Stochastic Volatility

Article *in* Handbook of Statistics · December 1995

DOI: 10.1016/S0169-7161(96)14007-4 · Source: RePEc

CITATIONS
714

READS
246

3 authors:

Eric Ghysels
University of North Carolina at Chapel Hill
313 PUBLICATIONS **15,431** CITATIONS
SEE PROFILE

Andrew C. Harvey
University of Cambridge
184 PUBLICATIONS **20,549** CITATIONS
SEE PROFILE

Eric michel Renault
The University of Warwick
176 PUBLICATIONS **7,380** CITATIONS
SEE PROFILE

Série Scientifique
Scientific Series

95s-49

Stochastic Volatility

Eric Ghysels, Andrew Harvey,
Eric Renault

Montréal
novembre 1995

CIRANO

Le CIRANO est une corporation privée à but non lucratif constituée en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de l'Industrie, du Commerce, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche. La *Série Scientifique* est la réalisation d'une des missions que s'est données le CIRANO, soit de développer l'analyse scientifique des organisations et des comportements stratégiques.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de l'Industrie, du Commerce, de la Science et de la Technologie, and grants and research mandates obtained by its research teams. The Scientific Series *fulfils one of the missions of CIRANO: to develop the scientific analysis of organizations and strategic behaviour.*

Les organisations-partenaires / The Partner Organizations

ISSN 1198-8177

Stochastic Volatility*

Eric Ghysels[†], *Andrew Harvey*[‡], *Eric Renault*[§]

Résumé / Abstract

Cet article, préparé pour le « Handbook of Statistics », vol. 14, Statistical Methods in Finance, passe en revue les modèles de volatilité stochastique. On traite les sujets suivants : volatilité des actifs financiers (volatilité instantanée des rendements d'actifs, volatilités implicites dans les prix d'options et régularités empiriques), modélisation statistique en temps discret et continu et enfin inférence statistique (méthodes de moments, pseudo-maximum de vraisemblance, méthodes bayesiennes et autres fondées sur la vraisemblance, inférence indirecte).

This paper, prepared for the " Handbook of Statistics ", vol.14, Statistical Methods in Finance, surveys the subject of Stochastic Volatility. The following subjects are covered : volatility in financial markets (instantaneous volatility of asset returns, implied volatilities in option prices and related stylized facts), statistical modelling in discrete and continuous time and finally statistical inference (methods of moments, Quasi-Maximum-Likelihood, Likelihood based and Bayesian Methods and Indirect Inference).

Mots Clés : rendements d'actifs financiers, hétéroscédasticité conditionnelle, prix d'option, modèle espace-état, processus de diffusion.

Keywords : Asset returns, Conditionnal heteroskedasticity, Option prices, State-Space models, Diffusion processus.

---

* We benefitted from helpful comments from Frank Diebold, René Garcia, Eric Jacquier and Neil Shephard on preliminary draft of the paper. The first author would like to acknowledge the financial support of FCAR (Québec), SSHRC (Canada) as well as the hospitality and support of CORE (Louvain-la-Neuve, Belgium). The second author wishes to thank the ESRC for financial support. The third author would like to thank the Institut Universitaire de France, the Fédération Française des Sociétés d'Assurance as well as CIRANO and C.R.D.E. for financial support. This text was prepared for *Handbook of Statistics*, Vol. 14 : Statistical Methods in Finance.

† C.R.D.E., Université de Montréal & CIRANO

‡ London School of Economics

§ GREMAQ et IDEI, Université des Sciences Sociales, Toulouse and Institut Universitaire de France

TABLE OF CONTENTS

1 Introduction

The class of stochastic volatility (SV) models has its roots both in mathematical finance and financial econometrics. In fact, several variations of SV models originated from research looking at very different issues. Clark (1973), for instance, suggested to model asset returns as a function of a random process of information arrival. This so-called time deformation approach yielded a time-varying volatility model of asset returns. Later Tauchen and Pitts (1983) refined this work proposing a mixture of distributions model of asset returns with temporal dependence in information arrivals. Hull and White (1987) were not directly concerned with linking asset returns to information arrival but rather were interesting in pricing European options assuming continuous time SV models for the underlying asset. They suggested a diffusion for asset prices with volatility following a positive diffusion process. Yet another approach emerged from the work of Taylor (1986) who formulated a discrete time SV model as an alternative to Autoregressive Conditional Heteroskedasticity (ARCH) models. Until recently estimating Taylor's model, or any other SV model, remained almost infeasible. Recent advances in econometric theory have made estimation of SV models much easier. As a result, they have become an attractive class of models and an alternative to other classes such as ARCH.

Contributions to the literature on SV models can be found both in mathematical finance and econometrics. Hence, we face quite a diverse set of topics. We say very little about ARCH models because several excellent surveys on the subject have appeared recently, including those by Bera and Higgins (1995), Bollerslev, Chou and Kroner (1992), Bollerslev, Engle and Nelson (1994) and Diebold and Lopez (1995). Furthermore, since this chapter is written for the *Handbook of Statistics*, we keep the coverage of the mathematical finance literature to a minimum. Nevertheless, the subject of option pricing figures prominently out of necessity. Indeed, section 2, which deals with definitions of volatility has extensive coverage of Black-Scholes implied volatilities. It also summarizes empirical stylized facts and concludes with statistical modelling of volatility. The reader with a greater interest in statistical concepts may want to skip the first three subsections of section 2 which are more finance oriented and start with section 2.4. Section 3 discusses discrete time models, while section 4 reviews continuous time models. Statistical inference of SV models is the subject of section 5. Section 6 concludes.

2 Volatility in Financial Markets

Volatility plays a central role in the pricing of derivative securities. The Black-Scholes model for the pricing of an European option is by far the most widely used formula even when the underlying assumptions are known to be violated. Section 2.1 will therefore take the Black-Scholes model as a reference point from which to discuss several notions of volatility. A discussion of stylized facts regarding volatility and option prices will appear next in section 2.2. Both sections set the scene for a formal framework defining stochastic volatility which is treated in section 2.3. Finally, section 2.4 introduces the *statistical* models of stochastic volatility.

2.1 The Black-Scholes Model and Implied Volatilities

More than half a century after the seminal work of Louis Bachelier (1900), continuous time stochastic processes have become a standard tool to describe the behavior of asset prices. The work of Black and Scholes (1973) and Merton (1990) has been extremely influential in that regard. In section 2.1.1 we review some of the assumptions that are made when modelling asset prices by diffusions, in particular to present the concept of instantaneous volatility. In section 2.1.2 we turn to option pricing models and the various concepts of implied volatility.

2.1.1 An Instantaneous Volatility Concept

We consider a financial asset, say a stock, with today's (time t) market price denoted by S_t.[1] Let the information available at time t be described by I_t and consider the conditional distribution of the return S_{t+h}/S_t of holding the asset over the period $[t, t+h]$ given I_t.[2] A maintained assumption throughout this chapter will be that asset returns have finite conditional expectation given I_t or :

$$E_t\left(S_{t+h}/S_t\right) = S_t^{-1} E_t S_{t+h} < +\infty \tag{2.1.1}$$

and likewise finite conditional variance given I_t, namely

$$V_t\left(S_{t+h}/S_t\right) = S_t^{-2} V_t S_{t+h} < +\infty \tag{2.1.2}$$

The continuously compounded expected rate of return will be characterized by $h^{-1}\log E_t\left(S_{t+h}/S_t\right)$. Then a first assumption can be stated as follows :

Assumption 2.1.1.A : The continuously compounded expected rate of return converges almost surely towards a finite value $\mu_S\left(I_t\right)$ when $h > 0$ goes to zero.
From this assumption one has $E_t S_{t+h} - S_t \sim h\mu_S\left(I_t\right) S_t$ or in terms of its differential representation :

$$\left.\frac{d}{d\tau} E_t\left(S_\tau\right)\right|_{\tau=t} = \mu_S\left(I_t\right) S_t \text{ almost surely} \tag{2.1.3}$$

where the derivatives are taken from the right. Equation (2.1.3) is sometimes loosely defined as : $E_t\left(dS_t\right) = \mu_S\left(I_t\right) S_t dt$. The next assumption pertains to the conditional variance and can be stated as :

Assumption 2.1.1.B : The conditional variance of the return $h^{-1} V_t\left(S_{t+h}/S_t\right)$ converges almost surely towards a finite value $\sigma_S^2\left(I_t\right)$ when $h > 0$ goes to zero.
Again, in terms of its differential representation this amounts to :

$$\left.\frac{d}{d\tau} Var_t\left(S_\tau\right)\right|_{\tau=t} = \sigma_S^2\left(I_t\right) S_t^2 \text{ almost surely} \tag{2.1.4}$$

and one loosely associates with the expression $V_t\left(dS_t\right) = \sigma_S^2\left(I_t\right) S_t^2 dt$.

[1] Here and in the remainder of the paper we will focus on options written on stocks or exchange rates. The large literature on the term structure of interest rates and related derivative securities will not be covered.

[2] Section 2.3 will provide a more rigorous discussion of information sets. It should also be noted that we will indifferently be using conditional distributions of asset prices S_{t+h} and of returns S_{t+h}/S_t since S_t belongs to I_t.

Both assumptions 2.1.1.A and B lead to a representation of the asset price dynamics by an equation of the following form :

$$dS_t = \mu_S(I_t)S_t dt + \sigma_S(I_t)S_t dW_t \tag{2.1.5}$$

where W_t is a standard Brownian Motion. Hence, every time a diffusion equation is written for an asset price process we have automatically defined the so-called instantaneous volatility process $\sigma_S(I_t)$ which from the above representation can also be written as :

$$\sigma_S(I_t) = \left[\lim_{h\downarrow o} h^{-1} V_t\left(S_{t+h}/S_t\right)\right]^{\frac{1}{2}} \tag{2.1.6}$$

Before turning to the next section we would like to provide a brief discussion of some of the foundations for the Assumptions 2.1.1.A and B. It was noted that Bachelier (1900) proposed Brownian Motion process as a model of stock price movements. In modern terminology this amounts to the random walk theory of asset pricing which claims that asset returns ought not to be predictable because of the informational efficiency of financial markets. Hence, it assumes returns on consecutive regularly sampled periods $[t+k,\, t+k+1]$, $k = 0, 2, ..., h-1$ are independently (identically) distributed. With such a benchmark in mind, it is natural to view the expectation and the variance of the continuously compounded rate of return $\log\left(S_{t+h}/S_t\right)$ as proportional to the maturity h of the investment.
Obviously we no longer use Brownian motions as a process for asset prices but it is nevertheless worth noting that Assumptions 2.1.1.A and B also imply that the expected rate of return and the associated squared risk (in terms of variance of the rate of return) of an investment over an infinitely-short interval $[t, t+h]$ is proportional to h. Sims (1984) provided some rationale for both assumptions through the concept of "local unpredictability".

To conclude, let us briefly discuss a particular special case of (2.1.5) predominantly used in theoretical developments and also highlight an implicit restriction we made. When $\mu_S(I_t) = \mu_S$ and $\sigma_S(I_t) = \sigma_S$ are constants for all t the asset price is a geometric Brownian motion. This process was used by Black and Scholes (1973) to derive their well-known pricing formula for European options. Obviously, since $\sigma_S(I_t)$ is a constant we no longer have an instantaneous volatility process but rather a single parameter σ_S - a situation which undoubtedly greatly simplifies many things including the pricing of options. A second point which needs to be stressed is that Assumptions 2.1.1.A and B allow for the possibility of discrete jumps in the asset price process. Such jumps are typically represented by a Poisson process and have been prominent in the option pricing literature since the work of Merton (1976). Yet, while the assumptions allow in principle for jumps, they do not appear in (2.1.5). Indeed, throughout this chapter we will maintain the assumption of sample path continuity and exclude the possibility of jumps as we focus exclusively on SV models.

2.1.2 Option Prices and Implied Volatilities

It was noted in the introduction that SV models originated in part from the literature on the pricing of options. We have witnessed over the past two decades a spectacular growth in options and other derivative security markets. Such markets are sometimes characterized as places where "volatilities are traded". In this section we will provide the rationale for such statements and study the relationship between so-called options implied volatilities and the concepts of instantaneous and averaged volatilities of the underlying asset return process.

The Black-Scholes option pricing model is based on a Log-Normal or Geometric Brownian Motion model for the underlying asset price:

$$dS_t = \mu_S S_t dt + \sigma_S S_t dW_t \tag{2.1.7}$$

where μ_S and σ_S are fixed parameters. A European call option with strike price K and maturity $t+h$ has a payoff:

$$[S_{t+h} - K]^+ = \begin{cases} S_{t+h} - K \text{ if } S_{t+h} \geq K \\ \\ 0 \text{ otherwise} \end{cases} \tag{2.1.8}$$

Since the seminal Black and Scholes (1973) paper, there is now a well established literature proposing various ways to derive the pricing formula of such a contract. Obviously, it is beyond the scope of this paper to cover this literature in detail.[3] Instead, the bare minimum will be presented here allowing us to discuss the concepts of interest regarding volatility.

With continuous costless trading assumed to be feasible, it is possible to form a portfolio using one call and a short-sale strategy for the underlying stock to eliminate all risk. This is why the option price can be characterized without ambiguity, using only arbitrage arguments, by equating the market rate of return of the riskless portfolio containing the call option with the risk-free rate. Moreover, such arbitrage-based option pricing does not depend on individual preferences[4].

This is the reason why the easiest way to derive the Black-Scholes option pricing formula is via a "risk-neutral world", where asset price processes are specified through a *modified probability measure*, referred to as the risk neutral probability measure denoted Q (as discussed more explicitly in section 4.2). This fictitious world where probabilities in general do not coincide with the Data Generating Process (DGP), is only used to derive the option price which remains valid in the objective probability setup. In the risk neutral world we have:

$$dS_t/S_t = r_t dt + \sigma_S dW_t \tag{2.1.9}$$

$$C_t = C\left(S_t, K, h, t\right) = B(t, t+h) E_t^Q (S_{t+h} - K)^+ \tag{2.1.10}$$

where E_t^Q is the expectation under Q, $B(t, t+h)$ is the price at time t of a pure discount bond with payoff one unit at time $t+h$ and

$$r_t = -\lim_{h \to 0} \frac{1}{h} Log\ B(t, t+h) \tag{2.1.11}$$

is the riskless instantaneous interest rate.[5] We have implicitly assumed that in this market interest rates are nonstochastic (W_t is the only source of risk) so that:

$$B(t, t+h) = exp\left[- \int_t^{t+h} r_\tau d_\tau \right]. \tag{2.1.12}$$

[3]See however Jarrow and Rudd (1983), Cox and Rubinstein (1985), Duffie (1989), Duffie (1992), Hull (1993) or Hull (1995) among others for more elaborate coverage of options and other derivative securities.

[4]This is sometimes refered to as *preference free option pricing*. This terminology may somewhat be misleading since individual preferences are implicitly taken into account in the market price of the stock and of the riskless bond. However, the option price only depends on individual preferences through the stock and bond market prices.

[5]For notational convenience we denote by the same symbol W_t a Brownian motion under P (in 2.1.7) and under Q (in 2.1.9). Indeed, Girsanov's theorem establishes the link between these two processes (see e.g. Duffie (1992) and section 4.2.1).

By definition, there are no risk premia in a risk neutral context. Therefore r_t coincides with the instantaneous expected rate of return of the stock and hence the call option price C_t is the discounted value of its terminal payoff $(S_{t+h} - K)^+$ as stated in (2.1.10).

The log-normality of S_{t+h} given S_t allows one to compute the expectation in (2.1.10) yielding the call price formula at time t:

$$C_t = S_t\phi(d) - KB(t, t+h)\phi(d - \sigma\sqrt{h}) \tag{2.1.13}$$

where ϕ is the cumulative standard normal distribution function while d will be defined shortly. Formula (2.1.13) is the so-called Black-Scholes option pricing formula. Thus, the option price C_t depends on the stock price S_t, the strike price K and the discount factor $B(t, t+h)$. Let us now define :

$$x_t = Log\ S_t/KB(t, t+h) \tag{2.1.14}$$

Then we have:

$$C_t/S_t = \phi(d) - e^{-x_t}\phi(d - \sigma\sqrt{h}) \tag{2.1.15}$$

with $d = \left(x_t/\sigma\sqrt{h}\right) + \sigma\sqrt{h}/2$. It is easy to see the critical role played by the quantity x_t, called the moneyness of the option.

- If $x_t = 0$, the current stock price S_t coincides with the present value of the strike price K. In other words, the contract may appear to be fair to somebody who would not take into account the stochastic changes of the stock price between t and $t+h$. We shall say that we have in this case an *at the money option.*
- If $x_t > 0$ (respectively $x_t < 0$) we shall say that the option is in the money (respectively out the money)[6].

It was noted before that the Black-Scholes formula is widely used among practitioners, even when its assumption are known to be violated. In particular the assumption of a constant volatility σ_S is unrealistic (see section 2.2 for empirical evidence). This motivated Hull and White (1987) to introduce an option pricing model with stochastic volatility assuming that the volatility itself is a state variable independent of W_t :[7]

$$\begin{cases} dS_t/S_t = r_t dt + \sigma_{St} dW_t \\ (\sigma_{St})_{t\in[0,T]}\,, (W_t)_{t\in[0,T]} \text{ independent Markovian} \end{cases} \tag{2.1.16}$$

It should be noted that (2..1.16) is still written in a risk neutral context since r_t coincides with the instantaneous expected return of the stock. On the other hand the exogenous volatility risk is not directly traded, which prevents us from defining unambiguously a risk neutral probability measure, as discussed in more detail in section 4.2. Nevertheless, the option pricing formula (2.1.10) remains valid provided the expectation is computed with respect to the joint probability distribution of the Markovian process (S, σ_S), given (S_t, σ_{St}).[8] We can then rewrite (2.1.10) as follows :

[6] We use here a slightly modified terminology with respect to the usual one. Indeed, it is more common to call at the money /in the money/ out of the money options, when $S_t = K/S_t > K/S_t < K$ respectively. From an economic point of view, it is more appealing to compare S_t with the present value of the strike price K.

[7] Other stochastic volatility models similar to Hull and White (1987) appear in Johnson and Shanno (1987), Scott (1987), Wiggins (1987), Chesney and Scott (1989), Stein and Stein (1991) and Heston (1993) among others.

[8] We implicitly assume here that the available information I_t contains the past values $(S_\tau, \sigma_\tau)_{\tau\leq t}$. This assumption will be discussed in section 4.2.

$$C_t = B(t,t+h)\,E_t(S_{t+h}-K)^+ = B(t,t+h)\,E_t\left\{E\left[(S_{t+h}-K)^+ \left| (\sigma_{S\tau})_{t\le\tau\le t+h}\right.\right]\right\} \tag{2.1.17}$$

where the expectation inside the brackets is taken with respect to the conditional probability distribution of S_{t+h} given I_t and a volatility path $\sigma_{S\tau}$, $t \le \tau \le t+h$. However, since the volatility process $\sigma_{S\tau}$ is independent from W_t, we obtain using (2.1.15) that :

$$B(t,t+h)\,E_t\left[(S_{t+h}-K)^+ \left| (\sigma_{S\tau})_{t\le\tau\le t+h}\right.\right] = S_t E_t\left[\phi(d_1) - e^{-x_t}\phi(d_2)\right] \tag{2.1.18}$$

where d_1 and d_2 are defined as follows :

$$\begin{cases} d_1 = \left(x_t/\gamma(t,t+h)\sqrt{h}\right) + \gamma(t,t+h)\sqrt{h}/2 \\ d_2 = d_1 - \gamma(t,t+h)\sqrt{h} \end{cases}$$

where $\gamma(t,t+h) > 0$ and :

$$\gamma^2(t,t+h) = \frac{1}{h}\int_t^{t+h} \sigma^2_{S\tau} d\tau. \tag{2.1.19}$$

This yields the so-called Hull and White option pricing formula :

$$C_t = S_t E_t\left[\phi(d_1) - e^{-x_t}\phi(d_2)\right], \tag{2.1.20}$$

where the expectation is taken with respect to the conditional probability distribution (for the risk neutral probability measure) of $\gamma(t,t+h)$ given σ_t.[9]

In the remainder of this section we will assume that observed option prices obey Hull and White's formula (2.1.20). Then option prices would yield two types of implied volatility concepts : (1) an instantaneous implied volatility and (2) an averaged implied volatility. To make this more precise, let us assume that the risk neutral probability distribution belongs to a parametric family, P_θ, $\theta \in \Theta$. Then, the Hull and White option pricing formula yields an expression for the option price as a function :

$$C_t = S_t F[\sigma_{St}, x_t, \theta_o] \tag{2.1.21}$$

where θ_o is the true unknown value of the parameters. Formula (2.1.21) reveals why it is often claimed that "option markets can be thought of as markets trading volatility" (see e.g. Stein (1989)). As a matter of fact, if for any given (x_t, θ), $F(\cdot, x_t, \theta)$ is one-to-one, then equation (2.1.21) can be inverted to yield an *implied instantaneous volatility* :[10]

$$\sigma^{imp}_\tau(\theta) = G[S_t, C_t, x_t, \theta] \tag{2.1.22}$$

Bajeux and Rochet (1992), by showing that this one-to-one relationship between option prices and instantaneous volatility holds, in fact formalize the use of option markets as an appropriate instrument to hedge volatility risk. Obviously implied instantaneous volatilities (2.1.22) could only be useful in practice for pricing or hedging derivative instruments when we know the true unknown value θ_o or , at least, are able to compute a sufficiently accurate estimate of it.

[9] The conditioning is with respect to σ_t since it summarizes the relevant information taken from I_t (the process σ is assumed to be Markovian and independent from W).

[10] The fact that $F(\cdot, x_t, \theta)$ is one-to-one is shown to be the case for any diffusion model on σ_t under certain regularity conditions, see Bajeux and Rochet (1992).

However, the difficulties involved in estimating SV models has for long prevented their wide spread use in empirical applications. This is the reason why practitioners often prefer another concept of implied volatility, namely the so-called *Black-Scholes implied volatility* introduced by Latane and Rendleman (1976). It is a process $\omega^{imp}(t, t+h)$ defined by :

$$\begin{cases} C_t = S_t\left[\phi(d_1) - e^{-x_t}\phi(d_2)\right] \\ d_1 = \left(x_t/\omega^{imp}(t,t+h)\sqrt{h}\right) + \omega^{imp}(t,t+h)\sqrt{h}/2 \\ d_2 = d_1 - \omega^{imp}(t,t+h)\sqrt{h} \end{cases} \tag{2.1.23}$$

where C_t is the observed option price.[11]

The Hull and White option pricing model can indeed be seen as a theoretical foundation for this practice; the comparison between (2.1.23) and (2.1.20) allows us to interpret the Black Scholes implied volatility $\omega^{imp}(t, t+h)$ as an implied averaged volatility since $\omega^{imp}(t, t+h)$ is something like a conditional expectation of $\gamma(t, t+h)$ (assuming observed option prices coincide with the Hull and White pricing formula). To be more precise, let us consider the simplest case of at the money options (the general case will be studied in section 4.2). Since $x_t = 0$ it follows that $d_2 = -d_1$ and therefore : $\phi(d_1) - e^{-x_t}\phi(d_2) = 2\phi(d_1) - 1$. Hence, $\omega_o^{imp}(t, t+h)$ (the index o is added to underline that we consider at the money options) is defined by :

$$\phi\left(\frac{\omega_o^{imp}(t,t+h)\sqrt{h}}{2}\right) = E_t\phi\left(\frac{\gamma(t,t+h)\sqrt{h}}{2}\right) \tag{2.1.24}$$

Since the cumulative standard normal distribution function is roughly linear in the neighborhood of zero, if follows that (for small maturities h) :

$$\omega_o^{imp}(t, t+h) \approx E_t\gamma(t, t+h)$$

This yields an interpretation of the Black-Scholes implied volatility $\omega_o^{imp}(t, t+h)$ as an implied average volatility :

$$\omega_o^{imp}(t, t+h) \approx E_t\left[\frac{1}{h}\int_t^{t+h}\sigma_\tau^2 d\tau\right]^{\frac{1}{2}} \tag{2.1.25}$$

2.2 Some Stylized Facts

The search for model specification and selection is always guided by empirical stylized facts. A model's ability to reproduce such stylized facts is a desirable feature and failure to do so is most often a criterion to dismiss a specification although one typically does not try to fit or explain all possible empirical regularities at once with a single model. Stylized facts about volatility have been well documented in the ARCH literature, see for instance Bollerslev, Engle and Nelson (1994). Empirical regularities regarding derivative securities and implied volatilities are also well covered for instance by Bates (1995a). In this section we will summarize empirical stylized facts, complementing and updating some of the material covered in the aforementioned references.

[11] We do not explicitly study here the dependence between $\omega^{imp}(t, t+h)$ and the various related processes : C_t, S_t, x_t. This is the reason why, for sake of simplicity, this dependence is not apparent in the notation $\omega^{imp}(t, t+h)$.

(a) *Thick tails*

Since the early sixties it was observed, notably by Mandelbrot (1963), Fama (1963, 1965), among others that asset returns have leptokurtic distributions. As a result, numerous papers have proposed to model asset returns as i.i.d. draws from fat-tailed distributions such as Paretian or Lévy.

(b) *Volatility clustering*

Any casual observations of financial time series reveals bunching of high and low volatility episodes. In fact, volatility clustering and thick tails of asset returns are intimately related. Indeed, the latter is a static explanation whereas a key insight provided by ARCH models is a formal link between dynamic (conditional) volatility behavior and (unconditional) heavy tails. ARCH models, introduced by Engle (1982) and the numerous extensions thereafter as well as SV models are essentially built to mimic volatility clustering. It is also widely documented that ARCH effects disappear with temporal aggregation, see e.g. Diebold (1988) and Drost and Nijman (1993).

(c) *Leverage effects*

A phenomenon coined by Black (1976) as the leverage effect suggests that stock price movements are negatively correlated with volatility. Because falling stock prices imply an increased leverage of firms it is believed that this entails more uncertainty and hence volatility. Empirical evidence reported by Black (1976), Christie (1982) and Schwert (1989) suggests, however, that leverage alone is too small to explain the empirical asymmetries one observes in stock prices. Others reporting empirical evidence regarding leverage effects include Nelson (1991), Gallant, Rossi and Tauchen (1992, 1993), Campbell and Kyle (1993) and Engle and Ng (1993).

(d) *Information arrivals*

Asset returns are typically measured and modeled with observations sampled at fixed frequencies such as daily, weekly or monthly observations. Several authors, including Mandelbrot and Taylor (1967) and Clark (1973) suggested to link asset returns explicitly to the flow of information arrivals. In fact it was already noted that Clark proposed one of the early examples of SV models. Information arrivals are non uniform through time and quite often not directly observable. Conceptually, one can think of asset price movements as the realization of a process $Y_t = Y^*_{Z_t}$ where Z_t is a so-called directing process. This positive nondecreasing stochastic process Z_t can be thought of as being related to the arrival of information. This idea of time deformation or subordinated stochastic processes was used by Mandelbrot and Taylor (1967) to explain fat tailed returns, by Clark (1973) to explain volatility and was recently refined and further explored by Ghysels, Gouriéroux and Jasiak (1995a). Moreover, Easley and O'Hara (1992) provide a microstructure model involving time deformation. In practice, it suggests a direct link between market volatility and (1) trading volume, (2) quote arrivals, (3) forecastable events such as dividend announcements or macroeconomic data releases, (4) market closures, among many other phenomena linked to information arrivals.

