

The Econometrics of Option Pricing*

René Garcia

Université de Montréal, CIRANO, CIREQ

Eric Ghysels

University of North Carolina and CIRANO

Éric Renault

Université de Montréal, CIRANO, CIREQ

First draft: November 2001

This version: August 1, 2003

Keywords: Stock Price Dynamics, Multivariate Jump-Diffusion Models, Latent variables, Stochastic Volatility, Objective and Risk Neutral Distributions, Nonparametric Option Pricing, Discrete-time Option Pricing Models, Risk Neutral Valuation, Preference-free Option Pricing.

JEL Classification: C1,C5,G1

* Address for correspondence: Département de Sciences Économiques, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, Québec, H3C 3J7, Canada. The first and the last authors gratefully acknowledge financial support from the Fonds de la Formation de Chercheurs et l'Aide à la Recherche du Québec (FCAR), the Social Sciences and Humanities Research Council of Canada (SSHRC), the Network of Centres of Excellence MITACS and the Institut de Finance Mathématique de Montréal (IFM2). The second author thanks CIRANO for financial support.

1 Introduction and overview

The growth of the option pricing literature parallels the spectacular developments of derivative securities and the rapid expansion of markets for derivatives in the last three decades. Writing a survey of option pricing models appears therefore like a formidable task. To delimit our focus we will put emphasis on the more recent contributions since there are already a number of surveys that cover the earlier literature. For example, Bates (1996b) wrote an excellent review, discussing many issues involved in testing option pricing models. Ghysels, Harvey and Renault (1996) and Shephard (1996) provide a detailed analysis of stochastic volatility modelling, while Renault (1997) explores the econometric modelling of option pricing errors. More recently, Sundaresan (2000) surveys the performance of continuous-time methods for option valuation. The material we cover obviously has many seminal contributions that pre-date the most recent work. Needless to say that due credit will be given to the seminal contributions related to the general topic of estimating and testing option pricing models. A last introductory word of caution: our survey deals almost exclusively with studies that have considered modelling the return process of a stock index and determining the price of European options written on this index.

One of the main advances that marked the econometrics of option pricing in the last five years has been the use of price data on both the underlying asset and options to jointly estimate the parameters of the process for the underlying and the risk premia associated with the various sources of risk. Even if important progress has been made regarding econometric procedures, the lesson that can be drawn from the numerous investigations, both parametric and nonparametric, in continuous time or in discrete time, is that the empirical performance still leaves much room for improvement. The empirical option pricing literature has revealed a considerable divergence between the risk-neutral distributions estimated from option prices after the 1987 crash and conditional distributions estimated from time series of returns on the underlying index. Three facts clearly stand out. First, the implied volatility extracted from at-the-money options differs substantially from the realized volatility over the lifetime of the option. Second, risk neutral distributions feature substantial negative skewness which is revealed by the asymmetric implied volatility curves when plotted against moneyness. Third, the shape of these volatility curves changes over time and maturities, in other words the skewness and the convexity are time-varying and maturity-dependent. Our survey will therefore explore possible explanations for the divergence between the objective and the risk neutral distributions. Modelling of the dynamics of the underlying asset price is an important part of the puzzle, while another essential element is the existence of time-varying risk premia. The last issue stresses the potentially

explicit role to be played by preferences in the pricing of options, a departure from the central tenet of the preference-free paradigm.

The main approach to modelling stock returns at the time prior surveys were written, was a continuous time stochastic volatility (henceforth SV) diffusion process possibly augmented with an independent jump process in returns. Heston (1993) proposed a SV diffusion model for which one could derive analytically an option pricing formula. Soon thereafter, see e.g. Duffie and Kan (1996), it was realized that Heston's model belonged to a rich class of affine jump diffusion processes for which one could obtain similar results. Duffie, Pan and Singleton (2000) discuss equity and fixed income derivatives pricing for the general class of affine jump diffusions. The evidence regarding the empirical fit of the affine class of processes is mixed, see e.g. Dai and Singleton (2000), Chernov, Gallant, Ghysels and Tauchen (2003), Ghysels and Ng (1998) for further discussion. There is a consensus that single volatility factor models, affine (like Heston (1993)) or non-affine (like Hull and White (1987) or Wiggins (1987)), do not fit the data (see Andersen, Benzoni and Lund (2002), Benzoni (1998), Chernov, Gallant, Ghysels and Tauchen (2003), Pan (2002), among others). How to expand single volatility factor diffusions to mimic the data generating process remains unsettled. Several authors augmented affine SV diffusions with jumps, see Andersen, Benzoni and Lund (2001), Bates (1996a), Chernov, Gallant, Ghysels and Tauchen (2003), Eraker, Johannes and Polson (2001), Pan (2002), among others. Bakshi, Cao and Chen (1997), Bates (2000) Chernov, Gallant, Ghysels and Tauchen (2003) and Pan (2002) show, however, that SV models with jumps in returns are not able to capture all the empirical features of observed option prices and returns. Bates (2000) and Pan (2002) argue that the specification of the volatility process should include jumps, possibly correlated with the jumps in returns. Chernov, Gallant, Ghysels and Tauchen (2003) maintain that a two-factor non-affine logarithmic SV diffusion model without jumps yields a superior empirical fit compared to affine one-factor or two factor SV processes, or SV diffusions with jumps. Alternative models were also proposed in recent years: they include volatility models of the Ornstein-Uhlenbeck type but with Lévy innovations (Barndorff-Nielsen and Shephard, 2001) and stochastic volatility models with long memory in volatility (Breidt, Crato and de Lima (1998)) and Comte and Renault (1998)).

The statistical fit of the underlying process and the econometric complexities associated with it should not be the only concern, however. An important issue for option pricing is whether or not the models deliver closed-form solutions. We will therefore discuss if and when there exists a trade-off between obtaining a good empirical fit or a closed-form option pricing formula. The dynamics of the underlying fundamental asset cannot be related to option prices without additional assumptions or information. One possibility is to assume

that the risks associated with stochastic volatility or jumps are idiosyncratic and not priced by the market. There is a long tradition of this, but most recent empirical work clearly indicates there are prices for volatility and jump risk (see e.g. Andersen, Benzoni and Lund (2002), Chernov and Ghysels (2000), Pan (2002), among others). One can simply set values for these premia and use the objective parameters to derive implications for option prices as in Andersen, Benzoni and Lund (2001). A more informative exercise is to use option prices to calibrate the parameters under the risk neutral process given some version of a nonlinear least-squares procedure as in Bakshi, Cao and Chen (1997) and Bates (2000). An even more ambitious program is to use both the time series data on stock returns and the panel data on option prices to characterize the dynamics of returns with stochastic volatility and with or without jumps as in Chernov and Ghysels (2000), Pan (2002), Poteshman (2000) and Garcia, Lewis and Renault (2001).

Whether one estimates the objective probability distribution, the risk neutral or both, there are many challenges in estimating the parameters of diffusions. The presence of latent volatility factors make maximum likelihood estimation computationally infeasible. This is the area where probably the most progress has been made in the last few years. Several methods have been designed for the estimation of continuous time dynamic state-variable models with the pricing of options as a major application. Simulation-based methods have been most successful in terms of empirical implementations. That includes the indirect inference and efficient methods of moments of Gouriéroux, Monfort and Renault (1993) and Gallant and Tauchen (1996) respectively, and several procedures discussed by Johannes and Polson (2002) as well as Aït-Sahalia, Hansen and Scheinkman (2002) in this Handbook. Another approach is to use implied state methods. While Pastorello, Patilea and Renault (2003) base an indirect inference approach on Black-Scholes implied volatilities, Pan (2002) uses the Fourier transform to derive a set of moment conditions pertaining to implied states. Renault and Touzi (1996), Patilea and Renault (1997) and Renault (1997) propose iterative and recursive procedures which extend the EM (expectation-maximization) methodology to maximum likelihood contexts where it usually does not apply. Pastorello, Patilea and Renault (2003) propose a general methodology of iterative and recursive estimation in structural non-adaptive models which nests all the previous implied state approaches.

Nonparametric methods have also been used extensively. Several studies aimed at recovering the risk-neutral probabilities or state-price densities implicit in option or stock prices. For instance, Rubinstein (1996) proposed an implied binomial tree methodology to recover risk-neutral probabilities which are consistent with a cross-section of option prices. Aït-Sahalia and Lo (1998) use a kernel estimator of the volatility function in a Black-Scholes type model. Stutzer (1996) uses an approach called canonical valuation which

uses past return data and possibly but not necessarily option price data to estimate the payoff distribution at expiration. Another approach consists in estimating directly the option pricing function with nonparametric methods. Hutchinson, Lo and Poggio (1994), Broadie, Detemple, Ghysels and Torrès (2000a,b), and Garcia and Gençay (2000) follow this route. An important issue with the model-free nonparametric approaches is that the recovered risk-neutral probabilities are not always positive and one may consider adding constraints on the pricing function or the state-price densities. For example, Ait-Sahalia and Duarte (2003) impose monotonicity and convexity restrictions using a nonparametric method based on locally polynomial estimators.

Bates (2000), among others, shows that risk-neutral distributions recovered from option prices before and after the crash of 1987 are fundamentally different whereas the objective distributions do not show such structural changes. Before the crash, both the risk neutral and the actual distributions look roughly lognormal. After the crash, the risk-neutral distribution is left skewed and leptokurtic. A possible explanation for the difference is a large change in the risk aversion of the average investor. Since risk aversion can be recovered empirically from the risk neutral and the actual distributions, Ait-Sahalia and Lo (2000), Jackwerth (2000) and Rosenberg and Engle (2002) estimate preferences for the representative investor using simultaneously S&P500 returns and options prices for contracts on the index. Preferences are recovered based on distance criteria between the model risk neutral distribution and the risk neutral distribution implied by option price data.

Another approach of recovering preferences is to set up a representative agent model and estimate the preference parameters from the first-order conditions using a GMM approach. While this has been extensively done with stock and Treasury bill return data (see Hansen and Singleton (1982), Epstein and Zin (1991) among others), it is only recently that Garcia, Luger and Renault (2003) estimated preference parameters in a recursive utility framework using option prices. In this survey we will discuss under which statistical framework option pricing formulas are preference-free and risk-neutral valuation relationships (Brennan, 1979) hold in a general stochastic discount factor framework (Hansen and Richard (1987)). When these statistical restrictions do not hold, it will be shown that preferences play a role. Bates (2001) argues that the overall industrial organization of the stock index option markets is not compatible with the idealized construct of a representative agent. He therefore proposes an equilibrium analysis with investor heterogeneity.

Apart from statistical model fitting, there are a host of other issues pertaining to the implementation of models in practice. A recent survey by Bates (2003) provides an overview of the issues involved in empirical option pricing, especially the questions surrounding data

selection, estimation or calibration of the model and presentation of results.

The price of a derivative security is determined by the risk factors affecting the dynamic process of the underlying asset. We start the survey with discrete time models based on the key notion of stochastic discount factor. The analysis in section 2 allows us to discuss many issues, both theoretical and empirical in a relatively simple and transparent setting. Sections 3 and 4 deal with continuous time processes. Section 3 is devoted to the subject of modelling the so-called objective probability measure, and section 4 discusses how to recover risk neutral probability densities in a parametric continuous time setting. Nonparametric approaches to pricing, hedging and recovering state price densities are reviewed in section 5.

2 Option pricing, market completeness and preferences

Since the seminal work of Black and Scholes (1973) it is customary to price options via perfect replication in complete markets and compute derivative prices in a hypothetical market in which agents have risk neutral preferences. Option pricing strategies are therefore based on the derivation of risk neutral valuation relationships (henceforth called RNVR) to adopt the terminology introduced by Brennan (1979). Since the work of Black and Scholes (1973) and Merton (1973) it is often argued that a continuous time setting is particularly suited for option pricing.¹ In the typical Black-Scholes and Merton continuous time setting of diffusions the RNVR principle consists, by virtue of the Girsanov theorem, in changing only the drift term of the underlying asset to the risk-free rate of return and keeping the diffusion term intact.² In the general diffusion case a number of sources of incompleteness are introduced via unhedgeable risks (like stochastic volatility and jump risk). In such cases the basic RNVR and Girsanov ideas remain, yet some well-suited risk premia associated with the various sources of unhedgeable risk are introduced. Consequently, the general diffusion setting naturally suggests that option pricing formulas only depend on potentially observable parameters and can be derived only from some rather weak assumptions about investor preferences. In particular, it is commonly assumed that the risk premium of the

¹See for instance recent discussions regarding continuous time diffusion models and option pricing by Bertsimas, Kogan and Lo (2000) who quantify the approximation errors of discrete time by characterizing the asymptotic distribution of the replication errors that arise from continuous time delta-hedging derivative securities in discrete time. Along different lines, Aït-Sahalia (2001) examines similar issues for a class of Markov diffusions while Andreou and Ghysels (2001) consider a general class of continuous time Lévy processes.

²See for example Duffie (2001), Merton (1992), among others, for a discussion of derivative security pricing in continuous time.

underlying asset (as characterized by the net expected return) does not play a direct role in determining the value of derivative contracts. Generally speaking, the link between market completeness and RNVR is often overstated, as the validity of the RNVR for diffusion models has much more to do with specific distributional features of diffusion models than with market completeness. This has been stressed recently by Duan (2001) who put forward a semi-recombined binomial lattice to show that market completeness does not imply RNVR.

Despite the perceived convenience of continuous time models we will use the discrete time approach to characterize all the relevant situations for option pricing: cases where RNVR prevails, violations of RNVR which do not preclude preference-free option pricing and finally, cases where an estimation of preference parameters, and possibly also of investors beliefs, is explicitly needed for option valuation.³

Following the modern approach in financial econometrics, as in the recent book by Cochrane (2001), we will rely on the unifying framework provided by the stochastic discount factor (SDF) paradigm put forward by Hansen and Richard (1987) to distinguish all these cases. While the common continuous time setting certainly simplifies a number of issues, it will be made clear that the SDF framework helps us understand some fundamental problems about option pricing regardless of the continuous time or discrete time settings. In particular, we will use discrete time modelling to argue that completeness is far to be necessary for RNVR.

This section is organized as follows. We will first summarize the SDF paradigm before considering the different situations of interest for option pricing. We start from the popular preference-free binomial option pricing model and show how implied binomial trees can be understood through the SDF approach. Second, we will discuss a quite large set of distributional assumptions (including GARCH option pricing) which guarantee RNVR. This characterization will motivate the typical violations of RNVR put forward in this chapter. These violations will basically involve the existence of a number of state variables, which are relevant for option valuation but are not instantaneously observed, neither by the econometrician nor by the investors. This will lead us to discuss the issues of leverage effects, investors' beliefs and learning in general multi-period discrete time models. To conclude the section we discuss how to nest common continuous time option pricing formulas in the general framework proposed here.

³See for instance Pliska (1997) for a comprehensive treatment of discrete time asset pricing models, including option pricing.

2.1 The SDF paradigm over one period

When asset markets are frictionless, Harrison and Kreps (1979) and Chamberlain and Rothschild (1983) show that portfolio prices can be characterized as a continuous linear valuation functional that assigns prices to the portfolio payoffs. Linearity and continuity correspond to the common “law of one price” assumption, that is a weak version of a no arbitrage condition involving limit series of payoffs. Hansen and Richard (1987) revisit this property in the presence of conditioning information. They work with a Hilbert space structure on the vectorial space G_{t+1} of possible payoffs at time $(t + 1)$. The space is characterized by the conditional scalar product, that is the L^2 structure corresponding to conditional expectations of payoffs at time $(t + 1)$ given the information J_t available to investors at time t . Their fundamental theorem shows that the price at time t of any payoff $g_{t+1} \in G_{t+1}$, denoted $\pi_t(g_{t+1})$ can be written as:

$$\pi_t[g_{t+1}] = E[(m_{t+1})(g_{t+1})|J_t] = E_t[(m_{t+1})(g_{t+1})]. \quad (2.1)$$

The variable m_{t+1} is by definition SDF admissible for the set G_{t+1} of payoffs. Note that only one among the admissible SDFs is an element of the space G_{t+1} of payoffs. It is the orthogonal projection m_{t+1}^* on G_{t+1} of any admissible SDF m_{t+1} . If we consider the set $G_{1+1}(g)$ of payoffs contingent to a given primitive payoff g_{t+1} , that is the set of all squared integrable functions $h [J_t, g_{t+1}]$ of this payoff, we have:

$$m_{t+1}^* = E_t[m_{t+1}|g_{t+1}] = E_t[m_{t+1}|J_t, g_{t+1}]. \quad (2.2)$$

Hansen and Richard (1987) show that it is sufficient to check that a particular admissible SDF is almost surely positive to be sure that the pricing functional $\pi_t(\cdot)$ has no arbitrage opportunities on G_{t+1} . Conversely, no arbitrage implies that the SDF m_{t+1}^* in (2.2) is positive almost surely. Therefore, without relying on any assumption of completeness, absence of arbitrage in the set of contingent claims corresponds to the existence of a unique positive SDF which is a function of the primitive payoffs. Note that these payoffs may have been generated by a stochastic volatility process with jumps, or any of other processes considered later in this chapter. Under market incompleteness we can define a risk neutral probability measure, as in Harrison and Kreps (1979), to compute the price of any contingent claim of payoff $h_{t+1} = h(J_t, g_{t+1})$ as: $\pi_t[h_{t+1}] = E_t[(m_{t+1}^*)(h_{t+1})] = E_t^*[B(t, t + 1)h_{t+1}]$, where $B(t, t + 1) = E_t[m_{t+1}^*]$ is the price at date t of a zero-coupon bond which delivers one dollar at date $(t + 1)$. The (conditional) risk neutral probability measure Q^* (which in turn defines the conditional expectation operator E_t^*) is defined from the historical one Q by the positive density function $(m_{t+1}^*)/B(t, t + 1)$.

2.2 Empirical pricing kernels

We discuss in this subsection some empirical strategies which have been proposed to calibrate the risk neutral probability measure from option prices data without relying on specific measures of the pricing kernel like some parametric functions of the aggregate consumption process or the market return. This general principle can be illustrated in the simplest case of a binomial model. If it is known at time t that the payoff g_{t+1} can take only two values $(u_t.g_t)$ and $(d_t.g_t)$, an application of the SDF formula for any contingent claim $h_{t+1} = a_t + b_t.g_{t+1}$ gives:

$$\pi_t[h_{t+1}] = E_t[(m_{t+1})(h_{t+1})] = a_t B(t, t+1) + b_t \pi_t(g_{t+1}).$$

where a_t and b_t are defined as solutions of : $h(J_t, u_t.g_t) = a_t + b_t.u_t.g_t$ and $h(J_t, d_t.g_t) = a_t + b_t.d_t.g_t$. Therefore, if one defines a ‘‘probability’’ p_t^* as the solution of:

$$\pi_t[g_{t+1}] = B(t, t+1)[p_t^*.u_t.g_t + (1 - p_t^*)d_t.g_t] \quad (2.3)$$

and uses this in the expression of $\pi_t[h_{t+1}]$, one obtains:

$$\pi_t[h_{t+1}] = B(t, t+1)[p_t^*h(J_t, u_t.g_t) + (1 - p_t^*)h(J_t, d_t.g_t)]. \quad (2.4)$$

Hence, the SDF approach provides a simple proof of classical option pricing formulas like the binomial option pricing formula of Cox, Ross and Rubinstein (1979). The risk neutral probabilities p_t^* and $(1 - p_t^*)$ of the two states of nature $(u_t.g_t)$ and $(d_t.g_t)$ can be easily computed from (2.3), that is from the observation of the underlying asset price, and then provide an empirical pricing kernel to value any derivative asset according to (2.4).

Typically, the relevant set of states contains more than two elements and the above methodology has to be generalized for the purpose of empirical pricing. The general framework of SDF pricing allows for any generalization within a one-period setting. The implied binomial tree method of Rubinstein(1994) involves possible outcomes $K_{it} = \alpha_{it}g_t, i = 0, 1, \dots, I + 1$ at time t for the underlying asset at time $(t + 1)$. They are obtained from the observed strike prices $K_{it}, i = 1, \dots, I (K_{1t} < K_{2t} < \dots < K_{It})$ of options quoted at time t with maturity $(t + 1)$ completed by $K_{I+1,t} > K_{It}$ and $K_{0I} < K_{1t}$.

The observation at time t of market prices $B(t, t + 1)$ of the bond, $\pi_t[g_{t+1}]$ of the underlying asset, and $\pi_t[Max(0, g_{t+1} - K_{it})]$ of European calls written on this asset with strike prices $K_{it}, i = 1, \dots, I$ is equivalent to observing the prices $\pi_t[g_{i,t+1}^*]$ of the $(I + 2)$ digital options whose payoffs $g_{i,t+1}^*$ are respectively $K_{it} = \alpha_{it}g_t, i = 0, 1, \dots, I + 1$ when $g_{t+1} = \alpha_{it}g_t$, and zero otherwise. The two sets of $(I + 2)$ payoffs are linked by a one-to-one linear relationship and, by the law of one price, their prices should fulfill the same

relationship. For the sake of notational simplicity, we will then consider, without loss of generality, that the prices of these digital options are observed instead of the prices of standard European calls. The argument developed above in the genuine binomial model ($I = 0$) can then be generalized to any contingent claim $h_{t+1} = a_t + \sum_{i=1}^{I+1} b_{it} \cdot g_{it}^*$. The pricing formula becomes:

$$\pi_t[h_{t+1}] = a_t B(t, t+1) + \sum_{i=1}^{I+1} b_{it} \pi_t[g_{it}^*] = B(t, t+1) E_t^*[h_{t+1}]$$

where the expectation operator $E_t^*[\cdot]$ is computed with the risk neutral probabilities $(p_{it}^*)_{0 \leq i \leq I+1}$, defined by: $\pi_t[g_{it}^*] = B(t, t+1)[p_{it}^* \cdot \alpha_{it} \cdot g_t + (1-p_{it}^*)0]$, $i = 1, 2, \dots, I+1$ and $p_{0t}^* = 1 - \sum_{i=1}^{I+1} p_{it}^*$. Hence, the observation of I option prices written on the same asset with the same maturity date, in addition to the observation of the bond price for the same maturity and the underlying asset price, allows us to recover a set of implied risk neutral probabilities $(p_{it}^*)_{0 \leq i \leq I+1}$ and in turn an empirical pricing kernel yielding the price $\pi_t[h_{t+1}]$ of any claim contingent on the same asset with the same maturity by :

$$\pi_t[h_{t+1}] = B(t, t+1) E_t^*[h_{t+1}] \tag{2.5}$$

The implied probabilities have been constructed under the restrictive assumption that the support of the risk neutral probability distribution is limited to the set of observed strikes of options quoted at time t with maturity $(t+1)$, augmented by the two extreme values K_{0t} and $K_{I+1,t}$ to rationalize respectively the positive price of the deepest in- and out-of-the money options. However, Rubinstein (1994) argues that the validity of the pricing kernel (2.5) is slightly more general, since possible values of the underlying asset payoff g_{t+1} may include other points than the values $\alpha_{it} g_t$, $i = 0, 1, \dots, I+1$ provided the corresponding risk neutral probabilities of any state between $K_{i,t}$ and $K_{i+1,t}$ are equal. This argument raises several issues that we will discuss in more detail in section 5 dedicated to nonparametric approaches.

2.3 Loglinear pricing

Along with the binomial model, loglinear pricing represents a central tenet of option pricing in discrete time. It is intimately related to the RNVR concept introduced by Brennan (1979). A RNVR consists of a pricing formula of the type:

$$\pi_t[h_{t+1}] = B(t, t+1) E_t[h(J_t, S_{t+1}^*)] \tag{2.6}$$

where S_{t+1}^* is a rescaled value of the underlying asset price defined by: $B(t, t+1) S_{t+1}^* = [S_{t+1}/E_t S_{t+1}] S_t$. In other words, S_{t+1}^* is proportional (given J_t) to S_{t+1} but the mean of its

distribution is $S_t/B(t, t+1)$. The other parameters of the conditional (given J_t) probability distribution of S_{t+1} are identical to the objective ones.

