Some aspects of modeling and statistical inference for financial models

K. Dzhaparidze*

Centre for Mathematics and Computer Science, Kruislaan 413, 1098 SJ Amsterdam

P. J. C. Spreij

Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam

J. H. van Zanten

Centre for Mathematics and Computer Science, Kruislaan 413, 1098 SJ Amsterdam

Modeling the stock price development as a geometric Brownian motion or, more generally, as a stochastic exponential of a diffusion, requires the use of specific statistical methods. For instance, the observations seldom reach us in the form of a continuous record and we are led to infer about diffusion coefficients from discrete time data. Next, often the classical assumption that the volatility is constant has to be dropped. Instead, a range of various stochastic volatility models is formed by the limiting transition from known volatility models in discrete time towards their continuous time counterparts. These are the main topics of the present survey. It is closed by a quick look beyond the usual Gaussian world in continuous time modeling by allowing a Lévy process to be the driving process.

1 Introduction

The amount of literature on quantitative aspects in finance is huge, even if one restricts the attention to papers that have a strong mathematical flavor. The purpose of the present paper is to give a survey of some (rather recent) approaches to problems of a statistical nature that arise in the context of mathematical finance. Due to space limitations it is impossible to give an account of this subject that does fully give justice to all the efforts of the many researchers in this field. Necessarily we had to give up striving for completeness. The choice of subjects that we present reflects our own, perhaps slightly biased, interests.

Nevertheless we attempt to present a rather coherent looking point of view based on the following principles. We mainly concentrate on models in continuous time

---

* kacha@cwi.nl

(however, these may be viewed as approximations of some explicit discrete time models that we also discuss). However, observations may reach us in the form of a continuous record or (more realistically) at discrete time instants only. The models we discuss are used to explain the behaviour of stock prices and of the volatility of these prices. Both are described by stochastic differential equations driven by Wiener processes. Parameters of interest are then found in the drift and diffusion coefficients and we will discuss various estimation methods that may be used in the different observations schemes that occur under different circumstances. We will only be concerned with finite dimensional parameters. We also discuss the relevance of using Wiener processes as the basic process on which the models are built and review alternatives.

In a paper that has been forgotten for many decades—as an illustration of this we mention that the English translation appeared 64 years later—BACHELIER (1900) introduced Brownian motion as a basic stochastic process to account for the fluctuations of stock prices. The model he used also contained an additive drift term (a trend). Although this model had some appeal it suffered from the substantial drawback that at all moments prices could be negative with positive probability. This led SAMUELSON (1965) to use the same model for the logarithm of the stock price as Bachelier did for the stock price itself. Formulated in another way, he suggested to use *geometric* Brownian motion to model the evolution of stock prices. His model is reflected in equation (1) below, that is nowadays often attributed to BLACK and SCHOLES (1973) who used it in their seminal paper on option pricing that appeared 8 years later than Samuelson's paper. If we denote by S_t the stock price at time t, the model suggested by Samuelson can be formulated in terms of the following *stochastic differential equation*

$$\mathrm{d}S_t = S_t(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t), \quad S_0 > 0 \tag{1}$$

where μ and σ are constants and W is a standard Brownian motion. The parameter σ is called the *volatility* parameter and measures how uncertain we are about future stock price movements. We find it back in the variance of infinitesimally small relative changes of the stock price. Informally, we have $\mathrm{Var}\,\mathrm{d}S_t/S_t = \sigma^2\,\mathrm{d}t$.

Using the Itô formula (see equation (3)) we find that $\log(S_t/S_0) = (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t$ so that S_t is always positive, and that the log-return $\log(S_{t+\Delta}/S_t)$ over an interval $[t, t+\Delta]$ is normally distributed and independent of the past values up to time t. Consequently we also get $\mathrm{Var}\,\log(S_{t+\Delta}/S_t) = \sigma^2\Delta$. Hence σ^2 is in this model not only the variance of log-returns over an interval of unit length, but also the conditional variance given the prices before such an interval. Other models that we will come across in the sequel are generalizations of the one in (1). This means that we will mainly work with *diffusions*.

For future reference we briefly recall some terminology from the theory of stochastic processes. A good account of this theory can be found in KARATZAS and SHREVE (1991).

By diffusion processes or diffusions we mean processes, X say, that can be represented by

$$\mathrm{d}X_t = b_t \, \mathrm{d}t + a_t \, \mathrm{d}W_t \tag{2}$$

with W a Brownian motion and b_t and a_t other stochastic processes that are usually of the form $b_t = b(X_t)$ and $a_t = a(X_t)$ for certain functions b and a. In equations of the type of (2) we call b the drift coefficient and a the diffusion coefficient.

One of the main reasons for modeling with diffusion processes is that functions of a diffusion are diffusions again. More precisely if f is a twice continuously differentiable function and X satisfies (2), then $f(X)$ satisfies an equation of the same type, namely

$$\mathrm{d}f(X_t) = \left(f'(X_t)b_t + \tfrac{1}{2}f''(X_t)a_t^2 \right) \mathrm{d}t + f'(X_t)a_t \, \mathrm{d}W_t. \tag{3}$$

The resulting differentiation formula in equation (3) is known as Itô's formula, an indispensable and powerful tool.

Consider now a process X satisfying (2), with $b_t = b(X_t)$ and $a_t = a(X_t)$, i.e. X is the solution of the following stochastic differential equation.

$$\mathrm{d}X_t = b(X_t) \, \mathrm{d}t + a(X_t) \, \mathrm{d}W_t. \tag{4}$$

Weak solutions of such equations are Markov processes and under a condition that guarantees the existence of an invariant distribution also ergodic (see e.g. GIHMAN and SKOROHOD (1972) or ROGERS and WILLIAMS (1997)). The invariant measure of an ergodic diffusion has a density that can be expressed in terms of the coefficients as follows. Let

$$s(x) = \exp\left(-2 \int \frac{b(x)}{a^2(x)} \, \mathrm{d}x \right),$$

where the notation $\int f(x) \, \mathrm{d}x$ is used to denote any function whose derivative is f. Then the stationary density is proportional to

$$\frac{1}{s(x)a^2(x)}. \tag{5}$$

The rest of the paper is organized as follows. In section 2 we abandon the assumption that σ in (1) is taken as a constant. Instead we discuss several ways of modeling σ as a stochastic process, both in discrete and in continuous time and the relations between these models in terms of weak convergence of discrete time processes to a diffusion limit. Then we treat in section 3 parameter estimation for diffusion processes with full observations, by which we mean that we observe all components of this process in case it is multidimensional. Most attention is paid to the case where the observations are available at discrete times only. In section 4 we will have a look at an estimation problem for a partially observed diffusion. The financial context is that of a bivariate diffusion, whose components are stock price and volatility, while only the stock price is observed. Whereas most of the models that we will treat up to that section are based on a diffusion we change this point of

departure in section 5 where we discuss statistical problems that give rise to use a certain Lévy process other than Brownian motion as a driving process.

2 Stochastic volatility models

The underlying model according to Samuelson, represented in (1), has a number of consequences. One is that the log-returns follow a normal distribution. Contrary to that it has been known since quite some time (MANDELBROT, 1963) that empirical analysis of financial data reveals excess kurtosis. Hence many authors tried to model returns of financial data directly by means of other distributions. We will come back to this issue in section 5.

Another issue is the constancy of the volatility parameter σ. This has been criticized by a number of authors. To explain some of the criticism we introduce some terminology. First we define the log price process $R_t = \log S_t$. Assuming (1), as we said before, it is given by $R_t = R_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t$. Notice that we have that $R_t - R_s$ is independent of the past up to time s, so that the conditional variance $\Sigma_t - \Sigma_s$ of $R_t - R_s$ given this past equals $\sigma^2(t - s)$. Hence for this model we find that $(\mathrm{d}/\mathrm{d}t)\Sigma_t = \sigma^2$. In more general models we define the volatility σ_t at time t as the conditional standard deviation of an infinitesimal small increment of R_t given the whole past just before time t. In general then, the σ_t form a stochastic process, called the volatility process, or just volatility. Let us make this notion precise. Suppose that S is given by the following extension of (1)

$$\mathrm{d}S_t = S_t(\mu_t \,\mathrm{d}t + \sigma_t \,\mathrm{d}W_t), \tag{6}$$

where μ and σ are appropriate stochastic processes. Applying Itô's rule we get for $R = \log S$

$$\mathrm{d}R_t = (\mu_t - \tfrac{1}{2}\sigma_t^2)\,\mathrm{d}t + \sigma_t \,\mathrm{d}W_t. \tag{7}$$

Let M be given by $M_t \equiv \int_0^t \sigma_s \,\mathrm{d}W_s$ and let $\Sigma = \langle M \rangle$, the quadratic variation process of M. Then $\Sigma_t \equiv \int_0^t \sigma_s^2 \,\mathrm{d}s$, with the interpretation that the conditional variance of $R_{t+\mathrm{d}t} - R_t$ given the past up to time t is given by $\Sigma_{t+\mathrm{d}t} - \Sigma_t = \sigma_t^2 \,\mathrm{d}t$. Hence the σ_t in equation (6) is indeed the volatility as we just defined it.

For discrete time models we follow a similar approach. Now the stock price is given by a discrete time process $S_0, S_1, \ldots$. As before R is the log-price process and the log-return process ΔR is defined by $\Delta R_t = \log(S_t/S_{t-1})$, with $\Delta R_t = R_t - R_{t-1}$. In this case we define the volatility σ_t at time t as the conditional standard deviation of ΔR_t given the past up to time $t - 1$. Again we will call the process $\{\sigma_t\}$ the volatility (process).