Regarding trading volume and volatility there are several papers documenting stylized facts notably linking high trading volume with market volatility, see for example Karpoff (1987) or Gallant, Rossi and Tauchen (1992).[12] The intraday patterns of volatility and market activity measured for instance by quote arrivals is also well-known and documented. Wood, McInish and Ord (1985) and Harris (1986) studied this phenomenon for securities markets and found a U-shaped pattern with volatility typically

[12] There are numerous models, theoretical and empirical, linking trading volume and asset returns which we cannot discuss in detail. A partial list includes Foster and Viswanathan (1993a,b), Ghysels and Jasiak (1994a,b), Hausman and Lo (1991), Huffman (1987), Lamoureux and Lastrapes (1990, 1993), Wang (1993).

high at the open and close of the market. The around the clock trading in foreign exchange markets also yields a distinct volatility pattern which is tied with the intensity of market activity and produces strong seasonal patterns. The intradaily patterns for FX markets are analyzed for instance by Müller et al. (1990), Baillie and Bollerslev (1991), Harvey and Huang (1991), Dacorogna et al. (1993), Andersen and Bollerslev (1995), Bollerslev and Ghysels (1994), Ghysels, Gouriéroux and Jasiak (1995b) among others. Another related empirical stylized fact is that of overnight and weekend market closures and their effect on volatility. Fama (1965) and French and Roll (1986) have found that information accumulates more slowly when the NYSE and AMEX are closed resulting in higher volatility on those markets after weekends and holidays. Similar evidence for FX markets has been reported by Baillie and Bollerslev (1989). Finally, numerous papers documented increased volatility of financial markets around dividend announcements (Cornell (1978), Patell and Wolfson (1979,1981)) and macroeconomic data releases (Harvey and Huang (1991, 1992), Ederington and Lee (1993)).

(e) *Long memory and persistence*

Generally speaking volatility is highly persistent. Particularly for high frequency data one finds evidence of near unit root behavior of the conditional variance process. In the ARCH literature numerous estimates of GARCH models for stock market, commodities, foreign exchange and other asset price series are consistent with an IGARCH specification. Likewise, estimation of stochastic volatility models show similar patterns of persistence (see for instance Jacquier, Polson and Rossi (1994)). These findings have led to a debate regarding modelling persistence in the conditional variance process either via a unit root or a long memory process. The latter approach has been suggested both for ARCH and SV models, see Baillie, Bollerslev and Mikkelsen (1993), Breidt et al. (1993), Harvey (1993) and Comte and Renault (1995). Ding, Granger and Engle (1993) studied the serial correlations of $|r(t,t+1)|^c$ for positive values of c where $r(t,t+1)$ is a one-period return on a speculative asset. They found $|r(t,t+1)|^c$ to have quite high autocorrelations for long lags while the strongest temporal dependence was for c close to one. This result initially found for daily S&P500 return series was also shown to hold for other stock market indices, commodity markets and foreign exchange series (see Granger and Ding (1994)).

(f) *Volatility comovements*

There is an extensive literature on international comovements of speculative markets. Concerns whether globalization of equity markets increase price volatility and correlations of stock returns has been the subject of many recent studies including, von Fustenberg and Jean (1989), Hamao, Masulis and Ng (1990), King, Sentana and Wadhwani (1994), Harvey, Ruiz and Sentana (1992), Lin, Engle and Ito (1994). Typically one uses factor models to model the commonality of international volatility, as in Diebold and Nerlove (1989), Harvey, Ruiz and Sentana (1992), Harvey, Ruiz and Shephard (1994) or explores so-called common features, see e.g. Engle and Kozicki (1993) and common trends as studied by Bollerslev and Engle (1993).

(g) *Implied volatility correlations*

Stylized facts are typically reported as model-free empirical observations.[13] Implied volatilities are obviously model-based as they are calculated from a pricing equation of a specific model, namely the Black and Scholes model as noted in section 2.1.3. Since they are computed on a daily basis there is obviously an internal inconsistency since the model presumes constant volatility. Yet, since many option prices are in fact quoted through their implied volatilities it is natural to study the time series behavior of the latter. Often one computes a composite measure since synchronous option prices with different strike prices and

[13]This is in some part fictitious even for macroeconomic data for instance when they are detrended or seasonal adjusted. Both detrending and seasonnaly adjustment are model-based. For the potentially severe impact of detrending on stylized facts see Canova (1992) and Harvey and Jaeger (1993) and for the effect of seasonal adjustment on empirical regularities see Ghysels et al. (1993).

maturities for the same underlying asset yield different implied volatilities. The composite measure is usually obtained from a weighting scheme putting more weight on the near-the-money options which are the most heavily traded on organized markets.[14]

The time series properties of implied volatilities obtained from stock, stock index and currency options are quite similar. They appear stationary and are well described by a first order autoregressive model (see Merville and Pieptea (1989) and Sheikh (1993) for stock options, Poterba and Summers (1986), Stein (1989), Harvey and Whaley (1992) and Diz and Finucane (1993) for the S&P 100 contract and Taylor and Xu (1994), Campa and Chang (1995) and Jorion (1995) for currency options). It was noted from equation (2.1.25) that implied (average) volatilities are expected to contain information regarding future volatility and therefore should predict the latter. One typically tests such hypotheses by regressing realized volatilities on past implied ones.

The empirical evidence regarding the predictable content of implied volatilities is mixed. The time series study of Lamoureux and Lastrapes (1993) considers options on non-dividend paying stocks and compared the forecasting performance of GARCH, implied volatility and historical volatility estimates and found that implied volatilities forecasts, though they are biased as one would expect from (2.1.25), outperform the others. In sharp contrast, Canina and Figlewski (1993) studied S&P 100 index call options for which there is an extremely active market. They found that implied volatilities were virtually useless in forecasting future realized volatilities of the S&P 100 index. In a different setting using weekly sampling intervals for S&P 100 option contracts and a different sample Day and Lewis (1992) not only found that implied volatilities had a predictive content but also were unbiased. Studies examining options on foreign currencies, such as Jorion (1995) also found that implied volatilities were predicting future realizations and GARCH as well as historical volatilities were not outperforming the implied measures of volatility.

(h) *The term structure of implied volatilities*

The Black-Scholes model predicts a flat term structure of volatilities. In reality, the term structure of at-the-money implied volatilities is typically upward sloping when short term volatilities are low and the reverse when they are high (see Stein(1989)). Taylor and Xu (1994) found that the term structure of implied volatilities from foreign currency options reverses slope every few months. Stein (1989) also found the actual sensitivity of medium to short term implied volatilities was greater than the estimated sensitivity from the forecast term structure and concluded that medium term implied volatilities overreacted to information. Diz and Finucane (1993) used different estimation techniques and rejected the overreaction hypothesis, even reported evidence suggesting underreaction.

(i) *Smiles*

If option prices in the market were conformable with the Black-Scholes formula, all the Black-Scholes implied volatilities corresponding to various options written on the same asset would coincide with the volatility parameter σ of the underlying asset. In reality this is not the case, and the Black-Scholes implied volatility $w^{imp}(t, t+h)$ defined by (2.1.23) heavily depends on the calendar time t, the time to maturity h and the moneyness $x_t = LogS_t/KB(t, t+h)$ of the option. This may produce various biases in option pricing or hedging when BS implied volatilities are used to evaluate new options with different strike prices K and maturities h . These price distortions, well-known to practitioners, are usually documented in the empirical literature under the terminology of the *smile effect,* where the so-called "smile" refers to the U-shaped pattern of implied volatilities across different strike prices. More precisely, the following stylized facts are extensively documented (see for instance Rubinstein (1985), Clewlow and Xu (1993), Taylor and

[14]Different weighting schemes have been suggested, see for instance Latane and Rendleman (1976), Chiras and Manaster (1978), Beckers (1981), Whaley (1982), Day and Lewis (1988), Engle and Mustafa (1992) and Bates (1995b).

Xu (1993)) :

- The U-shaped pattern of $w^{imp}(t, t+h)$ as a function of K (or $\log K$) has its minimum centered at near - the - money options (discounted K close to S_t, i.e. x_t close to zero).
- The volatility smile is often but not always symmetric as a function of $\log K$ (or of x_t).
- The amplitude of the smile increases quickly when time to maturity decreases. Indeed, for short maturities the smile effect is very pronounced (BS implied volatilities for synchronous option prices may vary between 15% and 25%) while it almost completely disappears for longer maturities.
- The smile can be asymmetric. This skewness effect can often be described as the addition of a monotonic curve to the standard symmetric smile: if a decreasing curve is added, implied volatilities tend to rise more for decreasing than for increasing strike prices and the implied volatility curve has its minimum out of the money. In the reverse case (addition of an increasing curve), implied volatilities tend to rise more with increasing strike prices and their minimum is in the money.

It is widely believed that volatility smiles have to be explained by a modelling of stochastic volatility. This is natural for several reasons: First, it is tempting to propose a model of stochastically time varying volatility to account for stochastically time varying BS implied volatilities. Moreover, the decreasing amplitude of the smile being a function of time to maturity is conformable with formula like (2.1.25). Indeed, it shows that, when time to maturity is increased, temporal aggregation of volatilities erases conditional heteroskedasticity, which decreases the smile phenomenon. Finally, the skewness itself may also be attributed to the stochastic feature of the volatility process and overall to the correlation of this process with the price process (the so-called leverage effect). Indeed, this effect, while sensible for stock prices data, is small for interest rate and exchange rate series which is why the skewness of the smile is more often observed for options written on stocks.

Nevertheless, it is important to be cautious about tempting associations: stochastic implied volatility and stochastic volatility; asymmetry in stocks and skewness in the smile. As will be discussed in section 4, such analogies are not always rigorously proven. Moreover, other arguments to explain the smile and its skewness (jumps, transaction costs, bid-ask spreads, non-synchronous trading, liquidity problems, ...) have also to be taken in account both for theoretical reasons and empirical ones. For instance, there exists empirical evidence suggesting that the most expensive options (the upper parts of the smile curve) are also the least liquid; skewness may therefore be attributed to specific configurations of liquidity in option markets.

2.3 Information sets

So far we left the specification of information sets vague. This was done on purpose to focus on one issue at the time. In this section we need to be more formal regarding the definition of information since it will allow us to clarify several missing links between the various SV models introduced in the literature and also between SV and ARCH models. We know that SV models emerged from research looking at a very diverse set of issues. In this section we will try to define a common thread and a general unifying framework. We will accomplish this through a careful analysis of information sets and associate notions of

non-causality in the Granger sense. These causality conditions will allows us to characterise in section 2.4 the distinct features of ARCH and SV models.[15]

2.3.1 State variables and information sets

The Hull and White (1987) model is a simple example of a derivative asset pricing model where the stock price dynamics are governed by some unobservable state variables, such as random volatility. More generally, it is convenient to assume that a multivariate diffusion process U_t summarizes the relevant state variables in the sense that:

$$\begin{cases} dS_t / S_t = \mu_t dt + \sigma_t dW_t \\ dU_t = \gamma_t dt + \delta_t dW_t^U \\ Cov\left(dW_t,\ dW_t^U\right) = \rho_t dt \end{cases} \tag{2.3.1}$$

where the stochastic processes $\mu_t, \sigma_t, \gamma_t, \delta_t$ and ρ_t are $I_t^U = [U_\tau, \tau \leq t]$ adapted (*Assumption 2.3.1*).

This means that the process U summarizes the whole dynamics of the stock price process S (which justifies the terminology "state" variable) since, for a given sample path $(U_\tau)_{0 \leq \tau \leq T}$ of state variables, consecutive returns $S_{t_{k+1}} / S_{t_k}, 0 \leq t_1 < t_2 < ... < t_k \leq T$ are stochastically independent and log-normal (as in the benchmark BS model).

The arguments of section 2.1.2 can be extended to the state variables framework (see Garcia and Renault (1995)) discussed here. Indeed, such an extension provides a theoretical justification for the common use of the Black and Scholes model as a standard method of quoting option prices via their implied volatilities.[16] In fact, it is a way of introducing neglected heterogeneity in the BS option pricing models (see Renault (1995) who draws attention to the similarities with introducing heterogeneity in microeconometric models of labor markets, etc.).

In continuous time models, available information at time t for traders (whose information determines option prices) is characterized by continuous time observations of both the state variable sample path and stock price process sample path; namely:

$$I_t = \sigma\left[U_\tau, S_\tau\ ;\ \tau \leq t\right] \tag{2.3.2}$$

2.3.2 Discrete sampling and Granger noncausality

In the next section we will treat explicitly discrete time models. It will necessitate formulating discrete time analogues of equation (2.3.1). The discrete sampling and Granger non causality conditions discussed here will bring us a step closer to building a formal framework for statistical modelling using discrete time data.

Clearly, a discrete time analogue of equation (2.3.1) is:

$$\log S_{t+1} / S_t = \mu\left(U_t\right) + \sigma\left(U_t\right) \varepsilon_{t+1} \tag{2.3.3}$$

[15] The analysis in this section has some features in common with Andersen (1992) regarding the use of information sets to clarify the difference between SV and ARCH type models.

[16] Garcia and Renault (1995) argued that Assumption 2.3.1 is essential to ensure the homogeneity of option prices with respect to the pair (stock price, strike price) which in turn ensures that BS implied volatilities do not depend on monetary units. This homogeneity property was first emphasized by Merton (1973).

provided we impose some restrictions on the process ε_t. The restrictions we want to impose must be flexible enough to accommodate phenomena such as leverage effects for instance. A setup that does this is the following :

Assumption 2.3.2.A : The process ε_t in (2.3.3) is i.i.d. and not Granger caused by the state variable process U_t.

Assumption 2.3.2.B : The process ε_t in (2.3.3) does not Granger cause U_t.

Assumption 2.3.2.B is useful for the practical use of BS implied volatilities as it is the discrete time analogue of Assumption 2.3.1 where it is stated that the coefficients of the process U are I_t^U adapted (for further details see Garcia and Renault (1995)). Assumption 2.3.2.A is important for the statistical interpretation of the functions $\mu(U_t)$ and $\sigma(U_t)$ respectively as trend and volatility coefficients. Namely,

$$\begin{aligned} &E\left[\log S_{t+1}/S_t \mid (S_\tau/S_{\tau-1}; \tau \leq t)\right] \\ &\quad = E\left[E\left[\log S_{t+1}/S_t \mid (U_\tau, \varepsilon_\tau; \tau \leq t)\right] \mid (S_\tau/S_{\tau-1}; \tau \leq t)\right] \\ &\quad = E\left[\mu(U_t) \mid (S_t/S_{t-1}; \tau \leq t)\right] \end{aligned} \tag{2.3.4}$$

since $E\left[\varepsilon_{t+1} \mid (U_\tau, \varepsilon_\tau; \tau \leq t)\right] = E\left[\varepsilon_{t+1} \mid \varepsilon_t; \tau \leq t\right] = 0$ due to the Granger noncausality from U_t to ε_t of Assumption 2.3.2.A. Likewise, one can easily show that

$$\begin{aligned} &Var\left[\log S_{t+1}/S_t - \mu(U_t) \mid (S_\tau/S_{\tau-1}; \tau \leq t)\right] \\ &\quad = E\left[\sigma^2(U_t) \mid (S_\tau/S_{\tau-1}; \tau \leq t)\right] \end{aligned} \tag{2.3.5}$$

Implicitly we have introduced a new information set in (2.3.4) and (2.3.5) which besides I_t defined in (2.3.2) will be useful as well for further analysis. Indeed, one often confines (statistical) analysis to information conveyed by a discrete time sampling of stock return series which will be denoted by the information set

$$I_t^R \equiv \sigma\left[S_\tau/S_{\tau-1} : \tau = 0, 1, ..., t-1, t\right] \tag{2.3.6}$$

where the superscript R stands for returns. By extending Andersen (1994), we shall adopt as the most general framework for univariate volatility modelling, the setup given by the Assumptions 2.3.2.A, 2.3.2.B and:
Assumption 2.3.2.C : $\mu(U_t)$ is I_t^R measurable.
Therefore in (2.3.4) and (2.3.5) we have essentially shown that :

$$E\left[\log S_{t+1}/S_t \mid I_t^R\right] = \mu(U_t) \tag{2.3.7}$$

$$Var\left[(\log S_{t+1}/S_t)\,\middle|\, I_t^R\right] = E\left[\sigma^2(U_t)\middle|\, I_t^R\right] \tag{2.3.8}$$

2.4 Statistical Modelling of Stochastic Volatility

Financial time series are observed at discrete time intervals while a majority of theoretical models are formulated in continuous time. Generally speaking there are two statistical methodologies to resolve this tension. Either one consider for the purpose of estimation statistical discrete time models of the continuous

time processes. Alternatively, the statistical model may be specified in continuous time and inference is done via a discrete time approximation. In this section we will discuss in detail the former approach while the latter will be introduced in section 4. The class of discrete time statistical models discussed here is general. In section 2.4.1 we introduce some notation and terminologies. The next section discuss the so-called stochastic autoregressive volatility model introduced by Andersen (1994) as a rather general and flexible semi-parametric framework to encompass various representations of stochastic volatility already available in the literature. Identification of parameters and requested additional restrictions are discussed in section 2.4.3

2.4.1 Notation and Terminology

In section 2.3, we left unspecified the functional forms which the trend $\mu(\cdot)$ and volatility $\sigma(\cdot)$ take. Indeed, in some sense we built a nonparametric framework recently proposed by Lezan, Renault and de Vitry (1995) which they introduced to discuss a notion of *stochastic volatility of unknown form*.[17] This nonparametric framework encompasses standard parametric models (see section 2.4.2 for more formal discussion). For the purpose of illustration let us consider two extreme cases, assuming for simplicity that $\mu(U_t) = 0$: (i) the discrete time analogue of the Hull and White model (2.1.16) is obtained when $\sigma(U_t) = \sigma_t$ is a stochastic process independent from the stock return standardized innovation process ε and (ii) σ_t may be a deterministic function $h(\varepsilon_t, \tau \leq t)$ of past innovations. The latter is the complete opposite of (i) and leads to a large variety of choices of parametrized functions for h yielding X-ARCH models (GARCH, EGARCH, QTARCH, Periodic GARCH, etc.).

Besides these two polar cases where Assumption 2.3.2.A is fulfilled in a trivial degenerate way, one can also accommodate leverage effects.[18] In particular the contemporaneous correlation structure between innovations in U and the return process can be nonzero, since the Granger non-causality assumptions deal with temporal causal links rather than contemporaneous ones. For instance, we may have $\sigma(U_t) = \sigma_t$ with:

$$\log S_{t+1}/S_t = \sigma_t \varepsilon_{t+1} \tag{2.4.1}$$

$$Cov\left(\sigma_{t+1}, \varepsilon_{t+1} | \mathrm{I}_t^R\right) \neq 0 \tag{2.4.2}$$

A negative covariance in (2.4.2) is a standard case of leverage effect, without violating the non-causality Assumptions 2.3.2.A and B.

A few concluding observations are worth making to deal with the burgeoning variety of terminologies in the literature. First, we have not considered the distinction due to Taylor (1994) between "lagged autoregressive random variance models" given by (2.4.1) and "contemporaneous autoregressive random variance models" defined by:

$$\log S_{t+1}/S_t = \sigma_{t+1} \varepsilon_{t+1} \tag{2.4.3}$$

[17] Lezan, Renault and de Vitry (1995) discuss in detail how to recover phenomena such as volatility clustering in this framework. As a nonparametric framework it also has certain advantages regarding (robust) estimation. They develop for instance methods that can be useful as a first estimation step for efficient algorithms assuming a specific parametric model (see Section 5).

[18] Assumption 2.3.2.B is fulfilled in the case (i) but may fail in the GARCH case (ii). When it fails to hold in the latter case it makes the GARCH framework not very well-suited for option pricing.

Indeed, since the volatility process σ_t is unobservable, the settings (2.4.1) and (2.4.3) are observationally equivalent as long as they are not completed by precise (non)-causality assumptions. For instance : (i) (2.4.1) and assumption 2.3.2.A together appear to be a correct and very general definition of a SV model possibly completed by Assumption 2.3.2.B for option pricing and (2.4.2) to introduction leverage effects, (ii) (2.4.3) associated with (2.4.2) would not be a correct definition of a SV model since in this case in general: $E\left[\log S_{t+1}/S_t \mid I_t^R\right] \neq 0$, and the model would introduce via the process σ a forecast which is related not only to volatility but also to the expected return.

For notational simplicity, the framework (2.4.3) will be used in section 3 with the leverage effect captured by $Cov(\sigma_{t+1}, \varepsilon_t) \neq 0$ instead of $Cov(\sigma_{t+1}, \varepsilon_{t+1}) \neq 0$. Another terminology was introduced by Amin and Ng (1993) for option pricing. Their distinction between "predictable" and "unpredictable" volatility is very close to the leverage effect concept and can also be analyzed through causality concepts as discussed in Garcia and Renault (1995). Finally, it will not be necessary to make a distinction between weak, semi-strong and strong definitions of SV models in analogy with their ARCH counterparts (see Drost and Nijman (1993)). Indeed, the class of SV models as defined here can accommodate parametrizations which are closed under temporal aggregation (see also section 4.1 on the subject of temporal aggregation).

2.4.2 Stochastic Autoregressive Volatility

For simplicity, let us consider the following univariate volatility process :

$$y_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1} \tag{2.4.4}$$

where μ_t is a measurable function of observables $y_t \in I_t^R$, $\tau \leq t$. While our discussion will revolve around (2.4.4), we will discuss several issues which are general and not confined to that specific model; extensions will be covered more explicitly in section 3.5. Following the result in (2.3.8) we know that :

$$Var\left[y_{t+1} \middle| I_t^R\right] = E\left[\sigma_t^2 \middle| I_t^R\right] \tag{2.4.5}$$

suggesting (1) that volatility clustering can be captured via autoregressive dynamics in the conditional expectation (2.4.5) and (2) that thick tails can be obtained in either one of three ways, namely (a) via heavy tails of the white noise ε_t distribution, (b) via the stochastic features of $E\left[\sigma_t^2 \middle| I_t^R\right]$ and (c) via specific randomness of the volatility process σ_t which makes it latent i.e. $\sigma_t \notin I_t^R$.[19] The volatility dynamics that follow from (1) and (2) are usually an AR(1) model for some nonlinear function of σ_t. Hence, the volatility process is assumed to be stationary and Markovian of order one but not necessarily linear AR(1) in σ_t itself. This is precisely what motivated Andersen (1994) to introduce the Stochastic Autoregressive Variance or SARV class of models where σ_t (or σ_t^2) is a polynomial function $g(K_t)$ of a Markov process K_t with the following dynamic specification :

$$K_t = w + \beta K_{t-1} + [\gamma + \alpha K_{t-1}]\, u_t \tag{2.4.6}$$

where $\tilde{u}_t = u_t - 1$ is zero-mean white noise with unit variance. Andersen (1994) discusses sufficient regularity conditions which ensure stationarity and ergodicity for K_t. Without entering into the details, let us note that the fundamental non-causality Assumption 2.3.2A implies that the u_t process in (2.4.6)

[19] Kim and Shephard (1994), using data on weekly returns on the S&P 500 Index , found that a t-GARCH model has an almost identical likelihood as the normal based SV model. This example shows that a specific randomness in σ_t may produce the same level of marginal kurtosis as a heavy tailed student distribution of the white noise ε.

does not Granger cause ε_t in (2.4.4). In fact, the non-causality condition suggests a slight modification of Andersen's (1994) definition. Namely, it suggests assuming ε_{t+1} independent of u_{t-j}, $j \geq 0$ for the conditional probability distribution, *given* ε_{t-j}, $j \geq 0$ rather than for the unconditional distribution. This modification does not invalidate Andersen's SARV class of models as the most general parametric statistical model studied so far in the volatility literature. The GARCH (1,1) model is straightforwardly obtained from (2.4.6) by letting $K_t = \sigma_t^2, \gamma = 0$ and $u_t = \varepsilon_t^2$. Note that the deterministic relationship $u_t = \varepsilon_t^2$ between the stochastic components of (2.4.4) and (2.4.6) emphasizes that, in GARCH models, there is no randomness specific to the volatility process. The Autoregressive Random Variance model popularized by Taylor (1986) also belongs to the SARV class. Here:

$$\log \sigma_{t+1} = \xi + \phi \log \sigma_t + \eta_{t+1} \tag{2.4.7}$$

where η_{t+1} is a white noise disturbance such that $Cov(\eta_{t+1}, \varepsilon_{t+1}) \neq 0$ to accommodate leverage effects. This is a SARV model with $K_t = \log \sigma_t$, $\alpha = 0$ and $\eta_{t+1} = \gamma u_{t+1}$.[20]

2.4.3 Identification of parameters

Introducing a general class of processes for volatility, like the SARV class discussed in the previous section prompts questions regarding identification. Suppose again that

$$\begin{array}{ll} y_{t+1} & = \sigma_t \varepsilon_{t+1} \\ \sigma_t^q & = g(K_t), \, q \in \{1, 2\} \\ K_t & = w + \beta K_{t-1} + [\gamma + \alpha K_{t-1}] u_t \, . \end{array} \tag{2.4.8}$$

Andersen (1994), noted the model is better interpreted by considering the zero-mean white noise process $\tilde{u}_t = u_t - 1$:

$$K_t = (w + \gamma) + (\alpha + \beta) K_{t-1} + (\gamma + \alpha K_{t-1}) \tilde{u}_t \, . \tag{2.4.9}$$

It is clear from the latter that it may be difficult to distinguish empirically the constant w from the "stochastic" constant γu_t. Similarly, the identification of the α and β parameters separately is also problematic as $(\alpha + \beta)$ governs the persistence of shocks to volatility. These identification problems are usually resolved by imposing (arbitrary) restrictions on the pairs of parameters (w, γ) and (α, β).

The GARCH(1,1) and Autoregressive Random Variance specifications assume that $\gamma = 0$ and $\alpha = 0$ respectively. Identification of all parameters without such restrictions generally requires additional constraints, for instance via some distributional assumptions on ε_{t+1} and u_t, which restrict the semi-parametric framework of (2.4.6) into a parametric statistical model.

To address more rigorously the issue of identification, it is useful to consider, according to Andersen (1994), the following reparametrisation (assuming for notational convenience that $\alpha \neq 0$) :

$$\left\{ \begin{array}{lcc} K & = & (w + \gamma) / (1 - \alpha - \beta) \\ \rho & = & \alpha + \beta \\ \delta & = & \gamma / \alpha \end{array} \right. \tag{2.4.10}$$

Hence equation (2.4.9) can be rewritten as :

[20] Andersen (1994) also shows that the SARV framework encompasses another type of random variance model that we have considered as ill-specified since it combines (2.4.2) and (2.4.3).

$$K_t = K + \rho\,(K_{t-1} - K) + (\delta + K_{t-1})\,\bar{U}_t$$

where $\bar{U}_t = \alpha\tilde{u}_t$.

It is clear from (2.4.10), that only three functions of the original parameters α, β, γ, w may be identified and that the three parameters K, ρ, δ are identified from the first three unconditional moments of the process K_t for instance.