Therefore, like the empirical pricing kernel of subsection 2.2, the RNVR provides a specification of the option pricing formula which no longer depends explicitly on preference parameters. These preference parameters are hidden, first in the price $B(t, t+1)$ of the pure discount bond, and second in the underlying asset through the risk premium $B(t, t+1)E_t S_{t+1}/S_t$. However, the two formulas (2.5) and (2.6) are significantly different and it is worthwhile to examine under what circumstances the RNVR might produce a well-founded alternative preference-free option pricing principle.

By focusing on the SDF provided by a time-separable utility function of a representative agent $m_{t+1} = \beta u'(C_{t+1})/u'(C_t)$, with the additional assumption that $[\log(C_{t+1}), \log(S_{t+1})]$ is jointly normal conditional on J_t , Brennan (1979) shows that a necessary and sufficient condition for RNVR is that the marginal utility function u' is a power function. However, one should note that the role of the power utility function is to obtain joint conditional normality of $[\log(u'(C_{t+1})), \log(S_{t+1})]$ from that of $[\log(C_{t+1}), \log(S_{t+1})]$. In other words, the crucial assumption to obtain RNVR in this one-period setting is:

Assumption 2.1: (Conditional Log Normality) $[\log(m_{t+1}), \log(S_{t+1}/S_t)]$ is conditionally normal given J_t .

The option pricing literature contains numerous well-known examples of models involving Assumption 2.1. When one writes explicitly the SDF corresponding to a one-period Black-Scholes option pricing model, one realizes (see e.g. Buraschi and Jackwerth (2001)), that $\log(m_{t+1})$ is an affine function of the normal log-return $\log(S_{t+1}/S_t)$. Therefore, Assumption 2.1 is fulfilled in the Black-Scholes case with a degenerate joint normal distribution. Moreover, as noted by Heston and Nandi (2000), the GARCH option pricing model of Duan (1995) assumes that the value of a call option one period prior to expiration obeys the Black-Scholes-Rubinstein formula. Hence, GARCH option pricing is also based on a degenerate conditional normal probability distribution of $[\log(m_{t+1}), \log(S_{t+1}/S_t)]$ given J_t . The only difference with the Black-Scholes geometric random walk is that the conditional variance of $\log(S_{t+1}/S_t)$ given J_t will now depend explicitly on J_t .⁴

We can conclude that all the aforementioned option pricing models are indeed simple applications of RNVR, since Assumption 2.1 implies RNVR. To prove this statement, it suffices to extend the proof in Brennan (1979), using the SDF formulation and Girsanov

⁴Garcia and Renault (1998a) and Kallsen and Taquq (1998) discuss in further detail the maintained assumptions of GARCH option pricing both in terms of equilibrium and arbitrage.

theorem (which makes clear the analogy with option pricing in the context of diffusion models; see Renault (2001)). It is based on the following formula implied by the joint normality assumption of $[\log(m_{t+1}), \log(S_{t+1})]$ given J_t . For $h_{t+1} = h[J_t, S_{t+1}] = h_t(S_{t+1})$:

$$\pi_t[h_{t+1}] = E_t[(m_{t+1})(h_{t+1})] = E_t(m_{t+1}) E_t\{h_t[S_{t+1} \exp[Cov_t(\log(m_{t+1}), \log(S_{t+1}))]]\}. \quad (2.7)$$

Using the bond pricing equation and the underlying asset pricing equation, one remarks that: $\exp[Cov_t(\log(m_{t+1}), \log(S_{t+1}))] = [B(t, t+1)E_t S_{t+1}]^{-1} S_t$, which is the scale factor put forward in (2.6). This confirms that option pricing formulas are preference-free since the relevant characteristics of the SDF are hidden in the bond price and in the underlying asset price.

For example, with $h_{t+1} = h(S_{t+1}) = \max[0, S_{t+1} - K]$, the payoff of an European call, we get:

$$\pi_t(h_{t+1}) = B(t, t+1) E_t \max[0, S_{t+1}^* - K]$$

where:

$$\log\left(\frac{S_{t+1}^*}{S_t}\right) = \left(\frac{S_{t+1}}{S_t}\right) - \log\frac{E_t(S_{t+1})B(t, t+1)}{S_t}$$

In a Black-Scholes (BS hereafter) world, $\log(S_{t+1}/S_t)$ follows a normal distribution with mean $\log(E_t(S_{t+1})/S_t) - \sigma^2/2$ and variance σ^2 . Then $\log(S_{t+1}^*/S_t)$ follows a normal distribution with mean $-\log B(t, t+1) - \sigma^2/2$ and variance σ^2 . Straightforward computations of expectations with normal distributions give the BS option pricing formula:

$$\pi_t(h_{t+1}) = BS[S_t, \sigma^2]$$

where:

$$BS[S_t, \sigma^2] = S_t \phi(d_1) - KB(t, t+1) \phi(d_2)$$

with: $d_1 = (1/\sigma) \log(S_t / KB(t, t+1)) + \sigma/2$ and $d_2 = d_1 - \sigma$.

Generally speaking, preference-free option pricing can be obtained within two alternative settings: either a linear factor model with an affine regression of m_{t+1} on S_{t+1} for the conditional expectation $E_t[m_{t+1}|S_{t+1}]$ given either a normality or a multinomial assumption, or a log-linear factor model with an affine regression of $\text{Log}(m_{t+1})$ on $\text{Log}(S_{t+1})$ for the conditional expectation $E_t[\log m_{t+1}|S_{t+1}]$. The basic intuition is that, without such a linearity property of the conditional expectation, the price of contingent claims with payoffs that are nonlinear functions of the underlying asset payoff, cannot be straightforwardly

deduced from the price of the underlying asset. Even though it is worth stressing that the aforementioned (log)normality properties are only conditional and therefore do not preclude conditional heteroskedasticity and unconditional leptokurticity of asset returns (as in the GARCH option pricing model), it is still empirically relevant to think about a way to relax these normality assumptions. Since the seminal paper of Clark (1973), a versatile tool to relax normality for asset (log) returns is to introduce mixture components.

2.4 Mixture models of option pricing

In this subsection we will further explore the log-linear model but we will add a mixture component to the conditioning set and maintain the following assumption:

Assumption 2.2: (Conditional Log Normality with mixture) There exists a latent state variable U_{t+1} such that: $[\log(m_{t+1}), \log(S_{t+1}/S_t)]$ is conditionally normal given J_t and U_{t+1} .

Hence, the conditional probability distribution of $[\log(m_{t+1}), \log(S_{t+1}/S_t)]$ given J_t is a mixture of normals with a mixture component U_{t+1} . The mixture model is quite standard and contains as particular cases the widely used stochastic volatility model first proposed by Taylor (1986) and first used for option pricing by Hull and White (1987) and extended by Heston (1993).⁵ However, in contrast with some standard option pricing models under stochastic volatility, we have to assume here, in order to relax RNVR pricing, that the latent variable U_{t+1} is unobserved not only by the econometrician but also by investors. An intuitive rationalization of this assumption is the genuine uncertainty surrounding the actual state of the economy or the business cycle at some points in time.

In this context, Girsanov's theorem can only be applied with the conditional probability distribution given U_{t+1} and the pricing formula is now given by:

$$\begin{aligned} \pi_t[h_{t+1}] &= E_t \{ E_t[m_{t+1}h_t(S_{t+1})|U_{t+1}] \} \\ &= E_t[m_{t+1}|U_{t+1}] E_t \{ h_t[S_{t+1} \exp[Cov_t[\log(m_{t+1}), \log(S_{t+1}/S_t)|U_{t+1}]]|U_{t+1}] \} \end{aligned} \quad (2.8)$$

By using both the bond and stock pricing equations, we now arrive at a pricing formula that generalizes the RNVR formula (2.6):

$$\pi_t[h_{t+1}] = E_t[B^*(t, t+1)] E_t \{ h_t[S_{t+1}^* \cdot \xi_{t,t+1}] | U_{t+1} \} \quad (2.9)$$

⁵Note that the class of Lévy-Stable distributions is also consistent with Assumption 2.2 when the mixture component is the variance of the normal viewed as a positive Lévy-stable distribution. Generally speaking, the mixture component is a way to capture observed fat tails in the conditional probability distribution of log-returns given J_t . For further discussion see next section.

where S_{t+1}^* is a rescaled value of the underlying asset price now defined by:

$$B^*(t, t+1)S_{t+1}^* = \frac{S_{t+1}}{E_t S_{t+1}} \cdot S_t. \quad (2.10)$$

with $B^*(t, t+1) = E_t[m_{t+1}|U_{t+1}]$ and $\xi_{t,t+1} = E_t[m_{t+1}(S_{t+1}/S_t)|U_{t+1}]$.

Note that $B^*(t, t+1)$ and $S_t \xi_{t,t+1}$ can be interpreted respectively as the bond and stock price in the fictitious world where the mixing variable U_{t+1} would be known at time t by the investors. By the law of iterated expectations, the expected value at time t of these fictitious prices coincide with the actual prices: $B(t, t+1) = E_t[B^*(t, t+1)]$ and $E_t[\xi_{t,t+1}] = 1$. Therefore the pricing formulas (2.9) and (2.10) generalize RNVR in two dimensions, in the sense that the observed bond and stock prices are both replaced by functions of the state variables. These functions will coincide with observed bond and stock prices if and only if there is no mixing effect in the joint probability distribution of the SDF and the stock price. Otherwise, the subordination of this distribution and in particular the characteristics of the SDF (conditional mean, variance and covariance with the asset return) will enter explicitly in the option pricing formula; the option pricing formula is no longer preference-free.⁶ This mixture model appears to be sufficiently versatile to capture some stylized facts about the volatility smile at a given point in time.

2.5 Volatility smiles, smirks and frowns in a mixture model

Following Garcia, Luger and Renault (2001, 2003) we apply in this subsection the mixture model to the pricing of a European call with payoff: $h(S_{t+1}) = \text{Max}[0, S_{t+1} - K]$. For simplicity, we will assume that there is no interest rate risk: $B^*(t, t+1) = B(t, t+1)$. Therefore, following (2.9), the price of a European call is given by:

$$\pi_t^h = E_t[E_t[B(t, t+1)\text{Max}[0, S_{t+1}^* \xi_{t,t+1} - K]|U_{t+1}]]$$

Given U_{t+1} , we are in a Black-Scholes risk neutral world. Therefore, we obtain the following Generalized Black-Scholes (GBS) option pricing formula:

$$\pi_t^h = E_t[BS[S_t \xi_{t,t+1}, \sigma^2(U_{t+1})]] \quad (2.11)$$

where, inside the standard BS option pricing formula, the current value of the stock price S_t has been replaced by $S_t \xi_{t,t+1}$ and the constant volatility parameter σ^2 has been replaced

⁶To make the link with linear factor pricing, note that the mixture effect in the probability distribution of the underlying asset return (including its “beta” coefficient with respect to the SDF) is also responsible of the violation of a CAPM-like pricing formula for the underlying asset. In particular, RNVR implying $\xi_{t,t+1} = 1$, means that: $E_t[S_{t+1}/S_t] = [B(t, t+1)]^{-1} \exp[\text{Cov}_t[\log(m_{t+1}), \log(S_{t+1}/S_t)]$.

by the stochastic volatility $\sigma^2(U_{t+1}) = Var_t[\log(S_{t+1}/S_t)|U_{t+1}]$. The expectation operator in (2.11) is with respect to the joint probability distribution of $\xi_{t,t+1}$ and $\sigma^2(U_{t+1})$ given the information J_t .

Even though the GBS formula (2.11) is derived in a discrete time context, a similar pricing formula was derived by Romano and Touzi (1997) in the context of risk neutral continuous time models of stochastic volatility with leverage, as in Heston (1993), and by Fouque, Papanicolaou and Sircar (2000), where the variable $\xi_{t,t+1}$ plays a similar role. However, the GBS formula (2.11) is even more convenient for empirical pricing since the expectation operator is considered with respect to the historical probability measure instead of the aforementioned formulas involving an equivalent martingale measure. Willard (1997) stressed in the context of the Heston (1993) model, that option pricing formulas viewed as expectations of Black-Scholes prices (and associated greeks derived as expectations of BS greeks) are particularly well-suited for Monte-Carlo simulation of option prices, since they become a conditional Monte Carlo where only the state variable process needs to be simulated.

The GBS formula is useful to interpret the observed shapes of the volatility curve obtained when BS implied volatilities $\sigma_{imp,t}$, defined by $\pi_t^h = BS[S_t, \sigma_{imp,t}(x_t)]$, are plotted against moneyness measured as $x_t = \text{Log}[S_t/KB(t, t+1)]$. Before 1987, this implied volatility curve had most often a U-shaped pattern and was dubbed volatility smile. Renault and Touzi (1996) and Renault (1997) find a rationalization of this symmetric curve. In the absence of the scaling factor $\xi_{t,t+1}$, they show that if:

$$\pi_t^h = E_t[BS[S_t, \sigma^2(U_{t+1})]], \quad (2.12)$$

then the volatility smile is U-shaped, symmetric ($\sigma_{imp,t}(x_t) = \sigma_{imp,t}(-x_t)$) and with the minimum at-the-money (when $x_t = 0$). It is important to note that the symmetry of the volatility smile must be considered with a logarithmic scale for moneyness. Hence, any pair of strike prices a with geometric average equal to the forward stock price produce the same option price.

Since the crash of 1987, observed volatility smiles are often skewed and even reversed, yielding smirks and frowns. In the context of GBS pricing, the observed asymmetries of the volatility curve are explained by the fact that $\xi_{t,t+1}$ is genuinely random, even though it is equal to one in expectation. Therefore, the asymmetries are due to a mixture effect either in the distribution of the return itself ($E_t[S_{t+1}|U_{t+1}] \neq E_t[S_{t+1}]$) or in its covariation with the SDF ($[Cov_t[\log(m_{t+1}), \log(S_{t+1}/S_t)|U_{t+1}]$, which create several kinds of leverage effects.

It is important to note that the option pricing formula is no longer preference-free since the characteristics of the SDF appear explicitly through its covariance with the underlying asset return. Therefore, preference parameters play an explicit role in the option pricing formula. Garcia, Luger and Renault (2001a) use an intertemporal option pricing model with recursive utility and show how asymmetries of the smile are related to the state variable process and to the preference parameters. In Garcia, Luger and Renault (2003), the same model is estimated with option and stock return data on the S&P 500. They provide convincing evidence in favor of the GBS model and the estimated values for the preference parameters are quite reasonable. Guidolin and Timmermann (1999) explain the empirical biases of the Black-Scholes option pricing model by Bayesian learning effects about the state variable, whereas the state variable was known after one period in the previous model. In a model with both learning and leverage effect, David and Veronesi (2002) show that investors' uncertainty about the drift of a firm's fundamentals affects option prices through its effect on stock volatility and the covariance between returns and volatility.

One way to understand the strength of these effects is to recall that the BS formula is convex with respect to the underlying asset current price. Therefore, by the convexity inequality, since $E_t[\xi_{t,t+1}] = 1$, the GBS price should be greater than the price (2.12) corresponding to symmetric volatility smiles, unless the correlation between $\xi_{t,t+1}$ and $\sigma^2(U_{t+1})$ would reverse the Jensen inequality. Therefore, GBS should be equivalent to replacing the current asset price by a greater (deterministic) value. Manaster and Rendleman (1982), Longstaff (1995), and Garcia, Luger and Renault (2002) provide some evidence of such implied index values often greater than the current spot value. Renault (1997) shows through simulations in a Hull and White (1987) model that even very small differences (1 percent) between the implied stock price and its value used to compute implied volatilities will produce severe asymmetries corresponding to widely observed smirks and frowns.

2.6 Multiperiod SDFs and semi-group pricing

The original analysis of Hansen and Richard (1987) was developed in the framework of one-period securities. In this subsection we show that this setting can be extended to multiperiod securities. The simplest is to consider a two-period environment. The price at time $(t + 1)$ of a time $t + 2$ payoff h_{t+2} will be characterized by some SDF m_{t+2} such that: $\pi_{t+1}[h_{t+2}] = E_{t+1}[(m_{t+2})h_{t+2}]$. By the law of one price over two periods, the time t price $\pi_t[h_{t+2}]$ for the time $(t + 2)$ h_{t+2} payoff should be:

$$\pi_t[h_{t+2}] = E_t[(m_{t+1})\pi_{t+1}[h_{t+2}]] = E_t[(m_{t+1})(m_{t+2})(h_{t+2})] \quad (2.13)$$

where m_{t+1} is any SDF able to price a set of payoffs including $\pi_{t+1}[h_{t+2}]$. Therefore, a SDF ${}_t m_{t+2}$ suitable to price at time t payoffs occurring at time $(t + 2)$ will be defined from one-period SDFs by:

$${}_t m_{t+2} = (m_{t+1})(m_{t+2}) \quad (2.14)$$

This simple property is the basis of semi-group pricing developed in the context of Markov models by Hansen and Scheinkman (2001).

2.6.1 Multiperiod state-dependent tree models

In order to extend to an intertemporal context the SDF pricing principle with state variables, we need to adapt the definition of contingent claims. This intertemporal setting needs to be adapted in the case of contingent claims. To parallel the one-period construct of a unique SDF which is mimicked by a derivative asset payoff, it is not enough in a multiperiod setting to condition on the gain process of the underlying asset. We need to introduce, along with the current value of the underlying asset and the path of the underlying asset returns, some state variables summarizing the information relevant for forecasting future returns and SDFs, unless some serial independence is assumed. Therefore, to apply the semi-group pricing (2.14) to the payoffs of a set of contingent claims, we allow for path-dependent contingent claims, which depend on this extended information set.

Let us define a J_t -adapted stochastic process $(U_t)_{t \geq 1}$ of state variables. A derivative asset payoff delivered at time $(t + 1)$ on a given non-dividend paying asset with price S_t at time t can be characterized by a function $h_{t+1} = h(U_1^{t+1}, S_{t+1})$, as opposed to the one-period setting where $h_{t+1} = h(J_t, g_{t+1})$. The main difference is that h_{t+1} not only depends on the past history of the state variables $U_1^t = (U_\tau)_{1 \leq \tau \leq t} \subset J_t$ but also on the future value U_{t+1} which is unknown at time t when the market price $\pi_t(h_{t+1})$ is set. The extension of the space of derivative payoffs ensures consistency between payoffs on consecutive dates. An application of the intertemporal pricing rule (2.13) to a payoff $h_{t+2} = h(U_1^{t+2}, S_{t+2})$ yields a price $\pi_{t+1}[h_{t+2}] = E_{t+1}[(m_{t+2})h_{t+2}]$ which should be written as a function of the state information at time $t + 1$, that is $\pi_{t+1}[h_{t+2}] = f(U_1^{t+1}, S_{t+1})$, for some measurable function f . This price can be written: $\pi_{t+1}[h_{t+2}] = E_{t+1}[(m_{t+2})h^*(U_1^{t+2}, (S_{t+2}/S_{t+1}), S_{t+1})]$, where in h^* the return on the stock appears explicitly. A sufficient condition to obtain the required form $f(U_1^{t+1}, S_{t+1})$ is to make the two following assumptions which amount to a nonparametric state-space model for asset returns and a given SDF:

Assumption 2.3: (Measurement) For any date t and a given sequence (m_t) of SDFs, the pair $[m_{t+2}, (S_{t+2}/S_{t+1})]$ is independent of J_{t+1} given U_1^{t+1} .

Assumption 2.4: (Transition) For any date t , U_{t+2} is independent of J_{t+1} given U_1^{t+1} .

Since $\pi_{t+1}[h_{t+2}]$ is a forecast at time $(t+1)$ of $[(m_{t+2})h_{t+2}] = [(m_{t+2})h^*(U_1^{t+2}, (S_{t+2}/S_{t+1}), S_{t+1})]$, it will in general involve some extra information relevant for forecasting returns and SFDSs but not conveyed by the current value S_{t+1} of the underlying asset price. Therefore, the main role of state variables is to summarize the dynamics of SDFs and asset returns (Assumption 2.3). Without such a summary, one would be unable to apply the SDF paradigm to multiperiod derivative asset pricing, except under the more restrictive assumption that asset returns and SDFs are serially independent. Therefore, a complete market assumption has been replaced by the assumption that the relevant sources of risk are summarized by some state variables for the purpose of defining contingent claim payoffs. Moreover, the state variables also summarize their own dynamics, namely Assumption 2.4 implies that asset returns and SDFs do not cause (in the Granger sense) the state variables. This is a fairly natural assumption, for state variables are supposed to describe the exogenous state of the environment. Assumption 2.4 is similar to the nonlinear factor structure introduced by Bansal and Viswanathan (1993) (see their Assumption 1) to provide a nonlinear extension of the Arbitrage Pricing Theory (APT) of Ross (1976).

Generally speaking, Assumptions 2.3 and 2.4 are valid for most option pricing models with state variables, including discrete and continuous time stochastic processes (the latter will be discussed in the next section). Assumptions 2.3 and 2.4 have been extensively discussed for discrete time applications by Garcia and Renault (2001). Their discussion elaborates in particular on Amin and Ng (1993), where the two forms of observability of state variables by investors, either from the beginning or only at the end of the period, were termed predictability and unpredictability respectively. Garcia and Renault (2001) also emphasize that a typical discrete time application of this idea is the Markov switching regime model of Hamilton (1989). Note also that, by analogy with the continuous time literature, the role of these state variables has not only been described in terms of stochastic volatility (when the state appears explicitly in the variance components) but also in terms of jumps (when the state appears in the mean components). In addition, a binomial model with one unobserved binary state variable is equivalent to the trinomial model considered by Boyle (1988) and multinomial models with K state variables (Kamrad and Ritchken (1991)) are also nested in this setting. The case of latent states in the context of Markov switching models has been studied more recently by Chourdakis and Tzavalis (2001).

In this general setting, the unique one-period SDF mimicked by a derivative asset payoff is $m_{t+1}^* = E_t[m_{t+1}|U_1^{t+1}, S_{t+1}]$. By virtue of the above assumptions, it is a deterministic

function $\lambda[U_1^{t+1}, (S_{t+1}/S_t), S_t]$ of the current value of the underlying asset price, of the underlying return and of the realized path U_1^{t+1} of state variables. Moreover, the replacement of m_{t+1} by m_{t+1}^* does not introduce a violation of the fundamental Assumption 2.3. The fact that the SDF m_{t+1}^* is a deterministic function only of a few “state variables” or “factors” U is similar to Bansal and Viswanathan (1993) Assumption 2, which they present as the key requirement for a nonlinear APT.

So far the intertemporal framework yields m_{t+1}^* , the unique SDF mimicked by a derivative asset payoff. However, since U_{t+1} has not been observed yet by investors at time t , a state-dependent analogue of (2.5) can only be derived under the following additional assumption.

Assumption 2.5: (No leverage effect) $(m_{t+1}, (S_{t+1}/S_t))$ is conditionally independent of U_{t+1} given U_1^t .

Leverage effect, as first pointed out by Black (1976), is akin to an instantaneous causality relationship between return and volatility and is a particular case of the effects precluded by Assumption 2.5. Option pricing in a more general setting which does not preclude this type of effect will be considered below as an extension of mixture models for the loglinear framework.

The fundamental implication of Assumptions 2.3 to 2.5 for the purpose of option pricing is the computation of the risk-neutral density function over two periods as the product of single-period density functions:

$$[B(t, t+1)B(t+1, t+2)]^{-1} {}_t m_{t+2}^* = [B(t, t+1)B(t+1, t+2)]^{-1} m_{t+1}^* m_{t+2}^* \quad (2.15)$$

The multiplicative pricing rule (2.15) results from the application of Assumptions 2.3, 2.4 and 2.5. It is particularly well-suited for tree-models, as it implies that the risk neutral probabilities can be multiplied along the tree. Hence, the role of Assumptions 2.3, 2.4 and 2.5 is to guarantee that the relevant forecasting environment remains multinomial in intertemporal settings (with a fixed number of states of the world for returns at each node) when one has to forecast at time t not only the underlying asset price but also the future price $\pi_{t+1}[h_{t+2}] = E_{t+1}[(m_{t+2})h_{t+2}]$ of contingent claims written on it. For instance, the intertemporal Cox, Ross and Rubinstein (1979) binomial option pricing model corresponds to (2.15) when the risk neutral densities are characterized by the probabilities p_t^* . Note that we have a very general version of this model since, at each date, the two possible values of the return and the associated risk neutral probability may depend on state variables. Without state variables, consecutive returns become independent and one recovers the standard model.

