CHAPTER 9

The Econometrics of Option Pricing

René Garcia*, Eric Ghysels,** *and* **Eric Renault*****

* Edhec Business School, Nice, France
** Department of Finance, Kenan-Flagler Business School and Department of Economics, University of North Carolina, Chapel Hill, NC
*** Department of Economics, University of North Carolina, Chapel Hill, NC

Contents

Abstract

Our survey will explore possible explanations for the divergence between the objective and the risk-neutral distributions. Modeling 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. 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 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 modeling 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.

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.

1. INTRODUCTION AND OVERVIEW

To delimit the focus of this survey, we will put emphasis on the more recent contributions because 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 et al. (1996) and Shephard (1996) provide a detailed analysis of stochastic volatility (SV) modeling, whereas Renault (1997) explores the econometric modeling 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 predate 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 modeling the return process for equity indices 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 10 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. Modeling 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 modeling stock returns at the time prior surveys were written was a continuous-time 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 (AJD) processes for which one could obtain similar results. Duffie et al. (2000) discuss equity and fixed income derivatives pricing for the general class of AJD. The evidence regarding the empirical fit of the affine class of processes is mixed, see, e.g., Dai and Singleton (2000), Chernov et al. (2003), and Ghysels and Ng (1998) for further discussion. There is a consensus that single volatility factor models, affine (like Heston, 1993) or nonaffine (like Hull and White, 1987 or Wiggins, 1987), do not fit the data (see Andersen et al., 2010; Benzoni, 1998; Chernov et al., 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 et al., 2001; Bates, 1996a; Chernov et al., 2003; Eraker et al., 2003; Pan, 2002, among others). Bakshi et al. (1997), Bates (2000), Chernov et al. (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 et al. (2003) maintain that a two-factor nonaffine logarithmic SV diffusion model without jumps yields a superior empirical fit compared with affine one-factor or two-factor SV processes or SV diffusions with jumps. Alternative models were also proposed: they include volatility models of the Ornstein–Uhlenbeck type but with Lévy innovations (Barndorff-Nielsen and Shephard, 2001) and SV models with long memory in volatility (Breidt et al., 1998; 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 SV or jumps are idiosyncratic and not priced by the market. There is a long tradition of this, but more recent empirical work clearly indicates there are prices for volatility and jump risk (see, e.g., Andersen et al., 2010; Chernov and Ghysels, 2000; Jones, 2003; 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 et al. (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 et al. (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 SV and with or without jumps as in Chernov and Ghysels (2000), Pan (2002), Poteshman (2000), and Garcia et al. (2009).

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 makes 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, which will be reviewed in this chapter.

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. 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.

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. Because risk aversion can be recovered empirically from the risk-neutral and the actual distributions, Aït-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 generalized method of moments (GMM) approach. Although this has been extensively done with stock and Treasury bill return data (see Epstein and Zin, 1991; Hansen and Singleton, 1982, among others), it is only more recently that Garcia et al. (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 (RNVRs) (Brennan, 1979) hold in a general stochastic discount factor (SDF) framework (Hansen and Richard, 1987). When these statistical restrictions do not hold, it will be shown that preferences play a role. Bates (2007) 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 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 SDF. 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 modeling 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. PRICING KERNELS, RISK-NEUTRAL PROBABILITIES, AND OPTION PRICING

The widespread acceptance among academics and investment professionals of the Black–Scholes (BS) option pricing formula as a benchmark is undoubtedly due to its usefulness for pricing and hedging options, irrespective of the unrealistic assumptions of the initial model. The purpose of econometrics of option pricing is not really to check the empirical validity of this model. It has been widely documented that by contrast with maintained assumptions of Black and Scholes geometric Brownian motion model, stock return exhibits both stochastic volatility and jumps. Thus, the interesting issue is not the validity of the model itself. In this section, we will rather set the focus on the assessment of the possible errors of the BS option pricing formula and on empirically successful strategies to alleviate them. After all, we can get the right pricing and hedging formula with a wrong model. This is the reason why the largest part of this section is focused on econometric modeling and inference about empirically valid extensions of the BS option pricing formula.

However, it is worth stressing even more generally the econometric content of arbitrage pricing. As first emphasized by Cox et al. (1979), there is a message of the Black and Scholes approach which goes beyond any particular specification of the underlying stochastic processes. Arbitrage-free pricing models generally allow to interpret derivative prices as expectations of discounted payoffs, when expectations are computed with respect to an equivalent martingale measure. It is worth stressing in this respect a nice correspondence between the theory of arbitrage pricing and econometrics of option pricing. Although option contracts are useful to complete the markets and so to get an unique equivalent martingale measure, the statistical observation of option prices is generally informative about the underlying equivalent martingale measure. Although only the historical probability distribution can be estimated from return data on the underlying asset, option prices data allow the econometrician to perform some statistical inference about the relevant martingale measure. This will be the main focus of interest of this chapter. For sake of expositional simplicity, as in Black and Scholes (1972) first empirical tests of their option pricing approach, the option contracts considered in this chapter will be mainly European calls written on stocks. Of course, in the same way, BS option pricing methodology has since been generalized to pricing of many other derivative securities, the econometric approaches sketched below can be extended accordingly.

2.1. Equivalent Martingale Measure and Volatility Smile

Assume that all stochastic processes of interest are adapted in a filtered probability space $(\Omega, (\mathcal{F}_t), P)$. Under standard regularity conditions, the absence of arbitrage is equivalent to the existence of an equivalent martingale measure Q. Without loss of generality, we will consider throughout that the payoffs of options of interest are attainable (see, e.g., Föllmer and Schied, 2004). Then, the arbitrage-free price of these options is defined without ambiguity as expectation under the probability measure Q of the discounted value of their payoff. Moreover, for an European call with maturity T, we will rather characterize its arbitrage price at time $t < T$ as the discounted value at time t of its expectation under the time t forward measure $Q_{t,T}$ for time T. By Bayes rule, $Q_{t,T}$ is straightforwardly defined as equivalent to the restriction of Q on $\mathcal{F}_t$. The density function $\mathrm{d}Q_{t,T}/\mathrm{d}Q$ is $[B(t,T)]^{-1}(B_t/B_T)$, where B_t stands for the value at time t of a bank account, whereas $B(t,T)$ is the time t price of a pure discount bond (with unit face value) maturing at time T. If K and S_t denote, respectively, the strike price and the price a time t of the underlying stock, the option price C_t a time t is

$$C_t = B(t,T)E^{Q_{t,T}}\mathrm{Max}[0, S_T - K]. \tag{2.1}$$

A formula such (2.1) provides a decomposition of the option price into two components:

$$C_t = S_t \Delta_{1t} - K\Delta_{2t}, \tag{2.2}$$

where

$$\Delta_{2t} = B(t,T)Q_{t,T}[S_T \geq K] \tag{2.3}$$

and

$$\Delta_{1t} = \Delta_{2t} E^{Q_{t,T}}\left[\frac{S_T}{S_t} \mid S_T \geq K\right] \tag{2.4}$$

It follows immediately (see Huang and Litzenberger, 1998, pp. 140, 169) that

$$\Delta_{2t} = -\frac{\partial C_t}{\partial K} \tag{2.5}$$

In other words, a cross-section at time t of European call option prices all maturing at time T, but with different strike prices, K is informative about the pricing probability measure $Q_{t,T}$. In the limit, a continuous observation of the function $K \longrightarrow C_t$ (or of its partial derivative $\partial C_t/\partial K$) would completely characterize the cumulative distribution function of the underlying asset return (S_T/S_t) under $Q_{t,T}$. Let us rather consider it through the probability distribution of the continuously compounded net return on the period $[t, T]$:

$$r_S(t,T) = \log\left[\frac{S_T B(t,T)}{S_t}\right]$$

With (log-forward) moneyness of the option measured by

$$x_t = \log\left[\frac{KB(t,T)}{S_t}\right],$$

the probability distribution under $Q_{t,T}$ of the net return on the stock $r_S(t,T)$ is characterized by its survival function deduced from (2.3) and (2.5) as

$$G_{t,T}(x_t) = -\exp(-x_t)\frac{\partial C_t}{\partial x_t}, \tag{2.6}$$

where

$$C_t(x_t) = \frac{C_t}{S_t} = E^{Q_{t,T}}\{\text{Max}[0, \exp(r_S(t,T)) - \exp(x_t)]\}. \tag{2.7}$$

For the purpose of easy assessment of suitably normalized orders of magnitude, practitioners often prefer to plot as a function of moneyness x_t the BS-implied volatility $\sigma^{\text{imp}}_{t,T}(x_t)$ rather than the option price $C_t(x_t)$ itself. The key reason that makes this sensible is that in the Black and Scholes model, the pricing distribution is indexed by a single volatility parameter σ. Under BS' assumptions, the probability distribution of the net return $r_S(t, T)$ under $Q_{t,T}$ is the normal with mean $(-1/2)(T - t)\sigma^2$ and variance $(T - t)\sigma^2$. Let us denote $\aleph_{t,T}(\sigma)$ this distribution.

Then, the BS-implied volatility $\sigma^{\text{imp}}_{T-t}(x_t)$ is defined as the value of the volatility parameter σ^2, which would generate the observed option price $C(x_t)$ as if the distribution of net return under $Q_{t,T}$ was the normal $\aleph_{t,T}(\sigma)$. In other words, $\sigma^{\text{imp}}_{T-t}(x_t)$ is characterized as solution of the equation:

$$C_t(x_t) = BS_h[x_t, \sigma^{\text{imp}}_h(x_t)] \tag{2.8}$$

where $h = T - t$ and

$$BS_h[x, \sigma] = N[d_1(x, \sigma, h)] - \exp(x)N[d_2(x, \sigma, h)], \tag{2.9}$$

where N is the cumulative distribution function of the standardized normal and

$$d_1(x, \sigma, h) = \frac{-x}{\sigma\sqrt{h}} + \frac{1}{2}h\sigma^2$$
$$d_2(x, \sigma, h) = \frac{-x}{\sigma\sqrt{h}} - \frac{1}{2}h\sigma^2.$$

It is worth reminding that the common use of the BS-implied volatility $\sigma^{\text{imp}}_h(x_t)$ by no mean implies that people think that the BS model is well specified. By (2.8), $\sigma^{\text{imp}}_h(x_t)$ is nothing but a known strictly increasing function of the observed option price $C_t(x_t)$. When plotting the volatility smile as a function $x_t \longrightarrow \sigma^{\text{imp}}_h(x_t)$ rather than $x_t \longrightarrow C_t(x_t)$, people simply consider a convenient rescaling of the characterization (2.6) of the pricing distribution. However, this rescaling depends on x_t and, by definition, produces a flat volatility smile whenever the BS pricing formula is valid in cross-section (for all moneynesses at a given maturity) for some specific value of the volatility parameter. Note that the validity of the BS model itself is only a sufficient but not necessary condition for that.

2.2. How to Graph the Smile?

When the volatility smile is not flat, its pattern obviously depends whether implied volatility $\sigma^{\text{imp}}_h(x_t)$ is plot against strike K, moneyness (K/S_t), forward moneyness $(KB(t, T)/S_t) = \exp(x_t)$, log-forward moneyness x_t, etc. The optimal variable choice of course depends on what kind of information people expect to be revealed immediately when plotting implied volatilities. The common terminology "volatility smile" seems to suggest that people had initially in mind a kind of U-shaped pattern, whereas words like "smirk" or "frown" suggest that the focus is set on frequently observed asymmetries with respect to a symmetric benchmark. Even more explicitly, because the volatility smile is supposed to reveal the underlying pricing probability measure $Q_{t,T}$, a common wisdom is that asymmetries observed in the volatility smile reveal a corresponding skewness in the distribution of (log) return under $Q_{t,T}$. Note in particular that as mentioned above, a flat volatility smile at level σ characterizes a normal distribution with mean $(-\sigma^2/2)$ and variance σ^2.

