

The $g(\cdot)$ function allows both the size and sign of its argument to influence its value. Consequently when $\varepsilon_{t-1} > 0, \partial h_t / \partial \varepsilon_{t-1} = w + \lambda$, while the derivative is $w - \lambda$ when $\varepsilon_{t-1} < 0$. As a result EGARCH responds non-symmetrically to shocks.

The Nelson (1991) paper is the first which models the conditional variance as a function of variables which are not solely squares of the observations. The asymmetry of information is potentially useful for it allows the variance to respond more rapidly to falls in a market than to corresponding rises. This is an important stylized fact for many assets (see Black, 1976; Schwert, 1989; Sentana, 1991; Campbell and Hentschel, 1992). The EGARCH model is used on UK stocks by Poon and Taylor (1992).

Although (1.9) looks somewhat complicated, its properties are quite easy to find. As $\varepsilon_{t-1} \sim$ i.i.d., so $g(\varepsilon_{t-1}) \sim$ i.i.d.. It also has zero mean and a constant variance (ε_t is uncorrelated with $\mid \varepsilon_t \mid - E \mid \varepsilon_t \mid$ due to the symmetry of ε_t). As a result, like (1.8), h_t is an autoregression and so is stationary if and only if $\mid \gamma_1 \mid < 1$. Notice that this allows $\rho_{y_t^2}(s)$ to be negative, unlike linear ARCH models. Hence EGARCH models can produce cycles in the autocorrelation function for the squares.

Like the ARCH model, EGARCH is built to allow the likelihood function to be easily evaluated. At present the limit theory for this model has not been rigorously examined, although it seems clear that asymptotic normality will be obtained if $\mid \gamma_1 \mid < 1$.

Decomposing IGARCH

GARCH can be extended to allow arbitrary numbers of lags on y_t^2 and σ_t^2 to enter the variance predictor. A difficulty with this approach is a lack of parsimony, due to the absence of structure in the model. Recently Engle and Lee (1992) have addressed this issue by parametrizing a GARCH model into permanent and transitory components, analogous to the Beveridge and Nelson (1981) decomposition for means. A simple example is

$$\begin{aligned} \sigma_t^2 &= \mu_t + \alpha_1 (y_{t-1}^2 - \mu_t) + \beta_1 (\sigma_{t-1}^2 - \mu_t) \\ \mu_t &= w + \mu_{t-1} + \phi (y_{t-1}^2 - \sigma_{t-1}^2). \end{aligned}$$

Here the intercept of the GARCH process, μ_t, changes over time. As $\Delta \mu_t - w$ is an MD, μ_t is a persistent process tracing the level of the volatility process while σ_t^2 deals with the temporary fluctuations.

It is possible to rewrite this model into its reduced form

$$\begin{aligned} \sigma_t^2 &= w(1 - \alpha_1 - \beta_1) + \{\alpha_1 + \phi(1 - \alpha_1 - \beta_1)\} y_{t-1}^2 - \alpha_1 y_{t-2}^2 \\ &+ \{1 + \beta_1 - \phi(1 - \alpha_1 - \beta_1)\} \sigma_{t-1}^2 - \beta_1 \sigma_{t-2}^2, \end{aligned}$$

which is a constrained IGARCH(2,2) model.

Fractionally integrated ARCH

Volatility tends to change quite slowly, with the effects of shocks taking a considerable time to decay (see Ding, Granger and Engle, 1993). This indicates that it might be useful to exploit a fractionally integrated model. The nonlinear autoregressive representation of ARCH suggests starting with:

$$(1-L)^d y_t^2 = \alpha_0 + v_t, \quad v_t = \sigma_t^2(\varepsilon_t^2 - 1), \quad d \in (-0.5, 0.5),$$

as the simplest fractionally integrated ARCH (FIARCH) model. Rewritten, this gives, say

$$\sigma_t^2 = \alpha_0 + \{1-(1-L)^d\}y_t^2 = \alpha_0 + \alpha(L)y_{t-1}^2.$$

Here $\alpha(L)$ is a polynomial in L which decays hyperbolically in lag length, rather than geometrically. Generalizations of this model introduced by Baillie, Bollerslev and Mikkelsen (1995), straightforwardly transform the ARFIMA models developed by Granger and Joyeux (1980) and Hosking (1981) into long-memory models of variance.

Although these models are observation-driven and so it is possible to write down $f(y_t \mid Y_{t-1})$, Y_{t-1} now has to contain a large amount of relevant data due to the slow rate of decay in the influence of old observations. This is worrying, because the likelihood of ARCH models usually conditions on some Y_0, working with $f(y_1, \ldots, y_T \mid Y_0)$. I think that for these models the construction of Y_0 may be important, although Baillie, Bollerslev and Mikkelsen (1995) argue this is not the case.

Weak GARCH

In this chapter emphasis has been placed on parametric models, which in the ARCH case means models of one-step-ahead prediction densities. Recently there has been some interest in weakening these assumptions, for a variety of reasons. One approach, from Drost and Nijman (1993), is to introduce a class of 'weak' GARCH models which do not build σ_t^2 out of $E(y_t^2 \mid Y_{t-1})$, but instead work with a best linear projection in terms of $1, y_{t-1}, y_{t-2}, \ldots, y_{t-1}^2, \ldots, y_{t-p}^2$.

Weak GARCH has been a useful tool in the analysis of temporally aggregated ARCH processes (see Drost and Nijman, 1993; Nijman and Sentana, 1993) and the derivation of continuous-time ARCH models (Drost and Werker, 1993). However, inference for these models is not trivial for it relies upon equating sample autocorrelation functions with their population analogues. This type of estimator can be ill behaved if y_t^2 is not covariance stationary (a tight condition).

Unobserved ARCH

A number of authors, principally Diebold and Nerlove (1989), Harvey, Ruiz and Sentana (1992), Gourieroux, Monfort and Renault (1993) and King, Sentana and Wadhwani (1994) have studied ARCH models observed with error:

$$y_t = f_t + \eta_t, \quad f_t = \varepsilon_t \sigma_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 f_{t-1}^2, \tag{1.10}$$

where ε_t and η_t are mutually independent and are normally and independently distributed. Their variances are 1 and σ_η^2 respectively. Unlike the other ARCH-type models outlined above, (1.10) is not easy to estimate for it is not possible to deduce $f(y_t \mid Y_{t-1})$ analytically. This is because f_{t-1} is not known given Y_{t-1}. Hence, it makes sense to think of these models as parameter-driven and so classify them as stochastic volatility models.

Approaches to tackling the problem of inference for this model are spelt out by Harvey, Ruiz and Sentana (1992). They employ a Kalman filtering approach based on the state space:

$$\begin{aligned} y_t &= f_t + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2), \\ f_t &= f_t, & f_t &\sim N(0, \alpha_0 + \alpha_1 f_{t-1}^2). \end{aligned}$$

It is possible to estimate f_t and η_t, using the unconditional distribution of f_t as the disturbance of the transition equation. The resulting filter gives a best linear estimator. However, it is inefficient because it ignores the dynamics.

An alternative approach is to use Y_t to estimate f_t and then use that estimate, $\widehat{f}_t$, to adapt the variance of f_{t+1} so that $f_{t+1} \mid Y_t \dot{\sim} N(0, +\alpha_0 + \alpha_1 \widehat{f}_t^{\,2})$. This is the approach of Diebold and Nerlove (1989). The approximation can be improved by noting that

$$f_t^2 = \widehat{f}_t^{\,2} + (f_t - \widehat{f}_t)^2 + 2\widehat{f}_t(f_t - \widehat{f}_t).$$

Taking expectations of this, given Y_t, and using the approximation $\widehat{f}_t \simeq E(f_t | Y_t)$, yields

$$E(f_t^2 \mid Y_t) \simeq \widehat{f}_t^{\,2} + p_t, \text{ where } p_t \simeq E\{(f_t - \widehat{f}_t)^2 \mid Y_t\}.$$

This delivers the improved approximation $f_{t+1} \mid Y_t \dot{\sim} N\{0, \alpha_0 + \alpha_1(\widehat{f}_t^{\,2} + p_t)\}$. As $\widehat{f}_t$ and p_t are in Y_t, if this were the true model, the resulting Kalman filter would be optimal; as it is, Harvey, Ruiz and Sentana (1992) use the phrase 'quasi-optimal' to describe their result. However, as it does not seem possible to prove any properties about this 'quasi-optimal' filter, perhaps a better name would be an 'approximate filter'.

A likelihood-based approach to this model is available via a Markov chain Monte Carlo (MCMC) method, since the model has a Markov random fields structure and so

$$f(f_t|f_{\backslash t}, y) \propto f(f_t|f_{t-1})f(f_{t+1}|f_t)f(y_t|f_t),$$

the notation $f_{\backslash t}$ meaning all elements of $f_1, \ldots, f_n$ except f_t. By continually simulating from $f_t|f_{\backslash t}, y$, for $t = 1, \ldots, n$, a Gibbs or Metropolis sampler can be constructed which converges to a sample from $f_1, \ldots, f_n|y$. These techniques will be spelt out in more detail in the next section.

Other ARCH specifications

There have been numerous alternative specifications for ARCH models. Some of the more influential include those based on:

- Absolute residuals. Suggested by Taylor (1986) and Schwert (1989), this puts:

$$\sigma_t = \alpha_0 + \alpha_1 \mid y_{t-1} \mid .$$

- Nonlinear. The NARCH model of Engle and Bollerslev (1986) and Higgins and Bera (1992) has the flavour of a Box–Cox generalization. It allows:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mid y_{t-1} \mid^{\gamma},$$

or a non-symmetric version:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mid y_{t-1} - k \mid^{\gamma} .$$

- Partially nonparametric model. Early works by Pagan and Schwert (1990) and Gourieroux and Monfort (1992) have tried to let the functional form of σ_t^2, as a response to y_{t-1}, be determined empirically. A simple approach is given in Engle and Ng (1993) who use a linear spline for σ_t^2:

$$\begin{aligned} \sigma_t^2 = \alpha_0 \quad &+ \quad \sum_{j=0}^{m+} \alpha_1^{+j} I(y_{t-1} - \tau_j > 0)(y_{t-1} - \tau_j) \\ &+ \quad \sum_{j=0}^{m^-} \alpha_1^{-j} I(y_{t-1} - \tau_j < 0)(y_{t-1} - \tau_{-j}), \end{aligned}$$

where $I(\cdot)$ are indicator functions and $(\tau_{-m}, \ldots, \tau_m)$ is an ordered set of knots typically set as $\tau_j = j\sqrt{\text{var}(y_t)}$ with $\tau_0 = 0$.

- Quadratic. A related QARCH model of Sentana (1991) has

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_1^* y_{t-1}. \tag{1.11}$$

Clearly there are constraints on the parameters to ensure $\sigma_t^2 \geq 0$. Again (1.11) is used to capture asymmetry.

- Threshold. Various TARCH models have been proposed which have different parameters for $y_{t-1} > 0$ and $y_{t-1} \leq 0$. Zakoian (1990) works with the absolute residuals, while in an influential paper Glosten, Jagannathan and Runkle (1993) work with a model of the type

$$\sigma_t^2 = \alpha_0 + \alpha_1^+ I(y_t > 0) y_{t-1}^2 + \alpha_1^- I(y_{t-1} \leq 0) y_{t-1}^2.$$

  This is used by Engle and Lee (1992), who allow asymmetry to enter the transitory component of volatility, but not the permanent part.

- ARCH in mean. A theoretically important characteristic of excess returns is the relationship between expected returns of a risky asset and the level of volatility. Engle, Lilien and Robins (1987) proposed the ARCH-M model

$$y_t = g(\sigma_t^2, \theta) + \varepsilon_t \sigma_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 \{y_{t-1} - g(\sigma_{t-1}^2, \theta)\}^2.$$

  A commonly used parametrization is the linear one: $g(\sigma_t^2, \theta) = \mu_0 + \mu_1 \sigma_t^2$. Its statistical properties are studied by Hong (1991). Likelihood inference is again straightforward.

1.2.6 Simple empirical illustrations

The simple ARCH-based models are quite easy to fit to data. To illustrate this we will briefly analyze the four series introduced in section 1.1 using GARCH and EGARCH models. Throughout we will work with the compounded return on the series $y_t = 100 \log x_t / x_{t-1}$. The models will be based on Gaussian and Student t distributions where the degrees of freedom are estimated by ML techniques. When the t distribution is used the innovations from the series will be mapped into normality by using the inverse Student t distribution function followed by computing the corresponding normal deviates. These will be used as inputs into the Box–Ljung statistics (Harvey, 1993b, p. 45) for the squares using 30 lags and the standardized fourth moment, or kurtosis, statistic. The first of these statistics should be centred around 30, the second around 3.

Table 1.3 gives the results for GARCH models. To benchmark each of the results I have presented two non-ARCH models: an NID model, whose diagnostics indicate failure because of large degrees of serial correlation in the squares and fat-tails; and an independently distributed (ID) Student t model, which eliminates most of the fat tails problems, but does not deal with the correlation in the squares.

The GARCH models do improve upon these benchmarks. They have two broad effects. First, they successfully deal with the serial correlation

Table 1.3 *Each column represents the empirical fit of a specific GARCH model to the denoted series. When the parameter estimates $\hat{\alpha}_1$ and $\hat{\beta}_1$ are missing, this means they are constrained to being zero. When $\hat{v}$, the degrees of freedom parameter, is missing, it is set to ∞ — giving a normal distribution. BL denotes the Box–Ljung statistic with 30 lags. K denotes the standardized fourth moment of the transformed innovations. K should be around 3*

		Nikkei 500 index		
$\hat{\alpha}_1$			0.198	0.161
$\hat{\beta}_1$			0.834	0.851
$\hat{v}$		3	4	
$\log L$	−3577	−3035	−2836	−3012
BL	828	828	51.5	40.6
K	36.2	3.28	3.14	13.0
		FTSE 100 index		
$\hat{\alpha}_1$			0.116	0.100
$\hat{\beta}_1$			0.879	0.820
$\hat{v}$		5	9	
$\log L$	−2933	−2681	−2595	−2697
BL	1375	1375	6.72	9.01
K	28.1	3.20	3.57	21.1
		German Deutsche Mark		
$\hat{\alpha}_1$			0.135	0.087
$\hat{\beta}_1$			0.902	0.896
$\hat{v}$		3	4	
$\log L$	−1351	−1121	−945.3	−1105
BL	638	638	15.8	20.4
K	10.1	2.90	3.20	9.00
		Japanese yen		
$\hat{\alpha}_1$			0.086	0.045
$\hat{\beta}_1$			0.939	0.945
$\hat{v}$		4	6	
$\log L$	−2084	−1983	−1879	−1937
BL	563	563	30.0	31.2
K	5.29	2.91	3.30	4.77

in the squares. Second, they reduce the fitted value of K in the normal-based model and increase the value of v in the Student t case. Both of these facts indicate that the GARCH model has explained a part of the fat tails in the distribution by a changing variance. However, I am impressed *al fit of a specific GARCH model stimates $\hat{\alpha}_1$ and $\hat{\beta}_1$ are missing, When $\hat{v}$, the degrees of freedom ı normal distribution. BL denotes s the standardized fourth moment ound 3*

) index	
0.198	0.161
0.834	0.851
4	
−2836	−3012
51.5	40.6
3.14	13.0
index	
0.116	0.100
0.879	0.820
9	
−2595	−2697
6.72	9.01
3.57	21.1
sche Mark	
0.135	0.087
0.902	0.896
4	
-945.3	−1105
15.8	20.4
3.20	9.00
yen	
0.086	0.045
0.939	0.945
6	
−1879	−1937
30.0	31.2
3.30	4.77

ted value of K in the normal- in the Student t case. Both of l has explained a part of the fat ınce. However, I am impressed

not by this result, but rather by the transpose of it. In the Nikkei 500 case, v still has to be under 4 for this model successfully to match up with the data. The other cases are nearly as extreme. Consequently, in terms of likelihood reduction, the use of the fat-tailed distribution is as important as the use of GARCH processes in modelling the data. I think this is disappointing, suggesting that GARCH models cannot deal with the extremely large movements in financial markets, even though they are good models of changing variance.

