WORKING PAPER n.02.10

December 2002

Stochastic Volatility Models: A Survey with Applications to Option Pricing and Value at Risk

M. Billio[a,b]
D. Sartore[a,b]

a. Università Ca' Foscari, Venice.
b. GRETA, Venice.

Stochastic Volatility Models: A Survey with Applications to Option Pricing and Value at Risk

by

Monica BILLIO* **and Domenico SARTORE****

(University Ca' Foscari of Venice and GRETA**)***

Abstract

This chapter presents an introduction to the current literature on stochastic volatility models. For these models the volatility depends on some unobserved components or a latent structure.

Given the time-varying volatility exhibited by most financial data, in the last two decades there has been a growing interest in time series models of changing variance and the literature on stochastic volatility models has expanded greatly. Clearly, this chapter cannot be exhaustive, however we discuss some of the most important ideas, focusing on the simplest forms of the techniques and models used in the literature.

The chapter is organised as follows. Section 1 considers some motivations for stochastic volatility models: empirical stylised facts, pricing of contingent assets and risk evaluation. While section 2 presents models of changing volatility, section 3 focuses on stochastic volatility models and distinguishes between models with continuous and discrete volatility, the latter depending on a hidden Markov chain. Section 4 is devoted to the estimation problem which is still an open question, then a wide range of possibility is given. Sections 5 and 6 introduce some extensions and multivariate models. Finally, in section 7 an estimation program is presented and some possible applications to option pricing and risk evaluation are discussed.

Readers interested in the practical utilisation of stochastic volatility models and in the applications can skip section 4.3 without hindering comprehension.

Keywords: Heteroskedasticity, Changing volatility models, Continuous and discrete volatility, Non-Gaussian parameter-driven models, Kalman filter, Hamilton filter, Quasi-maximum likelihood approach, Option pricing, Value at Risk.

** Monica Billio is Associate Professor of Econometrics at University Ca' Foscari of Venice (E-mail: billio@unive.it).*
*** Domenico Sartore is Full Professor of Econometrics at University Ca' Foscari of Venice and President of GRETA (E-mail: sartore@unive.it).*
**** GRETA – Gruppi di Ricerca Economica Teorica ed Applicata, San Marco 3870, 30124 Venezia, Italy.*
We gratefully acknowledge help from Michele Gobbo, Massimiliano Caporin and Andrea Giacomelli.

1. INTRODUCTION

In the last two decades there has been a growing interest in time series models of changing variance, given the time-varying volatility exhibited by most financial data. In fact, the empirical distributions of financial time series differ substantially from distributions obtained from sampling independent homoskedastic Gaussian variables. Unconditional density functions exhibit leptokurtosis and skewness; time series of financial returns show evidence of volatility clustering; and squared returns exhibit pronounced serial correlation whereas little or no serial dependence can be detected in the return process itself.

These empirical regularities suggest that the behaviour of financial time series may be captured by a model which recognizes the time-varying nature of return volatility, as follows:

$$y_t = \mu_t + \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \text{IID}(0,1), \qquad t = 1,2,...,T$$

where y_t denotes the return on an asset. A common way of modelling σ_t is to express it as a deterministic function of the squares of lagged residuals. Econometric specifications of this form are known as ARCH models and have achieved widespread popularity in applied empirical research (see Bollerslev, Chow and Kroner (1992), Bollerslev, Engle and Nelson (1993), Bera and Higgins (1993)).

Alternatively, volatility may be modelled as an unobserved component following some latent stochastic process, such as an autoregression. The resulting models are called stochastic volatility (SV) models and have been the focus of considerable attention in the recent years (Taylor (1994), Ghysels, Harvey and Renault (1996) and Shephard (1996)). These models present two main advantages over ARCH models. The first one is their solid theoretical background, as they can be interpreted as discretised versions of stochastic volatility continuous-time models put forward by modern finance theory (see Hull and White (1987)). The second is their ability to generalise from univariate to multivariate series in a more natural way, as far as their estimation and interpretation are concerned. On the other hand, SV models are more difficult to estimate than ARCH models, due to the fact that it is not easy to derive their exact likelihood function. For this reason, a number of econometric methods have been proposed to solve the problem of estimation of SV models.

The literature on SV models has expanded greatly in the last ten years, reaching considerable proportions; this chapter cannot therefore be exhaustive. We prefer 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 generalisations and technicalities. In the organisation of the structure of the present chapter, we have been inspired by the paper of Shephard (1996), who gave a very interesting survey on SV models updated to 1995.

To start, we will consider some motivations for stochastic volatility models: empirical stylised facts, pricing of contingent assets and risk evaluation.

1.1. Empirical stylised facts

To illustrate the models and to develop the examples we will work with three European stock indexes: the FTSE100, the CAC40 and the MIB30), which are market indexes for the London, Paris and Milan equity markets. These series run from 4th January 1999 to 12th August 2002, yielding 899 daily observations.

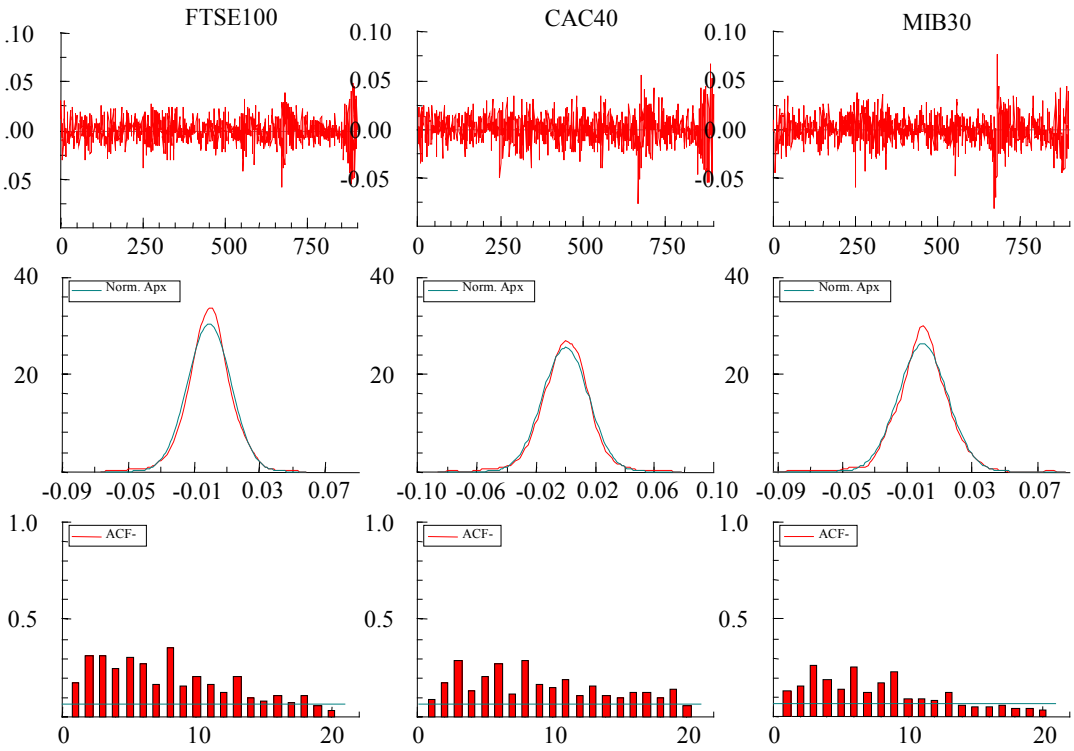

Figure 1: Summaries of the daily returns on three European stock indexes: the FTSE100, the CAC40 and the MIB30. Summaries are: time series of returns, nonparametric density estimate and normal approximation, correlogram of squared returns.

Throughout we will work with the compounded return[1] on the series

$$y_t = \log(x_t / x_{t-1})$$

where x_t is the value of stock index. Figure 1 displays some summaries of these three series. The raw time series of y_t 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, also reported in Figure 1, which shows significant correlations existing at quite extended lag lengths. This suggests that y_t^2 may follow a process close to an ARMA(1,1), for a simple AR

[1] An advantage of using a return series is that it helps in making the time series stationary, a useful statistical property (see footnote 4).

process cannot easily combine the persistence in shocks with the low correlation. A correlogram of y_t shows little activity and so is not given in this figure.

Figure 1 also gives a density estimate of the unconditional distribution of y_t together with the corresponding normal approximation[2]. This suggests that y_t is leptokurtic. This is confirmed by Table 1, which reports an estimate of the excess of kurtosis with respect to the normal distribution, which are significantly positive.

	FTSE100	**CAC40**	**MIB30**
Mean	-0.03634	-0.000219	-0.04313
Standard deviation	0.001463	0.015630	0.001689
Asymmetry	-0.2574	-0.223998	-0.23383
Excess of kurtosis	1.547015	4.624507	2.268374
Jarque-Bera test	97 (0.00)	106 (0.00)	197 (0.00)

Table 1: Summary statistics for the daily returns in Figure 1. In parentheses the p-value of the Jarque-Bera test.

Table 1 also reports the Jarque-Bera test for normality and the asymmetry coefficients evidencing that the distributions are negatively skewed, partially due to the period of analysis, and for all three the null hypothesis of normality is clearly rejected.

These stylized facts can be summarised as follows: non significant serial correlation in the levels of returns; volatility clustering, which implies a significant and positive serial correlation in the squares y_t^2; heavy tails and persistence of volatility.

Finally, there is some evidence that stock markets share periods of high volatility. This suggests that multivariate models will be important.

1.2. Pricing contingent assets

Consider an asset *C*, with expiring date $t+\tau$, which is a function of a generic underlying security *S*. Assume now that *S* can be described by the following geometric diffusion process:

$$dS = \mu S\, dt + \sigma S\, dz,$$

so that $d\ log\ S = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma\, dz$.

Economists term such an asset *C* as "contingent" or "derivative". A primary example of a derivative is an option, which entitles the owner the ability but not the obligation to trade the underlying security at a given price *K*, called strike price, in the future.

[2] The graphs are produced with the Ox software using some of its basic commands and default options. See also section 7.1.

European call options are the most known: the owner can buy the underlying asset at the strike price K only when the call expires, i.e. at date $t+\tau$. Its value at the expiration date will be:

$$C_{t+\tau} = \max\left(S_{t+\tau} - K, 0\right).$$

Its purchase value at time t is as yet unknown, but can be determined in different ways. One of these consists in calculating the discounted expected value of the option at time $t+\tau$:

$$\exp(-r\tau) E_{S_{t+\tau}|S_t}\left[\max\left(S_{t+\tau} - K, 0\right)\right],$$

where r is the free-risk interest rate. However, this completely ignores the fact this is a risky assets and traders expect higher returns than on riskless assets. This is the reason why the discounted expected value is not considered by the market as a correct method to evaluate an asset. To avoid this inconvenience it is opportune to introduce a utility function into the pricing of options, letting the dealers choose the risk-expected gain combination they prefer.

It turns out that the added complexity of a utility function can be avoided by assuming continuous and costless trading. This statement can be shown by creating a particular portfolio, which by construction is made up of owning θ of the underlying shares and by borrowing a single contingent asset C. If the investor properly selects θ at each time, the stochastic component of the process disappears and ensures the portfolio a riskless dynamic making its return a deterministic function of time (see Black and Scholes (1973)). As time passes, the portfolio will have to be continually adjusted to maintain risklessness, hence the need for continuous costless trading.

The return of this portfolio must be equal to the riskless interest rate r because the portfolio itself is risk-free, otherwise traders will have an arbitrage opportunity. This condition is necessary to obtain the stochastic differential equation followed by the contingent asset:

$$\frac{\partial C}{\partial t} + \frac{1}{2}\frac{\partial^2 C}{\partial S \partial S}\sigma^2 S^2 + r\frac{\partial C}{\partial S}S = rC, \quad \text{with end condition } C = \max\left(S - K, 0\right)$$

This equation is quite easy to solve and does not depend on the mean parameter μ nor on the risk preferences of the traders. Whatever the risk preferences may be, the evaluation of the option does not change. When solving the equation, risk neutral preferences are used to simplify calculations. With instantaneous variance σ^2, the following Black-Scholes valuation formula is obtained:

$$C_t^{BS}\left(\sigma^2\right) = S_t\Phi\left(d\right) - Ke^{-r\tau}\Phi\left(d - \sigma\sqrt{\tau}\right), \text{ where } d = \frac{\log\left(S_t/K\right) + \left(r + \sigma^2/2\right)\tau}{\sigma\sqrt{\tau}}$$

Note that σ^2 is the only unknown parameter: S_t and r are observed, while τ and K are usually given by institutional norms. The price depends strictly on σ^2 which is more important than the drift, as is often the case in finance, so that the price of the option can be considered an indicator of the volatility of the underlying asset.

Empirically, the Black-Scholes formula can be used in two ways: either by estimating σ^2 (the historical volatility) and then calculating the option price or by using real prices to determine a value for σ^2 (called implied volatility).

This type of analysis has a considerable shortcoming: the basic assumption that stock returns follow a geometric diffusion process is a poor one, as indicated in Figure 1, and can affect the valuation formula reducing the precision of the option pricing. This realization has prompted theoretical work into option pricing theory under various changing volatility regimes. The leading paper in this field is Hull and White (1987), to which we will return later.

1.3. Risk evaluation

VaR (Value at Risk) is the maximum amount that is expected to be lost over some target period, i.e. the maximum likely loss. It is a statistical risk measure and represents a percentile of the probability distribution of the variable of interest.

Generally speaking, VaR can be analytically defined as follows. Let x_t be a random variable of interest measure and $F_x(x_t)$ its cumulative distribution function,

$$a = \text{Prob}\left(x_t \le \bar{x}_t\right) = F_x\left(\bar{x}_t\right) = \int_{-\infty}^{\bar{x}_t} f\left(x_t\right) dx_t$$

VaR is the percentile defined by the relation:

$$VaR_x\left(1-a\right) = x_t^* = F_{x_t}^{-1}\left(a\right)$$

where (1-*a*) is the VaR confidence level, for instance 95% or 99%, and $F_{x_t}^{-1}\left(a\right)$ is the inverse of the cumulative distribution function.

Given its generality, the VaR method can be applied for different types of risk measurement, such as market risk, credit risk, operational risk and commodity risk (see Alexander (1996)). Moreover, for its versatility, VaR allows us to obtain an intuitive risk measure, to define homogeneous risk measures that permit a comparison among different financial instruments, to determine limiting positions and to construct risk-adjusted profitability measures.

Let us concentrate on its application to market risk. Market risk means the possibility that an unexpected variation of market factors (interest rates, exchange rates, stock

prices, etc.) causes an increase or a reduction in the value of a position or in the value of a financial portfolio. VaR, in this context, is the maximum expected loss of a marketable financial instruments portfolio which could be experienced, for a specified time horizon period and a specified confidence level.

We now consider a general portfolio model which allows us to set all the hypothesis discriminating a risk measurement model like VaR in a systematic manner, by paying particular attention to the role of the volatility.

Let x_τ be a random variable which represents the value of a portfolio in a future period τ. It is defined by the following relation:

$$x_\tau = \sum_{i=1}^{N} w_{i,t} P_{i,\tau}$$

where the random variables $P_{i,\tau}$ represent the future value of the N assets in the portfolio. If we suppose that the N assets will be subjected to K risk market factors $\chi_{j,\tau}$, the future value of the portfolio can be expressed as a function of the K stochastic risk factors by the following pricing formula:

$$x_\tau = \sum_{i=1}^{N} w_{i,t} P_{i,\tau}\left(\chi_{1,\tau}, \ldots, \chi_{K,\tau}\right).$$

The hypothesis characterizing the model therefore concerns: the endogenous variable choice; the pricing formula; the risk factors definition and their distributions; the risk factors volatility; the risk factors mapping; the confidence level and the choice of the time horizon.

In the literature, the following approaches are suggested to estimate the VaR (Best (1998)): parametric methods; historical simulation; Monte Carlo simulation; stress testing. Concerning the parametric methods and the Monte Carlo simulation, it is crucial to properly describe the volatility dynamics of the risk factors to obtain correct estimates of the VaR (see for example Lehar, Scheicher and Schittenkopf (2002)).

2. MODELS OF CHANGING VOLATILITY

Following Cox (1981) and Shephard (1996) models of changing volatility can be usefully partitioned into observation-driven and the parameter-driven models. They both can be generally expressed using the following parametric framework:

$$y_t | z_t \sim N\left(\mu_t, \sigma_t^2\right)$$

where μ_t is often set equal to zero (as we do not intend to focus on that feature of the model).

In the first class, i.e. in observation-driven models, z_t is a function of lagged values of y_t. The autoregressive conditional heteroskedasticity (ARCH) models introduced by Engle (1982) are the most representative example of observation-driven models. They describe the variance as a linear function of the squares of past observations

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

and so the model is defined by the conditional density (one-step-ahead forecast density)

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

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 and then one type of shock alone drives both the series itself and its volatility.

The use of models described by their one-step-ahead forecast offers remarkable advantages that are worth being highlighted. First, the likelihood expression can be simply obtained by combining these densities, making the estimation and testing easy to handle, at least in principle. Second, conditional densities imply the use of conditional moments which are used widely to specify finance theory, although this one is conditional to economic agents', if not the econometricians', information. Finally, the observation-driven models parallel the autoregressive and moving average ones which are commonly used for models of changing means.

In the second class, i.e. in parameter-driven (or parameter dynamic latent variable or state space) models, z_t is a function of an unobserved or latent component. The log-normal stochastic volatility model created by Taylor (1986) is the simplest and best-known example:

$$y_t | h_t \sim N\left(0, \exp\left(h_t\right)\right), \qquad h_t = \alpha + \beta h_{t-1} + \eta_t, \qquad \eta_t \sim NID\left(0, \sigma_\eta^2\right) \qquad (1)$$

where h_t represents the log-volatility, which is unobserved but can be estimated using the observations. With respect to the previous class, these models are driven by two types of shock, one of which influences the volatility (i.e. conditional variance equations). These models parallel the Gaussian state space models of means dealt with by Kalman (1960).

In spite of this, a shortcoming of parameter-driven volatility models is that they generally lack analytic one-step-ahead forecast densities $y_t | Y_{t-1}$, unlike the models of the mean which fit into the Gaussian state space form. Hence either an approximation or a numerically intensive method is required to deal with these models.

Although SV models are harder to handle statistically than the corresponding observation-driven models, there are still some good reasons for 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 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 with an Ornstein-Uhlenbeck process.

3. STOCHASTIC VOLATILITY MODELS

For these models the volatility depends on some unobserved components or a latent structure. 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 (Clark (1973)). The most popular of these parameter-driven stochastic volatility models, from Taylor (1986), puts

$$\begin{cases} y_t = \varepsilon_t \exp(h_t/2), \\ h_t = \alpha + \beta h_{t-1} + \eta_t \end{cases} \tag{2}$$

where ε_t and η_t are two independent Gaussian white noises, with variances 1 and σ_η^2, respectively. Due to the Gaussianity of η_t, this model is called a log-normal SV model.

Another possible interpretation for h_t is to characterise the regime in which financial markets are operating and then it could be described by a discrete valued variable. The most popular approach to modelling changes in regime is the class of Markov switching models introduced by Hamilton (1989) in the econometrics literature. In that case the simplest model is[3]:

$$\begin{cases} y_t = \varepsilon_t \exp(h_t/2), \\ h_t = \alpha + \beta s_t \end{cases} \tag{3}$$

where s_t is a two-state first-order Markov chain which can take values 0,1 and is independent of ε_t. The value of the time series s_t, for all *t*, depends only on the last value s_{t-1}, i.e. for *i,j*=0,1

$$P(s_t = j \mid s_{t\text{-}1} = i,\ s_{t\text{-}2} = i,\ \ldots) = P(s_t = j \mid s_{t\text{-}1} = i) = p_{ij}$$

The probabilities $(p_{ij})_{i,j=0,1}$ are called transition probabilities of moving from one state to the other. Obviously, we get that:

[3] The representation $y_t = \sigma_{s_t} \varepsilon_t$, with $\sigma_0 = \exp(\alpha/2)$ and $\sigma_1 = \exp((\alpha+\beta)/2)$, is clearly equivalent. To identify the regime 1 as the high volatility regime, we set $\beta > 0$.

$$p_{00} + p_{01} = p_{10} + p_{11} = 1$$

and these transition probabilities are collected in the transition matrix P

$$\mathrm{P} = \begin{bmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{bmatrix}$$

which fully describes the Markov chain.