Beyond the flat case, the common belief of a tight connection between smile asymmetries and risk-neutral skewness requires further qualification. First, the choice of variable must of course matter for discussion of smile asymmetries. We will argue below that log-forward moneyness x_t is the right choice, i.e., the smile asymmetry issue must be understood as a violation of the identity

$$\sigma_h^{\text{imp}}(x_t) = \sigma_h^{\text{imp}}(-x_t). \tag{2.10}$$

This identity is actually necessary and sufficient to deduce from (2.8) that the general option pricing formula (2.7) fulfills the same kind of symmetry property than the Black and Scholes one:

$$C_t(x) = 1 - \exp(x) + \exp(x)C_t(-x). \tag{2.11}$$

Although (2.11) is automatically fulfilled when $C_t(x) = BS_h[x, \sigma]$ [by the symmetry property of the normal distribution: $N(-d) = 1 - N(d)$], it characterizes the symmetry property of the forward measure that corresponds to volatility smile symmetry. It actually mimics the symmetry property of the normal distribution with mean $(-\sigma^2/2)$ and variance σ^2, which would prevail in case of validity of the Black and Scholes model. By differentiation of (2.11) and comparison with (2.6), it can be easily checked that *the volatility smile is symmetric in the sense of (2.10) if and only if, when $f_{t,T}$ stands for the probability density function of the log-return $r_S(t, T)$ under the forward measure $Q_{t,T}$, $\exp(x/2)f_{t,T}(x)$ is an even function of x.*

In conclusion, the relevant concept of symmetry amounts to consider pairs of moneynesses that are symmetric of each other in the following sense:

$$x_{1t} = \log\left[\frac{K_1 B(t, T)}{S_t}\right] = -x_{2t} = \log\left[\frac{S_t}{K_2 B(t, T)}\right].$$

In other words, the geometric mean of the two discounted strike prices coincides with the current stock price:

$$\sqrt{K_1 B(t, T)}\sqrt{K_2 B(t, T)} = S_t.$$

To conclude, it is worth noting that graphing the smile as a function of the log-moneyness x_t is even more relevant when one maintains the natural assumption that option prices are homogeneous functions of degree one with respect to the pair (S_t, K). Merton (1973) had advocated this homogeneity property to preclude any "perverse local concavity" of the option price with respect to the stock price. It is obvious from (2.7) that a sufficient condition for homogeneity is that as in the Black and Scholes case, the pricing probability distribution $Q_{t,T}$ does not depend on the level S_t of the stock price. This is the reason why, as discussed by Garcia and Renault (1998a), homogeneity holds with standard SV option pricing models and does not hold for GARCH option pricing.

For our purpose, the big advantage of the homogeneity assumption is that it allows to compare volatility smiles (for a given time to maturity) at different dates since then the implied volatility $\sigma_h^{\text{imp}}(x_t)$ depends only on moneyness x_t and not directly on the level S_t of the underlying stock price. Moreover, from the Euler characterization of homogeneity:

$$C_t = S_t \frac{\partial C_t}{\partial S_t} + K \frac{\partial C_t}{\partial K}$$

we deduce [by comparing (2.2) and (2.5)] that

$$\Delta_{1t} = \frac{\partial C_t}{\partial S_t} \tag{2.12}$$

is the standard delta-hedging ratio. Note that a common practice is to compute a proxy of Δ_{1t} by plugging $\sigma_h^{\text{imp}}(x_t)$ in the BS delta ratio. Unfortunately, this approximation suffers from a Jensen bias when the correct option price is a mixture of BS prices (see Section 2.5) according to some probability distribution of the volatility parameter. It is shown in Renault and Touzi (1996) and Renault (1997) that the BS delta ratio [computed with $\sigma_h^{\text{imp}}(x_t)$] underestimates (resp. overestimates) the correct ratio Δ_{1t} when the option is in the money (resp. out of the money), i.e., when $x_t < 0$ (resp. $x_t > 0$).

2.3. Stochastic Discount Factors and Pricing Kernels

Since Harrison and Kreps (1979), the so-called "fundamental theorem of asset pricing" relates the absence of arbitrage opportunity on financial markets to the existence of equivalent martingale measures.

The market model is arbitrage-free if and only if the set of all equivalent martingale measures is nonempty. It is a mild version of the old "efficient market hypothesis" that states that discounted prices should obey the fair game rule, i.e., to behave as martingales. Although Lucas (1978) had clearly shown that efficiency should not preclude risk-compensation, the notion of equivalent martingale measures reconciles the points of view. The martingale property and associated "risk-neutral pricing" is recovered for some distortion of the historical probability measure that encapsulates risk compensation. This distortion preserves "equivalence" by ensuring the existence of a strictly positive probability density function.

For the purpose of econometric inference, the concept of risk-neutral pricing may be less well suited because the characterization of a valid equivalent martingale measure depends in a complicated way of the time-span, the frequency of transactions, the filtration of information, and the list of primitive assets involved in self-financing strategies. Following Lucas (1978) and more generally the rational expectations literature, the econometrician rather sets the focus on investors' decisions at every given date, presuming that they know the true probability distributions over states of the world. In general, investors' information will be limited so that the true state of the world is not revealed to them at any point of time. Econometrician's information is even more limited and will always be viewed as a subset of investors' information. This is the reason why Hansen and Richard (1987) have revisited Harrison and Kreps (1979) Hilbert space methods to allow flexible conditioning on well-suited information sets. In a way, the change of probability measure is then introduced for a given date of investment and a given horizon, similarly to the forward equivalent measure.

The equivalent martingale measure approach allows to show the existence at any given date t and for any maturity date $T > t$ of an equivalent forward measure $Q_{t,T}$ such that the price π_t at time t of a future payoff g_T available at time T is

$$\pi_t = B(t, T)E^{Q_{t,T}}[g_T \mid (F_t)]. \tag{2.13}$$

Similarly, Hansen and Richard (1987) directly prove the existence of a strictly positive random variable $M_{t,T}$ such that

$$\pi_t = E_t[M_{t,T}g_T], \tag{2.14}$$

where the generic notation $E_t\{.\}$ is used to denote the historical conditional expectation, given a market-wide information set about which we do not want to be specific. Up to this degree of freedom, there is basically no difference between pricing equations (2.13) and (2.14). First note that (2.14), valid for any payoff, determines in particular the price $B(t, T)$ of a pure discount bond that delivers \$1 at time T:

$$B(t, T) = E_t[M_{t,T}]. \tag{2.15}$$

The discount factor in (2.13), equal to $B(t, T)$, is also the conditional expectation at time t of any variable $M_{t,T}$ conformable to (2.14). Such a variable is called a SDF. Thus, $\frac{M_{t,T}}{B(t,T)}$ is a probability density function that directly defines a forward measure from the historical measure. By contrast, a forward measure is usually defined in mathematical finance from its density with respect to an equivalent martingale measure. The latter involves the specification of the locally risk free spot rate. However, it is not surprising to observe that the issue of market incompleteness which is detrimental in mathematical finance due to the nonuniqueness of an equivalent martingale measure will also affect the econometric methodology of SDF pricing.

In the rest of this section, we discuss the properties of the SDF while overlooking the issue of its lack of uniqueness. It is first worth reminding the economic interpretation of the SDF. With obvious notations, plugging (2.15) into (2.14) allows to rewrite the latter as

$$B(t,T)E_t(R_{t,T}) = 1 - \mathrm{cov}_t\big[R_{t,T}, M_{t,T}\big], \tag{2.16}$$

where $R_{t,T} = \frac{g_T}{\pi_t}$ denotes the return over the period $[t, T]$ on the risky asset with terminal payoff g_T. In other words, the random features of the discounted risky return $B(t,T)R_{t,T}$ allow a positive risk premium (a discounted expected return larger than 1) in proportion of its covariance with the opposite of the SDF.

In the same way, the Bayes rule leads to see risk-neutral densities as multiplicative functionals over aggregated consecutive periods, and we must see the SDF as produced by the relative increments of an underlying pricing kernel process. Let $\tau < T$ be an intermediate trading date between dates t and T. The time T payoff g_T could be purchased at date t, or it could be purchased at date τ with a prior date t purchase of a claim to the date τ purchase price. The "law of one price" guarantees that these two ways to acquire the payoff g_T must have the same initial cost. This recursion argument implies a multiplicative structure on consecutive SDFs. There exists an adapted positive stochastic process m_t called the pricing kernel process such that

$$M_{t,T} = \frac{m_T}{m_t}. \tag{2.17}$$

Following Lucas (1978), a popular example of pricing kernel is based on the consumption process of a representative investor. Under suitable assumptions for preferences and endowment shocks, it is well known that market completeness allows us to introduce a representative investor with utility function U. Assuming that he or she can consume C_t at date t and C_T at the fixed future date T and that he or she receives a given portfolio of financial assets as endowment at date t, the representative investor adjusts the dollar amount invested in each component of the portfolio at each intermediary date to maximize the expected utility of his or her terminal consumption at time T. In equilibrium, the investor optimally invests all his or her wealth in the given portfolio and then consumes its terminal value C_T. Thus, the Euler first-order condition for optimality imposes that the price π_t at time t of any contingent claim that delivers the dollar amount g_T at time t is such that

$$\pi_t = E_t\left[\beta^{T-t}\frac{U'(C_T)}{U'(C_t)}g_T\right], \tag{2.18}$$

where β is the subjective discount parameter of the investor. For instance, with a constant relative risk aversion (CRRA) specification of the utility function, $U'(C) = C^{-a}$ where $a \geq 0$ is the Arrow–Pratt measure of relative risk aversion, and we have the consumption-based pricing-kernel process:

$$m_t = \beta^t C_t^{-a}. \tag{2.19}$$

2.4. Black–Scholes-Implied Volatility as a Calibrated Parameter

It is convenient to rewrite the call pricing equation (2.1) in terms of pricing kernel:

$$C_t = E_t\left[\frac{m_T}{m_t}\text{Max}[0, S_T - K]\right]. \tag{2.20}$$

It is then easy to check that the call pricing formula collapses into the BS one when the two following conditions are fulfilled:

- The conditional distribution given $\mathcal{F}_t$ of the log-return $\log\left[\frac{S_T}{S_t}\right]$ is normal with constant variance σ and
- The log-pricing kernel $\log(\frac{m_T}{m_t})$ is perfectly correlated to the log-return on the stock.

An important example of such a perfect correlation is the consumption-based pricing kernel described above when the investor's initial endowment is only one share of the stock such that he or she consumes the terminal value $S_T = C_T$ of the stock. Then,

$$\log\left[\frac{m_T}{m_t}\right] = -a\log\left[\frac{S_T}{S_t}\right] + (T - t)\log(\beta). \tag{2.21}$$

We will get a first interesting generalization of the BS formula by considering now that the log-return $\log\left[\frac{S_T}{S_t}\right]$ and the log-pricing kernel $\log(m_{t,T})$ may be jointly normally distributed given $\mathcal{F}_t$, with conditional moments possibly depending on the conditional information at time t. Interestingly enough, it can be shown that the call price computed from formula (2.20) with this joint conditional lognormal distribution will depend explicitly on the conditional moments only through the conditional stock volatility:

$$(T - t)\sigma_{t,T}^2 = \text{Var}_t\left[\log\left(\frac{S_T}{S_t}\right)\right]$$

More precisely, we get the following option pricing formula:

$$C_t = S_t BS_{T-t}[x_t, \sigma_{t,T-t}]. \tag{2.22}$$

The formula (2.22) is actually a generalization of the RNVR put forward by Brennan (1979) in the particular case (2.19). With joint lognormality of return and pricing kernel, we are back to a Black and Scholes functional form due to the Cameron–Martin formula (a static version of Girsanov's theorem), which tells us that when X and Y are jointly normal:

$$E\{\exp(X)g(Y)\} = E[\exp(X)]E\{g[Y + \text{cov}(X, Y)]\}$$

Although the term $E[\exp(X)]$ will give $B(t, T)$ [with $X = \log(m_{t,T})$], the term $\text{cov}(X, Y)$ (with $Y = \log[S_T/S_t]$) will make the risk-neutralization because $E\{\exp(X)\exp(Y)\}$ must be one as it equals $E[\exp(X)]E\{\exp[Y + \text{cov}(X, Y)]\}$.

From an econometric viewpoint, the interest of (2.22), when compared with (2.8), is to deliver a flat volatility smile but with an implied volatility level which may be time varying and corresponds to the conditional variance of the conditionally lognormal stock return. In other words, the time-varying volatility of the stock becomes observable as calibrated from option prices:

$$\sigma_{t,T} = \sigma^{\text{imp}}_{T-t}(x_t), \quad \forall x_t$$

The weakness of this approach is its lack of robustness with respect to temporal aggregation. In the GARCH-type literature, stock returns may be conditionally lognormal when they are considered on the elementary period of the discrete time setting ($T = t + 1$), whereas implied time-aggregated dynamics are more complicated. This is the reason why the GARCH-option pricing literature (Duan, 1995 and Heston and Nandi, 2000) maintains the formula (2.22) only for $T = t + 1$. Nonflat volatility smiles may be observed with longer times to maturity. Kallsen and Taqqu (1998) provide a continuous-time interpretation of such GARCH option pricing.