Table 1.4 *Each column represents the empirical fit of a specific EGARCH model to the denoted series. When the parameter estimate $\hat{\omega}$ is missing, this means it is constrained to being zero. When $\hat{v}$, the degrees of freedom parameter, is missing, it is set to ∞ — giving a normal distribution. BL denotes the Box–Ljung statistic with 30 lags. K denotes the standardized fourth moment of the transformed innovations. K should be around 3, even in the t-distribution case*

	Nikkei 500			FTSE 100		
γ_1	0.988	0.985	0.970	0.911	0.907	0.960
θ_1	−0.574	−0.478	−0.292	0.368	0.495	0.254
ω		−0.161	−0.158		−0.059	−0.031
λ	0.482	0.361	0.217	0.211	0.170	0.168
v			5			9
$\log L$	−3017	−2978	−2795	−2694	−2681	−2593
BL	130	46.2	38.6	19.2	21.3	12.9
K	10.9	18.7	3.67	17.8	13.4	3.44
	DM			Yen		
γ_1	0.955	0.969	0.983	0.985	0.981	0.986
θ_1	0.433	−0.139	−0.146	0.231	0.239	−0.170
ω		−0.079	−0.051		−0.027	−0.049
λ	0.154	0.176	0.232	0.089	0.080	0.141
v			4			5
$\log L$	−1119	−1104	−942	−1945	−1940	−1879
BL	19.1	22.1	15.3	33.6	32.2	30.0
K	9.75	9.27	3.61	4.78	4.76	3.20

The results of the GARCH models can be contrasted to the fit of the EGARCH models for these data sets. The results given in Table 1.4, where θ_1 denotes a moving average parameter added to equation (1.9), suggest that the use of the signs of the observations can very significantly improve the fit of the models. In this empirical work this seems to hold for both currencies and equities, although the effect is stronger in the latter. Although this is a standard result for equities (Nelson, 1991;

Brock, Lakonishok and LeBaron, 1992; Poon and Taylor, 1992), it is non-standard for currencies where the asymmetry effects are usually not significant.

1.3 Stochastic volatility

The basic alternative to ARCH-type modelling is to allow σ_t^2 to depend not on past observations, but on some unobserved components or latent structure. The most popular of these parameter-driven stochastic volatility models, from Taylor (1986), puts

$$y_t = \varepsilon_t \exp(h_t/2), \qquad h_{t+1} = \gamma_0 + \gamma_1 h_t + \eta_t,$$

although the alternative parametrization $y_t = \varepsilon_t \beta \exp(h_t/2)$ and $h_{t+1} = \gamma_1 h_t + \eta_t$ also has attractions. One interpretation for the latent h_t is to represent the random and uneven flow of new information, which is very difficult to model directly, into financial markets; this follows the work of Clark (1973) and Tauchen and Pitts (1983).

For the moment ε_t and η_t will be assumed independent of one another, Gaussian white noise. Their variances will be 1 and σ_η^2, respectively. Due to the Gaussianity of η_t, this model is called a **log-normal SV model.** Its major properties are discussed in Taylor (1986; 1994). Broadly speaking these properties are easy to derive, but estimation is substantially harder than for the corresponding ARCH models.

1.3.1 Basic properties

As η_t is Gaussian, h_t is a standard Gaussian autoregression. It will be (strictly and covariance) stationary if $|\gamma_1| < 1$ with:

$$\mu_h = E(h_t) = \frac{\gamma_0}{1-\gamma_1}, \qquad \sigma_h^2 = \text{var}(h_t) = \frac{\sigma_\eta^2}{1-\gamma_1^2}.$$

As ε_t is always stationary, y_t will be stationary if and only if h_t is stationary, for y_t is the product of two stationary processes. Using the properties of the log-normal distribution, if r is even, all the moments exist if h_t is stationary and are given by:

$$E(y_t^r) = E(\varepsilon_t^r) E\left\{\exp\left(\frac{r}{2}h_t\right)\right\} \tag{1.12}$$

$$= r! \exp\left(\frac{r}{2}\mu_h + r^2\sigma_h^2/8\right) / \left(2^{r/2}(r/2)!\right). \tag{1.13}$$

All the odd moments are zero. Of some interest is the kurtosis: $E(y_t^4)/(\sigma_y^2)^2 = 3\exp(\sigma_h^2) \geq 3$. This shows that the SV model has fatter tails than the corresponding normal distribution.

The dynamic properties of y_t are easy to find. First, as ε_t is iid, y_t is an MD and is WN if $|\gamma_1| < 1$. The squares have the product moments $E(y_t^2 y_{t-s}^2) = E\{\exp(h_t + h_{t-s})\}$. As h_t is a Gaussian AR(1),

$$\begin{aligned}\text{cov}(y_t^2, y_{t-s}^2) &= \exp\{2\mu_h + \sigma_h^2(1+\gamma_1^s)\} - \{E(y_t^2)\}^2 \\ &= \exp(2\mu_h + \sigma_h^2)\{\exp(\sigma_h^2\gamma_1^s) - 1\}\end{aligned}$$

and so (Taylor, 1986, pp. 74–75)

$$\rho_{y_t^2}(s) = \text{cov}(y_t^2 y_{t-s}^2)/\text{var}(y_t^2) = \frac{\exp(\sigma_h^2\gamma_1^s) - 1}{3\exp(\sigma_h^2) - 1} \simeq \frac{\exp(\sigma_h^2) - 1}{3\exp(\sigma_h^2) - 1}\gamma_1^s. \tag{1.14}$$

Notice that if $\gamma_1 < 0$, $\rho_{y_t^2}(s)$ can be negative, unlike the ARCH models. This is the autocorrelation function of an ARMA(1,1) process. Thus the SV model behaves in a manner similar to the GARCH(1,1) model.

The dynamic properties of the SV model can also be revealed by using logs. Clearly

$$\log y_t^2 = h_t + \log \varepsilon_t^2, \quad h_{t+1} = \gamma_0 + \gamma_1 h_t + \eta_t, \tag{1.15}$$

a linear process, which adds the iid $\log \varepsilon_t^2$ to the AR(1) h_t. As a result $\log y_t^2 \sim$ ARMA(1,1). If ε_t is normal then $\log \varepsilon_t^2$ has a mean of -1.27 and variance 4.93. It has a very long left-hand tail, caused by taking the logs of very small numbers (see Davidian and Carroll, 1987, p. 1088). The autocorrelation function for $\log y_t^2$ is

$$\rho_{\log y_t^2}(s) = \frac{\gamma_1^s}{(1 + 4.93/\sigma_h^2)}. \tag{1.16}$$

1.3.2 Estimation

The main difficulty of using SV models is that, unlike with ARCH models, it is not immediately clear how to evaluate the likelihood: the distribution of $y_t \mid Y_{t-1}$ is specified implicitly rather than explicitly. Like most non-Gaussian parameter-driven models, there are many different ways to perform estimation. Some involve estimating or approximating the likelihood, others use method-of-moments procedures.

Generalized method-of-moments (GMM)

In econometrics method-of-moments procedures are very popular. There seem to be two main explanations for this: economic theory often specifies that specific variables are uncorrelated and some econometricans are sometimes reluctant to make distributional assumptions. In the SV case neither of these explanations seems very persuasive, for the SV process is a fully specified parametric model.

In the SV case there are many possible moments to use in estimating the parameters of the model. This is because y_t^2 behaves like an ARMA(1,1) model and moving average models do not allow sufficient statistics which are of a smaller dimension than T. This suggests that the use of a finite number of moment restrictions is likely to lose information. Examples include those based on $y_t^2, y_t^4, \{y_t^2 y_{t-s}^2, s = 1, \ldots, S\}$, although there are many other possibilities. As a result, we may well want to use more moments than there are parameters to estimate, implying that they will have to be pooled. A reasonably sensible way of doing this is via generalized method of moments (GMM).

We will write, for example:

$$g_T = \begin{bmatrix} \{\frac{1}{T}\sum y_t^2\} - E(y_t^2) \\ \{\frac{1}{T}\sum y_t^4\} - E(y_t^4) \\ \\ \vdots \\ \\ \{\frac{1}{T}\sum y_t^2 y_{t-1}^2\} - E(y_t^2 y_{t-1}^2) \\ \{\frac{1}{T}\sum y_t^2 y_{t-S}^2\} - E(y_t^2 y_{t-S}^2) \end{bmatrix}$$

as the moment constraints. By varying θ, the $(S+2) \times 1$ vector g_T will be made small. The GMM approach of Hansen (1982) suggests measuring smallness by using the quadratic form $q = g_T' W_T g_T$.The weighting matrix W_T should reflect the relative importance given to matching each of the moments. A good discussion of GMM properties is given in Hamilton (1994, Chapter 14). Earlier versions of GMM include Cramér (1946, p. 425) and Rothenberg (1973).

Applications of this method to SV include Chesney and Scott (1989), Duffie and Singleton (1993), Melino and Turnbull (1990), Jacquier, Polson and Rossi (1994), Andersen (1995) and Andersen and Sorensen (1995). These papers use a variety of moments and weighting matrices W_T. In a rather different vein, Gallant, Hsieh and Tauchen (1994) use a semi-nonparametric ARCH model to provide scores to allow the fitting by 'efficient method of moments'.

It seems to me that there are a number of obvious drawbacks to the GMM approach to the estimation of the SV model:

- GMM can only be used if h_t is stationary. When γ_1 is close to unity, as we will find for many high-frequency financial data sets, we can expect GMM to work poorly.

- Parameter estimates are not invariant. By this I mean that if the model is reparametrized as $\tau = f(\theta)$, then $\widehat{\tau} \neq f(\widehat{\theta})$. This seems important as the direct parameters in the model, γ_0, γ_1 and σ_η^2, are not fundamentally more interesting than other possible parametrizations.
- The squares y_t^2 behave like an ARMA(1,1) model. Equation (1.14) indicates that if σ_η^2 is small (as we will find in practice), $\rho_{y_t^2}(s)$ will be small but positive for many s. Even if γ_1 is close to unity, this will hold. This implies that for many series, S will have to be very high to capture the low correlation/persistence in the volatility process.
- GMM does not deliver an estimate (filtered or smoothed) of h_t. Consequently a second form of estimation will be required to carry out that task.
- Conventional tests of the time series model, based on one-step-ahead prediction densities, are not available after the fitting of the model.

Quasi-likelihood estimation

A rather simpler approach can be based on (1.15). As $\log \varepsilon_t^2 \simeq$ i.i.d., $\log y_t^2$ can be written as a non-Gaussian but linear state space. Consequently the Kalman filter, given in the Appendix, can be used to provide the best linear unbiased estimator of h_t given Y_{t-1}^{l2}, where $Y_s^{l2} = (\log y_1^2, \ldots, \log y_s^2)'$. Further, the smoother gives the best linear estimator given Y_T^{l2}.

This way of estimating h_t is used by Melino and Turnbull (1990), after estimating θ by GMM. The parameters θ can also be estimated, following the suggestion of Harvey, Ruiz and Shephard (1994), using the quasi-likelihood (ignoring constants)

$$lq(\theta; y) = -\frac{1}{2}\sum_{t=1}^{T}\log F_t - \frac{1}{2}\sum_{t=1}^{T} v_t^2/F_t, \qquad (1.17)$$

where v_t is the one-step-ahead prediction error and F_t is the corresponding mean squared error from the Kalman filter. If (1.15) had been a Gaussian state space then (1.17) would be the exact likelihood. As this is not true, (1.17) is called a **quasi-likelihood** and can be used to provide a consistent estimator $\widehat{\theta}$ and asymptotically normal inference. The asymptotic distribution of $\widehat{\theta}$ is discussed in Harvey, Ruiz and Shephard (1994), who use the results of Dunsmuir (1979), and relies on the usual sandwich estimator of quasi-likelihood methods.

To gain some impression of the precision of this method we report, in the first row of Figure 1.2, 500 estimates resulting from the application

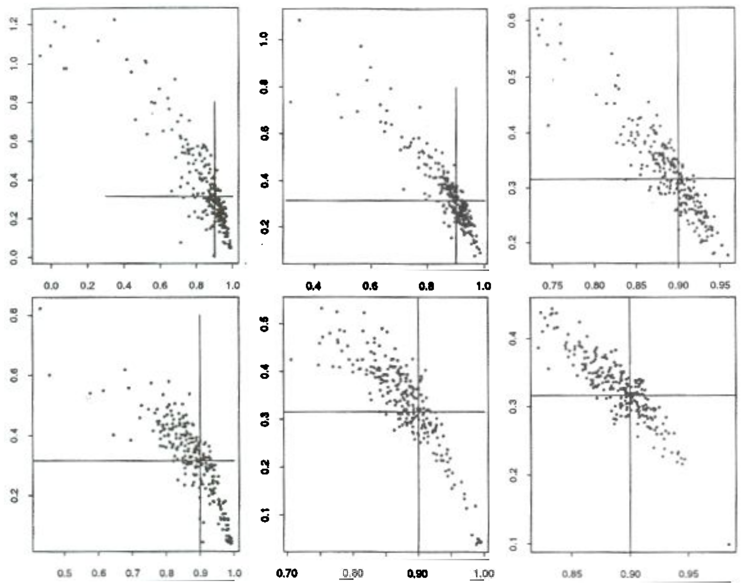
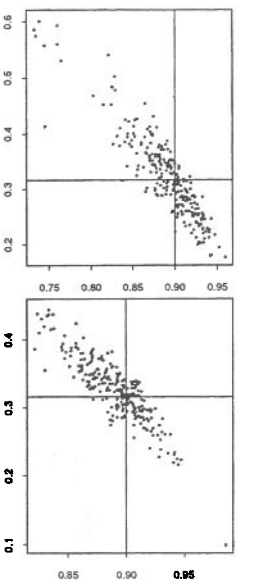


Figure 1.2 *QML and Bayes estimate of SV model. QML on top, Bayes in bottom row. Uses $T = 500$, $T = 1000$, $T = 2000$. T goes from left to right. The Y-axis is $\hat{\sigma}_\eta$ and on X-axis $\hat{\gamma}_1$. The crossing lines drawn on the graphs indicate the true parameter values.*

of the quasi-maximum likelihood (QML) method to simulations from the SV model with $\sigma_\eta^2 = 0.1$, $\gamma_0 = 0.0$, $\gamma_1 = 0.9$ for $T = 500, 1000$ and 2000, focusing on σ_η and γ_1. Notice the strong negative correlation between the two estimators. Later we will compare the properties of this estimator with two other likelihood suggestions. Recently, following a suggestion of Fuller (1996, Example 9.3.2), Breidt and Carriquiry (1995) have investigated modifying the $\log y_t^2$ transformation, to reduce the sensitivity of the estimation procedure to small values of y_t. Their work improves the small-sample performance of the QML estimator.

The mode

A satisfactory way of representing our knowledge of h_t is via its posterior distributions $f(h_t \mid Y_{t-1})$ and $f(h_t \mid Y_T)$. Unfortunately it is not possible to manipulate these densities into useful forms in order straightforwardly to learn their shapes. One approach to overcoming this is to report the mode of the 'smoothing density' $f(h_1, \ldots, h_T \mid y_1, \ldots, y_T)$,



ML on top, Bayes in bottom
from left to right. The Y-axis
the graphs indicate the true

thod to simulations from
= 0.9 for $T = 500, 1000$
rong negative correlation
pare the properties of this
ns. Recently, following a
idt and Carriquiry (1995)
formation, to reduce the
l values of y_t. Their work
QML estimator.

ge of h_t is via its posterior
rtunately it is not possible
in order straightforwardly
coming this is to report
$, \ldots, h_T \mid y_1, \ldots, y_T),$

suggested by Durbin (1992). The approach is based on some recent work by Whittle (1991), Fahrmeir (1992) and particularly Durbin and Koopman (1992) and Durbin (1996).