An empirical fact is that volatility evolves in clusters, there are periods of high volatility (a nervous market) followed by periods of low volatility (a calm market) and vice versa. Typically ARCH models (to be discussed in section 2.1) explain this behaviour in discrete time.

Another empirical finding is that volatility movements are negatively correlated with stock price movements. This is called the *leverage effect* (BLACK, 1976).

In practice people also work with what is called *implied volatility*. This is defined as follows. Suppose that S is modeled according to equation (1). The Black–Scholes price of an option (see BLACK and SCHOLES (1973) or BJÖRK (1998), page 90) for the explicit formula) is a function of the volatility parameter σ as well as a function of the exercise price, the time to maturity, the current price of the underlying asset and the interest rate. Keeping all the parameters constant except the σ (a *ceteris paribus* condition) one can invert the Black–Scholes formula to find σ as a function of the market price of an option. The resulting value of σ is called the implied volatility (parameter). Doing so for different options and different values of the exercise price, it has been observed that the implied volatility showed a typically not constant behaviour that is not compatible with the Samuelson–Black–Scholes assumption. As a matter of fact, if the exercise price is close to the actual stock price (this is called an 'at the money option') the implied volatility is relatively low as compared to the values obtained when the exercise price and stock price are far apart. This phenomenon is known as the *smile* pattern of implied volatility.

The qualitative properties of financial data that we mentioned above are often referred to as stylized facts or stylized features of these data. Other examples of stylized features are heavy-tailedness and long-range dependence. In the next subsections we will discuss a number of alternative approaches that are aimed at mimicking the practically observed non constant behaviour of the volatility. Although efforts have been undertaken to model volatility as a deterministic function (see e.g. DUPIRE (1994)), these models suffer from some drawbacks especially when it concerns predictive performance DUMAS, FLEMING and WHALEY (1995). Nevertheless these models also have a great advantage, since in this set up markets are complete. Therefore hedging strategies can be used to price derivatives and a unique price exists (see BJÖRK, 1998, chapter 9 for a discussion of these concepts). Henceforth we will concentrate on stochastic models.

2.1 *Volatility models in discrete time*

As we have argued at the beginning of this section, the assumption that the volatility is a constant cannot be held. There is a variety of models that give alternative descriptions of the volatility as a stochastic process. In this subsection we present some of the more popular ones in discrete time. In the next subsection we present continuous time models.

All models below belong to the class of so called (G)ARCH-processes or to some related class of processes. An introduction to modeling with (G)ARCH processes is given in the recent book by GOURIÉROUX (1997) as well as in the survey papers BOLLERSLEV, CHOU and KRONER (1992) (which contains an extensive discussion of qualitative aspects and an economic interpretation of the ARCH-like methodology of modeling) and BOLLERSLEV, ENGLE and NELSON (1994) (where more attention is paid to mathematical and probabilistic properties of these models).

Whereas in the 1970s modeling with ARMA processes (BOX and JENKINS, 1976, is a widely used reference), which concentrates on modeling conditional first moments of a time series, became very popular it was observed later on that a standard ARMA fit to financial data was poor. The by now classical paper of ENGLE (1982), where ARCH processes were introduced as a way to model conditional second order moments, was a first attempt to overcome this problem. Later on BOLLERSLEV (1986) observed that even the class of ARCH-processes was not always suitable for modeling, unless a large number of parameters was involved. In line with the parsimony doctrine of BOX and JENKINS (1976), he extended the concept of ARCH to that of GARCH, in a way similar to the extension of AR models to ARMA ones. Indeed, GARCH models with few parameters often give a good fit, see e.g. BOLLERSLEV, CHOU and KRONER (1992). Note that (G)ARCH is just one of the possible ways of nonlinear modeling of time series.

We continue with an explanation of (G)ARCH modeling. Suppose one is given some sequence of observations, $\{Y_t\}$ say, for instance $Y_t = \Delta R_t$. The basic model is to write

$$Y_t = \mu_t + \varepsilon_t \tag{8}$$

Here we take ε as a zero mean martingale difference sequence w.r.t some underlying filtration $\{\mathscr{F}_t\}$ and μ_t is then the conditional expectation of Y_t given the past up to time $t-1$. In the ARCH modeling framework a model is specified for the innovations process ε. The starting point for this is

$$\varepsilon_t = \sigma_t z_t \tag{9}$$

where σ is a nonnegative predictable process (i.e. σ_t is $\mathscr{F}_{t-1}$-measurable for all t) and z an *i.i.d.* sequence, usually standard normal. In this set up we find that σ_t^2 is the conditional variance of Y_t given the past. In particular, if $Y_t = \Delta R_t$ and assuming that $\sigma_t \geqslant 0$, we see that σ_t plays the role of the volatility process as we defined it for discrete time observations of the stock price.

In the rest of this section we assume that the z_t form a white noise sequence that is symmetric around zero, i.e. all the z_t have the same symmetric distribution with $Ez_t^2 = 1$ (and $Ez_t = 0$) and $Ez_t z_s = 0$ for $t \neq s$.

Recall that the kurtosis κ of the distribution of a random variable X is defined by $\kappa = [E(X - EX)^4]/(E(X - EX)^2)^2$ if $EX^2 < \infty$ and by ∞ if $EX^2 = \infty$. Notice that always $\kappa \geqslant 1$ and that for a normal distribution $\kappa = 3$.

One feature that one wants to capture in a model for financial data is leptokurtosis (as compared to a Gaussian sequence), one way of expressing that the tails of a probability distribution are fatter than those of a (standard) normal distribution. It already follows from the general set up of equation (8) and the specification of the innovations as in (9) that this is the case, as we shall see now.

Assuming that z is an i.i.d. sequence with finite fourth moments, we find that the conditional kurtosis of ε_t, defined as $E[\varepsilon_t^4|\mathscr{F}_{t-1}]/(E[\varepsilon_t^2|\mathscr{F}_{t-1}])^2$, is equal to the kurtosis of z_t, whereas the unconditional kurtosis of ε_t factorizes as the product of

the kurtosis of σ_t and that of z_t and is therefore bigger than the conditional kurtosis if σ_t is not deterministic. Moreover, assuming a standard normal law for the z_t, we find that the kurtosis of ε_t is equal to 3 times the kurtosis of σ_t and hence bigger than 3.

The contribution of Engle was to specify a functional relationship (called ARCH(q)) between σ_t^2 and past innovations as

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2, \tag{10}$$

with the a_i nonnegative, or equivalently with $u_t = \varepsilon_t^2 - \sigma_t^2$ as

$$\varepsilon_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + u_t, \tag{11}$$

where u is again a martingale difference sequence, provided that $E\sigma_t^2$ is finite for all t. So the squares ε_t^2 satisfy an equation of the autoregressive type and in particular they also form a positively correlated sequence, as opposed to the uncorrelated sequence ε.

Assuming (weak) stationarity of ε^2 (it is then required that $\sum_{i=1}^{q} a_i < 1$) we find $E\varepsilon_t^2 = a_0/(1 - \Sigma_{i=1}^{q} a_i)$.

It is not too difficult to compute the kurtosis of ε_t in the ARCH(1) case under the assumption that the z_t are standard normals, see e.g. GOURIÉROUX (1997). The result is $3(1 - a_1^2)/(1 - 3a_1^2)$, provided that $3a_1^2 < 1$. It is clear that apart from explaining excess kurtosis, ARCH models are also appropriate to describe other stylized facts for financial time series such as volatility clustering, due to the positive correlation in the ε_t^2. Typically we find the tendency of small (large) absolute values of the ε_t's to be followed by large (small) values.

The GARCH extension (G stands for generalized) introduced by Bollerslev then consists of adding past values of σ^2 with positive weights to equation (2.10) to obtain the GARCH(p, q) specification

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} b_i \sigma_{t-i}^2, \tag{12}$$

with the b_i nonnegative. With the u process as above this equation may be rewritten as

$$\varepsilon_t^2 = a_0 + \sum_{i=1}^{\max\{p,q\}} (a_i + b_i)\varepsilon_{t-i}^2 + u_t - \sum_{i=1}^{p} b_i u_{t-i}, \tag{13}$$

with the convention that $a_i = 0$ for $i > q$ and $b_i = 0$ for $i > p$. This gives an ARMA-like representation for the ε^2 process.

It has been reported that already a GARCH(1,1) model gave a good description of certain financial data (BOLLERSLEV, CHOU and KRONER, 1992).

Among the many variations on the (G)ARCH theme we mention one, the so called EGARCH model proposed by NELSON (1990), see also NELSON (1991). In this case

we have a specification of the logarithm of the conditional variance process, viz. with $v_t = \log \sigma_t^2$

$$v_t = a_0 + \sum_{i=1}^{p} a_i v_{t-i} + \sum_{i=1}^{q} b_i(|z_{t-i}| - E|z_{t-i}| + cz_{t-i}). \tag{14}$$

An additional feature of EGARCH models is that they are capable of explaining the leverage effect, whereas (G)ARCH models cannot. To see the latter, notice that due to the independence property of the z process, we have that the increments of σ^2 as modeled by a (G)ARCH process are uncorrelated with z_t, since the correlation between z_t and z_t^2 is zero. (G)ARCH depends on the past of ε only through ε^2, so the magnitude of ε, not its sign. Clearly, in EGARCH models for σ^2, this is different.

All these models that are used to describe or explain financial data are based on empirical analysis of these data. There is no structural dynamic economic theory that can be used as a basis for modeling.

2.2 *Stochastic volatility models in continuous time*

We continue with proposing some stochastic volatility models in continuous time. These models, all of them of the type of equation (4), can often be seen as diffusion approximations of discrete time models. This will be explained in subsection 2.3.