2.5. Black–Scholes-Implied Volatility as an Expected Average Volatility

To account for excess kurtosis and skewness in stock log-returns, a fast empirical approach amounts to consider that the option price a time t is given by a weighted average:

$$\alpha_t S_t BS_h[x_t, \sigma_{1t}] + (1 - \alpha_t) S_t BS_h[x_t, \sigma_{2t}]. \tag{2.23}$$

The rationale for (2.23) is to consider that a mixture of two normal distributions with standard errors σ_{1t} and σ_{2t} and weights α_t and $(1 - \alpha_t)$, respectively, may account for both skewness and excess kurtosis in stock log-return. The problem with this naive approach is that it does not take into account any risk premium associated to the mixture component. More precisely, if we want to accommodate a mixture of normal distributions with a mixing variable $U_{t,T}$, we can rewrite (2.20) as

$$C_t = E^P\{E^P\left[M_{t,T}\text{Max}[0, S_T - K] \mid \mathcal{F}_t, U_{t,T}\right] \big| \mathcal{F}_t\}, \tag{2.24}$$

where, for each possible value $u_{t,T}$ of $U_{t,T}$, a BS formula like (2.22) is valid to compute

$$E^P\left[M_{t,T}\text{Max}[0, S_T - K] \mid \mathcal{F}_t, U_{t,T} = u_{t,T}\right].$$

In other words, it is true that as in (2.23), the conditional expectation operator [given (F_t)] in (2.24) displays the option price as a weighted average of different BS prices with the weights corresponding to the probabilities of the possible values $u_{t,T}$ of the mixing variable $U_{t,T}$. However, the naive approach (2.23) is applied in a wrong way when forgetting that the additional conditioning information $U_{t,T}$ should lead to modify some key inputs in the BS option pricing formula. Suppose that investors are told that the mixing variable $U_{t,T}$ will take the value $u_{t,T}$. Then, the current stock price would no longer be

$$S_t = E^P\left[M_{t,T}S_T \mid \mathcal{F}_t\right]$$

but instead

$$S_t^*(u_{t,T}) = E^P\left[M_{t,T}S_T \mid \mathcal{F}_t, U_{t,T} = u_{t,T}\right]. \tag{2.25}$$

For the same reason, the pure discount bond that delivers \$1 at time T will no longer be priced at time t as

$$B(t, T) = E^P\left[M_{t,T} \mid \mathcal{F}_t\right]$$

but rather

$$B^*(t, T)(u_{t,T}) = E^P\left[M_{t,T} \mid \mathcal{F}_t, U_{t,T} = u_{t,T}\right]. \tag{2.26}$$

Hence, various BS option prices that are averaged in a mixture approach like (2.23) must be computed, no longer with actual values $B(t, T)$ and S_t of the current bond and stock prices but with values $B^*(t, T)(u_{t,T})$ and $S_t^*(u_{t,T})$ not directly observed but computed from (2.26) and (2.25). In particular, the key inputs, underlying stock price and interest rate, should be different in various applications of the BS formulas like $BS_h[x, \sigma_1]$ and $BS_h[x, \sigma_2]$ in (2.23). This remark is crucial for the conditional Monte Carlo approach, as developed for instance in Willard (1997) in the context of option pricing with SV. Revisiting a formula initially derived by Romano and Touzi (1997), Willard (1997) notes that the variance reduction technique, known as conditional Monte Carlo, can be applied even when the conditioning factor (the SV process) is instantaneously correlated with the stock return as it is the case when leverage effect is present. He stresses that "by conditioning on the entire path of the noise element in the volatility (instead of just the average volatility), we can still write the option's price as an expectation over Black-Scholes prices by appropriately adjusting the arguments to the Black-Scholes formula". Willard's (1997) "appropriate adjustment" of the stock price is actually akin to (2.25). Moreover, he does not explicitly adjust the interest rate according to (2.26) and works with a fixed risk-neutral distribution. The Generalized Black–Scholes (GBS) option pricing below makes the required adjustments explicit.

2.6. Generalized Black–Scholes Option Pricing Formula

Let us specify the continuous-time dynamics of a pricing kernel $M_{t,T}$ as the relative increment of a pricing kernel process m_t according to (2.17). The key idea of the mixture model is then to define a conditioning variable $U_{t,T}$ such that the pricing kernel process and the stock price process jointly follow a bivariate geometric Brownian motion under the historical conditional probability distribution given $U_{t,T}$. The mixing variable $U_{t,T}$ will typically show up as a function of a state variable path $(X_\tau)_{t\leq\tau\leq T}$. More precisely, we specify the jump-diffusion model

$$\mathrm{d}(\log S_t) = \mu(X_t)\mathrm{d}t + \alpha(X_t)\mathrm{d}W_{1t} + \beta(X_t)\mathrm{d}W_{2t} + \gamma_t \mathrm{d}N_t \tag{2.27}$$

$$\mathrm{d}(\log m_t) = h(X_t)\mathrm{d}t + a(X_t)\mathrm{d}W_{1t} + b(X_t)\mathrm{d}W_{2t} + c_t \mathrm{d}N_t, \tag{2.28}$$

where (W_{1t}, W_{2t}) is a two-dimensional standard Brownian motion, N_t is a Poisson process with intensity $\lambda(X_t)$ depending on the state variable X_t, and the jump sizes c_t and γ_t are i.i.d. independent normal variables independent of the state variable process (X_t). The Brownian motion (W_{1t}) is assumed to be part of the state variable vector (X_t) to capture the possible instantaneous correlation between ex-jump volatility of the stock [as measured by $V_t = \alpha^2(X_t) + \beta^2(X_t)$] and its Brownian innovation. More precisely, the correlation coefficient $\rho(X_t) = \frac{\alpha(X_t)}{\sqrt{V_t}}$ measures the so-called leverage effect.

The jump-diffusion model [(2.27) and (2.28)] is devised such that given the state variables path $(X_\tau)_{t\leq\tau\leq T}$ as well as the number $(N_T - N_t)$ of jumps between times t and T, the joint normality of $(\log S_T, \log m_T)$ is maintained. This remark allows us to derive a GBS option pricing formula by application of (2.24) and (2.22):

$$C_t = S_t E^P\left[\xi_{t,T} BS_{T-t}(x_t^*, \sigma_{t,T}) \mid \mathcal{F}_t\right], \tag{2.29}$$

where

$$\sigma_{t,T}^2 = \int_t^T \left[1 - \rho^2(X_\tau)\right] V_\tau d\tau + (N_T - N_t)\mathrm{Var}(\gamma_t) \tag{2.30}$$

and

$$x_t^* = \log\left[\frac{KB^*(t,T)}{S_t \xi_{t,T}}\right],$$

where $S_t\xi_{t,T}$ and $B^*(t,T)$ correspond, respectively, to $S_t^*(u_{t,T})$ and $B^*(t,T)(u_{t,T})$ defined in (2.25) and (2.26). General computations of these quantities in the context of a jump-diffusion model can be found in Yoon (2008). Let us exemplify these formulas when there is no jump. Then, we can define a short-term interest rate as

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

and then

$$B^*(t,T) = \exp\left[-\int_t^T r(X_\tau)\mathrm{d}\tau\right]\exp\left[\int_t^T a(X_\tau)\mathrm{d}W_{1\tau} - \frac{1}{2}\int_t^T a^2(X_\tau)\mathrm{d}\tau\right] \tag{2.31}$$

and

$$\xi_{t,T} = \exp\left[\int_t^T [a(X_\tau) + \alpha(X_\tau)]\mathrm{d}W_{1\tau} - \frac{1}{2}\int_t^T [a(X_\tau) + \alpha(X_\tau)]^2\mathrm{d}\tau\right]. \tag{2.32}$$

It may be easily checked in particular that

$$B(t,T) = E^P\left[B^*(t,T) \mid \mathcal{F}_t\right]$$

and

$$S_t = E^P\left[S_t\xi_{t,T} \mid \mathcal{F}_t\right].$$

Let us neglect for the moment the difference between $S_t\xi_{t,T}$ and $B^*(t,T)$ and their respective expectations S_t and $B(t,T)$. It is then clear that the GBS formula warrants the interpretation of the BS-implied volatility $\sigma_{T-t}^{\mathrm{imp}}(x_t)$ as approximatively an expected average volatility. Up to Jensen effects (nonlinearity of the BS formula with respect to volatility), the GBS formula would actually give

$$\left[\sigma_{T-t}^{\mathrm{imp}}(x_t)\right]^2 = E^P\left[\sigma_{t,T}^2 \mid \mathcal{F}_t\right]. \tag{2.33}$$

The likely impact of the difference between $S_t\xi_{t,T}$ and $B^*(t,T)$ and their respective expectations S_t and $B(t,T)$ is twofold. First, a nonzero function $a(X_\tau)$ must be understood as a risk premium on the volatility risk. In other words, the above interpretation of $\sigma_{T-t}^{\mathrm{imp}}(x_t)$ as approximatively an expected average volatility can be maintained by using risk-neutral expectations. Considering the BS-implied volatility as a predictor of volatility over the lifetime of the option is tantamount to neglect the volatility risk premium. Beyond this risk premium effect, the leverage effect $\rho(X_t)$ will distort this interpretation through its joint impact on $\sigma_{t,T}^2$ and on $\xi_{t,T}$ as well (through $\alpha(X_t) = \rho(X_t)\sqrt{V_t}$).

Although Renault and Touzi (1996) have shown that we will get a symmetric volatility smile in case of zero-leverage, Renault (1997) explains that with nonzero leverage, the implied distortion of the stock price by the factor $\xi_{t,T}$ will produce asymmetric volatility smirks. Yoon (2008) characterizes more precisely the cumulated impact of the two effects of leverage and shows that they compensate each other almost exactly for at the money options, confirming the empirical evidence documented by Chernov (2007). Finally, Comte and Renault (1998) long-memory volatility model explains that in spite of the time averaging in (2.30), (2.33) the volatility smile does not become flat even for long-term options.

It is worth stressing that the fact that $S_t\xi_{t,T}$ and $B^*(t, T)$ may not coincide with their respective expectations S_t and $B(t, T)$ implies that, by contrast with the standard BS option pricing, the GBS formula is not preference free. Although in preference-free option pricing, the preference parameters are hidden within the observed value of the bond price and the stock price, and the explicit impact of the volatility risk premium function $a(X_t)$ in the formulas (2.32) and (2.31) for $\xi_{t,T}$ and $B^*(t, T)$ is likely to result in an explicit occurrence of preference parameters within the option pricing formula (see Garcia et al., 2005, and references therein for a general discussion). Although Garcia and Renault (1988b) characterize the respective impacts of risk aversion and elasticity of intertemporal substitution on option prices, Garcia et al. (2003) set the focus on the converse property. Because the impact of preference parameters on option prices should be beyond their role in bond and stock prices, option price data are likely to be even more informative about preference parameters. This hypothesis is strikingly confirmed by their econometric estimation of preference parameters.

Although properly taking into account the difference between historical and risk-neutral expectations, the tight connection [(2.30) and (2.33)] between BS-implied volatility and the underlying volatility process ($\sqrt{V_t}$) has inspired a strand of literature on estimating volatility dynamics from option prices data. Pastorello et al. (2000) consider directly $[\sigma^{\text{imp}}_{T-t}(x_t)]^2$ as a proxy for squared spot volatility V_t and correct the resulting approximation bias in estimating volatility dynamics by indirect inference. The "implied-states approach" described in Section 4 uses more efficiently the exact relationship between $\sigma^{\text{imp}}_{T-t}(x_t)$ and V_t, as given by (2.29), (2.30) for a given spot volatility model, to estimate the volatility parameters by maximum likelihood or GMM.