The mode of $h \mid Y_T$ is the mode of the joint density of h, Y_T. The log of this density, setting $\gamma_0 = 0$ and $h_1 = 0$ for simplicity, is:

$$l = -\frac{1}{2}\sum_{t=1}^{T}\{y_t^2 \exp(-h_t) + h_t + (h_{t+1} - \gamma_1 h_t)^2/\sigma_\eta^2\}. \qquad (1.18)$$

Then $\partial l/\partial h_t$ is nonlinear in h and so we have to resort to solving $\partial l/\partial h = 0$ iteratively. The standard way of carrying this out (ignoring the Gaussian part due to the transition equation) is to start off the kth iteration at $h_1^{(k)}, \ldots, h_T^{(k)}$ and write

$$\begin{aligned}\frac{\partial \log f(y_t|h_t)}{\partial h_t} &= \frac{\partial \log f(y_t|h_t)}{\partial \exp(-h_t)}\frac{\partial \exp(-h_t)}{\partial h_t} \\ &= \{y_t^2 - \exp(h_t)\}\exp(-h_t)/2 \\ &\simeq \Big\{y_t^2 - \exp\left(h_t^{(k)}\right) \\ &- \left(h_t - h_t^{(k)}\right)\exp\left(h_t^{(k)}\right)\Big\}\exp\left(-h_t^{(k)}\right)/2.\end{aligned}$$

This linearized derivative has the same form as a Gaussian measurement model, with

$$\begin{aligned}y_t^2 - \exp\left(h_t^{(k)}\right) - h_t^{(k)}\exp\left(h_t^{(k)}\right) &= \exp\left(h_t^{(k)}\right)h_t + \varepsilon_t, \varepsilon_t \\ &\sim N\left[0, 2\exp\left(h_t^{(k)}\right)\right].\end{aligned}$$

Hence the Kalman filter and analytic smoother (see the Appendix for details), applied to this model, solves the linearized version of the approximation to $\partial l/\partial h_t = 0$. Repeated uses of this approximation will converge to the joint mode as (1.18) is concave in h_t.

Importance sampling

A more direct way of performing inference is to compute the likelihood, integrating out the latent h_t process:

$$f(y_1, \ldots, y_T) = \int f(y_1, \ldots, y_T \mid h) f(h) dh. \qquad (1.19)$$

As this integral has no closed form it has to be computed numerically, integrating over $T \times \dim(h_t)$ dimensional space, which is a difficult task. One approach to this problem is to use Monte Carlo integration, say by drawing from the unconditional distribution of h, with the

jth replication being written as h^j, and computing the estimate $(1/M)\sum_{j=1}^{M} f(y_1, \dots, y_T \mid h^j)$. This is likely to be a very poor estimate, even with very large M. Consequently it will be vital to use an importance sampling device (see Ripley, 1987, p. 122) to improve its accuracy. This rewrites (1.19) as

$$f(y) = \int \frac{f(y \mid h) f(h)}{g(h \mid y)} g(h \mid y) dh,$$

where it is easy to draw from some convenient $g(h \mid y)$. Replications from this density will be written as h^i, giving a new estimate, $(1/M)\sum_{i=1}^{M} f(y \mid h^i) f(h^i)/g(h^i \mid y)$. In some very impressive work, Danielsson and Richard (1993) and Danielsson (1994) have designed g functions which recursively improve (or accelerate) their performance, converging towards the optimal g_*. The details will not be dealt with here as they are quite involved even for the simplest model.

1.3.3 Markov chain Monte Carlo

Although importance sampling shows promise, it is likely to become less useful as T becomes large or $\dim(h_t)$ increases beyond 1. For such difficult problems a natural approach is one based on Markov chain Monte Carlo (MCMC). MCMC will be used to produce draws from $f(h \mid y)$ and sometimes, if a Bayesian viewpoint is taken, the posterior on the parameters $\theta \mid y$. For the moment we will focus on the first of these two targets.

Early work on using MCMC for SV models focused on **single-move** algorithms, drawing h_t individually, ideally from its conditional distribution $h_t \mid h_{\backslash t}, y$, where the notation $h_{\backslash t}$ means all elements of h except h_t. However, a difficulty with this is that although

$$\begin{aligned} f(h_t \mid h_{\backslash t}, y) &= f(h_t \mid h_{t-1}, h_{t+1}, y_t) \\ &\propto f(y_t \mid h_t) f(h_{t+1} \mid h_t) f(h_t \mid h_{t-1}), \quad (1.20) \end{aligned}$$

has an apparently simple form, the constant of proportionality is unknown. As a result it seemed difficult to sample directly from (1.20). Shephard (1993) used a random walk Metropolis algorithm to overcome this problem. A much better approach is suggested in Jacquier, Polson and Rossi (1994), building on the work of Carlin, Polson and Stoffer (1992).

Rejection Metropolis

Jacquier, Polson and Rossi (1994) suggest using a Metropolis algorithm built around an accept/reject kernel, which uses an approximation to

(1.20) which is easy to simulate from, written as $g(h_t \mid h_{\setminus t}, y)$. Then if there exists a c such that

$$f(h_t \mid h_{\setminus t}, y) \leq cg(h_t \mid h_{\setminus t}, y), \quad \forall h_t, \tag{1.21}$$

we could sample from f by drawing from g and accepting this with probability $f(h_t \mid h_{\setminus t}, y) \,/\{g(h_t \mid h_{\setminus t}, y)c\}$.

Jacquier, Polson and Rossi (1994) argued that it was difficult to find a valid g satisfying (1.21) for all h_t, but that this is not so important as we can use a g inside a Metropolis algorithm which overcomes this approximation. Their proposal takes on the form of moving from h_t to h_t^* with probability

$$\min\left[1, \frac{f(h_t^* \mid h_{\setminus t}, y)}{f(h_t \mid h_{\setminus t}, y)} \frac{\min\{f(h_t \mid h_{\setminus t}, y), cg(h_t \mid h_{\setminus t}, y)\}}{\min\{f(h_t^* \mid h_{\setminus t}, y), cg(h_t^* \mid h_{\setminus t}, y)\}}\right].$$

Note that our lack of knowledge of the constant of proportionality in (1.20) is now irrelevant as it cancels in this expression. The choice of g governs how successful this algorithm will be. If it is close to f, c can be close to 1 and the algorithm will almost always accept the moves. Jacquier, Polson and Rossi (1994) suggest using an inverse gamma distribution to approximate the distribution of $\exp(h_t) \mid h_{t-1}, h_{t+1}, y_t$ and a variety of *ad hoc* rules for selecting c.

In my discussion of single-move MCMC I am going to avoid using the Jacquier, Polson and Rossi (1994) method as I think there are now simpler ways of proceeding. One approach is to devise an accept/reject algorithm based around the prior. We write $h_t|h_{t-1}, h_{t+1} \sim N(h_t^*, \sigma_t^2)$, then $\log f(h_t|y_t, h_{t-1}, h_{t+1}) = \text{const} + \log f^*$ where

$$\begin{aligned}
\log f^* &= -\frac{1}{2}h_t - \frac{1}{2\sigma_t^2}(h_t - h_t^*)^2 - \frac{1}{2}\left\{y_t^2 \exp(-h_t)\right\} && (1.22)\\
&\leq -\frac{1}{2}h_t - \frac{1}{2\sigma_t^2}(h_t - h_t^*)^2 && (1.23)\\
&- \left(\frac{y_t^2}{2}\right)\{\exp(-h_t^*)(1 + h_t^*) - h_t \exp(-h_t^*)\} && (1.24)\\
&= \log g^* && (1.25)
\end{aligned}$$

is a bounding function. Hence it is a trivial matter to draw from f using an accept/reject algorithm. The proposal, drawn from the normalized version of g^*, a normal distribution, has mean and variance

$$\mu_t = h_t^* + \frac{\sigma_t^2}{2}\left[y_t^2 \exp(-h_t^*) - 1\right] \quad \text{and} \quad \sigma_t^2 = \sigma_\eta^2 / \left(1 + \gamma_1^2\right).$$

Hence we can sample from $h_t \mid h_{t-1}, h_{t+1}, y_t$ by proposing $h_t \sim N(\mu_t^*, \sigma_t^2)$ and accepting with probability f^*/g^*. This idea, suggested and extended in Pitt and Shephard (1995), gave a 99.5% acceptance rate in a Monte Carlo study using the true values $\gamma_1 = 0.9, \sigma_\eta^2 = 0.05, \gamma_0 = 0$, while it executed about 10 times faster than the code supplied by Jacquier, Polson and Rossi. It is possible also to include the second-order term in the Taylor expansion of $\exp(-h_t)$. This naturally leads to a Metropolis sampler (as it no longer bounds) with a very small rejection probability. Finally, the second-order expansion can be extended to allow a multimove algorithm.

An alternative to this algorithm is to note that $f^*(h_t \mid h_{t-1}, h_{t+1}, y_t)$ is log-concave, which means that the adaptive routines of Gilks and Wild (1992) and Wild and Gilks (1993) can be used.

Whichever of these two algorithms is used, we can now use the Gibbs sampler:

1. Initialize h.
2. Draw h_t^* from $f(h_t \mid h_{t-1}^*, h_{t+1}, y_t), \quad t = 1, \ldots, T$.
3. Write $h = h^*$; go to 2.

This sampler will converge to drawings from $h \mid y$ so long as $\sigma_\eta^2 > 0$. However, the speed of convergence may be slow, in the sense of taking a large number of loops. To illustrate these features, Figure 1.3 reports two sets of experiments, each computed using two sets of parameter values: $\gamma_1 = 0.9$, $\sigma_\eta^2 = 0.1$ and $\gamma_1 = 0.99$, $\sigma_\eta^2 = 0.01$. The first experiment reports the runs of 500 independent Gibbs samplers, each initialized at $h_t = 0$ for all t, as they iterated. The lines are the average of the samplers after given numbers of iterations. This experiment is designed to show how long the initial conditions last in the sampler and so reflect the memory or correlation in the sampler. The second experiment runs a single Gibbs sampler for 100 000 iterations, discarding the first 10 000 results, for the above two sets of parameter values. The resulting 90 000 drawings from $h_{50}|Y_{100}$ were inputted into a correlogram and are reported in Figure 1.3. The idea is to represent the correlation in the sampler once it has reached equilibrium.

The results of Figure 1.3 are very revealing for they show that as γ_1 increases (and similarly as $\sigma_\eta^2 \to 0$), so the sampler slows up, reflecting the increased correlation among the $h \mid y$. This unfortunate characteristic of the single-move Gibbs sampler is common to all parameter-driven models (see Carter and Kohn, 1994; Shephard, 1994b). If a component, such as h_t, changes slowly and persistently, the single-move sampler will be slow. In the limit, when $h_t = h_{t-1}$, the sampler will not converge at all. Given that volatility tends to move slowly, this suggests that this algorithm may be unreliable for real finance problems.

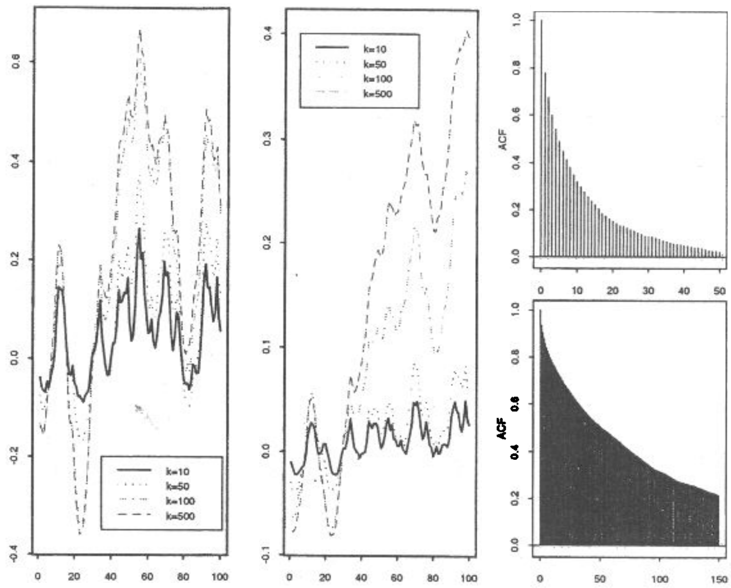


Figure 1.3 *Signal extraction of SV model using single-move Gibbs. Indicates rate of convergence. Left picture is $\gamma_1 = 0.9$ case. Middle gives $\gamma_1 = 0.99$. Correlograms: Top has $\gamma_1 = 0.9$, bottom has $\gamma_1 = 0.99$. In the first two graphs the X-axis is t, and the Y-axis is an estimate of $E(h_t|Y_T)$.*

Multimove samplers

A common solution in the MCMC literature to the problem of elements of $h \mid y$ being highly correlated is to avoid sampling single elements of $h_t \mid h_{\backslash t}, y$, by working with blocks (see Smith and Roberts, 1993, p. 8; Liu, Wong and Kong, 1994). In the context of time series models, early work on designing methods to sample blocks includes that by Carter and Kohn (1994) and Fruhwirth-Schnatter (1994) which has now been refined by de Jong and Shephard (1995). This work can be used to analyse the SV model by using the linear state-space representation:

$$\log y_t^2 = h_t + \log \varepsilon_t^2, \quad \varepsilon_t \sim NID(0,1). \tag{1.26}$$

The idea, suggested in Shephard (1994b) and later used by Carter and Kohn (1994) and Mahieu and Schotman (1994), is to approximate the $\log \varepsilon_t^2$ distribution by a mixture of normals so that:

$$\log \varepsilon_t^2 \mid w_t = j \sim N(\mu_j, \sigma_j^2), \quad j = 1, \dots, J. \tag{1.27}$$

Here the $w_t \sim$ i.i.d., with $\Pr(w_t = j) = \pi_j$. Kim and Shephard (1994) selected μ_j, σ_j^2, π_j for $j = 1, \ldots, 7$ to match up the moments of this approximation to the truth (and various other features of the $\log \chi_1^2$ distribution).

The advantage of this representation of the model is that, conditionally on w, the state space (1.26) is now Gaussian. It is possible to draw $h \mid Y_T$, w directly using the Gaussian simulation smoother given in the Appendix. Likewise, using (1.27) it is easy to draw $w \mid Y_T, h$ using uniform random numbers. This offers the possible **multimove** Gibbs sampler:

1. Initialize w.
2. Draw h^* from $f(h \mid Y_T, w)$.
3. Draw w^* from $f(w \mid Y_T, h^*)$.
4. Write, $w = w^*$; go to 2.

This sampler avoids the correlation in the h process. We might expect $h \mid Y_T$ and $w \mid Y_T$ to be less correlated, allowing rapid convergence. Figure 1.4 shows that this hope is justified, for it repeats the experiment of Figure 1.3 but now uses a multimove sampler. Convergence to the equilibrium distribution appears more rapid, while there is substantially less correlation in the sampler once equilibrium is obtained.

Although the multimoving can be carried out by transforming the model and using a mixture representation, it could be argued that this is only an approximation. It is a challenging problem to come up with multimove algorithms without transforming the model since a fast and, more importantly, reliable sampler will improve the usefulness of these MCMC techniques.