Out of the many models that are currently used we have selected the following ones that are of prime interest (for the present paper). They all start from the model (1) for the stock price behaviour, but instead of assuming σ to be constant a stochastic model in terms of a suitable function of σ is proposed for it. These models are

$$d\sigma_t^2 = b\sigma_t^2 \, dt + \delta\sigma_t^2 \, dW_t \tag{15}$$

$$d\log\sigma_t^2 = (b_1 - b_2\log\sigma_t^2) \, dt + \delta \, dW_t \tag{16}$$

$$d\sigma_t^2 = (b_1 - b_2\sigma_t^2) \, dt + \delta \, dW_t \tag{17}$$

$$d\sigma_t^2 = (b_1 - b_2\sigma_t^2) \, dt + \delta\sigma_t \, dW_t \tag{18}$$

$$d\sigma_t^2 = (b_1 - b_2\sigma_t^2) \, dt + \delta\sigma_t^2 \, dW_t \tag{19}$$

Equation (15) models the volatility as a geometric Brownian motion and was used by HULL and WHITE (1987). In equation (16) an Ornstein–Uhlenbeck process is used to model the logarithm of the volatility and was proposed as a model by WIGGINS (1987). It turned out to be an empirical relevant one, see BOLLERSLEV, ENGLE and NELSON (1994). We will come back to this model in subsection 2.3. The Ornstein–Uhlenbeck process of equation (17) is mainly used as a (popular) model for short term interest rates and has been proposed by VASICEK (1977). It is not suitable as a model for σ^2, since the solution of this equation takes on negative

values with positive probability. The model of equation (18) was suggested by HESTON (1993). It is the same as the one that Cox, INGERSOLL and ROSS (1985) used for the term structure of interest rates. Finally, equation (19) arises in a natural way as a limit of a GARCH(1,1) process, see subsection 2.3. From the examples above the conditions to ensure an ergodic solution are satisfied by all of them for proper choices of the parameters except for the geometric Brownian motion of (15).

After some years in which the ARCH-type processes in discrete time had enjoyed an immense popularity, the question arose as to how such processes had to be defined in continuous time. A definition has been proposed by DROST and WERKER (1996) based on temporal aggregation for discrete time models or equivalently (for continuous time models) sampling at different frequencies to get a discrete time process. They also showed that the process of equation (19) satisfies their definition of a continuous time GARCH process. We will not go into details and subtleties concerning this topic and refer the reader to DROST and WERKER (1996) and the references therein.

2.3 *Continuous time approximations*

Although there had been many results for diffusion limits of discrete time processes around (most famous is perhaps the functional Central Limit Theorem, or Donsker's invariance principle, see (KARATZAS and SHREVE, 1991, theorem 2.4.20 or (BILLINGSLEY, 1968, theorem 2.10.1), it was not until 1990, that limiting results for ARCH-type models were investigated. NELSON (1990) appears to be one of the first who made contributions in this direction and proved for instance a limit theorem for the GARCH(1,1) process.

As examples we summarize some results in the literature and confine ourselves in doing so to the models we have discussed previously.

The general scheme is the following. A sequence of random processes X^n, where $X^n = (X_1^n, X_2^n, \ldots)$ is available. With this sequence we associate a sequence ξ^n of continuous time processes as follows. Let (h_n) be a sequence that converges to zero (often $h_n = 1/n$) and consider

$$\xi_t^n = \sum_{k \leqslant t/h_n} X_k^n. \tag{20}$$

This way a sequence of processes is obtained that are (obviously) right continuous and have finite left limits at all time instants. The suitable space to consider weak convergence is therefore the Skorohod space $D[0, 1]$ (or $D[0, \infty)$). General theorems on weak convergence in this setting can be found in e.g. the books ETHIER and KURTZ (1986), JACOD and SHIRYAEV (1987), BILLINGSLEY (1968), or STROOCK and Varadham (1979).

We proceed with giving two examples, both taken from NELSON (1990). The first one concerns a GARCH (1,1) process with $\mu_t = Y_{t-1} + c\sigma_t^2$ and $\sigma_t^2 = a_0 + b\sigma_{t-1}^2 + a\varepsilon_{t-1}^2$. We assume that $0 < a + b < 1$. The basic model given by (8) and (9) is then specified as follows:

$$Y_t - Y_{t-1} = c\sigma_t^2 + \sigma_t z_t \tag{21}$$

and

$$\sigma_t^2 - \sigma_{t-1}^2 = a_0 - (1 - b - a)\sigma_{t-1}^2 + a\sigma_{t-1}^2(z_{t-1}^2 - 1) \tag{22}$$

where z is the *i.i.d.* sequence of section 2.1. This is a usual discrete time model with the time parameter t running through consecutive integers. A continuous time approximation to it may arise only if we partition time more and more finely. In this way the following sequence of related models, indexed by $n = 1, 2, \ldots$ is treated. The mesh of the partition in the nth model is h_n, like in (20). To describe thus the nth model, write for simplicity

$$Y_{kh_n} \equiv Y_k^n, \, \sigma_{kh_n} \equiv \sigma_k^n, \, z_{kh_n} \equiv z_k^n,$$

and give the upper index n to all the constants involved. From (21) and (22) we get at $t = kh_n$

$$Y_k^n - Y_{k-1}^n = c^n \sigma_k^{n2} + \sigma_k^n z_k^n \tag{23}$$

and

$$\sigma_k^{n2} - \sigma_{k-1}^{n2} = a_0^n - (1 - b^n - a^n)\sigma_{k-1}^{n2} + a^n \sigma_{k-1}^{n2}(z_{k-1}^{n2} - 1). \tag{24}$$

We have deliberately written the increments on the left-hand side of these equations. It is now easy to take the sum like in (20) to get

$$Y_{[t/h_n]}^n = \frac{c^n}{h_n} \sum_{k \leqslant t/h_n} \sigma_k^{n\,2} h_n + \sum_{k \leqslant t/h_n} \sigma_k^n z_k^n \tag{25}$$

and with $\varsigma_k^n = (z_k^{n2} - 1)/\sqrt{2}$ and the parameters $\delta^n \equiv \sqrt{2}\, a^n$, $\theta^n \equiv (1 - b^n - a^n)/h_n$

$$\sigma_{[t/h_n]}^n = a_0^n \left[\frac{t}{h_n}\right] - \theta^n \sum_{k \leqslant t/h_n} \sigma_{k-1}^{n2} h_n + \delta^n \sum_{k \leqslant t/h_n} \sigma_{k-1}^{n2} \varsigma_{k-1}^n \tag{26}$$

where all the coefficients are supposed to be convergent as $n \to \infty$ so that

$$\frac{c^n}{h_n} \to c, \, \delta^n \to \delta, \frac{a_0^n}{h_n} \to b_1, \, \theta^n \to b_2;$$

we retain the notations of (19) for the limiting coefficients. The system (25) and (26), treated continuously in time for $n = 1, 2, \ldots,$ yield the sequence of a bivariate piecewise constant process (Y_t^n, σ_t^{n2}) that may be imbedded in the Skorohod space $D([0, T], \mathbb{R}^2)$. To understand the weak limit of this sequence, first look at the sum in the first term on the right hand side of (25) or at the similar sum in the second term on the right hand side of (26). It is easy to recognize the usual Riemann sum which tends to converge to a corresponding Riemann integral. We shall display it soon. Look meanwhile at the second term on the right hand side of (25) or at the similar sum in the extreme right term of (26) – both weighted sums formed by the standardized iid sequences z and ς (note that if z is the sequence of

independent standard normal variables, then ς is a standardized iid sequence uncorrelated with z; this anticipates the independence of the two Gaussian components which these two sequences yield in the limit). Here the matter is a bit more delicate—we leave out the details—and one needs to appeal to Donsker's invariance principle which explains occurrence of two stochastic integrals with respect to two independent standard Brownian motions W^1 and W^2 in the following system of diffusion equations:

$$Y_t = c \int_0^t \sigma_s^2 \mathrm{d}S + \int_0^t \sigma_s \, \mathrm{d}W_s^1 \tag{27}$$

$$\sigma_t^2 = \int_0^t (b_1 - b_2 \sigma_s^2) \, \mathrm{d}S + \delta \int_0^t \sigma_s^2 \, \mathrm{d}W_s^2, \tag{28}$$

with the solution (Y, σ^2) that is the resulting weak limit of the bivariate sequence of piecewise constant process $(Y^n, \sigma^{n2})_{n=1,2,...}$. Note that the second of these equations is the integral version of the stochastic differential equation (19).

Under the conditions that $b_1 > 0$ and $2b_2 + \delta^2 > 0$ it follows from (28) that the process σ_t^{-2} has a stationary gamma distribution with parameters $1 + 2b_2/\delta^2$ and $2b_1/\delta^2$. Furthermore, it can be shown that for the case where $c = 0$ the t-distribution with $2 + 4b_2/\delta^2$ degrees of freedom is the invariant distribution of Y. We will return to this in subsection 4.2.

As a next example we briefly consider the EGARCH model (14) with $p = q = 1$ and set (for simplicity) μ_t identically zero. The approach is similar in spirit to the previous example and we only give the result. The limiting diffusion process satisfies (for an appropriate choice of the coefficients)

$$\log \sigma_t^2 = \int_0^t (b_1 - b_2 \log \sigma_s^2) \, \mathrm{d}s + W_t^2. \tag{29}$$

Unlike the situation of the previous example here the Brownian motion W^2 has non-zero correlation with the Brownian motion that drives the return process. As a consequence this model also captures the leverage effect. We refer to NELSON (1990) for details.

Equation (29) describes the logarithm of the volatility process in the same way as the model of WIGGINS, cf. (16).

3   Inference for diffusion processes with full observations

In this section we will consider parameter estimation when we have observations, either as a continuous record or as a finite set, from a diffusion process. We limit ourselves to estimation of finite dimensional parameters. Nonparametric procedures will not be treated. The reader who is interested in these is referred to work by AÏT-SAHALIA (1996), for instance, where the important topic of nonparametric estimation of the volatility function is treated.