3. MODELING ASSET PRICE DYNAMICS VIA DIFFUSIONS FOR THE PURPOSE OF OPTION PRICING

Because the seminal papers by 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 SV models introduced by Hull and White (1987) and reviewed in Ghysels et al. (1996). Although 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 by Heston (1993), who proposed a SV diffusion with an analytical option pricing formula, was generalized by Duffie and Kan (1996) and Duffie et al. (2000) to a rich class of AJD. This class will be reviewed in a first subsection. Alternative models, mostly nonaffine, 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 AJD models examined in detail by Duffie et al. (2000) 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 $\mathrm{d}\log S_t = \mathrm{d}U_{1t}$ 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, and the second term $\exp \Delta X_t - \iota$ represents discrete jumps, where X_t is a N-dimensional Lévy process and ι is a vector of ones. The process U_t is governed by the following equations:

$$\mathrm{d}U_t = \mu(U_t, t)\mathrm{d}t + \sigma(U_t, t)\mathrm{d}W_t + \exp \Delta X_t - \iota, \tag{3.1}$$

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 $\mu : D \to \mathbf{R}^N$ and $\sigma(y)\sigma(y)' = h + \sum_{j=1}^{N} y_j H^{(j)}$ where $\sigma : D \to \mathbf{R}^{N\times N}$. 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 $\mathbf{R}^N$. Although the first component of the U_t process relates to returns, the other components U_{it} for $i = 2, \ldots, N$ govern either the stochastic drift or volatility of returns.[1] 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, e.g., Chernov et al., 2000, 2003). 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:

$$\begin{pmatrix} \mathrm{d}Y_t \\ \mathrm{d}V_t \end{pmatrix} = \begin{pmatrix} \mu \\ \kappa(\theta - V_t) \end{pmatrix}\mathrm{d}t + \sqrt{V_t}\begin{pmatrix} 1 & 0 \\ \rho\sigma_v & \sqrt{(1-\rho^2)}\sigma_v \end{pmatrix}\mathrm{d}W_t + \xi \mathrm{d}N_t, \tag{3.2}$$

[1]All further details regarding the regularity conditions pertaining to the U_t are discussed by Duffie et al. (2000) and therefore omitted.

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 arrival intensities λ_y and λ_v, and $\xi = (\xi^y, \xi^v)'$ is a vector of jump sizes for returns and volatility, respectively.[2] We adopt the mnemonics used by Duffie et al. and Eraker et al. (2003): SV for SV models with no jumps in returns nor volatility $(\lambda_y = \lambda_v = 0)$, SVJ for SV models with jumps in returns only $(\lambda_y > 0, \lambda_v = 0)$, and SVJJ for SV 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 jump 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 papers have investigated the Heston (1993) SV model. Most papers (Andersen et al., 2010; Benzoni, 1998; Eraker et al., 2003) conclude that the SV model provides a much better fit of stock return than standard one-factor diffusions. In particular, the strong negative correlation around -0.4 found between the volatility shocks and the underlying stock return shocks captures well the negative skewness observed in stock returns. However, the model is rejected because it is unable to accommodate the excess kurtosis observed in the stock returns.[3] 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 because the continuous path SV 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 process.[4] Eraker et al. (2003) propose a likelihood-based methodology using Markov Chain Monte Carlo methods (see also Jones, 2003). 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 volatility of volatility, and the speed of mean reversion all fall dramatically. This is somewhat consistent with the results of Andersen et al. (2010) when they estimate the models from 1980 till 1996 but with less magnitude. However, in the latter study,

[2]A specification with βV_t in the drift of the returns equation was considered by Eraker et al. (2003). This additional term was found to be insignificant, in accordance with the findings of Andersen et al. (2001) and Pan (2002).

[3]Both Andersen et al. (2001) and Benzoni (1998) estimate a nonaffine 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.

[4]For a discussion of the different types of volatility filters, see Ghysels et al. (1996) and the chapter of Gallant and Tauchen (2010) in this Handbook.

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 et al. (2003), 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 et al. (2010) and Chernov et al. (2003). The introduction of jumps lowers the negative correlation between the innovations in returns and the innovations 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 resolve 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 SV component. Duffie et al. (2000) generalize this class, and Chernov et al. (2000), Eraker et al. (2003), and Pan (2002) estimate multifactor jump-diffusion models with affine stochastic jump intensity. The models considered by Duffie et al. 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 imply 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 et al. (2000) and Eraker et al. (2003) consider also a slightly more general class of processes:

$$\lambda(x, U) = \lambda_0(x, t) + \lambda_1(x, t) U_t, \tag{3.4}$$

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, e.g., 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 six additional parameters to determine the jump intensity, without specifying the size distribution of jump. Chernov et al. (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 eve with such long data sets.[5]

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 et al. (2000) provide very elegant general discussions of the class of AJDs with SV, 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.

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

$$\begin{aligned} \mathrm{d}Y_t &= (\alpha_{10} + \alpha_{12}U_{1t})\mathrm{d}t + \sqrt{\beta_{10} + \beta_{12}U_{2t} + \beta_{13}U_{3t}}(\mathrm{d}W_{1t} + \psi_{12}\mathrm{d}W_{2t} + \psi_{13}\mathrm{d}W_{3t}) \\ \mathrm{d}U_{1t} &= (\alpha_{20} + \alpha_{22}U_{1t})\mathrm{d}t + \beta_{20}\mathrm{d}W_{2t} \\ \mathrm{d}U_{it} &= (\alpha_{i0} + \alpha_{ii}U_{it})\mathrm{d}t + \sqrt{\beta_{i0} + \beta_{ii}U_{it}}\mathrm{d}W_{it}, \quad i = 2, 3. \end{aligned} \tag{3.5}$$

[5] Chernov et al. (2000) also examine nonaffine Lévy processes, which will be covered in the next subsection.

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, whereas 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 et al. (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 nonaffine 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. Nonaffine Index Models

Another way to capture the small and large movements in returns is to specify SV models with two factors as in Chernov et al. (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

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

$$\mathrm{d}U_{it} = (\alpha_{i0} + \alpha_{ii} U_{it})\,\mathrm{d}t + (\beta_{i0} + \beta_{ii} U_{it})\,\mathrm{d}W_{it}, \quad i = 2, 3 \tag{3.7}$$

Chernov et al. (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 for $i = 3$, there is feedback, a feature found to be important in Gallant et al. (1999) and Jones (2003). The exponential specification in (3.6) is of course not the only index function one can consider.

Chernov et al. (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 et al. (2002), Chacko and Viceira (1999), Gallant et al. (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 because 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 et al. (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, e.g., Barndorff-Nielsen and Shephard (2001), Clark (1973), Ghysels et al. (1997), Madan and Seneta (1990), and Tauchen and Pitts (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 Carr et al. (2003), Carr and Wu (2004), 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 Eq. (3.1) the class of affine jump-diffusion processes. Nicolato and Venardos consider a different class, namely

$$\mathrm{d}Y_t = (\mu + \beta\sigma_t^2)\mathrm{d}t + \sigma_t \mathrm{d}W_t + \rho \mathrm{d}Z_{\lambda t} \tag{3.8}$$

$$\mathrm{d}\sigma_t = -\delta\sigma_t^2 \mathrm{d}t + \mathrm{d}Z_{\delta t} \tag{3.9}$$

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, and called by Barndorff-Nielsen and Shephard (2001) 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

$$\sigma_t^2 = \exp -\delta t \sigma_0^2 + \int_0^t \exp t - s \mathrm{d}Z_{\delta s}. \tag{3.10}$$

The resulting dynamics of the stock price process are

$$\begin{aligned} \mathrm{d}S_t &= S_{t-}(b_t \mathrm{d}t + \sigma_t + \mathrm{d}M_t) \\ \mathrm{d}b_t &= \mu + \delta\kappa(\rho) +] \left(\beta + \frac{1}{2}\right)\sigma_t^2 \\ M_t &= \sum_{0<s\leq t} (\exp \rho \Delta Z_{\delta s} - 1) - \delta\kappa(\rho)t, \end{aligned} \tag{3.11}$$

where $\kappa(x)$ is the cumulant transform, i.e., $\kappa(x) = \log E[\exp x Z_1]$. 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 (2001) and Madan and Seneta (1990).

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 BS formula using so-called *effective* log-stock prices, namely

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

similar to an expression of Hull and White (1987) and similar to the GSB 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 nonaffine diffusions with leverage as well as jump-diffusions. To compute actual option prices, Nicolato and Venardos (2003) suggest to simulate the pair $(Y_{\text{eff}}, V_{\text{eff}})$ and provide the relevant references to do so.

The observation that asset prices actually display many small jumps on a fine time scale has led to the development of more general jump structures, which permit an infinite number of jumps to occur within any finite time interval. Examples of infinite activity jump models include the inverse Gaussian model of Barndorff-Nielsen (1998, 2001), the generalized hyperbolic class of Eberlein et al. (1998), the variance gamma (VG) model of Madan and Milne (1991), the generalization of VG in Carr et al. (2003), and the finite moment log-stable model of Carr and Wu (2003). Empirical work by these authors is generally supportive of the use of infinite-activity processes as a way to model returns in a parsimonious way. The recognition that volatility is stochastic has led to further extensions of infinite activity Lévy models by Barndorff-Nielsen and Shephard (2001) and by Carr et al. (2003). However, these models often assume that changes in volatility are independent of asset returns and consider the leverage effect only under special cases. Carr and Wu (2004) use time-changed Lévy processes which generalize the affine Poisson jump-diffusions by relaxing the affine structure and by allowing more general specifications of the jump structure. Since the pioneering work by Heston (1993), the literature has used the characteristic function for deriving option prices. Accordingly, Carr and Wu focus on developing analytic expressions for the characteristic function of a time-changed Lévy process. Carr et al. (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.

To conclude, it should be noted that much has been written on testing for jumps in the context of high-frequency financial data, see for instance Andersen et al. (2010) in this Handbook as well as the survey by Brockwell (2009) and Eberlein (2009).

3.2.3. Long-Memory in Continuous Time

In Section 2, 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 SV and its instantaneous correlation with the return of the underlying asset. However, as pointed out by Sundaresan (2000) in his survey of the performance of continuous-time methods for option valuation, the remaining puzzle is the so-called term structure of volatility smiles, i.e., 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 long-term ones, which is consistent with the SV 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 $(1/h)\int_t^{t+h}\sigma^2(u)\mathrm{d}u$ 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 SV 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:

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

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

$$E\left[R_t^2\,(h)\,/\sigma\right] = \int_t^{t+h} \sigma_u^2 \mathrm{d}u$$

and

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

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

$$k(h) = \frac{E\left[R_t^4(h)\right]}{\left(E\left[R_t^2(h)\right]\right)^2} = 3\left[1 + \frac{\mathrm{Var}\left[\frac{1}{h}\int_t^{t+h}\sigma_u^2\mathrm{d}u\right]}{\left(E(\sigma^2)\right)^2}\right]. \tag{3.13}$$

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 ($h \to 0$) and infinitely long time to maturity ($h \to \infty$). First, because $\frac{1}{h}\int_t^{t+h} \sigma_u^2 \mathrm{d}u$ converges in mean-square toward σ_t^2 when h tends to zero:

$$\lim_{h \to 0} k(h) = 3\left[1 + \frac{\operatorname{Var}(\sigma^2)}{\left(E(\sigma^2)\right)^2}\right]. \tag{3.14}$$

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 three 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 $V_t\left[\frac{1}{h}\int_t^{t+h} \sigma_u^2 \mathrm{d}u\right]$ 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 BS pricing for very short-term options cannot be captured within the one-factor SV 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 SV 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. To address this issue, Comte and Renault (1998) proposed a continuous-time SV model with long memory. Long memory in volatility dynamics is a well-documented empirical fact (see, e.g., Ding et al., 1993), which has given rise to various long-memory GARCH models (Baillie et al., 1996; Bollerslev and Mikkelsen, 1996; Robinson, 1991) and long-memory discrete time SV models (Breidt et al., 1998; Harvey, 1998).

To get a long-memory continuous-time SV model, the basic idea of Comte and Renault (1998) was to extend the lognormal SV model to fractional Brownian motion. The log-volatility process follows Ornstein–Uhlenbeck dynamics but with the standard Brownian motion replaced by a fractional one. Because the main strand of the volatility literature is now more oriented toward affine models, we rather present here an overview of the affine fractional SV of Comte et al. (2001). 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 SV model as in Heston (1993), $d\tilde{\sigma}^2(t) = k\left(\tilde{\theta} - \tilde{\sigma}^2(t)\right) dt + \gamma\tilde{\sigma}(t)dW^X(t)$, Comte et al. consider the centered process $X(t) = \tilde{\sigma}^2(t) - \tilde{\theta}$ and a fractional integration of it:

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

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

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

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 Comte et al. for a discussion), the volatility process is defined by $\sigma_t^2 = X^{(d)}(t) + \theta$ for some positive parameter θ. For $d = 0, \sigma_t^2$ is a standard affine volatility process:

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

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

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

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

and the autocorrelation function of σ_t^2 has only an hyperbolic rate of decay for large lags: $\rho\left[\sigma_{t+h}^2, \sigma_t^2\right] \sim (kh)^{2d-1}/\,\Gamma(2d)$ 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 accommodates much better the volatility smile puzzle for long-term options. Indeed, it can be shown that for $0 \leq d < 1/2$,

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

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 $\left(\gamma^2\tilde{\theta}/k^{2d+1}\right)$ 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 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 toward its Gaussian limit 3 at the some speed h^{2d-1} as the conditional variance goes to zero. In other words and by 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 may produce cumbersome statistics because the past information is very slowly forgotten. However, a convenient feature of the affine fractional SV model is that integrated volatility $\int_t^{t+h} \sigma_s^2 \mathrm{d}s$ over the lifetime of the option and BS-implied volatilities are fractionally cointegrated. Moreover, the conditional probability distribution of $\int_t^{t+h} \sigma_s^2 \mathrm{d}s - E_t\left[\int_t^{t+h} \sigma_s^2 \mathrm{d}s\right]$ given information available at time t only depends on the current value of the state variable $X(t)$.