Bayesian estimation

The ability to sample from $h \mid Y_T$ means parameter estimation is reasonably straightforward. The simplest approach to state is the Bayesian one: it assumes a known prior, $f(\theta)$, for $\theta = (\sigma_\eta^2, \gamma_1)$. Then the multimove sample, for example, becomes, when we write $g(h \mid Y_T, \theta)$ to denote a MCMC update using either a Gibbs sampler or multimove sampler:

1. Initialize θ.
2. Draw h^* from $g(h \mid Y_T, \theta)$.
3. Draw θ^* from $f(\theta \mid h^*)$,
4. Write $\theta = \theta^*$; go to 2.

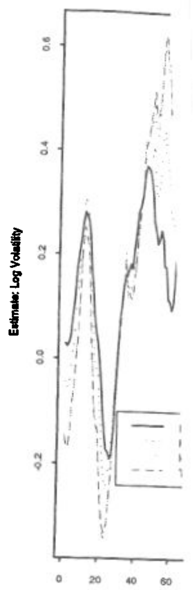


Figure 1.4 *Signal ex of convergence. Left hand picture indicat top corresponds to th high-persistence cas estimate of $Eh_t|Y_T$.*

As the likelihood normal-inverse gar and Rossi (1994). the stationarity con Gaussian prior on γ step 3 into two part

3a. Draw σ_η^{2*} fro
3b. Draw γ_1^* from

The likelihood $f($ prior for $\sigma_\eta^2 \mid h, \gamma_1$ y

$\chi_T^{-2} \{$

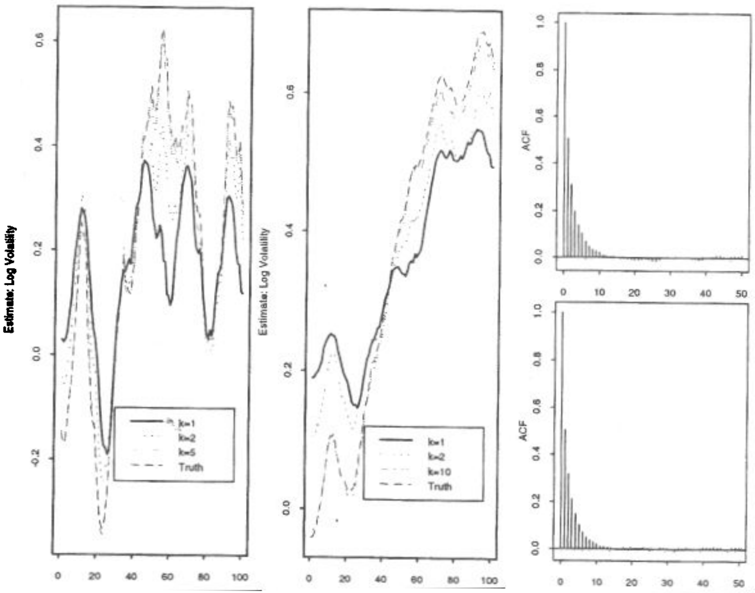


Figure 1.4 *Signal extraction of SV model using multimove Gibbs. Indicates rate of convergence. Left picture is* $\gamma_1 = 0.9$ *case. Middle gives* $\gamma_1 = 0.99$. *The right hand picture indicates serial correlation in the sampler. The correlogram on the top corresponds to the low-persistence case, the one on the bottom represents the high-persistence case. In the first two graphs the X-axis is t, and the Y-axis is an estimate of* $Eh_t|Y_T$. *Truth denotes the actual value of* $Eh_t|Y_T$.

As the likelihood $f(h^* \mid \theta)$ is Gaussian it is tempting to use the standard normal-inverse gamma conjugate prior for θ, as do Jacquier, Polson and Rossi (1994). However, I think there are advantages in enforcing the stationarity conditions on γ_1, and hence on the h process, which a Gaussian prior on γ_1 will not achieve. It can be carried out by dividing step 3 into two parts:

3a. Draw σ_η^{2*} from $f(\sigma_\eta^2 \mid h^*, \gamma_1)$.
3b. Draw γ_1^* from $f(\gamma_1 \mid h^*, \sigma_\eta^{2*})$.

The likelihood $f(h \mid \theta)$ suggests a simple non-informative conjugate prior for $\sigma_\eta^2 \mid h, \gamma_1$ yielding the posterior

$$\chi_T^{-2}\left\{\sum_{t=2}^{T}(h_t - \gamma_1 h_{t-1})^2 + h_1^2(1-\gamma_1^2)\right\},$$

where χ^{-2} denotes an inverse chi-squared distribution. The prior for $\gamma_1 \mid h, \sigma_\eta^2$ is harder due to the non-standard likelihood. However, as $f(h \mid \theta)$ is concave, when $f(\gamma_1)$ is log-concave the Wild and Gilks (1993) method can be used to draw from $\gamma_1 \mid h, \sigma_\eta^2$. A simple example of such a prior is the rescaled beta distribution, with $E(\gamma_1) = \{2\alpha/(\alpha+\beta)\} - 1$. In the analysis I give below, I will set $\alpha = 20$, $\beta = 3/2$ so that the prior mean is 0.86 and the standard deviation 0.11.

One way of thinking of this approach is to regard it as an empirical Bayes procedure, reporting the mean of the posterior distributions as an estimator of θ. This is the approach followed by Jacquier, Polson and Rossi (1994) who show empirical Bayes outperforms QML and GMM in the SV case. Here we confirm those results by repeating the QML experiments reported in section 1.3.2. Again the results are given in Figure 1.2. The gains are very substantial, even for quite large samples.

Simulated EM algorithm

Although the Bayesian approach is simple to state and computationally attractive, it requires the elicitation of a prior. This can be avoided, at some cost, by using MCMC techniques inside a simulated EM algorithm. This was suggested for the SV model by Kim and Shephard (1994). An excellent introduction to the statistical background for this procedure is given in Qian and Titterington (1991); see also the more recent work by Chan and Ledolter (1995). Earlier work on this subject includes Bresnahan (1981), Wei and Tanner (1990) and Ruud (1991).

The EM algorithm works with the mixture-of-normals representation given above, where w is the mixture number. Then

$$\log f(Y_T;\theta) = \log f(Y_T \mid w;\theta) + \log \Pr(w) - \log \Pr(w \mid Y_T;\theta)$$

As $Y_T \mid w$ is a Gaussian state space, its log-density can be evaluated using the Kalman filter

$$\log f(Y_T \mid w;\theta) = \text{const} - \frac{1}{2}\sum_{t=1}^{T}\log F_t - \frac{1}{2}\sum_{t=1}^{T} v_t^2/F_t.$$

As $f(w)$ is parameter-free, the next step of the EM algorithm is found as:

$$\theta^{(i+1)} = \arg\max_{\theta} \sum_{w} \log f(Y_T \mid w;\theta)\Pr(w \mid Y_T;\theta^{(i)}).$$

As $\Pr(w \mid Y_T;\theta^{(i)})$ is unknown, it is not possible to solve this maximization directly. It is replaced by an estimate, using simulations from $\Pr(w \mid Y_T;\theta^{(i)})$ drawn using MCMC techniques. Consequently,

ıtribution. The prior for
likelihood. However, as
ıe Wild and Gilks (1993)
imple example of such a
) = $\{2\alpha/(\alpha+\beta)\} - 1$.
} = 3/2 so that the prior

'egard it as an empirical
terior distributions as an
by Jacquier, Polson and
rforms QML and GMM
. by repeating the QML
the results are given in
for quite large samples.

ate and computationally
This can be avoided, at
ıimulated EM algorithm.
nd Shephard (1994). An
ınd for this procedure is
o the more recent work
ɔn this subject includes
Ruud (1991).
f-normals representation
ıen

$- \log \Pr(w \mid Y_T; \theta)$.

lensity can be evaluated

$- \frac{1}{2}\sum_{t=1}^{T} v_t^2/F_t$.

EM algorithm is found

$\Pr(w \mid Y_T; \theta^{(i)})$.

possible to solve this
mate, using simulations
chniques. Consequently,

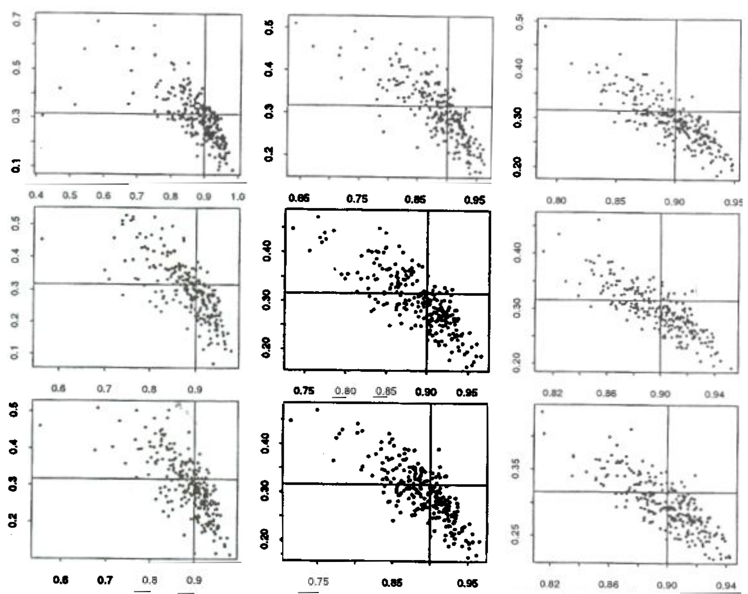


Figure 1.5 *Simulating EM estimate of SV model. Uses $T = 500$, $T = 1000$, $T = 2000$ and $M = 1$, $M = 3$, $M = 10$. T goes from right to left. $M = 1$ in the top row. The Y-axis is $\hat{\sigma}_\eta$ and the X-axis $\hat{\gamma}_1$. The crossing lines drawn on the graphs indicate the true parameter values.*

the function which is numerically maximized is $(1/M)\sum_{i=1}^{M} \log f(Y_T \mid w^i; \theta)$. As $M \to \infty$ so this algorithm converges to the EM algorithm. It may be possible to construct an asymptotic theory for the resulting iterated estimator even if M is finite, using the simulated scores argument of Hajivassiliou and McFadden (1990).

There is some hope that this EM algorithm will converge very quickly to the ML estimator since the 'missing' data w is not very informative about θ. The results, using $M = 1, 3, 10$ and employing 10 steps of the EM algorithm, are reported in Figure 1.5. Kim and Shephard (1994) have found little gain in taking M much bigger than 10, although for more complicated models the situation may be different.

Diagnostic checking

Although there is now a vast literature on fitting SV models, there is barely a word on checking them formally; a notable exception to this is the paper by Gallant, Hsieh and Tauchen (1994). This is an important deficiency. The QML approach offers some potential to close this hole, for the Kalman filter delivers quasi-innovations $\widetilde{v}_t$, which should be uncorrelated (not independent) and have mean squared error F_t. This allows a correlogram, and consequently a Box–Ljung statistic, to be constructed out of $\widetilde{v}_t/\sqrt{F_t}$. However, the distribution of the quasi-innovations is unknowable.

It seems natural to want to work with the true innovations, based on the one-step-ahead forecast's distribution. At first sight MCMC should be able to deliver this distribution, just as it gave us $h_t \mid Y_T$. However, although MCMC methods are good at smoothing, finding the filtering density $h_t \mid Y_{t-1}$ is a more difficult task. The following multimove sampler will work:

1. Initialize $w^{(t)}$.
2. Sample $h^{*(t+1)}$ from $f(h^{(t+1)} \mid Y_t, w^{(t)})$
3. Sample $w^{*(t)}$ from $f(w^{(t)} \mid Y_t, h^{*(t)})$
4. Write $w^{(t)} = w^{*(t)}$; go to 2.

Here $r^{(t)}$ generically denotes $(r_1, \ldots, r_t)$. This MCMC will allow us to sample from $h_{t+1} \mid Y_t$ and so estimate:

$$\Pr(y_{t+1} \leq x \mid Y_t) = F_{y_{t+1}|Y_t}(x) \doteq \frac{1}{M}\sum_{i=1}^{M} \Pr(y_{t+1} \leq x \mid h_{t+1}^{(i)}).$$

These distribution functions or probabilities are vital for they provide the natural analogue of the Gaussian innovations from a time series model. The first reference I know to them is Smith (1985) who noted that it is possible to map them into any convenient distribution to allow easy diagnostic checking. Examples of their use will be given later in section 1.3.5.

A significant difficulty with the MCMC approach is that if T is large it will be computationally expensive. The diagnostic simulation is $O(T^2)$, which is unsatisfactory. Some work on avoiding this has been carried out by Berzuini *et al.* (1994) and Geweke (1994). More work needs to be carried out on this important topic.

1.3.4 Extensions of SV

The basic log-normal SV model can be generalized in a number of directions. A natural framework might be based on adapting the Gaussian state space so that

$$y_t = \varepsilon_t \exp\{z_t' h_t/2\}, \quad h_{t+1} = T_t h_t + \eta_t, \quad \eta_t \sim N(0, H_t).$$

A straightforward generalization might allow h_{t+1} to follow a more complicated ARMA process. Perhaps more usefully, inspiration for new components can be found in the linear models of Harrison, West and Harvey. A simple example would be

$$z_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad h_{t+1} = \begin{pmatrix} \gamma_1 & 0 \\ 0 & 1 \end{pmatrix} h_t + \eta_t,$$

where

$$\eta_t \sim N\left\{0, \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_p^2 \end{pmatrix}\right\}.$$

Now h_{2t+1} is a random walk, allowing the permanent level of the volatility to change slowly. This is analogous to the Engle and Lee (1992) decomposition of shocks into permanent and transitory. A model along the same lines has been suggested by Harvey and Shephard (1993a) and Carter and Kohn (1993), who allow (ignoring the cyclical AR(1) component):

$$z_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad T_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad H_t = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_p^2 \end{pmatrix}.$$

This uses the Kitagawa and Gersch (1984) 'smooth trend' model in the SV context, which in turn is close to putting a cubic spline through the data. This may provide a good summary of historical levels of volatility, but it could be poor as a vehicle for forecasting as confidence intervals for forecasted h_{T+s} may grow very quickly with s. Another suggestion is to allow h_t to be a fractional process, giving the SV model long memory. This has been discussed by Harvey (1993a) and Breidt, Crato and de Lima (1993).

Asymmetric response

One motivation for the EGARCH model introduced by Nelson (1991) was to capture the non-symmetric response of the condition to shocks. A similar feature can be modelled using an SV model by allowing ε_t and η_t to be correlated. Notice ε_t is correlated with η_t, not η_{t-1}. The former model is an MD, the latter is not. If ε_t and η_t are negatively correlated, and if $\varepsilon_t > 0$, then $y_t > 0$ and h_{t+1} is likely to fall. Hence, a large y_t^2's effect on the estimated h_{t+1} will be accentuated by a negative sign on y_t, while its effect will be partially ameliorated by a positive sign.

This correlation between ε_t and η_t was suggested by Hull and White (1987) and estimated using GMM by Melino and Turnbull (1990) and Scott (1991). A simple quasi-likelihood method has been proposed recently by Harvey and Shephard (1993b). Jacquier, Polson and

Rossi (1995) have extended their single-move MCMC sampler to estimate this effect.

Finally, the Engle, Lilien and Robins (1987) ARCH-M model can be extended to the SV framework, by specifying $y_t = \mu_0 + \mu_1 \exp(h_t) + \varepsilon_t \exp(h_t/2)$. This model allows y_t to be moderately serially correlated. It is analysed in some depth by Pitt and Shephard (1995).

1.3.5 Simple empirical applications

To provide a simple illustration of the use of SV models we will repeat the analysis of ARCH-type models presented in the previous section. The SV models used will be the simple AR(1) log-normal-based process, with Gaussian measurement error. We will use a simulated EM algorithm and an empirical Bayes procedure to perform the estimation and use the diagnostic simulator to produce diagnostic checks.