It should be noted however, that the multiplicative pricing rule (2.15) does not provide a well-suited factor structure for long horizons when short term pricing is described by a linear factor model. The loglinear factor structure will be better suited for such a task.

2.6.2 Multiperiod loglinear pricing with state variables

To obtain a GBS formula in a multiperiod context, one only needs, in addition to previous Assumptions 2.3 and 2.4, a joint log-normality assumption of $m_{t,T}$ and S_T/S_t given a path U_1^T of state variables which extends Assumption 2.2.⁷ If the conditional probability distribution of $(\log m_{t+1}, \log S_{t+1}/S_t)$ given U_1^{t+1} is, for $t = 1, \dots, T-1$, assumed to be a bivariate normal:

$$\mathcal{N} \left[\begin{pmatrix} \mu_{mt+1} \\ \mu_{st+1} \end{pmatrix}, \begin{bmatrix} \sigma_{mt+1}^2 & \sigma_{mst+1} \\ \sigma_{mst+1} & \sigma_{st+1}^2 \end{bmatrix} \right].$$

one obtains the following intertemporal GBS option pricing formula which extends (2.11) to a multiperiod setting with possibly stochastic interest rates:

$$\frac{\pi_t}{S_t} = \pi_t(x) = E_t \left\{ \xi_{t,T} \Phi(d_1(x)) - \frac{B^*(t, T)}{B(t, T)} e^{-x} \Phi(d_2(x)) \right\} \quad (2.16)$$

where $x = \log S_t / (KB(t, T))$ and:

$$\begin{aligned} d_1(x) &= \frac{x}{\bar{\sigma}_{t,T}} + \frac{\bar{\sigma}_{t,T}}{2} + \frac{1}{\bar{\sigma}_{t,T}} \log \left[\xi_{t,T} \frac{B(t, T)}{B^*(t, T)} \right] \\ d_2(x) &= d_1(x) - \bar{\sigma}_{t,T} \text{ with } \bar{\sigma}_{t,T}^2 = \sum_{\tau=t}^{T-1} \sigma_{s\tau+1}^2. \end{aligned}$$

with

$$\xi_{t,T} = E \left[{}_t m_T \cdot \left(\frac{S_T}{S_t} \right) | U_1^T \right] = B^*(t, T) \exp \left(\sum_{\tau=t}^{T-1} \sigma_{m s \tau+1} \right) E \left[\frac{S_T}{S_t} | U_1^T \right]$$

and $B^*(t, T) = E [{}_t m_T | U_1^T] = \exp \left(\sum_{\tau=t}^{T-1} \mu_{m\tau+1} + \frac{1}{2} \sum_{\tau=t}^{T-1} \sigma_{m\tau+1}^2 \right)$.

To put this general option pricing formula into perspective, we will compare it to pricing formulas based on equilibrium or absence of arbitrage. Concerning the equilibrium

⁷This assumption is somehow a consequence of a standard conditional central limit argument which can be applied thanks to Assumption 2.3 and to the additivity property of the log SDF in (2.14) through an arbitrary time scale given a path of state variables. Since Clark (1973), there is a long tradition of this approach in financial econometrics. Clark (1973) stressed that non-normality is a puzzle when one has in mind the geometric temporal averaging of the returns and a corresponding central limit theorem argument. In this respect, log normality of returns can be invoked without any significant loss of generality once it is recovered after conditioning on a sufficient number of state variables.

approach, this setting is very general since it is based on a stochastic model for the SDF which does not rely on restrictive assumptions about preferences, endowments, or agent heterogeneity. Moreover, the semigroup property of the SDF is more general than the usual product of intertemporal marginal rates of substitution in time-separable utility models. Indeed, this factorization of the SDF accommodates non-separable or state-dependent preferences. The preference features will appear in general explicitly in the option pricing formula through the characteristics such as mean μ_m , standard deviation σ_m and covariation σ_{ms} of the SDF which are included in $B^*(t, T)$ and $\xi_{t,T}$. Note that $B^*(t, T)$ and $S_t \xi_{t,T}$ can be interpreted as respectively the bond price and the stock price in the fictitious world where the path of state variables U_t^T would be known at time t by the investors. By the law of iterated expectations, the expected value at time t of these fictitious prices coincide with the actual prices: $B(t, T) = E_t[B^*(t, T)]$ and $E_t(\xi_{t,T}) = 1$.

In contrast to the static mixture model described in subsection 2.4, it should be noted that $B^*(t, T)$ is not known at time t as it depends on the future path of state variables and consequently may not be preference-free. Preference parameters may be hidden in the bond and stock prices which will be observed along the lifetime of the option. To see this, first note that by virtue of Assumptions 2.3 and 2.4:

$$\begin{cases} B^*(t, T) = \prod_{\tau=t}^{T-1} B^*(\tau, \tau + 1) \\ \xi_{t,T} = \prod_{\tau=t}^{T-1} \xi_{\tau, \tau+1} \end{cases} \quad (2.17)$$

The one-period ‘‘fictitious pricing’’ $B^*(\tau, \tau + 1)$ and $S_\tau \xi_{\tau, \tau+1}$ at time τ will coincide with actual pricing $B(\tau, \tau + 1)$ and S_τ if and only if there is no mixture effect at horizon one, that is $B^*(\tau, \tau + 1)$ and $\xi_{\tau, \tau+1}$ only depend on the path U_τ^T of state variables already observed at time τ by the investors. Extending a terminology proposed by Amin and Ng (1993), this case where

$$B^*(\tau, \tau + 1) = E_\tau B^*(\tau, \tau + 1) = B(\tau, \tau + 1)$$

and

$$\xi_{(\tau, \tau+1)} = E_\tau(\xi_{\tau, \tau+1}) = 1$$

may be termed the predictable case. It is tantamount to Assumption 2.5 of no leverage effect. Under this assumption:

$$\begin{cases} B^*(t, T) = \prod_{\tau=t}^{T-1} B(\tau, \tau + 1) \\ \xi_{t,T} = 1 \end{cases} \quad (2.18)$$

and the GBS option pricing formula (2.16) is the conditional expectation of the Black-Scholes price, where the expectation is computed with respect to the joint probability

distribution of the rolling-over interest rate $\bar{r}_{t,T} = -\sum_{\tau=t}^{T-1} \log B(\tau, \tau + 1)$ and the cumulated volatility $\bar{\sigma}_{t,T}$. This observation implies that the GBS option pricing formula also nests the common option pricing formulas which are preference-free as only based on the absence of arbitrage. In the most basic case, interest rates and volatility are deterministic and one finds the Black and Scholes (1973) and Merton (1973) formula. When volatility is stochastic, $\bar{\sigma}_{t,T}^2 = \text{Var} [\log(S_T/S_t) | U_1^T]$ is the discrete time analog of the integrated volatility $\int_t^T \sigma_\tau^2 d\tau$ and the GBS formula coincides with the Hull and White (1987) stochastic volatility extension of Black-Scholes, in the particular case where the volatility risk is not compensated. We will even be able to show in subsection 2.7 below that the most popular continuous time stochastic volatility models with possible leverage effects and a premium for the volatility risk (Heston, (1993), Pan (2002)) are nested in the GBS framework. Finally, the GBS formula allows for stochastic interest rates as in Turnbull and Milne (1991) and Amin and Jarrow (1992).

All the aforementioned preference-free option pricing formulas are usually obtained by the absence of arbitrage in either a complete market setting, or an arbitrary choice of a risk neutral probability measure with zero risk premium for latent state variables.

The very general SDF-based option pricing considered in this section does not preclude incompleteness and points out in which cases this incompleteness may invalidate the preference-free paradigm. The only case of incompleteness which matters in this respect is the case of leverage effect which occurs when the conditions (2.18) are not fulfilled.

The violation of these conditions leads to an explicit role of the preference parameters in the GBS option pricing formula and generates mixture effects resulting in volatility smirks and frowns.

2.7 From discrete time to continuous time models

In the previous subsections we emphasized, through the SDF approach, the similarity between the formulas obtained in discrete and continuous time. In particular, we noted that a mixture variable could produce an option pricing model with stochastic volatility and jumps in returns. As we will see in the next section the jump-diffusion model is the canonical model proposed for options written on a stock index. The resulting formulas will look very much the same whether we set the model in discrete time or in continuous time. We already mentioned that Romano and Touzi (1997) obtained a formula which can be written as (2.11) in a Heston (1993) stochastic volatility diffusion model. The fundamental reason behind this similarity lies in the fact that the absence of arbitrage ensures the existence of a state-price density (and conversely) and hence of a stochastic discount factor. In the context of diffusion processes, Assumption 2.3 basically means (see

Garcia and Renault (1998b) for a formal characterization) that the drift and the diffusion terms of the log-stock index and the log-SDF are fixed when the state variables are fixed, in order to recover, conditional on the state variables, the serial independence property of the increments of the Brownian motion. Moreover, the argument can be extended to other Lévy processes, including the Poisson jump process (see Section 3 for a review). Generally speaking, the continuous time analogs of Assumptions 2.3 and 2.4 are fulfilled by all the common continuous time option pricing models and this leads to option pricing formulas similar to the GBS formula of (2.16).

To see this, let us just consider:

$$dS_t = [r(X_t) + P_t^S(X_t)] S_t dt + \sigma(X_t) S_t dW_t^S \quad (2.19)$$

where S_t is the stock index, W_t^S is a standard Brownian motion, $r(X_t)$ is the short term interest rate process, $P_t^S(X_t)$ is the risk premium associated with the diffusive price shock and X_t is a vector of exogenous state variables:

$$dX_t = f(X_t)dt + g(X_t)dW_t^X \quad (2.20)$$

For the sake of expositional simplicity, a dividend process for the stock has not been made explicit in the drift of (2.19). If such dividends exist, the genuine risk premium is the sum of $P_t^S(X_t)$ and of the dividend process. It is also worth noting that the exogeneity of the state variables process X_t means that the drift and diffusion coefficients in (2.20) only depend upon X_t and not on S_t or on the SDF m_t . The exogeneity assumption is the continuous time analog of the no-Granger causality assumption 2.4 (see Comte and Renault (1996) for a formal characterization) and does not preclude any kind of leverage effects through instantaneous non zero correlations between W_t^S and the components of W_t^X . The continuous time analog of Assumption 2.3 about the pricing kernel m_t amounts to a specification:

$$d\log(m_t) = h(X_t)dt + a(X_t)dW_t^X + b(X_t)dW_t^{X\perp}. \quad (2.21)$$

where $W_t^{X\perp}$ is defined by the instantaneous regression of the Brownian motion W_t^S on W_t^X

$$W_t^S = \rho_t W_t^X + (1 - \rho_t^2)^{1/2} W_t^{X\perp} \quad (2.22)$$

with $\rho_t dt = Cov_t [dW_t^S, dW_t^X]$. Note that the continuous time notation $(m_t)_{t \in [0, T]}$ leads to a slight change of notation with respect to the previous subsections. By virtue of the multiplicative rule (2.14), we will now denote by m_{t_2}/m_{t_1} the SDF which gives the price at time t_1 of payoffs occurring at time $t_2 > t_1$.

Equations (2.19), (2.20) and (2.21) when augmented with a Poisson jump component (which would be easy to introduce) nests all the commonly used jump-diffusion option pricing models that will be further discussed in Section 3 below. One can also easily derive the associated GBS option pricing formulas and discuss their dependence on preference characteristics, through the specification of risk premia associated to the various sources of risk. Let us do this exercise in the simplest case of only one state variable, that is a univariate process X_t .

Generally speaking, Assumptions 2.3 and 2.4 do not preclude the instantaneous correlation coefficient ρ_t to be a deterministic function of X_t (see Garcia and Renault (1998b)). We will consider it here as a constant ($\rho_t = \rho$) for expositional simplicity.

As previously noted some characteristics of the SDF are hidden in the bond and in the stock prices. To see this consider the pricing equation of a bank account:

$$m_t = E_t \left[m_{t+h} \exp \left(\int_t^{t+h} r(X_\tau) d\tau \right) \right] \quad (2.23)$$

which implies that:

$$h(X_t) = -r(X_t) - \frac{1}{2} [a^2(X_t) + b^2(X_t)] \quad (2.24)$$

Then, the stock pricing equation:

$$m_t S_t = E_t [m_{t+h} S_{t+h}] \quad (2.25)$$

implies that:

$$(1 - \rho^2)^{1/2} b(X_t) \sigma(X_t) = -\rho a(X_t) \sigma(X_t) - P_t^S(X_t) \quad (2.26)$$

Equation (2.26) says that $b(X_t)$ is fully defined when two risk premia are specified: (1) The risk premium $P_t^S(X_t)$ associated with the diffusive price shock, and (2) the risk premium $a(X_t)\sigma(X_t)$ associated with the volatility shock.

Moreover, from (2.24), observations of the risk free rate $r(X_t)$ characterize the SDF instantaneous rate of expected growth $h(X_t)$. Note also that, insofar as one considers that both the risk free rate $r(X_t)$ and the instantaneous expected rate of return on the stock ($r(X_t) + P_t^S(X_t)$) are observed, the only characteristic of the pricing kernel that is not fully hidden in the bond and stock prices is its state variable diffusion coefficient $a(X_t)$, or equivalently, the volatility risk premium $a(X_t)\sigma(X_t)$.⁸

⁸To draw on analogy with the previously described discrete time framework, note that if the SDF lognormal conditional probability distribution were characterized through a parametric specification of μ_m, σ_m and σ_{ms}, σ_{ms} would have been deduced from $\left(\mu_m + \frac{\sigma_m^2}{2}\right)$ through the stock pricing equation while μ_m could be deduced from σ_m^2 through the bond pricing equation. Then, in the discrete case too, the volatility coefficient of the SDF is the crucial channel for the additional role of preferences in option pricing.

It is then convenient, exactly as it has been done in discrete time, to define two variables $B^*(t, T)$ and $S_t \xi_{t,T}$ that can respectively be interpreted as the bond price and the stock price in the fictitious world where the path X_t^T of state variables would be known at time t by investors:

$$\begin{aligned} B^*(t, T) &= E_t \left[\frac{m_T}{m_t} \mid X_\tau, t \leq \tau \leq T \right] \\ &= \exp \left[- \int_t^T r(X_\tau) d\tau \right] \exp \left[\int_t^T a(X_\tau) dW_\tau^X - \frac{1}{2} \int_t^T a^2(X_\tau) d\tau \right] \end{aligned} \quad (2.27)$$

and:

$$S_t \xi_{t,T} = E_t \left[\frac{m_T}{m_t} \cdot S_T \mid X_\tau, t \leq \tau \leq T \right]$$

that is:

$$\xi_{t,T} = \exp \left[\int_t^T [a(X_\tau) + \rho \sigma(X_\tau)] dW_\tau^X \right] \exp \left[- \frac{1}{2} \int_t^T \{a(X_\tau) + \rho \sigma(X_\tau)\}^2 d\tau \right] \quad (2.28)$$

By the law of iterated expectations, the expected value at time t of these fictitious prices coincide with the actual prices:

$$B(t, T) = E_t [B^*(t, T)] \quad \text{and} \quad E_t [\xi_{t,T}] = 1.$$

Computing the price of an European call:

$$\pi_t = E_t \left[\frac{m_T}{m_t} \cdot \text{Max} [0, S_T - K] \right]$$

we obtain the GBS option pricing formula similar to (2.11) and (2.16):

$$\pi_t = E_t \left[\widetilde{BS}(S_t \xi_{t,T}, \bar{\sigma}_{t,T}^2) \right] \quad (2.29)$$

where: $\bar{\sigma}_{t,T}^2 = (1 - \rho^2) \int_t^T \sigma^2(X_\tau) d\tau$ and $\widetilde{BS}(\cdot, \cdot)$ is a BS-like option pricing formula where the standard discount factor $\exp \left[- \int_t^T r(X_\tau) d\tau \right]$ has been replaced by $B^*(t, T)$ as in (2.16).

In contrast to similar option pricing formulas derived by Romano and Touzi (1997) and Fouque, Papanicolaou and Sircar (2000), the expectation (2.29) is with respect to the historical distribution, while the risk-neutral probability distribution they consider corresponds to the particular case $a(X_t) \equiv 0$. In particular, they have not addressed the difference between $B^*(t, T)$ and $\exp \left[- \int_t^T r(X_\tau) d\tau \right]$.

It is also worth noticing that the distortion on the stock price through the scaling factor $\xi_{t,T}$ is now the result of two effects: (1) the nonzero volatility risk premium coefficient $a(X_t)$ and (2) the nonzero leverage effect coefficient ρ .

Two remarks are in order in this respect. On the one hand, the discrete time framework, although more general than the continuous time one, was unable to accommodate such distortions through volatility risk premia in the zero-leverage case. On the other hand, volatility risk premia that are not associated with leverage are somewhat immaterial in terms of volatility smile shapes.

Since the option price can always be interpreted as an expectation with respect to the risk neutral probability distribution, that is with $a(X_t) = 0$ (and $P_t^S(X_t) = 0$), zero-leverage will always imply that the volatility smile is U-shaped, symmetric and minimal at the money. The general result of Renault (1997) regarding (2.12) still applies. Therefore, in case of zero-leverage, the only way to assess the volatility risk premium is to statistically compare the risk neutral and the historical probability distributions.

The volatility smile is still useful in this respect since we know (see Breeden and Litzenberger (1978)) that the conditional cumulative risk-neutral distribution function is:

$$\begin{aligned} F_t(S_{t+1}) &= [B(t, t+1)]^{-1} \frac{\partial \pi_t}{\partial K} \Big|_{K=S_{t+1}} + 1 \\ &= [B(t, t+1)]^{-1} \Delta_t(\sigma_{imp,t}) \frac{\partial \sigma_{imp,t}}{\partial K} \Big|_{K=S_{t+1}} + 1 \end{aligned} \quad (2.30)$$

where $\Delta_t(\sigma)$ denotes the BS delta of the option for a value σ . In particular, while in case of zero-leverage the risk neutral distribution does not feature any excess skewness with respect to the historical one, the risk neutral return variance and the historical one may still differ (see Rosenberg (2000)).

3 Modelling asset price dynamics via diffusions for the purpose of option pricing

Since the seminal papers of Black and Scholes (1973) and Merton (1973), the greater part of option pricing models have been based on parametric continuous time models for the underlying asset. The overwhelming rejection of the constant variance geometric Brownian motion lead to a new class of stochastic volatility models introduced by Hull and White (1987) and reviewed in Ghysels, Harvey and Renault (1996). While the models in the SV class are by now well-established, there are still a number of unresolved issues about their empirical performance.

The work of Heston (1993), who proposed a SV diffusion with an analytical option pricing formula, was generalized by Duffie and Kan (1996) and Duffie, Pan and Singleton (2000) to a rich class of affine jump diffusions. This class will be reviewed in a first subsection. Alternative models, mostly non-affine, will be covered in the second subsection. A final subsection discusses option pricing without estimated prices of risk.

3.1 The affine jump-diffusion class of models

The general class of affine jump diffusion (AJD) models examined in detail by Duffie, Pan and Singleton (2000) (DPS hereafter) includes as special cases many option pricing models that have been the object of much econometric analysis in the past few years. To describe the class of processes consider the following return dynamics, where $d\log S_t = dU_{1t}$ with U_{1t} is the first element of a vector process N -dimensional U_t which represents the continuous path diffusion component of the return process, the second term $\exp \Delta X_t - \iota$ represents discrete jumps, where X_t is a N -dimensional Lévy process and ι a vector of ones. The process U_t is governed by the following of equations:

$$dU_t = \mu(U_t, t)dt + \sigma(U_t, t)dW_t + \exp \Delta X_t - \iota \quad (3.1)$$

where the process U_t is Markovian and takes values in an open subset D of \mathbf{R}^N , $\mu(y) = \Theta + \mathcal{K}y$ with $\mu : D \rightarrow \mathbf{R}^N$ and $\sigma(y)\sigma(y)' = h + \sum_{j=1}^N y_j H^{(j)}$ where $\sigma : D \rightarrow \mathbf{R}^{N \times N}$. Moreover, the vector Θ is $N \times 1$, the matrix \mathcal{K} is $N \times N$ whereas h and H are all symmetric $N \times N$ matrices. The process W_t is a standard Brownian motion in \mathbf{R}^N . While the first component of the U_t process relates to returns, the other components U_{it} for $i = 2, \dots, N$ either govern the stochastic drift or volatility of returns.⁹ This setup is a general affine structure that allows for jumps in returns (affecting the first component U_{1t}) and the less common situation of jumps in volatility factors (affecting the components U_{it} that determine volatility factors). Empirical models for equity have at most $N = 4$, where the U_{2t} affects the drift of U_{1t} and U_{3t} and U_{4t} affect either the volatility or jump intensity (see Chernov, Gallant, Ghysels and Tauchen (2000, 2003) for examples). We will start with single volatility factor models, followed by a discussion of jump diffusions and models with multiple volatility factors.

3.1.1 Models with a single volatility factor

The class is defined as the following system of stochastic differential equations:

$$\begin{pmatrix} dY_t \\ dV_t \end{pmatrix} = \begin{pmatrix} \mu \\ \kappa(\theta - V_t) \end{pmatrix} dt + \sqrt{V_t} \begin{pmatrix} 1 & 0 \\ \rho\sigma_v & \sqrt{(1-\rho^2)}\sigma_v \end{pmatrix} dW_t + \xi dN_t \quad (3.2)$$

where Y_t is the logarithm of the asset price S_t , $W_t = (W_{1t}, W_{2t})'$ is a vector of independent standard Brownian motions, $N_t = (N_t^y, N_t^v)'$ is a vector of Poisson processes with constant

⁹All further details regarding the regularity conditions pertaining to the U_t are discussed by DPS (2000) and therefore omitted.

arrival intensities λ_y and λ_v and $\xi = (\xi^y, \xi^v)'$ is a vector of jump sizes for returns and volatility respectively¹⁰. We adopt the mnemonics used by DPS and Eraker, Johannes and Polson (2001): SV for stochastic volatility models with no jumps in returns nor volatility ($\lambda_y = \lambda_v = 0$); SVJ, for stochastic volatility models with jumps in returns only ($\lambda_y > 0$, $\lambda_v = 0$); SVJJ, for stochastic volatility models with jumps in returns and volatility ($\lambda_y > 0$, $\lambda_v > 0$). In SVJ the jump size is distributed normally, $\xi^y \sim \mathcal{N}(\mu_y, \sigma_y^2)$. The SVJJ can be split into the SVIJ model (with independent jump arrivals in returns and volatility and independent jump sizes $\xi^y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ and $\xi^v \sim \exp(\mu_v)$) and the SVCJ model (with contemporaneous Poisson jumps arrivals in returns and volatility, $N_t^y = N_t^v$ with arrival rate λ_y and correlated sizes $\xi^v \sim \exp(\mu_v)$ and $\xi^y | \xi^v \sim \mathcal{N}(\mu_y + \rho_J \xi^v, \sigma_y^2)$).

A number of recent papers have investigated the Heston (1993) SV model. Most papers (Andersen, Benzoni and Lund (2002), Benzoni (1998), Eraker, Johannes and Polson (2001)) conclude that the stochastic volatility model provides a much better fit of stock returns than standard one-factor diffusions. In particular, the strong negative correlation around -0.4 found between the volatility shocks and the underlying stock returns shocks captures well the negative skewness observed in stock returns. However, the model is rejected since it is unable to accommodate the excess kurtosis observed in the stock returns¹¹. Basically, it cannot fit the large changes in stock prices occurring during crash-like events. In the SV model, there is a strong volatility persistence (the estimated value for the mean reversion parameter κ is in the order of 0.02).