In other words, 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 et al. (2010) and Eraker et al. (2003) 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 SV, 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, to price options in an arbitrage-free framework, one needs to specify a candidate state price density (SPD) or to characterize the transformation from the objective measure to the risk-neutral measure. In the presence of jump and SV risks, appropriate risk compensation must be incorporated in the risk-neutral dynamics. As already noted, there are potentially risk premia associated with SV, 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 SV 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 et al. (2001) and Eraker et al. (2003) make such an assumption and study the impact of SV 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 because 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, SV 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 et al. (2003) 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% points in the implied volatility for at-the-money options with 1 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 et al. (2001) who need to add risk premia to generate steep-implied volatility curves. This is mainly due to the fact that their estimates for the jump parameters are small compared with Eraker et al. (2003). 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

The concept of pricing kernel or SPD is central to the dynamic asset pricing theory, in particular to the pricing of derivatives. The price at time t of a claim paying an $\mathcal{F}_T$−measurable random variable V at time T is given by

$$\pi_t = \frac{1}{\theta_t} E[V\theta_T \mid \mathcal{F}_T]. \tag{4.1}$$

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

$$\theta_t = \exp\left(-\int_0^t r_\tau \mathrm{d}\tau\right) \exp\left(-\int_0^t \zeta_\tau \mathrm{d}W_\tau - \frac{1}{2}\int_0^t \zeta_\tau' \zeta_\tau \mathrm{d}\tau\right) \exp\left(\sum_{i,\tau_i \le t} \xi_i^\pi\right), \tag{4.2}$$

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

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

This specification of the market prices of risk makes the risk premia for the diffusive 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 (1996a, 2000) based on a log utility model for the representative investor.

The jump risks are priced by the jump components ξ_i^π in the SPD, assumed to be i.i.d. and normally distributed 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 $\theta_t \exp\left(\int_0^t r_\tau \mathrm{d}\tau\right)$. The SVJ model will be then written as

$$\begin{pmatrix} \frac{\mathrm{d}S_t}{S_t} \\ \mathrm{d}V_t \end{pmatrix} = \begin{pmatrix} r_t - \eta^s V_t - \lambda_y^* \mu_y^* \\ \kappa(\alpha - V_t) + \eta^v V_t \end{pmatrix} \mathrm{d}t + \sqrt{V_t} \begin{pmatrix} 1 & 0 \\ \rho\sigma_v & \sqrt{(1-\rho^2)}\sigma_v \end{pmatrix} \mathrm{d}W_t^* + \begin{pmatrix} \xi^y \mathrm{d}N_t^{Qy^*} \\ 0 \end{pmatrix} \tag{4.4}$$

The risk-neutral dynamics differs from the dynamics 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

$$W_t^* = W_t + \int_0^t \zeta_s \mathrm{d}s, \quad 0 \leq t \leq T. \tag{4.5}$$

The jump process N^{Q^*y} has the same distribution under Q^* than under Q except that $\xi^y \sim \mathcal{N}(\mu_y^*, \sigma_y^2)$, where $\mu_y^* = \mu_y + \sigma_y \sigma_\pi \rho_\pi$. It means that the model allows for a jump-size risk. It can also allow for a jump-timing risk because the λ_y^* can be different from $\lambda_y : \lambda_y^* = \lambda_y \exp(\mu_\pi + \sigma_\pi^2/2)$. In Bates (2000) and Pan (2002), the jump-size intensity is made volatility dependent with one and two factors in volatility.

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

$$\pi_t = \frac{1}{\theta_t} E_t\left[\theta_T (S_T - K)^+\right]. \tag{4.6}$$

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

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

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 detail in the next subsection the various issues raised by the estimation of such a model.

4.1. Econometric Model of Option Pricing Errors

Typically, such 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:

$$Y_t = \{h_i [X_t, \theta]\}_{1 \leq i \leq n} . \tag{4.8}$$

Note that when one loosely says asset "prices", one should rather understand "yields" in the case of bonds or "option premium by unit of spot price" in case of 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 nonstationary. In the context of options on equity, one may also replace (see, e.g., Chernov and Ghysels, 2000; Pastorello et al., 2000; Renault and Touzi, 1996) option prices by the corresponding BS-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 that is more 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 as instantaneous risk free rate, diffusive 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 modeling, this approach is particularly well suited to capture through K explanatory latent factors of the yield curve the relationships between n observed yields in cross-section. A large strand of literature, initiated in particular by Chen and Scott (1993), Pearson and Sun (1994), and Duan (1995), uses this indirect empirical modeling of bond yields through underlying latent factors. In contrast, explicit dynamic modeling 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 et al. (2003).

Of course, the simplest approach to estimating a K factors 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, 1995) 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 leads unmistakably to neglect the 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 option pricing on the S&P 500 index from the joint time series of the index and one near-the-money short dated option on it. One option price is sufficient to get a one-to-one relationship with the volatility factor, yet (see, e.g., Dumas et al., 1998), by taking into account the various possible moneynesses and maturities, the number of fairly liquid option prices on S&P 500 that can be observed at any given date may be about 10 or even more. Similarly, although common models of the yield curve involve $K = 1, 2$, or 3 factors, the number n of available maturities in the cross-section is about 30 or even more.

However, as emphasized by Renault (1997), when the number n of observed asset prices is larger than the number K of latent state variables, this produces some stochastic singularity and statistical estimation theory becomes irrelevant. If one takes the asset pricing model seriously, some parameters can be computed exactly. For example, in the BS case of no latent state variable, observing the price of one option will be enough to compute exactly the volatility of the process. In the case of SV 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. But different option prices would imply different values for the current state of the variance process. This fundamental inconsistency can be resolved either by 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 Rubinstein (1994) implied binomial tree methodology described in Section 5) or by admitting that these formulas are approximative 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 abstracting from microstructure effects and market frictions.

Therefore, the retained empirical specification of the asset pricing model (4.8) will be

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

Note that we consider at this stage that the n assets prices have been relabeled to get zero pricing errors for the K first ones, whereas the $(n-K)$ other ones 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, whereas error terms may be added to the other ones. We just say that because 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. Of course, such a specification needs to know a priori what are the K prices (or the K linear combinations of prices) that 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:

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

4.2. Maximum Likelihood-Based Inference

To present a variety of likelihood-based inference strategies, we follow here the presentation of implied-state maximum likelihood as first proposed by Renault and Touzi (1996) and Renault (1997). Pastorello et al. (2003) encompass a larger set of implied-state methodologies under the name of implied-state backfitting.

The conditional likelihood associated to a data set $\{Y_t, t = 1, \ldots, T\}$ (and an initial conditioning value Y_0) must be derived, through the Jacobian formula, from the latent one associated with the "latent data" set $\left\{Y_t^*, t = 1, \ldots, T\right\}$ produced by the latent realizations of a Markov process Y^* 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, thanks to $(n - K)$ equations that complete the K equations (4.10). A natural idea 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. Of course, their consistent estimation may be useful for improving the accuracy of the estimation of the parameters of interest θ. We do want to ensure, however, 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 Ω 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. Because 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 identified only 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 remain true to this philosophy, a better strategy is to define the latent vector Y_t^* and the associated function $g[Y_t^*, \theta]$ in the following way:

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

Note that $(n - K)$ among the n, so-called latent variables Y_t^* are actually observed, but 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^*:

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

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

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

where the last expression stresses the fact that this density function depends on the value of the unknown parameters only through θ_1. By the no instantaneous causality assumption, $l[V_t \,\big|X_t, Y_{t-1}^*]$ is simply obtained by a translation of size $e[X_t, \theta]$ applied to the conditional probability distribution $l[u_t \,\big|Y_{t-1}^*, \eta]$ of the error terms given the past. This probability density function depends on the value of the unknown parameters only through the nuisance parameters η.

Because 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 we will carry out inference about 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 $\Omega(\eta)$:

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

Several remarks are in order about the use of this quasi-likelihood. First, it 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. Second, we should not forget that the quasi-likelihood may differ from the true likelihood and that we just want to get a consistent estimator of the structural parameters of interest θ. The nuisance parameters η are likely to be poorly defined and not consistently estimated. However, a general specification of the covariance matrix $\Omega(\eta)$ should at least allow us to take into account the obvious strong cross-sectional patterns of correlation and heteroskedasticity among error terms (see Renault, 1997, for a general discussion).

Starting from an estimator η_T of the nuisance parameters and a corresponding estimator $\Omega_T = \Omega(\eta_T)$, we first plug it into (4.13) to define the latent criterion for extremum estimation of the structural parameters θ:

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

Up to recursive refinements, the backfitting (or iterative implied-state) methodology amounts to 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 $X_t\big(\theta^{(p)}\big) = h^{-1}\big[Z_t, \theta^{(p)}\big]$. This defines a sample-based criterion $Q_T\big(\theta, \theta^{(p)}\big)$.
- Compute the estimator $\theta^{(p+1)}$ as $\arg\max_\theta Q_T\big(\theta, \theta^{(p)}\big)$.

Because the nuisance parameters η have been introduced in a way that preserves adaptivity, the resulting asymptotic probability distribution of the backfitting estimator of θ will only depend on the probability limit of Ω_T and not on its accuracy as estimator of the (pseudo) true unknown value of $\Omega(\eta)$. However, at least in case where the conditional distribution of the error terms would be well specified, the most accurate backfitting estimator would be obtained when Ω_T is a consistent estimator of the true value of $\Omega(\eta)$. This is the reason why it is natural to think to a "quasi-generalized" version of backfitting in the following way.

Start from an arbitrary Ω_T (e.g., the identity matrix) and compute the corresponding backfitting estimator θ_T of θ. Then, use it to compute "estimated error terms" $u_t(\theta_T) = V_t - e[X_t(\theta_T), \theta_T]$ and to derive a consistent estimator $\eta(\theta_T)$ of the pseudo true value of η and in turn, a consistent estimator $\Omega_T^* = \Omega[\eta(\theta_T)]$ of the pseudo true value of Ω. Then, perform a second backfitting estimation of θ based on the criterion (4.16) where Ω_T has been replaced by W_T^*. Of course, such a procedure is costly because it implies several backfitting estimations. Fortunately, there exists a much faster procedure, i.e., in terms of estimation of θ, asymptotically equivalent to quasi-generalized backfitting, but in terms of computing time, equivalent to a simple backfitting.

This procedure that we term "extended backfitting" amounts to using each step $\theta^{(p)}$ of the backfitting iteration to compute a new estimator $\Omega[\eta(\theta^{(p)})]$ of the matrix Ω and to plug it into (4.16) in place of Ω_T to derive the next step estimator $\theta^{(p+1)}$ of θ. At first sight, extended backfitting is similar to standard backfitting applied to the augmented vector (θ, η) of unknown parameters. However, we do not refer to a general backfitting theory (in terms of an augmented vector of parameters) to justify this procedure. There is little hope to get a sequence that is contracting with respect to the nuisance parameters η, and this is the reason why the relevant convergence criterion of the approximation sequence for applications will only be based on the norm $||\theta^{(p+1)} - \theta^{(p)}||$.