The sequel of this section is divided into two parts. In the first part (subsection 3.1) we discuss estimation procedures based on the likelihood of a diffusion that is continuously observed. In the second part (subsection 3.2) inference procedures based on the likelihood of discrete time observations of a diffusion will be treated.

3.1 *Inference based on the continuous likelihood*

3.1.1 *MLE for continuous observations*

Consider the stochastic differential equation (SDE)

$$dX_t = b(X_t, \theta) \, dt + \sigma(X_t) \, dW_t, \quad X_0 = x, \quad 0 \leqslant t \leqslant T, \tag{30}$$

where W is a standard Brownian motion, $\sigma > 0$ and the drift b depends on an unknown parameter $\theta \in \Theta \subseteq \mathbb{R}$. We suppose that the SDE has a weak solution for all θ, and that this solution is unique in law. We denote the law of such a solution by $\mathbb{P}_\theta$, that is a probability measure on the Borel sets of $C[0, T]$. The aim is to estimate θ from the continuous observations $\{X_t\}_{t \leqslant T}$.

It is well known (see e.g. LIPTSER and SHIRYAYEV (1977)) that if for all θ

$$\int_0^T \frac{b^2(X_s, \theta)}{\sigma^2(X_s)} \, ds < \infty,$$

$\mathbb{P}_\theta$-a.s., then there exists a probability measure $\mathbb{P}_0$ on $C[0, T]$, equivalent to all the $\mathbb{P}_\theta$ and such that for all θ

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}_0}(X) = \exp\left[\int_0^T \frac{b(X_s, \theta)}{\sigma^2(X_s)} \, dX_s - \frac{1}{2} \int_0^T \frac{b^2(X_s, \theta)}{\sigma^2(X_s)} \, ds\right], \tag{31}$$

$\mathbb{P}_\theta$-a.s. This fact follows from Girsanov's theorem (KARATZAS and SHREVE, 1991, theorem 3.5.1 or JACOD and SHIRYAYEV, 1987, theorem III.3.24 for a more general result). The Radon-Nikodym derivative $L_T(\theta) = d\mathbb{P}_\theta/d\mathbb{P}_0$ is called the *likelihood function* and the *maximum likelihood estimator (MLE)* is defined as the point where $\theta \mapsto L_T(\theta)$ attains its maximum.

It is important to note that for the derivation of (31) it is essential that the diffusion coefficient a does not depend on the parameter. The dominating measure $\mathbb{P}_0$ is constructed by using Girsanov's theorem to remove the drift from X. In other words, $\mathbb{P}_0$ is the law of the weak solution of the SDE

$$dX_t = \sigma(X_t) \, dW_t, \quad X_0 = x.$$

Therefore, we can only use Girsanov's theorem to find the dominating measure $\mathbb{P}_0$ and the expression (31) when σ does not depend on the parameter θ. In subsections 3.1.3 and 3.2 we will discuss estimation problems when σ depends on θ.

For the study of the asymptotic properties of the MLE we refer to KUTOYANTS (1984). Suppose that the diffusion X is ergodic, with invariant law π_θ, and that b is differentiable with respect to θ, denote the derivative by $\dot{b}$. For all θ, suppose that the following quantity is finite:

$$I(\theta) = \int_{-\infty}^{\infty} \left(\frac{\dot{b}(x, \theta)}{\sigma(x)} \right)^2 \pi_\theta(\mathrm{d}x). \tag{32}$$

This quantity is called the Fisher information. If $\theta \mapsto I(\theta)$ is bounded away from 0, then (under some extra regularity conditions) the MLE is consistent, and we have for all θ that the law of $\sqrt{T}(\hat{\theta}_T - \theta)$ under $\mathbb{P}_\theta$ converges to the normal law $N(0, I(\theta)^{-1})$ as $T \to \infty$, i.e.

$$\mathscr{L}\{\sqrt{T}(\hat{\theta}_T - \theta)|\mathbb{P}_\theta\} \to N(0, I(\theta)^{-1}). \tag{33}$$

Hence, by the Cramér–Rao theorem, the estimator is asymptotically efficient.

3.1.2 *Discretization of the likelihood*

Consider the ergodic diffusion X defined by (30) with the drift b that again satisfies some smoothness and growth conditions. But unlike the previous section, suppose that X is observed at certain equidistant instants $t_i^n = i h_n$, $i = 0, 1, \ldots, n$; $h_n > 0$ is called the discretization step. Put $T = n h_n$, for simplicity. Let the sample size n increase unboundedly and consider again the problem of estimating the parameter θ of the drift.

The first approach to this problem is based on the straightforward discretization of the continuous likelihood (31). We simply substitute the Lebesgue integral by the corresponding Riemann sum, and the stochastic integral by the corresponding Itôsum. When the drift coefficient b is differentiable with respect to θ, we obtain the approximate score function

$$\tilde{l}_n(\theta) = \sum_{i=1}^{n} \frac{\dot{b}(X_{t_{i-1}^n}, \theta)}{\sigma^2(X_{t_{i-1}^n})} [X_{t_i^n} - X_{t_{i-1}^n} - h_n b(X_{t_{i-1}^n}, \theta)]. \tag{34}$$

We can now define the estimator $\tilde{\theta}_n$ as the zero of the estimating function $\theta \mapsto \tilde{l}_n(\theta)$. Asymptotic properties of the estimator strongly depend on the behaviour of the discretization step h_n as $n \to \infty$. As is shown in FLORENS–ZMIROU (1989) the estimator is consistent if $h_n \to 0$ and $T = n h_n \to \infty$ as $n \to \infty$. If, in addition, $n h_n^2 \to 0$, then the estimator is also asymptotically normal and asymptotically efficient as $T = n h_n \to \infty$ in the sense that

$$\mathscr{L}\{\sqrt{n h_n}(\tilde{\theta}_n - \theta)|\mathbb{P}_\theta\} \to N(0, I(\theta)^{-1}) \tag{35}$$

where $I(\theta)$ is the Fisher information (32); cf. (33). However, when h_n is bounded away from 0 as $n \to \infty$, the estimator is not even consistent.

3.1.3 *Method of Bibby and Sørensen*

As was mentioned in the preceding section, if the discretization step is independent of the sample size, i.e. $h_n \equiv h$, the estimation function (34) does not yield a consistent estimator. In this situation BIBBY and SØRENSEN suggest in BIBBY and SØRENSEN (1995) an adjustment of (34) to get a zero-mean martingale as a renewed

scoring function, and to remove thus the estimation bias. Departing in this manner from the likelihood function (31) and the corresponding score (34), Bibby and Sørensen allow the parameter θ also to appear in the diffusion coefficient σ. Thus the SDE considered in BIBBY and SØRENSEN (1995) is given by

$$dX_t = b(X_t, \theta) \, dt + \sigma(X_t, \theta) \, dW_t, \quad X_0 = x, \tag{36}$$

with the parameter θ to be estimated from the observations $\{X_{t_i}\}_{i=0,1,\ldots,n}$ at equidistant instants $t_i = ih$. This results in an estimating function of the following form:

$$\tilde{G}_n(\theta) = \sum_{i=1}^{n} \frac{\dot{b}(X_{t_{i-1}}, \theta)}{\sigma^2(X_{t_{i-1}}, \theta)} [X_{t_i} - m(X_{t_{i-1}}, \theta)],$$

where

$$m(x, \theta) = \mathbb{E}_\theta[X_h | X_0 = x]. \tag{37}$$

As it is well known, this conditional moment is related to the drift coefficient in the follow manner:

$$m(x, \theta) = x + hb(x, \theta) + o(h). \tag{38}$$

Its explicit expression, however, is often unknown, in which case one cannot utilize the estimating function $\tilde{G}_n$. We will turn back to this situation later.

Obviously, $\tilde{G}_n$ is one particular choice out of the entire class of martingale estimating functions of the form

$$G_n(\theta) = \sum_{i=1}^{n} g_{i-1}(\theta) [X_{t_i} - m(X_{t_{i-1}}, \theta)] \tag{39}$$

where $g_{i-1}(\theta)$ is a certain function of observations $X_0, \ldots, X_{t_{i-1}}$, depending on the unknown parameter θ. Under certain regularity conditions, these estimating functions define a class of consistent and asymptotically normal estimators, see BIBBY and SØRENSEN (1995). Furthermore, they show that there is an optimal estimating function within the class (39) that yields the smallest possible asymptotic confidence interval around θ. By using arguments of HEYDE (1997) (cf. also DZHAPARIDZE and SPREIJ, 1993), this optimal estimating function is constructed as follows:

$$G_n^*(\theta) = \sum_{i=1}^{n} \frac{\dot{m}(X_{t_{i-1}}, \theta)}{m_2(X_{t_{i-1}}, \theta)} [X_{t_i} - m(X_{t_{i-1}}, \theta)] \tag{40}$$

where

$$m_2(x, \theta) = \mathbb{E}_\theta[(X_h - m(x, \theta))^2 | X_0 = x] \tag{41}$$

is the conditional variance. It is related to the diffusion coefficient as follows:

$$m_2(x, \theta) = h\sigma^2(x, \theta) + o(h). \tag{42}$$

But its explicit expression is often unknown and the corresponding optimal estimator θ_n^*, the zero of (40), is impossible to find. This situation will be discussed in sections 3.2.2 and 3.2.3.