To give to these identification results an empirical content, it is essential to know : (1) how to go from the moments of the observable process Y_t to the moments of the volatility process σ_t and (2) how to go from the moments of the volatility process σ_t to the moments of the latent process K_t. The first point is easily solved by specifying the corresponding moments of the standardised innovation process ε. If we assume for instance a Gaussian probability distribution, we obtain that :

$$\left\{\begin{array}{rcll} E\,|y_t| & = & \sqrt{2/\pi} & E\sigma_t \\ E\,|y_t|\,|y_{t-j}| & = & 2/\pi & E\,(\sigma_t\sigma_{t-j}) \\ E\,|y_t^2|\,|y_{t-j}| & = & \sqrt{2/\pi} & E\,(\sigma_t^2\sigma_{t-j}) \end{array}\right. \tag{2.4.11}$$

The solution of the second point requires in general the specification of the mapping g and of the probability distribution of u_t in (2.4.6). For the so-called Log-normal SARV model, it is assumed that $\alpha = 0$ and $K_t = \log\sigma_t$ (Taylor's autoregressive random variance model) and that u_t is normally distributed (Log-normality of the volatility process). In this case, it is easy to show that :

$$\left\{\begin{array}{rcl} E\sigma_t^n & = & \exp\left[n\;EK_t + n^2\;VarK_t/\,2\right] \\ E\left(\sigma_t^m\sigma_{t-j}^n\right) & = & E\sigma_t^m E\sigma_{t-j}^n \exp\left[mnCov\,(K_t, K_{t-j})\right] \\ Cov\,(K_t, K_{t-j}) & = & \beta^j VarK_t \end{array}\right. \tag{2.4.12}$$

Without the normality assumption (i.e. QML, mixture of normal, Student distribution ...) this model will be studied in much more detail in sections 3 and 5 from both probabilistic and statistical points of view. Moreover, this is a template for studying other specifications of the SARV class of models. In addition, various specifications will be considered in section 4 as proxies of continuous time models.

3 Discrete Time Models

The purpose of this section will be to discuss the statistical handling of discrete time SV models, using simple univariate cases. We start by defining the most basic SV model corresponding the autoregressive random variance model discussed earlier in (2.4.7). We study its statistical properties in section 3.2 and provide a comparison with ARCH models in section 3.3. Section 3.4 is devoted to filtering, prediction and smoothing. Various extensions, including multivariate models, are covered in the last section. Estimation of the parameters governing the volatility process is discussed later in section 5.

3.1 The Discrete Time SV Model

The discrete time SV model may be written as

$$y_t = \sigma_t\varepsilon_t, \qquad t = 1, ..., T, \tag{3.1.1}$$

where y_t denotes the demeaned return process $y_t = \log(S_t/S_{t-1}) - \mu$ and $\log \sigma_t^2$ follows an AR(1) process. It will be assumed that ε_t is a series of independent, identically distributed random disturbances. Usually ε_t is specified to have a standard distribution so its variance, σ_ε^2, is unknown. Thus for a normal distribution σ_ε^2 is unity while for a t-distribution with ν degrees of freedom it will be $\nu/(\nu-2)$. Following a convention often adopted in the literature we write:

$$y_t = \sigma \varepsilon_t e^{0.5 h_t} \tag{3.1.2}$$

where σ is a scale parameter, which removes the need for a constant term in the stationary first-order autoregressive process

$$h_{t+1} = \phi h_t + \eta_t, \eta_t \sim IID(0, \sigma_\eta^2), \quad |\phi| < 1. \tag{3.1.3}$$

It was noted before that if ε_t and η_t are allowed to be correlated with each other, the model can pick up the kind of asymmetric behavior which is often found in stock prices. Indeed a negative correlation between ε_t and η_t induces a leverage effect. As in section 2.4, the timing of the disturbance in (**??**) ensures that the observations are still a martingale difference, the equation being written in this way so as to tie in with the state space literature.

It should be stressed that the above model is only an approximation to the continuous time model of section 2 observed at discrete intervals. The accuracy of the approximation is examined in Dassios (1995) using Edgeworth expansions.

3.2 Statistical Properties

The following properties of the SV model hold even if ε_t and η_t are contemporaneously correlated. Firstly, as noted, y_t is a martingale difference. Secondly, stationarity of h_t implies stationarity of y_t. Thirdly, if η_t is normally distributed, it follows from the properties of the lognormal distribution that $E[\exp(ah_t)] = \exp(a^2\sigma_h^2/2)$, where a is a constant and σ_h^2 is the variance of h_t. Hence, if ε_t has a finite variance, the variance of y_t is given by

$$Var(y_t) = \sigma^2 \sigma_\varepsilon^2 \exp(\sigma_h^2/2) \tag{3.2.1}$$

where σ_h^2 is the variance of h_t. Similarly if the fourth moment of ε_t exists, the kurtosis of y_t is $\kappa \exp(\sigma_h^2)$, where κ is the kurtosis of ε_t, so y_t exhibit more kurtosis than ε_t. Finally all the odd moments are zero.

For many purposes we need to consider the moments of powers of absolute values. Again, η_t is assumed to be normally distributed. Then for ε_t having a standard normal distribution, the following expressions are derived in Harvey (1993):

$$E \mid y_t \mid^c = \sigma^c 2^{c/2} \frac{, \left(\frac{c}{2} + \frac{1}{2}\right)}{, \left(\frac{1}{2}\right)} \exp\left(\frac{c^2}{8}\sigma_h^2\right), c > -1, c \neq 0 \tag{3.2.2}$$

and

$$Var \mid y_t \mid^c = \sigma^{2c} 2^c \exp\left(\frac{c^2}{2}\sigma_h^2\right) \left\{ \frac{, \left(c+\frac{1}{2}\right)}{, \left(\frac{1}{2}\right)} - \left[\frac{, \left(\frac{c}{2} + \frac{1}{2}\right)}{, \left(\frac{1}{2}\right)} \right]^2 \right\}, c > -0.5, c \neq 0$$

Note that , $(1/2) = \sqrt{\pi}$ and , $(1) = 1$. Corresponding expressions may be computed for other distributions of ε_t including Student's t and the General Error Distribution.

Finally, the square of the coefficient of variation of σ_t^2 is often used as a measure of the relative strength of the SV process. This is $Var(\sigma_t^2)/[E(\sigma_t^2)]^2 = \exp(\sigma_h^2) - 1$. Jacquier, Polson and Rossi (1994) argue that this is more easily interpretable than σ_η^2. In the empirical studies they quote it is rarely less than 0.1 or greater than 2.

3.2.1 Autocorrelation Functions

If we assume that the disturbances ε_t and η_t are mutually independent, and η_t is normal, the ACF of the absolute values of the observations raised to the power c is given by

$$\rho_\tau^{(c)} = \frac{E(\mid y_t \mid^c \mid y_{t-\tau} \mid^c) - \{E(\mid y_t \mid^c)\}^2}{E(\mid y_t \mid^{2c}) - \{E(\mid y_t \mid^c)\}^2} = \frac{\exp(\frac{c^2}{4}\sigma_h^2\rho_{h,\tau}) - 1}{\kappa_c \exp(\frac{c^2}{4}\sigma_h^2) - 1}, \tau \geq 1, c > -0.5, c \neq 0 \tag{3.2.3}$$

where κ_c is

$$\kappa_c = E(\mid y_t \mid^{2c})/\{E(\mid y_t \mid^c)\}^2, \tag{3.2.4}$$

and $\rho_{h,\tau}, \tau = 0, 1, 2, ...$ denotes the ACF of h_t . Taylor (1986) gives this expression for c equal to one and two and ε_t normally distributed. When $c = 2$, κ_c is the kurtosis and this is three for a normal distribution. More generally,

$$\kappa_c = , (c + \frac{1}{2}), (\frac{1}{2})/\{, (\frac{c}{2} + \frac{1}{2})\}^2, c \neq 0$$

For Student's t-distribution with ν degrees of freedom :

$$\kappa_c = \frac{, (c + \frac{1}{2}), (-c + \frac{\nu}{2}), (\frac{1}{2}), (\frac{\nu}{2})}{\{, (\frac{c}{2} + \frac{1}{2}), (-\frac{c}{2} + \frac{\nu}{2})\}^2}, |c| < \nu/2, c \neq 0 \tag{3.2.5}$$

Note that ν must be at least five if c is two.

The ACF, $\rho_\tau^{(c)}$, has the following features. First, if σ_h^2 is small and/or $\rho_{h,\tau}$ is close to one,

$$\rho_\tau^{(c)} \simeq \rho_{h,\tau} \frac{\exp(\frac{c^2}{4}\sigma_h^2) - 1}{(\kappa_c \exp(\frac{c^2}{4}\sigma_h^2) - 1)}, \tau \geq 1; \tag{3.2.6}$$

compare Taylor (1986, p. 74-5). Thus the shape of the ACF of h_t is approximately carried over to $\rho_\tau^{(c)}$ except that it is multiplied by a factor of proportionality, which must be less than one for c positive as κ_c is greater than one. Secondly, for the t-distribution, κ_c declines as ν goes to infinity. Thus $\rho_\tau^{(c)}$ is a maximum for a normal distribution. On the other hand, a distribution with less kurtosis than the normal will give rise to higher values of $\rho_\tau^{(c)}$.

Although (**??**) gives an explicit relationship between $\rho_\tau^{(c)}$ and c, it does not appear possible to make any general statements regarding $\rho_\tau^{(c)}$ being maximized for certain values of c. Indeed different values of σ_h^2 lead to different values of c maximizing $\rho_\tau^{(c)}$. If σ_h^2 is chosen so as to give values of $\rho_\tau^{(c)}$ of a similar size to those reported in Ding, Granger and Engle (1993) then the maximum appears to be attained for c slightly less than one. The shape of the curve relating $\rho_\tau^{(c)}$ to c is similar to the empirical relationships reported in Ding, Granger and Engle, as noted by Harvey (1993).

3.2.2 Logarithmic Transformation

Squaring the observations in (??) and taking logarithms gives

$$\log y_t^2 = \log \sigma^2 + h_t + \log \varepsilon_t^2. \tag{3.2.7}$$

Alternatively

$$\log y_t^2 = \omega + h_t + \xi_t, \tag{3.2.8}$$

where $\omega = \log \sigma^2 + E \log \varepsilon_t^2$, so that the disturbance ξ_t has zero mean by construction.

The mean and variance of $\log \varepsilon_t^2$ are known to be -1.27 and $\pi^2/2$ =4.93 when ε_t has a standard normal distribution; see Abramovitz and Stegun (1970). However, the distribution of $\log \varepsilon_t^2$ is far from being normal, being heavily skewed with a long tail.

More generally, if ε_t has a t-distribution with ν degrees of freedom, it can be expressed as:

$$\varepsilon_t = \zeta_t \kappa_t^{-0.5},$$

where ζ_t is a standard normal variate and κ_t is independently distributed such that $\nu\kappa_t$ is chi-square with ν degrees of freedom. Thus

$$\log \varepsilon_t^2 = \log \zeta_t^2 - \log \kappa_t$$

and again using results in Abramovitz and Stegun (1970), it follows that the mean and variance of $\log \varepsilon_t^2$ are -1.27 -$\psi(\nu/2) - \log(\nu/2)$ and $4.93 + \psi'(\nu/2)$ respectively, where $\psi(.)$ is the digamma function. Note that the moments of ξ_t exist even if the model is formulated in such a way that the distribution of ε_t is Cauchy, that is $\nu = 1$. In fact in this case ξ_t is symmetric with excess kurtosis two, compared with excess kurtosis four when ε_t is Gaussian.

Since $\log \varepsilon_t^2$ is serially independent, it is straightforward to work out the ACF of $\log y_t^2$ for h_t following any stationary process:

$$\rho_\tau^{(0)} = \rho_{h,\tau} / \{1 + \sigma_\xi^2 / \sigma_h^2\}, \tau \geq 1 \tag{3.2.9}$$

The notation $\rho_\tau^{(0)}$ reflects the fact that the ACF of a power of an absolute value of the observation is the same as that of the Box-Cox transform, that is $\{|y_t|^c - 1\}/c$, and hence the logarithmic transform of an absolute value, raised to any (non-zero) power, corresponds to $c = 0$. (But note that one cannot simply set $c = 0$ in (??)).

Note that even if η_t and ε_t are not mutually independent, the η_t and ξ_t disturbances are uncorrelated if the joint distribution of ε_t and η_t is symmetric, that is $f(\varepsilon_t, \eta_t) = f(-\varepsilon_t, -\eta_t)$; see Harvey, Ruiz and Shephard (1994). Hence the expression for the ACF in (??) remains valid.

3.3 Comparison with ARCH models

The GARCH(1,1) model has been applied extensively to financial time series. The variance in (??) is assumed to depend on the variance and squared observation in the previous time period. Thus

$$\sigma_t^2 = \gamma + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \qquad t = 1, ..., T. \tag{3.3.1}$$

The GARCH model was proposed by Bollerslev (1986) and Taylor (1986), and is a generalization of the ARCH model formulated by Engle (1982). The ARCH(1) model is a special case of GARCH(1,1) with $\beta = 0$. The motivation comes from forecasting; in an AR(1) model with independent disturbances, the optimal prediction of the next observation is a fraction of the current observation, and in ARCH(1) it is a fraction of the current squared observation (plus a constant). The reason is that the optimal forecast is constructed conditional on the current information and in an ARCH model the variance in the next period is assumed to be known. This construction leads directly to a likelihood function for the model once a distribution is assumed for ε_t. Thus estimation of the parameters upon which σ_t^2 depends is straightforward in principle. The GARCH formulation introduces terms analogous to moving average terms in an ARMA model, thereby making forecasts a function of a distributed lag of past squared observations.

It is straightforward to show that y_t is a martingale difference with (unconditional) variance $\gamma/(1-\alpha-\beta)$. Thus $\alpha + \beta < 1$ is the condition for covariance stationarity. As shown in Bollerslev (1986), the condition under which the fourth moment exists in a Gaussian model is $2\alpha^2 + (\alpha+\beta)^2 < 1$. The model then exhibits excess kurtosis. However, the fourth moment condition may not always be satisfied in practice. Somewhat paradoxically, the conditions for strict stationarity are much weaker and, as shown by Nelson (1990), even include the case $\alpha + \beta = 1$.

The specification of GARCH(1,1) means that we can write

$$y_t^2 = \gamma + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2 + v_t = \gamma + (\alpha+\beta)y_{t-1}^2 + v_t - \beta v_{t-1}$$

where $v_t = y_t^2 - \sigma_t^2$ is a martingale difference. Thus y_t^2 has the form of an ARMA(1,1) process and so its ACF can be evaluated in the same way. The ACF of the corresponding ARMA model seems to be indicative of the type of patterns likely to be observed in practice in correlograms of y_t^2.

The GARCH model extends by adding more lags of σ_t^2 and y_t^2. However, GARCH(1,1) seems to be the most widely used. It displays similar properties to the SV model, particularly if ϕ is close to one. This should be clear from (**??**) which has the pattern of an ARMA(1,1) process. Clearly ϕ plays a role similar to that of $\alpha+\beta$. The main difference in the ACFs seems to show up most at lag one. Jacquier et al. (1994, p373) present a graph of the correlogram of the squared weekly returns of a portfolio on the New York Stock Exchange together with the ACFs implied by fitting SV and GARCH(1,1) models. In this case the ACF implied by the SV model is closer to the sample values.

The SV model displays excess kurtosis even if ϕ is zero since y_t is a mixture of distributions. The σ_η^2 parameter governs the degree of mixing independently of the degree of smoothness of the variance evolution. This is not the case with a GARCH model where the degree of kurtosis is tied to the roots of the variance equation, α and β in the case of GARCH(1,1). Hence, it is very often necessary to use a non-Gaussian GARCH model to capture the high kurtosis typically found in a financial time series.

The basic GARCH model does not allow for the kind of asymmetry captured by an SV model with contemporaneously correlated disturbances, though it can be modified as suggested in Engle and Ng (1993). The EGARCH model, proposed by Nelson (1991), handles asymmetry by taking $\log\sigma_t^2$ to be a function of past squares and absolute values of the observations.

3.4 Filtering, Smoothing and Prediction

For the purposes of pricing options, we need to be able to estimate and predict the variance, σ_t^2, which of course, is proportional to the exponent of h_t. An estimate based on all the observations up to, and possibly including, the one at time t is called a *filtered* estimate. On the other hand an estimate based on

all the observations in the sample, including those which came after time t is called a *smoothed* estimate. *Predictions* are estimates of future values. As a matter of historical interest we may wish to examine the evolution of the variance over time by looking at the smoothed estimates. These might be compared with the volatilities implied by the corresponding options prices as discussed in section 2.1.3. For pricing 'at the money' options we may be able to simply use the filtered estimate at the end of the sample and the predictions of future values of the variance, as in the method suggested for ARCH models by Noh, Engle and Kane (1994). More generally, it may be necessary to base prices on the full distribution of future values of the variance, perhaps obtained by simulation techniques; for further discussion see section 4.2.

One can think of constructing filtered and smoothed estimates in a very simple, but arbitrary way, by taking functions (involving estimated parameters) of moving averages of transformed observations. Thus :

$$\widehat{\sigma_t}^2 = g\{\sum_{j=t-1}^{r} w_{tj} f(y_{t-j})\}, t = 1, .., T, \tag{3.4.1}$$

where $r = 0$ or 1 for a filtered estimate and $r = t - T$ for a smoothed estimate.

Since we have formulated a stochastic volatility model, the natural course of action is to use this as the basis for filtering, smoothing and prediction. For a linear and Gaussian time series model, the state space form can be used as the basis for optimal filtering and smoothing algorithms. Unfortunately, the SV model is nonlinear. This leaves us with three possibilities:

1. a.compute inefficient estimates based on a linear state space model;

   b. use computer intensive techniques to estimate the optimal filter to a desired level of accuracy;

   c.use an (unspecified) ARCH model to approximate the optimal filter.

We now turn to examine each of these in some detail.

3.4.1 Linear State Space Form

The transformed observations, the $\log y_t^{2\prime}s$, can be used to construct a linear state space model as suggested by Nelson (1988) and Harvey, Ruiz and Shephard (1994). The measurement equation is (**??**) while (**??**) is the transition equation. The initial conditions for the state, h_t, are given by its unconditional mean and variance, that is zero and $\sigma_\eta^2/(1-\phi^2)$ respectively.

While it may be reasonable to assume that η_t is normal, ξ_t would only be normal if the absolute value of ε_t were lognormal. This is unlikely. Thus application of the Kalman filter and the associated smoothers yields estimators of the state, h_t, which are only optimal within the class of estimators based on linear combinations of the $\log y_t^{2\prime}s$. Furthermore, it is not the $h_t^\prime s$ which are required, but rather their exponents.

Suppose $h_{t|T}$ denotes the smoothed estimator obtained from the linear state space form. Then exp($h_{t|T}$) is of the form (**??**), multiplied by an estimate of the scaling constant, σ^2. It can be written as a weighted geometric mean. This makes the estimates vulnerable to very small observations and is an indication of the limitations of this approach.

Working with the logarithmic transformation raises an important practical issue, namely how to handle observations which are zero. This is a reflection of the point raised in the previous paragraph, since obviously any weighted geometric mean involving a zero observation will be zero. More generally we wish to avoid very small observations. One possible solution is to remove the sample mean. A somewhat more satisfactory alternative, suggested by Fuller, and studied by Breidt and Carriquiry (1995), is to make the following transformation based on a Taylor series expansion:

$$\log y_t^2 \cong \log(y_t^2 + cs_y^2) - cs_y^2/(y_t^2 + cs_y^2), \quad t = 1, \cdots, T, \tag{3.4.2}$$

where s_y^2 is the sample variance of the $y_t's$ and c is a small number, the suggested value being 0.02. The effect of this transformation is to reduce the kurtosis in the transformed observations by cutting down the long tail made up of the negative values obtained by taking the logarithms of the 'inliers'. In other words it is a form of trimming. It might be more satisfactory, to carry out this procedure after correcting the observations for heteroskedasticity by dividing by preliminary estimates, $\widehat{\sigma}_t^{2\prime}s$. The log $\widehat{\sigma}_t^{2\prime}s$ are then added to the transformed observations. The $\widehat{\sigma}_t^{2\prime}s$ could be constructed from a first round or by using a totally different procedure, perhaps a nonparametric one.

The linear state space form can be modified so as to deal with asymmetric models. It was noted earlier that even if η_t and ε_t are not mutually independent, the disturbances in the state space form are uncorrelated if the joint distribution of ε_t and η_t is symmetric. Thus the above filtering and smoothing operations are still valid, but there is a loss of information stemming from the squaring of the observations. Harvey and Shephard (1993) show that this information may be recovered by conditioning on the signs of the observations. These signs are, of course, the same as the signs of the ε_t's. Let $E_+(E_-)$ denote the expectation conditional on ε_t being positive (negative), and assign a similar interpretation to variance and covariance operators. The distribution of ξ_t is not affected by conditioning on the signs of the ε_t's, but, remembering that $E(\eta_t \mid \varepsilon_t)$ is an odd function of ε_t,

$$\mu^* = E_+(\eta_t) = E_+[E\eta_t|\varepsilon_t] = -E_-(\eta_t),$$

and

$$\gamma^* = Cov_+(\eta_t, \xi_t) = E_+(\eta_t\xi_t) - E_+(\eta_t)E(\xi_t) = E_+(\eta_t\xi_t) = -Cov_-(\eta_t, \xi_t),$$

because the expectation of ξ_t is zero and

$$E_+(\eta_t\xi_t) = E_+[E(\eta_t|\varepsilon_t)\log\varepsilon_t] - \mu^* E(\log\varepsilon_t) = -E_-(\eta_t\xi_t).$$

Finally

$$Var_+\eta_t = E_+(\eta_t^2) - [E_+(\eta_t)]^2 = \sigma_\eta^2 - \mu^{*2}.$$

The linear state space form is now

$$\begin{aligned} \log y_t^2 &= \omega + h_t + \xi_t \\ h_{t+1} &= \phi h_t + s_t\mu^* + \eta_t^* \quad , \\ \begin{pmatrix} \xi_t \\ \eta_t^* \end{pmatrix} \mid s_t &\sim ID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\xi^2 & \gamma^* s_t \\ \gamma^* s_t & \sigma_\eta^2 - \mu^{*2} \end{pmatrix}\right). \end{aligned} \tag{3.4.3}$$

The Kalman filter may still be initialized by taking h_0 to have mean zero and variance $\sigma_\eta^2/(1-\phi^2)$.

The parameterization in (**??**) does not directly involve a parameter representing the correlation between ε_t and η_t. The relationship between μ^* and γ^* and the original parameters in the model can only be obtained by making a distributional assumption about ε_t as well as η_t. When ε_t and η_t are bivariate normal with $Corr(\varepsilon_t, \eta_t) = \rho$, $E(\eta_t|\varepsilon_t) = \rho\sigma_\eta\varepsilon_t$, and so

$$\mu^* = E_+(\eta_t) = \rho\sigma_\eta E_+(\varepsilon_t) = \rho\sigma_\eta\sqrt{2/\pi} = 0.7979\rho\sigma_\eta. \tag{3.4.4}$$

Furthermore,

$$\gamma^* = \rho\sigma_\eta E(|\varepsilon_t|\log\varepsilon_t^2) - 0.7979\rho\sigma_\eta E(\log\varepsilon_t^2) = 1.1061\rho\sigma_\eta. \tag{3.4.5}$$

When ε_t has a t-distribution, it can be written as $\zeta_t \kappa_t^{-0.5}$, and ζ_t and η_t can be regarded as having a bivariate normal distribution with correlation ρ, while κ_t is independent of both. To evaluate μ^* and γ^* one proceeds as before, except that the initial conditioning is on ζ_t rather than on ε_t, and the required expressions are found to be exactly as in the Gaussian case.

The filtered estimate of the log volatility h_t, written as $h_{t+1|t}$, takes the form:

$$h_{t+1|t} = \phi h_{t|t-1} + \frac{\phi(p_{t|t-1} + \gamma^* s_t)}{p_{t|t-1} + 2\gamma^* s_t + \sigma_\xi^2}(\log y_t^2 - \omega - h_{t|t-1}) + s_t \mu^*,$$

where $p_{t|t-1}$ is the corresponding mean square error of the $h_{t|t-1}$. If $\rho < 0$, then $\gamma^* < 0$, and the filtered estimator will behave in a similar way to the EGARCH model estimated by Nelson (1991), with negative observations causing bigger increases in the estimated log volatility than corresponding positive values.

3.4.2 Nonlinear Filters

In principle, an exact filter may be written down for the original (**??**) and (**??**), with the former taken as the measurement equation. Evaluating such a filter requires approximating a series of integrals by numerical methods. Kitagawa (1987) has proposed a general method for implementing such a filter and Watanabe (1993) has applied it to the SV model. Unfortunately, it appears to be so time consuming as to render it impractical with current computer technology.

As part of their Bayesian treatment of the model as a whole, Jacquier, Polson and Rossi (1994) show how it is possible to obtain smoothed estimates of the volatilities by simulation. What is required is the mean vector of the joint distribution of the volatilities conditional on the observations. However, because simulating this joint distribution is not a practical proposition, they decompose it into a set of univariate distributions in which each volatility is conditional on all the others. These distributions may be denoted $p(\sigma_t | \sigma_{-t}, y)$, where σ_{-t} denotes all the volatilities apart from σ_t. What one would like to do is to sample from each of these distributions in turn, with the elements of σ_{-t} set equal to their latest estimates, and repeat several thousand times. As such this is a Gibbs sampler. Unfortunately, there are difficulties. The Markov structure of the SV model may be exploited to write

$$p(\sigma_t | \sigma_{-t}, y) = p(\sigma_t | \sigma_{t-1}, \sigma_{t+1}, y_t) \propto p(y_t | h_t)\, p(h_t | h_{t-1})\, p(h_{t+1} | h_t)$$

but although the right hand side of the above expression can be written down explicitly, the density is not of a standard form and there is no analytic expression for the normalising constant. The solution adopted by Jacquier, Polson and Rossi is to employ a series of Metropolis accept/reject independence chains.

Kim and Shephard (1994) argue that the single mover algorithm employed by Jacquier, Polson and Rossi will be slow if ϕ is close to one and/or σ_η^2 is small. This is because σ_t changes slowly; in fact when it is constant, the algorithm will not converge at all. Another approach based on the linear state space form, is to capture the non-normal disturbance term in the measurement equation, ξ_t, by a mixture of normals. Watanabe (1993) suggested an approximate method based on a mixture of two moments. Kim and Shephard (1994) propose a multimove sampler based on the linear state space form. Blocks of the $h_t's$ are sampled, rather than taking them one at a time. The technique they use is based on mixing an appropriate number of normal distributions to get the required level of accuracy in approximating the disturbance in (3.2.7). Mahieu and Schotman (1994a) extend this approach by introducing more degrees

of freedom in the mixture of normals where the parameters are estimated rather than fixed a priori. Note that the distribution of the $\sigma_t's$ can be obtained from the simulated distribution of the $h_t's$.

Jacquier, Polson and Rossi (1994, p.416) argue that no matter how many mixture components are used in the Kim and Shephard method, the tail behavior of $\log \varepsilon_t^2$ can never be satisfactorily approximated. Indeed, they note that given the discreteness of the Kim and Shephard state space, not all states can have been visited enough in the small number of draws mentioned, i.e. the socalled inlier problem (see also section 3.4.1 and Nelson (1994)) is still present.

As a final point it should be noted that when the hyperparameters are unknown, the simulated distribution of the state produced by the Bayesian approach allows for their sampling variability.