Note that there is a problem with computing the ARCH likelihood. Usually the likelihood for ARCH is found by conditioning on some initial observations. In the previous section the ARCH likelihood was formed by conditioning on 20 initial observations to find σ^2_{20}, and computing a prediction decomposition using the observations from time index 21 to T. In Tables 1.5 and 1.6 we used the unconditional variance to initialize σ^2_0. This technique was used to make the computed likelihood comparable with that for the SV model. The SV model has a properly defined density $f(y_1, \ldots, y_T)$, as h_0 has a proper unconditional distribution. This accounts for the difference in the ARCH likelihoods reported in Tables 1.3, 1.5 and 1.6.

The fitted models are reported in Tables 1.5 and 1.6 and posterior distributions for the parameters are given in Figure 1.6. The approximate symmetry of the posteriors for γ_1 and β means that the empirical Bayes and simulated EM algorithms give very similar results for those parameters. The variance parameter, σ_η has a noticeable right-hand tail and so the result of the empirical Bayes solution having a higher value than the simulated EM algorithm is not surprising.

The results, given in Tables 1.5 and 1.6, suggest that the SV models are empirically more successful than the normal-based GARCH models. This should provide some assurance for option pricing theorists who price assets using this very simple SV model. However, the success of the SV model is accounted for by its better explanation of the fat-tailed behaviour of returns: SV is not a better model of volatility, it is a better model of the distribution of returns. The use of a t-distribution GARCH model overturns this SV outperformance, but not dramatically. The diagnostics of these two models seem similar. There is some evidence that the use nove MCMC sampler to

7) ARCH-M model can be
g $y_t = \mu_0 + \mu_1 \exp(h_t) +$
derately serially correlated.
hard (1995).

V models we will repeat the
he previous section. The SV
normal-based process, with
a simulated EM algorithm
n the estimation and use the
hecks.
iting the ARCH likelihood.
conditioning on some initial
RCH likelihood was formed
o find σ^2_{20}, and computing a
ons from time index 21 to T.
nal variance to initialize σ^2_0.
uted likelihood comparable
del has a properly defined
unconditional distribution.
RCH likelihoods reported in

s 1.5 and 1.6 and posterior
Figure 1.6. The approximate
3 means that the empirical
very similar results for those
s a noticeable right-hand tail
lution having a higher value
prising.
, suggest that the SV models
rmal-based GARCH models.
on pricing theorists who price
owever, the success of the SV
ion of the fat-tailed behaviour
atility, it is a better model of
-distribution GARCH model
dramatically. The diagnostics
is some evidence that the use

Table 1.5 *Empirical fits of SV models. BL denotes the Box–Ljung statistic with 30 lags. K denotes the standardized fourth moment. ARCH denotes the likelihood of the best normal ARCH models. The ARCH model is initialized using $\sigma_0^2 = \alpha_0/(1-\alpha_1-\beta_1)$. The SV model is initialized by the unconditional distribution of h_0. CI denotes a 95% Bayesian confidence interval*

SV	FTSE 100		
	SIEM	Bayes	CI
γ_1	0.945	0.944	[0.903, 0.962]
σ_η	0.212	0.215	[0.116, 0.269]
β	−0.452	−0.446	[−0.660, −0.054]
$\log L$	−2651		
BL	46.1		
K	5.05		
ARCH			
$\alpha_1 + \beta_1$	0.921	0.944	
v		9	
$\log L$	−2725	−2623	
BL	9.0	22.6	
K	21.2	3.21	
SV	Yen		
	SIEM	Bayes	CI
γ_1	0.967	0.957	[0.932, 0.977]
σ_η	0.190	0.223	[0.172, 0.280]
β	−1.14	−1.16	[−0.914, −1.393]
$\log L$	−1929		
BL	36.5		
K	3.67		
ARCH			
$\alpha_1 + \beta_1$	0.994	0.997	
		5	
$\log L$	−1963	−1903	
BL	31.4	37.5	
K	4.81	4.00	

Table 1.6 *Empirical fits of SV models. BL denotes the Box–Ljung statistic with 30 lags. K denotes the standardized fourth moment. ARCH denotes the likelihood of the best normal ARCH models. The ARCH model is initialized using* $\sigma_0^2 = \alpha_0/(1-\alpha_1-\beta_1)$. *The SV model is initialized by the unconditional distribution of* h_0. *CI denotes a 95% Bayesian confidence interval*

SV	Nikkei		
	SIEM	Bayes	CI
γ_1	0.936	0.936	[0.915, 0.968]
σ_η	0.424	0.426	[0.167, 0.261]
β	−0.363	−0.359	[−0.631,−0.260]
$\log L$	−2902		
BL	88.3		
K	5.51		
ARCH			
$\alpha_1+\beta_1$	0.989	0.978	
v		4	
$\log L$	−3036	−2853	
BL	48.4	34.3	
K	13.5	3.00	
SV	DM		
	SIEM	Bayes	CI
γ_1	0.951	0.947	[0.924, 0.967]
σ_η	0.314	0.333	[0.276, 0.390]
β	−2.08	−2.09	[−2.36,−1.81]
$\log L$	−1007		
BL	28.1		
K	4.08		
ARCH			
$\alpha_1+\beta_1$	0.986	0.998	
v		4	
$\log L$	−1127	−968	
BL	19.2	31	
K	9.07	2.96	

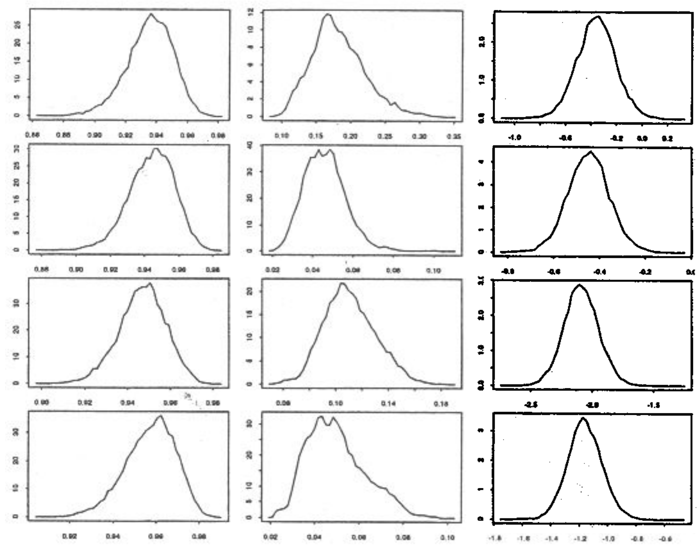


Figure 1.6 *Estimated posterior densities for γ_1, σ_η and β using the multimove Gibbs sampler. Top picture corresponds to the FTSE 100 case, then Nikkei 500, DM and yen.*

of a fat-tailed distribution on either ε_t or, the way I would prefer, on η_t would improve the fit of the SV model. Finally, fitting more complicated SV models, such as ones based on an ARMA(1,1) h_t process, could be empirically more successful for some assets.

An interesting stylized fact to emerge from this table is that the estimated γ_1 parameter is typically lower for SV models than the corresponding $\alpha_1 + \beta_1$ for GARCH. This fact has yet to be explained.

1.4 Multivariate models

Most of macro-economics and finance is about how variables interact, which, for multivariate volatility models, means it is important to capture changing cross-covariance patterns. Multivariate modelling of means is difficult and rather new: constructing multivariate models of covariance is much harder, dogged by extreme problems of lack of parsimony.

1.4.1 Multivariate ARCH

Multivariate ARCH models have existed nearly as long as the ARCH model itself. Kraft and Engle (1982) introduced the basic model which (here translated into a GARCH model) has for an $N \times 1$ series

$$y_t \mid Y_{t-1} \sim N(0, \Omega_t),$$

where

$$\text{vech}(\Omega_t) = \alpha_0 + \alpha_1 \text{vech}(y_{t-1}y'_{t-1}) + \beta_1 \text{vech}(\Omega_{t-1}),$$

where vech$(\cdot)$ denotes the column stacking operator of the lower portion of a symmetric matrix. Deceptively complicated, based on the expansion of the unique elements of Ω_t, this model has $\{N(N+1)/2\} + 2\{N(N+1)/2\}^2$ unknown parameters ($N = 5$ delivers 465 parameters). It is difficult to state the conditions needed for this model to ensure that Ω_t stays positive definite (see Engle and Kroner, 1995).

The multivariate model is virtually useless due to its lack of parsimony. The problem has encouraged a cottage industry of researchers who search for plausible constraints to place on this cumbersome model. An important example is Bollerslev, Engle and Wooldridge (1988), who constrain α_1 and β_1 to be diagonal.

Constant correlation matrix

One of the more empirically successful multivariate ARCH models is the constant correlation model of Bollerslev (1990), who allows the (i, j)th element of Ω_t, Ω_{tij}, to be

$$\Omega_{tij} = \rho_{ij} h_{iit}^{1/2} h_{jjt}^{1/2}, \text{ where } h_{iit} = \alpha_{0i} + \alpha_{1i} y_{it-1}^2 + \beta_{1i} h_{iit-1}.$$

This highly constrained model implies that corr$(y_{it}, y_{jt} \mid Y_{t-1})$ is constant over time. This is often found to be empirically reasonable (see Baillie and Bollerslev, 1990), but it does lack the flexibility required to address some interesting theoretical finance issues which relate to the importance of changing correlation.

1.4.2 Multivariate asset returns

The above multivariate models are either extremely unparsimonious or quite tightly constrained. It seems useful to see if we can look to economic theory to guide us in constructing some more useful models. To start with, I will follow King, Sentana and Wadhwani (1994) and work with an $N \times 1$ series of excess returns (over a riskless interest rate),

$$y_t = \mu_t + \eta_t.$$

Here, given some common information set z_t (perhaps lagged observations or some latent process), μ_t is the expected return of the asset above a safe interest rate, and η_t is the corresponding unexpected component. The covariance of returns will be modelled using a factor structure (see Bartholomew, 1987, pp. 8–9) for η_t, with

$$\eta_t = \sum_{j=1}^{K} b_j f_{jt} + v_t = B f_t + v_t.$$

Here $f_{1t}, \ldots, f_{Kt}$ and v_t will be assumed independent of one another. Then

$$\text{var}(\eta_t \mid z_t) = B \Lambda_t B' + \Psi_t,$$

where

$$\text{var}(f_{jt} \mid z_t) = \sigma_{jt}^2 \text{ and } \Lambda_t = \text{diag}(\sigma_{1t}^2, \ldots, \sigma_{Kt}^2).$$

The Ψ_t will be assumed to be diagonal and sometimes time-invariant. The $N \times (N-K)$ matrix of weights, B, will be called the factor loadings.

This framework can reveal that the covariance structure of the N assets influences the returns μ_t by using the arbitrage pricing theory from Ross (1976). It states that $\mu_t = B\pi_t$, where π_t is a vector whose jth element is the risk premium of a portfolio made up entirely of factor f_{jt}. Hence the risk premium of an asset is a linear combination of the risk premiums on the factors.

Unfortunately the Ross (1976) theory does not tell us how to measure risk premiums, although most finance theorists would put the risk premiums π_t as linear combinations of Λ_t. This has been justified in a formal setting by Hansen and Singleton (1983) in their work on consumption-based asset pricing theory; see also the Appendix of Engle, Ng and Rothschild (1990). In either case this delivers the model for asset returns

$$y_t = B\Lambda_t \tau + B f_t + v_t, \qquad (1.28)$$

where τ is a $K \times 1$ vector of constants. In the univariate model this delivers the ARCH-M and SV-M models outlined in the previous two sections.

From the econometrician's viewpoint (1.28) is a rather incomplete model, as z_t is unspecified. However, it can be completed by using observation-driven or parameter-driven processes, leading to factor ARCH and SV models. In this section the risk premium term $B\Lambda_t\tau$ will tend to be dropped for expositional reasons.

1.4.3 Factor ARCH models

The basis of the factor ARCH model will be

$$y_t = B f_t + \varepsilon_t, \text{ and } f_{it} \mid Y_{t-1} \sim N(0, \sigma_{it}^2), \quad i = 1, \ldots, K,$$

where $\varepsilon_t \sim NID(0, \Psi)$. The important feature of the model is that $f_t = (f_{it}, \dots, f_{Kt})'$ is observation-driven, i.e. conditionally on Y_{t-1} its distribution is specified for this delivers the likelihood.

This model, introduced by Engle, Ng and Rothschild (1990), potentially improves the parsimony problem if a good mechanism for σ_{it}^2 can be found, for now the complexity of the model is really only of the dimension of f_t. To understand the way Engle, Ng and Rothschild (1990) suggest driving σ_{it}^2, it is useful to study briefly some of the features of this model. Then

$$\Omega_t = B\Sigma_t B' + \Psi, \qquad \text{where } \Sigma_t = \text{diag}(\sigma_{it}^2, \dots, \sigma_{Kt}^2).$$

As $B\Sigma_t B'$ is deficient of rank by $N - K$ there exits an $N \times N$ matrix β such that $\beta B = (0 : I_k)$. Consequently, if we write $\beta = (\beta_1', \dots, \beta_N')$, this model allows $N - K$ portfolios $\beta_k' y_t$ to be formed which are homoscedastic. The other portfolios do not have time-varying covariances, just varying variances σ_{kt}^2.

This elegant result suggests forcing σ_{kt}^2 to vary as a GARCH model shocked by past portfolio values $\beta_k' y_t$. Consequently the K factor GARCH(1,1) model becomes:

$$\Omega_t = \Psi^* + \sum_{k=1}^{K} \alpha_k (b_k \beta_k' y_{t-1} y_{t-1}' \beta_k b_k') + \sum_{k=1}^{K} \gamma_k (b_k \beta_k' \Omega_{t-1} \beta_k b_k').$$

Here α_k and γ_k are $N \times N$ matrices. In this model the b_k and β_k are constrained so that $b_k' \beta_j = I(k = j)$ and $\beta_j' \iota = 1$,where ι is the unit vector. Estimation of this type of model is discussed at length by Lin (1992).

1.4.4 Unobserved ARCH

The use of the factor structure seems a real step forward, but the mechanism for driving the σ_{kt}^2 in the factor ARCH model seems quite involved. A simple structure could be obtained by allowing σ_{kt}^2 to be an ARCH process in the unobserved factors f_{it}. This is the suggestion of Diebold and Nerlove (1989) and has been refined by King, Sentana and Wadhwani (1994). The econometrics of this model is a straightforward generalization of the univariate case outlined in section 1.2.

ature of the model is that
i.e. conditionally on Y_{t-1}
e likelihood.
; and Rothschild (1990),
if a good mechanism for σ^2_{it}
model is really only of the
e, Ng and Rothschild (1990)
efly some of the features of

iiag$(\sigma^2_{it}, \ldots, \sigma^2_{Kt})$.

K there exits an $N \times N$
equently, if we write $\beta =$
rtfolios $\beta'_k y_t$ to be formed
os do not have time-varying

o vary as a GARCH model
Consequently the K factor

$$\sum_{k=1}^{K} \gamma_k (b_k \beta'_k \Omega_{t-1} \beta_k b'_k).$$

this model the b_k and β_k
and $\beta'_j \iota = 1$,where ι is the
el is discussed at length by

real step forward, but the
r ARCH model seems quite
ied by allowing σ^2_{kt} to be an
$_{it}$. This is the suggestion of
efined by King, Sentana and
s model is a straightforward
d in section 1.2.