3.2 *Inference based on the discrete likelihood*

3.2.1 *Transition probabilities*

Suppose meanwhile that not only the first two conditional moments (37) and (41) but the conditional distribution itself is known. That is the situation when necessary growth conditions on b and σ are satisfied so that the solution of the SDE (36) is a Markov process (see KARATZAS and SHREVE, 1991, section 5.4.C) which is characterized by the transition densities $p(t, x, y, \theta)$ known up to the parameter θ to be estimated from observations $\{X_{T_i}\}_{i=0,1,\ldots,n}$ at certain instants $0 = T_0 < T_1 < \cdots < T_n$. We have for all $t \geqslant 0$, $x \in \mathbb{R}$ and $B \in \mathscr{B}(\mathbb{R})$:

$$\mathbb{P}(X_t \in B | X_0 = x) = \int_B p(t, x, y, \theta) \, dy.$$

Given $X_0 = x_0$, the vector $(X_{T_1}, \ldots, X_{T_n})$ has the product density

$$(x_1, \ldots, x_n) \mapsto p(T_1 - T_0, x_0, x_1, \theta) \cdots p(T_n - T_{n-1}, x_{n-1}, x_n, \theta).$$

Hence the MLE $\hat{\theta}_n$ for the parameter θ renders the likelihood

$$\theta \mapsto L_n(\theta) := \prod_{i=1}^n p(T_i - T_{i-1}, X_{T_{i-1}}, X_{T_i}, \theta) \tag{43}$$

as large as possible. To establish the consistency, asymptotic normality and asymptotic efficiency of this estimator one can apply the classical theory of statistical inference about ergodic Markov chains (see BILLINGSLEY, 1961), by checking requirements on transition densities in terms of b and σ. For this purpose DACUNHA-CASTELLE and FLORENS-ZMIROV (1986) study the relationship between the diffusion X and the embedded Markov chain $\{X_{T_i}\}_{i=0,1,\ldots,n}$.

3.2.2 *Pedersen's method*

Consider the same estimation problem as in the preceding section, but now in a more realistic situation in which the transition densities of the process X are unknown. The matter becomes much more complicated. However, there are some possibilities to treat it. In this section we will describe one particular approach proposed by PEDERSEN (1995a, 1995b) and in the next section a different approach by KESSLER (1997).

The theoretical part of PEDERSEN's (1995a) paper consists of constructing a certain 'approximate log-likelihood', an approximation to the log-likelihood $l_n(\theta) = \log L_n(\theta)$ of the previous section, cf. (43). This is based on the method developed by KLOEDEN and PLATEN (1992) for constructing approximations to the transition densities $p(t, x, y, \theta)$. The procedure begins with discretizing the SDE (36) accord-

ing to the so-called *Euler scheme* as follows: for fixed $N \geqslant 1$, define $t_k = k(t/N)$, for $k = 0, \ldots, N$. Then the process $Y^N = \{Y^N_{t_k}\}_{k=0,\ldots,N}$ is defined recurrently by

$$Y^N_0 := x, \qquad Y^N_{t_{k+1}} := Y^N_{t_k} + b(Y^N_{t_k}, \theta)\frac{t}{N} + \sigma(Y^N_{t_k}, \theta)(W_{t_{k+1}} - W_{t_k}). \qquad (44)$$

The process Y^N is called *Euler approximation* to the process $\{X_s\}_{s \leqslant t}$, under $\mathbb{P}_\theta = \mathbb{P}_{\theta,x}$, see KLOEDEN and PLATEN (1992). Taking into consideration (44), we see that the random variable $Y^N_t = Y^N_{t_N}$ has a density $y \mapsto p_N(t, x, y, \theta)$ with respect to the Lebesgue measure, given by

$$p_1(t, x, y, \theta) = \phi\left(\frac{y - x - tb(x, \theta)}{\sqrt{T}\sigma(x, \theta)}\right) \qquad (45)$$

and

$$p_N(t, x, y, \theta) = \mathbb{E}_{\theta,x}\, p_1\left(\frac{t}{N}, Y^N_{t_{N-1}}, y, \theta\right), \ N \geqslant 2, \qquad (46)$$

where ϕ is the standard normal density. These $p_N(t, x, y, \theta)$ are the desired approximations to the respective transition densities $p(t, x, y, \theta)$. Furthermore, the corresponding approximate log-likelihood function

$$l^N_n(\theta) = \sum_{i=1}^n \log p_N(T_i - T_{i-1}, X_{T_{i-1}}, X_{T_i}, \theta) \qquad (47)$$

is such that $l^N_n(\theta) \to l_n(\theta)$ in $\mathbb{P}_{\theta,x}$-probability, as $N \to \infty$; see PEDERSEN (1995a) for more details.

Surely, this is not enough to guarantee nice asymptotic properties of the approximate MLE, a maximizer of (47), which should share the asymptotic properties of the MLE of the previous section. Necessary arguments towards this conclusion can be found in PEDERSEN (1995b) though in this part of his work only some particular models are rigorously treated.

Finally, note that in practice we need to evaluate $l^N_n(\theta)$, hence expressions of type $p_N(t - s, X_s, X_t, \theta)$ have to be evaluated. For $N = 1$ the expression (45) is explicit, so the problem arises for $N \geqslant 2$. But in the latter case we have (46). Taking this into consideration, Pedersen suggests to simulate according to (44) a large number of independent observations $\{Y_m\}_{m=1,\ldots,M}$ of $Y^N_{t_{N-1}}$ and then to form the average

$$\frac{1}{M}\sum_{m=1}^M p_1\left(\frac{t}{N}, Y_m, y, \theta\right)$$

that approaches, with increasing M, the desired expectation on the right hand side of (46). Surely, this procedure is quite demanding numerically, since at every evaluation of the function $\theta \mapsto l^N_n(\theta)$ the simulations are anew carried out.

3.2.3 *Kessler's method*

Unlike the previous sections, we now consider the diffusion model with the drift and diffusion coefficients b and σ depending on two different parameters θ and η, respectively. In KESSLER (1997) the method of estimating these parameters $\alpha := (\theta, \eta)$ is suggested by maximizing a certain contrast functional of observations, that is a functional of the discretized trajectory $\{X_{t_i^n}\}_{0 \leqslant i \leqslant n}$ at equidistant instants $t_i^n = ih_n$.

As in the previous section, the transition densities of the consecutive observations are supposed to be unspecified. This means, of course, that the first two conditional moments

$$m(X_{t_{i-1}^n}, \alpha) = \mathbb{E}_\alpha[X_{t_i^n}|X_{t_{i-1}^n}] \tag{48}$$

and

$$m_2(X_{t_{i-1}^n}, \alpha) = \mathbb{E}_\alpha[(X_{t_i^n} - m(X_{t_{i-1}^n}, \alpha))^2|X_{t_{i-1}^n}] \tag{49}$$

are unspecified as well, cf. (37) and (41). If these moments were specified, one would be able to utilize the following contrast function:

$$\sum_{i=1}^n \left[\frac{(X_{t_i^n} - m(X_{t_{i-1}^n}, \alpha))^2}{m_2(X_{t_{i-1}^n}, \alpha)} + \log m_2(X_{t_{i-1}^n}, \alpha) \right]. \tag{50}$$

The latter may be interpreted as a log-likelihood function, provided the unknown transition densities are approximated by Gaussian densities with the parameters (48) and (50).

In the case of unknown m and m_2, Kessler suggests to substitute in (50) their closed approximations. For instance, take into consideration (38) and (42) which yield the following approximations

$$m(x, \alpha) \simeq x + h_n b(x, \theta) \tag{51}$$

and

$$m_2(x, \alpha) \simeq h_n \sigma^2(x, \eta). \tag{52}$$

Upon this substitution, we get the contrast

$$\sum_{i=1}^n \left[\frac{(X_{t_i^n} - X_{t_{i-1}^n} - h_n b(X_{t_{i-1}^n}, \theta))^2}{h_n \sigma^2(X_{t_{i-1}^n}, \eta)} + \log h_n \sigma^2(X_{t_{i-1}^n}, \eta) \right]. \tag{53}$$

The estimators $\tilde{\alpha}_n = (\tilde{\theta}_n, \tilde{\eta}_n)$ for the parameters $\alpha = (\theta, \eta)$ are then obtained by maximizing the contrast (53). Note that with η fixed, the maximization of (53) with respect to θ leads to the same estimator $\tilde{\theta}_n$ as in section 3.1.2, defined as the zero of (34).

Suppose now that the sample size n increases unboundedly, while the discretization step h_n tends to 0 in such a way that $nh_n \to \infty$. The situation then is similar to that of section 3.1.2 and the maximum contrast estimators $\tilde{\alpha}_n = (\tilde{\theta}_n, \tilde{\eta}_n)$ are consistent, i.e. they converge in probability to the true value of the estimated parameters. If, in

addition, $nh_n^2 \to 0$, then the estimators are asymptotically normal and asymptotically efficient in the sense that $\sqrt{nh_n}(\tilde{\theta}_n - \theta)$ and $\sqrt{n}(\tilde{\eta}_n - \eta)$ are asymptotically independent and distributed according to $N(0, I(\theta)^{-1})$ and $N(0, I(\eta)^{-1})$, respectively. As usual $I(\theta)$ and $I(\eta)$ are the corresponding Fisher information quantities: $I(\theta)$ is given by (32), while

$$I(\eta) = \frac{1}{2} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \eta} \log \sigma^2(x, \eta) \right)^2 \pi_\theta(\mathrm{d}x). \tag{54}$$

For the exact formulation of this result, see KESSLER (1997), theorem 1 on p. 216, and remark 2 on the same page concerning the Fisher information quantity (54); cf. also GENON-CATALOT and JACOD (1993). This is in fact a simple particular case of Kessler's theorem (which is in full agreement with the discussion in section 3.1.2 – recall (35)). The general assertion is much more complicated. We do not want to carry the reader too far afield, therefore we restrict ourselves to a few remarks of a general nature.