3.4.3 ARCH Models as Approximate Filters

The purpose here is to draw attention to a subject that will be discussed in greater detail in section 4.3. In an ARCH model the conditional variance is assumed to be an exact function of past observations. As pointed out by Nelson and Foster (1994, p.32) this assumption is *ad hoc* on both economic and statistical grounds. However, because ARCH models are relatively easy to estimate, Nelson (1992) and Nelson and Foster (1994) have argued that a useful strategy is to regard them as filters which produce *estimates* of the conditional variance. Thus even if we believe we have a continuous time or discrete time SV model, we may decide to estimate a GARCH(1,1) model and treat the $\sigma_t^{2\prime}s$ as an approximate filter, as in (**??**). Thus the estimate is a weighted average of past squared observations. It delivers an estimate of the mean of the distribution of σ_t^2, conditional on the observations at time t-1. As an alternative, the model suggested by Taylor (1986) and Schwert (1989), in which the conditional standard deviation is set up as a linear combination of the previous conditional standard deviation and the previous absolute value, could be used. This may be more robust to outliers as it is a linear combination of past absolute values.

Nelson and Foster derive an ARCH model which will give the closest approximation to the *continuous time* SV formulation (see section 4.3 for more details). This does not correspond to one of the standard models, though it is fairly close to EGARCH. For *discrete time* SV models the filtering theory is not as extensively developed. Indeed, Nelson and Foster point out that a change from stochastic different equations to *difference* equations makes a considerable difference in the limit theorems and optimality theory. They study the case of near diffusions as an example to illustrate these differences.

3.5 Extensions of the Model

3.5.1 Persistence and Seasonality

The simplest nonstationary SV model has h_t following a random walk. The dynamic properties of this model are easily obtained if we work in terms of the logarithmically transformed observations, $\log y_t^2$. All we have to do is first difference to give a stationary process. The untransformed observations are nonstationary but the dynamic structure of the model will appear in the ACF of $\mid y_t/y_{t-1} \mid^c$, provided that $c < 0.5$.

The model is an alternative to IGARCH, that is (**??**) with $\alpha + \beta = 1$. The IGARCH model is such that the squared observations have some of the features of an integrated ARMA process and it is said to exhibit persistence; see Bollerslev and Engle (1993). However, its properties are not straightforward. For example it must contain a constant, γ, otherwise, as Nelson (1990) has shown, σ_t^2 converges almost surely to zero and the model has the peculiar feature of being strictly stationary but not weakly stationary. The nonstationary SV model, on the other hand, can be analyzed on the basis that h_t is a standard integrated of order one process.

Filtering and smoothing can be carried out within the linear state space framework, since $\log y_t^2$ is just a random walk plus noise. The initial conditions are handled in the same way as is normally done with nonstationary structural time series models, with a proper prior for the state being effectively formed from the first observation; see Harvey (1989). The optimal filtered estimate of h_t within the class of estimates which are linear in past $\log y_t^2$'s, that is $h_{t|t-1}$, is a constant plus an equally weighted moving average (EWMA) of past $\log y_t^2$'s. In IGARCH σ_t^2 is given exactly by a constant plus an EWMA of past squared observations.

The random walk volatility can be replaced by other nonstationary specifications. One possibility is the doubly integrated random walk in which $\Delta^2 h_t$ is white noise. When formulated in continuous time, this model is equivalent to a cubic spline and is known to give a relatively smooth trend when applied in levels models. It is attractive in the SV context if the aim is to find a weighting function which fits a smoothly evolving variance. However, it may be less stable for prediction.

Other nonstationary components can easily be brought into h_t. For example, a seasonal or intra-daily component can be included; the specification is exactly as in the corresponding levels models discussed in Harvey (1989) and Harvey and Koopman (1993). Again the dynamic properties are given straightforwardly by the usual transformation applied to $\log y_t^2$, and it is not difficult to transform the absolute values suitably. Thus if the volatility consists of a random walk plus a slowly changing, nonstationary seasonal as in Harvey (1989, p. 40-3), the appropriate transformations are $\Delta_s \log y_t^2$ and $\mid y_t/y_{t-s} \mid^c$ where s is the number of seasons. The state space formulation follows along the lines of the corresponding structural time series models for levels. Handling such effects is not so easy within the GARCH framework.

Different approaches to seasonality can also be incorporated in SV models using ideas of time deformation as discussed in a later sub-section. Such approaches may be particularly relevant when dealing with the kind of abrupt changes in seasonality which seem to occur in high frequency, like five minute or tick-by-tick, foreign exchange data.

3.5.2 Interventions and other deterministic effects

Intervention variables are easily incorporated into SV models. For example, a sudden structural change in the volatility process can be captured by assuming that

$$\log \sigma_t^2 = \log \sigma^2 + h_t + \lambda w_t$$

where w_t is zero before the break and one after and λ is an unknown parameter. The logarithmic transformation gives (**??**) but with λw_t added to the right hand side. Care needs to be taken when incorporating such effects into ARCH models. For example, in the GARCH(1,1) a sudden break has to be modelled as

$$\sigma_t^2 = \gamma + \lambda w_t - (\alpha + \beta)\lambda w_{t-1} + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$$

with λ constrained so that σ_t^2 is always positive.

More generally observable explanatory variables, as opposed to intervention dummies, may enter into the model for the variance.

3.5.3 Multivariate Models

The multivariate model corresponding to (**??**) assumes that each series is generated by a model of the form

$$y_{it} = \sigma_i \varepsilon_{it} e^{0.5 h_{it}}, t = 1, ..., T, \tag{3.5.1}$$

with the covariance (correlation) matrix of the vector $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ being denoted by Σ_ε . The vector of volatilities, h_t, follows a VAR(1) process, that is

$$h_{t+1} = \Phi h_t + \eta_t,$$

where $\eta_t \sim IID(0, \Sigma_\eta)$. This specification allows the movements in volatility to be correlated across different series via Σ_η. Interactions can be picked up by the off-diagonal elements of Φ.

The logarithmic transformation of squared observations leads to a multivariate linear state space model from which estimates of the volatilities can be computed as in section 3.4.1.

A simple nonstationary model is obtained by assuming that the volatilities follow a multivariate random walk, that is $\Phi = I$. If Σ_η is singular, of rank $K < N$, there are only K components in volatility, that is each h_{it} in (**??**) is a linear combination of $K < N$ common trends, that is

$$h_t = \Theta h_t^\dagger + \overline{h} \tag{3.5.2}$$

where $h_t^\dagger$ is the $K \times 1$ vector of common random walk volatilities, $\overline{h}$ is a vector of constants and Θ is an $N \times K$ matrix of factor loadings. Certain restrictions are needed on Θ and $\overline{h}$ to ensure identifiability; see Harvey, Ruiz and Shephard (1994). The logarithms of the squared observations are 'co-integrated' in the sense of Engle and Granger (1987) since there are $N - K$ linear combinations of them which are white noise and hence stationary. This implies, for example, that if two series of returns exhibit stochastic volatility, but this volatility is the same with $\Theta' = (1, 1)$, then the ratio of the series will have no stochastic volatility. The application of the related concept of 'co-persistence' can be found in Bollerslev and Engle (1993). However, as in the univariate case there is some ambiguity about what actually constitutes persistence.

There is no reason why the idea of common components in volatility should not extend to stationary models. The formulation of (**??**) would apply, without the need for $\overline{h}$, and with $h_t^\dagger$ modelled, for example, by a VAR(1).

Bollerslev, Engle and Wooldridge (1988) show that a multivariate GARCH model can, in principle, be estimated by maximum likelihood, but because of the large number of parameters involved computational problems are often encountered unless restrictions are made. The multivariate SV model is much simpler than the general formulation of a multivariate GARCH. However, it is limited in that it does not model changing covariances. In this sense it is analogous to the restricted multivariate GARCH model of Bollerslev (1986) in which the conditional correlations are assumed to be constant.

Harvey, Ruiz and Shephard (1994) apply the nonstationary model to four exchange rates and find just two common factors driving volatility. Another application is in Mahieu and Schotman (1994b). A completely different way of modelling exchange rate volatility is to be found in the latent factor ARCH model of Diebold and Nerlove (1989).

3.5.4 Observation intervals, aggregation and time deformation

Suppose that an SV model is observed every δ time periods. In this case, h_τ, where τ denotes the new observation (sampling) interval, is still AR(1) but with parameter ϕ^δ. The variance of the disturbance, η_t, increases, but σ_h^2 remains the same. This property of the SV model makes it easy to make comparisons

across different sampling intervals; for example it makes it clear why if ϕ is around 0.98 for daily observations, a value of around 0.9 can be expected if an observation is made every week (assuming a week has 5 days).

If averages of observations are observed over the longer period, the comparison is more complicated, as h_τ will now follow an ARMA(1,1) process. However, the AR parameter is still ϕ^δ. Note that it is difficult to change the observation interval of ARCH processes unless the structure is weakened as in Drost and Nijman (1993); see also section 4.4.1.

Since, as noted in section 2.4, one typically uses a discrete time approximation to the continuous time model, it is quite straightforward to handle irregularly spaced observations by using the linear state space form as described, for example, in Harvey (1989). Indeed the approach originally proposed by Clark (1973) based on subordinated processes to describe asset prices and their volatility fits quite well into this framework. The techniques for handling irregularly spaced observations can be used as the basis for dealing with time deformed observations, as noted by Stock (1988). Ghysels and Jasiak (1994a, b) suggest a SV model in which the operational time for the continuous time volatility equation is determined by the flow of information. Such time deformed processes may be particularly suited to dealing with high frequency data. If $\tau = g(t)$ is the mapping between calendar time τ and operational time t, then

$$dS_t = \mu S_t dt + \sigma\left(g(t)\right) S_t dW_{1t}$$

and

$$d\log\sigma(\tau) = a\left(\left(b - \log\sigma(\tau)\right)d\tau + cdW_{2\tau}\right.$$

where W_{1t} and $W_{2\tau}$ are standard, independent Wiener processes. The discrete time approximation generalizing (**??**), but including a term which in (**??**) is incorporated in the constant scale factor σ, is then

$$h_{t+1} = [1 - e^{-a\Delta g(t)}]b + e^{-a\Delta g(t)}h_t + \eta_t$$

where $\Delta g(t)$ is the change in operational time between two consecutive calendar time observations and η_t is normally distributed with mean zero and variance $c^2(1 - e^{-2a\Delta g(t)})/2a$. Clearly if $\Delta g(t) = 1$, $\phi = e^{-a}$ in (**??**). Since the flow of information, and hence $\Delta g(t)$, is not directly observable, a mapping to calendar time must be specified to make the model operational. Ghysels and Jasiak (1994a) discuss several specifications revolving around a scaled exponential function relating $g(t)$ to observables such as past volume of trade and past price changes with asymmetric leverage effects. This approach was also used by Ghysels and Jasiak (1994b) to model return-volume co-movements and by Ghysels, Gouriéroux and Jasiak (1995b) for modeling intra-daily high frequency data which exhibit strong seasonal patterns (cfr. section 3.5.1).

3.5.5 Long Memory

Baillie, Bollerslev and Mikkelsen (1993) propose a way of extending the GARCH class to account for long memory. They call their models Fractionally Integrated GARCH (FIGARCH), and the key feature is the inclusion of the fractional difference operator, $(1 - L)^d$, where L is the lag operator, in the lag structure of past squared observations in the conditional variance equation. However, this model can only be stationary when $d = 0$ and it reduces to GARCH. In a later paper, Bollerslev and Mikkelsen (1995) consider a generalization of the EGARCH model of Nelson (1991) in which $\log\sigma_t^2$ is modelled as a distributed lag of past $\varepsilon_t s$ involving the fractional difference operator. This FIEGARCH model is stationary and invertible if $\mid d \mid < 0.5$.

Breidt, Crato and de Lima (1993) and Harvey (1993) propose an SV model with h_t generated by fractional noise

$$h_t = \eta_t/(1-L)^d, \eta_t \sim NID(0, \sigma_\eta^2), 0 \leq d \leq 1. \tag{3.5.1}$$

Like the AR(1) model in (??), this process reduces to white noise and a random walk at the boundary of the parameter space, that is $d = 0$ and 1 respectively. However, it is only stationary if d<0.5. Thus the transition from stationarity to nonstationarity proceeds in a different way to the AR(1) model. As in the AR(1) case it is reasonable to constrain the autocorrelations in (??) to be positive. However, a negative value of d is quite legitimate and indeed differencing h_t when it is nonstationary gives a stationary 'intermediate memory' process in which $-0.5 \leq d \leq 0$.

The properties of the long memory SV model can be obtained from the formulae in sub-section 3.2. A comparison of the ACF for h_t following a long memory process with $d = 0.45$ and $\sigma_h^2 = 2$ with the corresponding ACF when h_t is AR(1) with $\phi = 0.99$ can be found in Harvey (1993). Recall that a characteristic property of long memory is a hyperbolic rate of decay for the autocorrelations instead of an exponential rate, a feature observed in the data (see section 2.2(e)). The slower decline in the long memory model is very clear and, in fact, for $\tau = 1000$, the long memory autocorrelation is still 0.14, whereas in the AR case it is only 0.000013. The long memory shape closely matches that in Ding, Granger and Engle (1993, p86-8).

The model may be extended by letting η_t be an ARMA process and/or by adding more components to the volatility equation.

As regards smoothing and filtering, it has already been noted that the state space approach is approximate because of the truncation involved and is relatively cumbersome because of the length of the state vector. Exact smoothing and filtering, which is optimal within the class of estimators linear in the $\log y_t^{2\prime}s$, can be carried out by a direct approach if one is prepared to construct and invert the T$\times T$ covariance matrix of the $\log y_t^{2\prime}s$.

4 Continuous Time Models

At the end of section 2 we presented a framework for statistical modelling of SV in discrete time and devoted the entire section 3 to specific discrete time SV models. To motivate the continuous time models we study first of all the exact relationship (i.e. without approximation error) between differential equations and SV models in discrete time. We examine this relationship in section 4.1 via a class of statistical models which are closed under temporal aggregation and proceed (1) from high frequency discrete time to lower frequencies and (2) from continuous time to discrete time. Next, in section 4.2, we study option pricing and hedging with continuous time models and elaborate on features such as the smile effect. The practical implementation of option pricing formulae with SV often requires discrete time SV and/or ARCH models as filters and forecasters of the continuous time volatility processes. Such filters, covered in section 4.3, are in general discrete time approximations (and not exact discretizations as in section 4.1) of continuous time SV models. Section 4.4 concludes with extensions of the basic model.

4.1 From discrete to continuous time

The purpose of this section is to provide a rigorous discussion of the relationship between discrete and continuous time SV models. The presentation will proceed first with a discussion of temporal aggregation in the context of the SARV class of models and focus on specific cases including GARCH models. This

material is covered in section 4.1.1. Next we turn our attention to the aggregation of continuous time SV models to yield discrete time representations. This is the subject matter of section 4.1.2.

4.1.1 Temporal Aggregation of Discrete Time Models

Andersen's SARV class of models was presented in section 2.4 as a general discrete time parametric SV statistical model. Let us consider the zero-mean case, namely :

$$y_{t+1} = \sigma_t \varepsilon_{t+1} \tag{4.1.1}$$

and σ_t^q for $q = 1$ or 2 is a polynomial function $\mathrm{g}(K_t)$ of the Markov process K_t with stationary autoregressive representation :

$$K_t = \omega + \beta K_{t-1} + \upsilon_t \tag{4.1.2}$$

where $|\beta| < 1$ and

$$\begin{array}{l} \mathrm{E}\left[\varepsilon_{t+1} \middle| \varepsilon_\tau, \upsilon_\tau \ \tau \leq t\right] = 0 \\ \mathrm{E}\left[\varepsilon_{t+1}^2 \middle| \varepsilon_\tau, \upsilon_\tau \ \tau \leq t\right] = 1 \\ \mathrm{E}\left[\upsilon_{t+1} \middle| \varepsilon_\tau, \upsilon_\tau \ \tau \leq t\right] = 0 \end{array} \tag{4.1.3}$$

The restrictions (4.1.3a-c) imply that υ is a martingale difference sequence with respect to the filtration $\mathcal{F}_\sqcup = \sigma\left[\varepsilon_\tau, \upsilon_\tau, \tau \leq \sqcup\right]$.[21] Moreover, the conditional moment conditions in (4.3.1a-c) also imply that ε in (4.1.1) is a white noise process in a semi-strong sense, i.e. $\mathrm{E}[\varepsilon_{t+1} | \varepsilon_\tau, \tau \leq t] = 0$ and $\mathrm{E}\left[\varepsilon_{t+1}^2 \middle| \varepsilon_\tau, \tau \leq t\right] = 1$, and is not Granger caused by υ.[22] From the very beginning of section 2 we choose the continuously compounded rate of return over a particular time horizon as the starting point for continuous time processes. Therefore, let y_{t+1} in (4.1.1) be the continuously compounded rate of return for $[t,\ t+1]$ of the asset price process S_t, consequently :

$$y_{t+1} = \log S_{t+1} / S_t \tag{4.1.4}$$

Since the unit of time of the sampling interval is to a large extend arbitrary, we would surely want the SV model defined by equations (4.1.1) through (4.1.3), (for given q and function g) to be closed under temporal aggregation. As rates of return are flow variables, closedness under temporal aggregation means that for any integer m :

$$y_{tm}^{(m)} \equiv \log S_{tm} / S_{tm-m} = \sum_{k=0}^{m-1} y_{tm-k}$$

is again conformable to a model of the type (4.1.1) through (4.1.3) for the same choice of q and g involving suitably adapted parameter values. The analysis in this section follows Meddahi and Renault (1995) who study temporal aggregation of SV models in detail, particularly the cases (1) $\sigma_t^2 = K_t$, i.e. $q = 2$ and g is the identity function and (2) $\sigma_t^2 = \exp(K_t)$ which is the leading discrete time SV model. We will focus

[21] Note that we do not use here the decomposition appearing in (2.4.9) namely, $\upsilon_t = [\gamma + aK_{t-1}]\,\tilde{u}_t$.

[22] The Granger noncausality considered here for ε_t is weaker than Assumption 2.3.2.A as it applies only to the first two conditional moments.

here on the former as it is related to the socalled continuous time GARCH approach of Drost and Werker (1994). Hence, we have (4.1.1) with:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \upsilon_t \tag{4.1.5}$$

With conditional moment restrictions (4.1.3a-c) this model is closed under aggregation. For instance, for $m = 2$:

$$y_{t+1}^{(2)} = y_{t+1} + y_t = \sigma_{t-1}^{(2)}\varepsilon_{t+1}^{(2)}$$

with :

$$\left(\sigma_{t-1}^{(2)}\right)^2 = w^{(2)} + \beta^{(2)}\left(\sigma_{t-3}^{(2)}\right)^2 + \upsilon_{t-1}^{(2)}$$

where :

$$\begin{cases} w^{(2)} &= 2\omega\,(1+\beta) \\ \beta^{(2)} &= \beta^2 \\ \upsilon_{t-1}^{(2)} &= (\beta+1)\left[\beta\upsilon_{t-2} + \upsilon_{t-1}\right]. \end{cases}$$

Moreover, it also worth nothing that whenever a leverage effect is present at the aggregate level, i.e. :

$$Cov\left[\upsilon_{t-1}^{(2)}, \varepsilon_{t-1}^{(2)}\right] \neq 0$$

with $\varepsilon_{t-1}^{(2)} = (y_{t-1} + y_{t-2})\Big/\sigma_{t-3}^{(2)}$ and $\upsilon_{t-1}^{(2)} = (\beta+1)(\beta\upsilon_{t-1} + \upsilon_{t-1})$, it necessarily appears at the disaggregate level, i.e. $Cov(\upsilon_t, \varepsilon_t) \neq 0$.

For the general case Meddahi and Renault (1995) show that model (4.1.5a-b) together with conditional moment restrictions (4.1.3a-c) is a class of processes closed under aggregation. Given this result, it is of interest to draw a comparison with the work of Drost and Nijman (1993) on temporal aggregation of GARCH. While establishing this link between Meddahi and Renault (1995) and Drost and Nijman (1993) we will also uncover issues of leverage properties in GARCH models. Indeed, contrary to what is often believed, we will find leverage effect restrictions in socalled *weak* GARCH processes defined below. Moreover, we will also find from the results of Meddahi and Renault that the class of weak GARCH processes includes certain SV models.

To find a class of GARCH processes which is closed under aggregation Drost and Nijman (1993) weakened the definition of GARCH, namely for a positive stationary process h_t :

$$h_t = w + ay_t^2 + bh_{t-1} \tag{4.1.6}$$

where $a + b < 1$, they defined :

- strong GARCH if $y_{t+1}\Big/\sqrt{h_t}$ is i.i.d. with mean zero and variance 1
- semi-strong GARCH if $\mathrm{E}[y_{t+1}|\, y_\tau, \tau \leq t] = 0$ and $\mathrm{E}\left[y_{t+1}^2\,|y_\tau, \tau \leq t\right] = h_t$
- weak GARCH if $\mathrm{EL}[y_{t+1}|\, y_\tau, y_\tau^2, \tau \leq t] = 0$; $\mathrm{EL}\left[y_{t+1}^2\,|y_\tau, y_\tau^2, \tau \leq t\right] = h_t$.[23]

[23]For any Hilbert space H of L^2, $\mathrm{EL}[x_t\,|z, z \in H]$ is the best linear predictor of x_t in terms of 1 and $z \in H$. It should be noted that a strong GARCH process is a fortiori semi-strong which itself is also a weak GARCH process.

Drost and Nijman show that weak GARCH processes temporally aggregate and provide explicit formula for their coefficients. In section 2.4 it was noted that the framework of SARV includes GARCH processes whenever there is no randomness specific to the volatility process. This property will allow us to show that the class of weak GARCH processes -as defined above- in fact includes more general SV processes which are strictly speaking not GARCH. The arguments, following Meddahi and Renault (1995), require a classification of the models defined by (4.1.3) and (4.1.5) according to the value of the correlation between v_t and y_t^2, namely :

(a) *Models with perfect correlation* : This first class, henceforth denoted $\mathcal{C}_\infty$, is characterized by a linear correlation between v_t and y_t^2 conditional on $(\varepsilon_\tau, v_\tau, \tau < t)$ which is either 1 or -1 for the model in (4.1.5a-b).

(b) *Models without perfect correlation* : This second class, henceforth denoted $\mathcal{C}_\in$ has the above conditional correlation less than one in absolute value.

The class $\mathcal{C}_\infty$ contains all semi-strong GARCH processes, indeed whenever $Var\left[y_t^2 \middle| \varepsilon_t, v_\tau, \tau < t\right]$ is proportional to $Var\left[v_t \middle| \varepsilon_\tau, v_\tau, \tau < t\right]$ in $\mathcal{C}_\infty$ we have a semi-strong GARCH. Consequently, a semi-strong GARCH processes is a model (4.1.5a-b) with (1) restrictions (4.1.3), (2) a perfect conditional correlation as in $\mathcal{C}_\infty$, and (3) restrictions on the conditional kurtosis dynamics. [24]

Let us consider now the following assumption :

Assumption 4.1.1 : The following two conditional expectations are zero :

$$\begin{aligned} E\left[\varepsilon_t v_t \middle| \varepsilon_\tau, v_\tau, \tau < t\right] &= 0 \\ E\left[\varepsilon_t^3 \middle| \varepsilon_\tau, v_\tau, \tau < t\right] &= 0. \end{aligned} \tag{4.1.7}$$

This assumption amounts to an absence of leverage effects, where the latter is defined in a conditional covariance sense to capture the notion of instantaneous causality discussed in section 2.4.1 and applied here in the context of weak white noise.[25] It should also parenthetically be noted that (4.1.7a) and (4.1.7b) are in general not equivalent except for the processes of class $\mathcal{C}_\infty$.

The class $\mathcal{C}_\in$ allows for randomness proper to the volatility process due to the imperfect correlation. Yet, despite this volatility-specific randomness one can show that under *Assumption 4.1.1* processes of $\mathcal{C}_\in$ satisfy the weak GARCH definition. A fortiori, any SV model conformable to (4.1.3a-c), (4.1.5a-b) and *Assumption 4.1.1* is a weak GARCH process. It is indeed the symmetry assumption (4.1.5a-b), or restrictions on leverage in GARCH, that makes that $\mathrm{EL}\left[y_{t+1}^2 \middle| y_\tau, y_\tau^2, \tau \leq t\right] = \sigma_t^2$ (together with the conditional moment restrictions (4.1.3a-c)) and yields the internal consistency for temporal aggregation found by Drost and Nijman (1993, example 2, p.915) for the class of socalled symmetric weak GARCH(1,1). Hence, this class of weak GARCH(1,1) processes can be viewed as a subclass of processes satisfying (4.1.3) and (4.1.5).[26]

---

[24] In fact, Nelson and Foster (1994) observed that the most commonly used ARCH models effectively assume that the variance of the variance rises linearly in σ_t^4, which is the main drawback of ARCH models to approximate SV models in continuous time (see also section 4.3).

[25] The conditional expectation (4.1.7b) can be viewed as a conditional variance between ε_t and ε_t^2. It is this conditional covariance which, if nonzero, produces leverage effects in GARCH.

[26] As noted before, the class of processes satisfying (4.1.3) and (4.1.5) is closed under temporal aggregation, including processes with leverage effects not satisfying *Assumption 4.1.1*.

4.1.2 Temporal aggregation of continuous time models

To facilitate our discussion we will specialize the general continuous time model (2.3.1) to processes with zero drift, i.e. :

$$\begin{aligned} & d\log S_t = \sigma_t dW_t \\ & d\sigma_t = \gamma_t dt + \delta_t dW_t^{\sigma} \\ & Cov\left(dW_t, dW_t^{\sigma}\right) = \rho_t dt \end{aligned} \tag{4.1.8}$$

where the stochastic processes $\sigma_t, \gamma_t, \delta_t$ and ρ_t are $\mathrm{I}_t^{\sigma} = [\sigma_\tau; \tau \leq t]$ adapted. To ensure that σ_t is a nonnegative process one typically follows either one of two strategies : (1) considering a diffusion for $\log \sigma_t^2$ or (2) describing σ_t^2 as a CEV process (or Constant Elasticity of Variance process following Cox (1975) and Cox and Ross (1976)).[27] The former is frequently encountered in the option pricing literature (see e.g. Wiggins (1987)) and is also clearly related to Nelson (1991), who introduced EGARCH, and to the log-Normal SV model of Taylor (1986). The second class of CEV processes can be written as

$$d\sigma_t^2 = k\left(\theta - \sigma_t^2\right)dt + \gamma\left(\sigma_t^2\right)^{\delta} dW_t^{\sigma} \tag{4.1.9}$$

where $\delta \geq \frac{1}{2}$ ensures that σ_t^2 is a stationary process with nonnegative values. Equation (4.1.9) can be viewed as the continuous time analogue of the discrete time SARV class of models presented in section 2.4. This observation establishes links with the discussion of the previous section 4.1.1 and yields exact discretization results of continuous time SV models. Here, as in the previous section it will be tempting to draw comparisons with the GARCH class of models, in particular the diffusions proposed by Drost and Werker (1994) in line with the temporal aggregation of weak GARCH processes.