Adding jumps in returns appears therefore natural since the continuous path stochastic volatility accommodates the clustered random changes in the returns volatility whereas the discrete Poisson jump captures the large infrequent price movements. However, jump components are difficult to estimate and complicate the extraction of the volatility process¹². Eraker, Johannes and Polson (2001) propose a likelihood-based methodology using Markov Chain Monte Carlo methods. Their estimation results for the period 1980-1999 show that the introduction of jumps in returns in the SVJ model has an important downward impact on the parameters of the volatility process. The parameters for average volatility, the

¹⁰A specification with βV_t in the drift of the returns equation was considered by Eraker, Johannes and Polson (2001). This additional term was found to be insignificant, in accordance with the findings of Andersen, Benzoni and Lund (2001) and Pan (2002).

¹¹Both Andersen, Benzoni and Lund (2001) and Benzoni (1998) estimate a non-affine specification with the log variance. The model fits slightly better than the affine SV model but it is still strongly rejected by the data. Jones (2003) estimates a SV model with CEV volatility dynamics but it generates too many extreme observations.

¹²For a discussion of the different types of volatility filters see Ghysels, Harvey and Renault (1996) and the chapter of Gallant and Tauchen (2001) in this Handbook.

volatility of volatility and the speed of mean reversion all fall dramatically. This is somewhat consistent with the results of Andersen, Benzoni and Lund (2002) when they estimate the models from 1980 till 1996, but with less magnitude. However, in the latter study, parameters associated with volatility change much less when the models are estimated over a longer period (1953 to 1996). The difference between the two latter studies is to be found in the estimates of the jump process. In Eraker, Johannes and Polson (2001), jumps arrive relatively infrequently, about 1.5 jumps per year and are typically large. The jump mean is -2.6% and the volatility is just over 4%. The large sizes of jumps are in contrast with the smaller estimates (μ_y of zero and σ_y less than 2%) obtained by Andersen, Benzoni and Lund (2002) and Chernov, Gallant, Ghysels and Tauchen (2003). The introduction of jumps lowers the negative correlation between the innovations in returns and in volatility. In all studies, the SVJ model appears to be less misspecified than the SV model.

All econometric complexities put aside, other issues remain. Adding jumps resolves the misfit of the kurtosis on the *marginal* distribution of returns, but one may suspect that the dynamic patterns of extreme events are not particularly well captured by an independent Poisson process. The stochastic structure of a one factor SV model augmented with a Poisson jump process implies that the day after a stock market crash another crash is equally likely as the day before. In addition, the occurrence of jumps is independent of volatility. Clearly, the independent Poisson process has unappealing properties and therefore some alternative models for jumps, i.e. alternative Lévy specifications, have been suggested. Bates (2000) estimated a class of jump-diffusions with random intensity for the jump process, more specifically where the intensity is an affine function of the stochastic volatility component. Duffie, Pan and Singleton (2000) generalize this class and Chernov, Gallant, Ghysels and Tauchen (2000), Eraker, Johannes and Polson (2001) and Pan (2002) estimate multi-factor jump-diffusion models with affine stochastic jump intensity. The models considered by DPS are:

$$\lambda(U_t) = \lambda_0(t) + \lambda_1(t)U_t \tag{3.3}$$

where the process U_t is of the affine class as V_t specified in (3.2). These structures may not be entirely suitable either to accommodate some stylized facts. Suppose one ties the intensity to the volatility factor V_t in (3.2), meaning that high volatilities implies high probability of a jump. This feature does not take into account an asymmetry one observes with extreme events. For instance the day before the 1987 stock market crash the volatility measured by the squared return on the S&P 500 index was roughly the same as the day after the crash. Therefore, in this case making the intensity of a crash a linear affine function of volatility would result in the probability of a crash the day after Black Monday being the same as

the trading day before the crash. Obviously, one could assign a factor specific to the jump intensity and governed by an affine diffusion. Hence, one adds a separate factor U_t that may be correlated with the volatility factor V_t . Pan (2002) examines such processes and provides empirical estimates. Chernov, Gallant, Ghysels and Tauchen (2000) and Eraker, Johannes and Polson (2001) consider also a slightly more general class of processes:

$$\lambda(x, U) = \lambda_0(x, t) + \lambda_1(x, t)U_t \quad (3.4)$$

where for instance $\lambda_i(x, t) = \lambda_i(t) \exp(G(x))$. This specification yields a class of jump Lévy measures which combines the features of jump intensities depending on, say volatility, as well as the size of the previous jump. The virtue of the alternative more complex specifications is that the jump process is no longer independent of the volatility process, and extreme events are more likely during volatile market conditions. There is, however, an obvious drawback to the introduction of more complex Lévy measures as they involve a much more complex parametric structure. Take for example the case where the jump intensity in (3.3) is a function of a separate stochastic factor U_t correlated with the volatility process V_t . Such a specification may involve up to 6 additional parameters to determine the jump intensity, without specifying the size distribution of jump. Chernov, Gallant, Ghysels and Tauchen (2000) endeavor into the estimation of various complex jump processes using more than a 100 years of daily Dow Jones data and find that it is not possible to estimate rich parametric specifications for jumps even with such long data sets.¹³

Despite all these reservations about jump processes, one has to note that various papers have not only examined the econometric estimation but also the derivative security pricing with such processes. In particular, Bakshi and Madan (2000) and Duffie, Pan and Singleton (2000) provide very elegant general discussions of the class of affine jump-diffusions with stochastic volatility which yield analytic solutions to derivative security pricing. One has nevertheless to bear in mind the empirical issues that are involved. A good example is the affine diffusion with jumps. In such a model there is a price of jump risk and a price of risk for jump size, in addition to the continuous path volatility factor risk price and return risk. Hence, there are many risk prices to be specified in such models. Moreover, complex specifications of the jump process with state-dependent jump intensity, result in an even larger number of prices of risk.

¹³Chernov, Gallant, Ghysels and Tauchen (2000) also examine non-affine Lévy processes, which will be covered in the next subsection.

3.1.2 Multiple volatility factors

Affine diffusion models are characterized by drift and variance functions which are linear functions of the factors. Instead of considering additional factors that govern jump intensities one might think of adding more continuous path volatility factors. Dai and Singleton (2000) discuss the most general specification of such models including the identification and admissibility conditions. Let us reconsider the specification of V_t in (3.2) and add a stochastic factor to the drift of returns, namely:

$$\begin{aligned}
 dY_t &= (\alpha_{10} + \alpha_{12}U_{1t})dt + \sqrt{\beta_{10} + \beta_{12}U_{2t} + \beta_{13}U_{3t}}(dW_{1t} + \psi_{12}dW_{2t} + \psi_{13}dW_{3t}) \\
 dU_{1t} &= (\alpha_{20} + \alpha_{22}U_{1t})dt + \beta_{20}dW_{2t} \\
 dU_{it} &= (\alpha_{i0} + \alpha_{ii}U_{it})dt + \sqrt{\beta_{i0} + \beta_{ii}U_{it}}dW_{it}, \quad i = 2, 3
 \end{aligned} \tag{3.5}$$

The volatility factors enter additively into the diffusion component specification. Hence, they could be interpreted as short and long memory components as in Engle and Lee (1999). The long memory (persistent) component should be responsible for the main part of the returns distribution, while the short memory component will accommodate the extreme observations. This specification allows feedback, in the sense that the volatilities of the volatility factors can be high via the terms $\beta_{ii}U_{it}$ when the volatility factors themselves are high. Adding a second volatility factor helps fitting the kurtosis, using arguments similar to those that explain why jumps help fitting the tails. The extra freedom to fit tails provided by an extra volatility factor has its limitations, however, as noted by Chernov, Gallant, Ghysels and Tauchen (2003). In fact, their best model, which does fit the data at conventional levels, is not an affine model (see next subsection).

Bates (2000) and Pan (2002) argue that the specification of the volatility process should include jumps, possibly correlated with the jumps in returns. This is an alternative to expanding the number of volatility factors. It has the advantage that one can fit the persistence in volatility through a regular affine specification of V_t and have extreme shocks to volatility as well through the jumps, hence capturing in a single volatility process enough rich features that simultaneously fit the clustering of volatility and the tails of returns. The drawback is that one has to identify jumps in volatility, a task certainly not easier than identifying jumps in returns.

3.2 Other continuous time processes

By other continuous time processes we mean a large class of processes that are either non-affine, or affine but do not involve the usual jump-diffusion processes but more general Lévy

processes or fractional Brownian motions. Three subsections describe the various models that have been suggested.

3.2.1 Non-Affine Index models

Another way to capture the small and large movements in returns is to specify stochastic volatility models with two factors as in Chernov, Gallant, Ghysels and Tauchen (2003). They propose to replace the affine setup (3.5) by some general volatility index function $\sigma(U_{2t}, U_{3t})$ able to disentangle the effects of U_{2t} and U_{3t} separately and therefore have a different effect of short and long memory volatility components. In particular they consider:

$$\sigma(U_{2t}, U_{3t}) = \exp(\beta_{10} + \beta_{12}U_{2t} + \beta_{13}U_{3t}) \quad (3.6)$$

$$dU_{it} = (\alpha_{i0} + \alpha_{ii}U_{it})dt + (\beta_{i0} + \beta_{ii}U_{it})dW_{it}, \quad i = 2, 3 \quad (3.7)$$

Chernov, Gallant, Ghysels and Tauchen (2003) study two different flavors of the logarithmic models, depending on the value of the coefficients β_{ii} . When $\beta_{ii} = 0$, the volatility factors are described by Ornstein-Uhlenbeck processes. In this case, the drift and variance of these factors are linear functions and, hence, the model can be described as logarithmic or log-affine. Whenever, $\beta_{ii} \neq 0$ either for $i = 2$ or 3 there is feedback, a feature found to be important in Gallant, Hsu, and Tauchen (1999), and Jones (2003). The exponential specification in (3.6) is of course not the only index function one can consider. Meddahi (2001) proposes another way to break the link between volatility persistence and fat tails by specifying a flexible form (linear combination of Hermite polynomials) for the diffusion coefficient in the returns equation.

Chernov, Gallant, Ghysels and Tauchen (2003) show that the exponential specification with two volatility factors (without jumps) yields a remarkably good empirical fit, i.e. the model is not rejected at conventional significance levels unlike the jump diffusion and affine two-factor models discussed in the previous section. Others have also found that such processes fit very well, see for instance Alizadeh, Brandt and Diebold (1999), Chacko and Viceira (1999), Gallant, Hsu and Tauchen (1999) and the two-factor GARCH model of Engle and Lee (1999). The fact that logarithmic volatility factors are used, instead of the affine specification, adds the flexibility of state-dependent volatility as noted by Jones (2003). In addition, an appealing feature of the logarithmic specification is the multiplicative effect of volatility factors on returns. One volatility factor takes care of long memory, whereas the second factor is fast mean-reverting (but not a spike like a jump). This property of logarithmic models facilitates mimicking the very short-lived but erratic

extreme event behavior through the second volatility factor. Neither one volatility factor models with jumps nor affine two-factor specifications are well equipped to handle such patterns typically found during financial crises.

It should also be noted that the two-factor logarithmic specification avoids several econometric issues. We noted that the presence of jumps also considerably complicates the extraction of the latent volatility and jump components since traditional filters no longer apply. In contrast, the continuous path two-factor logarithmic SV process does not pose any difficulties for filtering via reprojection methods as shown by Chernov, Gallant, Ghysels and Tauchen (2003). There is another appealing property to the two-factor logarithmic SV model: the model has a smaller number of risk factors compared to many of the alternative specifications, specifically those involving complex jump process features. The major drawback of this class of processes, however is the lack of an explicit option pricing formula: simulation-based option pricing is the only approach available.

3.2.2 Lévy processes and time deformation

It was noted before that one could easily relax normality in discrete time models through the introduction of mixture distributions. Likewise, in the context of continuous time models it was noted that one can replace Brownian motions by so called Lévy processes. The typical setup is through subordination, also referred to as time deformation, an approach suggested first in the context of asset pricing by Clark (1973) and used subsequently in various settings. The idea to use a Lévy process to change time scales and thus random changes in volatility can be interpreted as a clock ticking at the speed of information arrival in the market. For further discussion see for example Barndorff-Nielsen and Shephard (2001a,b, 2003), Clark (1973), Ghysels, Gouriéroux and Jasiak (1997), Madan and Seneta (1990), Tauchen and Pitt (1983), among many others.

The purpose of this section is to survey the option pricing implications of assuming the broader class of time deformed Lévy processes. Various authors have studied option pricing with this class of processes, including most recently Carr, Geman, Madan and Yor (2003), Carr and Wu (2002) and Nicolato and Venardos (2003). The latter follow closely the setup of Barndorff-Nielsen and Shephard, which we adopt here as well. We already introduced in equation (3.1) the class of affine jump-diffusion processes. Nicolato and Venardos consider a different class, namely:

$$dY_t = (\mu + \beta\sigma_t^2)dt + \sigma_t dW_t + \rho dZ_{\lambda t} \quad (3.8)$$

$$d\sigma_t = -\delta\sigma_t^2 dt + dZ_{\delta t} \quad (3.9)$$

with $\delta > 0$ and $\rho \leq 0$. The process $Z = (Z_{\delta t})$ is subordinator, independent of the Brownian

motion W_t , assumed to be a Lévy process with positive increments, called by Barndorff-Nielsen and Shephard (2001a) the background driving Lévy process. It is assumed that Z has no deterministic drift and its Lévy measure has a density λ . Note that the solution to (3.9) can be written as:

$$\sigma_t^2 = \exp -\delta t \sigma_0^2 + \int_0^t \exp t - s dZ_{\delta s} \quad (3.10)$$

The resulting dynamics of the stock price process are

$$\begin{aligned} dS_t &= S_{t-}(b_t dt + \sigma_t + dM_t) \\ db_t &= \mu + \delta \kappa(\rho) + (\beta + \frac{1}{2})\sigma_t^2 \\ M_t &= \sum_{0 < s \leq t} (\exp \rho \Delta Z_{\delta s} - 1) - \delta \kappa(\rho)t \end{aligned} \quad (3.11)$$

where $\kappa(x)$ is the cumulant transform, i.e. $\kappa(x) = \log E[\exp xZ_1]$. To build models of time deformation one exploits the property (see e.g. Sato (1999)) that for any self-decomposable probability distribution \mathcal{L} there exists a Lévy process Z such that the a *OU* process driven by Z has \mathcal{L} as marginal. Examples of self-decomposable distributions are the inverse Gaussian and Gamma distributions. Therefore, two popular models to specify the variance process are the so called *IG-OU* and $\Gamma-OU$ processes studied respectively by Barndorff-Nielsen and Shephard (2001a) and Madan and Seneta (1990). Carr, Geman, Madan and Yor (2003) study other processes as well, generalizations of Madan and Seneta and Normal inverse Gaussian models. De facto these processes are one-factor volatility models of the non-affine type, or view differently one has replaced Gaussian mixtures (in this case *OU* type processes) by processes that have a mixture representation.

The characteristic functions for the log of price can be derived in all the aforementioned cases and can be used to obtain option prices via the Fast Fourier transform. Equivalent martingale representations are obtained through measure changes within the class of *OU* process driven by Z . One interesting case that we would like to highlight is obtained by Nicolato and Venardos (2003), who express the call price of a European option as conditional expectation of the Black-Scholes formula using so called *effective* log-stock prices, namely:

$$\pi_t^h = E^{Q^*}[BS(Y_{eff}, V_{eff})|Y_t, \sigma_t^2] \quad (3.12)$$

similar to an expression of Hull and White (1987) and similar to the GBS formula discussed earlier, except that here (as in Hull and White) the expectation is taken under the risk neutral expectation. The *effective* log-price process X_{eff} is the original process X_t modified by the path of the future subordinator ($Z_{\delta T} - Z_{\delta t}$ where T is the maturity date of the

contract and V_{eff} is the (re-scaled) future realized volatility between t and T . Because of the processes involved, this formula applies to a wide variety of non-affine diffusions with leverage as well as jump-diffusions. To compute actual option prices, Nicolato and Venardos (2003) suggest to simulate the pair (Y_{eff}, V_{eff}) and provide the relevant references to do so. Carr, Geman, Madan and Yor (2003) construct option prices differently, following a method developed in Carr and Madan (1998) using a generalized Fourier transforms and parameters calibrated with cross-sections of option contracts.

3.2.3 Long-memory in Continuous Time

In section 2.5 we noted that numerous distorted smiles in the shapes of smirks or frowns are often inferred from market data since 1987 and provide an explanation in terms of stochastic volatility and its instantaneous correlation with the return of the underlying asset. However, as pointed out by Sundaresan (2000) in his recent survey of the performance of continuous-time methods for option valuation, the remaining puzzle is the so-called term structure of volatility smiles, that is, the fact that the volatility smile effect appears to be dependent, in a systematic way, on the maturity structure of options. Sundaresan (2000) first observes that the volatility smile appears to be stronger in short term options than in longer term ones, which is consistent with the stochastic volatility interpretation. When volatility is stochastic, the option price appears to be an expectation of the BS price with respect to the probability distribution of the so-called integrated volatility $(1/h) \int_t^{t+h} \sigma^2(u) du$ over the lifetime of the option (see Renault and Touzi (1996) in the context of the Hull and White (1987) model) or of a fraction of it in case of leverage effect (see Romano and Touzi (1997) in the context of the Heston (1993) model). Then, by a simple application of the law of large numbers to time averages of the volatility process (assumed to be stationary and ergodic), one realizes that the effects of the randomness of the volatility should vanish when the time to maturity of the option increases and therefore the volatility smile should be erased.

Nevertheless, as Sundaresan (2000) emphasizes, the term structure of implied volatilities still appears to have short term and long term patterns that cannot be so easily reconciled. Introducing long memory in the stochastic volatility process appears to be useful in this respect. To see this, it is worth revisiting the common claim that the convexity of the volatility smile is produced by the unconditional excess kurtosis of log returns. For notational simplicity, we consider that the log price has a zero deterministic drift and that there is no leverage effect, i.e. using the notations of subsection 2.7, the two Wiener processes

W^S and W^X are independent and the log return over the period $[t, t + h]$ can be written:

$$R_t(h) = \log \frac{S_{t+h}}{S_t} = \int_t^{t+h} \sigma_u dw W_u^s,$$

where the two stochastic processes σ and w^s are independent. Hence, given the volatility path, the log return is normal and we can write:

$$E [R_t^2(h) / \sigma] = \int_t^{t+h} \sigma_u^2 du$$

and

$$E [R_t^4(h) / \sigma] = 3 \left[\int_t^{t+h} \sigma_u^2 du \right]^2.$$

The unconditional kurtosis of the return over the period $[t, t + h]$ is therefore given by:

$$k(h) = \frac{E [R_t^4(h)]}{(E [R_t^2(h)])^2} = 3 \left[1 + \frac{\text{Var} \left[\frac{1}{h} \int_t^{t+h} \sigma_u^2 du \right]}{(E(\sigma^2))^2} \right] \quad (3.13)$$

Then, to address the issue of consistency between short term and long term patterns, it is worth considering the limit cases of infinitely short time to maturity ($h \rightarrow 0$) and infinitely long time to maturity ($h \rightarrow \infty$). First, since $\frac{1}{h} \int_t^{t+h} \sigma_u^2 du$ converges in mean-square towards σ_t^2 when h tends to zero:

$$\lim_{h \rightarrow 0} k(h) = 3 \left[1 + \frac{\text{Var}(\sigma^2)}{(E(\sigma^2))^2} \right] \quad (3.14)$$

Equation (3.14) is a specialization to very short term intervals of a well-known result since Clark (1973): the excess kurtosis is equal to 3 times the squared coefficient of variation of the stochastic variance. This excess kurtosis effect persists in the very short term even though the volatility smile evaporates and the conditional variance $V_t \left[\frac{1}{h} \int_t^{t+h} \sigma_u^2 du \right]$ tends to zero. This is a counterexample to the claim that convexity of the volatility smile is simply produced by unconditional excess kurtosis. As already previously noted, observed violations of Black-Scholes pricing for very short-term options cannot be captured within the one factor stochastic volatility framework without introducing a huge volatility risk premium which would become explosive in longer term options. This explains why jumps, multiple volatility factors or other nonlinearities have been introduced.

The focus of interest here is the remaining puzzle that stochastic volatility still appears to be significant for very long maturity options as documented by Bollerslev and Mikkelsen (1999). The implied level of volatility persistence to account for deep volatility smiles in long-term options is large in the framework of standard (short memory) models of volatility dynamics, even with a model of permanent and transitory component as in Engle and Lee

(1999). Moreover, this cannot be easily reconciled with the stylized fact that the sample autocorrelogram of squared asset returns generally decreases quite abruptly in the short term whereas it appears to converge slowly to zero in the long term. In order to address this issue, Comte and Renault (1998) proposed a continuous time stochastic volatility model with long memory. Long memory in volatility dynamics is a well-documented empirical fact (see e.g. Ding, Granger and Engle (1993)) which has given rise to various long-memory GARCH models, (Baillie, Bollerslev and Mikkelsen (1996), Bollerslev and Mikkelsen (1996), Robinson (1991)) and long-memory discrete time stochastic volatility models (Breidt, Crato and De Lima (1998), Harvey (1998)).

To get a long-memory continuous time stochastic volatility model, the basic idea of Comte and Renault (1998) was to extend the log-normal stochastic volatility model to fractional Brownian motion. The log-volatility process follows Ornstein-Uhlenbeck dynamics, but with the standard Brownian motion replaced by a fractional one. Since the main strand of the volatility literature is now more oriented towards affine models, we rather present here an overview of the affine fractional stochastic volatility of Comte, Coutin and Renault (2003) (CCR hereafter). The results are qualitatively similar to Comte and Renault (1998) but the affine setting provides closed form formulas useful for interpretation and option pricing applications as well. Starting from a CIR stochastic volatility model as in Heston (1993), $d\tilde{\sigma}^2(t) = k(\tilde{\theta} - \tilde{\sigma}^2(t))dt + \gamma\tilde{\sigma}(t)dW^X(t)$, CCR consider the centered process $X(t) = \tilde{\sigma}^2(t) - \tilde{\theta}$ and a fractional integration of it:

$$X^{(d)}(t) = \int_{-\infty}^t \frac{(t-s)^{d-1}}{\Gamma(d)} X(s) ds, 0 \leq d \leq 1. \quad (3.15)$$

To facilitate the interpretation, it is worth noting that a formal integration by part on (3.15) implies that, under some convergence conditions, one can rewrite $X^{(d)}(t)$ as:

$$X^{(d)}(t) = \int_{-\infty}^t \frac{(t-s)^d}{\Gamma(d+1)} dX(s) \quad (3.16)$$

It can be seen from (3.16) that $X^{(0)}(t) = X(t)$ and $X^{(1)}(t)$ corresponds to standard integration of $X(t)$ as in (3.15). It can be shown that for $0 \leq d < 1/2$, the process $X^{(d)}(t)$ is mean square stationary centered at zero. Then, up to positivity restrictions (see CCR for a discussion), the volatility process is defined by $\sigma_t^2 = X^{(d)}(t) + \theta$ for some positive parameter θ . For $d = 0$, σ_t^2 is a standard affine volatility process:

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \gamma\sqrt{\sigma_t^2 + \tilde{\theta} - \theta}dW^X(t).$$

While $Var(\sigma_t^2) = \tilde{\theta}\gamma^2/2k$ and the autocorrelation function of σ_t^2 has an exponential rate of decay: $\rho[\sigma_{t+h}^2, \sigma_t^2] = e^{-k|h|}$.

In contrast, for $0 < d < 1/2$, the volatility process is still mean-reverting, yet will feature some long range dependence. Moreover,

$$\text{Var}(\sigma_t^2) = \frac{\tilde{\theta}\gamma^2}{k^{2d+1}} \frac{\Gamma(1-2d)\Gamma(2d)}{\Gamma(1-d)\Gamma(d)} \quad (3.17)$$

and the autocorrelation function of σ_t^2 has only an hyperbolic rate of decay for large lags: $\rho[\sigma_{t+h}^2, \sigma_t^2] \sim (kh)^{2d-1} / \Gamma(2d)$ when h tends to infinity. In other words, a positive value of d allows to introduce much more volatility persistence, not only, as usual, through a small mean reversion parameter k , but also, even more importantly, through a rate of decay which is hyperbolic instead of exponential.