Note first that the condition $nh_n^2 \to 0$ tells us how frequent the observations should be. If the observations are less frequent so that only a less restrictive condition $nh_n^p \to 0$ with some $p > 2$ can be satisfied, then the contrast (53) fails and one has to construct finer approximation to (50). This construction, based on further expansion of m and m_2 in powers of h_n (the above approximations (51) and (52) involve only the first order term), is quite cumbersome to be presented here; cf. KESSLER (1997). We only note that the following result by FLORENS-ZMIROU (1989), lemma 1, plays a key rôle. Let L_α be the generator of the diffusion defined by

$$L_\alpha f(x) = b(x, \theta) \frac{\partial f}{\partial x}(x) + \frac{1}{2} \sigma^2(x, \eta) \frac{\partial^2 f}{\partial x^2}(x)$$

and L_α^k its kth iterate. Besides, let L_α^0 be the identity operator. Then we get the following power series expansion up to the power h_n^l say, of the conditional moments: for a sufficiently smooth function f

$$\mathbb{E}_\alpha[f(X_{t_i^n})|X_{t_{i-1}^n}] \simeq \sum_{j=0}^{l} \frac{h_n^j}{j!} L_\alpha^j f(X_{t_{i-1}^n}) \tag{55}$$

(for the remainder term, see FLORENS-ZMIROU (1989), formula (2.12)). Take, for instance, $f(x) = x$ to get by (48) and (55) the desired expansion

$$m(x, \alpha) \simeq \sum_{j=0}^{l} \frac{h_n^j}{j!} L_\alpha^j f(x)$$

where $L_\alpha^0 f(x) = x$, $L_\alpha^1 f(x) = b(x, \theta)$, $L_\alpha^2 f(x) = b(x, \theta)\dot{b}(x, \theta) + \frac{1}{2}\sigma^2(x, \eta)\ddot{b}(x, \theta)$ and so on.

4 Inference for partially observed diffusion processes

4.1 *Inference for continuous observations*

Consider the situation where one deals with a bivariate diffusion process with only one observable component. Then one has first to replace the semimartingale decomposition of the bivariate diffusion that is given w.r.t. some underlying filtration with the decomposition that is associated with the filtration generated by the observed component. The specification of the drift and diffusion coefficients in this decomposition (call them the filtered drift and filtered diffusion coefficients) belongs to the realm of (non-linear) filtering theory, see LIPTSER and SHIRYAYEV (1977). The technical problem one is immediately faced with is the computation of these coefficients. This is only possible in rare circumstances of which the case, where one observes a bivariate Gaussian process that is the solution of a linear stochastic differential equation, is the most important. In this case the Kalman-Bucy filter applies, although still closed form solutions are not available.

If one wants to estimate parameters, an obvious next step is to give an expression of the likelihood functional. This expression is similar to (31) and contains the filtered drift and diffusion coefficients, so a maximum likelihood approach is not straightforward.

In the next subsection we will focus our attention to the seemingly even more complex situation in which the observed process is known at discrete time points only. However, under conditions that ensure ergodicity, solutions to this inference problem exist.

4.2 *Inference for discrete observations*

In this subsection we continue to investigate estimation problems for discretely observed diffusions. The difference with subsections 3.1 and 3.2 is that we allow the diffusion coefficient (with the interpretation of volatility) of the observation process to be a stochastic process itself, again a diffusion and moreover that this volatility process itself is unobserved. So we find ourselves in the framework described at the beginning of this section.

In this part of the paper we focus on recent contributions by GENON-CATALOT et al. (1998, 1999) to estimate parameters in such a situation. Consider the following stochastic volatility model. For the log price process S we assume

$$\mathrm{d}S_t = \sigma_t \, \mathrm{d}B_t, \; S_0 = 0,$$

whereas

$$\mathrm{d}V_t = b(V_t, \theta) \, \mathrm{d}t + a(V_t, \theta) \, \mathrm{d}W_t,$$

with $V_t = \sigma_t^2$, describes the evolution of the volatility. Here B and W are assumed to be independent Brownian motions. Notice that this model is not capable of incorporating the leverage effect discussed in section 2; for this one needs correlated Brownian motions.

We moreover assume that V is an ergodic diffusion on $(0, \infty)$, with a stationary

density π_θ. The aim is to estimate the parameter θ based on the observations $S_0, S_h, \ldots, S_{nh}$.

Before we are going to discuss how to approach this problem, we wish to make a remark concerning the filtering problem that we briefly touched upon. Suppose, contrary to what we stated above, we would have a continuous record of observations from S_t, whereas V_t is still unobserved. In filtering one of the issues is to find a stochastic differential equation for the conditional density (assuming that it exists) of V_t given S_τ, $\tau \in [0, t]$. Now, since S is a Brownian martingale with $(d/dt)\langle S \rangle = V_t$, it follows that V_t is measurable w.r.t. the σ-algebra generated by S_τ, $\tau \in [0, t]$. In other words, along with S we also observe V. Hence the filtering problem in this situation becomes degenerate and for the parametric inference problem we can in principle use the methods of subsection 3.1.1. The situation with discrete observations of S that we introduced above is entirely different. There are interesting aspects of filtering problems, see e.g. BRIGO and HANZON (1998) for a recent contribution. In the present subsection filtering will not be considered anymore and we turn back to the parameter estimation problem outlined above.

Define for $k = 1, \ldots, n$ the random variables

$$X_k := \frac{1}{\sqrt{h}}(S_{kh} - S_{(k-1)h}) = \frac{1}{\sqrt{h}}\int_{(k-1)h}^{kh} \sigma_t \, dB_t.$$

Then, given $\mathscr{F}_\infty^V$, we find that X_k has a $N(0, \overline{V}_k)$ distribution, with

$$\overline{V}_k = \frac{1}{h}\int_{(k-1)h}^{kh} V_t \, dt.$$

Hence it holds for the characteristic function of X_k that

$$\mathbb{E}e^{iuX_k} = \mathbb{E}\mathbb{E}[e^{iuX_k}|\mathscr{F}_\infty^V] = \mathbb{E}e^{-\frac{1}{2}u^2\overline{V}_k}, \quad \forall u \in \mathbb{R}.$$

From this we see that for small h it holds that $\overline{V}_k \approx V_{kh}$ and hence we find that the characteristic function of X_k is approximately equal to

$$\mathbb{E}e^{-\frac{1}{2}u^2 V_{kh}} = \int_{\mathbb{R}_+} e^{-\frac{1}{2}u^2 v}\pi_\theta(v) \, dv,$$

which is just the characteristic function of $N\sqrt{V_0}$, where N stands for a $N(0, 1)$ distributed random variable, independent of V_0. For the density q_θ of $N\sqrt{V_0}$ it holds that

$$q_\theta(x) = \int_{\mathbb{R}_+} \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{x^2}{2v}\right)\pi_\theta(v) \, dv, \tag{56}$$

which yields the observation that X_k has approximately an absolutely continuous distribution with density q_θ, if h is small. Concerning the sampling time instants we make the following assumption.

$$h = h_n \to 0 \quad \text{and} \quad nh_n \to \infty. \tag{57}$$

Then it is shown in GENON-CATALOT et al. (1998) by means of an ergodic theorem that under certain regularity conditions

$$\frac{1}{n}\sum_{k=1}^{n} \phi(X_k) \xrightarrow{\mathrm{P}^\theta} \int \phi(x)q_\theta(x)\,\mathrm{d}x \tag{58}$$

for all ϕ in a class of (vector-valued) functions.

If one imposes moreover that $nh_n^2 \to 0$, then one can show using a central limit theorem for ergodic processes, that under $\mathbb{P}^\theta$

$$\sqrt{nh_n}\left(\frac{1}{n}\sum_{k=1}^{n} \phi(X_k) - \int \phi(x)q_\theta(x)\,\mathrm{d}x\right) \xrightarrow{\mathscr{L}} N(0, \Sigma) \tag{59}$$

where the covariance matrix Σ depends in a certain way on ϕ and π_θ (see GENON-CATALOT et al. (1999) for the relevant formulae).

These asymptotic results led GENON-CATALOT et al. (1999) to studying M-estimators (also referred to as contrast estimators) that maximize a criterion of the type

$$M_n(\theta) = \frac{1}{n}\sum_{k=1}^{n} m(\theta, X_k)$$

or, alternatively, are zeros of such a criterion. The results of (58) and (59) imply that these estimators enjoy desirable asymptotic properties like consistency and asymptotic normality for $h_n \to 0$, $nh_n \to \infty$ and $nh_n^2 \to 0$.

For example, based on the fact that X_k approximately has density q_θ for small h, Genon-Catalot et al. consider the estimator $\hat{\theta}_n$ that maximizes

$$\theta \mapsto \frac{1}{n}\sum_{k=1}^{n} \log q_\theta(X_k). \tag{60}$$

In this case, the criterion function depends on the parameter θ only through the stationary density π_θ (see equation (56)). This implies in particular that only the parameters that are evolved in π_θ can be identified by this criterion. Genon-Catalot et al. apply this for instance to the model described in section 2.3 by equations (27) and (28), with $c = 0$, $b_1 > 0$ and $2b_2 + \delta^2 > 0$. As is mentioned in section 2.3, in this case the stationary distribution of V is inverse gamma with parameters $\theta_1 = 1 + 2b_2/\delta^2$ and $\theta_2 = 2b_1/\delta^2$. In terms of the parameter $\theta = (\theta_1, \theta_2)$ the criterion (4.5) then takes the form

$$\theta \mapsto \frac{1}{n}\sum_{k=1}^{n}\left[\theta_1 \log\theta_2 - \log\frac{\Gamma(\theta_1)}{\Gamma(\theta_1 + \frac{1}{2})} - \left(\theta_1 + \frac{1}{2}\right)\log\left(\theta_2 + \tfrac{1}{2}X_k^2\right)\right].$$

If (57) holds as $n \to \infty$, the maximizer $\hat{\theta}_n$ of this expression is a consistent estimator of θ. If moreover $nh_n \to 0$ as $n \to \infty$, the estimator is asymptotically normal.