Firstly, one should note that the CEV process in (4.1.9) implies an autoregressive model in discrete time for σ_t^2 , namely :

$$\sigma_{t+\Delta t}^2 = \theta\left(1 - e^{-k\Delta t}\right) + e^{-k\Delta t}\sigma_t^2 + e^{-k\Delta t}\int_t^{t+\Delta t} e^{k(u-t)}\gamma\left(\sigma_u^2\right)^{\delta} dW_u^{\sigma} \tag{4.1.10}$$

Meddahi and Renault (1995) show that whenever (4.1.9) and its discretization (4.1.10) govern volatility then the discrete time process $\log S_{t+(k+1)\Delta t} \Big/ S_{t+k\Delta t}, k \in \mathbb{Z}$ is a SV process satisfying the model restrictions (4.1.3a-c) and (4.1.5a-b). Hence, from the diffusion (4.1.9) we obtain the class of discrete time SV models which is closed under temporal aggregation, as discussed in the previous section. To be more specific, consider for instance $\Delta t = 1$, then from (4.1.10) it follows that :

$$\begin{aligned} y_{t+1} &= \log S_{t+1} / S_t = \sigma_t \varepsilon_{t+1} \\ \sigma_t^2 &= w + \beta\sigma_{t-1}^2 + v_t \end{aligned} \tag{4.1.10}$$

where from (4.1.10) :

[27] Occasionally one encounters specifications which do not ensure nonnegativity of the σ_t process. For the sake of computational simplicity some authors for instance have considered Ornstein-Uhlenbeck processes for σ_t or σ_t^2 (see e.g. Stein and Stein (1991)).

$$\begin{cases} \beta & = & e^{-k}, w = \theta\left(1-e^{-k}\right), \\ \\ v_{t+1} & = & e^{-k}\int_{t}^{t+1} e^{k(u-t)}\gamma\left(\sigma_u^2\right)^{\delta} W_u^{\sigma}. \end{cases} \tag{4.1.11}$$

It is important to note from (4.1.11) that absence of leverage effect in continuous time, i.e. $\rho_t = 0$ in (4.1.8c), means no such effect at low frequencies and the two symmetry conditions of *Assumption 4.1.1* fulfilled. This line of reasoning also explains the temporal aggregation result of Drost and Werker (1994), but as noted in the previous section does not require absence of leverage. Indeed, following Meddahi and Renault (1995) one can interpret discrete time SV models with leverage effects as *exact* discretizations of continuous time SV models with CEV diffusions for volatility.

4.2 Option pricing and hedging

Section 4.2.1 is devoted to the basic option pricing model with SV or Hull and White model. This was introduced in section 2 but we are better equipped now to elaborate on its theoretical foundations. The practical implications appear in section 4.2.2 while 4.2.3 concludes with some extensions of the basic model.

4.2.1 The Basic Option Pricing Formula

Consider again formula (2.1.8) for a European option contract maturing at time $t+h=T$. As noted in section 2.1.2, we assume continuous and frictionless trading. Moreover no arbitrage profits can be made from trading in the underlying asset and riskless bonds ; interest rates are nonstochastic so that $B(t,T)$ defined by (2.1.12) denotes the time t price of a unit discount bond maturing at time T. Consider now the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, which is the fundamental space of the underlying asset price process S :

$$\begin{array}{lcl} dS_t/S_t & = & \mu(t,S_t,U_t)\,dt + \sigma_t dW_t^S \\ \sigma_t^2 & = & f(U_t) \\ dU_t & = & a(t,U_t)\,dt + b(t,U_t)\,dW_t^{\sigma} \end{array} \tag{4.2.1}$$

where $W_t = \left(W_t^S, W_t^U\right)$, is a standard two dimensional Brownian Motion (W_t^S and W_t^U are independent, zero-mean and unit variance) defined on $(\Omega, \mathcal{F}, \mathcal{P})$. The function f, called the volatility function, is assumed to be one-to-one. In this framework (under suitable regularity conditions) the *no free lunch* assumption is equivalent to the existence of a probability distribution Q on $(\Omega, \mathcal{F})$, equivalent to P, under which discounted price processes are martingales (see Harrison and Kreps (1979)). Such a probability is called an equivalent martingale measure and is unique if and only if the markets are complete (see Harrison and Pliska (1981)).[28] From the integral form of martingale representations (see Karatzas and Shreve (1988), problem 4.16, page 184), the (positive) density process of any probability measure Q equivalent to P can be written as :

$$M_t = \exp\left[-\int_0^t \lambda_u^S dW_u^S - \frac{1}{2}\int_0^t \left(\lambda_u^S\right)^2 du - \int_0^t \lambda_u^{\sigma} dW_u^{\sigma} - \frac{1}{2}\int_0^t \left(\lambda_u^{\sigma}\right)^2 du\right] \tag{4.2.2}$$

[28] Here, the market is seen as incomplete (before taking into account the market pricing of option) so that we have to characterize a set of equivalent martingale measures.

where the processes λ^S and λ^σ are adapted to the natural filtration $\sigma_t = \sigma[W_\tau, \tau \leq t], t \geq 0$, and satisfy the integrability conditions (almost surely) :

$$\int_0^t \left(\lambda_u^S\right)^2 du < +\infty \text{ and } \int_0^t \left(\lambda_u^\sigma\right)^2 du < +\infty$$

By Girsanov's theorem the process $\tilde{W} = \left(\tilde{W}^S, \tilde{W}^\sigma\right)'$ defined by :

$$\tilde{W}_t^S = \tilde{W}_t^S + \int_0^t \lambda_u^S du \text{ and } \tilde{W}_t^\sigma = W_t^\sigma + \int_0^t \lambda_u^\sigma du \tag{4.2.3}$$

is a two dimensional Brownian Motion under Q. The dynamic of the underlying asset price under Q is obtained directly from (4.2.1) and (4.2.3). Moreover, the discounted asset price process $S_t B(0,t), 0 \leq t \leq T$, is a Q-martingale if and only if for r_t defined in (2.1.11) :

$$\lambda_t^S = \frac{\mu(t, S_t, U_t) - r_t}{\sigma_t} \tag{4.2.4}$$

Since S is the only traded asset, the process λ^σ is not fixed. The process λ^S defined by (4.2.4) is called the asset risk premium. By analogy, any process λ^σ satisfying the required integrality condition can be viewed as a volatility risk premium and for any choice of λ^σ , the probability $Q(\lambda^\sigma)$ defined by the density process M in (4.2.2) is an equivalent martingale measure. Therefore, given the volatility risk premium process λ^σ :

$$C_t^{\lambda^\sigma} = B(t,T) E_t^{Q(\lambda^\sigma)} [Max[0, S_T - K]], 0 \leq t \leq T \tag{4.2.5}$$

is an admissible price process of the European call option.[29]

The Hull and White option pricing model relies on the following assumption, which restricts the set of equivalent martingale measures :

Assumption 4.2.1 : The volatility risk premium λ_t^σ only depends on the current value of the volatility process : $\lambda_t^\sigma = \lambda^\sigma(t, U_t), \forall t \in [0,T]$.

This assumption is consistent with an intertemporal equilibrium model where the agent preferences are described by time separable isoelastic utility functions (see He (1993) and Pham and Touzi (1993)). It ensures that $\tilde{W}^S$ and $\tilde{W}^\sigma$ are independent, so that the $Q(\lambda^\sigma)$ distribution of $\log S_T / S_t$, conditionally on $\mathcal{F}_\sqcup$ and the volatility path $(\sigma_t, 0 \leq t \leq T)$ is normal with mean $\int_t^T r_u du - \frac{1}{2}\gamma^2(t,T)$ and variance $\gamma^2(t,T) = \int_t^T \sigma_u^2 du$. Under *Assumption 4.2.1* one can compute the expectation in (4.2.5) conditionally on the volatility path, and we obtain finally:

$$C_t^{\lambda^\sigma} = S_t E_t^{Q(\lambda^\sigma)} \left[\phi(d_1) - e^{-x_t}\phi(d_2)\right] \tag{4.2.6}$$

with the same notation as in (2.1.20). To conclude it is worth noting that many option pricing formulae available in the literature have a feature common with (4.2.6) as they can be expressed as an expectation of the Black-Scholes price *over an heterogeneity distribution of the volatility parameter* (see Renault (1995) for an elaborate discussion on this subject).

[29] Here elsewhere $E_t^Q(\cdot) = E^Q(\cdot | \mathcal{F}_\sqcup)$ stands for the conditional expectation operator given $\mathcal{F}_\sqcup$ when the price dynamics are governed by Q.

4.2.2 Pricing and Hedging with the Hull and White model

The Markov feature of the process (S, σ) implies that the option price (4.2.6) only depends on the contemporaneous values of the underlying asset prices and its volatility. Moreover, under mild regularity conditions, this function is differentiable. Therefore, a natural way to solve the hedging problem in this stochastic volatility context is to hedge a given option of price C_t^1 by Δ_t^* units of the underlying asset and $\sum_t^*$ units of any other option of price C_t^2 where the hedging ratios solve :

$$\begin{cases} \partial C_t^1/\partial S_t - \Delta_t^* - \sum_t^* \partial C_t^2/\partial S_t = 0 \\ \partial C_t^1/\partial \sigma_t - \sum_t^* \partial C_t^2/\partial \sigma_t = 0 \end{cases} \tag{4.2.7}$$

Such a procedure, known as the delta-sigma hedging strategy, has been studied by Scott (1991). By showing that any European option completes the market, i.e. $\partial C_t^2/\partial \sigma_t \neq 0$, $0 \leq t \leq T$, Bajeux and Rochet justify the existence of an unique solution to the delta-sigma hedging problem (4.2.7) and the implicit assumption in the previous sections that the available information I_t contains the past values (S_t, σ_t), $\tau \leq t$.

Nevertheless, in practice, option traders focus on the risk due to the underlying asset price variations and consider the imperfect hedging strategy $\sum_t = 0$ and $\Delta_t = \partial C_t^1/\partial S_t$. Then, the Hull and White option pricing formula (4.2.6) provides directly the theoretical value of Δ_t :

$$\Delta_t = \partial C_t^{\lambda^\sigma}/\partial S_t = E_t^{Q(\lambda^\sigma)} \phi(d_1) \tag{4.2.8}$$

This theoretical value is hard to use in practice since : (1) even if we knew the $Q(\lambda^\sigma)$ conditional probability distribution of d_1 given I_t (summarised by σ_t), the derivation of the expectation (4.2.8) might be computationally intensive and (2) the conditional probability is directly related to the conditional probability distribution of $\gamma^2(t,T) = \int_t^T \sigma_u^2 du$ given σ_t, which in turn may involve nontrivially the parameters of the latent process σ_t. Moreover, these parameters are those of the conditional probability distribution of $\gamma^2(t,T)$ given σ_t under the risk-neutral probability $Q(\lambda^\sigma)$ which is generally different from the Data Generating Process P. The statistical inference issues are therefore quite complicated. We will argue in section 5 that only tools like simulation-based inference methods involving both asset *and* option prices data (via an option pricing model) may provide some satisfactory solutions.

Nevertheless, a practical way to avoid these complications is to use the Black-Scholes option pricing model, even though it is known to be misspecified. Indeed, option traders know that they cannot generally obtain sufficiently accurate option prices and hedge ratios by using the BS formula with historical estimates of the volatility parameters based on time series of the underlying asset price. However, the concept of Black-Scholes implied volatility (2.1.23) is known to improve the pricing and hedging properties of the BS model. This raises two issues : (1) what is the internal consistency of the simultaneous use of the BS model (which assumes constant volatility) and of BS implied volatility which is clearly time-varying and stochastic and (2) how to exploit the panel structure of option pricing errors ?[30]

Concerning the first issue, we noted in section 2 that the Hull and White option pricing model can indeed be seen as a theoretical foundation for this practice of pricing. Hedging issues and the panel structure of option *pricing errors* are studied in detail in Renault and Touzi (1992) and Renault (1995).

[30] The value of σ which equates the BS formula to the observed market price of the option heavily depends on the actual date t, the strike price K, the time to maturity $(T-t)$ and therefore creates a panel data structure.

4.2.3 Smile or Smurk?

As noted in section 2.2, the smile effect is now a well documented empirical stylized fact. Moreover the *smile* becomes sometimes a *smurk* since it appears more or less lopsided (the socalled skewness effect). We cautioned in section 2 that some explanations of the smile/smurk effect are often founded on tempting analogies rather than on rigorous proofs.

To the best of our knowledge, the state of the art is the following : (i) the first formal proof that a Hull and White option pricing formula implies a symmetric smile was provided by Renault and Touzi (1992), (ii) the first complete proof that the smile/smurk effects can alternatively be explained by liquidity problems (the upper parts of the smile curve, i.e. the most expensive options are the least liquid) was provided by Platten and Schweizer (1994) using a microstructure model, (iii) there is no formal proof that asymmetries of the probability distribution of the underlying asset price process (leverage effect, non-normality,...) are able to capture the observed skewness of the smile. A different attempt to explain the observed skewness is provided by Renault (1995). He showed that a slight discrepancy between the underlying asset price $\tilde{S}_t$ used to infer BS implied volatilities and the stock price S_t considered by option traders may generate an empirically plausible skewness in the smile. Such nonsynchronous $\tilde{S}_t$ and S_t may be related to various issues : bid-ask spreads, non-synchronous trading between the two markets, forecasting strategies based on the leverage effect, etc.

Finally, to conclude it is also worth noting that a new approach initiated by Gouriéroux, Monfort, Tenreiro (1994) and followed also by Ait-Sahalia, Bickel, Stoker (1994) is to explain the BS implied volatility using a nonparametric function of some observed state variables. Gouriéroux, Monfort, Tenreiro (1995) obtain for example a good nonparametric fit of the following form :

$$\sigma_t\left(S_t, K\right) = \mathrm{a}\left(K\right) + \mathrm{b}\left(K\right)\left(\log S_t / S_{t-1}\right)^2 .$$

A classical smile effect is directly observed on the intercept $\mathrm{a}(K)$ but an inverse smile effect appears for the path-dependent effect parameter $\mathrm{b}(K)$. For American options a different nonparametric approach is pursued by Broadie, Detemple, Ghysels and Torrès (1995) where besides volatility also exercise boundaries for the option contracts are nonparametrically obtained.[31]

4.3 Filtering and Discrete Time Approximations

In section 3.4.3 it was noted that the ARCH class of models could be viewed as filters to extract the (continuous time) conditional variance process from discrete time data. Several papers were devoted to the subject, namely Nelson (1990, 1992, 1995a,b) and Nelson and Foster (1994, 1995). It was one of Nelson's seminal contributions to bring together ARCH and continuous time SV. Nelson's first contribution in his 1990 paper was to show that ARCH models, which model volatility as functions of past (squared) returns, converge weakly to a diffusion process, either a diffusion for $\log \sigma_t^2$ or a CEV process as described in section 4.1.2. In particular, it was shown that a GARCH(1,1) model observed at finer and finer time intervals $\Delta t = h$ with conditional variance parameters $\omega_h = h\omega, \alpha_h = \alpha\left(h/2\right)^{\frac{1}{2}}$ and $\beta_h = 1 - \alpha\left(h/2\right)^{\frac{1}{2}} - \theta h$ and conditional mean $\mu_h = hc\sigma_t^2$ converges to a diffusion limit quite similar to equations (4.1.8a) combined with (4.1.9) with $\delta = 1$, namely

$$\begin{aligned} d\log S_t &= c\sigma_t^2 dt + \sigma_t dW_t \\ d\sigma_t^2 &= \left(\omega - \theta\sigma_t^2\right) dt + \sigma_t^2 dW_t^{\sigma} . \end{aligned}$$

[31] See also Bossaerts and Hillion (1995) for the use of a nonparametric hedging procedure and the smile effect.

Similarly, it was also shown that a sequence of AR(1)-EGARCH(1,1) models converges weakly to an Ornstein-Uhlenbeck diffusion for $\ln \sigma_t^2$:

$$d \ln \sigma_t^2 = \alpha \left(\beta - \ln \sigma_t^2 \right) dt + dW_t^{\sigma}.$$

Hence, these basic insights showed that the continuous time stochastic difference equations emerging as diffusion limits of ARCH models were no longer ARCH but instead SV models. Moreover, following Nelson (1992) even when misspecified ARCH models still kept desirable properties regarding extracting the continuous time volatility. The argument was that for a wide variety of misspecified ARCH models the difference between the ARCH filter volatility estimates and the true underlying diffusion volatilities converges to zero in probability as the length of the sampling time interval goes to zero at an appropriate rate. For instance the GARCH(1,1) model with ω_h, α_h and β_h described before estimates $\hat{\sigma}_t^2$ as follows :

$$\hat{\sigma}_t^2 = \omega_h \left(1 - \beta_h\right)^{-1} + \sum_{i=o}^{\infty} \alpha_h \beta_h^i y_{t-h(i+1)}^2$$

where $y_t = \log S_t / S_{t-h}$. This filter can be viewed as a particular case of equation (3.4.1). The GARCH(1,1) and many other models, effectively achieve consistent estimation of σ_t via a lag polynomial function of past squared returns close to time t.

The fact that a wide variety of misspecified ARCH models consistently extract σ_t from high frequency data raises questions regarding efficiency of filters. The answers to such questions are provided in Nelson (1995a,b) and Nelson and Foster (1994, 1995). In section 3.4 it was noted that the linear state space Kalman filter can also be viewed as a (suboptimal) extraction filter for σ_t. Nelson and Foster (1994) show that the asymptotically optimal linear Kalman filter has asymptotic variance for the normalized estimation error $h^{-\frac{1}{4}} \left[\ln \left(\hat{\sigma}_t^2 \right) - \ln \sigma_t^2 \right]$ equal to $\lambda \mathrm{Y}(1/2)^{\frac{1}{2}}$ where $\mathrm{Y}(x) = d \left[\ln , \ (x) \right] / dx$ and λ is scaling factor. A model, closely related to EGARCH of the following form :

$$\ln \left(\hat{\sigma}_{t+h}^2 \right) = \ln \left(\hat{\sigma}_t^2 \right) + \rho \lambda \left(S_{t+h} - S_t \right) \hat{\sigma}_t^{-1}$$
$$+ \lambda \left(1 - \rho^2\right)^{\frac{1}{2}} \left[, \ (1/2)^{\frac{1}{2}} , \ (3/2)^{\frac{1}{2}} \left| S_{t+2} - S_t \right| \hat{\sigma}_t^{-1} - 2^{-\frac{1}{2}} \right]$$

yields the asymptotically optimal ARCH filter with asymptotic variance for the normalized estimation error equal to $\lambda \left[2 \left(1 - \rho^2\right) \right]^{\frac{1}{2}}$ where the parameter ρ measure the leverage effect. These results also show that the differences between the most efficient suboptimal Kalman filter and the optimal ARCH filter can be quite substantial. Besides filtering one must also deal with smoothing an forecasting. Both of these issues were discussed in section 3.4 in discrete time SV models. The prediction properties of (misspecified) ARCH models were studied extensively by Nelson and Foster (1995). Nelson (1995) takes ARCH models a step further by studying smoothing filters, i.e. ARCH models involving not only lagged squared returns but also future realizations, i.e. $r = t - T$ in equation (3.4.1).

4.4 Long Memory

We conclude this section with a brief discussion of long memory in continuous time SV models. The purpose is to build continuous time long memory stochastic volatility models which are relevant for high frequency financial data and for (long term) option pricing. The reasons motivating the use of long memory models were discussed in sections 2.2 and 3.5.5. The advantage of considering *continuous time* long memory is their relative ability to provide a more structural interpretation of the parameters governing short term and long term dynamics. The first subsection defines fractional Brownian Motion. Next we will turn our attention to the fractional SV model followed by a section on filtering and discrete time approximations.

4.4.1 Stochastic integration with respect to fractional Brownian Motion

We recall in this subsection a few definitions and properties of fractional and long memory processes in continuous time, extensively studied for instance in Comte and Renault (1993). Consider the scalar process:

$$x_t = \int_0^t \mathrm{a}\,(t-s)\,dW_s \tag{4.4.1}$$

Such a process is asymptotically equivalent in quadratic mean to the stationary process :

$$y_t = \int_{-\infty}^t \mathrm{a}\,(t-s)\,dW_s \tag{4.4.2}$$

whenever $\int_0^{+\infty} a^2\,(x)\,dx < +\infty$. Such processes are called fractional processes if $a\,(x) = x^{\alpha}\tilde{a}\,(x)\,/\,,\ (1+\alpha)$for $|\alpha| < 1/2$, $'\tilde{a}$ where , $(1+\alpha)$ is a scaling factor useful for normalizing fractional derivation operators on $[0,T]$. Such processes admit several representations, and in particular that they can also be written :

$$x_t = \int_0^t c\,(t-s)\,dW_{\alpha s}, W_{\alpha t} = \int_0^t \frac{(t-s)^{\alpha}}{,\ (1+\alpha)} dW_s \tag{4.4.3}$$

where W_α is the so-called fractional Brownian Motion of order α (see Mandelbrot and Van Ness (1968)).

The relation between the functions a and c is one-to-one. One can show that W_α is *not* a semi-martingale (see e.g. Rogers (1995)) but stochastic integration with respect to W_α can be defined properly. The processes x_t are long memory if:

$$\begin{array}{l} \lim x\tilde{a}\,(x) = a_\infty\ ,\ 0 < \alpha < 1/2 \text{ and } 0 < a_\infty < +\infty. \\ x \to +\infty \end{array} \tag{4.4.4}$$

$$dx_t = -kx_t dt + \sigma dW_{\alpha t} \qquad x_t = 0, k > 0, 0 < \alpha < 1/2 \tag{4.4.5}$$

with its solution given by :

$$\begin{array}{l} x_t = \int_0^t (t-s)^{\alpha}\,(,\ (1+\alpha))^{-1}\,dx_t^{(\alpha)} \\ x_t^{(\alpha)} = \int_0^t e^{-k(t-s)}\sigma dW_s \end{array} \tag{4.4.6}$$

Note that, $x_t^{(\alpha)}$ the derivative of order α of x_t, is a solution of the usual SDE: $dz_t = -kz_t dt + \sigma dW_t$.

4.4.2 The fractional SV model

To facilitate comparison with both the FIEGARCH model and the fractional extensions of the log-Normal SV model discussed in section 3.5.5 let us consider the following fractional SV model (henceforth FSV) :

$$\begin{array}{l} dS_t\,/\,S_t = \sigma_t dW_t \\ d\log\sigma_t = -k\log\sigma_t dt + \gamma dW_{\alpha t} \end{array} \tag{4.4.7}$$

where $k > 0$ and $0 \le \alpha < 1/2$. If nonzero, the fractional exponent α will provide some degree of freedom in the order of regularity of the volatility process, namely the greater α the smoother the path of the volatility process. If we denote the autocovariance function of σ by $r_\sigma\,(\cdot)$ then:

$$\alpha > 0 \Rightarrow (r_\sigma\,(h) - r_\sigma\,(0))\,/h \to 0 \text{ as } h \to 0\ .$$

This would be incorrectly interpreted as near-integrated behavior, widely found in high frequency data for instance, when:

$$r_\sigma(h) - r_\sigma(0)/h = \left(\rho^h - 1\right)\Big/ h \to \log\rho \text{ as } h \to 0.$$

and σ_t is a continuous time AR(1) with correlation ρ near 1.

The long memory continuous time approach allows us to model persistence with the following features (1) the volatility process itself (and not only its logarithm) has hyperbolic decay of the correlogram ; (2) the persistence of volatility shocks yields leptokurtic features for return which vanishes with temporal aggregation at a slow hyperbolic rate of decay.[32] Indeed for rate of return on $[0,h]$:

$$\mathrm{E}\left[\log S_{t+h}/S_t - \mathrm{E}\left(\log S_{t+h}/S_t\right)\right]^4 \Big/ \left(\mathrm{E}\left[\log S_{t+h}/S_t - \mathrm{E}\left(\log S_{t+h}/S_t\right)\right]^2\right)^2 \to 3$$

as $h \to \infty$ at a rate $h^{2\alpha-1}$ if $\alpha \in]0,1/2[$ and a rate $\exp(-kh/2)$ if $\alpha = 0$.

4.4.3 Filtering and Discrete Time Approximations

The volatility process dynamics are described by the solution to the SDE (4.4.5), namely :

$$\log\sigma_t = \int_0^t (t-s)^\alpha /,\ (1+\alpha)\, d\log\sigma_s^{(\alpha)} \tag{4.4.6}$$

where $\log\sigma^{(\alpha)}$ follows the O-U process :

$$d\log\sigma_t^{(\alpha)} = -k\log\sigma_t^{(\alpha)}dt + \gamma dW_t \tag{4.4.7}$$

To compute a discrete time approximation one must evaluate numerically the integral (4.4.6) using only values of the process $\log\sigma^{(\alpha)}$ on a discrete partition of $[o,t]$ at point $j/n, j = 0,1\ldots,[nt]$.[33] A natural way to proceed is to use step functions, generating the following proxy process:

$$\log\hat{\sigma}_t^n = \sum_{j=1}^{[nt]} \left(t - (j-1)/n\right)^\alpha \Big/,\ (1+\alpha)\,\Delta\log\sigma_{j/n}^{(\alpha)} \tag{4.4.8}$$

where $\Delta\log\sigma_{j/n}^{(\alpha)} = \log\sigma_{j/n}^{(\alpha)} - \log\sigma_{(j-1)/n}^{(\alpha)}$. Comte and Renault (1995) show that $\log\hat{\sigma}_{nt}$ converges to the $\log\sigma_t$ process for $n \to \infty$ uniformly on compact sets. Moreover, by rearranging (4.4.8) one obtains:

$$\log\hat{\sigma}_{j/n}^n = \left[\sum_{i=0}^{j-1} \left(\left[(i+1)^\alpha - i^\alpha\right]/n^\alpha,\ (1+\alpha)\right) L_n^i\right] \log\sigma_{j/n}^{(\alpha)} \tag{4.4.9}$$

where L_n is the lag operator corresponding to the sampling scheme j/n, i.e. $L_n Z_{j/n} = Z_{(j-1)/n}$. With this sampling scheme $\log\sigma^{(\alpha)}$ is a discrete time AR(1) deduced from the continuous time process with the following representation :

$$(1 - \rho_n L_n)\log\sigma_{j/n}^{(\alpha)} = u_{j/n} \tag{4.4.10}$$

[32] With usual GARCH or SV models, it vanishes at an exponential rate (see Drost and Nijman (1993) and Drost and Werker (1994) for these issues in the short memory case).

[33] $[z]$ is the integer k such that $k \le z < k+1$.

where $\rho_n = \exp(-k/n)$ and $u_{j/n}$ is the associated innovations process. Since the process is stationary we are allowed to write (assuming $\log \sigma^{(\alpha)}_{j/n} = u_{j/n} = 0$ for $j \leq 0$) :

$$\log \hat{\sigma}^{(n)}_{j/n} = \left[\sum_{i=0}^{j-1} \frac{(i+1)^\alpha - i^\alpha}{n^\alpha,\ (1+\alpha)} L_n^i \right] (1 - \rho_n L_n)^{-1} u_{j/n} \tag{4.4.11}$$

which gives a parametrization of the volatility dynamics in two parts : (1) a long memory part which corresponds to the filter $\sum_{i=o}^{+\infty} \mathrm{a}_i\ L_n^i / n^\alpha$ with $\mathrm{a}_i = [(i+1)^\alpha - i^\alpha] /,\ (1+\alpha)$ and (2) a short memory part which is characterized by the AR(1) process : $(1 - \rho_n L_n)^{-1} u_{j/n}$. Indeed, one can show that the long memory filter is "long-term equivalent" to the usual discrete time long memory filters $(1-L)^{-\alpha}$ in the sense that there is a long term relationship (a cointegration relation) between the two types of processes. However, this long-term equivalence between the long-memory filter and the usual discrete time one $(1-L)^{-\alpha}$ *does no imply* that the standard parametrization FARIMA$(1, \alpha, 0)$ is well-suited in our framework. Indeed, one can show that the usual discrete time filter $(1-L)^{-\alpha}$ introduces some mixing between long and short term characteristics whereas the parsimonious continuous time model doesn't.[34] This feature clearly puts the FSV at an advantage with regard to the discrete time SV and GARCH long-memory models.