A two state Markov chain can easily be represented by a simple AR(1) process as follows:

$$s_t = (1 - p_{00}) + (-1 + p_{00} + p_{11}) s_{t-1} + v_t \tag{4}$$

where $v_t = s_t - E(s_t | s_{t-1}, s_{t-2}, \ldots)$. Although v_t can take only a finite set of values, on average v_t is zero. The innovation v_t is thus a martingale difference sequence. Given the autoregressive representation of the Markov chain, it is possible to rewrite the volatility equation of model (3) in the following way:

$$\begin{aligned} h_t &= \alpha + \beta s_t \\ &= \alpha + \beta[(1 - p_{00}) + (-1 + p_{00} + p_{11}) s_{t-1} + v_t] \\ &= \alpha(2 - p_{00} - p_{11}) + \beta(1 - p_{00}) + (-1 + p_{00} + p_{11}) h_{t-1} + \beta v_t \end{aligned}$$

The SV model with discrete volatility has therefore the same structure of the model (2) but with a noise that can take only a finite set of values:

$$\begin{cases} y_t = \varepsilon_t \exp(h_t / 2), \\ h_t = a + b\, h_{t-1} + \omega_t \end{cases} \tag{5}$$

Let us describe the basic properties of both type of models in the following sections.

3.1. SV models with continuous volatility

We consider ε_t and η_t independent, Gaussian white noises. The properties of model (2) are discussed in Taylor (1986) and Taylor (1994) (see also Shephard (1996)). Broadly speaking, given the product process nature of the model, these properties are easy to derive, but estimation is substantially harder than for the corresponding ARCH models.

As η_t is Gaussian, h_t is a standard Gaussian autoregression. It will be stationary (covariance[4] and strictly[5]) if $|\beta| < 1$ with:

[4] A stochastic process

$$\mu_h = E(h_t) = \frac{\alpha}{1-\beta},$$

$$\sigma_h^2 = Var(h_t) = \frac{\sigma_\eta^2}{1-\beta^2}.$$

As ε_t is always stationary, y_t will be stationary if and only if h_t is stationary, y_t being the product of two stationary processes. Using the properties of the log-normal distribution, all the moments exist if h_t is stationary and in particular the kurtosis is:

$$\frac{E(y_t^4)}{(E(y_t^2))^2} = 3\exp(\sigma_h^2) \geq 3$$

which shows that the SV model has fatter tails than the corresponding normal distribution and all the odd moments are zero.

The dynamic properties of y_t are easy to find. First, as ε_t is iid, y_t is a martingale difference[6] and is a white noise[7] if $|\beta| < 1$. As h_t is a Gaussian AR(1),

$$\begin{aligned} Cov(y_t^2, y_{t-r}^2) &= E(y_t^2 y_{t-r}^2) - (E(y_t^2))^2 \\ &= E(\exp(h_t + h_{t-r})) - (E(\exp(h_t)))^2 \\ &= \exp(2\mu_h + \sigma_h^2)(\exp(\sigma_h^2 \beta^r) - 1) \end{aligned}$$

and so

$$\rho_{y_t^2}(r) = \frac{Cov(y_t^2, y_{t-r}^2)}{Var(y_t^2)} = \frac{\exp(\sigma_h^2 \beta^r) - 1}{3\exp(\sigma_h^2) - 1} \cong \frac{\exp(\sigma_h^2) - 1}{3\exp(\sigma_h^2) - 1} \beta^r$$

---

y_t is covariance stationary if the degree of covariance amongst its observations depends only on the time gap between them, i.e. $Cov(y_t, y_{t+r}) = \gamma(r)$ for all *t*.

[5] For some processes there will exist no moments, even in cases where the corresponding unconditional distributions are perfectly well-behaved. The strict stationarity of y_t is then defined as follows: $F(y_{t+r}, y_{t+r+1}, ..., y_{t+r+p}) = F(y_t, y_{t+1}, ..., y_{t+p})$ for all p and r.

[6] y_t being a martingale difference stipulates that $E|y_t| < \infty$ and that $E(y_t | y_{t-1}, y_{t-2}, ...) = 0$. All martingale differences have zero means and are uncorrelated over time. If the unconditional variance of the martingale difference is constant over time, then the series is also a white noise.

[7] This means $E(y_t) = \mu, Var(y_t) = \sigma^2$ and $Cov(y_t, y_{t+r}) = 0$ for all $r \neq 0$. 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 white noise, a special case of which is independent and identically distributed (iid).

Hence, the memory of the y_t is defined by the memory of the latent h_t, in this case an AR(1). Moreover, note that if $\beta < 0$, $\rho_{y_t^2}(r)$ 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. Finally, note that there is no need for non-negativity constraints nor for bounded kurtosis constraints on the coefficients. This is a great advantage with respect to GARCH models.

Insights on the dynamic properties of the SV model can also be obtained by squaring and taking logs, getting

$$\begin{cases} \log\left(y_t^2\right) = h_t + \log\left(\varepsilon_t^2\right), \\ h_t = \alpha + \beta h_{t-1} + \eta_t \end{cases} \tag{6}$$

a linear process, which adds the iid $\log\left(\varepsilon_t^2\right)$ to the AR(1) h_t. As a result $\log\left(y_t^2\right)$~ARMA(1,1). If ε_t is Gaussian, then $\log\left(\varepsilon_t^2\right)$ has a mean of -1.27 and variance 4.93, but its distribution is far from being normal, as it is heavily skewed with a long left-hand tail, caused by taking the logs of very small numbers, an operation which generates outliers. The autocorrelation function for $\log\left(y_t^2\right)$ is

$$\rho_{\log\left(y_t^2\right)}(r) = \frac{\beta^r}{1 + \frac{4.93}{\sigma_h^2}}$$

3.2. SV models with discrete volatility

We consider a two-state Markov chain s_t independent of ε_t, which is Gaussian white noise.

Assuming stationarity[8], the unconditional probabilities to be in the regime 0 ($P(s_t = 0) = \pi_0$) or 1 ($P(s_t = 1) = \pi_1$) are defined as follows:

$$\begin{cases} \pi_0 = p_{00}\pi_0 + (1 - p_{11})\pi_1 \\ \pi_1 = (1 - p_{00})\pi_0 + p_{11}\pi_1 \end{cases}$$

with $\pi_0 + \pi_1 = 1$, or in a vector form:

$$\begin{cases} \pi = \mathrm{P}\pi \\ \mathbf{1}'\pi = 1 \end{cases}$$

[8] An ergodic Markov chain is a covariance stationary process. For some basic properties of Markov chains, see Hamilton (1994, chapter 22).

where $\mathbf{1} = (1,1)'$. Thus, they are:

$$\pi_0 = \frac{1 - p_{11}}{2 - p_{00} - p_{11}},$$
$$\pi_1 = \frac{1 - p_{00}}{2 - p_{00} - p_{11}}$$

From the definition of b in equation (5), we can note that when $p_{00} + p_{11} > 1$ the h_t process is likely to persist in its current state and it would be positively serially correlated. Its unconditional moments are:

$$\begin{aligned} E(h_t) &= \alpha + \beta E(s_t) \\ &= \alpha + \beta \pi_1 \\ Var(h_t) &= \beta^2 \pi_1 (1 - \pi_1). \end{aligned}$$

Under stationarity[9], as for the SV model with continuous volatility, all the moments exist, all the odd moments are zero and the kurtosis is:

$$\frac{E(y_t^4)}{\left(E(y_t^2)\right)^2} = 3\frac{(\pi_0 + \exp(2\beta)\pi_1)}{(\pi_0 + \exp(\beta)\pi_1)^2} \geq 3$$

Moreover, as ε_t is iid, y_t is a martingale difference and its dynamic properties are described by the covariances of squares:

$$\begin{aligned} Cov(y_t^2, y_{t-r}^2) &= E(y_t^2 y_{t-r}^2) - \left(E(y_t^2)\right)^2 \\ &= E(\exp(h_t + h_{t-r})) - (\exp(\alpha)\pi_0 + \exp(\alpha + \beta)\pi_1)^2 \\ &= \exp(2\alpha)P(s_t = 0, s_{t-r} = 0) + \exp(2\alpha + \beta)P(s_t = 0, s_{t-r} = 1) \\ &\quad + \exp(2\alpha + \beta)P(s_t = 1, s_{t-r} = 0) + \exp(2\alpha + 2\beta)P(s_t = 1, s_{t-r} = 1) \\ &\quad - (\exp(\alpha)\pi_0 + \exp(\alpha + \beta)\pi_1)^2 \end{aligned}$$

where, the vector of unconditional joint probabilities $P(s_t, s_{t-r})$ can be computed as follows:

$$\begin{aligned} P(s_t, s_{t-r}) &= P(s_t | s_{t-r}) P(s_{t-r}) \\ &= \mathrm{P}^r \pi \end{aligned}$$

with

[9] For this see Francq and Zakoïan (2001) and Francq, Roussignol and Zakoïan (2001).

$$\mathrm{P}^{\mathrm{r}} = \begin{bmatrix} \dfrac{(1-p_{11})+\lambda^r(1-p_{00})}{2-p_{00}-p_{11}} & \dfrac{(1-p_{11})+\lambda^r(1-p_{11})}{2-p_{00}-p_{11}} \\ \dfrac{(1-p_{00})+\lambda^r(1-p_{00})}{2-p_{00}-p_{11}} & \dfrac{(1-p_{00})+\lambda^r(1-p_{11})}{2-p_{00}-p_{11}} \end{bmatrix}$$

and $\lambda = -1 + p_{00} + p_{11}$.

Finally, it is useful to note that h_t is itself a Markov chain which can take the values α and $\alpha + \beta$ with the same transition matrix P.

4. ESTIMATION

The difficulties in estimating SV models lie in the latent nature of the volatility. Inference may be difficult, because the distribution of $y_t | Y_{t-1}$ is specified implicitly rather than explicitly and the likelihood function appears as a multivariate integral the size of which is equal to the number of observations multiplied by the size of the latent variables, which is 1 for the described models.

Like most non-Gaussian parameter-driven models, there are many different ways of performing estimation: some involve estimating or approximating the likelihood, others use the method of moments procedures (see Ghysels, Harvey and Renault (1996) and Shephard (1996)).

Let us first of all clearly state the problem of computing the likelihood function for the general class of parametric dynamic latent variable or non-linear and/or non-Gaussian state-space models.

4.1. A general filter for non-Gaussian parameter-driven models

Both SV models (with continuous and discrete volatility) fit in the following framework:

$$\begin{cases} y_t = \phi_t(h_t, \varepsilon_t; \theta) & \textit{measurement equation} \\ h_t = \varphi_t(h_{t-1}, \eta_t; \theta) & \textit{transition equation} \end{cases} \qquad (7)$$

where ε_t and η_t are independent white noises, with marginal distributions which may depend on θ, the vector of parameters. Let H^t and Y^t denote $(h_1, h_2, ..., h_t)'$ and $(y_1, y_2, ..., y_t)'$, respectively.

There are serious difficulties in computing the likelihood function; in fact, with *T* the number of observations, we have:

$$f\left(Y^T, H^T; \theta\right) = \prod_{t=1}^{T} f\left(y_t \middle| Y^{t-1}, H^t; \theta\right) f\left(h_t \middle| Y^{t-1}, H^{t-1}; \theta\right)$$

and the likelihood function is:

$$\ell_T(\theta) \equiv f\left(Y^T; \theta\right) = \int \prod_{t=1}^{T} f\left(y_t \middle| Y^{t-1}, H^t; \theta\right) f\left(h_t \middle| Y^{t-1}, H^{t-1}; \theta\right) \prod_{t=1}^{T} dh_t \qquad (8)$$

which is an integral whose size is equal to the number of observations multiplied by the dimension of the unobserved variable h_t, and thus it is practically unfeasible.

It is however possible to derive a general algorithm which allows the formal computation of the likelihood function by decomposing the calculation of integral (8) into a sequence of integrals of lower dimension.

Let $f\left(h_{t-1} \middle| Y^{t-1}\right)$ be the input of the iteration[10] t=1,2,…,T. First of all, we can decompose the joint conditional density of h_t, h_{t-1} into the product of the transition density by the input density:

1. $$f\left(h_t, h_{t-1} \middle| Y^{t-1}\right) = f\left(h_t \middle| h_{t-1}\right) f\left(h_{t-1} \middle| Y^{t-1}\right)$$

By marginalisation we obtain the prediction density of h_t

2. $$f\left(h_t \middle| Y^{t-1}\right) = \int f\left(h_t, h_{t-1} \middle| Y^{t-1}\right) dh_{t-1} = \int f\left(h_t \middle| h_{t-1}\right) f\left(h_{t-1} \middle| Y^{t-1}\right) dh_{t-1}$$

Let us now consider the joint density of y_t, h_t. It can be decomposed into the product of the measurement density and the prediction density,

3. $$f\left(y_t, h_t \middle| Y^{t-1}\right) = f\left(y_t \middle| h_t\right) f\left(h_t \middle| Y^{t-1}\right)$$

and, again, by marginalisation we obtain the one-step-ahead forecast density of y_t

4. $$f\left(y_t \middle| Y^{t-1}\right) = \int f\left(y_t, h_t \middle| Y^{t-1}\right) dh_t = \int f\left(y_t \middle| h_t\right) f\left(h_t \middle| Y^{t-1}\right) dh_t$$

which is particularly useful, since by the combination of these densities it is possible to obtain the likelihood function. Finally, by conditioning we obtain the filtering density (output)

---

[10] For the first iteration (t=1) it is possible to consider the unconditional distribution of h_t, $f\left(h_1\right)$. For the sake of simplicity we omit the dependence on the parameter θ.

5. $$f\left(h_t\middle|Y^t\right)=\frac{f\left(y_t,h_t\middle|Y^{t-1}\right)}{f\left(y_t\middle|Y^{t-1}\right)}=\frac{f\left(y_t|h_t\right)f\left(h_t\middle|Y^{t-1}\right)}{\int f\left(y_t|h_t\right)f\left(h_t\middle|Y^{t-1}\right)dh_t}$$

which ends the iteration.

The previous algorithm allows us to obtain several important elements. Step 2 gives the estimation of h_t given all the information available until *t*-1 (prediction density). Step 5 provides the estimation of h_t given all the information currently available (filtering density). Finally, step 4 by providing the one-step-ahead forecast density, allows us to compute the likelihood function.

Unfortunately, only in very special cases is it possible to obtain analytic recursive algorithms[11] from this general filtering algorithm: the Kalman filter in the Gaussian and linear case and the Hamilton filter in the Markovian and discrete case.

In the Gaussian and linear cases, the initial input $f\left(h_1\middle|Y^0\right)$ and the measurement and transition densities are assumed to be Gaussian and at each step of the algorithm Gaussianity is preserved, then also all the outputs are Gaussian. The Normal distribution is completely described by its first two moments and then the algorithm can be rewritten by relating means and variances of the different densities involved. This is the Kalman filter.

For the switching regime models introduced by Hamilton (1989), which represent the Markovian and discrete case, the integrals which appear at steps 2 and 4 become a simple sum over the possible regimes, and then the whole algorithm is analytically tractable.

In all the other cases, it is necessary to consider approximated solutions or simulation-based methods. Examples of approximations are the extended Kalman filter (Anderson and Moore (1979), Harvey (1989), Fridman and Harris (1998)), the Gaussian sum filter (Sorenson and Alspach (1971)), the numerical integration (Kitagawa (1987)), the Monte Carlo integration (Tanizaki and Mariano (1994, 1998)), or the particle filter (Gordon, Salmond and Smith (1993), Kitagawa (1996), Pitt and Shephard (1999a)). The simulation-based solutions are certainly more time consuming and demanding in terms of computing, but they are definitely more general. We will see these methods in greater detail later.

However, for the two presented models (2) and (3), the general filter introduced here is useful for estimation. In fact, for the linearised version (6), the Kalman filter allows a quasi maximum likelihood estimation of the parameters and the discrete version (3) is a particular case of switching regime models for which the Hamilton filter gives the

[11] See also Shephard (1994a) for another particular case in which h_t is set to be a random walk and $\exp(\eta_t)$ a highly contrived scaled beta distribution. This delivers a one-step-ahead prediction distribution which has some similarities to the ARCH models.

likelihood function.

4.1.1. The Kalman filter for quasi maximum likelihood (QML) estimation of continuous SV models

We can consider the log-transformation (6) of the continuous SV model. As $\log\left(\varepsilon_t^2\right)$~*iid*, we obtain a linear state space model.

Let $LY^{\tau} = \left(\log\left(y_1^2\right), \log\left(y_2^2\right), ..., \log\left(y_{\tau}^2\right)\right)'$, $\hat{h}_{t/\tau} = E\left(h_t \middle| LY^{\tau}\right) = E\left(h_t \middle| Y^{\tau}\right)$ and $Q_{t/\tau} = MSE\left(h_t \middle| LY^{\tau}\right) = MSE\left(h_t \middle| Y^{\tau}\right)$. The Kalman filter (see for example[12] Harvey (1989)) computes these quantities recursively for t=1,…,T,

$$\hat{h}_{t/t-1} = \alpha + \beta \hat{h}_{t-1/t-1}$$
$$Q_{t/t-1} = \beta^2 Q_{t-1/t-1} + \sigma_{\eta}^2$$

$$e_{t/t-1} = \log\left(y_t^2\right) - \hat{h}_{t/t-1}$$
$$F_{t/t-1} = Q_{t/t-1} + \pi^2 \big/ 2$$

$$\hat{h}_{t/t} = \hat{h}_{t/t-1} + K_t e_{t/t-1}$$
$$Q_{t/t} = (1 - K_t)^2 Q_{t/t-1}$$

where $K_t = Q_{t/t-1} F_{t/t-1}^{-1}$ is the Kalman gain.

However, as $\log\left(\varepsilon_t^2\right)$ is not Gaussian, the Kalman filter can be used to provide the best linear unbiased estimator of h_t given Y^t.

Moreover, if (6) were a Gaussian state space model, the Kalman filter would provide the exact likelihood function. In fact, a bi-product of the filter are the innovations $e_{t/t-1}$, which are the one-step-ahead forecast errors and their corresponding mean square errors, $F_{t/t-1}$. Together they deliver the likelihood (ignoring constants):

$$\ell_T(\theta) = -\frac{1}{2}\sum_{t=1}^{T} \log F_{t/t-1} - \frac{1}{2}\sum_{t=1}^{T} \frac{e_{t/t-1}^2}{F_{t/t-1}}$$

As the state space is linear but not Gaussian, the filter gives a quasi likelihood function which can be used to obtain a consistent estimator $\hat{\theta}$ and asymptotically normal inference (see Ruiz (1994)).

[12] See also Carraro and Sartore (1987).

This way of estimating h_t is used by Melino and Turnbull (1990), after estimating θ by the generalised method of moments (see section 4.3.1). Harvey, Ruiz and Shephard (1994) examine the QML estimator.

4.1.2. The Hamilton filter for maximum likelihood estimation of discrete SV models

The discrete SV model (3) is a non-linear and non-Gaussian state-space model. In the two-regimes case, the transition equation can be written in a linear form (see (5)) and the measurement equation can be linearised by the log transformation, but both the equations are non-Gaussian. However, the joint process (y_t, h_t) is Markovian and thus the general filter presented in section 4.1 gives an analytic recursion, since the integrals become simple sums over the possible values of h_t. The input is the filtered probability[13] $P\left(h_{t-1}\middle|Y^{t-1}\right)$ and the algorithm gives the prediction probability, the one-step-ahead forecast density and the subsequent filtered probability:

$$P\left(h_t\middle|Y^{t-1}\right)=\sum_{h_{t-1}} P\left(h_t|h_{t-1}\right)P\left(h_{t-1}\middle|Y^{t-1}\right)$$

$$f\left(y_t\middle|Y^{t-1}\right)=\sum_{h_t} f\left(y_t|h_t\right)P\left(h_t\middle|Y^{t-1}\right)$$

$$P\left(h_t\middle|Y^{t}\right)=\frac{f\left(y_t|h_t\right)P\left(h_t\middle|Y^{t-1}\right)}{\sum_{h_t} f\left(y_t|h_t\right)P\left(h_t\middle|Y^{t-1}\right)}$$

The combination of the one-step-ahead forecast densities:

$$\ell_T(\theta)=\prod_{t=1}^{T} f\left(y_t\middle|Y^{t-1}\right)$$

provides the likelihood function, the maximization of which gives the maximum likelihood estimators of the parameters.