1.4.5 Multivariate SV models

Some multivariate SV models are easy to state. Harvey, Ruiz and Shephard (1994) used quasi-likelihood Kalman filtering techniques on

$$y_{it} = \varepsilon_{it} \exp(h_{it}/2), \quad i = 1, \ldots, N,$$

where

$$\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{NT})' \sim NID(0, \Sigma_\varepsilon),$$

in which Σ_ε is a correlation matrix (this model can be viewed as a generalization of the discounted dynamic models of Quintana and West, 1987). They allow $h_t = (h_{it}, \ldots, h_{NT})'$ to follow a multivariate random walk, although more complicated linear dynamics could be handled (Harvey, 1989, Chapter 8). The approach again relies on linearizing (this time with loss of information) by writing $\log y^2_{it} = h_{it} + \log \varepsilon^2_{it}$. The components of the vector of $\log \varepsilon^2_{it}$ are iid, all with means -1.2704, and a covariance matrix which is a known function of Σ_ε (Harvey, Ruiz and Shephard, 1994, p. 251). Consequently Σ_ε and the parameters indexing the dynamics of h_t can be estimated.

There are two basic points about this model. First, it allows common trends and cycles in volatility by placing reduced rank constraints on h_t, paralleling the work of Harvey and Stock (1988) on the levels of income and consumption. Second, the model is one of changing variances, rather than changing correlation, similar to the Bollerslev (1990) model of constant conditional correlation. Consequently, this model can be empirically successful, but it is of limited interest since it cannot model some theoretically important features of the data. Other work on this model includes Mahieu and Schotman (1994), while Jacquier, Polson and Rossi (1995) look at using an MCMC sampler on this model.

1.4.6 Multivariate factor SV model

Perhaps a more attractive multivariate SV model can be built out of factors. The simplest one-factor model puts

$$\begin{aligned} y_t &= \beta f_t + w_t, \quad w_t \sim NID(0, \Sigma_w) \\ f_t &= \varepsilon_t \exp(h_t/2), \text{ where } h_{t+1} = \gamma_1 h_t + \eta_t, \quad \eta_t \sim NID(o, \sigma^2_\eta). \end{aligned}$$

Here w_t obscures the scaled univariate SV model f_t. Typically Σ_w will be assumed diagonal, perhaps driven by independent SV models. It is similar in spirit to the Diebold and Nerlove (1989) model.

Direct Kalman filtering methods do not seem effective on these models as there is no obvious linearizing transformation. MCMC methods do not suffer this drawback and are explored in Jacquier, Polson and Rossi (1995) and Pitt and Shephard (1995).

1.5 Option pricing with changing volatility

There is now a considerable literature on computing option prices on assets with variances which change over time. Most – for example Taylor (1986, Ch. 9), Hull and White (1987), Scott (1987), Wiggins (1987), Melino and Turnbull (1990), Stein and Stein (1991) and Heston (1993) – use an SV framework (all but Taylor in continuous-time). Recently there has been some work on option pricing based on ARCH-driven models; see Engle and Mustafa (1992), Engle, Hong, Kane and Noh (1993), Engle, Kane and Noh (1993), Heynen, Kemna and Vorst (1994), Noh, Engle and Kane (1994) and Duan (1995). In both cases it is not possible to use the risk-elimination methods discussed earlier for the Black–Scholes constant variance solution, for there is no direct market which trades volatility. However, under the assumption that volatility is uncorrelated with aggregate consumption, progress can be made (Hull and White, 1987, p. 283). In this discussion we will also assume that volatility is uncorrelated with the stock price itself, a less satisfactory assumption except for currencies, although this can be removed.

1.5.1 Stochastic volatility

Hull and White (1987) set up a risk-neutral world and exploit the standard diffusion for $dS = rSdt + \sigma Sdz$, the stock price, and add $dV = \lambda V dt + \theta V dw$, allowing λ to depend on $V = \log \sigma^2$. Then they show a fair European call option price would be

$$\begin{aligned} hw_v &= e^{-rv} \int \max\{S(T+v) - K, 0\} \\ &\qquad f\{S(T+v) \mid S(T), \sigma^2(T)\} dS(T+v), \end{aligned}$$

assuming $\sigma^2(T)$ is observable.

This integral has been solved analytically, using characteristic function inversions, in a number of recent papers (Stein and Stein, 1991; Heston 1993). There are, however, substantial benefits from simplification which can be used if simple simulation techniques are to be used to approximate it. If we write σ^2 as the sample path of the volatility, then

$$\log\{S(T+v)/S(T)|\sigma^2\} \sim N(rv - v\overline{\sigma}^2/2, v\overline{\sigma}^2),$$

where

$$\overline{\sigma}^2 = \frac{1}{v}\int_T^{T+v} \sigma^2(t)dt,$$

is the average level of volatility during the option. This implies

$$hw_v = e^{-rv}\int\int \max\{S(\tau+v)-K,0\}f\{S(T+v)\mid S(T),\overline{\sigma}^2\}$$
$$dS(T+v)f\{\overline{\sigma}^2\mid\sigma^2(T)\}d\overline{\sigma}^2$$

The inner of these integrals is the Black–Scholes pricing formula (1.2) replacing σ^2 by $\overline{\sigma}^2$. This leaves the integral in its 'Rao–Blackwellized' form,

$$hw_v = \int bs_v(\overline{\sigma}^2)f\{\overline{\sigma}^2\mid\sigma^2(T)\}d\overline{\sigma}^2, \tag{1.29}$$

which is much easier to solve by simulation.

In practice, the diffusions will be discretized, perhaps into the SV models discussed earlier. Then we will need to take into account that $\sigma^2(T)$ is unobserved.

A simple approach to this problem is to draw m MCMC replications from $f(h \mid y)$ to construct a fixed population of initial conditions $h_T^1,\ldots,h_T^m$. Then, sampling with replacement from this population, we can initialize a draw from a whole sequence of future $h_{T+1}^j,\ldots,h_{T+v}^j$. Each sequence draws a mean

$$\overline{\sigma}^{2j} = \frac{1}{v}\sum_{t=1}^{v}\exp(h_{T+t}^j)$$

to provide an estimate of (1.29) $(1/R)\sum_{j=1}^R bs_v(\overline{\sigma}^{2j})$.

As Hull and White (1987) noted, it is possible to use antithetic variables productively (Ripley, 1987, pp. 129–132) for the shocks η_t^j in the autoregression, since h_t^j appear monotonically in the replications of $\overline{\sigma}^{2j}$. This means that if η_t^j are negatively correlated across j then the resulting $\overline{\sigma}^{2j}$ will also be negatively correlated (see Ripley, 1987, Theorem 5.2, p. 129), reducing the variance of the Monte Carlo estimate. An obvious example is to draw double replications based on

$$h_{T+t+1}^{j+} = \gamma_1 h_{T+t}^{j+} + \eta_{T+t} \quad \text{and} \quad h_{T+t+1}^{j-} = \gamma_1 h_{T+t}^{j-} - \eta_{T+t},$$

starting off $h_T^{j+} = h_T^{j-}$ at the same point.

1.5.2 ARCH modelling

It is not so straightforward to carry out option pricing based on ARCH models of the form

$$\log S_{t+1} = r + \log S_t + y_t, \quad \text{where } y_t \sim \text{GARCH}.$$

The main reason for this is that $\log\{S_{T+v}/S_T|\sigma^2\}$ is no longer normally distributed. The implication of this is that the ARCH option pricing formula has to be estimated by the path-dependent summation

$$arch_v = e^{-rv}\sum_{j=1}^{R}\max\{S^j_{T+v} - K, 0\},$$

where S^j_{T+v} are GARCH simulations into the future starting off using Y_T. This is much less satisfactory, implying R will have to be much higher for this model than for the corresponding SV model as it depends on the sample path of the observations rather than the volatility.

To overcome this difficulty, Noh, Engle and Kane (1994) use a 'plug-in' estimate of future average volatility to deliver

$$arch_v \simeq bs_v(\sigma^{(v)}_{t+1|t}) \quad \text{where } \sigma^{(v)}_{t+1|t} = \frac{1}{v}\sum_{i=1}^{v} Ey^2_{t+i}|Y_T.$$

Although this is an approximation, based on the expectation of a function being approximately the function of the expectation, Noh, Engle and Kane (1994) provide some evidence that this is a sufficiently good improvement over existing technology to provide a trading profit on at-the-money (where $K = y_T$) Standard and Poor's 500 index European options. Further, from a theoretical viewpoint Cox and Rubinstein (1985, p. 218) show that the Black–Scholes pricing equation is essentially a linear function of the standard deviation around at-the-money prices. Hence the plug-in approximation is likely to be good for at-the-money options, although at other prices it could be poor.

1.5.3 Applying volatility models

Even though it is not possible to predict the way the market is going to move, it may be possible to use the volatility models to trade profitably in the options market. Suppose we believe that in the next v periods the market will be more volatile than do other traders. We should buy both call and put options: if the market goes down we would exercise the put option, if it goes up the call option would be used. Thus this **straddle** is of value in volatile times. If the opposite is true, we can sell a straddle.

This strategy of using straddles potentially allows us economically to value the use of the various volatility models: implied volatility, ARCH and stochastic volatility. The first two of these are compared in Noh, Engle and Kane (1994). I have not seen any work on using straddles to value SV models. Of course in practice this type of trading, particularly the selling of straddles, can be very risky; see the report in the *Financial Times* on the collapse of Barings Bank by Martin (1995).

1.6 Continuous-time models

Continuous-time models play a crucial role in finance and some economic theory, often providing very simple and powerful solutions to difficult problems. An example of this is the Hull and White (1987) generalization of the Black–Scholes option pricing formula to allow the instantaneous variance of returns to change over time. Two obvious questions are:

- Should we use continuous-time (tick-by-tick) data to estimate continuous- and discrete-time models?
- How do we use daily or weekly data to estimate continuous-time models?

I think the answer to the first of these is probably no. The reason for this is that, as with continuous-time models in other fields, continuous-time models are grossly untrue over short periods of time due to institutional factors. An example of this is the large seasonal patterns of volatility which occur during the day (see Foster and Viswanathan, 1990), due to the opening, lunch hour and closing of various markets around the world. These types of feature introduce high-dimensional unstable nuisance features into the modelling process. As they are of little interest, it seems sensible to abstract from them.

This criticism of the use of continuous-time data suggests that the highly impressive work of Nelson (1992; 1995) and Nelson and Foster (1994), who have used the continuous record asymptotics, may not be a particularly fruitful approach to the direct estimation of models (although the asymptotics are useful at indicating the link between different models). This is because that work relies on building ARCH-type models by studying approximations (for example in Nelson, 1990b; 1995) of continuous-time SV models as the time gap between observations falls in a particular way. Thus, at least in theory, there is a belief that the continuous-time data could be used to perform filtering and smoothing.

The second question is the subject of considerable research attention at the moment which we will briefly discuss here.

1.6.1 Direct estimation

There is little work on the direct estimation of continuous-time models using discrete-time data. The reason for this is the nonlinear nature of the volatility models which make the tools developed for linear models (see Bergstrom, 1983; Harvey and Stock, 1985) not directly useful.

Recently there has been an upswing of interest in using generalized method of moments criteria to estimate continuous-time SV models. Leading work on this topic includes that by Duffie and Singleton (1993) and Ho, Perraudin and Sorensen (1993). Given the gross inefficiency of GMM in discrete time, it is doubtful that these approaches will be particularly effective at dealing with continuous-time models.

1.6.2 Indirect inference

A promising approach to using a discrete-time model to perform inference on continuous models has been suggested by Smith (1993) and elaborated by Gourieroux, Monfort and Renault (1993) and Gallant and Tauchen (1995). Gourieroux, Monfort and Renault (1993) call this procedure **indirect inference**.

The idea is to use an approximate model's objective function, $Q(y;\beta)$, such as a likelihood or quasi-likelihood from a discrete-time model, as the basis of inference for a fully parametric continuous-time model, indexed by parameters θ, where the likelihood is difficult to calculate. We write

$$\widehat{\beta} = \arg\max_{\beta} Q(y;\beta) \quad \text{and} \quad \widetilde{\beta}^h(\theta) = \arg\max_{\beta} Q(\widetilde{y}^h(\theta);\beta)$$

where $\widetilde{y}^h(\theta)$ is the hth simulation using the continuous-time model with the parameter θ. Then we solve

$$\widehat{\theta} = \arg\max_{\theta} \left\{\widehat{\beta} - \frac{1}{H}\sum_{h=1}^{H}\widetilde{\beta}^h(\theta)\right\}' \Omega^{-1} \left\{\widehat{\beta} - \frac{1}{H}\sum_{h=1}^{H}\widetilde{\beta}^h(\theta)\right\},$$

for some choice of Ω. The asymptotic properties of $\widehat{\theta}$ are studied in Gourieroux, Monfort and Renault (1993).

A slight variant of this approach can be found in Gallant and Tauchen (1995), who work with

$$\widehat{\theta} = \arg\max_{\theta} \left\{\frac{1}{H}\sum_{h=1}^{H}\frac{\partial Q}{\partial \beta'}(\widetilde{y}^h(\theta);\widehat{\beta})\right\}' \Sigma^{-1} \left\{\frac{1}{H}\sum_{h=1}^{H}\frac{\partial Q}{\partial \beta}(\widetilde{y}^h(\theta);\widehat{\beta})\right\}.$$

In the case where the dimension of β equals that of θ, this is equivalent to solving for $\widehat{\theta}$ the equation

$$\frac{1}{H}\sum_{h=1}^{H}\frac{\partial Q}{\partial \beta}(\widetilde{y}^h(\widehat{\theta});\widehat{\beta}) = 0.$$

Note that the data only play a role through $\widehat{\beta}$.

To illustrate this procedure we will first work with a toy example. Suppose $y_i \sim N(\mu, 1)$, and that we use a correctly specified model. Then, setting $x_i^h \sim NID(0,1)$ and writing θ as the mean under estimation, the indirect inference estimator becomes

$$\widehat{\theta} = \arg\max_{\theta} \{\overline{y} - (\theta + \overline{x})\}^2,$$

yielding $\widehat{\theta} = \overline{y} + \overline{x}$. The Gallant and Tauchen (1995) approach yields the same estimator, via

$$\frac{\partial Q}{\partial \mu} = \sum_{h=1}^{H}\sum_{i=1}^{T}(y_i^h - \overline{y}) = TH(\theta - \overline{y}) + \sum_{h=1}^{H}\sum_{i=1}^{T} x_i^h, \text{ where } y_i^h = \theta + x_i^h.$$

This implies $\widehat{\theta} \sim N\left(\theta, \frac{1}{T}(1 + \frac{1}{H})\right)$. For more complicated problems $\widehat{\theta}$ remains consistent for finite H as $T \to \infty$.

The indirect inference method can be used to estimate SV models using an ARCH likelihood. One approach to estimating the continuous-time SV models is to use a discrete-time GARCH(1,1) model as a template. This is convenient as the GARCH model has an analytic likelihood. It is also straightforward to simulate from the continuous-time SV model by taking a very fine discrete-time approximation (e.g. split each day into ten parts). This would give consistent estimation of the continuous-time models. This agenda has recently been followed by Engle and Lee (1994). The efficiency of these types of procedure compared to some likelihood-based methods is still open to question.

1.7 Concluding remarks

I have tried to develop a balanced introduction to the current literature on ARCH and SV models. The work on ARCH models is somewhat more mature than the corresponding SV analysis, but both areas of research are vibrant.

I think it is clear from my survey that the development of univariate discrete-time ARCH is a thoroughly mined area. It would be surprising if there were any really major contributions which could be made to this subject. However, multivariate models are still in their infancy and

I believe the best work on this subject has yet to be written. The key to achieving parsimony must be the combination of economic and time series insight.

SV models are much newer. Their close connection to continuous-time models provides a strong motivation; building an appropriate econometric toolbox to treat them is still an ongoing project. Fast MCMC algorithms need to be developed, while the interplay of good empirical model checking and building has only just started for univariate models. There is much to be done.