The relevant argument is the following. Irrespective of the choice of the weighting matrix Ω_T in (4.16), the backfitting estimator is a consistent estimator of the true unknown value of θ. Therefore, it is clear that the limit of the sequence $\theta^{(p)}$ produced by the extended backfitting algorithm also provides a consistent estimator of θ, and in turn, the limit of the sequence $\Omega[\eta(\theta^{(p)})]$ provides a consistent estimator of the true unknown value of $\Omega[\eta]$. Because the asymptotic probability distribution of the backfitting estimator of θ only depends on the probability limit of Ω_T, it is then clear that we get an estimator asymptotically equivalent to the quasi-generalized backfitting. Let us briefly sketch a comparison with the maximum likelihood based competitors also well suited for inference on such 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 has been first proposed in the context of affine models of the yield curve by Duan and Simonato (1999) and systematically developed by De Jong (2000). Of course, severe nonlinearities or nonnormality 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 to be linear as in the case of affine models of the yield curve, the two approaches may be competitors. Roughly speaking, the Kalman filtering approach can be seen as a quick and dirty procedure to check the validity of our possibly more accurate but also more risky approach. Typically, the backfitting approach seeks to get more efficient estimators and filters by taking the risk to specify exact nonlinear relationships 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). Their idea is quite simple. Even though the latent model is conceived to be simpler than the observable one, the hard 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 if we substitute 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 risk of misspecification bias in case of 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 SV.

Moreover, as neatly put forward by Duffee (2002), "another advantage of QML (which it shares with maximum likelihood and related techniques) is that $(\cdots)$ a model estimated with QML will guarantee that the time-t state vector implied by time-t yields is in the state vector's admissible space (to avoid a likelihood zero). By contrast, $(\cdots)$ techniques such as efficient method of moments (EMM) $(\cdots)$ do not require that the estimated term structure model be sufficiently flexible to reproduce the term structure shapes in the data. The parameters of the model in Dai and Singleton (2000), which were estimated with EMM, illustrate this point." This point is actually an important motivation to prefer implied-state-based likelihood rather than simulation-based minimum chi-square competitors like indirect inference or EMM.

As far as efficiency is concerned, several remarks are in order. First, contrary to common belief, the fact that 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 do as if the state variables were directly observable. The crucial point is that the one-to-one relationship (4.12) between latent variables Y^* and observable variables Y does depend on the unknown parameters θ. Therefore, nobody knows whether the Cramer-Rao bound $\left(I^*\right)^{-1}$ 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 described above must not give the fallacious feeling that the Cramer-Rao bound associated with the maximization of the log-likelihood $\sum_{t=1}^{T} \log L\left[Y_t^* \mid Y_{t-1}^*, \theta\right]$ has been reached. This maximization is actually infeasible, and the backfitting iterative scheme is based on the sequence:

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

As shown in Pastorello et al. (2003), the cost of this necessary iteration is to multiply the Cramer-Rao bound $\left(I^*\right)^{-1}$ by a matrix-form factor, which is all the less detrimental than the mapping $\theta^{(p)} \to \theta^{(p+1)}$ is more strongly contracting. This theory is based on a well-defined choice of the number $p(T)$ of iterations (as a function of the sample size T) to define a backfitting estimator $\theta^{p(T)+1}$. Of course, 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

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

but rather

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

where $|Jg^{-1}(Y_t,\theta)|$ denote the absolute value of the Jacobian of the transformation g. 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 even more importantly, the direct maximization of (4.18) will lead to look for a maximizer θ, which should simultaneously meet two requirements. On the one hand, it has to give a large value to the latent likelihood, as it is natural to require. But, on the other hand, θ will tend to be chosen to select the most likely implied-state values $g^{-1}(Y_t,\theta)$. In many circumstances, such a selection appears to be a fairly risky strategy. For instance, Pastorello et al. (2003) observe that in the case of application of Aït-Sahalia (2003) likelihood expansions for affine-type diffusion processes, this will perversely push $g^{-1}(Y_{t-1},\theta)$ toward 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 of likelihood maximization rather than directly maximizing the possibly unpalatable log-likelihood (4.18).

Indirect inference and EMM are often presented as appealing alternatives to maximum likelihood, precisely when the likelihood function becomes unpalatable due to some unobserved state variables. Because 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 et al. (2000) propose to avoid the backfitting iteration by simply using BS-implied volatilities as proxies of implied states in a one-factor SV model. Thanks to the matching of estimated parameter or fitted-score vectors on simulated data, the indirect inference principle (see Gouriéroux et al., 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 although 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, whereas 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) allows one to filter spot volatilities via reprojection, i.e., to 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 (2010) in this handbook.

4.3. Implied-State GMM

Taking advantage of 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 et al. (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^* were observed:

$$E\left[\Psi\left(Y_t^*,\theta\right) \mid Y_{t-1}^*\right] = 0 \tag{4.19}$$

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

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

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

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

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

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

Of course, this estimator is infeasible because Y_t^* is not observed. Then, two strategies may be imagined. The implied-state backfitting of Pastorello et al. (2003) still amounts to replace every occurrence of Y^* in $M_{t-1}(\theta)$ and $\Psi\big(Y_t^*,\theta\big)$ by $g^{-1}\left(Y_t,\theta^{(p)}\right)$ where $\theta^{(p)}$ comes from a previous step estimation. Insofar as such iterations converge, they will converge toward Pan's (2003) IS-GMM estimator, which is actually the second strategy: directly solve (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 $\Psi\big(Y_t^*,\theta\big)$ but also inside any occurrence of $Y_t^*=g^{-1}(Y_t,\theta)$.

By contrast, Pastorello et al. (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. Then, the choice between the two strategies is just a matter of computational convenience, depending whether one consider that the backfitting iterations simplify or not the solution of the IS-GMM fixed point problem.

Moreover, as stressed by Pan (2003) in her discussion of Pastorello et al. (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, same error terms are added and the vector Y^* includes same observed asset prices which do identify the risk premium parameters. Then, implied-state backfitting works. In any case, 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:

$$E\big[\Psi\left(g^{-1}\left(Y_t,\theta\right),\theta\right)\mid Y_{t-1}\big]=0. \tag{4.21}$$

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, although 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, because 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. More contracting it is, smaller is the efficiency loss.

4.4. Joint Estimation of Risk-Neutral and Objective Distributions

The area of joint estimation of risk-neutral and objective measures is probably where most of the progress took place over the last five years. The stage was set in the early 1990s with the considerable advances made regarding estimation of diffusion processes. Exploiting the EMM estimation procedure of Gallant and Tauchen (1996) for the estimation of diffusions, Chernov and Ghysels (2000) propose a generic procedure for estimating and pricing options using simultaneously the fundamental price S_t and a set of option contracts $[(\sigma_{it}^I)_{i=1,m}]$ where $m \geq 1$ and σ_{it}^I is the BS-implied volatility. The procedure consists of two steps. The first one fits a SNP density of $[S_t, (\sigma_{it}^I)_{i=1,m}]$ conditional on its own past $[S_\tau, (\sigma_{i\tau}^I)_{i=1,m}]$ for $\tau < t$. Second, one simulates the fundamental price and option prices and calibrates the parameters of the diffusion and its associated option pricing model to fit the conditional density of the market data dynamics. 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, whereas 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.

Computing the prices of risk involves parameters of the objective measure, the risk-neutral measure, and the latent volatility process. The univariate specifications consist of models only using the fundamental (i.e., the usual setup) and models using only options data. It should be noted, however, that the knowledge of the estimated model parameters is not sufficient to compute an option price or a hedge ratio. We have to know the latent spot volatility as well. Because the option price is a one-to-one function of the current value of the volatility process (Renault and Touzi, 1996), one can recover it via an inversion of the option pricing formula. However, this procedure is computationally cumbersome, except if one relies on approximations by series expansions (Garcia et al., 2009; Lewis, 2000).

Another possible strategy is to use 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.

Pan (2002) examines also a joint time series model of the S&P 500 index and near-the-money short-term option prices in the context of the jump-diffusion model described at the beginning of this section. She uses an implied-state GMM approach to estimate the model. For a given set of model parameters ϑ, she replaces the unobserved volatility V_t by an option-implied volatility V_t^{ϑ} inverted numerically from the spot price S_t and a near-the-money short-term option price π_t based on the option pricing formula implied by the jump-diffusion model.[6] The interest of such a method is to take advantage of the analytical tractability of the state variables S and V compared with the complicated joint dynamics of the pair S and π, given the nonlinear nature of the option pricing function. The usual GMM procedure can be applied to the moments of the pair of state variables S_t and V_t^{ϑ}, but now one of the state variables is parameter-dependent. The closer ϑ is to the true model parameter vector ϑ_0, the more accurate is the corresponding option-implied volatility V_t^{ϑ}.

Garcia et al. (2009) propose an estimation procedure that uses both option prices and high-frequency spot price feeds to estimate jointly the objective and risk-neutral parameters of SV 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 SV 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 (2000), which corresponds to the BS 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)

[6]The numerical procedure to compute the model-based implied volatility is described in Appendix B of Pan (2002).

and option prices for computing implied volatilities (the risk-neutral part of the estimation) allows 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 (2004) 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 because the quantities of interest are functions, whether it is the risk-neutral distribution or 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 et al. (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 BS formula. Next, we will see a maximum entropy approach initiated 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), Dupire (1994), or Rubinstein (1994). Third, we will survey the purely nonparametric approaches such as kerned-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.[7]

[7] See also Yatchew and Härdle (2005), Birke and Pilz (2009), and Fan and Mancini (2008).

Last, we will describe how to recover preferences from the estimates of the SPD as proposed by Jackwerth (2000), Aït-Sahalia and Lo (2000), Rosenberg and Engle (2002), and Chabi-Yo et al. (2008).

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 the S&P 100-based contracts) have been avoided to circumvent early exercise premia and boundaries. It is worth noting that nonparametric methods are particularly suited to handle American-type options. Broadie et al. (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 BS 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. Although 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:

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

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 $_i$ is the BS volatility implied by the observed price of option i. A call pricing function can then be estimated as

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

From this function, one can also obtain estimators for the option's delta and the SPD by taking the appropriate partial derivatives according to (2.6) and (2.12):

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

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

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 BS model, Aït-Sahalia and Lo (1998) show that the estimation errors for all nonparametric quantities (option price, option delta, and SPD) remain within 1% 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 $\widehat{\sigma}(F_{t,\tau}, X, \tau)$ generates a strongly asymmetric volatility smile with respect to moneyness, confirming several sources of evidence according to which out-of-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, whereas the contrary is obtained for the positive kurtosis.

A somewhat less ambitious approach has been advocated by Erikkson et al. (2009) and Ghysels and Wang (2009). They suggest to use the normal inverse Gaussian (NIG) family to approximate an unknown distribution risk-neutral density. The appeal of the NIG family of distributions is that they are characterized by the first four moments: mean, variance, skewness, and kurtosis. These are the moments we care about in many applications – including derivative pricing. The unknown density function is approximated by matching the cumulants. The latter are obtained from the cross-section of option prices using methods proposed by Bakshi et al. (2003). One strength of their approach is that they link the pricing of individual derivatives to the moments of the risk-neutral distribution, which has an intuitive appeal in terms of how volatility, skewness, and kurtosis of the risk-neutral distribution can explain the behavior of the derivative prices. Erikkson et al. (2009) show that the approximation errors are minor when compared to several option pricing models that have known densities. Another approach, advocated by Figlewski (2009), consists of estimating the central part of the distribution only with options and extrapolating the tails via extreme value distributions.

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.[8] But, as we have extensively documented in the previous sections, there is evidence of jumps and SV in the underlying stock index process. Therefore, we need procedures that extract the asset probability distribution directly from observed prices either on the asset itself or on

[8]In fact, the semiparametric approach could also be valid for i.i.d. jump processes as in Merton (1976) or Bates (1991)

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 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 $S(t)$, $t = -1, -2, \ldots, -H$, one can construct a rolling historical time series of T-period gross returns:

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

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

$$S^h = SR(-h), \quad h = 1, 2, \ldots, H-T. \tag{5.6}$$

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 $\widehat{p}(h) = \frac{1}{H-T}$. The problem is to find the risk-neutral probabilities p^*, which are the closest to the empirical probabilities $\widehat{p}$ in the Kullback–Leibler Information Criterion (KLIC) distance:

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

and which obey the nonarbitrage economic constraint (assuming a constant interest rate):

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

The solution to this problem is

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

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

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

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

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

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

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

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 3 to 6 month range for $x > 0.02$ using only the historical data on S&P 500 returns starting in 1987 and without incorporating market option prices in the valuation process.[9]

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

[9] Gray and Norman (2005) apply canonical valuation of options in the presence of SV. Haley and Walker (2007) propose alternative tilts (or probability distortions) based on the Cressie-Read divergence family.

risk-neutral probabilities say $\widetilde{p}_j^*$ and find the risk-neutral probabilities p_j^* associated with the binomial terminal stock price S_T that are the closest to $\widetilde{p}_j^*$ but price correctly an existing set of options and the underlying stock. The risk-neutral probabilities p_j^* are solutions to the following program:

$$\min_{p_j^*} \sum_j \left(p_j^* - \widetilde{p}_j^*\right)^2 \text{ subject to} \tag{5.13}$$

$$\sum_j p_j^* = 1 \text{ and } p_j^* \geq 0 \text{ for } j = 0, \ldots, n$$

$$S^b \leq S \leq S^a \text{ where } S = \left(\sum_j p_j^* S_j\right) \Big/ r^\tau$$

$$C_i^b \leq C_i \leq C_i^a \text{ where } C_i = \left(\sum_j p_j^* \max[0, S_j - K_i]\right) \Big/ r^\tau \text{ for } i = 1, \ldots, m,$$

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, C_i^a and C_i^b are 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.[10] Although the maximum entropy criterion appears the best one from a theoretical point of view, because 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 $\min_{p_j^*} \sum_j (p_j^* - \widetilde{p}_j^*)^2 / \widetilde{p}_j^*$ 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 by minimizing $\sum_j (p_{j-1}^* - 2p_j^* + p_{j+1}^*)^2$, as in Jackwerth and Rubinstein (1996) 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.