4.2. A general smoother for non-Gaussian parameter-driven models

We also might want to obtain the estimation of h_t given all the information available, that is conditional on Y^T. Such a procedure is called smoothing and as before it is possible to derive a formal backward algorithm which delivers the smoothed densities $f\left(h_t\middle|Y^T\right)$.

[13] The initial probability $P\left(h_0\middle|Y^0\right)$ can be taken equal to the unconditional (ergodic) probability $P(h_0)=\pi$.

Let $f\left(h_{t+1}\middle|Y^T\right)$ be the input of the iteration[14] $t=T-1,T-2,\ldots,2,1$. We can decompose the joint density of h_{t+1},h_t, conditional on the information set Y^t, in the product of the transition density by the filtered density (available from the filter):

$$\textbf{1.}\quad f\left(h_{t+1},h_t\middle|Y^t\right)=f\left(h_{t+1}\middle|h_t\right)f\left(h_t\middle|Y^t\right)$$

By conditioning with the prediction density obtained from the filter, we obtain the following conditional density:

$$\textbf{2.}\quad f\left(h_t\middle|h_{t+1},Y^t\right)=\frac{f\left(h_{t+1},h_t\middle|Y^t\right)}{f\left(h_{t+1}\middle|Y^t\right)}$$

The joint density of h_{t+1},h_t, conditional on the information set Y^T, is given by the product of the conditional density $f\left(h_t\middle|h_{t+1},Y^T\right)$ by the input of the algorithm $f\left(h_{t+1}\middle|Y^T\right)$. The information set h_{t+1},Y^T is included in the information set $h_{t+1},Y^t,\varepsilon_{t+1}^T,\eta_{t+2}^T$, where $\varepsilon_{t+1}^T=\left(\varepsilon_{t+1},\ldots,\varepsilon_T\right)'$ and $\eta_{t+2}^T=\left(\eta_{t+2},\ldots,\eta_T\right)'$. Given that $\varepsilon_{t+1}^T,\eta_{t+2}^T$ is independent of h_t,h_{t+1},Y^t, we can conclude that $f\left(h_t\middle|h_{t+1},Y^T\right)=f\left(h_t\middle|h_{t+1},Y^t\right)$ (computed at step 2) and then

$$\textbf{3.}\quad f\left(h_{t+1},h_t\middle|Y^T\right)=f\left(h_t\middle|h_{t+1},Y^T\right)f\left(h_{t+1}\middle|Y^T\right)=f\left(h_t\middle|h_{t+1},Y^t\right)f\left(h_{t+1}\middle|Y^T\right)$$

Finally, by marginalisation we obtain the smoothed density of h_t (output):

$$\textbf{4.}\quad f\left(h_t\middle|Y^T\right)=\int f\left(h_{t+1},h_t\middle|Y^T\right)dh_{t+1}=\int f\left(h_t\middle|h_{t+1},Y^t\right)f\left(h_{t+1}\middle|Y^T\right)dh_{t+1}$$

Again, only in the linear and Gaussian case, and in the Markovian and discrete case is it possible to obtain an analytic backward recursion: the Kalman smoother and the Kim smoother (Kim (1994)).

4.2.1. The Kalman smoother for continuous SV models

Let $\hat{h}_{t+1/T}=E\left(h_{t+1}\middle|LY^T\right)=E\left(h_{t+1}\middle|Y^T\right)$ and $Q_{t+1/T}=MSE\left(h_{t+1}\middle|LY^T\right)=MSE\left(h_{t+1}\middle|Y^T\right)$. The Kalman smoother[15] computes these quantities recursively for $t=T-1,T-2,\ldots,2,1$,

[14] For the first iteration ($t=T-1$), the input is simply the final output of the filter $f\left(h_T\middle|Y^T\right)$.

$$\hat{h}_{t/T} = \hat{h}_{t/t} + \beta Q_{t/t} Q_{t+1/t}^{-1} \left(\hat{h}_{t+1/T} - \hat{h}_{t+1/t} \right)$$
$$Q_{t/T} = Q_{t/t} + \beta^2 Q_{t/t}^2 Q_{t+1/t}^{-2} \left(Q_{t+1/T} - Q_{t+1/t} \right)$$

where $\hat{h}_{t/t}$, $Q_{t/t}$, $\hat{h}_{t+1/t}$, $Q_{t+1/t}$ are stored from the Kalman filter.

For the log-transformation of the continuous SV model (6), the Kalman smoother is useful in estimating the unobserved log-volatility, in fact it provides the best linear unbiased estimator of h_t given $\left(y_1, y_2, ..., y_T\right)'$.

4.2.2. The Kim smoother for discrete SV models

The input is the smoothed probability $P\left(h_{t+1} \middle| Y^T\right)$ and the recursion is simply:

$$P\left(h_t \middle| Y^T\right) = \sum_{s_{t+1}} \frac{P\left(h_{t+1} \middle| h_t\right) P\left(h_t \middle| Y^t\right) P\left(h_{t+1} \middle| Y^T\right)}{P\left(h_{t+1} \middle| Y^t\right)}$$

where $P\left(h_t \middle| Y^t\right)$ and $P\left(h_{t+1} \middle| Y^t\right)$ are stored from the Hamilton filter.

4.3. Other estimation methods for continuous SV models

For the discrete SV model the Hamilton filter allows us to obtain the maximum likelihood estimator of the parameters. On the contrary, for the continuous SV models, the Kalman filter provides only an approximation of the likelihood function. Let us review some other possible estimation methods useful for the continuous SV model.

Like most non-Gaussian parameter-driven models, there are many different ways to perform estimation. Some involve estimating the likelihood; others use method of moments procedures.

4.3.1. Method of moments

The simplest approach is the method of moments, based on matching empirical and theoretical moments. 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 lead to loss of information. Examples include those based on y_t^2, y_t^4, $y_t^2 y_{t-r}^2$, although there are many other possibilities. As a result, we may well want to use more moments than there are parameters to estimate,

[15] See also de Jong (1989).

implying that they will have to be pooled. A reasonably sensible way of doing this is via the Generalised Method of Moments (GMM).

We can consider, for example, the vector g_T of the first *r* autocovariances of y_t^2 or of $\log\left(y_t^2\right)$ as the moment constraints. There are more moments than parameters and the issue is how to weight all the available information. The GMM approach of Hansen (1992) suggests minimising the quadratic form $g_T' W_T g_T$ by varying the parameters θ and the weighting matrix W_T should reflect the relative importance given to matching each of the chosen moments. Applications of this method to SV models are the seminal Melino and Turnbull (1990) and the extensive study of Andersen and Sørensen (1996).

The main advantage of the GMM approach comes from the fact that it does not require distributional assumptions. However, this is not useful for the SV model since it is a fully specified parametric model. On the contrary, as argued by Shephard (1996), there are a number of drawbacks to the GMM estimation of the SV model. First of all, GMM can only be used if h_t is stationary; if β is close to one (as we will find for many high frequency financial data sets), we can expect GMM to work poorly. Second, parameter estimates are not invariant to the parameterization and the model (2) is not fundamentally more interesting than

$$\begin{cases} y_t = \varepsilon_t \gamma \exp(h_t/2), \\ h_t = \beta h_{t-1} + \eta_t \end{cases}$$

Third, as already observed, the squares y_t^2 behave like an ARMA(1,1) model; if σ_η^2 is small (as we will find in practice), $\rho_{y_t^2}(r)$ will be small but positive for many *r*. This implies that for many series the number of moments to be considered will have to be very high to capture the low correlation in the volatility process. Finally, GMM does not deliver an estimate (filtered or smoothed) of h_t, consequently a second form of estimation will be required.

The GMM and QML approaches are the simplest way of estimating the SV models and they are about equally efficient, with the relative performance being dependent on the specific parameter values (see on this Andersen and Sørensen (1997)).

4.3.2. Simulation-based methods

All the others estimation approaches are based on simulation techniques. In the last ten years there has been a growing interest in simulation[16]-based methods which propose several ways of resolving the inference problem for this class of models (see Billio (1999) and Billio (2002b)). In fact, it is clear that one can easily recursively simulate (path simulations) from the system (2) for any given value of parameters, θ.

[16] Simulation techniques make use of sequences of pseudo-random numbers which are generated by a computer procedure.

A first approach relies on simulation-based methods which are relatively simple to implement, but which are less efficient than the maximum likelihood approach: see, for example, the Simulated Method of Moments (Duffie and Singleton (1993)), the Indirect Inference Method (Gouriéroux, Monfort and Renault (1993)) or the Efficient Method of Moments (Gallant and Tauchen (1996), Gallant, Hsieh and Tauchen (1997)). A second approach considers the problem of the computation (or of the approximation) of the likelihood and then of the maximum likelihood estimator through importance sampling methods (Danielsson and Richard (1993), Danielsson (1994), Durbin and Koopman (1997)). In a Bayesian framework, a third approach considers Markov Chain Monte Carlo (MCMC) techniques based on the data augmentation principle, which yields samples out of the joint posterior distribution of the latent volatility and all model parameters, and allows the parameter estimates and the latent volatility dynamics to be obtained (Jacquier, Polson and Rossi (1994), Kim, Shephard and Chib (1998), Chib, Nardari and Shephard (2002)). Finally, a fourth approach utilizes MCMC methods to compute (or approximate) the maximum likelihood estimator (see the Simulated Expectation Maximisation (Shephard (1993) and Geyer (1994), Geyer (1996), Billio, Monfort and Robert (1998)).

In practice, the choice between these different simulation-based approaches depends on several criteria, such as efficiency and computing time. Unfortunately, in general there is a trade off between these criteria. Methods like the Simulated Maximum Likelihood and the Simulated Likelihood Ratio have several advantages in the estimation of SV models. Since they are likelihood methods, the classical theory of maximum likelihood carries over to the simulated case and standard likelihood ratio tests can be constructed. MCMC-based approaches are certainly more time consuming, but also allow estimation of the latent volatility dynamics by simulating from the smoothing/posterior distribution of h_t.

Let us briefly introduce part of these methods and their application to SV models.

4.3.2.1. Indirect Inference approach

The so-called Indirect Inference methodology was recently introduced in the literature by Smith (1993), Gouriéroux, Monfort and Renault (1993), Gallant and Tauchen (1996), for a simulation-based inference on generally intractable structural models through an auxiliary model, conceived as easier to handle. This methodology allows the use of somewhat misspecified auxiliary models, since the simulation process in the well-specified structural model and the calibration of the simulated paths against the observed one through the same auxiliary model will provide an automatic misspecification bias correction. There are several ways of implementing this idea[17].

The original approach is the Indirect Inference Method of Gouriéroux, Monfort and Renault (1993). Consider an auxiliary model $f_a\left(y_t \middle| Y^{t-1}; \pi\right)$ for the observed data (for example[18] the general linear state-space model obtained by the log-transformation

[17] For all these methods, it is necessary to recycle the random numbers used in the calculation when θ changes, in order to have good numerical and statistical properties of the estimators based on these simulations.

[18] Another possible auxiliary model is an ARMA(*p,q*) on the logarithms of the squared data (see Monfardini (1998)).

(6)). Let $\hat{\pi}_T = \Pi_T\left(Y^T\right)$ denote the QML estimator of π based on f_a as a function $\Pi_T(\cdot)$ of the observed data set Y^T. The Indirect Inference estimator of structural parameters θ is given by:

$$\hat{\theta}_{II} = \arg\min_{\theta}\left[\hat{\pi}_T - \tilde{\pi}_{NT}(\theta)\right]' W_T \left[\hat{\pi}_T - \tilde{\pi}_{NT}(\theta)\right]$$

where W_T is a weighting matrix and $\tilde{\pi}_{NT}(\theta)$ is the π estimator obtained on a simulated path of $\tilde{Y}^{NT}$ for a given value of θ (i.e. that is given by the binding function $\tilde{\pi}_{NT}(\theta) = \lim_{N\to\infty} \Pi_{NT}\left(\tilde{Y}^{NT}\right)$, which is approximated by $\Pi_{NT}\left(\tilde{Y}^{NT}\right)$ for large *N*). This approach may be very computationally demanding as one needs to evaluate the binding function $\tilde{\pi}_{NT}(\theta)$ for each value of θ appearing in the numerical optimisation algorithm.

The estimator of Gallant and Tauchen (1996) circumvents the need to evaluate the binding function by using the score vector $\frac{\partial}{\partial \pi} f_a\left(y_t \middle| Y^{t-1}; \pi\right)$ (score generator) to define the matching conditions. If the auxiliary model $f_a\left(y_t \middle| Y^{t-1}; \pi\right)$ is chosen flexibly with a suitable nonparametric interpretation, then the estimator achieves the asymptotic efficiency of maximum likelihood and has good power properties for detecting misspecification (Gallant and Long (1997), Tauchen (1997)), hence the term Efficient Method of Moments (EMM). EMM delivers consistent estimates of the structural parameter vector under weak conditions on the choice of the auxiliary model. However, extrapolating from the Generalised Method of Moments evidence, it is natural to conjecture that the quality of inference may hinge on how well the auxiliary model approximates the salient features of the observed data. This intuition is formalized by Gallant and Long (1997), who show that a judicious selection of the auxiliary model, ensuring that the quasi-scores asymptotically span the true score vector, will result in full asymptotic efficiency[19].

Andersen, Chung and Sorensen (1999) perform an extensive Monte Carlo study of EMM estimation of a stochastic volatility model. They examine the sensitivity to the choice of auxiliary model using ARCH, GARCH, and EGARCH models for the score as well as nonparametric extensions. EMM efficiency approaches that of maximum likelihood for larger sample sizes, while inference is sensitive to the choice of auxiliary model in small samples, but robust in larger samples[20].

The Indirect Inference theory, however, crucially depends on the correct specification

[19] In fact, as the score generator approaches the true conditional density, the estimated covariance matrix for the structural parameter approaches that of maximum likelihood. This result embodies one of the main advantages of EMM, since it prescribes a systematic approach to the derivation of efficient moment conditions for estimation in a general parametric setting.

[20] Care must be taken, however, to avoid over-parameterization of the auxiliary model, as convergence problems may arise if the quasi-score is extended to the point where it begins to fit the purely idiosyncratic noise in the data.

assumption concerning the structural model. There is now an emerging literature (see, for example, Dridi and Renault (2000) and Dridi (2000)) which focuses on procedures more robust to the structural model specification. In particular, Dridi and Renault (2000) propose an extension to the Indirect Inference methodology to semiparametric settings and show how the Semiparametric Indirect Inference works on basic examples using SV models.

4.3.2.2. Importance sampling

A more direct way of performing inference is to compute the likelihood by integrating out the latent h_t process. As previously seen, the integral (8) has no closed form and it has to be computed numerically. However, the likelihood function naturally appears as the expectation of the function $\prod_{t=1}^{T} f\left(y_t \middle| Y^{t-1}, H^t; \theta\right)$ with respect to the p.d.f. P defined by[21] $\prod_{t=1}^{T} f\left(h_t \middle| Y^{t-1}, H^{t-1}; \theta\right)$, from which it is easy to recursively draw. Therefore, an unbiased simulator of the whole likelihood function $\ell_T(\theta)$ is $\prod_{t=1}^{T} f\left(y_t \middle| Y^{t-1}, {}^n\widetilde{H}^t; \theta\right)$ where ${}^n\widetilde{H}^t$ are recursively drawn from the auxiliary p.d.f. P.
The likelihood is then approximated by the empirical mean:

$$\frac{1}{N}\sum_{n=1}^{N}\prod_{t=1}^{T} f\left(y_t \middle| Y^{t-1}, {}^n\widetilde{H}^t; \theta\right).$$

and this simulated likelihood can be numerically maximised. However, this basic simulator may be very slow, in the sense that the simulator may have a very large variance and then some accelerating technique is needed. One solution is to consider the general method of importance sampling based on a sequence of conditional p.d.f.'s $q\left(h_t \middle| Y^T, H^{t-1}1\right)$. Let us denote this probability distribution by Q and the corresponding expectation by E_Q. We have:

$$\ell_T(\theta) = E_P\left[\prod_{t=1}^{T} f\left(y_t \middle| Y^{t-1}, H^t; \theta\right)\right]$$
$$= E_Q\left[\prod_{t=1}^{T} \frac{f\left(y_t \middle| Y^{t-1}, H^t; \theta\right) f\left(h_t \middle| Y^{t-1}, H^{t-1}; \theta\right)}{q\left(h_t \middle| Y^T, H^{t-1}; \theta\right)}\right]$$

Therefore, an unbiased simulator of $\ell_T(\theta)$ is:

---

[21] It is important to note that this p.d.f. is neither $f\left(H^T; \theta\right)$, except when y_t does not cause h_t, nor $f\left(H^T \middle| Y^T; \theta\right)$.

$$E_Q\left[\prod_{t=1}^{T}\frac{f\left(y_t\left|Y^{t-1},{}^n\widetilde{H}^t;\theta\right.\right)f\left({}^n\widetilde{h}_t\left|Y^{t-1},{}^n\widetilde{H}^{t-1};\theta\right.\right)}{q\left({}^n\widetilde{h}_t\left|Y^T,{}^n\widetilde{H}^{t-1};\theta\right.\right)}\right]$$

where ${}^n\widetilde{H}^{-T}$ is drawn in Q. The problem is then how to choose the importance function: the natural answer is by reducing the Monte Carlo variance. It is easy to calculate the theoretical optimal choice $f\left(H^T\left|Y^T;\theta\right.\right)=\prod_{t=1}^{T}f\left(h_t\left|Y^T,H^{t-1}\right.\right)$ (i.e. the smoothing density of h_t), for which one simulation is sufficient, but it is clearly not computable. Then it is possible to consider the smoothing density of an approximating model, and fix a parametric family of importance functions, choosing the member that minimizes the Monte Carlo variance (which is eventually computed in an approximated way). For the SV model (2), the first solution is proposed by Sandmann and Koopman (1998) by using as approximating model the linearised version (6). In the aim of the second solution, Danielsson and Richard (1993) propose a sequentially optimized importance sampling, which Danielsson (1994) applies to the SV model[22]. In both cases, the Simulated Maximum Likelihood estimates of model parameters are obtained by numerical optimization of the logarithm of the simulated likelihood[23].

4.3.2.3. Bayesian approach

In the Bayesian setting, there are also serious difficulties in estimating the SV model. In general, the posterior density $f\left(\theta\left|Y^T\right.\right)$ and the posterior expectation of θ cannot be computed in a closed form. Again, this complex setting requires a simulation-based approach. The data augmentation principle, which considers the latent variable h_t as nuisance parameters, and the utilisation of Gibbs sampling (Gelfand and Smith (1990)), by iterating simulations from $f\left(H^T\left|Y^T,\theta\right.\right)$ (data augmentation step) and $f\left(\theta\left|Y^T,H^T\right.\right)$ (parameter simulation step), allow simulation from the joint posterior distribution $f\left(H^T,\theta\left|Y^T\right.\right)$, derivation of the distribution of interest as the marginal distribution of θ and approximation of the posterior expectation by a sample average. When conditional distributions cannot be directly simulated, the corresponding steps in the Gibbs algorithm are replaced by Metropolis-Hastings steps[24]. Moreover, the prior modelling on the parameters is usually quasi non-informative.

One way of considering 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.

---

[22] The details will not be dealt with here as they are quite involved, even for the simplest model.

[23] As for non-efficient methods, numerical and statistical accuracy is obtained by recycling the random numbers used in the calculation for each parameter value.