This long memory model of volatility accomodates much better the volatility smile puzzle for long-term options. Indeed, it can be shown that for $0 \leq d < 1/2$:

$$\text{Var}_t \left[\frac{1}{h} \int_t^{t+h} \sigma_s^2 ds \right] \sim \frac{\gamma^2 \tilde{\theta}}{k^{2d+1}} \frac{(hk)^{2d-1}}{(d+1)\Gamma(d+1)^2} \quad (3.18)$$

when h tends to infinity. Hence, we can clearly disentangle two effects in the explanation of the volatility smile: (i) the first one, independent of the maturity is simply produced by the stochastic feature of volatility and is proportional to its unconditional variance through the term $(\gamma^2 \tilde{\theta} / k^{2d+1})$ and (ii) the second one captures the erosion of the volatility smile when the time to maturity increases. It is given by the term $(hk)^{2d-1}$ where, for a given long memory parameter d , the time to maturity h is scaled by the mean reversion parameter k .

The second effect is important to understand the phenomenon that long-term options still feature deep volatility smiles. For instance, a moderate level of long memory in the volatility process, $d = 1/4$ say, would imply that the conditional variance (3.18) would be divided by a factor of ten when the time to maturity h of the option contract is multiplied by 100. In contrast, the same factor 100 would divide the variance in the short memory case ($d = 0$).

Finally, note that the kurtosis coefficient $k(h)$ will converge towards its Gaussian limit 3 at the some speed h^{2d-1} as the conditional variance goes to zero. In other words, contrast with the short term case, the volatility smile and the excess kurtosis vanish at the same speed when time to maturity increases to infinity. Of course, long memory processes may produce non-trivial econometric issues as the past information is very slowly forgotten. However, a convenient feature of the affine fractional stochastic volatility model is that integrated volatility $\int_t^{t+h} \sigma_s^2 ds$ over the lifetime of the option and BS implied volatilities are fractionally cointegrated. Moreover, the conditional probability distribution of $\int_t^{t+h} \sigma_s^2 ds - E_t \left[\int_t^{t+h} \sigma_s^2 ds \right]$ given information available at time t only depends on the current value of the state variable $X(t)$. Therefore, all the long memory features relevant for option pricing

are encapsulated in the expected integrated volatility, and can be captured by BS implied volatilities. Note in particular that the fractional cointegration relationship justifies the widely used predicting regressions of realized volatilities on BS implied volatilities. See Bandi and Perron (2003) for an empirical illustration of fractional cointegration in this context. Indeed, it can even be shown that there is a deterministic relationship between expected integrated volatility and BS implied volatilities for very long-term options. Beyond that, all the residual variations of BS implied volatilities across moneyness (volatility smile) and across maturities (volatility term structure) are well described by the short memory dynamics of the state variables.

3.3 Pricing options based on objective parameters

A number of papers such as Andersen, Benzoni and Lund (2002) and Eraker, Johannes and Polson (2001) have derived the option pricing implications of jump-diffusion models relying only on returns data for the underlying asset. This exercise aims at evaluating the economic significance of statistical differences across models. Understanding how the various factors such as stochastic volatility, jumps in returns or jumps in volatility determine the conditional distribution as a function of time to maturity and level of volatility is equivalent to understanding how option prices change with respect to these factors. Indeed, options with different strike prices and times to maturity are affected by different attributes of the conditional distribution of returns. However, in order to price options in an arbitrage-free framework, one needs to specify a candidate state price density or to characterize the transformation from the objective measure to the risk neutral measure. In the presence of jump and stochastic volatility risks, appropriate risk compensation must be incorporated in the risk neutral dynamics. As already noted, there are potentially risk premia associated with stochastic volatility, mean jump sizes, volatility of jump sizes and jump timing. Separating the various risk premia is not an easy task. Assumptions have to be made. The crudest assumption consists in setting at zero all risk premia associated with stochastic volatility and jumps. Under this assumption, the change from the objective measure to the risk neutral measure affects only the drift of the stock index returns which is equal to the interest rate minus the dividend yield. Andersen, Benzoni and Lund (2001) and Eraker, Johannes and Polson (2001) make such an assumption and study the impact of stochastic volatility and jumps on the levels of implied volatility as well as on the shapes of the implied volatility curves.

Jumps in returns affect mainly the tails of the conditional distribution and induce excess kurtosis. As shown by Das and Sundaram (1999) among others, this effect is strongest for short maturity options since the degree of excess kurtosis in a jump model decreases with

maturity. With jump processes, the implied volatility smile flattens out very quickly. Unlike jumps, stochastic volatility affects the conditional distribution the most at longer horizons. For typical parameterizations such as a slow-moving mean reverting volatility, the term structure of kurtosis is increasing over a reasonable horizon. Eraker, Johannes and Polson (2001) produce a figure of implied volatility curves for the models SV, SVJ, SVIJ and SVCJ for four different times to maturity (2 weeks, 2 months, 6 months and 1 year). The results indicate that there are differences both in the levels of implied volatility and in the shapes of the implied volatility curves. Regarding the volatility level, the main difference between the models comes from the estimates of the spot volatility. The spot volatility estimates for the S&P 500 are 15.10%, 14.32%, 15.18% and 15.51% for SV, SVJ, SVCJ and SVIJ respectively. This translates into a level difference of almost 2 percentage points in the implied volatility for at-the-money options with one year to maturity. There are a number of noteworthy results for the shapes of the volatility curves. First, the implied volatility curves produced by the SV model are flat. Second, adding jumps in returns steepens the implied volatility curves at all maturities. With a sizable negative mean jump estimate for all the models, the implied volatility curves are downward sloping to the right and not U-shaped. Third, the addition of jumps in volatility fattens further the tails of the conditional distributions and makes the implied volatility curves steeper. Therefore, even without any risk premia, jumps and especially jumps in volatility have an important impact on option prices, which translates into term structures and cross-sections of implied volatility more consistent with data. These results are in contrast with Andersen, Benzoni and Lund (2001) who need to add risk premia to generate steep implied volatility curves. This is due mainly to the fact that their estimates for the jump parameters are small compared to Eraker, Johannes and Polson (2001). However, all studies concur in finding a flattening out of the implied volatility curves as maturity increases for all the models. Indeed, the skewness and kurtosis of the conditional distribution at longer horizons are due mainly to the volatility process and not to the jump processes.

To assess the actual quantitative importance of risk premia for option pricing, one needs to estimate these risk premia along with the parameters of the model. The option market provides us with prices which can be used, along with stock returns, to estimate these risk premia. However, to achieve this, one needs additional assumptions to characterize the form of these risk premia as well as an econometric model of option pricing errors.

4 Implied risk-neutral probabilities

Let us reconsider the fundamental pricing equation and write the price at time t of a claim paying a J_T -measurable random variable g_T at time T as:

$$\pi_t[g_T] = \frac{1}{m_t} E[g_T m_T | J_T] \quad (4.1)$$

In the context of the jump-diffusion model described in the previous section, markets are incomplete and this state-price density is not unique. For a SVJ model, Pan (2002) proposes a candidate state-price density of the following form:

$$m_t = \exp\left(-\int_0^t r_\tau d\tau\right) \exp\left(-\int_0^t \zeta_\tau dW_\tau - \frac{1}{2} \int_0^t \zeta_\tau' \zeta_\tau d\tau\right) \exp\left(\sum_{i, \tau_i \leq t} \xi_i^\pi\right) \quad (4.2)$$

where ζ represents a vector of the market prices of risk for the price and volatility shocks respectively and ξ_i^π the market price of jump risk. The market prices of risk are defined by:

$$\zeta_t^{(1)} = \eta^s \sqrt{V_t}, \quad \zeta_t^{(2)} = -\frac{1}{\sqrt{1-\rho^2}} (\rho \eta^s \frac{\eta^v}{\sigma^v}) \sqrt{V_t} \quad (4.3)$$

This specification of the market prices of risk makes the risk premia for the diffusion price shock and the volatility shock proportional to V_t and equal to $\eta^s V_t$ and $\eta^v V_t$ respectively. These forms of the risk premia have been suggested by Bates (1996, 2000) based on a log utility model for the representative investor. Jump risks are priced by the components ξ_i^π in the state-price density, assumed to be i.i.d. and Gaussian with mean μ_π and variance σ_π^2 and independent of W . The random jump sizes ξ_i^π and ξ_i^y are allowed to be correlated with a constant correlation ρ_π but are independent at different jump times.

It is more common to transform the model to write it under a risk neutral measure Q^* which is defined from a density $m_t \exp\left(\int_0^t r_\tau d\tau\right)$, yielding the following representation of the SVJ model:

$$\begin{pmatrix} dS_t/S_t \\ dV_t \end{pmatrix} = \begin{pmatrix} r_t - \eta^s V_t - \lambda_y^* \mu_y^* \\ \kappa(\alpha - V_t) + \eta^v V_t \end{pmatrix} dt + \sqrt{V_t} \begin{pmatrix} 1 & 0 \\ \rho \sigma_v & \sqrt{(1-\rho^2)} \sigma_v \end{pmatrix} dW_t^* + \begin{pmatrix} \xi^y dN_t^{Qy^*} \\ 0 \end{pmatrix} \quad (4.4)$$

The risk-neutral dynamics differ from those under the objective measure by the drift terms which incorporate the risk premia and by replacing $W_t = (W_{1t}, W_{2t})'$ by $W_t^* = (W_{1t}^*, W_{2t}^*)'$, a vector of independent standard Brownian motions under Q^* defined by:

$$W_t^* = W_t + \int_0^t \zeta_s ds, \quad 0 \leq t \leq T. \quad (4.5)$$

The jump process N^{Q^*y} has the same distribution under Q^* and Q except that $\xi^y \sim \mathcal{N}(\mu_y^*, \sigma_y^2)$, where $\mu_y^* = \mu_y + \sigma_y \sigma_\pi \rho_\pi$. This means that the model allows for a jump-size

risk. It can also allow for a jump-timing risk since the λ_y^* can be different from λ_y : $\lambda_y^* = \lambda_y \exp(\mu_\pi + \sigma_\pi^2/2)$. In Bates (2000) and Pan (2002), the jump-size intensity is made volatility dependent with one and two factors in volatility respectively.

The price of a European option expiring at T with a strike price of K is given by:

$$\pi_t = \frac{1}{m_t} E_t[m_T(S_T - K)^+]. \quad (4.6)$$

A Fourier transform-based approach is adopted to calculate this expectation, as in Heston (1993), Bates (1996, 2000), Bakshi, Cao and Chen (1997), Bakshi and Madan (2000) and Duffie, Pan and Singleton (2000). The explicit formula is given in these papers. For our purpose, let us characterize the solution as a function f :

$$\pi_t = S_t f(V_t, \vartheta, r_t, T - t, K/S_t) \quad (4.7)$$

where $\vartheta = (\kappa, \alpha, \sigma_v, \rho, \eta^s, \eta^v, \lambda_y, \lambda_y^*, \mu_y, \sigma_y, \mu_y^*)$ is the vector of model parameters. We will digress in the next subsections on the various issues raised by the estimation of such a model.

4.1 Econometric model of option pricing errors

Typically, a theoretical asset pricing model explains an observed stationary process Y_t of n asset “prices” as a known function of the current value X_t of K latent state variables and p unknown parameters θ :

$$Y_t = \{h_i[X_t, \theta]\}_{1 \leq i \leq n} \quad (4.8)$$

Note that when one loosely says asset “prices”, one should rather understand “yields” for bonds or “option premium by unit of spot price” for options on equity or any other transformation well-suited to build a n -dimensional stationary time series Y_t from the observation of time series of asset prices, likely to be non-stationary. In the context of options on equity, one may also replace (see e.g. Renault and Touzi (1996), Pastorello, Renault and Touzi (2000, 2003), Chernov and Ghysels (2000)) option prices by the corresponding Black-Scholes implied volatilities.

With respect to the most general formulation of empirical asset pricing models presented in Section 2, we focus here on a more specific approach common in the arbitrage-free asset pricing literature. First, the pricing kernel is not explicitly included in the list of latent state variables. Instead, it is defined as a known function of a collection X_t of relevant risk factors such as instantaneous risk free rate, diffusion return shocks, volatility shocks and jump events as well as a collection of risk premium parameters θ_2 that define the compensation for the various risk factors. Then, the dynamics of the latent risk factors

X_t only identify a set θ_1 of unknown “statistical” parameters while the risk premium parameters θ_2 must be added to define the complete vector θ of structural parameters of interest for asset pricing $\theta = [\theta'_1, \theta'_2]'$.

For empirical option pricing on equity, the above approach is typically the one followed by Heston (1993), Bates (2000), Chernov and Ghysels (2000), and Pan(2002) among others. For term structure modelling, this approach is particularly well-suited to capture through K explanatory latent factors of the yield curve the relationships between n observed yields in a cross-section. A large strand of literature, initiated in particular by Chen and Scott (1993), Pearson and Sun (1994) and Duan (1994), uses this indirect empirical modelling of bond yields through underlying latent factors. In contrast, explicit dynamic modelling of the joint stochastic process of asset returns and pricing kernel can be found in the consumption-based equilibrium asset pricing literature (see e.g. Aït-Sahalia and Lo (2000), Jackwerth (2000), Rosenberg and Engle (2002) for applications to option pricing) or, in an even more general way in Constantinides (1992) and Garcia, Luger and Renault (2003).

The simplest approach to estimating a K factor model is to select $n = K$ asset prices in the cross-section and to exploit the one-to-one relationship between prices and factors to get either the exact likelihood (Chen and Scott (1993), Pearson and Sun (1994), Duan (1994)) or an expansion of it (Aït-Sahalia and Kimmel (2002)) or implied moments (Pan(2002)) or a simulated score (Dai and Singleton(2000)). This approach neglects potentially useful information conveyed by a number of observed related prices in the cross-section. For instance Pan (2002) estimates a stochastic volatility model for S&P 500 index option pricing from the joint time series of the index and one near-the-money short maturity option. One option price is sufficient to obtain a one-to-one relationship with the volatility factor, yet (see e.g. Dumas, Fleming and Whaley (1998)), the number of fairly liquid option prices on S&P 500 observed at any given date may be ten or even more across maturity and moneyness. Similarly, while common models of the yield curve involve $K = 1, 2$ or 3 factors, the number n of available maturities in the cross-section is typically thirty or more.

As emphasized by Renault (1997), when the number n of observed asset prices is larger than the number K of latent state variables, some stochastic singularity emerges and statistical estimation theory becomes irrelevant. If one takes the asset pricing model at face value, some parameters can be computed exactly. For example, in the Black-Scholes model without any latent state variable, observing the price of one option will be enough to compute exactly the volatility of the process. In the case of stochastic volatility models, one can recover the exact value of the current state of the variance process by matching observed prices with the pricing formulas after elimination of unknown parameters. Different option prices imply, however, different values for the current state of the variance process. This

fundamental inconsistency can be resolved either by (1) increasing ad infinitum the number of state variables and match perfectly the observed paths or cross-sections of option prices (this nonparametric approach is in the spirit of the implied binomial tree methodology of Rubinstein (1994) describes in section 5 below) or (2) accepting that the pricing formulas are approximations and that the observed price is the price given by the formula plus an error term. The presence of this error term is not difficult to justify by simply recognizing that any model is intrinsically misspecified whether it is in its assumptions about the stochastic process followed by the underlying or in its simplistic description of market structure (e.g. abstracting from microstructure effects and other market frictions). Therefore, the empirical specification of the asset pricing model (4.8) will be replaced by:

$$\begin{aligned} Z_t &= (Y_{it})_{1 \leq i \leq K} = h[X_t, \theta] = [h_i(X_t, \theta)]_{1 \leq i \leq K} \\ V_t &= (Y_{it})_{K+1 \leq i \leq n} = e[X_t, \theta] + u_t = [h_i(X_t, \theta)]_{K+1 \leq i \leq n} + [u_{it}]_{K+1 \leq i \leq n}. \end{aligned} \quad (4.9)$$

Note that we consider at this stage that the n assets prices have been relabelled in order to get zero pricing errors for the K first ones while the remaining $(n - K)$ differ from their theoretical values by error terms u_{it} . Hence, we do not really maintain the arbitrary assumption that exactly K prices coincide with their theoretical values while error terms may be added to the other ones. Equation (4.9) is interpreted as saying that, since the structural model already involves K latent factors, there is no reason to introduce more than $(n - K)$ error terms, while at least K independent linear combinations should be observed without error. Note that such a specification assumes a priori which K prices (or the K linear combinations of prices) are observed without error. This is mainly an empirical question.

Let us first set the stage for inference on (4.9) in the context of maximum likelihood-based inference strategies. A maintained assumption will be that the error terms u_{it} have a zero unconditional mean and that the first K equations provide a one-to-one relationship between the vector Z_t of the K prices observed without error and the vector X_t of structural state variables:

$$Z_t = (Y_{it})_{1 \leq i \leq K} = h[X_t, \theta] \Leftrightarrow X_t = h^{-1}[Z_t, \theta] \quad (4.10)$$

4.2 Maximum Likelihood based inference

In order to present a variety of likelihood-based inference strategies, we follow the presentation of implied-state maximum likelihood as first proposed by Renault and Touzi (1996) and Renault (1997). Pastorello, Patilea and Renault (2003) encompasses a larger set of implied state methodologies under the name of implied-state backfitting.

The conditional likelihood associated to a data set $\{Y_t, t = 1, \dots, T\}$ (and an initial conditioning value Y_0) must be derived, through the Jacobian formula, from the latent conditional likelihood associated with the “latent data” set $\{Y_t^*, t = 1, \dots, T\}$ produced by the latent realizations of a Markov process Y^* obtained as a one-to-one function of Y :

$$Y_t = g[Y_t^*, \theta] \Leftrightarrow Y_t^* = g^{-1}[Y_t, \theta] \quad (4.11)$$

Typically, (4.11) must be defined by n equations, with $(n - K)$ equations that complete the K equations (4.10). A natural approach would be to define the state vector Y_t^* by augmenting the vector X_t of K structural factors with the vector u_t of $(n - K)$ error terms. However, an alternative approach is better suited for two reasons.

First, the parameters η that would define the probability distribution of the error term u_t are not the focus of interest. While their consistent estimation may be useful for improving accuracy of the parameters of interest θ , we want to ensure that even if η is not consistently estimated, we obtain a consistent estimator of θ . Typically, in case of Gaussian errors, the vector of nuisance parameters η consists of the unconditional covariance matrix - of the $(n - K)$ error terms u_t and possibly the parameters defining the conditional mean and variance dynamics. The mere fact that these error terms are added ex post and not rationalized within a structural asset pricing model with additional state variables implies that we have no structural information about their dynamics. Since from (4.9) we note that the estimation of the dynamics of the error terms may contaminate the estimation of the dynamics of the structural factors, it is important to define a procedure that focuses only on the structural parameters θ and not on the augmented vector (θ, η) .

Second, the implied-state identification condition for θ would be problematic if we defined the latent state vector Y_t^* as $Y_t^* = (X_t, u_t)$. The empirical asset pricing model (4.9) provides a one-to-one relationship between observed prices Y_t and latent variables (X_t, u_t) but the risk premium parameters θ_2 are only identified by the relationship itself and not by the probability distribution of the latent process (X_t, u_t) . However, the philosophy of the implied-state methodology is precisely to assume that the latent model (the transition equation of the state variables) carries more information about the unknown parameters of interest than their occurrence in the measurement equation. To materialize this, a better strategy is to define the latent vector Y_t^* and the associated function $g[Y_t^*, \theta]$ as:

$$Y_t^* = [X_t', V_t']', Y_t = [Z_t', V_t']' \quad \text{with:} \quad Y_t = g[X_t, V_t, \theta] = [h'(X_t, \theta), V_t']'. \quad (4.12)$$

Note that $(n - K)$ among the n so-called latent variables Y_t^* are actually observed, yet this is not a reason for not applying the general implied-state methodology. In this context, the transition density function of the Markov process Y_t^* :

$$l[Y_t^* | Y_{t-1}^*] = l[X_t | Y_{t-1}^*] l[V_t | X_t, Y_{t-1}^*] \quad (4.13)$$

will be specified under the maintained assumption that error terms do not cause structural factors, neither in the Granger sense nor instantaneously. This assumption is natural since, if one imagines its violation, one implicitly endows the error terms with some structural interpretation. Then, by the no-Granger causality assumption:

$$l[X_t | Y_{t-1}^*] = l[X_t | X_{t-1}] = l[X_t | X_{t-1}, \theta_1] \quad (4.14)$$

where the last expression emphasizes the dependence of the density on the value of the unknown parameters only through θ_1 . Since we maintain the assumption that all the structural content of the model is captured by the factors X_t , we do not really want to specify the dynamics of the error terms and therefore carry out inference about the structural parameters through a latent quasi-likelihood, written as the likelihood of a latent model where the error terms would be i.i.d. Gaussian with a covariance matrix specified as a function - (η) :

$$l[u_t | Y_{t-1}^*, \eta] = l[u_t | \eta] = (2\pi)^{-(n-K)/2} [\det(\eta)]^{-1/2} \exp[-\frac{1}{2} u_t' \eta^{-1}(\eta) u_t]. \quad (4.15)$$

The above expression is well-suited only if the scale Y_t used to measure asset prices is consistent with conditional normality like for instance log-returns or log-implied volatilities. Moreover, it should also be noted that the quasi-likelihood may differ from the true likelihood and that we just want to plug it into (4.13) to estimate consistently the structural parameters of interest θ (see Renault (1997) for further discussion).

Starting with the nuisance parameters estimator η_T and a corresponding estimator $\eta_T^{-1}(\eta_T)$, and using (4.13) the latent criterion for extremum estimation of the structural parameters θ is obtained as:

$$Q_T^*(\theta) = \sum_{t=2}^T \log l[X_t | X_{t-1}, \theta_1] - \frac{1}{2} \sum_{t=1}^T [V_t - e(X_t, \theta)]' \eta_T^{-1} [V_t - e(X_t, \theta)]. \quad (4.16)$$

Up to recursive refinements, the backfitting (or iterative implied-state) methodology amounts defining a sequence $\theta^{(p)}$ of estimators in the following way:

- Start from an estimator $\theta^{(1)}$ provided by a quick procedure.
- For $\theta^{(p)}$ given, replace in (4.16) the unknown factor values X_t by $X_t(\theta^{(p)}) = h^{-1}[Z_t, \theta^{(p)}]$. This defines a sample based criterion $Q_T(\theta, \theta^{(p)})$.
- Compute the estimator $\theta^{(p+1)}$ as $\arg \max_{\theta} Q_T(\theta, \theta^{(p)})$.

Pastorello, Patilea and Renault (2003) provide further details and improvements upon the above described procedure. To conclude, let us compare with the maximum likelihood based competitors also well-suited for empirical asset pricing models with latent factors.

A first competitor is the Kalman filter based quasi-maximum likelihood. The most popular strategy is to introduce n error terms instead of $(n - K)$. This was first proposed in the context of affine models of the yield curve by Duan and Simonato (1999) and systematically developed by De Jong (2000). Severe nonlinearities or non-normality of the structural model are likely to alter the validity of the Kalman filter. Generally speaking, the Kalman filter should not be used for highly nonlinear models and the backfitting filtering strategy should be much better suited. However, in the context of return dynamics that are not too far from linearity as in the case of affine models of the yield curve, the two approaches may be close competitors. Typically, the backfitting approach seeks to get more efficient estimators and filters via the specification of an exact nonlinear relationship between prices and factors with K zero error terms.

Another quasi-maximum likelihood approach for factor models of the yield curve has been applied by Fisher and Gilles (1996) and Duffee (2002). The difficult part of the latent log-likelihood (4.16) is the transition density function of the structural factors X_t . This function is in general produced by a continuous time model and may be hard to compute or simply unknown. However, consistent (albeit inefficient) estimates can still be obtained by substituting the true theoretical transition density with a Gaussian one, provided that the first two conditional moments of X_t are correctly specified. Besides its potential inefficiency, this alternative QML approach also suffers from a potential misspecification bias due to a nonlinear mapping g between the latent variables and the observables. In such a case, the Jacobian formula applied to a latent Gaussian quasi-likelihood may not yield a correct quasi-likelihood for observables. This drawback is not detrimental in the case of affine (Fisher and Gilles (1996)) or essentially affine (Duffee (2002)) term structure models but would be an issue in the case of option prices on equity with stochastic volatility.