In general an analytic expression for q_θ will not be available. But, since q_θ depends on π_θ through (56) and π_θ depends on the known functions $b(\cdot, \theta)$ and $a(\cdot, \theta)$ through (5), numerical integration and optimization can in principle be carried out. The same remark applies to the alternative approach outlined below.

Another approach is to estimate the moments of the stationary distribution. Observe that it follows from equation (56) that under the condition that π_θ has a finite moment of order α

$$\int x^{2\alpha} q_\theta(x)\,\mathrm{d}x = C_{2\alpha} D_\theta(\alpha),\tag{61}$$

where $C_{2\alpha}$ is the moment of order 2α of the $N(0, 1)$ distribution and $D_\theta(\alpha)$ the moment of order α of π_θ. Then, if $n \to \infty$ (57) it follows from (58) that

$$\frac{1}{nC_{2\alpha}} \sum_{k=1}^{n} X_k^{2\alpha} \xrightarrow{\mathrm{P}^\theta} D_\theta(\alpha),$$

and consequently

$$\hat{D}(\alpha)_{n,h} := \frac{1}{nC_{2\alpha}} \sum_{k=1}^{n} X_k^{2\alpha}$$

is a consistent estimator of $D_\theta(\alpha)$, that is moreover asymptotically normal under the condition that $nh_n^2 \to \infty$. Because (58) and (59) are valid for vector-valued functions ϕ, one is able to estimate several moments of π_θ simultaneously with consistent, asymptotically normal estimators. If the parameter θ happens to be a smooth function of some moments of π_θ, then we can apply the delta-method to find a consistent, asymptotically normal estimator of θ if we define $\hat{\theta}$ as the same smooth function of the sample moments.

5   Outside the Brownian world

As we have already mentioned in section 2, the Black–Scholes–Samuelson framework has been proved to be too rigid to account for various phenomena that have been observed in practice. In this section we focus on one of the alternatives/ relaxations of this model, although it should be said that there are plenty of alternatives. Not only in stochastic modeling (stable processes, long range dependence, fractionally integrated processes), but also in deterministic modeling (dynamical systems, chaos).

The idea that we follow in this section, is to replace the Brownian motion with another process, or rather a member of a class of processes of which Brownian Motion is a member too, the Lévy processes. We like to point out at this stage that the Lévy process that is going to be used (the hyperbolic Lévy motion) is not a generalization of Brownian motion, since it has no Gaussian component, as we shall see later on.

Consider again the stock price evolution according to the SDE that describes S as a geometric Brownian motion:

$$\mathrm{d}S_t = S_t(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t).$$

Alternatively, we have that $\log(S_t/S_s)$ follows a normal distribution with mean $(\mu - \frac{1}{2}\sigma^2)(t - s)$ and variance $\sigma^2(t - s)$ for all $t > s$, the log-returns are normally distributed. Many attempts have been made to model these returns by means of other distributions, e.g. Pareto distributions as suggested by MANDELBROT (1963). These attempts were not always successful, but recently a new class of distributions has enjoyed some popularity and promising results have been reported. This class is that of the hyperbolic distributions and was introduced by Barndorff–Nielsen in a context (turbulence theory) that is completely different from financial modeling (see e.g. BARNDORFF–NIELSEN (1986) In EBERLEIN and KELLER (1995) have made an attempt to use this class to fit financial data. In the sequel we outline their approach.

First we have to observe that typical path properties of Brownian motion are invariant under time and space scaling. Hence Eberlein and Keller argue, that if one uses Brownian motion as a vehicle to model stock prices, we should observe a similar invariance under change of time scale. In particular, they looked at intradaily data, and what one observes then is (not surprisingly of course) a piecewise constant behaviour of stock prices. This is a first indication to use jump processes for modeling.

Eberlein and Keller used daily BASF and Deutsche Bank data to test the normality hypothesis by different methods. Qualitatively, QQ-plots and density plots already showed a large discrepancy between the empirical and fitted normal distributions. But also χ^2 tests and tests based on kurtosis and skewness rejected the normal hypothesis at very small significance levels. Therefore we have to look for alternatives, for instance α-stable distributions. As a reminder we mention that symmetric α-stable distributions are those whose characteristic functions ϕ are of the form $\phi(t) = \exp(-(c|t|)^\alpha)$, with $c > 0$ and $0 < \alpha \leq 2$. Notice that for $\alpha = 2$ we get a normal and for $\alpha = 1$ we get a Cauchy distribution. A survey of modeling with Pareto and stable distributions and distributions with other stability properties can be found in MITTNICK and RACHEV (1993).

One of the tests to see whether stable distributions are capable of explaining data is based on the following observation. If $X_1, \ldots, X_n$ are independent drawings from the same α-stable distribution, then also linear combinations $\sum_{i=1}^{n} a_i X_i$ have a stable distribution with same coefficient α. What several authors have done is to base tests on estimators of α that use sums of data of different lengths. It turned out that the estimators of α had a tendency to grow with the sum length and quite often this resulted in numerical values close to 2 for relatively moderate lengths, indicating that (at least from this point of view) normality seems to be a reasonable assumption for modeling data over longer time intervals.

We now turn to hyperbolic distributions. From the observation that the log density

of the normal distribution is a parabola, the name of the hyperbolic distribution stems from the fact that the log density is a hyperbola. Explicitly, the formula for this density is given by

$$\frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})}\exp(-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)).$$

In this expression K_1 is the so-called modified Bessel function of the third kind with index 1 (a more precise description is not relevant for the present paper). The parameters α and β satisfy the restriction $\alpha > \beta$ and are called the *shape* parameters. The δ and μ are the location and scale parameters. One often uses a different parametrization by leaving μ and δ unchanged and by using $\xi = (1 + \delta\sqrt{\alpha^2 - \beta^2})^{-1/2}$ and $\chi = \xi\beta/\alpha$. This has the advantage that the new shape parameters ξ and χ are location and scale invariant. They satisfy the restrictions $|\chi| < \xi < 1$, describing the so called 'shape triangle'. It follows from properties of the Bessel function and the characteristic function of the hyperbolic distribution that in the limit situations corresponding to the cases $\xi \to 0$, $\xi \to 1$, $\chi \to \pm\xi$ one gets the normal distribution, the Laplace distribution and the generalized inverse Gaussian distribution. This shows that hyperbolic distributions form a rather flexible class of distributions, which already (partly) explains why good data fit has been found in a number of circumstances.

For future use we mention that hyperbolic distributions can be obtained as mixtures of normals, where the mixing distribution (a generalized inverse Gaussian distribution) influences both mean and variance of the normals.

Estimation of the parameters using the same data set that has been reported above yielded the following conclusions. The estimated hyperbolic distribution was nearly symmetric and the shape parameters were far from the values that correspond to normality. The QQ and density plots corresponding to the fitted hyperbolic distribution showed a remarkable fit and using chi-square and Kolmogorov–Smirnov tests the hypothesis that the underlying distribution was a hyperbolic one turned out to be acceptable as well. Besides, larger data sets were used to carry out similar estimation procedures and one of the conclusions was that for aggregation (or increasing the time lags over which the returns were calculated) the fitted hyperbolic distribution was close to a normal one (estimated ξ and χ close to zero), in agreement with the estimation results for the stability index of a stable distribution under temporal aggregation that we mentioned above.

However one must take care with fitting hyperbolic (or other light-tailed) distributions, especially when outliers have been observed, possibly caused by the fact that the underlying distribution is heavy-tailed. In this case one still often gets a good fit of a such a distribution, especially when it has many parameters, in the center of the observations, but in the tails it may be poor.

Having established the fact that hyperbolic distributions form a flexible class that showed a good fit to the data, the next step Eberlein and Keller undertook was to

come up with a dynamical model that explains a hyperbolic distribution of the data. Barndorff–Nielsen and Halgreen (see also HALGREEN (1979)) showed infinite divisibility of the generalized inverse normal distribution and this together with the characterization of a hyperbolic distribution as a mixture of normals with the generalized inverse normal distribution as the directing one yields that also the hyperbolic distributions are infinitely divisible. This property calls for the use of Lévy processes as a dynamical model.

To simplify things we will in view of the previously observed fit of a symmetric distribution henceforth assume that the hyperbolic distributions are centered around the origin, so $\beta = \mu = 0$. With $\zeta = \alpha\delta$ we get the following density

$$\mathrm{hyp}_{\zeta,\delta}(x) = \frac{1}{2\delta K_1(\zeta)} \exp\left(-\zeta\sqrt{1 + \frac{x^2}{\delta^2}}\right).$$

Consider now the Lévy process Z (which by definition has stationary and independent increments) that is such that $Z_0 = 0$ and such that Z_1 has the hyperbolic density $\mathrm{hyp}_{\zeta,\delta}$. The choice of $t = 1$ as the time instant to impose the hyperbolic distribution is motivated as follows. Observations come to us at equidistant times. Then we normalize the time intervals such that they have unit length. Hence all $Z_t - Z_{t-1}$ now have a hyperbolic distribution and for integers $t \neq s$ the random variables $Z_s - Z_{s-1}$ and $Z_t - Z_{t-1}$ are independent; in particular the conditional distribution of $Z_t - Z_{t-1}$ given $Z_1, \ldots, Z_{t-1}$ is the same as the unconditional one. Notice that we now also determined the distribution of a (finite) sequence $Z_1, \ldots, Z_n$ for any $n \in \mathbb{N}$.