[24] Such hybrid algorithms are validated in Tierney (1994).

In Jacquier, Polson and Rossi (1994) the posterior distribution of the parameters was sampled by MCMC methods using a one-move approach (i.e. the latent variables h_t were sampled each at time from $f\left(h_t \middle| Y^T, H^{-t}, \alpha, \beta, \sigma_\eta^2\right)$, where H^{-t} denotes all the elements of H^T excluding h_t). Although this algorithm is conceptually simple, it is not particularly efficient from a simulation perspective, as is shown by Kim, Shephard and Chib (1998), who develop an alternative, more efficient, multi-move MCMC algorithm. The efficiency gain in the Kim, Shephard and Chib (1998) algorithm arises from the joint sampling of H^T in one block conditioned on everything else in the model. Finally, Chib, Nardari and Shephard (2002) develop efficient Markov Chain Monte Carlo algorithms for estimating generalized models of SV defined by heavy-tailed Student-t distributions, exogenous variables in the observation and volatility equations, and a jump component in the observation equation (see section 5.1).

4.3.2.4. A MCMC approach to maximum likelihood estimation

Although the Bayesian approach is straightforward to state and computationally attractive, it requires the elicitation of a prior, which is often regarded by some econometricians as being difficult in dynamic models. Even if this is not an insurmountable problem, alternatives are available which allow us to perform maximum likelihood estimation using MCMC methods.

The first possibility is the Simulated Expectation Maximisation (SEM) algorithm proposed by Shephard (1993). The EM algorithm exploits the following decomposition of the log likelihood function:

$$\begin{aligned}\log f\left(Y^T;\theta\right) &= \log f\left(Y^T, H^T;\theta\right) - \log f\left(H^T \middle| Y^T;\theta\right) \\ &= E\left[\log f\left(Y^T, H^T;\theta\right) \middle| Y^T\right] - E\left[\log f\left(H^T \middle| Y^T;\theta\right) \middle| Y^T\right]\end{aligned}$$

and iterates:

$$\theta^{i+1} = \arg\max_\theta E_{\theta^i}\left[\log f\left(Y^T, H^T;\theta\right) \middle| Y^T\right]$$

This is an increasing algorithm such that the sequence θ^i converges to the ML estimator. The problem is that, although $\log f\left(Y^T, H^T;\theta\right)$ has in general a closed form, the same is not true for its conditional expectation. In the SEM algorithm this expectation is replaced by an approximation based on simulations. Thus, the problem is now to be able to draw in the conditional distribution of H^T given Y^T and θ. Shephard (1993), in the context of a non-linear state space model, uses the Hastings-Metropolis algorithm to solve this problem, and applies it to the SV model.

Another possible approach is the Simulated Likelihood Ratio (SLR) method proposed

by Billio, Monfort and Robert (1998). The general principle is:

$$\frac{f\left(Y^T;\theta\right)}{f\left(Y^T;\bar{\theta}\right)} = E_{\bar{\theta}}\left[\left.\frac{f\left(Y^T,H^T;\theta\right)}{f\left(Y^T,H^T;\bar{\theta}\right)}\right| Y^T\right] \tag{9}$$

where $\bar{\theta}$ is an arbitrary fixed value of the parameters. Obviously,

$$\arg\max_{\theta} f\left(Y^T;\theta\right) = \arg\max_{\theta} \frac{f\left(Y^T;\theta\right)}{f\left(Y^T;\bar{\theta}\right)}$$

and with ${}^n\tilde{H}^T$, $n = 1,2,...,N$, simulated paths in the conditional distribution $f\left(H^T\middle|Y^T;\theta\right)$, the SLR method amounts to maximising:

$$\frac{1}{N}\sum_{n=1}^{N}\frac{f\left({}^n\tilde{H}^T,Y^T;\theta\right)}{f\left({}^n\tilde{H}^T,Y^T;\bar{\theta}\right)}$$

with respect to θ. The method can be implemented by simulating the conditional distribution[25] $f\left(H^T\middle|Y^T;\bar{\theta}\right)$. As already noted, it is impossible to simulate directly this distribution, thus a Hastings-Metropolis approach is suggested.

Contrary to the SEM approach, the SLR method allows for the computation of the likelihood surface and then of likelihood ratio test statistics. It needs only one optimisation run and not a sequence of optimisations; it is possible to store the simulated paths, and then only one simulation run is required. Moreover, as the simulation is made for only one value of the parameter, the objective function will be smooth with respect to θ, even if simulations involve rejection methods.

Billio, Monfort and Robert (1998) apply the SLR method also to the SV model (2).

5. EXTENSIONS OF SV MODELS

The basic SV models can be generalised in a number of directions. Straightforward generalisations might allow ε_t to have heavy-tailed Student-t distributions and exogenous variables in the observation and volatility equations.

Moreover, the ARCH in mean model of Engle, Lilien and Robins (1987) can be extended to the SV framework, by specifying $y_t = \mu_0 + \mu_1 \exp(h_t) + \varepsilon_t \exp(h_t/2)$.

[25] The resulting Monte Carlo approximation of (9) could be only locally good around $\bar{\theta}$, and so Geyer (1996) suggests updating $\bar{\theta}$ to the maximiser of the Monte Carlo likelihood and repeating the Monte Carlo procedure using the new $\bar{\theta}$. By updating $\bar{\theta}$ a few times, one should obtain better approximations of the relative likelihood function near the true maximum likelihood estimate.

This model allows y_t to be moderately serially correlated, but in the discrete SV model the Hamilton filter no longer works, because y_t, h_t are not jointly Markovian.

5.1. Extensions of continuous SV models

In the context of continuous SV models, Harvey, Ruiz and Shephard (1994) concentrated their attention on models based on Student-t error; Mahieu and Schotman (1998) analysed the possibility of using a mixture distribution. Jacquier, Polson and Rossi (1995) have computed the posterior density of the parameters of a Student-t-based SV model. This particular type of model in fact can be viewed as a Euler discretisation of a Student-t based Levy process but with additional stochastic volatility effects; further articles are available in (continuous-time) mathematical options and risk assessment literature[26]. By building on the work of Kim, Shephard and Chib (1998), Chib, Nardari and Shephard (2002) develop efficient Markov Chain Monte Carlo algorithms for estimating these models. They also consider a second type of models which contain a jump component[27] in the observation equation to allow for large, transient movements.

Moreover, a natural framework for extension of continuous SV models might be based on adapting the Gaussian state space so that:

$$\begin{cases} y_t = \varepsilon_t \exp\left(z_t' h_t/2\right), \\ h_t = T_t h_{t-1} + \eta_t, \qquad \eta_t \sim N\left(0, \mathrm{H}_t\right) \end{cases}$$

and then on allowing h_t to follow a more complicated ARMA process. Another simple example would be:

$$z_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad h_t = \begin{pmatrix} \beta & 0 \\ 0 & 1 \end{pmatrix} h_{t-1} + \eta_t, \quad \eta_t \sim N\left\{0, \begin{pmatrix} \sigma_{\eta_1}^2 & 0 \\ 0 & \sigma_{\eta_2}^2 \end{pmatrix}\right\}$$

Now, the second component of h_t is a random walk, allowing the permanent level of the volatility to slowly change. 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 (1993), who allow (ignoring the cyclical AR(1) component):

$$z_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad T_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathrm{H}_t = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_{\eta_2}^2 \end{pmatrix}$$

[26] Leading references include Eberlein (2001) and Eberlein and Prause (2001). The extension to allow for stochastic volatility effects is discussed in Eberlein and Prause (2001) and Eberlein, Kallsen and Kristen (2001).

[27] Jump models are quite popular in continuous time models of financial asset pricing. See, for example, Merton (1976), Ball and Torous (1985), Bates (1996), Duffie, Pan and Singleton (2000).

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 volatilities h_{t+r} may grow very quickly with *r*.

Another suggestion is to allow h_t to be a fractional process, giving the long-memory SV model. For financial time series, there is strong evidence that the effect of a shock to volatility persists (i.e. is not absorbed) for a long number of periods (see e.g. Andersen and Bollerslev (1997), Lobato and Savin (1998), Harvey (1998), Bollerslev and Jubinski (1999), Bollerslev and Mikkelsen (1999), Bollerslev and Wright (2000) and Ray and Tsay (2000)), thus the concept of long memory seems suitable and has been suggested by Breidt, Crato and de Lima (1998). A covariance stationary time series y_t has long memory if:

$$\sum_{r=0}^{\infty} \left| Cov\left(y_t, y_{t-r}\right) \right| = \infty$$

with $Var\left(y_t\right) < \infty$. Basically, it says that the autocovariances do decay as the lag increases but very slowly, usually hyperbolically.

Currently there exist four approaches to estimate the long-memory SV model. The quasi maximum likelihood estimator of Breidt, Crato and de Lima (1998), the GMM approach of Wright (1999), the widely-used semiparametric, log-periodogram estimator of Geweke and Porter-Hudak (1983) (see e.g. Andersen and Bollerslev (1997), Ray and Tsay (2000), Wright (2000), Deo and Hurvich (2001) and the recent developments of Hurvich and Ray (2001), Hurvich, Moulines and Soulier (2001)) and the Bayesian estimator based on the Markov Chain Monte Carlo sampler (Chan and Petris (1999)) and eventually the wavelet representation of the log-squared returns (Jensen (1999), (2000), (2001)).

The quasi MLE of the long-memory SV model is known to be strongly consistent, but requires the order of the short-memory autoregressive and moving average parameters to be correctly identified, as does the GMM estimator. The difference between the quasi MLE and GMM is that when the fractional order of integration is smaller than 1/4, the asymptotic properties in addition to consistency are known for the GMM estimator. Unlike the quasi MLE of a short-memory stochastic volatility model whose asymptotic properties are known (Harvey, Ruiz and Shepard (1994) and Ruiz (1994)), these other asymptotic properties are not yet known for the quasi MLE of the long-memory SV model. However, in simulation experiments, Wright (1999) finds that neither estimator's finite-sample properties dominate the other; the GMM estimator of the long-memory parameter generally produces smaller standard errors but with a significant downward bias. From these simulations Wright (1999) admonishes developing alternative estimators of the long-memory SV model that are more efficient and less biased. In fact, even if Deo and Hurvich (2001) find the asymptotic properties for the log-periodogram estimator of volatility to be similar to those proved by Robinson (1995) for the same estimator in the mean, correct inference about the degree of long-memory relies on the number of Fourier frequencies in the regression

growing at a rate that is dependent on the value of the unknown long-memory parameter. It thus seems that neither the quasi MLE nor the log-periodogram estimator of the long-memory volatility model lend themselves nicely to the construction of confidence intervals or hypothesis testing of the long-memory parameter estimate.

Finally, it could be useful to allow the SV model to capture the non-symmetric response to shocks. This feature can be modelled by allowing ε_{t-1} and η_t to be correlated. If ε_{t-1} and η_t are negatively correlated, and if $\varepsilon_{t-1} > 0$, then $y_{t-1} > 0$ and h_t is likely to fall. Hence, a large effect of y_{t-1}^2 on the estimated h_t will be accentuated by a negative sign on y_{t-1}, while its effect will be partially ameliorated by a positive sign. This correlation was suggested by Hull and White (1987) and estimated using GMM by Melino and Turnbull (1990) and Scott (1991). A simple quasi maximum likelihood estimator has been proposed by Harvey and Shephard (1996). Jacquier, Polson and Rossi (1995) have extended their single move MCMC sampler to estimate this effect.

5.2. Extensions of discrete SV models

In the discrete case, the basic model might be extended by considering different Markov chains, which can allow the decomposition of shocks into permanent and transitory as in the continuous case:

$$\begin{cases} y_t = \varepsilon_t \exp\left(z_t' h_t / 2\right), \\ h_t = \mu + T S_t, \end{cases} \tag{10}$$

where S_t represents a vector of Markov chains. However, the vector of Markov chains can be easily represented by a single Markov chain with a sufficient number of states and then the model (10) formally reduces to the basic model (3).

Finally, the two SV models can be combined by allowing the continuous latent volatility to be governed by a first-order Markov chain. In that case, the estimation is very difficult. So, Lam and Lee (1998) therefore propose Bayesian estimators which are constructed by Gibbs sampling.

6. MULTIVARIATE MODELS

Most macroeconomics and finance is about how variables interact, thus a multivariate approach is very important. For multivariate stochastic volatility models this means that it is essential to capture changing cross-covariance patterns. Multivariate modelling of covariance is rather new and difficult because it is afflicted by extreme problems of lack of parsimony. From a modelling point of view, the multivariate SV models are easier to extend than the ARCH models, but the estimation problem remains.

6.1. Multivariate continuous SV models

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

$$y_{it} = \varepsilon_{it} \exp(h_{it}/2), \quad i = 1,...,M, \quad \varepsilon_t = (\varepsilon_{1t},...,\varepsilon_{Mt})' \sim Niid(0, \Sigma_\varepsilon) \tag{11}$$

where Σ_ε is a correlation matrix and $h_t = (h_{1t},...,h_{Mt})'$ a multivariate random walk, although more complicated linear dynamics could be handled. The approach again relies on linearising, this time with loss of information, by writing $\log y_{it}^2 = h_{it} + \log \varepsilon_{it}^2$. The vector of $\log \varepsilon_{it}^2$ is iid, all with means -1.27, and a covariance matrix which is a known function of Σ_ε. Consequently Σ_ε and the parameters indexing the dynamics of h_t can be estimated.

It is worthwhile pointing out two aspects of this model. If rank constraints are imposed on h_t, common trends and cycles will be allowed into the process describing the volatility. Furthermore, the model is similar to Bollerslev's (1990) model which is characterised by constant conditional correlation. Hence the model is better defined as one of changing variances rather than of changing correlation. Consequently, it fails to represent important features of the data and so it is of limited interest.

Perhaps a more attractive multivariate SV model can be obtained by introducing factors. The simplest one-factor model is:

$$\begin{cases} y_t = \lambda f_t + w_t, \quad w_t \sim Niid(0, \Sigma_w) \\ f_t = \varepsilon_t \exp(h_t/2), \quad h_t = \beta h_{t-1} + \eta_t \quad \eta_t \sim Niid(0, \sigma_\eta^2) \end{cases}$$

where y_t is perturbed by w_t and explained by the scaled univariate SV model f_t. Typically Σ_w will be assumed diagonal, perhaps driven by independent SV models.

The lack of an obvious linearising transformation for these models prevents us from effectively using Kalman filtering methods. MCMC methods do not suffer this drawback and are explored in Jacquier, Polson and Rossi (1995) and Pitt and Shephard (1999b).

6.2. Multivariate discrete SV models

The multivariate extension of the discrete stochastic volatility model (3) is easy to state. We can consider the multivariate framework (11) and allow each component of $h_t = (h_{1t},...,h_{Mt})'$ to follow a two state Markov chain, i.e.

$$\begin{cases} y_{it} = \varepsilon_{it} \exp(h_{it}/2), \;\; i = 1,...,M, \;\; \varepsilon_t = (\varepsilon_{1t},...,\varepsilon_{Mt})' \sim Niid(0, \Sigma_\varepsilon) \\ h_{it} = \alpha + \beta s_{it} \end{cases}$$

In order to apply the Hamilton filter and to obtain the likelihood function, it is useful to define a new Markov chain S_t with 2^M states, which represents the *M* Markov chains governing the dynamics of h_t.

If the *M* Markov chains are independent, the transition probabilities of S_t are simply obtained by multiplying the probabilities that drive the different Markov chains. Accordingly, the transition probability matrix will be $\mathrm{Q} = \mathrm{P}_1 \otimes \mathrm{P}_2 \otimes ... \otimes \mathrm{P}_\mathrm{M}$, where $\otimes$ indicates the Kronecker product and P_i the transition matrix of s_{it}, i=1,2,…,*M*. In that case, the number of states rises exponentially with the dimension of h_t, but the number of parameters describing the Markov chains grows linearly with *M* and is $2M$.

A more general specification does not make any a priori assumptions about the relations between the different Markov chains. The transition probabilities of the composite Markov chain S_t are then given by:

$$q_{ij} = (S_t = j | S_{t-1} = i), \qquad i, j = 1,2,...,2^M$$

which requires $2^M \left(2^M - 1\right)$ parameters. To understand the dimension of the problem, with *M*=2 (and two states), the independent case requires 4 parameters, while the general specification requires 12 parameters.

Clearly the general specification becomes quickly unfeasible but, in some applications, the independent case is not useful to understand the causality between the volatility of different assets. Billio (2002a) proposes considering several correlated cases with a number of parameters comprised between $2M$ and $2^M \left(2^M - 1\right)$ by exploiting the concept of Granger causality.

As for the continuous SV model, a more interesting multivariate extension can be obtained by introducing a latent factor structure where the latent factors are characterised by discrete stochastic volatility. Unfortunately, in that case the joint process of the observable variable y_t and of the latent Markov chains is no longer Markovian, and then the Hamilton filter no longer works. For the estimation it is thus necessary to use some approximation or to use simulation-based methods (see Billio and Monfort (1998) and Kim and Nelson (1999)).

7. EMPIRICAL APPLICATIONS

To provide simple illustrations of the usefulness of SV models, the two basic models are estimated and their output is used to develop standard option pricing and to calculate the Value-at-Risk of an asset or a portfolio.

7.1. The Volatility program

There do not exist statistical packages to easily and directly estimate[28] SV models and thus the necessary routines have been developed with Ox (version 3.20), a programming language created mainly by Jurgen A. Doornik[29]. These routines can also be used within the package GiveWin.

The files required for running the Volatility program[30] are "volatilitymain.ox", the main program file, "volatility.oxo", a compiled file containing the definition of the functions, and the header file "volatility.h", containing the lists of global variables and functions. In Ox or GiveWin it is sufficient to load the main program file, to select the appropriate options and then to run the program (for the details of the commands and options see the enclosed readme.txt file).

Depending on which commands are commented out (// in front of the command) the program can:

- estimate a basic continuous or discrete SV model on a user provided series;
- simulate a basic continuous or discrete SV model;
- estimate a basic continuous or discrete model on a user provided series and then simulate an alternative path with the estimated parameters.

It shall be stressed that GiveWin is not needed to estimate the models but only to display graphs. This program can easily be used with the freeware version of Ox in conjunction with any text editor, however we recommend the use of OxEdit since it integrates with Ox; both packages can be downloaded from Doornik's website (see footnote 29). All the graphic windows presented in this chapter are taken from GiveWin.

---

[28] Linear state-space models can be estimated with the Kalman filter in EViews, with the GAUSS package FANPAC or the OX package SSFPack (see also STAMP). Thus the linearised version (6) could be estimated with a quasi-maximum likelihood approach. For the switching regime models, see also MSVAR, an Ox package developed by H.M. Krolzig and designed for the econometric modelling of univariate and multiple time series subject to shifts in regime (http://www.economics.ox.ac.uk/research/hendry/krolzig/).

[29] Ox is an object-oriented matrix programming language with a comprehensive mathematical and statistical function library whose major features are speed, extensive library and well-designed syntax, leading to programs which are easy to maintain. In particular, this program takes advantage of the concept of class: it is possible to create new classes based on existing ones and to use their functions, therefore avoiding the need to rewrite them for the new class. In our case, the program is built on the Database class, which is the class designed for handling databases, samples, names of variables, etc… The Database class is used as a starting point for the more specific class of Stochastic Volatility Model: the functions, both to estimate and to simulate the models and to store the results, are totally rewritten, while the functions that manage the time series are part of the Database class.
More information on OX can be found at http://www.nuff.ox.ac.uk/users/doornik/. See also Doornik (2001).