As far as efficiency is concerned several remarks are in order. First, contrary to a common belief, the fact that one can invert any vector of n asset prices into the n state variables and use the implied state variables in the estimation does not mean that one can proceed *as if* the state variables were directly observable. The crucial point is that the one-to-one relationship (4.11) between the latent variables Y^* and the observable variables Y depends upon the unknown parameters θ . Therefore, one does not know whether the Cramer-Rao bound $(I^*)^{-1}$ for efficient estimation associated with the hypothetical observation of Y^* would be smaller or larger than the Cramer-Rao bound $(I)^{-1}$ associated with the actual observation Y . The backfitting strategy should not give the fallacious

feeling that the Cramer-Rao bound associated with the maximization of the log-likelihood $\sum_{t=1}^T \log L [Y_t^* | Y_{t-1}^*, \theta]$ has been reached. This maximization is actually infeasible and the backfitting iterative scheme is based on the sequence:

$$\theta^{(p+1)} = \arg \max_{\theta} \sum_{t=1}^T \log L \left[g^{-1} \left(Y_t, \theta^{(p)} \right) | g^{-1} \left(Y_{t-1}, \theta^{(p)} \right), \theta \right]$$

As shown in Pastorello, Patilea and Renault (2003), the cost of this necessary iteration is the scaling of the Cramer-Rao bound $(I^*)^{-1}$ by a matrix factor closer to an identity matrix when the mapping $\theta^{(p)} \rightarrow \theta^{(p+1)}$ is more strongly contracting.

If one wants to avoid such iterations and directly maximize the actual log-likelihood to reach the Cramer-Rao bound I^{-1} , one should not maximize:

$$\sum_{t=1}^T \log L \left[g^{-1} (Y_t, \theta) | g^{-1} (Y_{t-1}, \theta), \theta \right] \quad (4.17)$$

but rather the likelihood:

$$\sum_{t=1}^T \log L \left[g^{-1} (Y_t, \theta) | g^{-1} (Y_{t-1}, \theta), \theta \right] + \sum_{t=1}^T \log | Jg^{-1} (Y_t, \theta) | \quad (4.18)$$

where $|Jg^{-1} (Y_t, \theta)|$ denotes the absolute value of the Jacobian of the transformation g^{-1} . This can be done in some cases but will often be involved for several reasons.

First, the function g is provided by the asset pricing model. It is in general highly nonlinear and even not available in a closed form formula. Computing the Jacobian matrix can then be cumbersome. Second, and more importantly, the direct maximization of (4.18) will lead to a maximizer θ which should simultaneously meet two requirements. Namely, it has to give a large value to the latent likelihood. Yet, on the other hand, θ will tend to be chosen in order to select the most likely implied-state values $g^{-1} (Y_t, \theta)$. In many circumstances, such a selection strategy is problematic. For instance, Pastorello, Patilea and Renault (2003) observe that when applying the Aït-Sahalia (2003) likelihood expansions for affine-type diffusion processes, this will perversely push $g^{-1} (Y_{t-1}, \theta)$ towards the frontier of the domain where the likelihood (as provided by its expansion) is infinite. This is the reason why one may prefer to perform the backfitting strategy for likelihood maximization rather than directly maximizing the log-likelihood (4.18).

To conclude it should be noted that Indirect Inference and Efficient Method of Moments (EMM) are often presented as appealing alternatives to maximum likelihood, precisely when the likelihood function becomes untractable due to some unobserved state variables. Since the chapter by Gallant and Tauchen in this handbook is devoted to these techniques, we just sketch here some specific applications for option pricing.

Pastorello, Renault and Touzi (2003) propose to avoid the backfitting iteration by simply using BS implied volatilities as proxies of implied states in a one-factor stochastic volatility model. Thanks to the matching of estimated parameter or fitted-score vectors on simulated data, the indirect inference principle (see Gouriéroux, Monfort and Renault (1993)) will correct for the misspecification bias due to the use of BS implied volatilities as proxies of actual spot volatilities which are unobserved. The main drawback of this approach is that, while a fully parametric model is needed for the purpose of simulation, nobody knows the efficiency loss due to the use of an auxiliary model (here, the model on BS implied volatilities) to simplify the likelihood.

By matching a SemiNonParametric (SNP) score generator, EMM aims at correcting for this efficiency loss. The EMM procedure allows estimating the model parameters under both objective and risk-neutral probability measures if one uses implied volatilities and the underlying asset data jointly. Time series of the underlying asset provide estimators under the objective probability measure while risk-neutral parameters can be retrieved from options. Chernov and Ghysels (2000) adopt the Heston model, which has a closed-form option pricing formula, and compare univariate and multivariate models in terms of pricing and hedging performance. An extension of the SNP/EMM methodology introduced in Gallant and Tauchen (1998) which allows one to filter spot volatilities via reprojection, i.e. compute the expected value of the latent volatility process using a SNP density conditioned on the observable processes such as returns and/or options data. The results in Chernov and Ghysels (2000) show that the univariate approach only involving options by and large dominates. A by-product of this finding is that they uncover a remarkably simple volatility extraction filter based on a polynomial lag structure of implied volatilities. The bivariate approach appears useful when the information from the cash market provides support via the conditional kurtosis to price options. This is the case for some long-term options. Another solution to the efficiency problem may be provided by Markov Chain Monte Carlo techniques as described by Johannes and Polson in this handbook.

4.3 Implied-State GMM

Using the explicitly known moment-generating function of return and volatility in an affine model, Pan (2003) also advocates an implied-state methodology to focus directly on the joint dynamics of the state variables rather than the market observables which could be highly nonlinear functions of state variables. In this respect, the approach still belongs to the general class of backfitting methodologies as studied by Pastorello, Patilea and Renault (2003) but the convenience of the GMM setting introduces some additional simplifications. The basic idea is to start from conditional moment restrictions which would provide a

feasible GMM if the latent variable Y^* would be observed:

$$E [\Psi (Y_t^*, \theta) | Y_{t-1}^*] = 0 \quad (4.19)$$

Following Hansen (1985) , Pan (2003) uses the optimal instrument matrix provided by:

$$M_{t-1}(\theta) = E \left[\frac{\partial \Psi'}{\partial \theta} (Y_t^*, \theta) | Y_{t-1}^* \right] (Var [\Psi (Y_t^*, \theta) | Y_{t-1}^*])^{-1}.$$

Then, one would like to work with the just identified unconditional moment restrictions:

$$E [M_{t-1}(\theta) \Psi (Y_t^*, \theta)] = 0$$

and to look for the estimator $\hat{\theta}_T$ solution of:

$$\frac{1}{T} \sum_{t=1}^T M_{t-1}(\hat{\theta}_T) \Psi (Y_t^*, \hat{\theta}_T) = 0 \quad (4.20)$$

this estimator is unfeasible since Y_t^* is not observed. Two strategies can be considered. The implied-state backfitting of Pastorello, Patilea and Renault (2003) amounts to replacing every occurrence of Y^* in $M_{t-1}(\theta)$ and $\Psi (Y_t^*, \theta)$ by $g^{-1}(Y_t, \theta^{(p)})$ where $\theta^{(p)}$ comes from a previous step estimation. Insofar as such iterations converge, they will converge towards the second strategy: Pan's (2003) IS-GMM estimator. The latter consists of directly solving (4.20) when Y_t^* is replaced by $g^{-1}(Y_t, \theta)$. Then, the unknown θ appears not only in the occurrences of θ in $M_{t-1}(\theta)$ and $\Psi (Y_t^*, \theta)$ but also inside any occurrence of $Y_t^* = g^{-1}(Y_t, \theta)$.

In contrast, Pastorello, Patilea and Renault (2003) define a number $p(T)$ of iterations (as a function of the number T of observations) such that the backfitting estimator $\theta^{p(T)+1}$ is asymptotically equivalent to the Pan (2003) IS-GMM estimator. The choice between the two strategies is just a matter of computational convenience, depending on whether one considers that the backfitting iterations simplify or not the solution of the IS-GMM fixed point problem.

Moreover, as noted by Pan (2003) in her discussion of Pastorello, Patilea and Renault (2003), there is a case where IS-GMM may work while IS-backfitting does not work. This is the case where θ would not be fully identified from state variables dynamics Y^* , for instance due to some risk premium parameters which do not appear in the factor dynamics. Even in such a case, one may hope that IS-GMM still identifies θ . It is however worth reminding that, when as in subsection 4.2 there are more observed prices than latent state variables, some error terms are added and the vector Y^* includes some observed asset prices which do identify the risk premium parameters. Then, implied-state backfitting works.

As in the implied-state likelihood methodology of subsection 4.2, efficiency is not guaranteed by this kind of implied-state approaches. In the context of (4.19), semiparametric

efficiency would involve the computation of optimal instruments for the conditional moment restrictions:

$$E [\Psi (g^{-1} (Y_t, \theta), \theta) | Y_{t-1}] = 0 \quad (4.21)$$

Then, the Jacobian matrix of the moment conditions needed for computing optimal instruments involves differentiation with respect to the two occurrences of θ in (4.21) and not only the second one - as acknowledged by Pan (2003), we sacrifice efficiency and gain analytical tractability by ignoring the dependence of Y_t^* on θ . As already mentioned in the likelihood case, it may indeed be challenging to look simultaneously for the “optimal” value of the implied states and for the best fit in the latent model. However, while backfitting was really needed in the likelihood case because, otherwise, forgetting the Jacobian term may imply inconsistency of the estimator, there is no such consistency problem with GMM. The only consequence of not taking into account the complete Jacobian term is that the efficiency of the optimal instrument scheme may be “limited”, as acknowledged by Pan (2003). Indeed, since the two estimators IS-GMM and IS-backfitting are asymptotically equivalent, this limit to efficiency is tightly related to the contracting feature of the backfitting correspondence. The more contracting it is, the smaller is the efficiency loss.

Garcia, Lewis and Renault (2001) propose an estimation procedure which uses both option prices and high-frequency spot price feeds to estimate jointly the objective and risk-neutral parameters of stochastic volatility models. This procedure is based on series expansions of option prices and implied volatilities and on a method-of-moment estimation that uses analytical expressions for the moments of the integrated volatility. In a stochastic volatility model, with or without correlation, the option pricing formula involves the computation of a conditional expectation of a highly nonlinear integral function of the volatility process. To simplify this computation, the authors propose to use an expansion of the option pricing formula in the neighborhood of $\sigma_V = 0$, as in Lewis[30], which corresponds to the Black-Scholes deterministic volatility case. The coefficients of this expansion are well-defined functions of the conditional moments of the joint distribution of the underlying asset returns and integrated volatilities, which are also derived analytically. These analytical expansions allow to compute very quickly implied volatilities which are functions of the parameters of the processes and of the risk premia. A two-step GMM approach using intraday returns for computing approximate integrated volatilities (the objective part of the estimation) and option prices for computing implied volatilities (the risk-neutral part of the estimation) allow to recover the volatility risk premia λ . The main attractive feature of this method is its simplicity once analytical expressions for the various conditional moments of interest are available. The great advantage of the affine diffusion model is precisely to allow an analytical treatment of the conditional moments of interest. Eraker (2001) applies

a Markov chain Monte Carlo based approach to joint time-series data on spot and options also for a jump-diffusion model.

5 Nonparametric approaches

The financial theoretical models of the previous sections are based on parametric dynamic processes for stock returns. Despite the great deal of complexity put into these processes to capture the features of the data, they remain usually misspecified. Therefore, nonparametric methods, which are so called model-free and make minimal assumptions about the underlying asset price process, appear as a promising tool to apply in the context of derivative pricing. Moreover, these methods are well adapted to the financial problems at hand since the quantities of interest are functions, whether it is the risk neutral distribution or state price density (SPD), the distribution function for hedging or else the value-at-risk quantile function of the conditional distribution of returns.

Nonparametric methods have been applied to all the above-mentioned financial problems of interest. We will discuss in this section how nonparametric methods can be used to recover a pricing function, a hedging ratio and a risk-neutral distribution. As a way to make the transition between the parametric and nonparametric approaches, we will first consider a semiparametric approach proposed by Aït-Sahalia and Lo (1998) and Gouriéroux, Monfort and Tenreiro (1994). The main idea is to recover risk-neutral distribution using a nonparametric deterministic volatility function while maintaining that the derivative pricing function is given by the parametric Black-Scholes formula. Next, we will see a maximum entropy approach proposed by Buchen and Kelly (1996) and Stutzer (1996) to recover a risk-neutral distribution from a set of option and stock prices, as well as the implied binomial tree method of Derman and Kani (1994) or Rubinstein (1994). Third, we will survey the purely nonparametric approaches such as kernel-based techniques or learning networks used to estimate an option pricing function and recover the other quantities of interest with option price data. We will underline several potential problems associated with these purely nonparametric approaches such as negative risk neutral probabilities and argue following Garcia and Gençay (2000) and Aït-Sahalia and Duarte (2003) that imposing weak constraints on the shape and properties of the pricing function can improve the performance of the statistical model in several dimensions. Last, we will describe how to recover preferences from the estimates of the state price density as recently proposed by Jackwerth (2000), Aït-Sahalia and Lo (2000) and Rosenberg and Engle (2002).

Most empirical studies of option pricing focus on European contracts. In contrast, American options, while actively traded and very liquid in some cases (such as for example

the S&P 100 based contracts) have been avoided in order to circumvent early exercise premia and boundaries. It is worth noting that nonparametric methods are particularly suited to handle American-type options. Broadie, Detemple, Ghysels and Torrès (2000a,b) use nonparametric techniques to estimate pricing functions as well as early exercise boundaries for American options.

5.1 Semiparametric approaches to derivative pricing

One of the reasons why option price data do not conform to the Black-Scholes model is that volatility is not constant. One can still maintain the assumption of a one-factor diffusion process but make the diffusion coefficient a deterministic function of the available information such as the exercise price, the underlying price and the time to maturity. While Shimko (1993) proposed a polynomial function of these variables for the volatility, Aït-Sahalia and Lo (1998) modeled the volatility function using kernel methods. The strategy is to construct a nonparametric estimator of the expectation of volatility given the information available on the underlying stock price S_t (or the futures price $F_{t,\tau_i} = S_t e^{(r_{t,\tau} - \delta_{t,\tau})\tau}$, with r and δ the interest rate and the dividend rate), the exercise price X_i and the time to maturity τ_i associated with n traded options:

$$\widehat{\sigma}(F_{t,\tau}, X, \tau) = \frac{\sum_{i=1}^n k_F \left(\frac{F_{t,\tau} - F_{t,\tau_i}}{h_F} \right) k_X \left(\frac{X - X_i}{h_X} \right) k_\tau \left(\frac{\tau - \tau_i}{h_\tau} \right) \sigma_i}{\sum_{i=1}^n k_F \left(\frac{F_{t,\tau} - F_{t,\tau_i}}{h_F} \right) k_X \left(\frac{X - X_i}{h_X} \right) k_\tau \left(\frac{\tau - \tau_i}{h_\tau} \right)} \quad (5.1)$$

where the multivariate kernel is formed as a product of three univariate kernels k_F , k_X and k_τ , each with their own bandwidth value, with respect to the three variables of interest, and where σ_i is the Black-Scholes volatility implied by the observed price of option i . A call pricing function can then be estimated as:

$$\widehat{\pi}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = \pi_{BS}(F_{t,\tau}, X, \tau, r_{t,\tau}, \widehat{\sigma}(F_{t,\tau}, X, \tau)) \quad (5.2)$$

From this function, one can also obtain estimators for the option's delta and the state price density by taking the appropriate partial derivatives:

$$\widehat{\Delta}_t = \frac{\partial \widehat{\pi}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial S_t} \quad (5.3)$$

$$\widehat{f}_t^*(S_T) = e^{r_{t,\tau}\tau} \left[\frac{\partial^2 \widehat{\pi}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial X^2} \right]_{|X=S_T} \quad (5.4)$$

The fact that the SPD can be recovered from option prices is a result of Banz and Miller (1978) and Breeden and Litzenberger (1978). They show that the SPD is the second

derivative, normalized to have an integral of one, of a call option-pricing formula with respect to the strike price. Of course, in nonparametric methods, higher order derivatives are estimated at a slower rate of convergence. This is known as the curse of differentiation. However, in a simulation framework based on a Black-Scholes model, Aït-Sahalia and Lo (1998) show that the estimation errors for all nonparametric quantities (option price, option delta and SPD) remain within 1 percent of their theoretical counterparts. Aït-Sahalia and Lo (1998) apply their method to the estimation of these quantities for S&P 500 European option price data. Their sample period is January 4, 1993 to December 31, 1993. Their nonparametric estimator of volatility $\hat{\sigma}(F_{t,\tau}, X, \tau)$ generates a strongly asymmetric volatility smile with respect to moneyness, confirming several sources of evidence according to which out-of-the-money put prices have been consistently bid up since the crash of 1987. The shape of the smile changes as time to maturity increases. The one-month smile is the steepest: volatility curves are flatter for longer times to maturity. Strong skewness and kurtosis effects are present in the semiparametrically estimated SPDs. The (negative) skewness in returns diminishes as the maturity increases, while the contrary is obtained for the positive kurtosis.

An important aspect of this methodology is the unconditional nature of the estimated quantities. This approach assumes that these quantities are fixed functions of a vector of state variables over the estimation period and has to be contrasted with methods such as the implied binomial tree which rely on current cross-sections of option prices to infer the SPD.

5.2 Canonical valuation and implied binomial trees

The semiparametric approach we just described still depends on the assumptions that there is just one state variable and that it is governed by an Itô process. But, as we have extensively documented in the previous sections, there is evidence of jumps and stochastic volatility in the underlying stock index process. Therefore, we need procedures which extract the asset probability distribution directly from observed prices either on the asset itself or on options written on the asset. We will describe first a procedure based on the maximum entropy principle which has been proposed by Buchen and Kelly (1996) and by Stutzer (1996) and contrast it with the binomial tree approach of Rubinstein (1994). Both the former procedure, called canonical valuation by Stutzer (1996), and the latter assume that a set of financial instruments are priced correctly and can be used to recover the asset distribution from an expectation pricing model. As we will see, the differences between the two approaches lie in the choice of objective function.

5.2.1 Canonical valuation

We want to estimate the payoff distribution of the underlying asset at expiration of the option from a set of available asset and option prices. To illustrate the method, we will take the simplest case of one underlying asset that does not pay dividends, which will be used to price derivative securities expiring T periods from now. Following Stutzer (1996), we start by using only returns on the underlying asset, then we will add price information coming from options. The method involves three steps. First, starting with the current price S and a historical time series $S(t), t = -1, -2, \dots, -H$, one can construct a rolling historical time series of T -period gross returns:

$$R(-h) = \frac{S(-h)}{S(-h-T)}, \quad h = 1, 2, \dots, H-T. \quad (5.5)$$

Then, the asset's price T -periods from now is:

$$S^h = SR(-h), \quad h = 1, 2, \dots, H-T. \quad (5.6)$$

In other words, the past realized returns are used to construct possible prices at T for the underlying asset, each with estimated objective (actual) probability $\hat{p}(h) = \frac{1}{H-T}$. The problem is to find the risk neutral probabilities p^* which are the closest to the empirical probabilities \hat{p} in the Kullback-Leibler Information Criterion (KLIC) distance:

$$\hat{p}^* = \arg \min_{p^*(h) > 0, \sum_h p^*(h) = 1} I(p^*, \hat{p}) = \sum_{h=1}^{H-T} \hat{p}(h) \log \frac{p^*(h)}{\hat{p}(h)} \quad (5.7)$$

and which obey the non-arbitrage economic constraint (assuming a constant interest rate):

$$\sum_{h=1}^{H-T} \frac{R(-h)}{r^T} \frac{p^*(h)}{\hat{p}(h)} \hat{p}(h) = 1 \quad (5.8)$$

The solution to this problem is:

$$\hat{p}^*(h) = \frac{\exp \left[\gamma^* \frac{R(-h)}{r^T} \right]}{\sum_h \exp \left[\gamma^* \frac{R(-h)}{r^T} \right]}, \quad h = 1, 2, \dots, H-T \quad (5.9)$$

where γ^* is found as the arg min of $\sum_h \exp \left[\gamma \left(\frac{R(-h)}{r^T} - 1 \right) \right]$. The last step is of course to use the $p^*(h)$ to value say a call option with exercise price X expiring at T by:

$$C = \sum_h \frac{\max[SR(-h) - X, 0]}{r^T} \hat{p}^*(h) \quad (5.10)$$

The methodology is easily extendable to compute risk-neutral probabilities based on more than one underlying asset. One can also ensure that a subset of derivative securities is correctly priced at a particular date. For example, if we wanted to ensure the correct pricing of a particular call option expiring at date T with exercise price X and market price C , we would need to find a vector γ^* of two elements (γ_1^*, γ_2^*) such that:

$$[\gamma_1^*, \gamma_2^*] = \arg \min_{\gamma} \sum_h \exp \left[\gamma_1 \left(\frac{R(-h)}{r^T} - 1 \right) + \gamma_2 \left(\frac{\max[SR(-h) - X, 0]}{r^T} - C \right) \right] \quad (5.11)$$

These values would then be used to compute the estimated risk-neutral probabilities as:

$$\hat{p}^*(h) = \frac{\exp \left[\gamma_1^* \left(\frac{R(-h)}{r^T} \right) + \gamma_2^* \left(\frac{\max[SR(-h) - X, 0]}{r^T} \right) \right]}{\sum_h \exp \left[\gamma_1^* \left(\frac{R(-h)}{r^T} \right) + \gamma_2^* \left(\frac{\max[SR(-h) - X, 0]}{r^T} \right) \right]}, h = 1, 2, \dots, H - T. \quad (5.12)$$

Stutzer (1996) uses this methodology to evaluate the impact of the 1987 crash on the risk-neutral probabilities first using only S&P 500 returns. As many other papers, he finds that the left-hand tail of the canonical distribution estimated with data including the crash extends further than the tail of the distribution without crash data. A useful diagnostic test is the skewness premium proposed by Bates (1991). It is the percentage difference of the price of a call that is x percent (> 0) out-of-the-money (relative to the current forward index value for delivery at the option's expiration) to the price of a put that is also x percent out-of-the-money. The canonical valuation passes this diagnostic test for options in the three to six month range for $x > 0.02$ using only the historical data on S&P 500 returns starting in 1987.

5.2.2 Implied binomial trees

The implied binomial tree methodology proposed by Rubinstein (1994) aims also at recovering the risk-neutral probabilities that will come closest to pricing correctly a set of derivative securities at a given date. The idea is to start with a prior guess for the risk-neutral probabilities say \tilde{p}_j^* and find the risk-neutral probabilities p_j^* associated with the binomial terminal stock price S_T that are the closest to \tilde{p}_j^* but price correctly an existing set of options and the underlying stock. The risk-neutral probabilities p_j^* are solutions to the following program:

$$\min_{p_j^*} \sum_j (p_j^* - \tilde{p}_j^*)^2 \text{ subject to} \quad (5.13)$$

$$\begin{aligned} \sum_j p_j^* &= 1 \text{ and } p_j^* \geq 0 \text{ for } j = 0, \dots, n \\ S^b &\leq S \leq S^a \text{ where } S = (\sum_j p_j^* S_j) / r^\tau \\ C_i^b &\leq C_i \leq C_i^a \text{ where } C_i = (\sum_j p_j^* \max[0, S_j - K_i]) / r^\tau \text{ for } i = 1, \dots, m \end{aligned}$$

where j indexes the ending binomial nodes from lowest to highest, S_j is the underlying asset prices (supposing no dividends) at the end of a standard binomial tree, S^b and S^a are the current observed bid and ask underlying asset price and C_i^a and C_i^b the current observed bid and ask call option prices with striking price K_i , r is the observed annualized riskless return and τ is the time to expiration.