The use of continuous-time models in theoretical finance provides a large incentive for econometricians to think about how to make inferences on these models using discrete-time data. This is going to be an extremely active area of research in the next decade.

1.8 Appendix

This Appendix focuses on the computational issues of filtering and smoothing. It provides no direct insight into ARCH or SV models.

1.8.1 Gaussian state-space form

The state-space form

$$\begin{aligned} y_t &= Z_t\alpha_t + G_t u_t, \quad u_t \sim NID(0, I), \\ \alpha_{t+1} &= T_t\alpha_t + H_t u_t, \\ \alpha_1|Y_0 &\sim N(a_{1|0}, P_{1|0}), \end{aligned}$$

has a prominent role in modern time series (see Hannan and Deistler, 1988; Harvey, 1993b). It provides, by appropriate selection of Z_t, G_t, T_t and H_t, a unified representation of all linear Gaussian time series models. For simplicity we assume that $G_t'H_t = 0$ and we write the non-zero rows of H_t as M_t, $G_tG_t' = \Sigma_{\epsilon t}$ and $H_tH_t' = \Sigma_{\eta t}$. An example of this is an AR(2) with measurement error. Using an obvious notation,

$$Z_t = \begin{pmatrix} 1 & 0 \end{pmatrix}, G_t = \begin{pmatrix} \sigma_\varepsilon & 0 \end{pmatrix}, T_t = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix},$$

$$H_t = \begin{pmatrix} 0 & \sigma_\eta \\ 0 & 0 \end{pmatrix}, M_t = \begin{pmatrix} 0 & \sigma_\eta \end{pmatrix}.$$

1.8.2 Kalman filter

The Kalman filter (Harvey, 1989) plays a crucial computational role in the analysis of models in state-space form. In particular, if we write

$a_{t|t-1} = E\alpha_t|Y_{t-1}$ and $P_{t|t-1} = \text{MSE}(\alpha_t|Y_{t-1})$, then the Kalman filter computes these quantities recursively for $t = 1, \ldots, T$,

$$\begin{aligned} a_{t+1|t} &= T_t a_{t|t-1} + K_t v_t, & P_{t+1|t} &= T_t P_{t|t-1} L_t' + \Sigma_{\eta t}, \\ v_t &= y_t - Z_t a_{t|t-1}, & F_t &= Z_t P_{t|t-1} Z_t' + \Sigma_{\varepsilon t}, \\ K_t &= T_t P_{t|t-1} Z_t' F_t^{-1}, & L_t &= T_t - K_t Z_t. \end{aligned} \tag{1.30}$$

A by-product of the filter are the innovations v_t, which are the one-step-ahead forecast errors, and their corresponding mean squared errors, F_t. Together they deliver the likelihood, such as (1.17).

1.8.3 Simulation smoother

Traditionally the posterior density of α given Y_T is called the **smoothing density**. In the Gaussian case there are algorithms which compute the mean and covariance terms of this highly multivariate distribution. I call these algorithms **analytic smoothers**. More recently a number of algorithms, labelled **simulation smoothers**, have been proposed for drawing random numbers from $\alpha|Y_T$. This is useful in Gibbs sampling problems. The first algorithms were proposed by Carter and Kohn (1994) and Fruhwirth-Schnatter (1994). We exploit the approach of de Jong and Shephard (1995), which is computationally simpler, more efficient and avoids singularities. This requires that F_t^{-1}, v_t and K_t be stored from the run of the Kalman filter.

Setting $r_T = 0$ and $N_T = 0$, we run for $t = T, \ldots, 1$

$$\begin{aligned} C_t &= M_t M_t' - M_t H_t' N_t H_t M_t', \\ r_{t-1} &= Z_t' F_t^{-1} v_t + L_t' r_t - L_t' N_t H_t M_t' C_t^{-1} \kappa_t, \\ \kappa_t &\sim N(0, C_t), \\ N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' N_t L_t + L_t' N_t H_t M_t' C_t^{-1} M_t H_t' N_t L_t, \end{aligned} \tag{1.31}$$

storing the simulated $M_t u_t$ as $\widehat{M_t u_t} = M_t H_t' r_t + \kappa_t$. It will then be convenient to add to $\widehat{M_t u_t}$ the corresponding zero rows so that we simulate from the whole $H_t u_t$ vector (recall H_t is M_t plus some rows of zeros). We will write this as $\widehat{\eta}_t$.

The end condition $\widehat{\eta}_0$ is calculated by

$$\begin{aligned} C_0 &= P_{1|0} - P_{1|0} N_0 P_{1|0}, \quad \kappa_0 \sim N(0, C_0), \\ \widehat{\eta}_0 &= P_{1|0} r_0 + \kappa_0. \end{aligned}$$

The α vector to be simulated via the forward recursion, starting with $\alpha_0 = 0$, is given by

$$\alpha_{t+1} = T_t \alpha_t + \widehat{\eta}_t, \quad t = 0, \ldots, T-1. \tag{1.32}$$

1.8.4 Analytic smoothing

Analytic smoothing is useful in estimating the unobserved log-volatility in the SV models. In particular, the analytic smoother, due to de Jong (1989), computes $a_{t|T} = E\alpha_t|Y_T$ and $P_{t|T} = \text{MSE}(\alpha_t|Y_T)$. It requires that $\alpha_{t+1|T}, P_{t+1|T}, F_t^{-1}, v_t$ and K_t be stored from the Kalman filter. Then, setting $r_T = 0$ and $N_t = 0$, it computes backwards:

$$r_{t-1} = Z_t'F_t^{-1}v_t + L_t'r_t, \quad N_{t-1} = Z_t'F_t^{-1}Z_t + L_t'N_tL_t, t = T, \ldots, 1.$$

Then it records

$$\alpha_{t+1|T} = \alpha_{t|t-1} + P_{t|t-1}r_{t-1} \text{ and } P_{t|T} = P_{t|t-1} - P_{t|t-1}N_{t-1}P_{t|t-1}.$$

1.9 Computing and data sources

There is very little publicly available software to fit ARCH and SV models. The most developed set of programs are those available in TSP and SAS which perform GARCH-M estimation using a normal target. EVIEWS allows some analysis of multivariate GARCH models. The unobserved components time series software STAMP allows a quasi-likelihood analysis of SV models (see Koopman *et al.*, 1995).

All the calculations reported in this paper were performed using my own FORTRAN code compiled using WATCOM 9.5. Throughout I used NAG subroutines to generate random numbers and E04JAF to do standard numerical optimization. I thank Wally Gilks and Peter Rossi for sending me their code to perform random number generation and MCMC analysis of SV models, respectively.

All series reported in this paper are taken from the UK's DATASTREAM. DATASTREAM does not record the exchange rates at weekends (even though there is some very thin trading), and so gives roughly 261 observations a year. The series records the previous day's value if the market is closed during a weekday. Examples of this are Christmas day and bank holidays. These have not been taken out in the analysis presented in this paper.

Acknowledgements

I wish to thank the organizing committee of the Séminaire Européen de Statistique for their invitation to give the lectures on which this paper is based. I am very grateful to Ole Barndorff-Nielsen, David Cox, Eric Ghysels, Andrew Harvey, David Hendry, David Hinkley, Wilfrid Kendall, Lars Korsholm, Aurora Manrique, Mike Pitt, Tina Rydberg, Enrique Sentana and Stephen Taylor for their comments on a previous version of

observed log-volatility
ic smoother, due to
$_{t|T} = \mathrm{MSE}(\alpha_t|Y_T)$. It
tored from the Kalman
tes backwards:

$L_t' N_t L_t, t = T, \ldots, 1.$

$- P_{t|t-1} N_{t-1} P_{t|t-1}.$

to fit ARCH and SV
those available in TSP
using a normal target.
GARCH models. The
TAMP allows a quasi-
et al., 1995).
e performed using my
9.5. Throughout I used
1 E04JAF to do standard
Peter Rossi for sending
on and MCMC analysis

ken from the UK's
ord the exchange rates
n trading), and so gives
ords the previous day's
. Examples of this are
ot been taken out in the

Séminaire Européen de
es on which this paper
ielsen, David Cox, Eric
Hinkley, Wilfrid Kendall,
Tina Rydberg, Enrique
on a previous version of

this paper and to Jim Durbin for allowing me to quote his suggestion of finding the mode in the SV problem. I would like to thank the ESRC for their financial support and Maxine Brant for typing the first draft of this paper.

References

Andersen, T. (1995) Return volatility and trading volume: an information flow interpretation of stochastic volatility. *Journal of Finance*. Forthcoming.

Andersen, T. and B. Sorensen (1995) GMM estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Business and Economic Statistics*. Forthcoming.

Baillie, R.T. and T. Bollerslev (1990) A multivariate generalized ARCH approach to modelling risk premium in forward foreign exchange rate markets. *Journal of International Money and Finance*, **9**, 309–324.

Baillie, R.T. and T. Bollerslev (1992) Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics*, **52**, 91–113.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1995). Fractionally integrated generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*. Forthcoming.

Bartholomew, D.J. (1987) *Latent Variable Models and Factor Analysis*. Oxford University Press, New York.

Bera, A.K. and M.L. Higgins (1995) On ARCH models: properties, estimation and testing. In L. Oxley, D.A.R. George, C.J. Roberts, and S. Sayer (eds), *Surveys in Econometrics*. Oxford: Blackwell. Reprinted from *Journal of Economic Surveys*.

Bergstrom, A.R. (1983) Gaussian estimation of structural parameters in higher order continuous time dynamic models. *Econometrica*, **51**, 117–151.

Berzuini, C., N.G. Best, W.R. Gilks, and C. Larizza (1994) Dynamic graphical models and Markov chain Monte Carlo methods. Unpublished paper: MRC Biostatistics Unit, Cambridge.

Beveridge, S. and C.R. Nelson (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics*, **7**, 151–174.

Black, F. (1976) Studies of stock price volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177–181.

Black, F. and M. Scholes (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637–654.

Bollerslev, T. (1986) Generalised autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **51**, 307–327.

Bollerslev, T. (1987) A conditional heteroscedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, **69**, 542–547.

Bollerslev, T. (1990) Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach. *Review of Economics and Statistics*, **72**, 498–505.

Bollerslev, T., R.Y. Chou, and K.F. Kroner (1992) ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, **52**, 5–59.

Bollerslev, T. and R.F. Engle (1993) Common persistence in conditional variances. *Econometrica*, **61**, 167–186.

Bollerslev, T., R.F. Engle and D.B. Nelson (1995) ARCH models. In R.F. Engle and D. McFadden (eds), *The Handbook of Econometrics, Volume 4*. North-Holland, Amsterdam. Forthcoming.

Bollerslev, T., R.F. Engle and J.M. Wooldridge (1988) A capital asset pricing model with time varying covariances. *Journal of Political Economy*, **96**, 116–131.

Bollerslev, T. and J.M. Wooldridge (1992) Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews*, **11**, 143–172.

Bougerol, P. and N. Picard (1992) Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics*, **52**, 115–128.

Breidt, F.J. and A.L. Carriquiry (1995) Improved quasi-maximum likelihood estimation for stochastic volatility models. Unpublished paper: Department of Statistics, Iowa State University.

Breidt, F.J., N. Crato and P. de Lima (1993) Modelling long-memory stochastic volatility. Unpublished paper: Statistics Department, Iowa State University.

Bresnahan, T.F. (1981) Departures from marginal-cost pricing in the American automobile industry, estimates from 1977–1978. *Journal of Econometrics*, **17**, 201–227.

ity changes. *Proceedings Business and Economic*

of options and corporate , 637–654.

:oregressive conditional :, **51**, 307–327.

edastic time series model *Review of Economics and*

ice in short-run nominal ARCH approach. *Review*

1992) ARCH modeling in rical evidence. *Journal of*

persistence in conditional

(1995) ARCH models. In *andbook of Econometrics,* thcoming.

lge (1988) A capital asset nces. *Journal of Political*

uasi maximum likelihood odels with time varying 3–172.

ity of GARCH processes *urnal of Econometrics*, **52**,

mproved quasi-maximum ility models. Unpublished University.

) Modelling long-memory tatistics Department, Iowa

arginal-cost pricing in the rom 1977–1978. *Journal of*

Brock, W., J. Lakonishok and B. LeBaron (1992) Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, **47**, 1731–1764.

Campbell, J.Y. and L. Hentschel (1992) No news is good news: an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, **31**, 281–318.

Carlin, B.P., N.G. Polson and D. Stoffer (1992) A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *Journal of the American Statistical Association*, **87**, 493–500.

Carter, C.K. and R. Kohn (1993) On the applications of Markov chain Monte Carlo methods to linear state space models. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 131–136.

Carter, C.K. and R. Kohn (1994) On Gibbs sampling for state space models. *Biometrika*, **81**, 541–553.

Chan, K.S. and J. Ledolter (1995) Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, **89**, 242–252.

Chesney, M. and L.O. Scott (1989) Pricing European options: a comparison of the modified Black–Scholes model and a random variance model. *Journal of Financial and Qualitative Analysis*, **24**, 267–284.

Clark, P.K. (1973) A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica*, **41**, 135–156.

Cox, D.R. (1981) Statistical analysis of time series: some recent developments. *Scandanavian Journal of Statistics*, **8**, 93–115.

Cox, J.C. and M. Rubinstein (1985) *Options Markets*. Prentice Hall, Englewood Cliffs, NJ.

Cramér, H. (1946) *Mathematical Methods of Statistics*. Princeton University Press, Princeton, NJ.

Danielsson, J. (1994) Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics*, **61**, 375–400.

Danielsson, J. and J.F. Richard (1993) Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics*, **8**, S153–S174.

Davidian, M. and R.J. Carroll (1987) Variance function estimation. *Journal of the American Statistical Association*, **82**, 1079–1091.

de Jong, P. (1989) Smoothing and interpolation with the state space model. *Journal of the American Statistical Association*, **84**, 1085–1088.

de Jong, P. and N. Shephard (1995) The simulation smoother for time series models. *Biometrika*, **82**, 339–350.

Demos, A. and E. Sentana (1994) Testing for GARCH effects: a one sided approach. Unpublished: CEMFI, Madrid.

Diebold, F.X. and J.A. Lopez (1995) ARCH models. In K. Hoover (ed.), *Macroeconomics: Developments, Tensions and Prospects*.

Diebold, F.X. and M. Nerlove (1989) The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics*, **4**, 1–21.

Ding, Z., C.W.J. Granger, and R.F. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, **1**, 83–106.

Drost, F.C. and T.E. Nijman (1993) Temporal aggregation of GARCH processes. *Econometrica*, **61**, 909–927.

Drost, F.C. and B. Werker (1993) Closing the GARCH gap: continuous time GARCH modelling. Unpublished paper: CentER, Tilburg University.

Duan, J.C. (1995) The GARCH option pricing model. *Mathematical Finance*, **6**.

Duffie, D. and K.J. Singleton (1993) Simulated moments estimation of Markov models of asset process. *Econometrica*, **61**, 929–952.

Dunsmuir, W. (1979) A central limit theorem for parameter estimation in stationary vector time series and its applications to models for a signal observed with noise. *Annals of Statistics*, **7**, 490–506.

Durbin, J. (1992) Personal correspondence to Andrew Harvey and Neil Shephard.

Durbin, J. (1996) *Time Series Analysis Based on State Space Modelling for Gaussian and Non-Gaussian Observations*. Oxford: Oxford University Press. RSS Lecture Notes Series.

Durbin, J. and S.J. Koopman (1992) Filtering, smoothing and estimation for time series models when the observations come from exponential family distributions. Unpublished paper: Department of Statistics, London School of Economics.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Engle, R.F. (1995) *ARCH*. Oxford: Oxford University Press.