[30] Updated versions of the program will be available for download at the address www.greta.it (under the Working Papers section). The package is free of charge for academic and research purposes. For commercial use, please contact the author (mgobbo@greta.it).

```
#import "volatility"


main()
{
    decl stochobj = new Stochastic();                          // create object

    stochobj.Load("c:/programs/stocvol/series.xls");           // load series
    stochobj.Info();                                           // database info

    stochobj.Select(Y_VAR, { "FTSE100", 0, 0 } );              // variable selection

    stochobj.SetSelSample(-1, 1, -1, 1);                       // full sample
    MaxControl(-100, 100);                                     // maximization control

    stochobj.Estimate(0, 2, <-0.28; 0.54; 0.15>);              // estimate the model

    stochobj.Simulation(0, 0, 1500, 0);                        // simulate the model

    stochobj.SeriesEst("c:/programs/stocvol/FTSE100E.xls"); // save estimated data to file

    stochobj.SeriesSim("c:/programs/stocvol/FTSE100S.xls"); // save simulated data to file

    stochobj.Graph(0, "c:/programs/stocvol/FTSE100.ps");       // save graph to file


    delete stochobj;

}
```

Figure 2: The main program "volatilitymain.ox" loaded with GiveWin with the full list of commands.

The first line of code in Figure 2, just before the "main" command, imports the "volatility.oxo" file, which contains the functions, recalls the Database class and other Ox packages such as the graphic, the probabilistic and the maximisation ones[31].

The program is then organised as follows:

- in a first step the time series of interest is requested in an Excel spreadsheet. The user has to indicate the exact path of the file to be loaded (other data formats can be used, see the readme.txt file for additional details). Data passed to the programs must be the price levels, the necessary transformations are directly carried out by the estimation routines;
- in a second step the model is chosen, estimated and, if desired, simulated;
- finally, the outputs of the model are printed and graphed to the screen and saved.

The available variables in the Excel file "series.xls" are the daily stock indexes analyzed in section 1.1, i.e. the FTSE100, the CAC40 and the MIB30 indexes. In the example developed in this chapter attention will be focused on the modelling of the FTSE100 index.
In the first part of the program, these variables are loaded in Excel format. Thus the full sample of FTSE100 is selected in order to start the analysis and perform the estimation. The command "Estimate" is quite complex and requires inputs by the user: the type of model, the choice of the initialisation of the parameter values and their values if the user wants to define them (see figure 3 and the readme.txt file).

[31] Of course, it is possible to modify the program adding functions belonging to the loaded Database class or to different others. In this case, the class containing the desired functions must be loaded by adding a line of code (#import "…") before the "main" command.

```
main()
{
    ...
    // load data file into object, the example consider Excel file
    // recall that in Ox are available different load commands
    // depending on file type see for further information Ox references
    stochobj.Load("c:/programs/stocvol/series.xls");
    ...
    // select the variable of interest from the database
    // remember to insert the same names as in the database, in the
    // example we consider the FTSE100 series
    stochobj.Select(Y_VAR, { "FTSE100", 0, 0 } );
    ...
    // estimation command, input required:
    // a -> model type 0: ARSV, 1:SRSV
    // b -> starting values 0:random, 1:user, 2:data driven
    // c -> set of initial parameters value, 0:if b=0 or 2,
    //      if b=1 -> 3x1:if a=2, 5x1:if a=1
    // in the example ARSV, user provided starting values
    stochobj.Estimate(0, 2, <-0.28; 0.54; 0.15>);
    ...
    delete stochobj;
}
```

Figure 3: The “Load”, “Select” and “Estimate” commands.

7.1.1. Estimation

The package allows the analysis of the two basic models, i.e. the log-normal SV model (2), called ARSV in the program, which is estimated by quasi maximum likelihood with the Kalman filter, and the two regimes switching model (3), called SRSV, which is estimated by maximum likelihood with the Hamilton filter.

The first model is:

$$\begin{cases} y_t = \mu + \varepsilon_t \exp(h_t/2), \\ h_t = \alpha + \beta h_{t-1} + \eta_t \end{cases}$$

with ε_t and η_t independent Gaussian white noises. Their variances are 1 and σ_η^2, respectively. The volatility equation is characterised by the constant parameter α, the autoregressive parameter β and the variance σ_η^2 of the volatility noise. The mean is either imposed equal to zero or estimated with the empirical mean of the series (see below equation (12)).

Since the specification of the conditional volatility is an autoregressive process of order one, the stationary condition is $|\beta| < 1$. Moreover, the volatility σ_η must be strictly positive. In the estimation procedure the following logistic and logarithm

reparameterisations $\beta = 2\left(\frac{\exp(b)}{1+\exp(b)}\right)-1$, $\sigma_\eta = \exp(s_\eta)$ have been considered in order to satisfy the above constraints.

The second model is a particular specification of the regime switching model introduced by Hamilton. Precisely the distribution of the returns is described by two regimes with same mean but different variances and by constant transition matrix:

$$y_t = \begin{cases} \mu + \sigma_0 \varepsilon_t, & if \;\; s_t = 0 \\ \mu + \sigma_1 \varepsilon_t, & if \;\; s_t = 1 \end{cases} \qquad \mathrm{P} = \begin{bmatrix} p_{00} & 1-p_{11} \\ 1-p_{00} & p_{11} \end{bmatrix}$$

where[32] s_t is a two state Markov chain independent of ε_t, which is a Gaussian white noise with unit variance. The parameters of this model are the mean μ, the low and high standard deviation σ_0, σ_1 and the transition probabilities p_{00}, p_{11} (also called regime persistence probabilities). As for the log-normal SV model, the logarithm and the logistic transformations ensure the positiveness of the volatilities and constraint the transition probabilities to assume values in the (0,1) interval.

Before starting the estimation it is necessary to transform the raw time series, which are expressed in level, in the logarithmic returns[33] and to set the starting values of the parameters in the maximisation algorithm[34]. Moreover, for the log-normal SV model the returns are modified as follows:

$$y_t^* = \log\left(y_t - \bar{y}_t\right)^2 + 1.27 \qquad (12)$$

where $\bar{y}_t$ is the empirical mean. Thus, for the log-normal SV model the mean is not estimated but it is simply set equal to the empirical mean.

While these transformations are automatically done by the estimation procedure, the setting of the starting parameter values requires a choice by the user from the following options:

- random initialisation: a range of possible values of the parameters is fixed, where necessary, and a value is randomly extracted. This method is useful when the user has no idea about the possible value of the parameters but wants to better investigate the parametric space. The drawback of this option is that the optimisation algorithm may be quite time-consuming, because it needs more iterations to converge and the probability that it does not converge to the global maximum increases and then several optimization run (with different random starting values) may be required;

[32] According to the model (3), $\sigma_0 = \exp(\alpha/2)$ and $\sigma_1 = \exp((\alpha+\beta)/2)$.

[33] In the Excel file "series.xls" there are 899 observations of the daily stock indexes analyzed in section 1.1. In the estimation we therefore consider 898 daily return observations.

[34] Recall that data transformations are directly carried out by the programs that require input price levels.

- data driven initialisation: the starting values of the parameters are calculated considering the time series analyzed. For example, the sample mean is used as an approximation of the mean of the switching regime model and the empirical variance multiplied respectively by appropriate factors is used for the high and low variance. This alternative helps the user to speed up the convergence even if he has no opinion on the possible values of the parameters;
- user initialisation: the starting values of the parameters are directly inserted by the user.

In the example, the data driven initialisation has been selected.

During the estimation it is possible to control each step of the algorithm through the command MaxControl (see the readme.txt file for more information). The estimation output is then given by the estimated values of the parameters, their standard errors and relative t-Student statistics[35].

Figure 4 shows the final output of the log-normal SV model for the FTSE100 index. In this example the numerical optimisation ends after 76 iterations, which take 8.32 seconds[36], and the log-likelihood[37] is -1153.7. The volatility of the FTSE100 index is very persistent, in fact the autoregressive coefficient of the volatility equation (β) is equal to 0.956. In practice, for financial time series this coefficient is very often bigger than 0.9.

Figure 5 exemplifies the graphic output, which consists of the estimated volatility for the FTSE100 index along with the historical return series. The estimated volatility is obtained by using the Kalman smoother $\hat{h}_{t/T} = E\left(h_t \middle| Y^{*T}\right)$, which is however not immediately useful. In fact, we are interested in $E\left(\sigma_t \middle| Y^T\right) = E\left(\exp\left(\frac{h_t}{2}\right)\middle| Y^T\right)$, but $E\left(\exp\left(\frac{h_t}{2}\right)\middle| Y^T\right) \neq \exp\left(E\left(\frac{h_t}{2}\middle| Y^T\right)\right)$. Thus, we consider a first-order Taylor expansion of $\exp(h_t/2)$ around $\hat{h}_{t/T}$, and compute the conditional mean and estimate the volatility in the following way:

$$\hat{\sigma}_{t/T} = E\left(\exp\left(\frac{h_t}{2}\right)\middle| Y^T\right) \cong \exp\left(\frac{\hat{h}_{t/T}}{2}\right) + \frac{1}{8}\exp\left(\frac{\hat{h}_{t/T}}{2}\right) Q_{t/T}\,.$$

This computation is performed directly by the program and figure 5 presents the estimated volatility.

---

[35] The standard errors are calculated following Ruiz (1994) for the log-normal SV model and as the inverse of the Information matrix for the switching regime model. In both cases the *z*-statistics asymptotically follow a *N*(0,1) distribution.

[36] On a Pentium III 933 MHz.

[37] For the log-normal SV model the log-likelihood is computed with the Kalman filter for the transformed series y_t^*, see equation (12).

```
---- Database information ----
Sample:    1 - 899 (899 observations)
Frequency: 1
Variables: 3

Variable         #obs   #miss         min         mean         max      std.dev
FTSE100           899       0      3777.1      5879.6      6930.2        623.5
CAC40             899       0      3023.7      5091.7      6922.3       892.33
MIB30             899       0       23564       37752       51093       6472.4

Starting values
parameters
  -0.00037495     0.020098      -5.9420
gradients
      -1628.7      0.56669      0.50753
Initial function =      -8311.86544015

Position after 76 BFGS iterations
Status: Strong convergence
parameters
     -0.39130       3.7872      -3.1959
gradients
  4.5475e-007 -2.2737e-007 -2.2737e-008
function value =      -1153.70477274

Stochastic Volatility Model, version 1.00
Strong convergence

                     parameters value    standard error         z-statistic
costant                      -0.391298          0.195784          -1.99862
AR part                       0.955684          0.0220615          43.3191
standard deviation            0.19665           0.0578519           3.3992
elapsed time 8.32 secs               loglikelihood -1153.7
forecasted volatility 0.0189846
```

Figure 4: Estimation output of the log-normal SV model for the FTSE100 index.

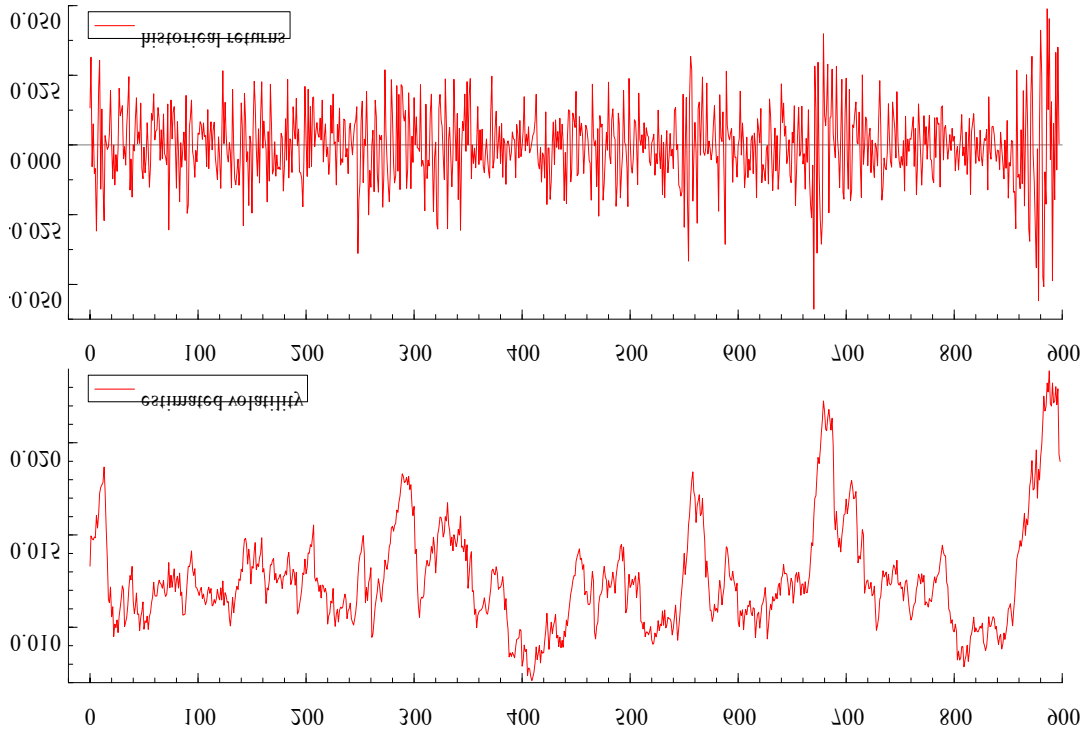

Figure 5: Historical returns and estimated volatility for the FTSE100 index obtained with the log-normal SV model.

Figure 6 shows the final output[38] of the switching regime model for the FTSE100 index. In this case the numerical optimisation ends after 26 iterations, which take 7.65 seconds, and the log-likelihood[39] is -2680.18. For this model we can judge the persistence of the volatility by the value taken by the transition (or persistence) probabilities p_{00}, p_{11}. They are very high (0.99 and 0.96), confirming the high persistence of the volatility of the FTSE100 index. Moreover, the levels of the high and low volatility are perfectly in line with the values of the volatility estimated with the log-normal SV model.

```
---- Database information ----
Sample:    1 - 899 (899 observations)
Frequency: 1
Variables: 3

Variable        #obs  #miss         min        mean         max     std.dev
FTSE100          899      0      3777.1      5879.6      6930.2       623.5
CAC40            899      0      3023.7      5091.7      6922.3      892.33
MIB30            899      0       23564       37752       51093      6472.4

Starting values
parameters
  -0.00036337     -0.69315     -0.69315      -4.6792      -4.0521
gradients
       433.59       6.8030     -0.17620      -19.442      -54.401
Initial function =       2626.32194786

Position after 26 BFGS iterations
Status: Strong convergence
parameters
  -0.00017494       4.9389       3.2807      -4.5279      -3.7532
gradients
  5.3524e-005      0.00000 -4.5475e-008  4.3201e-006 -8.6402e-007
function value =       2680.17860015

Stochastic Volatility Model, version 1.00
Strong convergence

                       parameters value    standard error         z-statistic
mean                       -0.000174935       0.000387296         -0.451684
low persistence prob.          0.992889        0.00435136           228.179
high persistence prob.         0.963762         0.0227076           42.4422
low volatility reg.            0.0108036       0.000399697          27.0294
high volatility reg.           0.0234427        0.00231372           10.132
elapsed time  7.65 secs                  loglikelihood -2680.18
forecasted volatility 0.0196862
```

Figure 6: Estimation output of the switching regime model for the FTSE100 index.

In figure 7 the graphic output of the switching regime model is presented. It consists of the historical return series, the weighted or estimated volatility and the estimated switches between regimes[40].

To estimate the volatility we consider the output of the Kim smoother. Since $\sigma_t = \exp(\alpha/2)(1-s_t) + \exp((\alpha+\beta)/2)s_t = \sigma_0(1-s_t) + \sigma_1 s_t$, we can compute:

[38] It is important to underline that the *z*-statistics for the transition probabilities are not useful for testing $p_{ii} = 0,\ p_{ii} = 1,\ i = 0,1$. In fact, these tests are not standard since they imply testing for the presence of two regimes (see Davies (1977, 1987) and Hansen (1992, 1996)).

[39] In this case the log-likelihood is computed with the Hamilton filter for the return series and thus it is not directly comparable with the log-likelihood of the log-normal SV model.

[40] The regime is 0 if $P\left(h_t = \alpha \middle| Y^T\right) \geq 0.5$ and 1 otherwise.

$$\hat{\sigma}_{t/T} = E\left(\sigma_t \middle| Y^T\right) = \sigma_0 P\left(s_t = 0 \middle| Y^T\right) + \sigma_1 P\left(s_t = 1 \middle| Y^T\right) \qquad (13)$$

where $P\left(s_t = 0 \middle| Y^T\right) = P\left(h_t = \alpha \middle| Y^T\right)$ and $P\left(s_t = 1 \middle| Y^T\right) = P\left(h_t = \alpha + \beta \middle| Y^T\right)$.

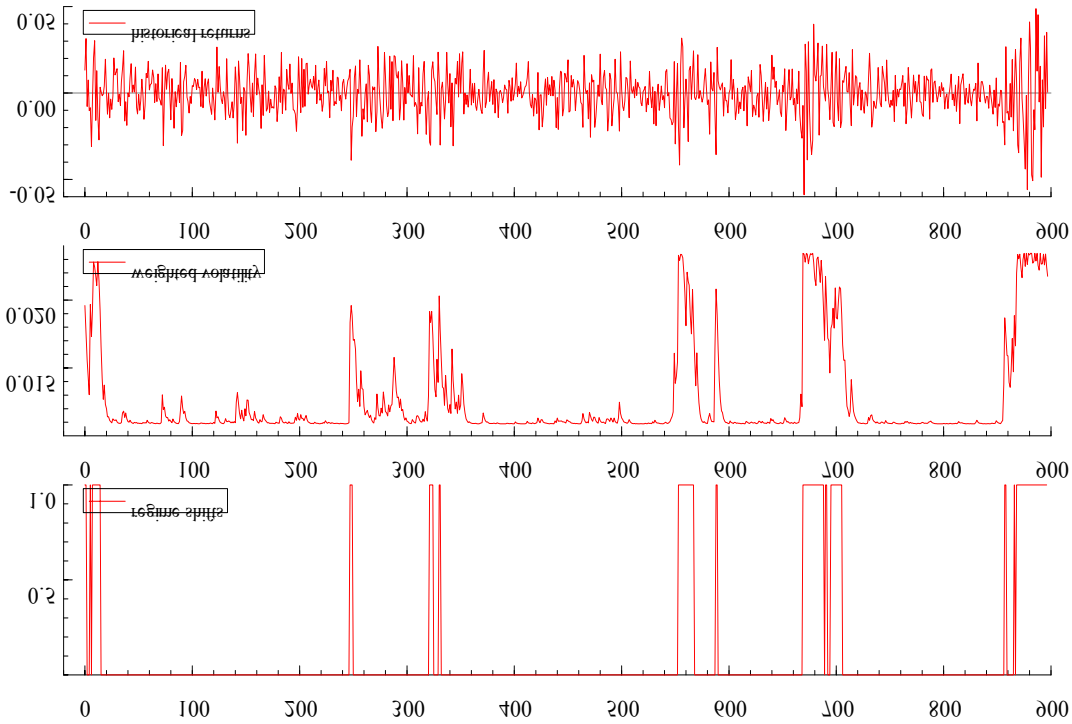

Figure 7: Historical returns, weighted volatility and estimated switches between regimes for the FTSE100 index obtained with the regime switching model.

Finally, it is possible to save the estimated volatility. Since the visualisation of the graphs is possible only with the commercial version of Ox, if this is not available the program allows only the saving of the estimated volatility series. The "SeriesEst" command allows the saving of the following series in an Excel format[41]: historical returns, estimated volatilities and for the switching regime model the smoothed and filtered probabilities of the high volatility regime and the regime shifts. Moreover, the graphs can be directly saved in a postscript format with the "Graph" command. In both cases the user should provide a path including the name and extension of the destination file. The "Graph" command includes also an additional control variable to choose whether or not to plot the series (see figure 8).