[10] Cont and Tankov (2004) use a relative entropy criterion with respect to a chosen prior model to find a risk-neutral exponential Lévy model that reproduces observed option prices.

5.2.3. A SDF Alternative to Implied Binomial Trees

One might also measure closeness as the distance between pricing kernels and not between risk-neutral probabilities by looking for the SDF m^*_{t+1} defined by

$$m^*_{i,t+1} = B(t, t+1)\left(\frac{p^*_{it}}{p_{it}}\right), \quad i = 0, 1, \ldots, I+1$$

which is closest to a prior SDF

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

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

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

Therefore, the Hansen and Jagannathan (1997) measure of closeness (5.14) between SDFs and the goodness of fit criterion between probabilities $\sum_{i=0}^{I+1}(1/\tilde{p}^*_{it})\left(p^*_{it} - \tilde{p}^*_{it}\right)^2$ 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 (2004).

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 because these small probabilities appear at the denominator of SDFs and therefore, have a large weight in the effective computation of derivative asset prices. Almeida and Garcia (2008) generalize the quadratic Hansen and Jagannathan (1997) measure of closeness by choosing the Cressie-Read family of discrepancy measures. Because this family includes the KLIC and the empirical likelihood divergence criteria, this extension makes clear the links between all the nonparametric approaches adopted to recover risk-neutral probabilities or pricing kernels to price options.

All of 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

Because 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. Because we cannot really say that one approach is better than the other, we can only sketch the advantages and disadvantages of both methods following Aït-Sahalia and Lo (1998).

We will compare the implied binomial tree method of Rubinstein (1994) to the semiparametric 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, whereas the second exploits the dependencies in the data around a given date. Implied binomial trees are less data-intensive, whereas 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 with 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.

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 2 months to maturity on a day in 1999, they compare several estimators (unconstrained Nadaraya–Watson, unconstrained locally linear, quadratic and cubic, shape-constrained locally linear) in terms of price function, first derivative with respect to the strike price and SPD (second derivative). 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, whereas 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 violate the first derivative constraint, whereas for the SPD (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 subsections, we will revisit these constrained and unconstrained approaches in the SNP context. A first way to enforce the shape restrictions is to use a parametric model for the SDF while remaining nonparametric for the historical distribution. It is the main motivation of the Extended Method of Moments (XMM). A second strategy is to directly fit a SNP model for the option pricing function. Then sieve estimators and especially neural networks are well suited to take into account shape restrictions.

5.4. Extended Method of Moments

The GMM was introduced by Hansen (1982) and Hansen and Singleton (1982) to estimate a structural parameter θ identified by Euler conditions:

$$p_{i,t} = E_t\big[M_{t,t+1}(\theta)p_{i,t+1}\big], \quad i = 1, \ldots, n, \ \forall t, \tag{5.15}$$

where $p_{i,t}, i = 1, \ldots, n$, are the observed prices of n financial assets, E_t denotes the expectation conditional on the available information at date t, and $M_{t,t+1}(\theta)$ is the stochastic discount factor. Model (5.15) is semiparametric. The GMM estimates parameter θ regardless of the conditional distribution of the state variables. This conditional distribution however becomes relevant when the Euler conditions (5.15) are used for pricing derivative assets. Indeed, when the derivative payoff is written on $p_{i,t+1}$ and its current price is not observed on the market, the derivative pricing requires the joint estimation of parameter θ and the conditional distribution of the state variables.

The XMM estimator of Gagliardini et al. (2008) extends the standard GMM to accommodate a more general set of moment restrictions. The standard GMM is based on uniform conditional moment restrictions such as (5.15), which are valid for any value of the conditioning variables. The XMM can handle uniform moment restrictions, as well as local moment restrictions, that are only valid for a given value of the conditioning variables. This leads to a new field of application to derivative pricing, as the XMM can be used for reconstructing the pricing operator on a given day, by using the information in a cross section of observed traded derivative prices and a time series of underlying asset returns. To illustrate the principle of XMM, consider an investor at date t_0 is interested in estimating the price $c_{t_0}(h, k)$ of a call option with time-to-maturity h and moneyness strike k that is currently not (actively) traded on the market. She has data on a time series of T daily returns of the S&P 500 index, as well as on a small cross section of current option prices $c_{t_0}(h_j, k_j)$, $j = 1, \ldots, n$, of n highly traded derivatives. The XMM approach provides the estimated prices $\hat{c}_{t_0}(h, k)$ for different values of moneyness strike k and time-to-maturity h, which interpolate the observed prices of highly traded derivatives and satisfy the hypothesis of absence of arbitrage opportunities. These estimated prices are consistent for a large number of dates T, but a fixed, even small, number of observed derivative prices n.

We are interested in estimating the pricing operator at a given date t_0, i.e., the mapping that associates any European call option payoff $\varphi_{t_0}(h, k) = \left(\exp R_{t_0,h} - k\right)^+$ with its price $c_{t_0}(h, k)$ at time t_0, for any time-to-maturity h and any moneyness strike k. We denote by r_t the logarithmic return of the underlying asset between dates $t-1$ and t. We assume that the information available to the investors at date t is generated by the random vector X_t of state variables with dimension d, including the return r_t as the first component, and that X_t is also observable by the econometrician. The process (X_t) on $\mathcal{X} \subset \mathbf{R}^d$ is supposed to be strictly stationary and Markov under the historical probability with transition density $f(x_t|x_{t-1})$. Besides the cross section of option prices $c_{t_0}(h_j, k_j)$, $j = 1, \ldots, n$ the available data consist in T serial observations of the state variables X_t corresponding to the current and previous days $t = t_0 - T + 1, \ldots, t_0$. The no-arbitrage assumption implies two sets of moment restrictions for the observed asset prices. The constraints concerning the observed derivative prices at t_0 are given by

$$c_{t_0}(h_j, k_j) = E\left[M_{t,t+h_j}(\theta)(\exp R_{t,h_j} - k_j)^+ | X_t = x_{t_0}\right], \quad j = 1, \ldots, n. \tag{5.16}$$

The constraints concerning the risk free asset and the underlying asset are

$$\begin{cases} E[M_{t,t+1}(\theta) | \; X_t = x] = B(t, t+1), & \forall x \in \mathcal{X}, \\ E[M_{t,t+1}(\theta) \exp r_{t+1} | \; X_t = x] = 1, & \forall x \in \mathcal{X}, \end{cases} \tag{5.17}$$

respectively, where $B(t, t+1)$ denotes the price at time t of the short-term risk free bond. The conditional moment restrictions (5.16) are local because they hold for a single value of the conditioning variable only, namely the value x_{t_0} of the state variable at time t_0. This is because we consider only observations of the derivative prices $c_{t_0}(h_j, k_j)$ at date t_0.

Conversely, the prices of the underlying asset and the risk free bond are observed for all trading days. Therefore, the conditional moment restrictions (5.17) hold for all values of the state variables. They are called the uniform moment restrictions. The distinction between the uniform and local moment restrictions is a consequence of the differences between the trading activities of the underlying asset and its derivatives. Technically, it is the essential feature of the XMM that distinguishes this method from its predecessor GMM.

The XMM estimator presented in this section is related to the literature on the information-based GMM (e.g., Imbens et al., 1998; Kitamura and Stutzer, 1997). It provides estimators of both the SDF parameter θ and the historical transition density $f(y|x)$. By using the parameterized SDF, the information-based estimator of the historical transition density defines the estimated SPD for pricing derivatives.

The XMM approach involves a consistent nonparametric estimator of the historical transition density $f(y|x)$, such as the kernel density estimator:

$$\hat{f}(y|x) = \frac{1}{h_T^{\tilde{d}}} \sum_{t=1}^{T} \tilde{K}\left(\frac{y_t - y}{h_T}\right) K\left(\frac{x_t - x}{h_T}\right) \Big/ \sum_{t=1}^{T} K\left(\frac{x_t - x}{h_T}\right), \tag{5.18}$$

where K (resp. $\tilde{K}$) is the d-dimensional (resp. $\tilde{d}$-dimensional) kernel, h_T is the bandwidth, and (x_t, y_t), $t = 1, \ldots, T$, are the historical sample data.[11] Next, this kernel density estimator is improved by selecting the conditional pdf that is the closest to $\widehat{f}(y|x)$ and satisfies the moment restrictions as defined below.

The XMM estimator $\left(\hat{f}^*(\cdot|x_0), \hat{f}^*(\cdot|x_1), \ldots, \hat{f}^*(\cdot|x_T), \widehat{\theta}\right)$ consists of the functions $f_0, f_1, \ldots, f_T$ defined on $Y \subset R^{\tilde{d}}$, and the parameter value θ that minimize the objective function:

$$L_T = \frac{1}{T} \sum_{t=1}^{T} \int \frac{\left[\widehat{f}(y|x_t) - f_t(y)\right]^2}{\widehat{f}(y|x_t)} \mathrm{d}y + h_T^d \int \log\left[\frac{f_0(y)}{\widehat{f}(y|x_0)}\right] f_0(y)\mathrm{d}y,$$

subject to the constraints:

$$\int f_t(y)\mathrm{d}y = 1, \quad t = 1, \ldots, T, \qquad \int f_0(y)\mathrm{d}y = 1,$$

$$\int g\left(y;\theta\right) f_t(y)\mathrm{d}y = 0, \quad t = 1, \ldots, T, \qquad \int g_2\left(y;\theta\right) f_0(y)\mathrm{d}y = 0. \tag{5.19}$$

[11] For expository purpose, the dates previous to t_0, at which data on (X, Y) are available, have been reindexed as $t = 1, \ldots, T$ and accordingly the asymptotics in T correspond to a long history before t_0.

The objective function L_T has two components. The first component involves the chi-square distance between the density f_t and the kernel density estimator $\widehat{f}(.|x_t)$ at any sample point x_t, $t = 1, \ldots, T$. The second component corresponds to the KLIC between the density f_0 and the kernel estimator $\widehat{f}(.|x_0)$ at the given value x_0. In addition to the unit mass restrictions for the density functions, the constraints include the uniform moment restrictions written for all sample points and the whole set of local moment restrictions. The combination of two types of discrepancy measures is motivated by computational and financial reasons. The chi-square criterion evaluated at the sample points allows for closed form solutions $f_1(\theta), \ldots, f_T(\theta)$ for a given θ. Therefore, the objective function can be easily concentrated with respect to functions $f_1, \ldots, f_T$, which reduces the dimension of the optimization problem. The KLIC criterion evaluated at x_0 ensures that the minimizer f_0 satisfies the positivity restriction (see, e.g., Kitamura and Stutzer, 1997; Stutzer, 1996). The positivity of the associated SPD at t_0 guarantees the absence of arbitrage opportunities in the estimated derivative prices. The estimator of $\hat{\theta}$ minimizes the concentrated objective function:

$$\mathcal{L}_T^c(\theta) = \frac{1}{T}\sum_{t=1}^{T} \widehat{E}\big(g(\theta)|x_t\big)' \, \widehat{V}\big(g(\theta)|x_t\big)^{-1} \widehat{E}\left(g(\theta)|x_t\right) - h_T^d \log \widehat{E}\big(\exp\big(\lambda(\theta)' g_2(\theta)\big)\,|x_0\big), \quad (5.20)$$

where the Lagrange multiplier $\lambda(\theta) \in \mathbf{R}^{n+2}$ is such that

$$\widehat{E}\left[g_2(\theta)\exp\left(\lambda\,(\theta)'\,g_2(\theta)\right)|x_0\right] = 0, \quad (5.21)$$

for all θ, and $\widehat{E}\left(g(\theta)|x_t\right)$ and $\widehat{V}\left(g(\theta)|x_t\right)$ denote the expectation and variance of $g(Y;\theta)$, respectively, w.r.t. the kernel estimator $\widehat{f}(y|x_t)$. The first part of the concentrated objective function (5.20) is reminiscent from the conditional version of the continuously updated GMM (Ai and Chen, 2003; Antoine et al., 2007). The estimator of $f(y|x_0)$ is given by

$$\widehat{f}^*(y|x_0) = \frac{\exp\left(\lambda\left(\hat{\theta}\right)' g_2(y;\hat{\theta})\right)}{\widehat{E}\left[\exp\left(\lambda\left(\hat{\theta}\right)' g_2(\hat{\theta})\right)|x_0\right]}\widehat{f}(y|x_0), \;\; y \in \mathcal{Y}. \quad (5.22)$$

This conditional density is used to estimate the pricing operator at time t_0.