5 Statistical Inference

Evaluating the likelihood function of ARCH models is a relatively straightforward task. In sharp contrast for SV models it is impossible to obtain explicit expressions for the likelihood function. This is a generic feature common to almost all nonlinear latent variable models. The lack of estimation procedures for SV models made them for a long time an unattractive class of models in comparison to ARCH. In recent years, however, remarkable progress has been made regarding the estimation of nonlinear latent variable models in general and SV models in particular. A flurry of methods are now available and are up and running on computers with ever increasing CPU performance. The early attempts to estimate SV models used a GMM procedure. A prominent example is Melino and Turnbull (1990). Section 5.1 is devoted to GMM estimation in the context of SV models. Obviously, GMM is not designed to handle continuous time diffusions as it requires discrete time processes satisfying mixing conditions. A continuous time GMM approach, developed by Hansen and Scheinkman (1994), involves moment conditions directly drawn from the continuous time representation of the process. This approach is discussed in Section 5.3. In between, namely in section 5.2, we discuss the QML approach suggested by Harvey, Ruiz and Shephard (1994) and Nelson (1988). It relies on the fact that the nonlinear (Gaussian) SV model can be transformed into a linear non-Gaussian state space model as in Section 3, and from this a Gaussian quasi-likelihood can be computed . None of the methods covered in Sections 5.1 through 5.3 involve simulation. However, increased computer power has made simulation-based estimation techniques increasingly popular. The simulated method of moments, or simulation-based GMM approach proposed by Duffie and Singleton (1993), is a first example which is covered in Section 5.4. Next we discuss the indirect inference approach of Gouriéroux, Monfort and Renault (1993) and the moment matching methods of Gallant and Tauchen (1994) in Section 5.5. Finally, Section 5.6 covers a very large class of estimators using computer intensive Markov Chain Monte Carlo methods applied in the context of SV models by Jacquier, Polson and Rossi

[34] Namely, $(1-L_n)^\alpha \log \hat{\sigma}^n_{j/n}$ is not an AR(1) process.

(1994) and Kim and Shephard (1994), and simulation based ML estimation proposed in Danielsson (1994) and Danielsson and Richard (1993).

In each section we will only try to limit our focus to the use of estimation procedures in the context of SV models and avoid details regarding econometric theory. Some useful references to complement the material which will be covered are (1) Hansen (1992), Gallant and White (1988), Hall (1993) and Ogaki (1993) for GMM estimation, (2) Gouriéroux and Monfort (1993b) and Wooldridge (1994) for QMLE, (3) Gouriéroux and Monfort (1995) and Tauchen (1995) for simulation based econometric methods including indirect inference and moment matching, and finally (4) Geweke (1995) and Shephard (1995) for Markov Chain Monte Carlo methods.

5.1 Generalized Method of Moments

Let us consider the simple version of the discrete time SV as presented in equations (3.1.2) and (3.1.3) with the additional assumption of normality for the probability distribution of the innovation process (ε_t, η_t). This log-normal SV model has been the subject of at least two extensive Monte Carlo studies on GMM estimation of SV models. They were conducted by Andersen and Sørensen (1993) and Jacquier, Polson and Rossi (1994). The main idea is to exploit the stationary and ergodic properties of the SV model which yield the convergence of sample moments to their unconditional expectations. For instance, the second and fourth moments are simple expressions of σ^2 and σ_h^2, namely $\sigma^2\exp(\sigma_h^2/2)$ and $3\sigma^4\exp(2\sigma_h^2)$ respectively. If these moments are computed in the sample, σ_h^2 can be estimated directly from the sample kurtosis, $\widehat{\kappa}$, which is the ratio of the fourth moment to the second moment squared. The expression is just $\widehat{\sigma}_h^2 = \log(\widehat{\kappa}/3)$. The parameter σ^2 can then be estimated from the second moment by substituting in this estimate of σ_h^2. We might also compute the first-order autocovariance of y_t^2, or simply the sample mean of $y_t^2 y_{t-1}^2$ which has expectation $\sigma^4\exp(\{1+\phi\}\sigma_h^2)$ and from which, given the estimate of σ^2 and σ_h^2 , it is straightforward to get an estimate of ϕ.

The above procedure is an example of the application of the method of moments. In general terms, m moments are computed. For a sample of size T, let $g_T(\beta)$ denotes the $m \times 1$ vector of differences between each sample moment and its theoretical expression in terms of the model parameters β. The generalized method of moments (GMM) estimator is constructed by minimizing the criterion function

$$\hat{\beta}_T = Arg \min_{\beta} g_T(\beta)' W_T g_T(\beta)$$

where W_T is an $m \times m$ weighting matrix reflecting the importance given to matching each of the moments. When ε_t and η_t are mutually independent, Jacquier, Polson and Rossi (1994) suggest using 24 moments. The first four are given by (**??**) for $c = 1, 2, 3, 4$, while the analytic expression for the others is:

$$E[\mid y_t^c y_{t-\tau}^c \mid] = \left\{\sigma^{2c} 2^c [,\ \left(\tfrac{c}{2} + \tfrac{1}{2}\right)]^2/\pi\right\} \exp(\tfrac{c^2}{4}\sigma_h^2[1+\phi^\tau]), c = 1, 2, \tau = 1, 2, .., 10.^{35}$$

In the more general case when ε_t and η_t are correlated, Melino and Turnbull (1990) included estimates of : $E[\mid y_t \mid y_{t-\tau}]$,$\tau = 0, \pm 1, \pm 2, ..., 10$. They presented an explicit expression in the case of $\tau = 1$ and show that its sign is entirely determined by ρ.

[35] A simple way to derive these moment conditions is via a two-step approach similar in spirit to (2.4.8) and (2.4.9) or (3.2.3).

The GMM method may also be extended to handle a non-normal distribution for ε_t. The required analytic expressions can be obtained as in section 3.2. On the other hand, the analytic expression of unconditional moments presented in section 2.4 for the general SARV model may provide the basis of GMM estimation in more general settings (see Andersen (1994)).

From the very start we expect the GMM estimator not to be efficient. The question is how much inefficiency should be tolerated in exchange for its relative simplicity. The generic setup of GMM leaves unspecified the number of moment conditions, except for the minimal number required for identification, as well as the explicit choice of moments. Moreover, the computations of the weighting matrix is also an issue since many options exist in practice. The extensive Monte Carlo studies of Andersen and Sørensen (1993) and Jacquier, Polson and Rossi (1994) attempted to answer these many outstanding questions. In general they find that GMM is a fairly inefficient procedure primarily stemming from the stylized fact, noted in section 2.2, that ϕ in equation (3.1.3) is quite close to unity in most empirical findings because volatility is highly persistent. For parameter values of ϕ close to unity convergence to unconditional moments is extremely slow suggesting that only large samples can rescue the situation. The Monte Carlo study of Andersen and Sørensen (1993) provides some guidance on how to control the extent of the inefficiency, notably by keeping the number of moment conditions small. They also provide specific recommendations for the choice of weighting matrix estimators with data-dependent bandwidth using the Bartlett kernel.

5.2 Quasi Maximum Likelihood Estimation

5.2.1 The Basic Model

Consider the linear state space model described in sub-section 3.4.1, in which (**??**) is the measurement equation and (**??**) is the transition equation. The QML estimators of the parameters ϕ, σ_η^2 and the variance of ξ_t, σ_ξ^2, are obtained by treating ξ_t and η_t as though they were normal and maximizing the prediction error decomposition form of the likelihood obtained via the Kalman filter. As noted in Harvey, Ruiz and Shephard (1994), the quasi maximum likelihood (QML) estimators are asymptotically normal with covariance matrix given by applying the theory in Dunsmuir(1979, p. 502). This assumes that η_t and ξ_t have finite fourth moments and that the parameters are not on the boundary of the parameter space.

The parameter ω can be estimated at the same time as the other parameters. Alternatively, it can be estimated as the mean of the log y_t^2's, since this is asymptotically equivalent when ϕ is less than one in absolute value.

Application of the QML method does not require the assumption of a specific distribution for ε_t. We will refer to this as unrestricted QML. However, if a distribution is assumed, it is no longer necessary to estimate σ_ξ^2, as it is known, and an estimate of the scale factor, σ^2, can be obtained from the estimate of ω. Alternatively, it can be obtained as suggested in sub-section 3.4.1.

If unrestricted QML estimation is carried out, a value of the parameter determining a particular distribution within a class may be inferred from the estimated variance of ξ_t. For example in the case of the Student's t, ν may be determined from the knowledge that the theoretical value of the variance of ξ_t is $4.93 + \psi'(\nu/2)$ (where $\Psi(\cdot)$ is the digamma function introduced in section 3.2.2).

5.2.2 Asymmetric Model

In an asymmetric model, QML may be based on the modified state space form in (**??**). The parameters σ_ξ^2, σ_η^2, ϕ, μ^*, and γ^* can be estimated via the Kalman filter without any distributional assumptions, apart

from the existence of fourth moments of η_t and ξ_t and the joint symmetry of ξ_t and η_t. However, if an estimate of ρ is wanted it is necessary to make distributional assumptions about the disturbances, leading to formulae like (??) and (??). These formulae can be used to set up an optimization with respect to the original parameters $\sigma^2, \sigma_\eta^2, \phi$ and ρ. This has the advantage that the constraint $|\rho| < 1$ can be imposed. Note that any t-distribution gives the same relationship between the parameters, so within this class it is not necessary to specify the degrees of freedom.

Using the QML method with both the original disturbances assumed to be Gaussian, Harvey and Shephard (1993) estimate a model for the CRSP daily returns on a value weighted US market index for 3rd July 1962 to 31st December 1987. These data were used in the paper by Nelson (1991) to illustrate his EGARCH model. The empirical results indicate a very high negative correlation, suggesting that the Black-Scholes option pricing equation will be quite badly biased.

5.2.3 QML in the Frequency Domain

For a long memory SV model, QML estimation in the time domain becomes relatively less attractive because the state space form (SSF) can only be used by expressing h_t as an autoregressive or moving average process and truncating at a suitably high lag. Thus the approach is cumbersome, though the initial state covariance matrix is easily constructed, and the truncation does not affect the asymptotic properties of the estimators. If the autoregressive approximation, and therefore the SSF, is not used, time domain QML requires the repeated construction and inversion of the $T \times T$ covariance matrix of the $\log y_t^{2\prime}s$; see Sowell (1992). On the other hand, QML estimation in the frequency domain is no more difficult than it is in the AR(1) case. Cheung and Diebold (1994) present simulation evidence which suggests that although time domain estimation is more efficient in small samples, the difference is less marked when a mean has to be estimated.

The frequency domain (quasi) log-likelihood function is, neglecting constants,

$$\log L = -\frac{1}{2}\sum_{j=1}^{T-1} \log g_j - \pi \sum_{j=1}^{T-1} I(\lambda_j)/g_j \tag{5.2.1}$$

where $I(\lambda_j)$ is the sample spectrum of the $\log y_t^{2\prime}s$ and g_j is the spectral generating function (SGF), which for (??) is

$$g_j = \sigma_\eta^2[2(1-\cos\lambda_j)]^{-d} + \sigma_\xi^2$$

Note that the summation in (??) is from $j = 1$ rather than $j = 0$. This is because g_0 cannot be evaluated for positive d . However, the omission of the zero frequency does remove the mean. The unknown parameters are $\sigma_\eta^2, \sigma_\xi^2$ and d, but σ_ξ^2 may be concentrated out of the likelihood function by a reparameterisation in which σ_η^2 is replaced by the signal-noise ratio $q = \sigma_\eta^2/\sigma_\xi^2$. On the other hand if a distribution is assumed for ε_t, then σ_ξ^2 is known. Breidt, Crato and de Lima (1993) show the consistency of the QML estimator.

When d lies between 0.5 and one, h_t is nonstationary, but differencing the $\log y_t^{2\prime}s$ yields a zero mean stationary process, the SGF of which is

$$g_j = \sigma_\eta^2[2(1-\cos\lambda_j)]^{1-d} + 2(1-\cos\lambda_j)\sigma_\xi^2$$

One of the attractions of long memory models is that inference is not affected by the kind of unit root issues which arise with autoregressions. Thus a likelihood based test of the hypotheses that $d = 1$ against the alternative that it is less than one can be constructed using standard theory; see Robinson (1993).

5.2.4 Comparison of GMM and QML

Simulation evidence on the finite sample performance of GMM and QML can be found in Ruiz (1994), Harvey and Shephard (1996), Jacquier, Polson and Rossi (1994), Andersen and Sørensen (1993) and Breidt and Carriquiry (1995). The general conclusion seems to be that QML gives estimates with a smaller MSE when the volatility is relatively strong as reflected in a high coefficient of variation. This is because the normally distributed volatility component in the measurement equation, (??), is large relative to the non-normal error term. With a lower coefficient of variation, GMM dominates. However, in this case Jacquier, Polson and Rossi (1994, p. 383) observe that "...the performance of both the QML and GMM estimators deteriorates rapidly." In other words the case for one of the more computer intensive methods outlined in Section 5.6 becomes stronger.

Other things being equal, an AR coefficient, ϕ, close to one tends to favor QML because the autocorrelations are slow to die out and are hence captured less well by the moments used in GMM. For the same reason, GMM is likely to be rather poor in estimating a long memory model.

The attraction of QML is that it is very easy to implement and it extends easily to more general models, for example nonstationary and multivariate ones. At the same time, it provides filtered and smoothed estimates of the state, and predictions. The one-step ahead prediction errors can also be used to construct diagnostics, such as the Box-Ljung statistic, though in evaluating such tests it must be remembered that the observations are non-normal. Thus even if the hyperparameters are eventually estimated by another method, QML may have a valuable role to play in finding a suitable model specification.

5.3 Continuous Time GMM

Hansen and Scheinkman (1995) propose to estimate continuous time diffusions using a GMM procedure specifically tailored for such processes. In section 5.1 we discussed estimation of SV models which are either explicitly formulated as discrete time processes or else are discretizations of the continuous time diffusions. In both cases inference is based on minimizing the difference between unconditional moments and their sample equivalent. For continuous time processes Hansen and Scheinkman (1995) draw directly upon the diffusion rather than its discretization to formulate moment conditions. To describe the generic setup of the method they proposed let us consider the following (multivariate) system of n diffusion equations:

$$dy_t = \mu(y_t;\theta)dt + \sigma(y_t;\theta)dW_t \tag{5.3.1}$$

A comparison with the notation in section 2 immediately draws attention to certain limitations of the setup. First, the functions $\mu_\theta(\cdot) \equiv \mu(\cdot;\theta)$ and $\sigma_\theta(\cdot) \equiv \sigma(\cdot;\theta)$ are parameterized by y_t only which restricts the state variable process U_t in section 2 to contemporaneous values of y_t. The diffusion in (5.3.1) involves a general vector process y_t, hence y_t could include a volatility process to accommodate SV models. Yet, the y_t vector is assumed observable. For the moment we will leave these issues aside, but return to them at the end of the section. Hansen and Scheinkman (1995) consider the infinitesimal operator A defined for a class of square integrable functions φ: $\boldsymbol{R}^n \to \boldsymbol{R}$ as follows:

$$A_\theta\varphi(y) = \frac{d\varphi(y)}{dy'}\mu_\theta(y) + \frac{1}{2}Tr\left(\sigma_\theta(y)\,\sigma_\theta'(y)\,\frac{d^2\varphi(y)}{dydy'}\right). \tag{5.3.2}$$

Because the operator is defined as a limit, namely :

$$A_\theta\varphi(y) = \lim_{t\to 0} t^{-1}[\,\boldsymbol{E}(\varphi(y_t)|\,y_o = y) - y],$$

it does not necessarily exist for all square integrable functions φ but only for a restricted domain D. A set of moment conditions can now be obtained for this class of functions $\varphi \in D$. Indeed, as shown for instance by Revuz and Yor (1991), the following equalities hold :

$$EA_\theta \varphi(y_t) = 0, \tag{5.3.3}$$

$$E\left[A_\theta \varphi(y_{t+1})\,\tilde{\varphi}(y_t) - \varphi(y_{t+1})\,A_\theta^* \tilde{\varphi}(y_t)\right] = 0, \tag{5.3.4}$$

where A_θ^* is the adjoint infinitesimal operator of A_θ for the scalar product associated with the invariant measure of the process y.[36] By choosing an appropriate set of functions, Hansen and Scheinkman exploit moment conditions (5.3.3) and (5.3.4) to construct a GMM estimator of θ.

The choice of the function $\varphi \in D$ and $\tilde{\varphi} \in D^*$ determines what moments of the data are used to estimate the parameters. This obviously raises questions regarding the choice of functions to enhance efficiency of the estimator but first and foremost also the identification of θ via the conditions (5.3.3) and (5.3.4). It was noted in the beginning of the section that the multivariate process y_t, in order to cover SV models, must somehow include the latent conditional variance process. Gouriéroux and Monfort (1994, 1995) point out that since the moment conditions based on φ and $\tilde{\varphi}$ cannot include any latent process it will often (but not always) be impossible to attain identification of all the parameters, particularly those governing the latent volatility process. A possible remedy is to augment the model with observations indirectly related to the latent volatility process, in a sense making it observable. One possible candidate would be to include in y_t both the security price and the Black-Scholes implied volatilities obtained through option market quotations for the underlying asset. This approach is in fact suggested by Pastorello, Renault and Touzi (1993) although not in the context of continuous time GMM but instead using indirect inference methods which will be discussed in section 5.5.[37] Another possibility is to rely on the time deformation representation of SV models as discussed in the context of continuous time GMM by Conley et al. (1995).

5.4 Simulated Method of Moments

The estimation procedures discussed so far do not involve any simulation techniques. From now on we cover methods combining simulation and estimation beginning with the simulated method of moments (SMM) estimator, which is covered by Duffie and Singleton (1993) for time series processes.[38] In section 5.1 we noted that GMM estimation of SV models is based on minimizing the distance between a set of chosen sample moments and unconditional population moments expressed as analytical functions of the model parameters. Suppose now that such analytical expressions are hard to obtain. This is particularly the case when such expressions involve marginalizations with respect to a latent process such a stochastic volatility process. Could we then simulate data from the model for a particular value of the parameters and match moments from the simulated data with sample moments as a substitute? This strategy is precisely what SMM is all about. Indeed, quite often it is fairly straightforward to simulate processes and therefore take advantage of the SMM procedure. Let us consider again as point of reference and illustration the

[36] Please note that A_θ^* is again associated with a domain D^* so that $\varphi \in D$ and $\tilde{\varphi} \in D^*$ in (5.3.4).

[37] It was noted in section 2.1.3 that implied volatilities are biased. The indirect inference procedures used by Pastorello, Renault and Touzi (1993) can cope with such biases, as will be explained in section 5.5. The use of option price data is further discussed in section 5.7.

[38] SMM was originally proposed for cross-section applications, see Pakes and Pollard (1989) and McFadden (1989). See also Gouriéroux and Monfort (1993a).

(multivariate) diffusion of the previous section (equation (5.3.1)) and conduct H simulations $i = 1, ..., H$ using a discretization:

$\Delta \widehat{y}_t^i(\theta) = \mu\left(\widehat{y}_t^i(\theta); \theta\right) + \sigma\left(\widehat{y}_t^i(\theta); \theta\right) \varepsilon_t$ and $i = 1, ..., H$ and $t = 1, ..., T$

where $\widehat{y}_t(\theta)$ are simulated given a parameter θ and ε_t is i.i.d. Gaussian.[39] Subject to identification and other regularity conditions one then considers

$$\widehat{\theta}_T^H = Arg \min_{\theta} \parallel f(y_1, ...y_T) - \frac{1}{H} \sum_{i=1}^{H} f\left(\widehat{y}_1^i(\theta), ..., \widehat{y}_T^i(\theta)\right) \parallel$$

with a suitable choice of norm, i.e. weighting matrix for the quadratic form as in GMM, and function f of the data, i.e. moment conditions. The asymptotic distribution theory is quite similar to that of GMM, except that simulation introduces an extra source of random error affecting the efficiency of the SMM estimator in comparison to its GMM counterpart. The efficiency loss can be controlled by the choice of H.[40]

5.5 Indirect Inference and Moment Matching

The key insight of the indirect inference approach of Gouriéroux, Monfort and Renault (1993) and the moment matching approach of Gallant and Tauchen (1994) is the introduction of an auxiliary model parameterized by a vector, say β, in order to estimate the model of interest. In our case the latter is the SV model.[41] In the first subsection we will describe the general principle while a second one will focus exclusively on estimating diffusions.

5.5.1 The Principle

We noted at the beginning of section 5 that ARCH type models are relatively easy to estimate in comparison to SV models. For this reason an ARCH type model may be a possible candidate as an auxiliary model. An alternative strategy would be to try to summarize the features of the data via a SNP density as developed by Gallant and Tauchen (1989). This empirical SNP density, or more specifically its score, could also fulfill the role of auxiliary model. Other possibilities could be considered as well. The idea is then to use the auxiliary model to estimate β, so that:

$$\widehat{\beta}_T = Arg \max_{\beta} \sum_{t=1}^{T} \log f^*\left(y_t \mid y_{t-1}, \beta\right) \tag{5.5.1}$$

where we restrict our attention here to a simple dynamic model with one lag for the purpose of illustration. The objective function f^* in (5.5.1) can be a pseudo-likelihood function when the auxiliary model is deliberately misspecified to facilitate estimation. As an alternative f^* can be taken from the class of SNP densities.[42] Gouriéroux, Monfort and Renault then propose to estimate the same parameter vector β not

[39] We discuss in detail the simulation techniques in the next section. Indeed, to control for the discretization bias, one has to simulate with a finer sampling interval.

[40] The asymptotic variance of the SMM estimator depends on H through a factor$(1 + H^{-1})$, see e.g. Gouriéroux and Monfort (1995).

[41] It is worth noting that the simulation based inference methods we will describe here are applicable to many other types of models for cross-sectional, time series and panel data.

[42] The discussion should not leave the impression that the auxiliary model can only be estimated via ML-type estimators. Any root T consistent asymptotically normal estimations procedure may be used.

using the actual sample data but instead using samples $\{\widehat{y}_t^i(\theta)\}_{t=1}^T$ simulated $i = 1, ...H$ times drawn from the model of interest given θ. This yields a new estimator of β, namely:

$$\widehat{\beta}_{HT}(\theta) = Arg\max_{\beta}(1/H)\sum_{i=1}^{H}\sum_{t=1}^{T}\log f^*\left(\widehat{y}_t^i(\theta) \mid \widehat{y}_{t-1}^i(\theta), \beta\right). \tag{5.5.2}$$

The next step is to minimize a quadratic distance using a weighting matrix W_T to choose an indirect estimator of θ based on H simulation replications and a sample of T observations, namely:

$$\widehat{\theta}_{HT} = Arg\min_{\beta}\left(\widehat{\beta}_T - \widehat{\beta}_{HT}(\theta)\right)' W_T\left(\widehat{\beta}_T - \widehat{\beta}_{HT}(\theta)\right) \tag{5.5.3}$$

The approach of Gallant and Tauchen (1994) avoids the step of estimating $\widehat{\beta}_{HT}(\theta)$ by computing the score function of f^* and minimizing a quadratic distance similar to (5.5.3) but involving the score function evaluated at $\widehat{\beta}_T$ and replacing the sample data by simulated series generated by the model of interest. Under suitable regularity conditions the estimator $\widehat{\theta}_{HT}$ is root T consistent and asymptotically normal. As with GMM and SMM there is again an optimal weighting matrix. The resulting asymptotic covariance matrix depends on the number of simulations in the same way the SMM estimator depends on H.

Gouriéroux, Monfort and Renault (1993) illustrated the use of indirect inference estimator with a simple example that we would like to briefly discuss here. Typically AR models are easy to estimate while MA models require more elaborate procedures. Suppose the model of interest is a moving average model of order one with parameter θ. Instead of estimating the MA parameter directly from the data they propose to estimate an AR(p.) model involving the parameter vector β. The next step then consists of simulating data using the MA model and proceed further as described above.[43] They found that the indirect inference estimator for $\widehat{\theta}_{HT}$ appeared to have better finite sample properties than the more traditional maximum likelihood estimators for the MA parameter. In fact the indirect inference estimator exhibited features similar to the median unbiased estimator proposed by Andrews (1993). These properties were confirmed and clarified by Gouriéroux, Renault and Touzi (1994) who studied the second order asymptotic expansion of indirect inference estimators and their ability to reduce finite sample bias.

5.5.2 Estimating Diffusions

Let us consider the same diffusion equation as in section 5.3 which dealt with continuous time GMM, namely:

$$dy_t = \mu(y_t;\theta)\,dt + \sigma(y_t;\theta)\,dW_t \tag{5.5.4}$$

In section 5.3 we noted that the above equation holds under certain restrictions such as the functions μ and σ being restricted to y_t as arguments. While these restrictions were binding for the setup of section 5.3 this will not be the case for the estimation procedures discussed here. Indeed, equation (5.5.4) is only used as an illustrative example. The diffusion is then simulated either via exact discretizations or some type of approximate discretization (e.g. Euler or Mil'shtein, see Pardoux and Talay (1985) or Kloeden and Platten (1992) for further details). More precisely we define the process $y_t^{(\delta)}$ such that:

[43] Again one could use a score principle here, following Gallant and Tauchen (1994). In fact in a linear Gaussian setting the SNP approach to fit data generated by a MA (1) model would be to estimate on AR(p) model. Ghysels, Khalaf and Vodounou (1994) provide a more detailed discussion of score-based and indirect inference estimators of MA models as well as their relation with more standard estimators.

$$y^{(\delta)}_{(k+1)\delta} = y^{(\delta)}_{k\delta} + \mu\left(y^{(\delta)}_{k\delta};\theta\right)\delta + \sigma\left(y^{(\delta)}_{k\delta};\theta\right)\delta^{1/2}\varepsilon^{(\delta)}_{(k+1)\delta} \tag{5.5.5}$$

Under suitable regularity conditions (see for instance Strook and Varadhan (1979)) we know that the diffusion admits a unique solution (in distribution) and the process $y^{(\delta)}_t$ converges to y_t as δ goes to zero. Therefore one can expect to simulate y_t quite accurately for δ sufficiently small. The auxiliary model may be a discretization of (5.5.4) choosing $\delta = 1$. Hence, one formulates a ML estimator based on the nonlinear AR model appearing in (5.5.5) setting $\delta = 1$. To control for the discretization bias one can simulate the underlying diffusion with $\delta = 1/10$ or $1/20$, for instance, and aggregate the simulated data to correspond with the sampling frequency of the DGP. Broze, Scaillet and Zakoïan (1994a) discuss of the simulation step size on the asymptotic distribution.

The use of simulation-based inference methods becomes particularly appropriate and attractive when diffusions involve latent processes, such as is the case with SV models. Gouriéroux and Monfort (1994, 1995) discuss several examples and study their performance via Monte Carlo simulation. It should be noted that estimating the diffusion at a coarser discretization is not the only possible choice of auxiliary model. Indeed, Pastorello, Renault and Touzi (1993), Engle and Lee (1994) and Gallant and Tauchen (1994) suggest the use of ARCH-type models.