[41] In the output file, for the ARSV model, Var1 indicates the historical returns and Var2 the estimated volatilities. For the SRSV, Var1 indicates the historical returns, Var2 the estimated volatilities, Var3 and Var4 the smoothed and filtered probabilities of the high volatility regime and Var5 the regime shifts.

```
...
// simulation command, input required:
// a -> simulate? 0:yes 1:no
// b -> simulated model 0:ARSV, 1:SRSV
// c -> length of simulated series
// d -> parameters for the simulation:
//      3x1 if b=0  ex. <-0.28; 0.54; 0.15>
//      5x1 if b=1  ex. <-0.01; 0.8; 0.8; 0.2; 0.4>
//      if 0 use the last estimated values, check that model type of
//      simulation and estimation correspond
// in the example, simulate ARSV with last estimated parameters
stochobj.Simulation(0, 0, 2500, 0);

// save data to file
// historical returns; estimated volatility;
// conditional probability of high regime; regime shifts.
stochobj.SeriesEst("c:/programs/stocvol/FTSE100E.xls");

// save data to file
// simulated returns; simulated volatility;
// conditional probability of high regime; regime shifts.
stochobj.SeriesSim("c:/programs/stocvol/FTSE100S.xls");

// Graph command, input required
// a -> 0: visualize, 1: don't visualize
// save graph to file, path of the destination
stochobj.Graph(0, "c:/programs/stocvol/graphs.ps");
...
```

Figure 8: The “Simulation”, “SeriesEst”, “SeriesSim” and “Graph” commands.

7.1.2. Simulation

The Volatility program also allows simulation of both the models. The “Simulation” command gives the possibility to choose the type of model, the values of the parameters and the length of the simulated series. If the user wants to simulate the model characterised by the parameters just estimated, the last input of the “Simulation” command must be set to 0, otherwise it has to be replaced by the column vector of the desired parameters.

The graphic output of the simulation is composed by the simulated series and their volatilities. A final possibility is to plot both the estimation and simulation phases (see figures 9 and 10). In particular, for the switching regime model the program plots the simulated volatility, which jumps between the low and high level, and the smoothed (weighted) simulated volatility, which is computed in the same way as the estimated volatility (see equation (13)).

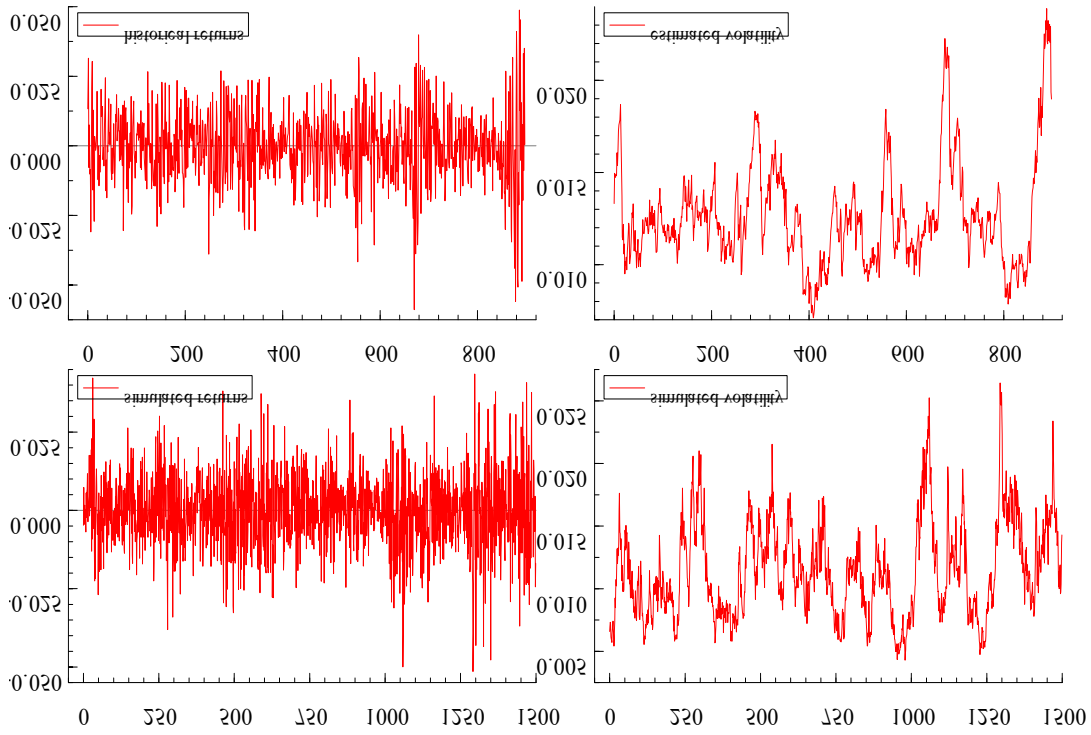

Figure 9: Estimation and simulation graphic output of the log-normal SV model.

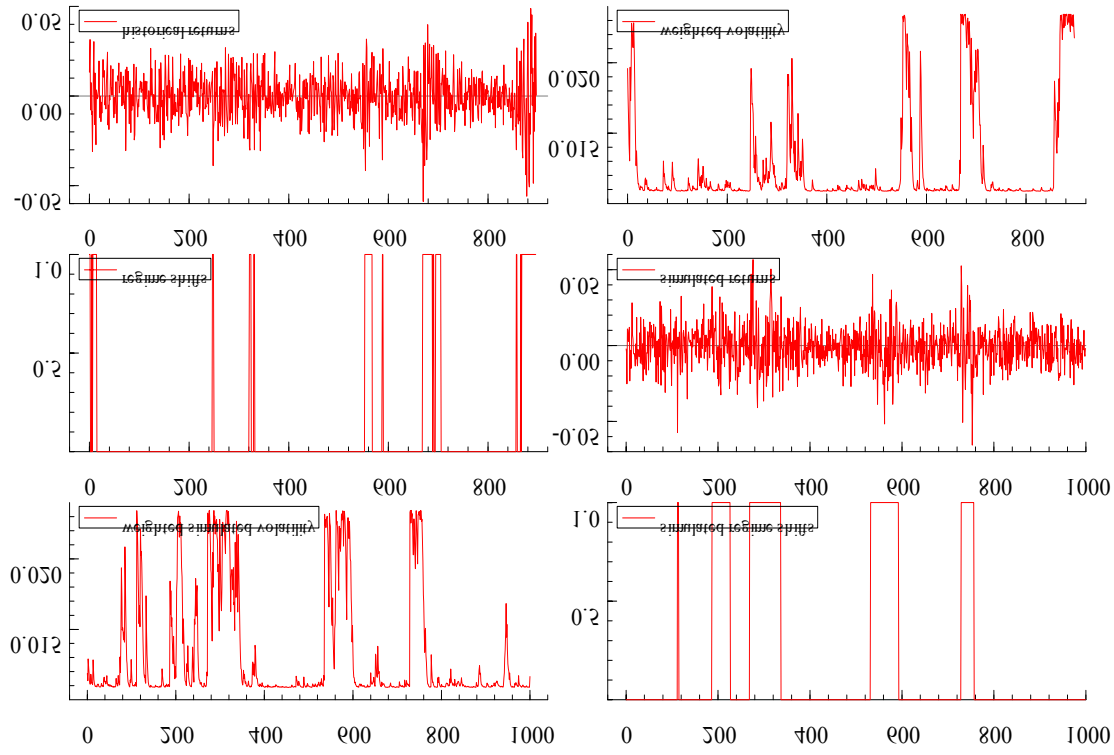

Figure 10: Estimation and simulation graphic output of the switching regime model.

Finally, it is possible to save the simulated volatility. The "SeriesSim" command allows the saving of the following series in an Excel format[42]: simulated returns, simulated volatilities and for the switching regime model the smoothed probabilities of the high volatility regime and the regime shifts.

7.1.3. Forecasting

The final output given by the Volatility program is the forecasted volatility for the following period. The Kalman and Hamilton filters also give the prediction density of h_{t+1} then it is possible to forecast the next value of the volatility.

For the log-normal SV model, we consider a first-order Taylor expansion of $\exp(h_t/2)$ around $\hat{h}_{T+1/T}$ and by taking the conditional expectation we forecast the volatility in the following way:

$$\hat{\sigma}_{T+1/T} = E\left(\exp\left(\frac{h_{T+1}}{2} \right) \middle| Y^T \right) \cong \exp\left(\frac{\hat{h}_{T+1/T}}{2} \right) + \frac{1}{8} \exp\left(\frac{\hat{h}_{T+1/T}}{2} \right) Q_{T+1/T}$$

With regard to the switching regime model, since $\sigma_t = \sigma_0 (1 - s_t) + \sigma_1 s_t$, we can forecast the volatility as follows:

$$\hat{\sigma}_{T+1/T} = E\left(\sigma_{T+1} \middle| Y^T \right) = \sigma_0 P\left(s_{T+1} = 0 \middle| Y^T \right) + \sigma_1 P\left(s_{T+1} = 1 \middle| Y^T \right)$$

where $P\left(s_{T+1} = 0 \middle| Y^T \right) = P\left(h_{T+1} = \alpha \middle| Y^T \right)$, $P\left(s_{T+1} = 1 \middle| Y^T \right) = P\left(h_{T+1} = \alpha + \beta \middle| Y^T \right)$ are the prediction probabilities[43] obtained with the last iteration of the Hamilton filter.

The forecasted volatility is evidenced in the output and it is saved as the last value of the estimated volatility[44].

Let us now consider some practical utilisations of the estimated volatilities.

7.2. Option pricing

As seen in section 1.2, the option price in the Black and Scholes framework can be expressed as a conditional expectation given the current price of the underlying asset:

---

[42] For the ARSV model, Var1 indicates the simulated returns and Var2 the simulated volatilities. For the SRSV, Var1 indicates the simulated returns, Var2 the simulated volatilities, Var3 the smoothed probabilities of the high volatility regime and Var4 the regime shifts.

[43] It is possible to obtain the prediction probabilities by multiplying the transition matrix P by the filtered probabilities (which are saved in the estimation output file as Var4), i.e.

$$\begin{bmatrix} P\left(s_{t+1} = 0 \middle| Y^t \right) \\ P\left(s_{t+1} = 1 \middle| Y^t \right) \end{bmatrix} = \mathrm{P} \begin{bmatrix} P\left(s_t = 0 \middle| Y^t \right) \\ P\left(s_t = 1 \middle| Y^t \right) \end{bmatrix}.$$

[44] In the estimation output file, in the last row the only non-zero value is the forecasted volatility (all the other variables are set equal to zero).

$$C_t^{BS} = \exp(-r\tau) E_{S_{t+\tau}|S_t} \left[\max\left(S_{t+\tau} - K, 0\right)\right]$$

where the dynamic of the asset is described by a geometric diffusion process and the expectation is taken with respect to the risk-neutral probability measure.

Since the Black and Scholes formula can be expressed as a function only of the volatility, a great effort has been made in modelling its behaviour. While Black and Scholes assume that it is constant over the life of the option, a series of models proposed in the late 1980s supposes that it varies through time in a deterministic or stochastic way, in the attempt to capture the empirical features of the option prices. In fact, an analysis of the volatility implied in the market option prices (the so-called implied volatility) highlights that the volatility is neither constant through time nor independent of the strike price (the so-called "smile" and "sneer" effect) (see Rubinstein (1985)).

A very simple approach consists in using the volatility estimated with stochastic volatility models as input of the Black and Scholes formula. In that case, it is sufficient to consider the forecasted volatility $\hat{\sigma}_{t/t-1}$ as the volatility parameter in the formula to obtain the option price:

$$C_t = C_t^{BS}\left(\hat{\sigma}_{t/t-1}\right)$$

In the following we review some stochastic volatility models for the option pricing, which consider the volatility as an exogenous stochastic process.

The path-breaking work on stochastic volatility models applied to option pricing is the paper by Hull and White (1987). The authors assume that both the underlying security S and the variance σ^2 follow a geometric diffusion process:

$$dS = \mu S\, dt + \sigma S\, dz$$
$$d\sigma^2 = \phi\sigma^2 dt + \xi\sigma^2 d\omega$$

where the correlation ρ between the two Brownian motions $dz, d\omega$ is a constant with modulus less than one. Hull and White take $\rho \equiv 0$. Scott (1987) considers the case in which the volatility follows an Ornstein Uhlenbeck process and also imposes the restriction $\rho \equiv 0$. Finally, Heston (1993) proposes the familiar mean reverting square root process for the volatility:

$$dS = \mu S\, dt + \sigma S\, dz$$
$$d\sigma^2 = \gamma\left(\phi - \sigma^2\right)dt + \xi\sigma\, d\omega$$

where ϕ is the long run average variance and he takes the assumption $\rho \neq 0$.

Introducing stochastic volatility in the definition of the stochastic differential equation of the underlying asset creates several complications. A dynamic portfolio with only

one option and one underlying asset is not sufficient to create a riskless investment strategy. The problem arises since the stochastic differential equation for the option contains two sources of uncertainty. Unfortunately, it is impossible to eliminate volatility market risk premium and correlation parameters from the partial differential equation using only one option and one underlying asset. Moreover, these parameters are difficult to estimate[45] and extensive use of numerical techniques is required to solve the two-dimensional partial differential equation.

In the Hull and White (1987) formula, the option price is determined assuming that the volatility market risk premium is zero and there is zero correlation between the two Brownian motions describing the underlying asset and the volatility, i.e. the volatility is uncorrelated with the asset price. With these assumptions and using a risk-neutral valuation procedure, they show that the price of an option with stochastic volatility is the Black and Scholes price integrated over the distribution of the mean volatility:

$$C_t^{HW} = \int C_t^{BS}\left(\overline{\sigma}^2\right) g\left(\overline{\sigma}^2 \mid \sigma_t^2\right) d\overline{\sigma}^2$$

where

$$\overline{\sigma}^2 = \frac{1}{\tau}\int_t^{t+\tau} \sigma^2(u)\,du$$

and $g\left(\overline{\sigma} \mid \sigma_t^2\right)$ is the conditional probability density of the mean variance $\overline{\sigma}^2$ over the period τ.

In the more general case of non-zero correlation, the framework becomes more complex, allowing only numerical solutions.

It can be observed that continuous time stochastic volatility provides an attractive and intuitive explanation for observed volatility patterns and for observed biases in implied volatility. Precisely, smiles, skews, upward and downward term structures of implied volatility arise naturally from a stochastic volatility model. However the fact that stochastic volatility models fit empirical patterns does not mean that those models are correct and the biases in market prices may be the results of other factors, non considered, such as liquidity problems.

A work related to that of Hull and White is Naik (1993). While in the Hull and White specification the volatility follows a continuous diffusion process, Naik analyses the case where the instantaneous variance of an asset is subject to random discontinuous shifts. In particular, the volatility is described with a right-continuous Markov chain process: it remains in the same state for a random amount of time and then shifts to another state with transition probabilities determined by a matrix. In the case of a two-state volatility process the transition matrix is simply:

[45] An exception occurs when the volatility is a deterministic function of the asset price or time. In this case it is possible to easily find a solution to the partial differential equation.

$$\mathrm{P} = \begin{bmatrix} p_{00} & 1-p_{11} \\ 1-p_{00} & p_{11} \end{bmatrix}$$

Assuming that the underlying process is continuous, the risk of a shift in volatility is diversifiable, and therefore not priced, and that the two processes are uncorrelated, the option valuation equation can be expressed in a closed form as follows:

$$C_t^N(\sigma_1) = \int_0^\tau C_t^{BS}\left(\bar{\sigma}^2(x)\right) g(x \mid \sigma_1) dx$$

where σ_1 indicates the high volatility level, $\bar{\sigma}^2(x) = \dfrac{\sigma_1^2 x + \sigma_0^2(\tau - x)}{\tau}$, $0 \le x \le \tau$ and $g(x \mid \sigma_1)$ denotes the unconditional density of the time spent by the volatility process in the high volatility state, given the current high volatility state. In the same way it is possible to determine the option price conditional on the current low volatility state ($C_t^N(\sigma_0)$ with σ_0 the low volatility level).

As in the Hull and White model, the option price is the expectation of the Black and Scholes formula computed for the average future volatility, given the current state. Since two regimes are considered, the final option price can be obtained by a weighted average of the two conditional values $C_t^N(\sigma_1)$ and $C_t^N(\sigma_0)$.

This analysis can be extended by considering multiple states and correlation between changes of the underlying and shifts in volatility, but unfortunately in these cases the option price can be obtained only via numerical methods. In this kind of procedure a discrete time Markov chain is used as an approximation of the volatility process.

We briefly present two examples of this approach, due to Billio and Pelizzon (1997) and Bollen (1998) (see also Bollen, Gray and Whaley (2000)). Both these works are based on the hypothesis that the returns of the underlying asset follow a switching regime model. The distribution is characterised by a mixture of distribution with different variance, where the weights depend on a hidden Markov chain process which represents possible different volatility regimes of the market. To obtain the numerical solution, the basic idea is to approximate the distribution through a multinomial approach, considering a binomial tree for each of the two distributions characterised by different variance.

Following Cox, Ross, Rubinstein (1979), the future value of the underlying process can be expressed as:

$$S_{t+\Delta t} = \begin{cases} uS_t & p \\ dS_t & 1-p \end{cases}$$

One possible specification of the parameters that guaranties asymptotic normality and convergence to the desired mean and variance of the continuously compounded returns is $u = e^{\sqrt{\sigma^2 \Delta t + r^2 \Delta t^2}}$, $d = u^{-1}$ and $p = \frac{e^{r\Delta t} - d}{u - d}$. In this case the process is fully characterised by the parameter representing the variance σ^2, since the step size Δt and the free risk interest rate r are given. Once the variance is estimated it is possible to construct a tree to represent the possible future paths of the variable and hence the distribution of returns (at the maturity).

If a regime switching model is considered, the distribution of the returns is simply a mixture of the distributions characterizing each state. Therefore a discrete process can be used to approximate the continuous time process, and hence the distribution of returns, in each state. In this case, two binomial distributions are used as an approximation of the mixture of the two distributions.

$$S_{t+\Delta t} = \begin{cases} u_1 S_t \mid s_t = 1 & p_{11} \\ u_0 S_t \mid s_t = 0 & p_{00} \\ d_0 S_t \mid s_t = 0 & 1 - p_{00} \\ d_1 S_t \mid s_t = 1 & 1 - p_{11} \end{cases}$$

where s_t denotes the regime.

In this case it is necessary to invoke the risk neutral hypothesis to determine the values of the parameters that are $u_1 = e^{\sqrt{\sigma_1^2 \Delta t + r^2 \Delta t^2}}$, $d_1 = u_1^{-1}$, $p_1 = \frac{e^{r\Delta t} - d_1}{u_1 - d_1}$ for the high volatility regime and $u_0 = e^{\sqrt{\sigma_0^2 \Delta t + r^2 \Delta t^2}}$, $d_0 = u_0^{-1}$ and $p_0 = \frac{e^{r\Delta t} - d_0}{u_0 - d_0}$ for the low volatility regime. This model is usually called quadrinomial tree (or lattice). The inner two branches correspond to the low volatility regime while the outer ones correspond to the high volatility regime. Each set of probabilities $(p_{ii}, 1 - p_{ii})$ must be interpreted as the branch probability conditional on the regime i, with $i = 0, 1$.

Although the quadrinomial tree represents accurately both distributions, its branches do not recombine efficiently, exhibiting an exponential growth of the computational time as the number of steps increases.

Bollen (1998) proposes a method to overcome this problem and develops a pentanomial tree. The definition of the parameters is modified to yield both the possibility to recombine the tree and the convergence of the discrete process to the mixture of distributions. This is obtained by approximating each regime density by a trinomial distribution instead of the binomial one. The modified tree has five evenly spaced branches because the step size of the two regimes are in 1 over 2 ratio

($u_1 = e^{\sqrt{\sigma_1^2 \Delta t + r^2 \Delta t^2}}$, $u_0 = e^{\frac{1}{2}\sqrt{\sigma_1^2 \Delta t + r^2 \Delta t^2}}$) and the middle branch is shared by the two regimes:

$$S_{t+\Delta t} = \begin{cases} u_1 S_t \mid s_t = 1 & p_{1,u} \\ u_0 S_t \mid s_t = 0 & p_{0,u} \\ S_t & p_m \\ d_0 S_t \mid s_t = 0 & p_{0,d} \\ d_1 S_t \mid s_t = 1 & p_{1,d} \end{cases}$$

where $p_{s_t,u}$, $p_{s_t,d}$ are the probabilities to go up and down conditional to the regime s_t and p_m is the probability to remain at the same level price S_t.