The process Z is called the hyperbolic Lévy motion. It can be shown that Z_t has moments of all orders, whereas only the distribution of Z_1 belongs to the hyperbolic ones. Another feature of this Lévy process is that it has no Gaussian part, so it is a pure jump process.

It is possible (see EBERLEIN and KELLER (1995)) to give an expression for the density g of the Lévy measure. This expression is rather complicated in that it involves two more Bessel functions. But one can deduce that in a neighborhood of the origin $g(x)$ behaves like $1/x^2$, indicating that Z (like many other Lévy processes) has infinitely many small jumps in every time interval. This already shows that such a Lévy process is a natural candidate to model financial data that evolve in such a way that small changes frequently occur.

In principle there are two natural candidate models for stock price evolution based on a hyperbolic Lévy motion. The first model arises when we replace the Brownian motion in the Samuelson model by the hyperbolic Lévy motion, thus

$$\mathrm{d}S_t = \rho S_{t-}\, \mathrm{d}t + S_{t-}\, \mathrm{d}Z_t. \tag{62}$$

The problem with this equation is that log-returns over intervals of unit length do not have the hyperbolic distribution we started with and that S also takes on negative values. So we need to use another model that does not have these drawbacks. This happens for

$$S_t = S_0 \exp(\rho t + Z_t). \tag{63}$$

Eberlein and Keller suggest to use this model, but they immediately point out a drawback to use it. This model describes an incomplete market, in the sense that there is no unique equivalent martingale measure and hence a problem arises if one uses this model to price financial derivatives like options. In EBERLEIN and KELLER (1995) one particular equivalent martingale measure is constructed using so called Esscher transforms and a pricing formula for a European call option is given. In EBERLEIN and JACOD (1987) the whole set of possible option prices corresponding to the set of equivalent martingales measures is characterized.

Not only hyperbolic Lévy motion as a background driving process results in hyperbolic distribution of the returns. In BIBBY and SØRENSEN (1998) another model is proposed to get hyperbolic distributions. They assume that the price process is described by an ergodic diffusion, where the drift and diffusion coefficients are such that the invariant distribution is hyperbolic. To be explicit, we assume that

$$dS_t = b(S_t)\, dt + \sigma(S_t)\, dW_t.$$

Then as we mentioned in subsection 2.2, under certain stationarity and ergodicity conditions, the solution has an absolutely continuous stationary distribution with density proportional to

$$\frac{1}{s(x)\sigma^2(x)}. \tag{64}$$

What has to be done then, is to choose drift and diffusion terms in such a way that (64) is the density of a hyperbolic distribution up to a normalizing constant. This can be accomplished by choosing

$$b(x) = \frac{1}{2} v(x) \frac{d}{dx} \log(f(x)v(x)) \tag{65}$$

$$\sigma^2(x) = v(x), \tag{66}$$

where f is a function proportional to the density of a hyperbolic distribution and v is such that the conditions for ergodicity are satisfied (which happens if $v(x) = f(x)^\gamma$ for some $\gamma \in [0, 1]$). With the special choice $\gamma = 1$ and estimated parameters a good fit of a hyperbolic distribution to prices of VW stock was found in BIBBY and SØRENSEN (1998).

References

AÏT-SAHALIA, Y. (1996), Nonparametric pricing of interest derivative securities, *Econometrica* **64**, 527–560.

BACHELIER, L. (1900), Théorie de la spéculation. Reprinted in P. COOTNER (ed.), *The random character of stock market prices*, 17–78. Cambridge, MA: MIT Press (1964).

BARNDORFF–NIELSEN, O. E. (1986), Sand, wind and statistics, *Acta Mechanica* **64**, 1–18.

BARNDORFF–NIELSEN, O. E. and O. HALGREEN (1977), Infinite divisibility of the hyperbolic and

generalized inverse Gaussian distributions, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **38**, 309–312.

BIBBY, B. M. and M. SØRENSEN (1995), Martingale estimation functions for discretely observed diffusion processes, *Bernoulli* **1**, 17–39.

BIBBY, B. M. and M. SØRENSEN (1998), Simplified estimating functions for diffusion models with a high dimensional parameter, *working paper* 17, Centre for Analytical Finance, University of Aarhus.

BILLINGSLEY, P. (1961), *Statistical inference for Markov processes*, Wiley.

BILLINGSLEY P. (1968), *Convergence of probability measures*, Wiley.

BJöRK, T. (1998), *Arbitrage theory in continuous time*, Oxford University Press.

BLACK, F. (1976), Studies in stock price volatility change, in *Proceedings of the 1976 Business of the Business and Economics Statistics Selection, American Statistical Association*, 171–181.

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

BOLLERSLEV, T. (1986), Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* **31**, 307–321.

BOLLERSLEV, T., R. Y. CHOU and K. F. KRONER (1992), ARCH modelling in finance, *Journal of Econometrics* **52**, 5–59.

BOLLERSLEV, T., R. F. ENGLE and D. B. NELSON (1994), ARCH models, *Handbook of econometrics* Vol. IV, ch. 49, 2959–3038.

BOX, G. E. P. and G. M. JENKINS (1976), *Times series analysis, forecasting and control*, Wiley.

BRIGO, D. and B. HANZON (1998), On some filtering problems arising in mathematical finance, *Insurance Mathematics and Economics*, to appear.

COX, J., J. INGERSOLL and S. ROSS (1985), A theory of the term structure of interest rates, *Econometrica* **53**, 385–407.

DACUNHA-CASTELLE D. and D. FLORENS-ZMIROV (1986), Estimation of the coefficients of a diffusion from discrete observations, *Stochastics* **20**, 263–284.

DROST, F. C. and B. J. M. WERKER (1996), Closing the GARCH gap: continuous time GARCH modelling, *Journal of Econometrics* **74**, 31–57.

DUMAS, B., J. FLEMING and R. E. WHALEY (1995), Implied volatility functions: empirical tests, preprint Duke University.

DUPIRE, B. (1994), Pricing with a smile, *RISK* **7**, 1–28.

DZHAPARIDZE, K. and P. J. C. SPREIJ, (1993), On optimality of regular projective estimators in semimartingale models, *Stochastics* **43**, 161–178.

EBERLEIN E. and U. KELLER (1995), Hyperbolic distribution in finance, *Bernoulli* **1**, 281–299.

ENGLE, R. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, *Econometrica* **50**, 987–1008.

ETHIER, S. N. and T. G. KURTZ (1986), *Markov processes: characterization and convergence*, Wiley.

FLORENS-ZMIROV F. (1989), Approximate discrete-time schemes for statistics of diffusion processes, *Statistics* **20**, 547–557.

GENON-CATALOT, V., T. JEANTHEAU and C. LAREDO (1998), Limit theorems for discretely observed stochastic volatility models, *Bernoulli* **4**, 283–303.

GENON-CATALOT, V., T. JEANTHEAU and C. LAREDO (1999), Parameter estimation for discretely observed stochastic volatility models, *Bernoulli* **5**, 855–872.

GENON-CATALOT, V. and J. JACOD (1993), On the estimation of the diffusion coefficient for multidimensional diffusion processes, *Annales Henri Poincaré* **29**, 119–151.

GIHMAN, I. I. and A. V. SKOROHOD (1972), *Stochastic differential equations*, Springer.

GOURIÉROUX, C. (1997), *ARCH models and financial economics*, Springer.

HALGREEN, C. (1979), Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandke Gebiete* **47**, 13–17.

HESTON, S. L. (1993), A closed-form solution for options with stochastic volatility with applications to Bond and Currency options, *The Review of Finacial Studies* **6**, 327–343.

HEYDE, C. C. (1997), *Quasi-likelihood and its application*, Springer.

HULL, D. and A. WHITE (1987), The pricing of options on assets with stochastic volatilities, *Journal of Finance* **42**, 281–300.

EBERLEIN, E. and J. JACOD (1987), On the range of option prices, *Finance and Stochastics* **1**, 131–140.

JACOD, J. and A. N. SHIRYAYEV (1987), *Limit theorems for stochastic processes*, Springer.

KARATZAS, I. and S. E. SHREVE (1991), *Brownian motion and stochastic calculus*, 2nd Edition, Springer.

KESSLER, M. (1997), Estimation of an ergodic diffusion from discrete observations, *Scandinavian Journal of Statistics* **24**, 211–229.

KLOEDEN, P. E. and E. PLATEN (1992), *Numerical solution of stochastic differential equations*, Springer.

KUTOYANTS, Yu. A. (1984), *Parameter estimation for stochastic processes*, Heldermann.

LIPTSER, R. Sh. and A. N. SHIRYAYEV (1977), *Statistics of random processes I*, Springer.

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

MITTNIK, S. and S. T. RACHEV (1993), Modelling asset returns with alternative stable distributions, *Econometric Reviews* **12**, 261–330.

NELSON, D. B. (1990), ARCH models as diffusion approximations, *Journal of Econometrics* **45**, 7–38.

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

PEDERSEN, A. R. (1995a), A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scandinavian Journal of Statistics*, **22**, 55–71.

PEDERSEN, A. R. (1995b), Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes, *Bernoulli* **1**, 257–279.

ROGERS, L. C. G. and D. WILLIAMS (1997), *Diffusions, Markov processes and martingales*, Wiley.

SAMUELSON, P. (1965), Rational theory of warrant pricing, *Industrial Management Review* **6**, 13–32.

STROOCK, D. W. and S. R. S. VARADHAN (1979), *Multidimensional diffusion processes*, Springer.

VASICEK, O. (1977), An equilibrium characterization of the term structure, *Journal of Financial Economics* **5**, 177–188.

WIGGINS, J. B. (1987), Option valuation under stochastic volatility, *Journal of Financial Economics* **19**, 351–372.