There have been several successful applications of these methods to financial time series. They include Broze et al.(1994b), Engle and Lee (1994), Gallant, Hsieh and Tauchen (1994), Gallant and Tauchen (1994, 1995), Ghysels, Gouriéroux and Jasiak (1995b), Ghysels and Jasiak (1994a and b), Pastorello et al. (1993), among others.

5.6 Likelihood-based and Bayesian Methods

In a Gaussian linear state space model the likelihood function is constructed from the one step ahead prediction errors. This prediction error decomposition form of the likelihood is used as the criterion function in QML, but of course it is not the exact likelihood in this case. The exact filter proposed by Watanabe (1993) will, in principle, yield the exact likelihood. However, as was noted in section 3.4.2, because this filter uses numerical integration, it takes a long time to compute and if numerical optimization is to be carried out with respect to the hyperparameters it becomes impractical.

Kim and Shephard (1994) work with the linear state space form used in QML but approximate the $\log(\chi^2)$ distribution of the measurement error by a mixture of normals. For each of these normals, a prediction error decomposition likelihood function can be computed. A simulated EM algorithm is used to find the best mixture and hence calculate approximate ML estimates of the hyperparamaters.

The exact likelihood function can also be constructed as a mixture of distributions for the observations conditional on the volatilities, that is

$$L(y;\phi,\sigma_\eta^2,\sigma^2) = \int p(y|h)p(h)dh$$

where y and h contain the T elements of y_t and h_t respectively. This expression can be written in terms of the $\sigma_t^{2\prime}s$, rather than their logarithms, the $h_t's$, but it makes little difference to what follows. Of course the problem is that the above likelihood has no closed form, so it must be calculated by some kind of simulation method. Excellent discussions can be found in Shephard (1995) and in Jacquier, Polson and Rossi (1994), including the comments. Conceptually, the simplest approach is to use Monte Carlo integration by drawing from the unconditional distribution of h for given values of the parameters,$(\phi,\sigma_\eta^2,\sigma^2)$, and estimating the

likelihood as the average of the $p(y|h)'s$. This is then repeated, searching over ϕ, σ_η^2 until the maximum of the simulated likelihood is found. As it stands this procedure is not very satisfactory, but it may be improved by using ideas of importance sampling. This has been implemented for ML estimation of SV models by Danielsson and Richard (1993) and Danielsson (1994). However, the method becomes more difficult as the sample size increases.

A more promising way of attacking likelihood estimation by simulation techniques is to use Markov Chain Monte Carlo (MCMC) to draw from the distribution of volatilities conditional on the observations. Ways in which this can be done were outlined in sub-section 3.4.2 on nonlinear filters and smoothers. Kim and Shephard (1994) suggest a method of computing ML estimators by putting their multimove algorithm within a simulated EM algorithm. Jacquier, Polson and Rossi (1994) adopt a Bayesian approach in which the specification of the model has a hierarchical structure in which a prior distribution for the hyperparameters, $\varphi = (\sigma_\eta, \phi, \sigma)'$, joins the conditional distributions, y| h and h | φ. (Actually the $\sigma_t's$ are used rather than the $h_t's$). The joint posterior of h and φ is proportional to the product of these three distributions, that is $p(h, \varphi|y) \propto p(y|h)p(h|\varphi)\,p(\varphi)$. The introduction of h makes the statistical treatment tractable and is an example of what is called data augmentation; see Tanner and Wong (1987). From the joint posterior, $p(h, \varphi|y)$, the marginal $p(h|y)$ solves the smoothing problem for the unobserved volatilities, taking account of the sampling variability in the hyperparameters. Conditional on h, the posterior of $\varphi, p(\varphi|h, y)$ is simple to compute from standard Bayesian treatment of linear models. If it were also possible to sample directly from $p(h|\varphi, y)$ at low cost, it would be straightforward to construct a Markov chain by alternating back and forth drawing from $p(\varphi|h, y)$ and $p(h|\varphi, y)$. This would produce a cyclic chain, a special case of which is the Gibbs sampler. However, as was noted in sub-section 3.4.2, Jacquier, Polson and Rossi (1994) show that it is much better to decompose $p(h|\varphi, y)$ into a set of univariate distributions in which each h_t, or rather σ_t, is conditioned on all the others.

The prior distribution for ω, the parameters of the volatility process in JPR (1994), is the standard conjugate prior for the linear model, a (truncated) Normal-Gamma. The priors can be made extremely diffuse while remaining proper. JPR conduct an extensive sampling experiment to document the performance of this and more traditional approaches. Simulating stochastic volatility series, they compare the sampling performances of the posterior mean with that of the QML and GMM point estimates. The MCMC posterior mean exhibit root mean squared errors anywhere between half and a quarter of the size of the GMM and QML point estimates. Even more striking are the volatility smoothing performance results. The root mean squared error of the posterior mean of h_t produced by the Bayesian filter is 10% smaller than the point estimate produced by an approximate Kalman filter supplied with the *true* parameters.

Shephard and Kim in their comment of JPR (1994) point out that for very high ϕ and small σ_η, the rate of convergence of the JPR algorithm will slow down. More draws will then be required to obtain the same amount of information. They propose to approximate the volatility disturbance with a discrete mixture of normals. The benefit of the method is that a draw of the vector h is then possible, faster than T draws from each h_t. However this is at the cost that the draws navigate in a much higher dimensional space due to the discretisation effected. Also, the convergence of chains based upon discrete mixtures is sensitive to the number of components and their assigned probability weights. Mahieu and Schotman (1994) add some generality to the Shephard and Kim idea by letting the data produce estimates of the characteristics of the discretized state space (probabilities, mean and variance).

The original implementation of the JPR algorithm was limited to a very basic model of stochastic volatility, AR(1) with uncorrelated mean and volatility disturbances. In a univariate setup, correlated disturbances are likely to be important for stock returns, i.e., the so called leverage effect. The evidence in Gallant, Rossi, and Tauchen (1994) also points at non normal conditional errors with both skewness and

kurtosis. Jacquier, Polson, and Rossi (1995a) show how the hierarchical framework allows the convenient extension of the MCMC algorithm to more general models. Namely, they estimate univariate stochastic volatility models with correlated disturbances, and skewed and fat-tailed variance disturbance, as well as multivariate models. Alternatively, the MCMC algorithm can be extended to a factor structure. The factors exhibit stochastic volatility and can be observable or non observable.

5.7 Inference and Option Price Data

Some of the continuous time SV models currently found in the literature were developed to answer questions regarding derivative security pricing. Given this rather explicit link between derivates and SV diffusions it is perhaps somewhat surprising that relatively little attention has been paid to the use of option price data to estimate continuous time diffusions. Melino (1994) in his survey in fact notes: "*Clearly, information about the stochastic properties of an asset's price is contained both in the history of the asset's price and the price of any options written on it. Current strategies for combining these two sources of information, including implicit estimation, are uncomfortably ad hoc. Statistically speaking, we need to model the source of the prediction errors in option pricing and to relate the distribution of these errors to the stock price process.* For example implicit estimation, like computation of BS implied volatilities, is certainly uncomfortably ad hoc from a statistical point of view. In general, each observed option price introduces one source of prediction error when compared to a pricing model. The challenge is to model the *joint* nondegenerate probability distribution of options and asset prices via a number of unobserved state variables. This approach has been pursued in a number of recent papers, including Christensen (1992), Renault and Touzi (1992), Pastorello et al. (1993), Duan (1994) and Renault (1995).

Christensen (1992) considers a pricing model for n assets as a function of a state vector x_t which is $(l+n)$ dimensional and divided in a l-dimensional observed (z_t) and n-dimensional unobserved (ω_t) component. Let p_t be the price vector of the n assets, then:

$$p_t = m\left(z_t, \omega_t, \theta\right) \tag{5.7.1}$$

Equation (5.7.1) provides a one-to-one relationship between the n latent state variables ω_t and the n observed prices p_t, for given z_t and θ. From a financial viewpoint, it implies that the n assets are appropriate instruments to complete the markets if we assume that the observed state variables z_t are already mimicked by the price dynamics of other (primitive) assets. Moreover, from a statistical viewpoint it allows full structural maximum likelihood estimation provided the log-likelihood function for observed prices can be deduced easily from a statistical model for x_t. For instance, in a Markovian setting where, conditionally on x_0, the joint distribution of $x_1^T = (x_t)_{1\leq t\leq T}$ is given by the density:

$$f_x\left(x_1^T \left| x_0, \theta\right.\right) = \prod_{t=1}^{T} f\left(z_t, \omega_t \left| z_{t-1}, \omega_{t-1}, \theta\right.\right) \tag{5.7.2}$$

the conditional distribution of data $D_1^T = (p_t, z_t)_{1\leq t\leq T}$ given $D_0 = (p_0, z_0)$ is obtained by the usual Jacobian formula:

$$\begin{aligned} f_D\left(D_1^T \left| D_0, \theta\right.\right) &= \prod_{t=1}^{T} f\left[z_t, m_\theta^{-1}\left(z_t, p_t\right) \middle| z_{t-1}, m_\theta^{-1}\left(z_{t-1}, p_{t-1}\right), \theta\right] x \\ &\quad \left|\nabla_\omega m\left(z_t, m_\theta^{-1}\left(z_t, p_t\right), \theta\right)\right|^{-1} \end{aligned} \tag{5.7.3}$$

where $m_\theta^{-1}(z,.)$ is the ω-inverse of $m(z,.,\theta)$ defined formally by $m_\theta^{-1}(z, m(z,\omega,\theta)) = \omega$ while $\nabla_\omega m(\cdot)$ represents the columns corresponding to ω of the Jacobian matrix. This MLE using price data of derivatives was proposed independently by Christensen (1992) and Duan (1994). Renault and Touzi (1992) were instead more specifically interested in the Hull and White option pricing formula with: $z_t = S_t$ observed underlying asset price, and $\omega_t = \sigma_t$ unobserved stochastic volatility process. Then with the joint process $x_t = (S_t, \sigma_t)$ being Markovian we have a call price of the form:

$$C_t = m(x_t, \theta, K)$$

where $\theta = \left(\alpha', \gamma'\right)$ involves two types of parameters: (1) the vector α of parameters describing the dynamics of the joint process $x_t = (S_t, \sigma_t)$ which under the equivalent martingale measure allows to compute the expectation with respect to the (risk-neutral) conditional probability distribution of $\gamma^2(t, t+h)$ given σ_t; and (2) the vector γ of parameters which characterize the risk premia determining the relation between the risk neutral probability distribution of the x process and the Data Generating Process.

Structural MLE is often difficult to implement. This motivated Renault and Touzi (1992) and Pastorello, Renault and Touzi (1993) to consider less efficient but simpler and more robust procedures involving some proxies of the structural likelihood (5.7.3).

To illustrate these procedures let us consider the standard log-normal SV model in continuous time:

$$d\log\sigma_t = k(a - \log\sigma_t)\,dt + c dW_t^\sigma. \tag{5.7.4}$$

Standard option pricing arguments allow us to ignore misspecifications of the drift of the underlying asset price process. Hence, a first step towards simplicity and robustness is to isolate from the likelihood function the volatility dynamics, namely:

$$\prod_{i=1}^{n} \left(2\pi c^2\right)^{-\frac{1}{2}} \exp\left[-\left(2c^2\right)^{-1}\left(\log\sigma_{t_i} - e^{-k\Delta t}\log\sigma_{t_{i-1}} - a\left(1 - e^{-k\Delta t}\right)\right)\right]^2 \tag{5.7.5}$$

associated with a sample $\sigma_{t_i}, i = 1, \ldots, n$ and $t_i - t_{i-1} = \Delta t$. To approximate this expression one can consider a *direct* method, as in Renault and Touzi (1992) or an *indirect* method, as in Pastorello et al. (1993). The former involves calculating implied volatilities from the Hull and White model to create pseudo samples σ_{t_i} parameterized by k, a and c and computing the maximum of (5.7.5) with respect to those three parameters.[44] Pastorello et al. (1993) proposed several indirect inference methods, described in section 5.5, in the context of (5.7.5). For instance, they propose to use an indirect inference strategy involving GARCH(1,1) volatility estimates obtained form the underlying asset (also independently suggested by Engle and Lee (1994)). This produces asymptotically unbiased but rather inefficient estimates. Pastorello et al. indeed find that an indirect inference simplification of the Renault and Touzi direct procedure involving option prices is far more efficient. It is a clear illustration of the intuition that the use of option price data paired with suitable statistical methods should largely improve the accuracy of estimating volatility diffusion parameters.

5.8 Regression Models with Stochastic Volatility

A single equation regression model with stochastic volatility in the disturbance term may be written

[44] The direct maximization of (5.7.5) using BS implied volatilities has also been proposed, see e.g. Heynen, Kemna and Vorst (1994). Obviously the use of BS implied volatility induces a misspecification bias due to the BS model assumptions.

$$y_t = x_t'\beta + u_t, \quad t = 1, ..., T, \tag{5.8.1}$$

where y_t denotes the $t-th$ observation, x_t is a $k \times 1$ vector of explanatory variables, β is a $k \times 1$ vector of coefficients and $u_t = \sigma\varepsilon_t \exp(0.5h_t)$ as discussed in section 3. As a special case, the observations may simply have a non-zero mean so that $x_t'\beta = \mu \forall t$.

Since u_t is stationary, an OLS regression of y_t on x_t yields a consistent estimator of β. However it is not efficient.

For given values of the SV parameters, ϕ and σ_η^2, a smoothed estimator of h_t, $h_{t|T}$, can be computed using one of the methods outlined in section 3.4. Multiplying (5.8.1) through by $\exp(-.5h_{t|T})$ gives

$$\tilde{y}_t = \tilde{x}_t'\beta + \tilde{u}_t, t = 1, ..., T \tag{5.8.2}$$

where the $\widetilde{u_t}$'s can be thought of as heteroskedasticity corrected disturbances. Harvey and Shephard (1993) show that these disturbances have zero mean, constant variance and are serially uncorrelated and hence suggest the construction of a feasible GLS estimator

$$\tilde{\beta} = \left[\sum_{t=1}^{T} e^{-h_{t|T}} x_t x_t'\right]^{-1} \sum_{t=1}^{T} e^{-h_{t|T}} x_t y_t \tag{5.8.3}$$

In the classical heteroskedastic regression model h_t is deterministic and depends on a fixed number of unknown parameters. Because these parameters can be estimated consistently, the feasible GLS estimator has the same asymptotic distribution as the GLS estimator. Here h_t is stochastic and the MSE of its estimator is of $O(1)$. The situation is therefore somewhat different. Harvey and Shephard (1993) show that, under standard regularity conditions on the sequence of x_t, $\widetilde{\beta}$ is asymptotically normal with mean β and a covariance matrix which can be consistently estimated by

$$\widetilde{avar}(\widetilde{\beta}) = \left[\sum_{t=1}^{T} e^{-h_{t|T}} x_t x_t'\right]^{-1} \sum_{t=1}^{T} (y_t - x_t'\tilde{\beta})^2 e^{-2h_{t|T}} x_t x_t' \left[\sum_{t=1}^{T} e^{-h_{t|T}} x_t x_t'\right]^{-1} \tag{5.8.4}$$

When $h_{t|T}$ is the smoothed estimate given by the linear state space form, the analysis in Harvey and Shephard (1993) suggests that, asymptotically, the feasible GLS estimator is almost as efficient as the GLS estimator and considerably more efficient than the OLS estimator. It would be possible to replace $\exp(h_{t|T})$ by a better estimate computed from one of the methods described in section 3.4 but this may not have much effect on the efficiency of the resulting feasible GLS estimator of β.

When h_t is nonstationary, or nearly nonstationary, Hansen (1995) shows that it is possible to construct a feasible adaptive least squares estimator which is asymptotically equivalent to GLS.

6 Conclusions

No survey is ever complete. There are two particular areas we expect to flourish in the years to come but which we were not able to cover. The first is the area of market microstructures which is well surveyed in a recent review paper by Goodhart and O'Hara (1995). With the ever increasing availability of high frequency data series, we anticipate more work involving game theoretic models. These can now be estimated because of recent advances in econometric methods, similar to those enabling us to estimate diffusions. Another area where we expect interesting research to emerge is that involving nonparametric procedures to estimate

SV continuous time and derivative securities models. Recent papers include Ait-Sahalia (1994), Ait-Sahalia et al. (1994), Bossaerts, Hafner and Härdle (1995), Broadie et al. (1995), Conley et al. (1995), Elsheimer et al. (1995), Gouriéroux, Monfort and Tenreiro (1994), Gouriéroux and Scaillet (1995), Hutchinson, Lo and Poggio (1994), Lezan et al. (1995), Lo (1995), Pagan and Schwert (1992).

Research into the econometrics of Stochastic Volatility models is relatively new. As our survey has shown, there has been a burst of activity in recent years drawing on the latest statistical technology. As regards the relationship with ARCH, our view is that SV and ARCH are not necessarily direct competitors, but rather complement each other in certain respects. Recent advances such as the use of ARCH models as filters, the weakening of GARCH and temporal aggregation and the introduction of nonparametric methods to fit conditional variances, illustrate that a unified strategy for modelling volatility needs to draw on both ARCH and SV.

References

[1] Abramowitz, M. and N.C. Stegun (1970), *Handbook of Mathematical Functions.* New York : Dover Publications Inc.

[2] Ait-Sahalia, Y. (1994), "Nonparametric Pricing of Interest Rate Derivative Securities", Discussion Paper, Graduate School of Business, University of Chicago.

[3] Ait-Sahalia, Y., S.J. Bickel and T.M. Stoker (1994), "Goodness-of-Fit Tests for Regression Using Kernel Methods", Discussion Paper, University of Chicago.

[4] Amin, K.L. and V. Ng (1993) "Equilibrium Option Valuation with Systematic Stochastic Volatility", *Journal of Finance* 48, 881-910.

[5] Andersen, T.G. (1992), "Volatility", Discussion paper, Northwestern University.

[6] Andersen, T.G. (1994). "Stochastic Autoregressive Volatility : A Framework for Volatility Modeling", *Mathematical Finance* 4, 75-102.

[7] Andersen, T.G. and T. Bollerslev (1995), "Intraday Seasonality and Volatility Persistence in Financial Markets", Discussion Paper, Northwestern University.

[8] Andersen, T.G. and B. Sørensen (1993), "GMM Estimation of a Stochastic Volatility Model : A Monte Carlo Study", Discussion paper, Northwestern University.

[9] Andrews, D.W.K. (1993), "Exactly Median-Unbiased Estimation of First Order Autoregressive Unit Root Models", *Econometrica*, 61, 139-165.

[10] Bachelier, L. (1900): "Théorie de la spéculation", *Ann. Sci. Ecole Norm. Sup. 17, 21-86, [On the Random Character of Stock Market Prices (Paul H. Cootner, ed.) The MIT Press, Cambridge, Mass. 1964].*

[11] Baillie, R.T. and T. Bollerslev (1989), "The Message in Daily Exchange Rates : A Conditional Variance Tale", *Journal of Business and Economic Statistics* 7, 297-305.

[12] Baillie, R.T. and T. Bollerslev (1991), "Intra Day and Inter Day Volatility in Foreign Exchange Rates", *Review of Economic Studies* 58, 565-585.

[13] Baillie, R.T., T. Bollerslev and H. O. Mikkelsen (1993), "Fractionally integrated generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics* (forthcoming).

[14] Bajeux, I. and J.C. Rochet (1992), "Dynamic Spanning : Are Options an Appropriate Instrument?", *Mathematical Finance*, (forthcoming).

[15] Bates, D.S. (1995a), "Testing Option Pricing Models", Discussion Paper, (Wharton School, University of Pennsylvania).

[16] Bates, D.S. (1995b), "Jumps and stochastic volatility : exchange rate processes implicit in PHLX Deutschemark options", *Review of Financial Studies*, forthcoming.

[17] Beckers, S. (1981), " Standard deviations implied in option prices as predictors of future stock price variability", *Journal of Banking and Finance* 5, 363-381.

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

[19] Black, F. (1976), "Studies in Stock Price Volatility Changes", *Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section, American Statistical Association*, 177-181.

[20] Black, F. and M. Scholes (1973), "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy 81,* 637-654.

[21] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* 31, 307-327.

[22] Bollerslev, T. (1993), "Long Memory in Stochastic Volatility", Discussion paper, Northwestern Unviersity.

[23] Bollerslev, T., Y.C. Chou and K. Kroner (1992), "ARCH Modelling in Finance : A Selective Review of the Theory and Empirical Evidence", *Journal of Econometrics* 52, 201-224.

[24] Bollerslev, T. and R. Engle (1993), "Common Persistence in Conditional Variances", *Econometrica* 61, 166-187.

[25] Bollerslev, T., R. Engle and D. Nelson (1994), "ARCH Models", in R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, (North-Holland, Amsterdam).

[26] Bollerslev, T., R. Engle and J. Wooldridge (1988), "A Capital Asset Pricing Model with Time Varying Covariances", *Journal of Political Economy* 96, 116-131.

[27] Bollerslev, T. and E. Ghysels (1994), "On Periodic Autoregression Conditional Heteroskedasticity", *Journal of Business and Economic Statistics* (forthcoming).

[28] Bollerslev, T. and H. O. Mikkelsen (1995), "Modeling and Pricing Long-Memory in Stock Market volatility", *Journal of Econometrics* (forthcoming).

[29] Bossaerts, P., C. Hafner and W. Härdle (1995), "Foreign Exchange Rates Have Surprising Volatility", Discussion Paper, CentER, University of Tilburg.

[30] Bossaerts, P. and P. Hillion (1995), "Local Parametric Analysis of Hedging in Discrete Time", *Journal of Econometrics* (forthcoming).

[31] Breidt, F. J., N. Crato and P. de Lima (1993), "Modeling Long-Memory Stochastic Volatility", Discussion paper, Iowa State University.

[32] Breidt, F. J. and A.L. Carriquiry (1995), "Improved Quasi-Maximum Likelihood Estimation for Stochastic Volatility Models". Mimeo, Department of Statistics, University of Iowa.

[33] Broadie, M., J. Detemple, E. Ghysels and O. Torrès (1995), "American Options with Stochastic Volatility : A Nonparametric Approach", Discussion Paper, CIRANO.

[34] Broze, L., O. Scaillet and J.M. Zakoian (1994a), "Quasi Indirect Inference for Diffusion Processes", Discussion Paper CORE.

[35] Broze, L., O. Scaillet and J.M. Zakoian (1994b), "Testing for continuous time models of the short term interest rate", *Journal of Empirical Finance* (forthcoming).

[36] Campa, J. M. and P.H.K. Chang (1995), "Testing the expectations hypothesis on the term structure of implied volatilities in foreign exchange options", *Journal of Finance* 50, (forthcoming).

[37] Campbell, J.Y. and A.S. Kyle (1993), "Smart Money, Noise Trading and Stock Price Behaviour", *Review of Economics Studies* 60, 1-34.

[38] Canina, L. and S. Figlewski (1993), "The informational content of implied volatility", *Review of Financial Studies* 6, 659-682.

[39] Canova, F. (1992), "Detrending and Business Cycle Facts", Discussion Paper, European University Institute, Florence.

[40] Chesney, M. and L. Scott (1989), "Pricing European Currency Options : A comparison of the Modified Black-Scholes Model and a Random Variance Model", *Journal of Financial and Quantitative Analysis* 24, 267-284.

[41] Cheung, Y.-W. and F.X. Diebold (1994), "On Maximum Likelihood Estimation of the Differenciary Parameter of Fractionally - Integrated Noise with Unknown Mean", *Journal of Econometrics* 62, 301-316.

[42] Chiras, D.P. and S. Manaster (1978), "The information content of option prices and a test of market efficiency", *Journal of Financial Economics* 6, 213-234.

[43] Christensen, B.J. (1992), "Asset Prices and the Empirical Martingale Model", Discussion Paper, New York University.

[44] Christie, A. A. (1982), "The Stochastic Behavior of Common Stock Variances : Value, Leverage, and Interest Rate Effects", *Journal of Financial Economics* 10, 407-432.

[45] Clark, P.K. (1973), "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", *Econometrica* 41, 135-156.

[46] Clewlow, L and X. Xu (1993), "The Dynamics of Stochastic Volatility", Discussion Paper, University of Warwick.

[47] Comte, F. and E. Renault (1993), "Long memory continuous time models", *Journal of Econometrics* (forthcoming).

[48] Comte, F. and E. Renault (1995), "Long memory continuous time Stochastic Volatility Models", Paper presented at the HFDF-I Conference, Zürich.

[49] Conley, T., L.P. Hansen, E. Luttmer and J. Scheinkman (1995), "Estimating Subordinated Diffusions from discrete Time Data" Discussion paper, University of Chicago.

[50] Cornell, B. (1978), "Using the Options Pricing Model to Measure the Uncertainty Producing Effect of Major announcements", *Financial Management* 7, 54-59.

[51] Cox, J.C. (1975), "Notes on Option Pricing I: Constant Elasticity of Variance Diffusions", Discussion Paper, Stanford University.

[52] Cox, J.C. and S. Ross (1976), "The Valuation of Options for Alternative Stochastic Processes", *Journal of Financial Economics* 3, 145-166.

[53] Cox, J.C. and M. Rubinstein (1985), *Options Markets*, (Englewood Cliffs, Prentice-Hall, New Jersey).

[54] Dacorogna, M.M., U.A. Müller, R.J. Nagler, R.B. Olsen and O.V. Pictet (1993), "A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market", *Journal of International Money and Finance* 12, 413-438.

[55] Danielsson, J. (1994), "Stochastic Volatility in asset Prices : Estimation with Simulated Maximum Likelihood", *Journal of Econometrics* 61, 375-400.

[56] Danielsson, J. and J.F. Richard (1993), "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models", *Journal of Applied Econometrics* 3, S153-S174.

[57] Dassios, A. (1995), "Asymptotic expressions for approximations to stochastic variance models". mimeo, London School of Economics.

[58] Day, T.E. and C.M. Lewis (1988), "The behavior of the volatility implicit in the prices of stock index options", *Journal of Financial Economics* 22, 103-122.

[59] Day, T.E. and C.M. Lewis (1992), "Stock market volatility and the information content of stock index options", *Journal of Econometrics* 52, 267-287.

[60] Diebold, F.X. (1988), *Empirical Modeling of Exchange Rate Dynamics*, (Springer Verlag, New York).

[61] Diebold, F.X. and J.A. Lopez (1995), "Modeling Volatility Dynamics", in K. Hoover (ed), *Macroeconomics : Developments, Tensions and Prospects.*

[62] Diebold, F.X. and M. Nerlove (1989), "The Dynamics of Exchange Rate Volatility : A Multivariate Latent Factor ARCH Model", *Journal of Applied Econometrics* 4, 1-22.

[63] Ding, Z., C.W.J. Granger and R.F. Engle (1993), "A Long Memory Property of Stock Market Returns and a New Model", *Journal of Empirical Finance* 1, 83-108.

[64] Diz, F. and T.J. Finucane (1993), "Do the options markets really overreact?", *Journal of Futures Markets* 13, 298-312.

[65] Drost, F.C. and T. E. Nijman (1993), "Temporal Aggregation of GARCH Processes", *Econometrica* 61, 909-927.

[66] Drost, F.C. and B.J.M. Werker (1994), "Closing the GARCH Gap : Continuous Time GARCH Modelling", Discussion Paper CentER, University of Tilburg.

[67] Duan, J.C. (1994), "Maximum Likelihood Estimation Using Price Data of the Derivative Contract", *Mathematical Finance* 4, 155-167.