Once the tree is generated, the option value is calculated operating backward from the terminal values, i.e. the payoffs, to the valuation time. In the regime switching approaches, for simplicity two conditional option values are calculated at each step, where the conditioning information is the regime at the valuation time t:

$$\begin{cases} C_t(s_t = 0) = \left[p_{00}\, C_{t+1}(s_{t+1} = 0) + (1 - p_{00}) C_{t+1}(s_{t+1} = 1)\right] e^{-r\Delta t} \\ C_t(s_t = 1) = \left[(1 - p_{11}) C_{t+1}(s_{t+1} = 0) + p_{11}\, C_{t+1}(s_{t+1} = 1)\right] e^{-r\Delta t} \end{cases}$$

At the valuation time the value of the option is obtained as a weighted average of the two conditional option values, where the weights depend on the knowledge of the current regime. If the regime is unknown, the unconditional probabilities ($\frac{1 - p_{11}}{2 - p_{00} - p_{11}}, \frac{1 - p_{00}}{2 - p_{00} - p_{11}}$) are used.

Let us take an example of the pentanomial approach by considering the parameters estimated in section 7.1.1.

We deal with a European call option on the FTSE100 index quoted at LIFFE on 22nd August 2000 (when the FTSE100 index quoted 6584.82), with maturity June 2001 and strike price 5900. The risk free interest rate r is approximated with three month Libor[46] and the time to maturity in terms of trading days is 213. In this example $\Delta t = 1$.

Taking the estimated volatilities of section 7.1.1, the parameters of the pentanomial model can be computed as follows:

[46] London Interbank Offered Rate.

$$p_{1,u} = \frac{e^{r} - e^{-2\sqrt{\sigma_0^2 + r^2}}}{e^{2\sqrt{\sigma_0^2 + r^2}} - e^{-2\sqrt{\sigma_0^2 + r^2}}}$$

$$p_{0,u} = \frac{e^{r} - e^{-\sqrt{\sigma_0^2 + r^2}}}{e^{\sqrt{\sigma_0^2 \Delta t + r^2 \Delta t^2}} - e^{-\sqrt{\sigma_0^2 + r^2}}}$$

$$u_1 = e^{2\sqrt{\sigma_0^2 + r^2}}$$

$$u_0 = e^{\sqrt{\sigma_0^2 + r^2}}$$

$$p_m = 1 - 4\left(\frac{\sqrt{\sigma_0^2 + r^2}}{\sqrt{\sigma_0^2 + r^2}}\right)^2$$

$$d_0 = e^{-\sqrt{\sigma_0^2 + r^2}}$$

$$p_{0,d} = 1 - p_{0,u}$$

$$d_1 = e^{-2\sqrt{\sigma_0^2 + r^2}}$$

$$p_{1,d} = 1 - p_{1,u} - p_m$$

Once the tree is generated, the payoffs are calculated with the usual procedure (see appendix A for the details). The conditional option values $C_t(s_t = 0), \quad C_t(s_t = 1)$ are obtained using the estimated transition probabilities, while the final value of the option is a weighted average where the weights are the unconditional probabilities $\frac{1 - p_{11}}{2 - p_{00} - p_{11}}, \frac{1 - p_{00}}{2 - p_{00} - p_{11}}$.

The pentanomial option value is therefore 1055.14, while the market value was 1047.5.

7.3. Value at Risk

VaR is a very intuitive measure to evaluate market risk because it indicates the maximum potential loss at a given level of confidence (a) for a portfolio of financial assets over a specified time horizon (h).

In practice, the value of a portfolio is expressed as a function of K risk factors, $x_\tau = \sum_{i=1}^{N} w_{i,t} P_{i,\tau}(\chi_{1,\tau}, \ldots, \chi_{K,\tau})$. The factors influencing the portfolio value are usually identified with some market variables such as interest rates, exchange rates or stock indexes. If their distribution is known in a closed form, we need to estimate the distribution of the future value of the portfolio conditional to the available information and the VaR is then the solution to:

$$a = \int_{-\infty}^{VaR(h,a)} f(x_{t+h}) dx$$

Different parametric models can be used to forecast portfolio return distribution. The simple way to calculate VaR involves assuming that the risk factor returns follow a multivariate normal distribution conditional to the available information. If the portfolio return is linearly dependent on them, its probability distribution is also

normal and the VaR is simply the quantile of this analytic distribution. If the linear assumption is inappropriate, the portfolio return can be approximated as a quadratic function of the risk factor returns.

An alternative way to handle the non-linearity is to use Monte Carlo simulation. The idea is to simulate repeatedly the random processes governing the risk factors. Each simulation gives us a possible value for the portfolio at the end of our time horizon. If enough of these simulations are considered, it is possible to infer the VaR, as the relevant quantile of the simulated distribution.

Since market risk factors have usually fatter tails than the normal distribution, it is also possible to use historical simulation rather than a parametric approach. The idea behind this technique is to use the historical distribution of returns to the assets in the portfolio to simulate the portfolio's VaR, on the hypothetical assumption that we held this portfolio over the period of time covered by our historical data set. Thus, the historical simulation involves collecting historic asset returns over some observation period and using the weights of the current portfolio to simulate the hypothetical returns we would have had if we had held our current portfolio over the observation period. It is then assumed that this historical distribution of returns is also a good proxy for the portfolio return distribution it will face over the next holding period and VaR is calculated as the relevant quantile of this distribution.

The advantage of the parametric approach is that the factors variance covariance matrix can be updated using a general model of changing or stochastic volatility. The main disadvantage is that the factor returns are usually assumed to be conditionally normal, losing the possibility to take into account nonlinear correlations among them. Historical simulation has the advantage of reflecting the historical multivariate probability distribution of the risk factor returns, avoiding ad hoc assumptions. However the method suffers a serious drawback. Its main disadvantage is that it does not incorporate volatility updating. Moreover extreme quantiles are difficult to estimate, as extrapolation beyond past observations is impossible. Finally, quantile estimates tend to be very volatile whenever a large observation enters the sample and the database is not sufficiently large.

The advantage of the parametric approach to update the volatility suggests the simplest utilisation of the SV models for the VaR computation. Chosen the asset or portfolio distribution (usually the normal one), it is possible to use the forecasted volatility to characterise the future return distribution. Thus, $\hat{\sigma}_{T+1/T}$ can be used to calculate the VaR over the next period.

A different approach using the SV model is to devolatilise the observed returns series and to revolatilise it with an appropriate forecasted value, obtained with a particular model of changing volatility. This approach is considered in several recent works (Barone-Adesi, Burgoin and Giannopoulos (1998), Hull and White (1998)) and is a way of combining different methods and partially overcoming the drawbacks of each.

To make the historical simulation consistent with empirical findings, the log-normal SV model and the regime switching model may be considered to describe the volatility behaviour. Past returns are standardised by the estimated volatility to obtain standardised residuals. Statistical tests can confirm that these standardised residuals

behave approximately as an iid series which exhibits heavy tails. Historical simulation can then be used. Finally, to adjust them to the current market conditions, the randomly selected standardised residuals are multiplied by the forecasted volatility obtained with the SV model.

For example, table 2 shows the results obtained with the FTSE100 index return by considering 1,000,000 historical simulations[47] (see appendix B for the details).

Confidence level	Log-normal SV model	Regime switching model
0.1	2.5442	2.7503
0.05	3.9298	3.5874
0.01	5.3417	4.6502

Table 2: VaR at different confidence levels for the FTSE100 index return.

Clearly, this approach allows a wide range of stochastic and changing volatility models, such as ARCH-GARCH models, to be considered. Moreover, it must be pointed out that instead of using historical simulation, an appropriate standard distribution can also be considered to model the transformed returns and then several probability distributions can be assumed for the unconditional returns (McNeil and Frey (2000), Eberlein, Kallsen and Kristen (2001)).

Another example of parametric approach to VaR calculation considers the hypothesis that a regime switching model governs the asset returns (Billio and Pelizzon 2000):

$$y_t = \begin{cases} \mu + \sigma_0 \varepsilon_t & s_t = 0 \\ \mu + \sigma_1 \varepsilon_t & s_t = 1 \end{cases}$$

where $\varepsilon_t \sim N(0,1)$ is independent of s_t.

To calculate the VaR it is necessary to determine the value of the conditional distribution for which the cumulative density is *a*, i.e.

$$a = \sum_{s_{t+h}=0,1} P(s_{t+h} \mid I_t) \int_{-\infty}^{VaR(h,a)} f_{s_{t+h}}(y_{t+h} \mid I_t) dy$$

where $f_{s_{t+h}}(y_{t+h} \mid I_t)$ is the probability density of y_{t+h} when the regime is s_{t+h} and conditional to the available information set I_t (usually containing past returns), $P(s_{t+h} \mid I_t)$ is the prediction probability obtained by the Hamilton filter.

Given the parameters estimated in section 7.1.1 for the switching regime model and the prediction probabilities at time *t+h* (obtained by the product of the transition

[47] Each operation takes about 12 seconds.

matrix, for h-1 steps, and the conditional probabilities $P(s_{t+1}|I_t)$ given by the Hamilton filter), the VaR is the relevant quantile of the mixture distribution.

The model can be generalised to the case of N risky assets providing an explicit link between the return on the asset and the return on the market index, thus by explicitly modelling the correlation between different assets. The Multivariate Switching Regime Beta Model (Billio and Pelizzon (2002)) is a sort of market model or better a single factor model in the Arbitrage Pricing Theory framework where the return of a single stock i is characterized by the regime switching of the market index and the regime switching of the specific risk of the asset. It can be written as:

$$\begin{cases} y_{mt} = \mu_m(s_t) + \sigma_m(s_t)\varepsilon_t & \varepsilon_t \sim IIN(0,1) \\ y_{1t} = \mu_1(s_{1t}) + \beta_1(s_t, s_{1t})y_{mt} + \sigma_1(s_{1t})\varepsilon_{1t} & \varepsilon_{1t} \sim IIN(0,1) \\ y_{2t} = \mu_2(s_{2t}) + \beta_2(s_t, s_{2t})y_{mt} + \sigma_2(s_{2t})\varepsilon_{2t} & \varepsilon_{2t} \sim IIN(0,1) \\ \vdots & \\ y_{Nt} = \mu_N(s_{Nt}) + \beta_N(s_t, s_{Nt})y_{mt} + \sigma_N(s_{Nt})\varepsilon_{Nt} & \varepsilon_{Nt} \sim IIN(0,1) \end{cases}$$

where y_{mt} is the market return, the market regime variable s_t and the single stock regime variables s_{jt}, $j = 1,..,N$ are independent Markov chains, ε_t and ε_{jt}, $j = 1,..,N$, are independently distributed.

Using this approach it is possible to take into account the correlation between different assets. In fact, the variance-covariance matrix between the two assets i and j is:

$$\Sigma(s_t, s_{it}, s_{jt}) = \begin{bmatrix} \beta_i^2(s_t, s_{it})\sigma_m^2(s_t) + \sigma_i^2(s_{it}) & \beta_i(s_t, s_{it})\beta_j(s_t, s_{jt})\sigma_m^2(s_t) \\ \beta_i(s_t, s_{it})\beta_j(s_t, s_{jt})\sigma_m^2(s_t) & \beta_j^2(s_t, s_{jt})\sigma_m^2(s_t) + \sigma_j^2(s_{jt}) \end{bmatrix}$$

then the correlation between different assets depends on the extent to which each asset is linked, through the factor loading β, to the market index.

To calculate VaR for a portfolio based on N assets it is sufficient to use the approach presented above. In particular, considering two assets and assuming that the number of regimes is 2 for all three Markov chains we have:

$$a = \sum_{s_{t+h}=0,1} \sum_{s_{i,t+h}=0,1} \sum_{s_{j,t+h}=0,1} P(s_{t+h}, s_{i,t+h}, s_{j,t+h} \mid I_t) \int_{-\infty}^{VaR(h,a)} f_{s_{t+h}, s_{i,t+h}, s_{j,t+h}}(y \mid I_t) dy$$

where $f_{s_{t+h}, s_{i,t+h}, s_{j,t+h}}(y \mid I_t)$ is the probability density of the portfolio return y when the regimes are $s_{t+h}, s_{i,t+h}, s_{j,t+h}$ and conditional to the available information set I_t. This distribution has mean $\mathbf{w}'\boldsymbol{\mu}(s_{t+h}, s_{i,t+h}, s_{j,t+h})$ and variance

$\mathbf{w}'\Sigma(s_{t+h}, s_{i,t+h}, s_{j,t+h})\mathbf{w}$ where $\mathbf{w}$ is the vector of the percentage of wealth invested in the two assets and $\boldsymbol{\mu}(s_{t+h}, s_{i,t+h}, s_{j,t+h})$ is the vector of risky asset mean returns, i.e.

$$\boldsymbol{\mu}(s_{t+h}, s_{i,t+h}, s_{j,t+h}) = \begin{cases} \mu_i(s_{i,t+h}) + \beta_i(s_{t+h}, s_{i,t+h})\mu_m(s_{t+h}) \\ \mu_j(s_{j,t+h}) + \beta_j(s_{t+h}, s_{j,t+h})\mu_m(s_{t+h}) \end{cases}$$

The drawback of this approach is that it requires the estimation of a number of parameters that grows exponentially with the number of the assets. In fact, the number of possible regimes generated by this model is 2^{N+1}.

One possible solution is to consider the idiosyncratic risk distributed as $IIN(0, \sigma_i^2)$ (without a specific Markov chain dependency) and to characterize the systematic risk with more than one source of risk. This approach is in line with the Arbitrage Pricing Theory model where the risky factors are characterized by switching regime processes. Formally, we can write this model as:

$$\begin{cases} F_{jt} = \alpha_j(s_{jt}) + \theta_j(s_{jt})\varepsilon_{jt}, & \varepsilon_{jt} \sim IIN(0,1),\ j = 1,2,...,K \\ y_{it} = \mu_i + \sum_{j=1}^{K} \beta_{ij}(s_{jt})F_{jt} + \sigma_i \varepsilon_{it}, & \varepsilon_{it} \sim IIN(0,1),\ i = 1,2,...,N \end{cases}$$

where F_{jt} is the value of factor j at time t, $(j = 1,2..,K)$, $\beta_{ij}(s_{jt})$ is the factor loading of the asset i on factor j, s_{jt}, $j = 1,2..,K$, are independent Markov chains, and $\varepsilon_{jt}, j = 1,2..,K$, and ε_{jt}, $j = 1,2..,N$, are independently distributed.

This model is more parsimonious, in fact the introduction of an extra asset implies that only $K + 2$ parameters need to be estimated. This approach is valid when the number of assets in the portfolio is high and the specific risk is easily eliminated by diversification.

8. CONCLUDING REMARKS

We have tried to develop an introduction to the current literature on stochastic volatility models. Other than the classical log-normal model introduced by Taylor (1986), we have also presented the discrete volatility model in which the latent stochastic structure of the volatility is described by a Markov chain.

Both models (with continuous and discrete volatility) fit in the framework of a non linear and non Gaussian state-space model, thus the estimation and smoothing problems are developed along the same lines. Only for the discrete case does the general algorithm introduced allow us to compute the likelihood function and then to obtain maximum likelihood estimates. In the continuous case, approximations or simulations must be introduced.

Some extensions and multivariate models are also presented, however there is still a great deal of work to be done.

Finally, the estimation program presented considers the two basic models and allows an estimation of the latent volatility. Some possible applications are suggested and discussed.

APPENDIX A: APPLICATION OF THE PENTANOMIAL MODEL

The example considers a European call option on the FTSE100 index, with maturity June 2001 and strike price 5900 traded at LIFFE on 22nd August 2000. Three month Libor is used as an approximation of the free risk interest rate r. The FTSE100 index quoted 6584.82 and Libor was 6.22%.

The FTSE100 index being a weighted average of the prices of 100 stocks, the dividend effect must be considered. In the example, this parameter is considered constant[48] and equal to $q = 3\%$. The time to maturity in terms of trading days is 213 and the step size is the single trading day ($\Delta t = 1$).

Taking the estimated volatilities of section 7.1.1 ($\hat{\sigma}_0 = 0.0108036$ and $\hat{\sigma}_1 = 0.0234427$), the parameters of the pentanomial model can be computed as follows:

$$u_1 = e^{\sqrt{\hat{\sigma}_1^2 + (r-q)^2}} = 1.02372$$

$$u_0 = e^{\frac{1}{2}\sqrt{\hat{\sigma}_1^2 + (r-q)^2}} = 1.01179$$

$$d_0 = e^{-\frac{1}{2}\sqrt{\hat{\sigma}_1^2 + (r-q)^2}} = 0.98835$$

$$d_1 = e^{-\sqrt{\hat{\sigma}_1^2 + (r-q)^2}} = 0.97683$$

$$p_{1,u} = \frac{e^{r-q} - e^{-\sqrt{\hat{\sigma}_1^2 + (r-q)^2}}}{e^{\sqrt{\hat{\sigma}_1^2 + (r-q)^2}} - e^{-\frac{1}{2}\sqrt{\hat{\sigma}_1^2 + (r-q)^2}}} = 0.49682926$$

$$p_{0,u} = \frac{e^{r-q} - e^{-\frac{1}{2}\sqrt{\hat{\sigma}_1^2 + (r-q)^2}}}{e^{\frac{1}{2}\sqrt{\hat{\sigma}_1^2 + (r-q)^2}} - e^{-\frac{1}{2}\sqrt{\hat{\sigma}_1^2 + (r-q)^2}}} = 0.42770369$$

$$p_m = 1 - 4\left(\frac{\sqrt{\hat{\sigma}_0^2 + (r-q)^2}}{\sqrt{\hat{\sigma}_1^2 + (r-q)^2}}\right)^2 = 0.15037295$$

$$p_{0,d} = 1 - p_{0,u} - p_m = 0.42192336$$

$$p_{1,d} = 1 - p_{1,u} = 0.50317074$$

Given u_1, u_0, d_0, d_1, the possible one step ahead values for the FTSE100 index can be calculated as follows:

[48] During the period of analysis the dividend yield was nearly constant. For details, see http://www.londonstockexchange.com.

$$
\begin{array}{rl}
 & \rightarrow u_1 S_t = 6741.01 \\
 & \rightarrow u_0 S_t = 6662.45 \\
S_{t+1} & \rightarrow \quad S_t = 6584.82 \\
 & \rightarrow d_0 S_t = 6508.11 \\
 & \rightarrow d_1 S_t = 6432.25
\end{array}
$$

and the tree is recursively generated.

The payoffs at maturity are calculated with the usual formula:

$$C_T = \max\left(S_T - K, 0\right)$$

and the values of the option in the previous periods are obtained operating backwards. In particular:

- at time T-1, the conditional values of the option at the i^{th} node, $C_{T-1}^{i}\left(s_{T-1}=0\right)$, $C_{T-1}^{i}\left(s_{T-1}=1\right)$, are given by

$$
\begin{cases}
C_{T-1}^{i}\left(s_{T-1}=0\right) = \left[\left(1-p_{0,u}-p_m\right)C_T^{i-1} + p_m\, C_T^{i} + p_{0,u} C_T^{i+1}\right]e^{-r} \\
C_{T-1}^{i}\left(s_{T-1}=1\right) = \left[\left(1-p_{1,u}\right)C_T^{i-2} + p_{1,u}\, C_T^{i+2}\right]e^{-r}
\end{cases}
$$

  Note that the nodes ± 1 consider the option values obtained with u_0 and d_0, while the nodes ± 2 consider u_1 and d_1. The calculation is repeated for all the nodes at time T-1 and we obtain two sets of conditional option values $C_{T-1}^{i}\left(s_{T-1}=0\right)$, $C_{T-1}^{i}\left(s_{T-1}=1\right)$.