The two methods are therefore very similar, the main difference being the distance criterion used. While the maximum entropy criterion appears the best one from a theoretical point of view, since it selects the posterior that has the highest probability of being correct given the prior, there does not seem to be a statistical criterion behind the quadratic distance. A goodness of fit criterion given by $\min_{p_j^*} \sum_j (p_j^* - \tilde{p}_j^*)^2 / \tilde{p}_j^*$ seems more natural and is closer to the criterion used by Hansen and Jagannathan (1997) (see subsection 5.2.3). The goodness of fit criterion places greater weight on states with lower probabilities. Another criterion used is to maximize smoothness $\sum_j (p_{j-1}^* - 2p_j^* + p_{j+1}^*)^2$, as in Jackwerth and Rubinstein (1996) in order to avoid the overfitting associated with exactly pricing the options. With the smoothness criterion, there is a trade-off between smoothing the risk-neutral distribution and explaining the option prices. All these approaches will produce risk-neutral distributions that have much more weight in the lower left tail than the lognormal case after the 1987 crash, but they will distribute the probability differently in the tail.

5.2.3 Misspecification in implied binomial trees and stochastic discount factor models

In this section we argue that to fit the pricing kernel, a more relevant measure of closeness is a distance between pricing kernels and not between risk neutral probabilities, which are only a tool for representing these pricing kernels. In other words, one might want to look for the SDF m_{t+1}^* defined by:

$$m_{i,t+1}^* = B(t, t+1) \left(\frac{p_{it}^*}{p_{it}} \right), \quad i = 0, 1, \dots, I+1$$

which is closest to a prior SDF:

$$\tilde{m}_{i,t+1}^* = B(t, t+1) \left(\frac{p_{it}^*}{p_{it}} \right)$$

For instance, according to Hansen and Jagannathan (1997), one can choose the L^2 -distance between SDFs:

$$E_t[m_{t+1}^* - \tilde{m}_{t+1}^*]^2 = B^2(t, t+1) \sum_{i=0}^{I+1} \frac{1}{p_{it}} (p_{it}^* - p_{it}^*)^2. \quad (5.14)$$

Therefore, the Hansen and Jagannathan (1997) measure of closeness (5.14) between SDFs and the goodness of fit criterion between probabilities $\sum_{i=0}^{I+1} (1/p_{it}^*) (p_{it}^* - p_{it}^*)^2$ will lead to similar conclusions if and only if the prior risk neutral probabilities p_{it}^* are close to the objective probability distribution p_{it} . However, risk neutral probabilities may include agents' anticipations about rare risks which are not apparent in a historical estimation of objective probabilities. This is the well-documented peso problem which has been discussed in the context of option pricing by Eraker (2001).

This discussion makes clear the potential drawback of the Euclidian distance (5.13) between probabilities. It does not put a sufficient weight on extreme events with small probabilities. This may lead to severe pricing errors since these small probabilities appear at the denominator of SDFs and therefore, have a large weight in the effective computation of derivative asset prices.

All the methodologies we have described in this section are geared toward extracting conditional risk-neutral distributions in the sense that they fit cross-sections of option prices and in that sense have to be opposed to the unconditional approach of the previous section. In the next section we summarize the advantages and disadvantages of both methods.

5.3 Comparing the unconditional and conditional methodologies for extracting risk-neutral distributions

Since the canonical valuation or the implied tree methodologies aim at obtaining risk neutral probabilities that come closest to pricing correctly the existing options at a single point in time, the risk-neutral distribution will change over time. On the contrary, a nonparametric kernel estimator aims at estimating the risk-neutral distribution as a fixed function of variables such as the current stock price, the exercise price, the riskless rate and other variables of interest. The functional form of the estimated risk-neutral distribution should be relatively stable over time. Since we cannot really say that one approach is better than the other, we can only sketch the advantages and disadvantages of both methods following Ait-Sahalia and Lo (1998).

We will compare the implied binomial tree method of Rubinstein (1994) to the semi-parametric estimate of the risk-neutral distribution of Aït-Sahalia and Lo (1998). The first method produces a distribution that is completely consistent with all option prices at each date, but it is not necessarily consistent across time. The second may fit poorly for a cross-section of option prices at some date but is consistent across time. However, being a fixed function of the relevant variables, the variation in the probabilities has to be captured by the variation in these variables. Another consideration is the intertemporal dependency in the risk-neutral distributions. The first method ignores it totally while the second exploits the dependencies in the data around a given date. Implied binomial trees are less data-intensive while the kernel method requires many cross-sections. Finally, smoothness has to be imposed for the first method, whereas the second method delivers a smooth function by construction. The stability of the risk-neutral distribution obtained with the kernel-based estimate should lower the out-of-sample forecasting errors at the expense of deteriorating the in-sample fit. Aït-Sahalia and Lo (1998) compare the out-of-sample forecasting performance of their semiparametric method to the implied tree method of Jackwerth and Rubinstein (1996) and conclude that, at short horizons (up to 5 days) the implied tree forecasting errors are lower, but that at horizons of 10 days and longer, the kernel method performance is better.

Recently, Aït-Sahalia and Duarte (2003) proposed a nonparametric method to estimate the risk-neutral density from a cross-section of option prices. This might appear surprising given that we know that nonparametric methods require a large quantity of data. Their nonparametric method is based on locally polynomial estimators that impose shape restrictions on the option pricing function. From the absence of arbitrage, we know that the price of a call option must be a decreasing and convex function of the strike price. The method consists therefore in two steps, first a constrained least square regression to impose monotonicity and convexity, followed by a locally polynomial kernel smoothing that preserves the constraints imposed in the first step. In a Monte Carlo analysis, Aït-Sahalia and Duarte (2003) show these constrained nonparametric estimates are feasible in the small samples encountered in a typical daily cross-section of option prices. In an application to S&P 500 call option data with about two months to maturity on a day in 1999, they compare several estimators for the price function, the first derivative with respect to the strike price and the state price density. The comparison emphasizes that the price function is well estimated near the money but that for high values of the strike, the locally quadratic and cubic estimators are highly variable, while the unconstrained Nadaraya-Watson estimator violates the convexity constraint on prices for low values of the strike. These poor properties show even more in the first and the second derivatives. For the first derivative,

all estimators except the constrained and unconstrained locally linear locally linear violate the first derivative constraint, while for the state price density (the second derivative) all the unconstrained estimators violate the positivity constraint in the left tail of the density or are too flat at the globally optimal bandwidth. This nonparametric approach with shape restrictions appears therefore promising but more evidence and comparisons are needed. In the next section we will revisit these constrained and unconstrained approaches in the seminonparametric context of sieve estimators, especially neural networks.

5.4 Seminonparametric estimators

In the seminonparametric approach, the nonlinear relationship f between the price of an option π and the various variables that affect its price, say Z , is approximated by a set of basis functions g :

$$f(Z, \cdot) = \sum_{n=1}^{\infty} \alpha_n g_n(Z, \cdot) \quad (5.15)$$

The term seminonparametric is explained by the fact that the basis functions are parametric, yet the parameters are not the object of interest since we need an infinity of them to estimate the function in the usual nonparametric sense. The methods vary according to the basis functions chosen. Hutchinson, Lo and Poggio (1994) propose various types of learning networks, Gouriéroux and Monfort (2001) consider approximations of the pricing kernel through splines, while Abadir and Rockinger (1998) investigates hypergeometric functions. In what follows, we will develop the neural network approach and see how one can choose the basis to obtain a valid state price density function. The basis chosen for neural networks will be:

$$g_n(Z, \alpha_n) = \frac{1}{1 + \exp(-\alpha_n Z)} \quad (5.16)$$

which is a very flexible sigmoid function. Then the function can be written as:

$$f(Z, \theta) = \beta_0 + \sum_{i=1}^d \beta_i \frac{1}{1 + \exp(\gamma_{i,0} - \gamma_{i,1} Z)} \quad (5.17)$$

where the vector of parameters $\theta = (\beta, \gamma)$ and the number d of units remains to be determined as the bandwidth in kernel methods. In neural network terminology, this is called a single hidden-layer feedforward network. Many authors have investigated the universal approximation properties of neural networks (see in particular Gallant and White (1988, 1992)). Using a wide variety of proof strategies, all have demonstrated that under general

regularity conditions, a sufficiently complex single hidden-layer feedforward network can approximate a large class of functions and their derivatives to any desired degree of accuracy where the complexity of a single hidden layer feedforward network is measured by the number of hidden units in the hidden layer. One of the requirements for this universal approximation property is that the activation function has to be a sigmoidal such as the logistic function presented above.

One nice property of this basis function is that the derivatives can be expressed in closed form. If we denote $h(Z) = \frac{1}{1+e^Z}$, then:

$$\begin{aligned} h'(Z) &= h(Z).(1 - h(Z)) \\ h''(Z) &= h(Z).(1 - h(Z)).(1 - 2h(Z)) \end{aligned}$$

Therefore, once the parameters of the pricing function are estimated for a given number of units, we can compute the hedge ratio or the risk-neutral distribution. Hutchinson, Lo and Poggio (1994) show using simulations that such an approach can learn the Black-Scholes formula. To reduce the number of inputs, Hutchinson, Lo and Poggio (1994) divide the function and its arguments by X and write the pricing function as a function of moneyness (S/X) and time-to-maturity (τ):

$$\frac{\pi_t}{X} = f\left(\frac{S_t}{X}, 1, \tau\right). \quad (5.18)$$

While they kept the number of units fixed, it is usually necessary as with any non-parametric method to choose it in some optimal way. The familiar trade-off is at play. Increasing the number of units d given a sample of data will lead to overfit the function in sample and cause a loss of predictive power out of sample. A way to choose the number of units is to use a cross-validation type of method on a validation period as proposed in Garcia and Gençay (2000). Although it is not mentioned in Hutchinson, Lo and Poggio (1994), even if we estimate well the pricing function, large errors are committed for the derivatives of the function and most notably, negative probabilities are obtained. This is consistent with what Aït-Sahalia and Duarte (2003) have found with local polynomial estimators based on a small sample of data, except that these bad properties are also present in large samples used for estimating the function over along time period.

A partial and imperfect way to better estimate the hedge ratio and the risk-neutral distribution is to use a network that will capture the homogeneity of the pricing function as in Garcia and Gençay (2000). The form in (5.18) assumes the homogeneity of degree one in the asset price and the strike price of the pricing function f . Another technical reason for dividing by the strike price is that the process S_t is nonstationary while the variable

S_t/X is stationary as strike prices bracket the underlying asset price process. This point is emphasized in Ghysels, Renault and Patilea (1997). From a theoretical point of view, the homogeneity property is obtained under unconditional or conditional independence of the distribution of returns from the level of the asset price (see Merton (1973) or Garcia and Renault (1998b)). Garcia and Gencay (2000) estimate a network of the form:

$$\frac{C_t}{X} = \beta_0 + \sum_{i=1}^d \beta_i^1 h(\gamma_{i,0}^1 + \gamma_{i,1}^1 \frac{S_t}{X} + \gamma_{i,2}^1 \tau) \quad (5.19)$$

$$-e^{-\alpha\tau} \sum_{i=1}^d \beta_i^2 h(\gamma_{i,0}^2 + \gamma_{i,1}^2 \frac{S_t}{X} + \gamma_{i,2}^2 \tau) \quad (5.20)$$

with $h(Z) = (1 + e^Z)^{-1}$. This has a similar structure than the Black-Scholes formula (which is itself homogeneous), except that the distribution function of the normal is replaced by neural network functions. Garcia and Gencay (2000) show that this structure improves the pricing performance compared to an unconstrained network, but that it does not improve the hedging performance. In fact, this network suffers (albeit slightly less) from the same deficiencies in terms of derivatives. To impose monotonicity and convexity on the function and ensuring that the resulting risk-neutral distribution is a proper density function as in Aït-Sahalia and Duarte (2003), we need to choose an appropriate structure for the network. The following basis function proposed in Dugas, Bengio, Bélisle, Nadeau and Garcia (2001):

$$\xi(Z) = \log(1 + e^Z) \quad (5.21)$$

is always positive and has its minimum at zero. Its first derivative:

$$\xi'(Z) = \frac{e^Z}{1 + e^Z} = h(Z) \quad (5.22)$$

is always positive and between 0 and 1 and therefore qualifies for a distribution function. Finally, its second derivative:

$$\xi''(Z) = h'(Z) = h(Z).(1 - h(Z)) \quad (5.23)$$

is always positive, becomes 0 when $h \rightarrow 0$ ($Z \rightarrow -\infty$) or when $h \rightarrow 1$ ($Z \rightarrow +\infty$) and has its maximum at $h = 1/2$ ($Z = 0$). These properties qualify for a density function.

Abadir and Rockinger (1998) with hypergeometric functions, Gottschling, Haefke and White (2000) with neural networks and Gouriéroux and Monfort (2001) with splines on the log-pricing kernel are three other ways to make sure that the estimated option pricing function always lead to a valid density, that is nonnegative everywhere and integrating to

one. Härdle and Yatchew (2001) also use nonparametric least squares to impose a variety of constraints on the option pricing function and its derivatives. Their estimator uses least squares over sets of functions bounded in Sobolev norm which offers a simple way of imposing smoothness on derivatives.

There is a need for a comparison of these methods which impose constraints on the estimation. Bondarenko (2003) proposes a new nonparametric method called positive convolution approximation which chooses among a rich set of admissible (smooth and well-behaved) densities the one that provides the best fit to the option prices. He conducts a Monte Carlo experiment to compare this method to seven other methods, parametric and nonparametric, which recover risk neutral densities. Daghli (2003) also provides a comparison between parametric and nonparametric methods for American options.

5.4.1 An economic application of nonparametric methods: extraction of preferences

Since, in a continuum of states, the state price density or risk neutral density corresponds to the Arrow-Debreu prices, it contains valuable information about the preferences of the representative investor. Indeed, the ratio of the state price density to the conditional objective probability density is proportional to the marginal rate of substitution of the representative investor, implying that preferences can be recovered given estimates of the state price density and the conditional objective distribution. A measure of relative risk aversion is given by:

$$\rho_t(S_T) = S_T \left(\frac{f'_t(S_T)}{f_t(S_T)} - \frac{f_t^{*'}(S_T)}{f_t^*(S_T)} \right) \quad (5.24)$$

where $f_t(S_T)$ and $f_t^*(S_T)$ denote respectively the conditional objective probability density and the state price density. This measure assumes that S_T , the value of the index at the maturity of the option, approximates aggregate consumption, the payoff on the market portfolio. Recently, Aït-Sahalia and Lo (2000), Jackwerth (2000) and Rosenberg and Engle (2002) followed this route to extract preferences based on semiparametric approaches. While Aït-Sahalia and Lo (2000) and Jackwerth (2000) estimate an average risk aversion function by using a nonparametric estimate of the objective density, Rosenberg and Engle (2002) compute a conditional risk aversion function with an asymmetric GARCH estimate of the conditional objective probability. The estimates obtained by Jackwerth (2000) reveal an oddly shaped risk aversion since marginal utility increases with wealth. In their parametric framework, Rosenberg and Engle (2002) find results that differ from Jackwerth (2000), in particular they do not find negative risk aversions when they use a power pricing

kernel. However, Rosenberg and Engle (2002) find results similar to Jackwerth's results when they use an orthogonal polynomial pricing kernel. In particular, they find that there is a region of negative risk aversion over the range from -4% to 2% for returns.

As noticed by Rosenberg and Engle (2002), the interpretation of the risk aversion function is debatable since the estimation technique of the implied binomial tree is based on time-aggregated data. This is the reason why Rosenberg and Engle (2002) propose to estimate the pricing kernel as a function of contemporaneously observed asset prices and a predicted asset payoff density based on an asymmetric GARCH model. The price to pay for this generality is the need to refer to a parametric model for the SDF. They propose:

$$m_{t+1}^* = E_t\left[\frac{m_{t+1}}{g_{t+1}}\right] = \theta_{0t}(g_{t+1})^{-\theta_{1t}} \quad (5.25)$$

The parameters of interest θ_{0t} and θ_{1t} are then estimated at each date t in order to minimize the sum of squared pricing errors that is differences between observed derivative prices (in a cross-section of derivatives all written on the same payoff g_{t+1}) and prices computed with the model SDF (5.25). As in the multinomial example described in section (2.2), there is some arbitrariness created by the choice of this particular quadratic measure of closeness. First, as discussed in Renault (1997), one may imagine that the pricing errors are severely heteroskedastic and mutually correlated. A GMM distance should get rid of this better than the uniform weighting. However, as stressed by Hansen and Jagannathan (1997), the GMM distance is probably not optimal to rank various misspecified SDFs since it gives an unfair advantage to the most volatile SDFs.

As explained above, Hansen and Jagannathan (1997) propose to consider directly a L^2 distance between SDFs. They show that it leads to a weighting matrix for pricing errors which is only defined by the covariance matrix of the net returns of interest and not by the product of returns with the SDF as in efficient GMM. Indeed, Buraschi and Jackwerth (2001) observe that the δ -metric of Hansen and Jagannathan (1997) has to be preferred to the GMM metric to select the best option pricing model since it is model independent whereas the optimal GMM weighting matrix is model dependent and asymptotic chi-squares tests typically reward models that generate highly volatile pricing errors.

Irrespective of the choice of a particular measure of closeness, the interpretation of parameters θ_{0t} and θ_{1t} which have been estimated from (5.25) may be questionable, except if a very specific model is postulated for the agent preferences. To illustrate this point, let us consider the general family of SDFs provided by the Epstein and Zin (1989) model of recursive utility:

$$m_{t+1} = \beta \left[\frac{C_{t+1}}{C_t} \right]^{\gamma(\rho-1)} \left[\frac{W_{t+1}}{(W_t - C_t)} \right]^{\gamma-1}, \quad (5.26)$$

where $\rho = 1 - 1/\sigma$ with σ the elasticity of intertemporal substitution, $\gamma = \alpha/\rho$ and $a = 1 - \alpha$ the index of comparative relative risk aversion. The variables C_t and W_t denote respectively the optimal consumption and wealth paths of the representative agent. They obey the following relationship:

$$\left[\frac{C_t}{W_t} \right] = [A(J_t)]^{1-\sigma},$$

where $V_t = A(J_t) \cdot W_t$ denotes the value at time t of the maximized recursive utility function. This value V_t is proportional to the wealth W_t available at time t for consumption and investment (homothetic preferences) and the coefficient of proportionality generally depends upon the information J_t available at time t . Therefore:

$$m_{t+1} = \beta \left[\frac{W_{t+1}}{W_t} \right]^{-a} \left[\frac{A(J_{t+1})}{A(J_t)} \right]^{1-a} [1 - A(J_t)^{1-\sigma}]^{\gamma-1} \quad (5.27)$$

Let us imagine, following Rosenberg and Engle (2002), that the agent wealth is proportional to the underlying asset payoff. Then:

$$m_{t+1}^* = E_t[m_{t+1}|g_{t+1}] = E_t[m_{t+1}|W_{t+1}]$$

will depend in general in a complicated way on the forecast of the value function $A(J_{t+1})$ as a function of W_{t+1} . For instance, we see that:

$$E_t[\log m_{t+1}|g_{t+1}] = B(J_t) - a \log \left[\frac{W_{t+1}}{W_t} \right] + (1 - a) E_t[\log A(J_{t+1})|W_{t+1}].$$

This illustrates that, except in the particular case $a = 1$ (logarithmic utility) or in a case where $A(J_{t+1})$ would not be correlated with W_{t+1} given J_t , the parameter θ_{1t} cannot be interpreted as risk aversion parameter and is not constant insofar as conditional heteroskedasticity will lead to time varying regression coefficients in $E_t[\log A(J_{t+1})|W_{t+1}]$. In other words, the intertemporal features of preferences which lead the agent to a non-myopic behavior prevent one to conclude that the risk aversion parameter is time-varying simply because one finds that the parameter θ_{1t} is time-varying. More generally, this analysis carries over to any missing factor in the parametric SDF model.

The general conclusion is that empirical pricing kernels which are computed without a precise account of the state variables which enter into the value function $A(J_t)$ cannot provide valuable insights on intertemporal preferences. For example, Chabi-Yo, Garcia and Renault (2002) show that in an economy with regime changes either in fundamentals or

preferences, an application of the nonparametric methodology used by Jackwerth (2000) to recover the absolute risk aversion will lead to similar negative estimates of the risk aversion function in some states of wealth even though the risk aversion functions are consistent with economic theory within each regime.

Of course, one can also question the representative agent framework. For example, Bates (2001) points out that the industrial organization of the stock index options market does not seem to be compatible with the representative agent construct and proposes a general equilibrium model in which crash-tolerant market makers insure crash-averse investors.

6 Conclusion

We have tried in this survey to offer a unifying framework to the prolific literature aimed at extracting useful and sometimes profitable economic information from derivatives markets. The stochastic discount factor methodology is by now the central tool in finance to price assets and provides a natural framework to integrate contributions in discrete and continuous time. Since most models are written in continuous time in option pricing we have established the link between these models and the discrete time approaches trying to emphasize the fundamental unity underlying both methodologies. To capture the empirical features of the stock market returns, which is the main underlying empirically studied in the option pricing literature, models have gained in complexity from the standard geometric Brownian motion of the seminal Black and Scholes (1973) model. Jump-diffusion models with various correlation effects have become increasingly complex to estimate. A main difficulty is the interplay of the latent variables which are everywhere present in the models and the inherent complex nonlinearities of the pricing formulas. This is the main aspect of the estimation methods on which we put some emphasis since the estimation of continuous time models is the object of another chapter in this Handbook.

Another major thread which underlies the survey is the interplay between preferences and option pricing. Even though the preference-free nature of the early formulas was often cited as a major advantage, it was not clear where this feature was coming from. We have made a special effort to specify the statistical assumptions that are needed to obtain this feature and to characterize the covariance or leverage effects which reintroduce preferences. In an equilibrium framework, the role of preferences appears clearly. In approaches based on the absence of arbitrage, these preferences are hidden in risk premia and it is harder to account for the links they impose between the risk premia of the numerous sources of risk. Researchers often treat these risk premia as free parameters and manage to capture some empirical facts but a deeper economic explanation is lacking. The extraction of preferences

from option prices using nonparametric methods is even more problematic. The puzzles associated with this literature often come from the fact that state variables have been omitted in the analysis.

Despite the length of the survey, there are a host of issues that we left unattended, especially issues pertaining to the implementation of models in practice. First, it is often difficult to obtain synchronized price data for derivatives and underlying fundamentals. This leads researchers to use theoretical relationships such as the put-call parity theorem to infer forward prices for the index. The same theorem is sometimes also used to infer prices for some far in-the-money options for which the reliability of the reported price is questionable because of staleness or illiquidity. Other types of filters such as taking out close-to-maturity options or options with close-to-zero prices are also imposed. All these data transformations have certainly an effect on model estimation and testing. A second issue concerns the final objective of the modelling exercise. Is the model intended to forecast future prices (or equivalently the moneyness and term structure of volatilities), to compute hedge ratios (or other greeks) or to recover risk-neutral probabilities for a certain horizon in order to price other derivatives on the same underlying asset? This is important both for estimation and testing of the model. Estimating a model according to a statistical criterion or to a financial objective leads to different estimates and performance in case of specification errors. Third, is the model taken at face value or do we recognize that it is fundamentally misspecified? Often, affine jump-diffusion models are reestimated every day or week and parameters can vary considerably from one cross-section to the other. Is it better to assume some latent structure instead of letting parameters vary from one period to the next? When agents make their financial decisions do they know the parameters or do they have to learn them? Is parameter uncertainty important? Do they try to make robust decisions? Finally, instead of exploiting fully-specified models, are the prices or bounds obtained by imposing weak economic restrictions useful? A recent retrospective by Bates (2003) addresses some of these issues and presents the current challenges that lie ahead.