[68] Duan, J.C. (1995), "The GARCH Option Pricing Model", *Mathematical Finance* 5, 13-32.

[69] Duffie, D. (1989), *Futures Markets*, (Prentice-Hall International Editions).

[70] Duffie, D. (1992), *Dynamic Asset Pricing Theory*, (Princeton University Press).

[71] Duffie, D. and K. J. Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices", *Econometrica* 61, 929-952.

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

[73] Easley, D. and M. O'Hara (1992), "Time and the Process of Security Price Adjustment", *Journal of Finance*, 47, 577-605.

[74] Ederington, L.H. and J.H. Lee (1993), "How markets process information : news releases and volatility", *Journal of Finance* 48, 1161-1192.

[75] Elsheimer, B., M. Fisher, D. Nychka and D. Zirvos (1995), "Smoothing Splines Estimates of the Discount Function based on US Bond Prices", Discussion Paper Federal Reserve, Washington, D.C.

[76] Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica* 50, 987-1007.

[77] Engle, R.F. and C.W.J. Granger (1987), "Co-Integration and Error Correction : Representation, Estimation and Testing", *Econometrica* 55, 251-576.

[78] Engle, R.F. and S. Kozicki (1993), "Testing for Common Features", *Journal of Business and Economic Statistics* 11, 369-379.

[79] Engle, R.F. and G.G.J Lee (1994), "Estimating diffusion models of stochastic volatility", Discussion Paper, Univeristy of California at San Diego.

[80] Engle, R.F. and C. Mustafa (1992), "Implied ARCH models from option prices", *Journal of Econometrics* 52, 289-311.

[81] Engle, R.F. and V.K. Ng (1993), "Measuring and Testing the Impact of News on Volatility", *Journal of Finance* 48, 1749-1801.

[82] Fama, E.F. (1963), "Mandelbrot and The Stable Paretian Distribution", *Journal of Business* 36, 420-429.

[83] Fama, E.F. (1965), "The Behavior of Stock Market Prices", *Journal of Business* 38, 34-105.

[84] Foster, D. and S. Viswanathan (1993a), "The Effect of Public Information and Competition on Trading Volume and Price Volatility", *Review of Financial Studies* 6, 23-56.

[85] Foster, D. and S. Viswanathan (1993b), "Can Speculative Trading Explain the Volume Volatility Relation", Discussion paper, Fuqua School of Business, Duke University.

[86] French, K. and R. Roll (1986), "Stock Return Variances : The Arrival of Information and the Reaction of Traders", *Journal of Financial Economics* 17, 5-26.

[87] Gallant, A.R., D.A. Hsieh and G. Tauchen (1994), "Estimation of Stochastic Volatility Models with Suggestive Diagnostics", Discussion paper, Duke University.

[88] Gallant, A.R., P.E. Rossi and G. Tauchen (1992), "Stock Prices and Volume", *Review of Financial Studies*, 5, 199-242.

[89] Gallant, A.R., P.E. Rossi and G. Tauchen (1993), "Nonlinear Dynamic Structures", *Econometrica* 61, 871-907.

[90] Gallant, A.R. and G. Tauchen (1989), "Semiparametric Estimation of Conditionally Constrained Heterogeneous Processes : Asset Pricing applications ", *Econometrica* 57, 1091-1120.

[91] Gallant, A.R. and G. Tauchen (1992), "A Nonparametric Approach to Nonlinear Time Series Analysis: Estimation and Simulation", in E. Parzen, D. Brillinger, M. Rosenblatt, M. Taqqu, J. Geweke and P. Caines (eds.), *New Dimensions in Time Series Analysis*, Springer-Verlag, New York.

[92] Gallant, A.R. and G. Tauchen (1994), "Which Moments to Match", *Econometric Theory* (forthcoming).

[93] Gallant, A.R. and G. Tauchen (1995), "Estimation of Continuous Time Models for Stock Returns and Interest Rates", Discussion Paper, Duke University.

[94] Gallant, A.R. and H. White (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, (Basil Blackwell, Oxford).

[95] Garcia, R. and E. Renault (1995), "Risk Aversion, Intertemporal Substitution and Option Pricing", Discussion Paper CIRANO.

[96] Geweke, J. (1994), "Comment on Jacquier, Polson and Rossi", *Journal of Business and Economics Statistics* 12, 397-399.

[97] Geweke, J. (1995), "Monte Carlo Simulation and Numerical Integration", in H. Amman, D. Kendrick and J. Rust (ed.) *Handbook of Computational Economics* (North Holland).

[98] Ghysels, E., C. Gouriéroux and J. Jasiak (1995a), "Market Time and Asset Price Movements: Theory and Estimation", Discussion paper CIRANO and C.R.D.E., Univeristé de Montréal.

[99] Ghysels, E., C. Gouriéroux and J. Jasiak (1995b), "Trading Patterns, Time Deformation and Stochastic Volatility in Foreign Exchange Markets", Paper presented at the HFDF Conference, Zürich.

[100] Ghysels, E. and J. Jasiak (1994a), "Comments on Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics* 12, 399-401.

[101] Ghysels, E. and J. Jasiak (1994b), "Stochastic Volatility and Time Deformation An Application of Trading Volume and Leverage Effects", Paper presented at the Western Finance Association Meetings, Santa Fe.

[102] Ghysels, E., L. Khalaf and C. Vodounou (1994), "Simulation Band Inference in Moving Average Models", Discussion Paper, CIRANO and C.R.D.E.

[103] Ghysels, E., H.S. Lee and P. Siklos (1993), "On the (Mis)Specification of Seasonality and its Consequences : An Empirical Investigation with U.S. Data", *Empirical Economics* 18, 747-760.

[104] Goodhart, C.A.E. and M. O'Hara (1995), "High Frequency Data in Financial Markets : Issues and Applications", Paper presented at HFDF Conference , Zürich.

[105] Gouriéroux, C. and A. Monfort (1993a), "Simulation Based Inference : A Survey with Special Reference to Panel Data Models", *Journal of Econometrics* 59, 5-33.

[106] Gouriéroux, C. and A. Monfort (1993b), "Pseudo-Likelihood Methods" in Maddala et al. (ed.) *Handbook of Statistics* Vol. 11, (North Holland, Amsterdam).

[107] Gouriéroux, C. and A. Monfort (1994), "Indirect Inference for Stochastic Differential Equations", Discussion Paper CREST, Paris.

[108] Gouriéroux, C. and A. Monfort (1995), *Simulation-Based Econometric Methods*,(CORE Lecture Series, Louvain-la-Neuve).

[109] Gouriéroux, C., A. Monfort and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics* 8, S85-S118.

[110] Gouriéroux, C., A. Monfort and C. Tenreiro (1994), "Kernel M-Estimators: Nonparametric Diagnostics for Structural Models", Discussion Paper, CEPREMAP.

[111] Gouriéroux, C., A. Monfort and C. Tenreiro (1995), "Kernel M-Estimators and Functional Residual Plots", Discussion Paper CREST - ENSAE, Paris.

[112] Gouriéroux, C., E. Renault and N. Touzi (1994), "Calibration by Simulation for Small Sample Bias Correction", Discussion Paper CREST.

[113] Gouriéroux, C. and O. Scaillet (1994), "Estimation of the Term Structure from Bond Data", *Journal of Empirical Finance*, forthcoming.

[114] Granger, C.W.J. and Z. Ding (1994), "Stylized Facts on the Temporal and distributional Properties of Daily Data for Speculative Markets", Discussion Paper, University of California, San Diego.

[115] Hall, A.R. (1993), "Some Aspects of Generalized Method of Moments Estimation" in Maddala et al. (ed.) *Handbook of Statistics* Vol. 11, (North Holland, Amsterdam).

[116] Hamao, Y., R.W. Masulis and V.K. Ng (1990), " Correlations in Price Changes and Volatility Across International Stock Markets", *Review of Financial Studies* 3, 281-307.

[117] Hansen, B.E. (1995), "Regression with Nonstationary Volatility", *Econometrica* 63, 1113-1132.

[118] Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica* 50, 1029-1054.

[119] Hansen, L.P. and J.A. Scheinkman (1995), "Back to the Future : Generating Moment Implications for Continuous-Time Markov Processes", *Econometrica* 63, 767-804.

[120] Harris, L. (1986), "A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns", *Journal of Financial Economics* 16, 99-117.

[121] Harrison, M. and D. Kreps (1979), "Martingale and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory* 20, 381-408.

[122] Harrison, J.M. and S. Pliska (1981), "Martingales and Stochastic Ingegrals in the Theory of Continuous Trading", *Stochastic Processes and Their Applications* 11, 215-260.

[123] Harrison, P.J. and C.F. Stevens (1976), "Bayesian Forecasting" (with discussion), *J. Royal Statistical Society*, Ser. B, 38, 205-247.

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

[125] Harvey, A.C. and A. Jaeger (1993), "Detrending Stylized Facts and the Business Cycle", *Journal of Applied Econometrics* 8, 231-247.

[126] Harvey, A.C. (1993), "Long Memory in Stochastic Volatility", Discussion Paper, London School of Economics.

[127] Harvey, A.C. and S.J. Koopman (1993), "Forecasting Hourly Electricity Demand using Time-Varying Splines", *Journal of the American Statistical Association* 88, 1228-1236.

[128] Harvey, A.C., E. Ruiz and E. Sentana (1992), "Unobserved Component Time Series Models with ARCH Disturbances", *Journal of Econometrics* 52, 129-158.

[129] Harvey, A.C., E. Ruiz and N. Shephard (1994), "Multivariate Stochastic Variance Models", *Review of Economic Studies* 61, 247-264.

[130] Harvey, A.C. and N. Shephard (1993), "Estimation of Testing of Stochastic Variance Models", STICERD Econometrics, Discussion paper, EM93/268, London School of Economics.

[131] Harvey, A.C. and N. Shephard (1996), "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns", *Journal of Business and Economic Statistics* (forthcoming).

[132] Harvey, C.R. and R.D. Huang (1991), "Volatility in the Foreign Currency Futures Market", *Review of Financial Studies* 4, 543-569.

[133] Harvey, C.R. and R.D. Huang (1992), "Information Trading and fixed Income Volatility, Discussion Paper, Duke University.

[134] Harvey, C.R. and R.E. Whaley (1992), "Market volatility prediction and the efficiency of the S&P 100 index option market", *Journal of Financial Economics* 31, 43-74.

[135] Hausman, J.A. and A. W. Lo (1991), "An Ordered Probit Analysis of Transaction Stock Prices", Discussion paper, Wharton School, University of Pennsylvania.

[136] He, H. (1993), "Option Prices with Stochastic Volatilities : An Equilibrium Analysis", Discussion Paper, University of California, Berkeley.

[137] Heston, S.L. (1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options", *Review of Financial Studies* 6, 327-343.

[138] Heynen, R., A. Kemna and T. Vorst (1994), "Analysis of the term structure of implied volatility, *Journal of Financial Quantitative Analysis*.

[139] Hull, J. (1993), *Options, futures and other derivative securities* 2nd ed., (Prentice-Hall International Editions, New Jersey).

[140] Hull, J. (1995), *Introduction to Futures and Options Markets*, 2nd ed., (Prentice-Hall, Englewood Cliffs, New Jersey).

[141] Hull, J. and A. White (1987), "The Pricing of Options on Assets with Stochastic Volatilities", *Journal of Finance* 42, 281-300.

[142] Huffman, G.W. (1987), "A Dynamic Equilibrium Model of Asset Prices and Transactions Volume", *Journal of Political Economy* 95, 138-159.

[143] Hutchinson, J.M., A.W. Lo and T. Poggio (1994), "A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks", *Journal of Finance* 49, 851-890.

[144] Jacquier, E., N.G. Polson and P.E. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models" (with discussion), *Journal of Business and Economic Statistics* 12, 371-417.

[145] Jacquier, E., N.G. Polson and P.E. Rossi (1995a), "Multivariate and Prior Distributions for Stochastic Volatility Models", Discussion Paper CIRANO.

[146] Jacquier, E., N.G. Polson and P.E. Rossi (1995b),"Stochastic Volatility: Univariate and Multivariate Extensions", Rodney White Center for Financial Research Working Paper 19-95, The Wharton School, University of Pennsylvania.

[147] Jacquier, E., N.G. Polson and P.E. Rossi (1995c), "Efficient Option Pricing under Stochastic Volatility", Manuscript, The Wharton School, University of Pennsylvania.

[148] Jarrow, R. and Rudd (1983), *Option Pricing*, (Irwin, Homewood III).

[149] Johnson, H. and D. Shanno (1987), "Option Pricing when the Variance is Changing", *Journal of Financial and Quantitative Analysis* 22, 143-152.

[150] Jorion, P. (1995), "Predicting volatility in the foreign exchange market", *Journal of Finance* 50, (forthcoming).

[151] Karatzas, I. and S.E. Shreve (1988), *Brownian Motion and Stochastic Calculus*. Springer-Verlag: New York, NY.

[152] Karpoff, J. (1987), "The Relation between Price Changes and Trading Volume : A Survey", *Journal of Financial and Quantitative Analysis* 22, 109-126.

[153] Kim, S. and N. Shephard (1994), "Stochastic Volatility : Optimal Likelihood Inference and Comparison with ARCH Model", Discussion Paper, Nuffield College, Oxford.

[154] King, M., E. Sentana and S. Wadhwani (1994), "Volatility and Links Between National Stock Markets", *Econometrica* 62, 901-934.

[155] Kitagawa, G. (1987), "Non-Gaussian State Space Modeling of Nonstationary Time Series" (with discussion), *Journal of the American Statistical Association* 79, 378-389.

[156] Kloeden, P.E. and E. Platten (1992), *Numerical Solutions of Stochastic Differential Equations* (Springer-Verlag, Heidelberg).

[157] Lamoureux, C. and W. Lastrapes (1990), "Heteroskedasticity in Stock Return Data : Volume versus GARCH Effect", *Journal of Finance* 45, 221-229.

[158] Lamoureux, C. and W. Lastrapes (1993), "Forecasting stock-return variance : towards an understanding of stochastic implied volatilities", *Review of Financial Studies* 6, 293-326.

[159] Latane, H. and R. Jr. Rendleman (1976), "Standard Deviations of Stock Price Ratios Implied in Option Prices", *Journal of Finance* 31, 369-381.

[160] Lezan, G., E. Renault and T. deVitry (1995) "Forecasting Foreign Exchange Risk", Paper presented at 7th World Congres of the Econometric Society, Tokyo.

[161] Lin, W.L., R.F. Engle and T. Ito (1994), "Do Bulls and Bears Move Across Borders? International Transmission of Stock Returns and Volatility as the World Turns", *Review of Financial Studies*, forthcoming.

[162] Lo, A.W. (1995), "Statistical Inference for Technical Analysis Via Nonparametric Estimation", Discussion Paper, MIT.

[163] Mahieu, R. and P. Schotman (1994a), "Stochastic volatility and the distribution of exchange rate news", Discussion Paper, University of Limburg.

[164] Mahieu, R. and P. Schotman (1994b), "Neglected Common Factors in Exchange Rate Volatility". *Journal of Empirical Finance* 1, 279-311.

[165] Mandelbrot, B.B. (1963), "The Variation of Certain Speculative Prices", *Journal of Business* 36, 394-416.

[166] Mandelbrot, B. and H. Taylor (1967), "On the Distribution of Stock Prices Differences", *Operations Research* 15, 1057-1062.

[167] Mandelbrot, B.B. and J.W. Van Ness (1968), "Fractal Brownian Motions, Fractional Noises and Applications", *SIAM Review* 10, 422-437.

[168] McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration", *Econometrica* 57, 1027-1057.

[169] Meddahi, N. and E. Renault (1995), "Aggregations and Marginalisations of GARCH and Stochastic Volatility Models", Discussion Paper, GREMAQ.

[170] Melino, A. and M. Turnbull (1990), "Pricing Foreign Currency Options with Stochastic Volatility", *Journal of Econometrics* 45, 239-265.

[171] Melino, A. (1994), "Estimation of Continuous Time Models in Finance", in C.A. Sims (eds.) *Advances in Econometrics* (Cambridge University Press).

[172] Merton, R.C. (1973), "Rational Theory of Option Pricing", *Bell Journal of Economics and Management Science* 4, 141-183.

[173] Merton, R.C. (1976), "Option Pricing when Underlying Stock Returns are Discontinuous", *Journal of Financial Economics* 3, 125-144.

[174] Merton, R.C. (1990), *Continuous Time Finance*, (Basil Blackwell, Oxford).

[175] Merville, L.J. and D.R. Pieptea (1989), "Stock-price volatility, mean-reverting diffusion, and noise", *Journal of Financial Economics* 242, 193-214.

[176] Metropolis, N.,A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller (1954), "Equation of State Calculations by Fast Computing Machines", *The Journal of Chemical Physics* 21, 1087-1092.

[177] Müller, U.A., M.M. Dacorogna, R.B. Olsen, W.V.Pictet, M. Schwarz and C. Morgenegg (1990), "Statistical Study of Foreign Exchange Rates. Empirical Evidence of a Price change Scaling Law and Intraday analysis", *Journal of Banking and Finance* 14, 1189-1208.

[178] Nelson, D.B. (1988), "Time Series Behavior of Stock Market Volatility and Returns", Ph.D. dissertation, MIT.

[179] Nelson, D.B. (1990), "ARCH Models as Diffusion Approximations", *Journal of Econometrics* 45, 7-39.

[180] Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset returns : A New Approach", *Econometrica* 59, 347-370.

[181] Nelson, D.B. (1992), "Filtering and Forecasting with Misspecified ARCH Models I : Getting the Right Variance with the Wrong Model", *Journal of Econometrics* 25, 61-90.

[182] Nelson, D.B. (1994), "Comment on Jacquier, Polson and Rossi", *Journal of Business and Economic Statistics* 12, 403-406.

[183] Nelson, D.B. (1995a), "Asymptotic Smoothing Theory for ARCH Models", *Econometrica* (forthcoming).

[184] Nelson, D.B. (1995b), "Asymptotic Filtering Theory for Multivariate ARCH Models", *Journal of Econometrics* (forthcoming).

[185] Nelson, D.B. and D.P. Foster (1994), "Asymptotic Filtering Theory for Univariate ARCH Models", *Econometrica* 62, 1-41.

[186] Nelson, D.B. and D.P. Foster (1995), "Filtering and Forecasting with Misspecified ARCH Models II : Making the Right Forecast with the Wrong Model", *Journal of Econometrics* (forthcoming).

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

[188] Ogaki, M. (1993), "Generalized Method of Moments : Econometric Applications", in Maddala et al. (ed.) *Handbook of Statistics* Vol. 11, (North Holland, Amsterdam).

[189] Pagan, A.R. and G.W. Schwert (1990), "Alternative Models for Conditional Stock Volatility", *Journal of Econometrics* 45, 267-290.

[190] Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators", *Econometrica* 57, 995-1026.

[191] Pardoux, E. and D. Talay (1985), "Discretization and Simulation of Stochastic Differential Equations", *Acta Applicandae Mathematica* 3, 23-47.

[192] Pastorello, S., E. Renault and N. Touzi (1993), "Statistical Inference for Random Variance Option Pricing", Discussion Paper, CREST.

[193] Patell, J.M. and M.A. Wolfson (1981), "The Ex-Ante and Ex-Post Price Effects of quarterly Earnings Announcement Reflected in Option and Stock Price", *Journal of Accounting Research* 19, 434-458.

[194] Patell, J.M. and M.A. Wolfson (1979), "Anticipated Information Releases Reflected in Call Option Prices", *Journal of Accounting and Economics* 1, 117-140.

[195] Pham, H. and N. Touzi (1993), "Intertemporal Equilibrium Risk Premia in a Stochastic Volatility Model", *Mathematical Finance* (forthcoming).

[196] Platten, E. and Schweizer (1995), "On Smile and Skewness", Discussion Paper, Australian National University, Canberra.

[197] Poterba, J. and L. Summers (1986), "The persistence of volatility and stock market fluctuations", *American Economic Review* 76, 1142-1151.

[198] Renault, E. (1995), "Econometric Models of Option Pricing Errors", Invited Lecture presented at 7th W.C.E.S., Tokyo, August.

[199] Renault, E. and N. Touzi (1992), "Option Hedging and Implicit Volatility", *Mathematical Finance* (forthcoming).

[200] Revuz, A. and M. Yor (1991), *Continuous Martingales and Brownian Motion* (Springer Verlag, Berlin).

[201] Robinson, P. (1993), "Efficient tests of nonstationary hypotheses", mimeo, London School of Economics.

[202] Rogers, L.C.G. (1995), "Arbitrage with Fractional Brownian Motion", University of Bath, Discussion paper.

[203] Rubinstein, M. (1985), "Nonparametric Tests of Alternative Option Pricing Models Using all Reported Trades and Quotes on the 30 Most Active CBOE Option Classes from August 23, 1976 through August 31, 1978", *Journal of Finance* 40, 455-480.

[204] Ruiz, E. (1994), "Quasi-maximum Likelihood Estimation of Stochastic Volatility Models", *Journal of Econometrics* 63, 289-306.

[205] Schwert, G.W. (1989), "Business Cycles, Financial Crises, and Stock Volatility", *Carnegie-Rochester Conference Series on Public Policy* 39, 83-126.

[206] 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.

[207] Scott, L. (1991), "Random Variance Option Pricing", *Advances in Futures and Options Research*, Vol. 5, 113-135.

[208] Sheikh, A.M. (1993), "The behavior of volatility expectations and their effects on expected returns", *Journal of Business* 66, 93-116.

[209] Shephard, N. (1995), "Statistical Aspect of ARCH and Stochastic volatility", Discussion Paper 1994, Nuffield College, Oxford University.

[210] Sims, A. (1984), "Martingale-Like Behavior of Prices", University of Minnesota.

[211] Sowell, F. (1992), "Maximum likelihood estimation of stationary univariate fractionally integrated time series models", *Journal of Econometrics* 53, 165-188.

[212] Stein, J. (1989): "Overreactions in the Options Market", *The Journal of Finance* 44, 1011-1023.

[213] Stein, E.M. and J. Stein (1991), "Stock Price distributions with Stochastic Volatility : An Analytic Approach", *Review of Financial Studies* 4, 727-752.

[214] Stock, J.H. (1988), "Estimating Continuous Time Processes Subject to Time Deformation", *Journal of the American Statistical Association* 83, 77-84.

[215] Strook, D.W. and S.R.S. Varadhan (1979), *Multi-dimensional Diffusion Processes,* (Springer Verlag, Heidelberg).

[216] Tanner, T. and W. Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation", *Journal of the American Statistical Association* 82, 528-549.

[217] Tauchen, G. (1995), "New Minimum Chi-Square Methods in Empirical Finance", Invited Paper presented at the 7th World Congress of the Econometric Society, Tokyo.

[218] Tauchen, G. and M. Pitts (1983), "The Price Variability-Volume Relationship on Speculative Markets", *Econometrica* 51, 485-505.

[219] Taylor, S.J. (1986), *Modeling Financial Time Series*, (John Wiley : Chichester).

[220] Taylor, S.J. (1994), "Modeling Stochastic Volatility : A Review and Comparative Study", *Mathematical Finance* 4, 183-204.

[221] Taylor, S.J. and X. Xu (1994), "The Term structure of volatility implied by foreign exchange options", *Journal of Financial and Quantitative Analysis* 29, 57-74.

[222] Taylor, S.J. and X. Xu (1993), "The Magnitude of Implied Volatility Smiles : Theory and Empirical Evidence for Exchange Rates", Discussion Paper, University of Warwick.

[223] Von Furstenberg, G.M. and B. Nam Jeon (1989), "International Stock Price Movements: Links and Messages", *Brookings Papers on Economic Activity* I, 125-180.

[224] Wang, J. (1993), "A Model of Competitive Stock Trading Volume", Discussion Paper, MIT.

[225] Watanabe, T. (1993), "The Time Series Properties of Returns, Volatility and Trading Volume in Financial Markets", Ph.D. Thesis, Department of Economics, Yale University.

[226] West, M. and J. Harrison (1990), "*Bayesian Forecasting and Dynamic Models*", (Springer Verlag, Berlin).

[227] Whaley, R.E. (1982), "Valuation of American call options on dividend-paying stocks. *Journal of Financial Economics* 10, 29-58.

[228] Wiggins, J.B. (1987), "Option Values under Stochastic Volatility : Theory and Empirical Estimates", *Journal of Financial Economics* 19, 351-372.

[229] Wood, R.T. McInish and J.K. Ord (1985), "An Investigation of Transaction Data for NYSE Stocks", *Journal of Finance* 40, 723-739.

[230] Wooldridge, J.M. (1994), "Estimation and Inference for Dependent Processes" in R.F. Engle and D. McFadden (ed.) *Handbook of Econometrics* Vol. 4, (North Holland, Amsterdam).

Liste des publications au CIRANO

Cahiers CIRANO / *CIRANO Papers* (ISSN 1198-8169)

94c-1 Faire ou faire faire : La perspective de l'économie des organisations / par Michel Patry

94c-2 Commercial Bankruptcy and Financial Reorganization in Canada / par Jocelyn Martel

94c-3 L'importance relative des gouvernements : causes, conséquences, et organisations alternatives / par Claude Montmarquette

95c-1 La réglementation incitative / par Marcel Boyer

95c-2 Anomalies de marché et sélection des titres au Canada / par Richard Guay, Jean-François L'Her et Jean-Marc Suret

Série Scientifique / *Scientific Series* (ISSN 1198-8177)

95s-35 Capacity Commitment Versus Flexibility: The Technological Choice Nexus in a Strategic Context / Marcel Boyer et Michel Moreaux

95s-36 Some Results on the Markov Equilibria of a class of Homogeneous Differential Games / Ngo Van Long et Koji Shimomura

95s-37 Dynamic Incentive Contracts with Uncorrelated Private Information and History Dependent Outcomes / Gérard Gaudet, Pierre Lasserre et Ngo Van Long

95s-38 Costs and Benefits of Preventing Worplace Accidents: The Case of Participatory Ergonomics / Paul Lanoie et Sophie Tavenas

95s-39 On the Dynamic Specification of International Asset Pricing Models / Maral kichian, René Garcia et Eric Ghysels

95s-40 Vertical Integration, Foreclosure and Profits in the Presence of Double Marginalisation / Gérard Gaudet et Ngo Van Long

95s-41 Testing the Option Value Theory of Irreversible Investment / Tarek M. Harchaoui et Pierre Lasserre

95s-42 Trading Patterns, Time Deformation and Stochastic Volatility in Foreign Exchange Markets / Eric Ghysels, Christian Gouriéroux et Joanna Jasiak

95s-43 Empirical Martingale Simulation for Asset Prices / Jin-Chuan Duan et Jean-Guy Simonato

95s-44 Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter / Jin-Chuan Duan et Jean-Guy Simonato

95s-45 Costs and Benefits of Preventing Workplace Accidents : Going from a Mechanical to a Manual Handling System / Paul Lanoie et Louis Trottier

95s-46 Cohort Effects and Returns to Seniority in France / David N. Margolis

95s-47 Asset and Commodity Prices with Multiattribute Durable Goods / Jérôme Detemple et Christos I. Giannikos

95s-48 Is Workers' Compensation Disguised Unemployment Insurance? / Bernard Fortin, Paul Lanoie et Christine Laporte

95s-49 Stochastic Volatility / Eric Ghysels, Andrew Harvey et Eric Renault