- at time T-2,

– for each node i, the values of the option are obtained conditional on the regime in T-1:

$$
\begin{cases}
C_{T-2/T-1}^{i}\left(s_{T-1}=0\right) = \left[\left(1-p_{0,u}-p_m\right)C_{T-1}^{i-1}\left(s_{T-1}=0\right) + p_m\, C_{T-1}^{i}\left(s_{T-1}=0\right) + p_{0,u} C_{T-1}^{i+1}\left(s_{T-1}=0\right)\right] \\
C_{T-2/T-1}^{i}\left(s_{T-1}=1\right) = \left[\left(1-p_{1,u}\right)C_{T-1}^{i-2}\left(s_{T-1}=1\right) + p_{1,u}\, C_{T-1}^{i+2}\left(s_{T-1}=1\right)\right]
\end{cases}
$$

– using the estimated transition probabilities ($\hat{p}_{00}$= 0.992889, $\hat{p}_{11}$=0.963762), they are then discounted considering the possibility that a switch occurs between time T-1 and T-2:

$$
\begin{cases}
C_{T-2}^{i}\left(s_{T-2}=0\right) = \left[\hat{p}_{00}\, C_{T-2/T-1}^{i}\left(s_{T-1}=0\right) + \left(1-\hat{p}_{00}\right)C_{T-2/T-1}^{i}\left(s_{T-1}=1\right)\right]e^{-r} \\
C_{T-2}^{i}\left(s_{T-2}=1\right) = \left[\left(1-\hat{p}_{11}\right)C_{T-2/T-1}^{i}\left(s_{T-1}=0\right) + \hat{p}_{11}\, C_{T-2/T-1}^{i}\left(s_{T-1}=1\right)\right]e^{-r}
\end{cases}
$$

Again we obtain two sets of conditional option values $C_{T-2}^{i}(s_{T-2}=0)$, $C_{T-2}^{i}(s_{T-2}=1)$.

- This computation is iterated for *T*-3, *T*-4,…, *t*.

At the evaluation time *t*, we obtain two conditional values $C_t(s_t=0)$, $C_t(s_t=1)$, which are respectively 1047.61 and 1093.54. Finally, the value of the option is calculated as a weighted average of these two values, where the weights depend on the knowledge on the current regime. If the regime is unknown, the estimated unconditional probabilities $\hat{p}_0=\frac{1-\hat{p}_{11}}{2-\hat{p}_{00}-\hat{p}_{11}}$, $\hat{p}_1=\frac{1-\hat{p}_{00}}{2-\hat{p}_{00}-\hat{p}_{11}}$ can be used.

The pentanomial option value is therefore 1055.14 while the market value was 1047.5. Another possibility, probably better from a methodological point of view as it uses all the available information, is to consider the filtered probabilities obtained as output of the estimation step, i.e. $P\left(s_t=0\middle|Y^t\right)$=0.98636043 and $P\left(s_t=1\middle|Y^t\right)$=0,01363957. In that case, the pentanomial option value is 1048.24.

As an exercise, readers may wish to replicate the following examples:
1) European put option quoted on 23/10/2000, strike price: 7000, FTSE100: 6315.9, Libor: 6.1475%, days to maturity: 39, option price: 700.
Results: Conditional option values: 756.15 and 679.89. Option value considering unconditional probabilities: 692.4.
2) European put option quoted on 03/11/2000, strike price: 5500, FTSE100: 6385.44, Libor: 6.12625%, days to maturity: 160, option price: 117.
Results: Conditional option values: 143.29 and 101.08. Option value considering unconditional probabilities: 108.

APPENDIX B: APPLICATION TO VALUE AT RISK

Consider a portfolio which perfectly replicates the composition of the FTSE100 index. Given the estimated volatility of the stochastic volatility models, the VaR of this portfolio can be obtained following the procedure proposed in Barone-Adesi, Burgoin and Giannopoulos (1998).

The historical portfolio returns are rescaled by the estimated volatility series to obtain the standardised residuals $u_t=\frac{y_t}{\sigma_t}$, *t*=1,…,*T* (in our case *T*=898, see footnote 33).
The historical simulation can be performed by bootstrapping the standardised returns to obtain the desired number of residuals $u_j^*,\ j=1,...,M$, where *M* can be arbitrarily large. To calculate the next period returns, it is sufficient to multiply the simulated residuals by the forecasted volatility $\hat{\sigma}_{T+1/T}$:

$$y_j^* = u_j^* \hat{\sigma}_{T+1/T} \cdot$$

The VaR for the next day, at the desired level of confidence h, is then calculated as the Mh^{th} element of these returns sorted in ascending order.

References

Alexander, C. (1996), *The Handbook of Risk Management and Analysis*. New York, Toronto, Singapore, John Wiley & Sons.

Andersen, T.G. and T. Bollerslev (1997), 'Heterogeneous information arrivals and return volatility dynamics: Uncovering the long run in high frequency returns', *Journal of Finance*, 52, 975-1005.

Andersen, T.G., H. Chung and B.E. Sørensen B. (1999), 'Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study', *Journal of Econometrics*, 91, 61–87.

Andersen, T.G. and B.E. Sørensen (1996), 'GMM estimation of a stochastic volatility model: A Monte Carlo study', *Journal of Business and Economic Statistics*, 14, 328–352.

Andersen, T.G. and B.E. Sørensen (1997), 'GMM and QML asymptotic standard deviations in stochastic volatility models: A comment on Ruiz (1994)', *Journal of Econometrics*, 76, 397–403.

Anderson, B. and J. Moore (1979), *Optimal Filtering*, Prentice Hall, New York.

Ball, C. and W. Torous (1985), 'On jumps in common stock prices and their impact on call option pricing', *Journal of Finance*, 40, 155-173.

Barone-Adesi, G., F. Burgoin and K. Giannopoulos (1998), 'Don't look back', *Risk,* 11, 100-104.

Bates, D.S. (1996), 'Jumps and stochastic volatility: exchange rate processes implicit in Deutsche mark options', *The Review of Financial Studies*, 9, 69-107.

Bera, A. and M. Higgins (1993), 'Arch models: properties, estimation and testing', *Journal of Economic Surveys*, 7, 305–366.

Best, P. (1998), *Implementing Value at Risk*. New York, Toronto, Singapore, John Wiley & Sons.

Billio, M. (1999), Simulation based methods for non linear state space models, Ph.D. thesis, Paris Dauphine.

Billio, M. (2002a), 'Correlated Markov chain', *WP* GRETA, Venice.

Billio, M. (2002b), 'Simulation based methods for financial time series', *Atti della XLI Riunione Scientifica of Italian Statistical Society*, Università di Milano-Bicocca, 5-7 June.

Billio, M. and A. Monfort (1998), 'Switching state space models: likelihood, filtering and smoothing', *Journal of Statistical Planning and Inference*, 68/1, 65-103.

Billio, M. and L. Pelizzon (1997), 'Pricing options with switching volatility', in *Money, Finance, Banking and Insurance,* ed. C. Hipp, Verlang.

Billio, M. and L. Pelizzon (2000), 'Value-at-Risk: a multivariate switching regime approach', *Journal of Empirical Finance*, 7, 531-554.

Billio, M., A. Monfort A. and C.P. Robert (1998), 'The simulated likelihood ratio (SLR) method', *Document de Travail du CREST* 9828, Paris.

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

Bollen, N. (1998), 'Valuating option in regime switching models', *Journal of Derivatives*, 6, 38-49.

Bollen, N., S. Gray and R. Whaley R. (2000), 'Regime-Switching in Foreign Exchange Rates: Evidence from Currency Options', *Journal of Econometrics*, 94, 239-276.

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. and D. Jubinski (1999), Equity trading volume and volatility: Latent information arrivals and common long-run dependencies', *Journal of Business and Economic Statistics*, 17, 9-21.

Bollerslev, T. and H.O. Mikkelsen (1999), 'Long-term equity anticipation securities and stock market volatility dynamics', *Journal of Econometrics*, 92, 75-99.

Bollerslev, T. and J. Wright (2000), 'Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data', *Journal of Econometrics*, 98, 81-106.

Bollerslev, T., R. Chow and K. Kroner (1992), 'Arch modelling in finance: a review of the theory and empirical evidence', *Journal of Econometrics*, 52, 5–59.

Bollerslev, T., R.F. Engle and D. Nelson (1993), 'Arch models', in: *Handbook of Econometrics*, Engle R. and McFadden D., eds., North-Holland, Amsterdam, volume IV.

Breidt, F.J., N. Crato and P. de Lima (1998), 'On the detection and estimation of long memory in stochastic volatility', *Journal of Econometrics*, 83, 325-348.

Carraro C. and D. Sartore (1987), 'Square Root Iterative Kalman Filter: Theory and Applications to Regression Models', *Annales d'Economie et de Statistique*, 6/7.

Chan, N.H. and G. Petris (1999), 'Bayesian analysis of long memory stochastic volatility models', Technical report Department of Statistics, Carnegie Mellon University.

Chib, S., F. Nardari and N. Shephard (2002), 'Markov Chain Monte Carlo methods for stochastic volatility models', *Journal of Econometrics*, forthcoming.

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', *Scandinavian Journal of Statistics,* 8, 93-115.

Cox, J.C., S. A. Ross and M. Rubinsten (1979), 'Option pricing: A simplified approach', *Journal of Financial Economics,* 7, 229-263.

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

Danielsson, J. and J. Richard (1993), 'Accelerated Gaussian importance sampler with application to dynamic latent variable models', *Journal of Applied Econometrics*, 8, 153–173.

Davies, R.B. (1977), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika*, 64, 247-54.

Davies, R.B. (1987), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika*, 74, 33-43.

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

Deo, R.S. and C.M. Hurvich (2001), 'On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models', *Econometric Theory*, 17, 686-710.

Doornik, J.A. (2001), *Ox: An Object-Oriented Matrix Language* (4th edition), London: Timberlake Consultants Press.

Dridi, R. (2000), 'Simulated asymptotic least squares theory', Working Paper 396, London School of Economics - Suntory Toyota, Econometrics.

Dridi, R. and E. Renault (2000), 'Semi-parametric indirect inference', Working Paper 392, London School of Economics - Suntory Toyota, Econometrics.

Duffie, D., J. Pan and K. Singleton (2000), 'Transform analysis and asset pricing for affine jump-diffusions', *Econometrica,* 68, 1343-1376.

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

Durbin, J. and S. Koopman (1997), 'Monte Carlo maximum likelihood estimation for non Gaussian state space models', *Biometrika*, 84, 669–684.

Eberlein, E. (2001), 'Application of generalized hyperbolic Lévy motion to finance', in *Lévy Processes - Theory and Applications*, Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S. (eds.), Birkhauser, Boston, pp. 319-37.

Eberlein, E. and K. Prause (2001), 'The generalized hyperbolic model: financial derivatives and risk measures', in *Proceedings of the Year 2000 Congress of the European Mathematical Society*.

Eberlein, E., J. Kallsen and J. Kristen (2001), 'Risk management based on stochastic volatility', FDM Preprint 72, University of Freiburg.

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

Engle, R.F. and G.G.J. Lee (1992), 'A permanent and transitory component model of stock return volatility', UCSD Economics Discussion Papers 92-44, 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.

Francq, C. and J.M. Zakoïan (2001), 'Stationarity of Multivariate Markov-switching ARMA Models', *Journal of Econometrics,* 102, 339-364.

Francq, C., M. Roussignol and J.M. Zakoïan (2001), 'Conditional Heteroskedasticity Driven by Hidden Markov Chains'*, Journal of Time Series Analysis,* 22/2, 197–220.

Fridman, M. and L. Harris (1998), 'A maximum likelihood approach for non-Gaussian stochastic volatility models', *Journal of Business and Economic Statistics*, 16, 284–291.

Gallant, A. and G. Tauchen (1996), 'Which moments to match?', *Econometric Theory*, 12, 657–681.

Gallant, A. and J. Long (1997), 'Estimating stochastic differential equations efficiently by minimum chi-square', *Biometrika*, 84, 125–141.

Gallant, A., D. Hsieh and G. Tauchen (1997), 'Estimation of stochastic volatility models with diagnostics', *Journal of Econometrics*, 81, 159–192.

Gelfand, A. and A. Smith (1990), 'Sampling based approaches to calculating marginal densities', *Journal of the American Statistical Association*, 85, 398–409.

Geweke, J. and S. Porter-Hudak (1983), 'The estimation and application of long-memory time series models', *Journal of Time Series Analysis,* 4, 221-238.

Geyer, C. (1994), 'On the convergence of Monte Carlo maximum likelihood calculations', *Journal of the Royal Statistical Society*, B 65, 261–274.

Geyer, C. (1996), 'Estimation and optimization of functions', in: *Markov Chain Monte Carlo in Practice*, Gilks W., Richardson S. and Spiegelhalter D., eds., Chapman and Hall, London, 241–258.

Ghysels, E., A. Harvey and E. Renault (1996), 'Stochastic volatility', in: *Statistical Methods in Finance*, Rao C. and Maddala G., eds., North-Holland, Amsterdam.

Gordon, N., D. Salmond and A. Smith (1993), 'A novel approach to non-linear and non Gaussian Bayesian state estimation', *IEE Proceedings*, F140, 107–133.

Gouriéroux, C., A. Monfort and E. Renault (1993), 'Indirect inference', *Journal of Applied Econometrics*, 8, 85–118.

Hamilton, J.D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, 57/2, 357-384.

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

Hansen, B.E. (1992), 'The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP', *Journal of Applied Econometrics*, 7, S61-S82.

Hansen, B. E. (1996), 'Erratum: The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP', *Journal of Applied Econometrics*, 11/2, 195-198.

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

Harvey, A.C. (1998), 'Long-memory in stochastic volatility', in *Forecasting Volatility in the Financial Market*, J. Knight, and S. Sachell, eds., Butterworth-Heineman, London.

Harvey, A.C. and N. Shephard (1993), 'Estimation and Testing Stochastic Variance Models', London School of Economics - Suntory Toyota, Econometrics Series n. 93-268.

Harvey, A.C. and N. Shephard (1996), 'Estimation of an asymmetric model of asset prices', *Journal of Business and Economic Statistics,* 14/4, 429-434.

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

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

Hull J. and A. White (1998), 'Incorporating volatility updating into the historical simulation method for Value-at-Risk', *The Journal of Risk*, 1, 5-19.

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

Hurvich, C.M. and B.K. Ray (2001), 'The local Whittle estimator of long memory stochastic signal volatility', Working paper Department of Statistics and Operations Research, New York University, http://www.stern.nyu.edu/sor/research/wp01.html.

Hurvich, C.M., E. Moulines and P. Soulier (2001), 'Estimation of long memory in stochastic volatility', Preprint 153, Université d'Evry-Val d'Essonne, http://www.tsi.enst.fr/~soulier/paper.html.

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

Jacquier, E., N.G. Polson and P. Rossi (1995), 'Models and Priors for Multivariate Stochastic Volatility', CIRANO Working Papers 95-18.

Jensen, M.J. (1999), 'An approximate wavelet MLE of short and long memory parameters', *Studies in Nonlinear Dynamics and Econometrics,* 3, 239-353.

Jensen, M.J. (2000), 'An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets', *Journal of Economic Dynamics and Control,* 24, 361-387.

Jensen, M.J. (2001), 'Bayesian Inference of Long-Memory Stochastic Volatility via Wavelets', WP 01-02 Department of Economics, University of Missouri-Columbia**.**

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, C.J. (1994), 'Dynamic linear models with Markov-switching', *Journal of Econometrics,* 60, 1-22.

Kim, C.J. and C. R. Nelson (1999), *State-Space Models with Regime Switching*, Cambridge, MIT Press.

Kim, S., N. Shephard and S. Chib (1998), 'Stochastic volatility: likelihood inference and comparison with arch models', *Review of Economic Studies*, 65, 361–393.

Kitagawa, G. (1987), 'Non-Gaussian state space modeling of nonstationary time series' (with discussion), *Journal of the American Statistical Association*, 82, 400, 1032–1063.

Kitagawa, G. (1996), 'Monte Carlo filter and smoother for non-Gaussian non linear state space models', *Journal of Computational and Graphical Statistics*, 5, 1–25.

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.

Lehar, A., M. Scheicher and C. Schittenkopf (2002), 'Garch vs Stochastic Volatility: Option Pricing and Risk Management', *Journal of Banking and Finance*, 26(2-3), 323-345.

Lobato, I. and N.E. Savin (1998), 'Real and spurious long-memory properties of stock-market data', *Journal of Business and Economic Statistics,* 16, 261-268.

Mahieu, R. and P. Schotman (1998), 'An empirical application of stochastic volatility models', *Journal of Applied Econometrics,* 16, 333-59.

McNeil, A.J. and R. Frey (2000), 'Estimation of Tail-Related Risk Measures for Heteroskedastic Financial Time Series: An Extreme Value Approach', *Journal of Empirical Finance,* 7, 271-300.

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

Merton, R.C. (1976), 'Option pricing when underlying stock returns are discontinuous', *Journal of Financial Economics* 3, 125-144.

Monfardini, C. (1998), 'Estimating stochastic volatility models through indirect inference', *Econometrics Journal*, 1, 113–128.

Naik, V. (1993), 'Option valuation and hedging strategies with jumps in the volatility of asset returns', *Journal of Finance*, 48, 1993.

Pitt, M. and N. Shephard N. (1999a), 'Filtering via simulation: auxiliary particle filter', *Journal of the American Statistical Association*, 94, 590-599.

Pitt, M. and N. Shephard (1999b), 'Time varying covariances: a factor stochastic volatility approach' with discussion, in *Bayesian Statistics*, Bernardo, J., Berger, J.O., Dawid, A.P., Smith, A.F.M. (eds.), Vol. 6. Oxford University Press, Oxford, 547-570.

Ray, B.K. and R. Tsay, (2000), 'Long-range dependence in daily stock volatilities', *Journal of Business and Economic Statistics,* 18, 254-262.

Robinson, P.M. (1995), 'Log-periodogram regression of time series with long range dependence', *Annals of Statistics,* 23, 1043-1072.

Rubinstein, M. (1985), 'Nonparemetric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE options classes from August 23, 1976 through August 21, 1978', *The Journal of Finance*, 40, 455-480.

Ruiz, E. (1994), 'Quasi maximum likelihood estimation of stochastic volatility models', *Journal of Econometrics,* 63, 289-306.

Sandmann, G. and S. Koopman (1998), 'Estimation of stochastic volatility models via Monte Carlo maximum likelihood', *Journal of Econometrics*, 87, 271–301.

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

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

Shephard, N. (1993), 'Fitting non-linear time series models, with applications to stochastic variance models', *Journal of Applied Econometrics*, 8, 135–152.

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.

Shephard, N. (1996), 'Statistical aspects of ARCH and stochastic volatility', in *Time Series Models with Econometric, Finance and Other Applications*, Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E. (eds.), Chapman & Hall, London, 1-677.

Smith, A.A. (1993), 'Estimating non linear time series models using simulated vector autoregressions', *Journal of Applied Econometrics,* 8, S63-S84.

So, M.K.P., K. Lam and W.K. Lee (1998), 'A stochastic volatility model with Markov switching', *Journal of Business and Economic Statistics,* 16, 244-53.

Sorenson, H. and D. Alspach (1971), 'Recursive Bayesian estimation using Gaussian sums', *Automatica*, 7, 465–479.

Tanizaki, H. and R. Mariano (1994), 'Prediction, filtering and smoothing in non-linear and non-normal cases using Monte Carlo integration', *Journal of Applied Econometrics*, 9, 163–179.

Tanizaki, H. and R. Mariano (1998), 'Nonlinear and nonnormal state-space modeling with Monte Carlo stochastic simulation', *Journal of Econometrics*, 83, 1-2, 263–290.

Tauchen, G. (1997), 'New minimum chi-square methods in empirical finance', in: *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, Kreps D. and Wallis K., eds., Cambridge University Press, Cambridge.

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

Taylor, S.J. (1994), 'Modelling stochastic volatility: a review and comparative study', *Mathematical Finance*, 4, 183–204.

Tierney, L. (1994), 'Markov chains for exploring posterior distributions' (with discussion), *Annals of Statistics*, 22, 1701–1786.

Wright, J.H. (1999), 'A new estimator of the fractionally integrated stochastic volatility model', *Economics Letter,* 63, 295-303.

Wright, J.H. (2000), 'Log-periodogram estimation of long memory volatility dependencies with conditionally heavy tailed returns', Board of Governors of the Federal Reserve System, International Finance Discussion Papers n. 685, http://www.bog.frb.fed.us.