Engle, R.F. and T. Bollerslev (1986) Modelling the persistence of conditional variances. *Econometric Reviews*, **5**, 1–50, 81–87.

Engle, R.F. and G. Gonzalez-Rivera (1991) Semiparametric ARCH models. *J. Economics and Business Statist.*, **9**, 345–359.

Engle, R.F., D.F. Hendry and D. Trumble (1985) Small-sample properties of ARCH estimators and tests. *Canadian Journal of Economics*, **18**, 66–93.

Engle, R.F., C.-H. Hong, A. Kane and J. Noh (1993) Arbitrage valuation of variance forecasts with simulated options. *Advances in Futures and Options Research*, **6**, 393–415.

Engle, R.F., A. Kane and J. Noh (1993) Option-index pricing with stochastic volatility and the value of accurate variance forecasts. Unpublished paper: Department of Economics, University of California at San Diego.

Engle, R.F. and K.F. Kroner (1995) Multivariate simultaneous generalized ARCH. *Econometric Theory*, **11**, 122–150.

Engle, R.F. and G.G.J. Lee (1992) A permanent and transitory component model of stock return volatility. Unpublished paper: Department of Economics, University of California at San Diego.

Engle, R.F. and G.G.J. Lee (1994) Estimating diffusion models of stochastic volatility. Unpublished paper: Department of Economics, University of California at San Diego.

Engle, R.F., D.M. Lilien and R.P. Robins (1987) Estimating time-varying risk premium in the term structure: the ARCH-M model. *Econometrica*, **55**, 391–407.

Engle, R.F. and C. Mustafa (1992) Implied ARCH models for options prices. *Journal of Econometrics*, **52**, 289–311.

Engle, R.F. and V.K. Ng (1993) Measuring and testing the impact of news on volatility. *Journal of Finance*, **48**, 1749–1801.

Engle, R.F., V.K. Ng and M. Rothschild (1990) Asset pricing with a factor ARCH covariance structure: empirical estimates for Treasury bills. *Journal of Econometrics*, **45**, 213–238.

Engle, R.F. and J.R. Russell (1994) Forecasting transaction rates: the autogressive conditional duration model. Unpublished paper: Department of Economics, University of California at San Diego.

Evans, M., N. Hastings, and B. Peacock (1993) *Statistical Distributions* (2nd edn) John Wiley and Sons, New York.

Fahrmeir, L. (1992) Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalised linear models. *Journal of the American Statistical Association*, **87**, 501–509.

Fama, E. (1965) The behaviour of stock market prices. *Journal of Business*, **38**, 34–105.

Foster, F.D. and S. Viswanathan (1990) A theory of the interday variations in volume, variance and trading costs in securities markets. *Rev. Financial Studies*, **3**, 593–624.

Fruhwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, 183–202.

Fuller, W. A. (1996) *Introduction to Time Series* (2nd edn). John Wiley, New York. Forthcoming.

Gallant, A.R., D. Hsieh and G. Tauchen (1991) On fitting recalcitrant series: the pound/dollar exchange rate, 1974–83. In W.A. Barnett, J. Powell and G. Tauchen (eds), *Nonparametric and Semiparametric Methods in Economics and Statistics*. Cambridge University Press, Cambridge.

Gallant, A.R., D. Hsieh and G. Tauchen (1994) Estimation of stochastic volatility models with diagnostics. Unpublished paper: Department of Economics, Duke University.

Gallant, A.R. and G. Tauchen (1995) Which moments to match. *Econometric Theory*, **11**. Forthcoming.

Geweke, J. (1986) Modelling the persistence of conditional variances: a comment. *Econometric Reviews*, **5**, 57–61.

Geweke, J. (1989) Exact predictive densities in linear models with ARCH disturbances. *Journal of Econometrics*, **44**, 307–325.

Geweke, J. (1994) Bayesian comparison of econometric models. Unpublished paper: Federal Reserve Bank of Minneapolis.

Gilks, W.R. and P. Wild (1992) Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41**, 337–348.

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**, 1779–1802.

Gourieroux, C. and A. Monfort (1992) Qualitative threshold ARCH models. *Journal of Econometrics*, **52**, 159–200.

Gourieroux, C., A. Monfort and E. Renault (1993) Indirect inference. *Journal of Applied Econometrics*, **8**, S85–S118.

Granger, C. W.J. and R. Joyeux (1980) An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15–39.

Hajivassiliou, V. and D. McFadden (1990) The method of simulated scores, with application to models of external debt crises. Cowles Foundation Discussion Paper, 967.

Hamilton, J. (1994) *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Hannan, E.J. and M. Deistler (1988) *The Statistical Theory of Linear Systems*. John Wiley, New York.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029–54.

Hansen, L.P. and K.J. Singleton (1983) Stochastic consumption, risk aversion and the temporal behavour of asset returns. *Journal of Political Economy*, **91**, 249–265.

Harrison, J. and C.F. Stevens (1976) Bayesian forecasting (with discussion) *Journal of the Royal Statistical Society, Series B*, **38**, 205–247.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Harvey, A.C. (1993a) Long memory and stochastic volatility. Submitted.

Harvey, A.C. (1993b) *Time Series Models* (2nd edn) Philip Allan, New York.

Harvey, A.C. and C.Fernandes (1989) Time series models for count data or qualitative observations. *Journal of Business and Economic Statistics*, **7**, 407–417.

Harvey, A.C., E. Ruiz and E. Sentana (1992) Unobserved component time series models with ARCH disturbances. *Journal of Econometrics*, **52**, 129–158.

Harvey, A.C., E. Ruiz and N. Shephard (1994) Multivariate stochastic variance models. *Review of Economic Studies*, **61**, 247–264.

Harvey, A.C. and N. Shephard (1993a) Estimation and testing of stochastic variance models. STICERD Econometrics Discussion Paper, London School of Economics.

Harvey, A.C. and N. Shephard (1993b) The estimation of an asymmetric stochastic volatility model for asset returns. Unpublished paper: Statistics Department, London School of Economics.

Harvey, A.C. and J.H. Stock (1985) The estimation of higher order continuous time autoregressive models. *Econometric Theory*, **1**, 97–112.

Harvey, A.C. and J.H. Stock (1988) Continuous time autoregressive models with common stochastic trends. *Journal of Economic Dynamics and Control*, **12**, 365–384.

Heston, S. L. (1993) A closed-form solution for options with stochastic volatility, with applications to bond and currency options. *Review of Financial Studies*, **6**, 327–343.

Heynen, R., A. Kemna and T. Vorst (1994) Analysis of the term structure of implied volatility. *Journal of Financial Quantitative Analysis*, **29**, 31–56.

Higgins, M.L. and A.K. Bera (1992) A class of nonlinear ARCH models. *International Economic Review*, **33**, 137–158.

Ho, M.S., W.R.M. Perraudin, and B.E. Sorensen (1993) Multivariate tests of a continuous time equilibrium arbitrage pricing theory with conditional heteroscedasticity and jumps. Unpublished paper: Department of Applied Economics, Cambridge University.

Hong, P.Y. (1991) The autocorrelation structure for the GARCH-M process. *Economic Letters*, **37**, 129–132.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika*, **68**, 165–176.

Hull, J. (1993) *Options, Futures, and Other Derivative Securities* (2nd edn). Prentice Hall International Editions, Englewood Cliffs, NJ.

Hull, J. and A. White (1987) The pricing of options on assets with stochastic volatilities. *Journal of Finance*, **42**, 281–300.

Ingersoll, J. E. (1987) *Theory of Financial Decision Making*. Rowman & Littlefield, Savage, MD.

Jacquier, E., N.G. Polson, and P.E. Rossi (1994) Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics*, **12**, 371–417.

Jacquier, E., N.G. Polson, and P.E. Rossi (1995) Models and prior distributions for multivariate stochastic volatility. Unpublished paper: Graduate School of Business, University of Chicago.

Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASMA, Series D*, **82**, 35–45.

Kim, S. and N. Shephard (1994) Stochastic volatility: Optimal likelihood inference and comparison with ARCH models. Unpublished paper: Nuffield College, Oxford.

King, M., E. Sentana and S. Wadhwani (1994) Volatility and links between national stock markets. *Econometrica*, **62**, 901–933.

Kitagawa, G. and W. Gersch (1984) A smoothness prior — state space modeling of time series with trend and seasonality. *Journal of the American Statistical Association*, **79**, 378–89.

Koenker, R., P. Ng and S. Portnoy (1994) Quantile smoothing splines. *Biometrika*, **81**, 673–80.

Koopman, S.J., A.C. Harvey, J.A. Doornik, and N. Shephard (1995) *STAMP 5.0: Structural Time Series Analyser, Modeller and Predictor*. Chapman & Hall, London.

Kraft, D.F. and R.F. Engle (1982) Autoregressive conditional heteroscedasticity in multiple time series models. Unpublished paper: Department of Economics, University of California at San Diego.

Lee, J.H.H. and M.L. King (1993) A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business and Economic Statistics*, **11**, 17–27.

Lee, S.-W. and B.E. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, **10**, 29–52.

Lin, W.-L. (1992) Alternative estimators for factor GARCH models — a Monte Carlo comparison. *Journal of Applied Econometrics*, **7**, 259–279.

Linton, O. (1993) Adaptive estimation in ARCH models. *Econometric Theory*, **9**, 539–569.

Liu, J., W.H. Wong and A. Kong (1994) Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes. *Biometrika*, **81**, 27–40.

Lumsdaine, R.L. (1991) Asymptotic properties of the quasi-maximum likelihood estimator in the GARCH(1,1) and IGARCH(1,1) models. Unpublished paper: Department of Economics, Princeton University.

Mahieu, R. and P. Schotman (1994) Stochastic volatility and the distribution of exchange rate news. Unpublished paper: Department of Finance, University of Limburg.

Mandelbrot, B. (1963) The variation of certain speculative prices. *Journal of Business*, **36**, 394–419.

Martin, P. (1995) Blunders that bust the bank. *Financial Times, 23 March*, p. 24.

McCullagh, P. and J. A. Nelder (1989) *Generalized Linear Models* (2nd edn). Chapman & Hall, London.

Melino, A. and S.M. Turnbull (1990) Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, **45**, 239–265.

Nelson, D.B. (1990a) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, **6**, 318–334.

Nelson, D.B. (1990b) ARCH models as diffusion approximations. *J. Econometrics*, **45**, 7–38.

Nelson, D.B. (1991) Conditional heteroscedasticity in asset pricing: a new approach. *Econometrica*, **59**, 347–370.

Nelson, D.B. (1992) Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model. *Journal of Econometrics*, **52**, 61–90.

Nelson, D.B. (1994) Asymptotically optimal smoothing with ARCH models. *Econometrica*, **63**. Forthcoming.

Nelson, D.B. and D.P. Foster (1994) Asymptotic filtering theory for univariate ARCH models. *Econometrica*, **62**, 1–41.

Nijman, T.E. and E. Sentana (1993) Marginalization and contemporaneous aggregation in multivariate GARCH processes. Discussion paper, CentER, Tilburg University.

Noh, J., R.F. Engle, and A. Kane (1994) Forecasting volatility and option pricing of the S&P 500 index. *Journal of Derivatives*, 17–30.

Pagan, A.R. and G.W. Schwert (1990) Alternative models for conditional stock volatility. *Journal of Econometrics*, **45**, 267–290.

Phillips, P. C.B. and S. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies*, **53**, 473–495.

Pitt, M. and N. Shephard (1995) Parameter-driven exponential family models. Unpublished paper: Nuffield College, Oxford.

Poon, S. and S.J. Taylor (1992) Stock returns and volatility: an empirical study of the UK stock market. *Journal of Banking and Finance*, **16**, 37–59.

Poterba, J. and L. Summers (1986) The persistence of volatility and stock market fluctuations. *American Economic Review*, **76**, 1124–1141.

Qian, W. and D.M. Titterington (1991) Estimation of parameters in hidden Markov Chain models. *Philosophical Transactions of the Royal Society of London, Series A*, **337**, 407–428.

Quintana, J.M. and M. West (1987) An analysis of international exchange rates using multivariate DLM's. *The Statistician*, **36**, 275–281.

Ripley, B.D. (1987) *Stochastic Simulation*. Wiley, New York.

Ross, S.A. (1976) The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13**, 341–360.

Rothenberg, T.J. (1973) *Efficient Estimation with A Priori Information*. Yale University Press, New Haven, CT.

Ruud, P. (1991) Extensions of estimation methods using the EM algorithm. *Journal of Econometrics*, **49**, 305–341.

Schwert, G. W. (1989) Why does stock market volatility change over time? *Journal of Finance*, **44**, 1115–1153.

Scott, L. (1987) Options pricing when the variance changes randomly: theory, estimation and an application. *Journal of Financial and Quantitative Analysis*, **22**, 419–438.

Scott, L. (1991) Random-variance option pricing. *Advances in Future and Options Research*, **5**, 113–135.

Sentana, E. (1991) Quadratic ARCH models: a potential re-intepretation of ARCH models. Unpublished paper: CEMFI, Madrid.

Shephard, N. (1993) Fitting non-linear time series models, with applications to stochastic variance models. *Journal of Applied Econometrics*, **8**, S135–S152.

Shephard, N. (1994a) Local scale model: state space alternative to integrated GARCH processes. *Journal of Econometrics*, **60**, 181–202.

Shephard, N. (1994b) Partial non-Gaussian state space. *Biometrika*, **81**, 115–131.

Smith, A.A. (1993) Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics*, **8**, S63–S84.

Smith, A. F.M. and G. Roberts (1993) Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, **55**, 3–23.

Smith, J. Q. (1985) Diagnostic checks of non-standard time series models. *Journal of Forecasting*, **4**, 283–291.

Smith, R.L. and J.E. Miller (1986) A non-Gaussian state space model and application to prediction records. *Journal of the Royal Statistical Society, Series B*, **48**, 79–88.

Steigerwald, D. (1991) Efficient estimation of models with conditional heteroscedasticity. Unpublished paper: Department of Economics, University of California at Santa Barbara.

Stein, E.M. and J. Stein (1991) Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies*, **4**, 727–752.

Tauchen, G. and M. Pitts (1983) The price variability volume relationship on speculative markets. *Econometrica*, **51**, 485–505.

Taylor, S.J. (1986) *Modelling Financial Time Series*. John Wiley, Chichester.

Taylor, S.J. (1994) Modelling stochastic volatility. *Mathematical Finance*, **4**, 183–204.

Uhlig, H. (1992) Bayesian vector autoregressions with time varying error covariances. Unpublished paper: Department of Economics, Princeton University.

Wei, G. C.G. and M.A. Tanner (1990) A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**, 699–704.

Weiss, A.A. (1986) Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory*, **2**, 107–131.

West, M. and J. Harrison (1989) *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York.

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.

White, H. (1994) *Estimation, Inference and Specification Analysis*. Cambridge University Press, Cambridge.

Whittle, P. (1991) Likelihood and cost as path integrals. *Journal of the Royal Statistical Society, Series B*, **53**, 505–538.

Wiggins, J.B. (1987) Option values under stochastic volatilities. *Journal of Financial Economics*, **19**, 351–372.

Wild, P. and W.R. Gilks (1993) AS 287: Adaptive rejection sampling from log-concave density functions. *Applied Statistics*, **42**, 701–709.

Xu, X. and S.J. Taylor (1994) The term structure of volatility implied by foreign exchange options. *Journal of Financial and Quantitative Analysis*, **29**, 57–74.

Zakoian, J.-M. (1990) Threshold heteroscedastic models. Unpublished paper: CREST, INSEE.

Zeger, S.L. and B.
series, a quasi lik

Zeger, S.L. and B. Qaqish (1988) Markov regression models for time series, a quasi likelihood approach. *Biometrics*, **44**, 1019–1032.