References

- [1] Abadir, K. M., and M. Rockinger (1998), “Density-embedding Functions”, Working Paper, University of York and Groupe HEC.
- [2] Aït-Sahalia, Y.(2001), Telling From Discrete Data Whether the Underlying Continuous-Time Model is a Diffusion”, *Journal of Finance*, forthcoming.
- [3] Aït-Sahalia, Y., and J. Duarte (2003), “Nonparametric Option Pricing under Shape Restrictions”, *Journal of Econometrics*, 116, 9-47.
- [4] Aït-Sahalia, Y., L.P. Hansen and J. Scheinkman (2002), this Handbook.
- [5] Aït-Sahalia, Y., and A.W. Lo (2000), “Nonparametric Risk Management and Implied Risk Aversion”, *Journal of Econometrics*, 94, 9-51.
- [6] Aït-Sahalia, Y. and A. W. Lo (1998), “Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices”, *Journal of Finance*, 53, 2, 499-547.
- [7] Alizadeh, S., M. Brandt and F.X. Diebold (2002), “Range-based estimation of stochastic volatility models”, *Journal of Finance* 57, 1047-1091.
- [8] Amin, K.I. and R. Jarrow (1992), “Pricing Options in a Stochastic Interest Rate Economy,” *Mathematical Finance*, 3(3), 1-21.
- [9] Amin, K.I. and V.K. Ng (1993), “Option Valuation with Systematic Stochastic Volatility”, *Journal of Finance*, 48, 881-910.
- [10] Andersen, T.G., L. Benzoni and J. Lund (2001), “Towards an Empirical Foundation of Continuous-Time Equity Return Models”, *Journal of Finance*, 57, 1239-1284.
- [11] Andersen, T.G., Bollerslev, T. and F.X. Diebold (2002), “Parametric and Nonparametric Volatility Measurement”, in L.P. Hansen and Y. Aït-Sahalia (eds.), *Handbook of Financial Econometrics*. Amsterdam: North-Holland, forthcoming.
- [12] Andreou, E. and E. Ghysels (2002), “When do Jump Diffusions Approximate Returns? Nonparametric Test Procedures Based on Price Rigidity and Discreteness”, work in progress.
- [13] Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996), “Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity”, *Journal of Econometrics* 74, 3-30.

- [14] Bakshi, G., C. Cao and Z. Chen (1997), “Empirical Performance of Alternative Option Pricing Models”, *Journal of Finance*, 94, 1-2, 277-318.
- [15] Bakshi, G. and D. B. Madan (1999), “Crash Discovery in Stock and Options Markets”, Working Paper, University of Maryland.
- [16] Bakshi, G. and D. Madan (2000), “Spanning and Derivative-Security Valuation”, *Journal of Financial Economics*, 55, 205-238.
- [17] Bandi, F. M. and B. Perron (2001), “Long Memory and the Relation between Implied and Realized Volatility,” Working Paper, Université de Montréal
- [18] Bansal, R. and S. Viswanathan (1993), “No Arbitrage and Arbitrage Pricing: A New Approach”, *The Journal of Finance*, 4, 1231-1262.
- [19] Banz, R. and M. Miller (1978), “Prices for State-Contingent Claims: Some Estimates and Applications”, *Journal of Business*, 51, 653-672.
- [20] Barndorff-Nielsen, O.E. (2001), “Superposition of Ornstein-Uhlenbeck type processes”, *Theory of Probability and Its Applications* 45(2), 175-194.
- [21] Barndorff-Nielsen, O.E and Shephard, N. (2001), “Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion)”, *Journal of the Royal Statistical Society - Series B*, 63, 167-241.
- [22] Barndorff-Nielsen, O.E and Shephard, N. (2001a), “Modelling by Lévy processes for financial econometrics”, in *Lévy processes*, pp. 283-318. Boston, MA: Birkhäuser Boston.
- [23] Barndorff-Nielsen, O.E and Shephard, N. (2003), *Financial Volatility: Stochastic Volatility and Lévy based processes*, Cambridge University Press (forthcoming).
- [24] Bates, D. (1991), “The Crash of '87: Was it expected?”, *Journal of Finance*, 46, 1009-1044.
- [25] Bates, D. (1996a), “Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options”, *Review of Financial Studies*, 9, 69-107.
- [26] Bates, D. (1996b), “Testing Option Pricing Models”, in Maddala and Rao (eds.), *Handbook of Statistics*, vol. 14: Statistical Methods in Finance, 567-611.

- [27] Bates, D. (2000), "Post-'87 Crash Fears in the S&P 500 Futures Option", *Journal of Econometrics*, 94, 181-238.
- [28] Bates, D. (2001), "The Market for Crash Risk," NBER Working Paper 8557.
- [29] Bates, D. (2003), "Empirical Option Pricing: a retrospection," *Journal of Econometrics*, 116, 387-404.
- [30] Benzoni, L. (1998), "Pricing Options under Stochastic Volatility: an Econometric Analysis", Unpublished Manuscript, Kellogg Graduate School of Management.
- [31] Bertsimas Dimitris, Kogan Leonid, Lo Andrew W. (2000), "When is time continuous?", *Journal of Financial Economics*, (55)2 pp.173-204.
- [32] Black, F. (1976), "Studies of Stock Price Volatility Changes", Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section, 177-181.
- [33] Black, F. and M. Scholes (1973), "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81, 637-659.
- [34] Bollerslev, T. and H.O. Mikkelsen (1996), "Modelling and Pricing Long Memory in Stock Market Volatility", *Journal of Econometrics*, 73, 151-184.
- [35] Bollerslev, T. and H.O. Mikkelsen (1999), "Long-Term Equity Anticipation Securities and Stock Market Volatility Dynamics", *Journal of Econometrics*, 92, 75-99.
- [36] Bondarenko, O. (2003), "Estimation of risk-neutral densities using positive convolution approximation," *Journal of Econometrics* 116, 85-112.
- [37] Boyle, P.P. (1988) "A Lattice Framework for Option Pricing with Two State Variables", *Journal of Financial and Quantitative Analysis*, 23-1, 1-12.
- [38] Breeden, D. and R. Litzenberger (1978), "Prices of State Contingent Claims Implicit in Option Prices", *Journal of Business*, 51, 621-651.
- [39] Breidt, J., N. Crato and P. de Lima (1998) "On the detection and estimation of long memory in stochastic volatility", *Journal of Econometrics*, 73, 325-334.
- [40] Brennan, M. J. (1979), "The Pricing of Contingent Claims in Discrete-Time Models", *Journal of Finance*, 34, 53-68.

- [41] Broadie, M., Detemple, J., Ghysels, E. and Torrès, O. (2000a), “American options with stochastic dividends and volatility: A nonparametric investigation”, *Journal of Econometrics* 94, 53-92.
- [42] Broadie, M., Detemple, J., Ghysels, E. and Torrès, O. (2000b), “Nonparametric Estimation of American Options Exercise Boundaries and Call Prices”, *Journal of Economic Dynamics & Control* 24, 1829-1857.
- [43] Buchen, P. W. and M. Kelly (1996), “The Maximum Entropy Distribution of an Asset Inferred from Option Prices”, *Journal of Financial and Quantitative Analysis*, 31, 143-159.
- [44] Buraschi, A. and J. Jackwerth (2001), “The Price of a Smile: Hedging and Spanning in Option Markets”, *Review of Financial Studies*, 14, 495-527.
- [45] Campa, J. M. and P. H. K. Chang (1995), “Testing the Expectations Hypothesis of the Term Structure of Volatilities in Foreign Exchange Options”, *Journal of Finance*, 50, 529-547.
- [46] Carr, P., Geman, H., Madan, D.B. and M. Yor (2003), “Stochastic Volatility for Lévy Processes”, *Mathematical Finance*, Vol. 13, No. 3, 345-382.
- [47] Carr, P. and L. Wu (2002), “Time-changed Lévy Processes and Option Pricing”, *Journal of Financial Economics*, forthcoming.
- [48] Chabi-Yo, F., R. Garcia and E. Renault (2002), “State-Dependent Fundamentals or Preferences Explain Pricing Kernel Puzzle,” Working Paper, Université de Montréal.
- [49] Chacko, G. and L. Viceira (1999) “Spectral GMM Estimation of Continuous-Time Processes”, Working Paper, Graduate School of Business, Harvard University.
- [50] Chamberlain, G. and M. Rothschild (1983), “Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets”, *Econometrica*, 51, 1281-1304.
- [51] Chernov, M., R. Gallant, E. Ghysels and G. Tauchen (2000), “A New Class of Stochastic Volatility Models with Jumps: Theory and Estimation”, Discussion paper.
- [52] Chernov, M., R. Gallant, E. Ghysels and G. Tauchen (2003), “Alternative Models for Stock Price Dynamics”, *Journal of Econometrics*.

- [53] Chernov, M. and E. Ghysels (2000), “A Study towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation”, *Journal of Financial Economics*, 56(3), 407-458.
- [54] Chourdakis, K. and E. Tzavalis (2001), “Option Pricing under Discrete-Time Regime Shifts in Stock Returns”, Working Paper, Department of Economics, Queen Mary, University of London.
- [55] Clark, P. K. (1973), “A subordinated stochastic process model with finite variance for speculative prices”, *Econometrica*, 41,135-156.
- [56] Cochrane, J. (2001), *Asset Pricing*, Princeton University Press.
- [57] Comte, F. and E. Renault (1996), “Non-Causality in Continuous Time Models”, *Econometric Theory*, 12, n° 2, 1996, 215-256.
- [58] Comte, F. and E. Renault (1998), “Long Memory Continuous Time Stochastic Volatility Models”, *Mathematical Finance* 8, n° 4, 291-323. .
- [59] Comte, F. L. Coutin and E. Renault (2001), “Affine Fractional Stochastic Volatility Models with Application to Option Pricing”, Discussion Paper, Université de Montréal.
- [60] Cox, J. C., S. A., Ross and M. Rubinstein (1979), “Option Pricing: A Simplified Approach”, *Journal of Financial Economics*, 7, 229-263.
- [61] Daglish, T. (2003), “A Pricing and Hedging Comparison of Parametric and Nonparametric Approaches for American Index Options,” forthcoming *Journal of Financial Econometrics*.
- [62] Dai, Q. and K. Singleton (2000), “Specification Analysis of Affine Term Structure Models”, *Journal of Finance*, 5, 55, 1943-1978.
- [63] Das, S. R. and R. K. Sundaram (1999), “Of Smiles and Smirks: A Term Structure Perspective”, *Journal of Financial and Quantitative Analysis* 34, 211-239.
- [64] David, A. and P. Veronesi (2002), “Option Prices with Uncertain Fundamentals: Theory and Evidence on the Dynamics of Implied Volatilities,” Working Paper, Graduate School of Business, University of Chicago.
- [65] Derman, E. and I. Kani (1994), “Riding on the Smile”, *Risk*, 7, 32-39.

- [66] Duan, J.C. (1995), “The GARCH Option Pricing Model”, *Mathematical Finance* 5, 13-32.
- [67] Duan, J. C. (2001), “Risk Premium of Derivatives in Complete Markets”, Discussion Paper, Rotman School of Management, University of Toronto.
- [68] Duffie, D. (2001), *Dynamics Asset Pricing*, Princeton University Press.
- [69] Duffie, D. and R. Kan (1996), “A Yield-Factor Model of Interest Rates”, *Mathematical Finance*, 6, 379-406.
- [70] Duffie, D., Pan, J. and Singleton, K. (2000), “Transform Analysis and Asset Pricing for Affine Jump-Diffusions”, *Econometrica*, 68, 6, 1343-1376.
- [71] Dugas, C., Y. Bengio, F. Belisle, C. Nadeau, and R. Garcia (2001), “Incorporating Second-Order Functional Knowledge for Better Option Pricing,” *Advances in Neural Information Processing Systems*, volume 13.
- [72] Engle, R. F. and G. G. J. Lee (1999), “A Permanent and Transitory Component Model of Stock Return Volatility,” in R. F. Engle and H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, 475-497. Oxford, UK: Oxford University Press.
- [73] Epstein, L. and S. Zin (1989), “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework”, *Econometrica*, 57, 937-969.
- [74] Epstein, L. and S. Zin (1991), “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns I: An Empirical Analysis”, *Journal of Political Economy*, 99, 263-286.
- [75] Eraker, B. (2001), “Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices,” Manuscript, Department of Economics, Duke University.
- [76] Eraker, B., M. S. Johannes and N. G. Polson (2001), “The Impact of Jumps in Returns and Volatility”, *Journal of Finance*, forthcoming.
- [77] Fouque J-P., G. Papanicolaou and K. R. Sircar (2000), “Derivatives in Financial Markets With Stochastic Volatility”, Princeton University Press.

- [78] Gallant, A.R., C. Hsu and G. Tauchen (1999), “Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance”, *The Review of Economics and Statistics*, 81(4), 617-631.
- [79] Gallant, A.R. and G. Tauchen (1996), “Which Moments to Match?” *Econometric Theory*, 12, 657-681.
- [80] Gallant, A.R. and G. Tauchen (1998), “Reprojecting Partially Observed Systems with Applications to Interest Rate Diffusions”, *Journal of the American Statistical Association*, 93, 10-24.
- [81] Gallant, A.R. and G. Tauchen (2001), this Handbook.
- [82] Gallant, A. R. and H. White, (1988), There exists a neural network that does not make avoidable mistakes, Proceedings of the Second Annual IEEE Conference on Neural Networks, San Diego, CA, I.657-I.664, New York: IEEE Press.
- [83] Gallant, A. R. and H. White (1992), On learning the derivatives of an unknown mapping with multilayer feedforward networks, *Neural Networks*, 5, 129–138.
- [84] Garcia, R., and R. Gençay (2000), “Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint”, *Journal of Econometrics*, 94, 93-115.
- [85] Garcia, R., M. A. Lewis, and E. Renault (2001), “Estimation of Objective and Risk-Neutral Distributions Based on Moments of Integrated Volatility”, Working Paper, CIRANO, Université de Montréal.
- [86] Garcia, R., R. Luger and E. Renault (2003), “Empirical Assessment of an Intertemporal Option Pricing Model with Latent Variables”, *Journal of Econometrics*, 116, 49-83.
- [87] Garcia, R. and E. Renault (1998a), “A Note on GARCH Option Pricing and Hedging”, *Mathematical Finance*, 8,2, 153-161.
- [88] Garcia, R. and E. Renault (1998b), “Risk Aversion, Intertemporal Substitution and Option Pricing”, Working Paper 98s-02, CIRANO.
- [89] Garcia, R. and E. Renault (2001), “State Variables Models for Stochastic Discount Factors”, *Handbooks of Mathematical Finance: Topics in Option Pricing, Interest Rates and Risk Management*, J. Cvitanic, E. Jouini and M. Musiela eds., Cambridge University Press.

- [90] Ghysels, E., Gouriéroux, C. and J. Jasiak (1997), “Market Time and Asset Price Movements: Theory and Estimation”, in *Statistics in Finance*, D. Hand and S. Jacka (eds.), Edward Arnold, London, Chapter 15, 307-332.
- [91] Ghysels, E., A. Harvey and E. Renault (1996), “Stochastic Volatility,” in G.S. Maddala (ed.) *Handbook of statistics*, Vol. 14, Statistical Methods in Finance, North Holland, Amsterdam.
- [92] Ghysels, E. and S. Ng (1998), “A Semi-Parametric Factor Model of Interest Rates and Tests of the Affine Term Structure”, *Review of Economics and Statistics*, 80, 535-548.
- [93] Ghysels, E., Patilea, V., Renault, E. and O. Torrès (1997), “Nonparametric Methods and Option Pricing” in *Statistics in Finance*, D. Hand and S. Jacka (eds.), Edward Arnold, London, 261-282.
- [94] Gottschling, A. , C. Haefke and H. White (2000), “Closed Form Integration of Artificial Neural Networks with Some Applications to Finance”, Working Paper, University of California, San Diego.
- [95] Gouriéroux, C. and A. Monfort (2001), “Pricing with Splines”, Working Paper, Crest-Insee (Paris).
- [96] Gouriéroux, C., A. Monfort and E. Renault (1993), “Indirect Inference”, *Journal of Applied Econometrics*, 8, S85-S118.
- [97] Gouriéroux, C., A. Monfort and C. Tenreiro (1994), “Kernel M-Estimators: Nonparametric Diagnostics for Structural Models”, Working Paper 9405, CEPREMAP, Paris.
- [98] Guidolin, M. and A. Timmermann (1999), “Option Prices and Implied Volatility Dynamics under Bayesian Learning,” forthcoming *Journal of Economic Dynamics and Control*.
- [99] Hamilton, J. D. (1989), “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle”, *Econometrica*, 57, 357-84.
- [100] Hansen, L. P. and R. Jagannathan (1997), “Assessing Specification Errors in Stochastic Discount Factor Models”, *Journal of Finance*, 52, 557-590.

- [101] Hansen, L.P. and S. Richard (1987), “The role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models”, *Econometrica*, 55, 587-613.
- [102] Hansen, L. P. and J. Scheinkman (2001), “Semigroup Models of Markov Pricing,” CIRANO Workshop on Financial Mathematics and Econometrics, Montreal, June 2001.
- [103] Hansen, L. and K. Singleton (1982), “Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models,” *Econometrica*, 50, 1269-1285.
- [104] Härdle, W. and A. Yatchew (2001), “Dynamic Nonparametric State Price Density Estimation,” Working Paper, University of Toronto
- [105] Harvey, A.C. (1998), ”Long Memory in Stochastic Volatility, ” in J. Knight and J. Satchell (eds.), *Forecasting Volatility in Financial Markets*. London UK: Butterworth-Heinemann.
- [106] Harrison, J. M. and D. Kreps (1979), “Martingales and Arbitrage in Multiperiod Securities Markets”, *Journal of Economic Theory* 20, 381-408.
- [107] Heston, S. (1993), “A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options”, *The Review of Financial Studies*, 6, 327-343.
- [108] Heston, S. and S. Nandi (2000), “A Closed-Form GARCH Option Valuation Model”, *Review of Financial Studies*, 13(3), 585-625.
- [109] Hull, J. and A. White (1987), “The Pricing of Options on Assets with Stochastic Volatilities”, *Journal of Finance*, 42, 281-300.
- [110] 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*, 3, 851-889.
- [111] Jackwerth, J. (2000), “Recovering Risk Aversion from Option Prices and Realized Returns”, *Review of Financial Studies*, 13, 433-451.
- [112] Jackwerth, J. and M. Rubinstein (1996), “Recovering Probabilities Distributions from Option Prices”, *Journal of Finance*, 51, 1611-1631.
- [113] Johannes and Polson (2002), this Handbook

- [114] Jones, C. (2003), “The Dynamics of Stochastic Volatility: evidence from underlying and options markets”, *Journal of Econometrics*, 116, 181-224.
- [115] Kallsen, J. and M. S. Taqqu (1998), “Option Pricing in ARCH-type Models,” *Mathematical Finance*, 8, 13-26.
- [116] Kamrad, B. and P. Ritchken (1991), “Multinomial Approximating Models for Options with k State Variables,” *Management Science* 37 (12), 1640-1652.
- [117] Lewis, A. (2000), *Option Valuation Under Stochastic Volatility with Mathematica Code*, Finance Press.
- [118] Longstaff, F. A. (1995), “Option Pricing and the Martingale Restriction,” *Review of Financial Studies*, 8, 4, 1091-1124.
- [119] Madan, D.B. and F. Milne (1991), “Option Pricing with VG Martingale Components”, *Mathematical Finance*, 1, 39-56.
- [120] Madan, D.B. and E. Seneta (1990), “The Variance Gamma (VG) Model for Share Market Returns”, *Journal of Business*, 63, 511-524.
- [121] Manaster, S. and R. J. Rendleman, JR. (1982), “Option Prices as Predictors of Equilibrium Stock Prices”, *The Journal of Finance*, 37, n° 4.
- [122] Meddahi, N. (2001), “An Eigenfunction Approach for Volatility Modelling”, Working Paper, CRDE, Université de Montréal.
- [123] Meddahi, N. and E. Renault (2000), “Temporal Aggregation of Volatility Models,” forthcoming in *Journal of Econometrics*.
- [124] Merton, R. C. (1973), “Theory of Rational Option Pricing”, *Bell Journal of Economics and Management Science* 4, 141-183.
- [125] Merton, R. C. (1992), *Continuous-Time Finance*, Basil Blackwell.
- [126] Nicolato, E. and E. Venardos (2002), “Option Pricing in Stochastic volatility Models of the Ornstein-Uhlenbeck type”, *Mathematical Finance*, forthcoming.
- [127] Pan, J. (2002), “The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study,” *Journal of Financial Economics*, 63, 3-50.

- [128] Pastorello S., E. Renault, and N. Touzi (2000), “Statistical Inference for Random Variance Option Pricing,” *Journal of Business and Economic Statistics*, 18(3), 358-367.
- [129] Pastorello, S., Patilea, V. and E. Renault (2003), “Iterative and Recursive Estimation of Structural Non Adaptive Models”, *Journal of Business and Economic Statistics*, forthcoming.
- [130] Patilea, V. and E. Renault (1997) “Continuously Updated Extremum Estimators,” CORE Discussion Paper 97-76.
- [131] Pliska, S. (1997), *Introduction to Mathematical Finance: Discrete Time Models*, Basil Blackwell.
- [132] Poteshman, A. M. (2000), “Estimating a General Stochastic Variance Model from Option Prices,” Working Paper, University of Illinois at Urbana-Champaign.
- [133] Renault, E. (1997), “Econometric Models of Option Pricing Errors”, in *Advances in Economics and Econometrics: Theory and Applications*, Vol. 3, Kreps, D. and Wallis, K. eds., Cambridge University Press.
- [134] Renault, E. (2001), “Option Pricing Models Made Useful for Econometric Analysis”, CIRANO Workshop on Financial Mathematics and Econometrics, Montréal, June 2001.
- [135] Renault, E. and N. Touzi (1996), “Option Hedging and Implied Volatilities in a Stochastic Volatility Model”, *Mathematical Finance*, 6, 279-302.
- [136] Robinson, P.M. (1991), ”Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression,” *Journal of Econometrics*, 47, 67-84.
- [137] Rogers, L.C.G. (1997), “Arbitrage with Fractional Brownian Motion,” *Mathematical Finance* 7, 95-105.
- [138] Romano, M. and Touzi, N. (1997), “Contingent Claims and Market Completeness in a Stochastic Volatility Model,” *Mathematical Finance* 7, 399-412.
- [139] Ross, S. A. (1976), “The Arbitrage Theory of Capital Asset Pricing,” *Journal of Economic Theory* 13, 341-360.

- [140] Rosenberg, J.V. (2000), “Asset Pricing Puzzles: Evidence from Option Markets”, Working Paper, NYU Stern School of Business.
- [141] Rosenberg, J.V. and R.F. Engle (2002), “Empirical Pricing Kernels”, *Journal of Financial Economics* 64, 341-372.
- [142] Rubinstein, M. (1994), “Implied Binomial Trees”, *Journal of Finance*, vol. XLIX, 3, 771-818.
- [143] Rubinstein, M. (1996), “The Valuation of Uncertain Income Streams and the Pricing of Options”, *Bell Journal of Economics and Management Science*, 407-425.
- [144] Rubinstein, M. and J. Jackwerth (1997), “Recovering Probabilities and Risk Aversion from Option Prices and Realized Returns”, Working Paper, University of Berkeley.
- [145] Sato, K. (1999), *Lévy Processes and Infinite Divisibility*, Cambridge University Press.
- [146] Shephard, N. (1996), “Statistical aspects of ARCH and stochastic volatility” in D.R. Cox, D.V. Hinkley, and O.E. Barndorff-Nielsen (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, 1-67. London : Chapman & Hall.
- [147] Shimko, D. (1993), “Bounds on Probabilitiy”, *Risk*, 6, 33-37.
- [148] Stutzer, M. (1996), “A Simple Nonparametric Approach to Derivative Security Valuation”, *Journal of Finance* 51(5), 1633-1652.
- [149] Sundaresan, S. (2000), “Continuous-Time Methods in Finance: A Review and an Assessment,” *Journal of Finance*, 55(4), 1569-1622.
- [150] Tauchen, G. and M. Pitts (1983), “The price variability-volume relationship on speculative markets”, *Econometrica* 51, 485–505.
- [151] Taylor, S.J. (1986), *Modelling Financial Time Series*, John Wiley, Chichester.
- [152] Turnbull, S., and F. Milne (1991), “A Simple Approach to Interest-Rate Option Pricing,” *Review of Financial Studies*, 4, 87-121.
- [153] Wiggins, J. (1987), “Option Values under Stochastic Volatility,” *Journal of Financial Economics*, 19(2), 351-372.
- [154] Willard, G.A. (1997), “Calculating Prices and Sensitivities for Path-independent Derivative Securities in Multifactor Models”, *The Journal of Derivatives*, Fall 1997, 45-61.