The XMM estimator of the derivative price $c_{t_0}(h,k)$ is

$$\hat{c}_{t_0}(h,k) = \int M_{t_0,t_0+h}(\hat{\theta})\left(\exp R_{t_0,h} - k\right)^+ \hat{f}^*\left(y|x_0\right) \mathrm{d}y, \quad (5.23)$$

for any time-to-maturity $h \leq \bar{h}$ and any moneyness strike k. The constraints (5.19) imply that the estimator $\hat{c}_{t_0}(h,k)$ is equal to the observed option price $c_{t_0}(h_j,k_j)$ when $h = h_j$ and $k = k_j, j = 1,\ldots,n$.

The large sample properties of estimators $\hat{\theta}$ and $\hat{c}_{t_0}(h,k)$ are examined in Gagliardini et al. (2008). These estimators are consistent and asymptotically normal for large samples T of the time series of underlying asset returns, but a fixed number n of observed derivative prices at t_0. The linear combinations of θ that are identifiable from uniform moment restrictions on the risk free asset and the underlying asset only are estimated at the standard parametric rate $\sqrt{T}$. Any other direction η_2^* in the parameter space and the derivative prices as well are estimated at the rate $\sqrt{Th_T^d}$ corresponding to nonparametric estimation of conditional expectations given $X = x_0$. The estimators of derivative prices are (nonparametrically) asymptotically efficient.

5.5. Other SNP Estimators

In the SNP 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:

$$f(Z,.) = \sum_{n=1}^{\infty} \alpha_n g_n(Z,.). \tag{5.24}$$

The term SNP is explained by the fact that the basis functions are parametric, yet the parameters are not the object of interest because 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 et al. (1994) propose various types of learning networks, Gouriéroux and Monfort (2001) consider approximations of the pricing kernel through splines, whereas Abadir and Rockinger (1998) investigate hypergeometric functions. In what follows, we will develop the neural network approach and see how one can choose the basis to obtain a valid SPD function. The basis chosen for neural networks will be

$$g_n(Z,\alpha_n) = \frac{1}{1+\exp(-\alpha_n Z)}, \tag{5.25}$$

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

$$f(Z,\theta) = \beta_0 + \sum_{i=1}^{d} \beta_i \frac{1}{1+\exp(\gamma_{i,0} - \gamma_{i,1} Z)}, \tag{5.26}$$

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

$$h'(Z) = h(Z).(1 - h(Z))$$
$$h''(Z) = h(Z).(1 - h(Z)).(1 - 2h(Z)).$$

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 et al. (1994) show using simulations that such an approach can learn the BS formula. To reduce the number of inputs, Hutchinson et al. (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):

$$\frac{\pi_t}{X} = f\left(\frac{S_t}{X}, 1, \tau\right). \tag{5.27}$$

Although 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).[12] Although it is not mentioned in Hutchinson et al. (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 a long-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 (5.27) assumes the homogeneity of degree

[12] Gençay and Qi (2001) studied the effectiveness of cross-validation, Bayesian regularization, early stopping, and bagging to mitigate overfitting and improving generalization for pricing and hedging derivative securities.

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, whereas the variable S_t/X is stationary as strike prices bracket the underlying asset price process. This point is emphasized in Ghysels et al. (1996). 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 Gençay (2000) estimate a network of the form

$$\frac{C_t}{X} = \beta_0 + \sum_{i=1}^{d} \beta_i^1 h\left(\gamma_{i,0}^1 + \gamma_{i,1}^1 \frac{S_t}{X} + \gamma_{i,2}^1 \tau\right) \tag{5.28}$$

$$- \mathrm{e}^{-\alpha\tau} \sum_{i=1}^{d} \beta_i^2 h\left(\gamma_{i,0}^2 + \gamma_{i,1}^2 \frac{S_t}{X} + \gamma_{i,2}^2 \tau\right) \tag{5.29}$$

with $h(Z) = \left(1 + \mathrm{e}^Z\right)^{-1}$. This has a similar structure than the BS formula (which is itself homogeneous), except that the distribution function of the normal is replaced by neural network functions.[13] Garcia and Gençay (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 et al. (2001)

$$\xi(Z) = \log(1 + \mathrm{e}^Z) \tag{5.30}$$

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

$$\xi'(Z) = \frac{\mathrm{e}^Z}{1 + \mathrm{e}^Z} = h(Z) \tag{5.31}$$

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

$$\xi''(Z) = h'(Z) = h(Z).(1 - h(Z)) \tag{5.32}$$

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

[13] This is what distinguishes this SNP approach from the semiparametric approach of Aït-Sahalia and Lo (2000), who use the BS formula with a nonparametric estimator of volatility.

Abadir and Rockinger (1998) with hypergeometric functions, Gottschling et al. (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, i.e., 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. Birke and Pilz (2009) propose a completely kernel-based estimate of the call price function, which fulfills all constraints given by the no-arbitrage principle. Fan and Mancini (2008) propose a new nonparametric method for pricing options based on a nonparametric correction of pricing errors induced by a given model.

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. Daglish (2003) also provides a comparison between parametric and nonparametric methods for American options.

5.6. An Economic Application of Nonparametric Methods: Extraction of Preferences

Because, in a continuum of states, the SPD 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 SPD 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 SPD and the conditional objective distribution. A measure of relative risk aversion is given by

$$\rho_t(S_T) = S_T\left(\frac{f_t'(S_T)}{f_t(S_T)} - \frac{f_t^{*\prime}(S_T)}{f_t^*(S_T)}\right), \tag{5.33}$$

where $f_t(S_T)$ and $f_t^*(S_T)$ denote, respectively, the conditional objective probability density and the SPD. 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.

Several researchers have extracted risk aversion functions or preference parameters from observed asset prices. Aït-Sahalia and Lo (2000) and Jackwerth (2000) have proposed nonparametric approaches to recover risk aversion functions across wealth states from observed stock and option prices. Rosenberg and Engle (2002), Garcia et al. (2003), and Bliss and Panigirtzoglou (2004) have estimated preference parameters based on parametric asset pricing models with several specifications of the utility function.

These efforts to exploit prices of financial assets to recover fundamental economic parameters have produced puzzling results. Aït-Sahalia and Lo (2000) find that the nonparametrically implied function of relative risk aversion varies significantly across the range of S&P 500 index values, from 1 to 60, and is U-shaped. Jackwerth (2000) finds also that the implied absolute risk aversion function is U-shaped around the current forward price but even that it can become negative. Parametric empirical estimates of the coefficient of relative risk aversion also show considerable variation. Rosenberg and Engle (2002) report values ranging from 2.36 to 12.55 for a power utility pricing kernel across time, whereas Bliss and Panigirtzoglou (2004) estimate average values between 2.33 and 11.14 for the same S&P 500 index for several option maturities.[14] Garcia et al. (2003) estimate a consumption-based asset pricing model with regime-switching fundamentals and Epstein and Zin (1989) preferences. The estimated parameters for risk aversion and intertemporal substitution are reasonable with average values of 0.6838 and 0.8532, respectively, over the 1991–1995 period.[15]

As noticed by Rosenberg and Engle (2002), the interpretation of the risk aversion function is debatable because 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

$$m_{t+1}^{*} = E_t\left[\frac{m_{t+1}}{g_{t+1}}\right] = \theta_{0t}(g_{t+1})^{-\theta 1t}. \tag{5.34}$$

The parameters of interest θ_{0t} and θ_{1t} are then estimated at each date t to minimize the sum of squared pricing errors, i.e., 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 (5.34). As in the multinomial example, 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

[14] Rosenberg and Engle (2002) also estimate an orthogonal polynomial pricing kernel and find that it exhibits some of the risk-aversion characteristics noted by Jackwerth (2000), with a region of negative absolute risk aversion over the range from 4 to 2% for returns and an increasing absolute risk aversion for returns greater than 4%.

[15] The authors also estimate a CCRA-expected utility model and find a similar variability of the estimates as in the related studies. The average value is 7.2 over the 1991–1995 period with a standard deviation of 4.83.

is probably not optimal to rank various misspecified SDFs because 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 because it is model independent, whereas the optimal GMM weighting matrix is model dependent and asymptotic chi-square, 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 (5.34) 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:

$$m_{t+1} = \beta \left[\frac{C_{t+1}}{C_t} \right]^{\gamma(\rho-1)} \left[\frac{W_{t+1}}{(W_t - C_t)} \right]^{\gamma-1}, \tag{5.35}$$

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 on the information J_t available at time t. Therefore,

$$m_{t+1} = \beta \left[\frac{W_{t+1}}{W_t} \right]^{-a} \left[\frac{A(J_{t+1})}{A(J_t)} \right]^{1-a} \left[1 - A(J_t)^{1-\sigma} \right]^{\gamma-1}. \tag{5.36}$$

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 that lead the agent to a nonmyopic 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 that are computed without a precise account of the state variables and enter into the value function $A(J_t)$ cannot provide valuable insights on intertemporal preferences. For example, Chabi-Yo et al. (2008) show that in an economy with regime changes either in fundamentals or in 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 (2007) 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 SDF 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. Because 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 because the estimation of continuous-time models is the object of another chapter in this Handbook.

Another major thread that 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 modeling 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 to price other derivatives on the same underlying asset? This is important both for estimation and for 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, AJD 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 retrospective by Bates (2003) addresses some of these issues.

ACKNOWLEDGMENTS

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. We would like to thank Lars Hansen for invaluable comments on a previous draft of this chapter.

REFERENCES

Abadir, K.M. and M. Rockinger (1998) Density-Embedding Functions, Working Paper, University of York and Groupe HEC.

Ai, C. and X. Chen (2003) Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions, *Econometrica*, 71, 1795–1843.

Aït-Sahalia, Y. (2003) Telling From Discrete Data Whether the Underlying Continuous-Time Model is a Diffusion, *Journal of Finance*, 57, 2075–2112.

Aït-Sahalia, Y. and J. Duarte (2003) Nonparametric Option Pricing Under Shape Restrictions, *Journal of Econometrics*, 116, 9–47.

Aït-Sahalia, Y. and R. Kimmel (2007) Maximum Likelihood of Stochastic Volatility Models, *Journal of Financial Economics*, 83, 413–452.

Aït-Sahalia, Y. and A.W. Lo (1998) Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices, *Journal of Finance*, 53, 499–547.

Aït-Sahalia, Y. and A.W. Lo (2000) Nonparametric Risk Management and Implied Risk Aversion, *Journal of Econometrics*, 94, 9–51.

Aït-Sahalia, Y., L.P. Hansen and J. Scheinkman (2010) Operator Methods for Continuous-Time Markov Processes, in Y. Aït Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, Amsterdam: North-Holland. In press.

Almeida, C. and R. Garcia (2008) Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators, mimeo, FGV (Rio de Janeiro) and Edhec Business School.

Antoine, B., H. Bonnal and E. Renault (2007) On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Euclidean Empirical Likelihood, *Journal of Econometrics*, 138, 461–487.

Alizadeh, S., M. Brandt and F.X. Diebold (2002) Range-based Estimation of Stochastic Volatility Models, *Journal of Finance*, 57, 1047–1091.

Amin, K.I. and R. Jarrow (1992) Pricing Options in a Stochastic Interest Rate Economy, *Mathematical Finance*, 3, 1–21.

Amin, K.I. and V.K. Ng (1993) Option Valuation with Systematic Stochastic Volatility, *Journal of Finance*, 48, 881–910.

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.

Andersen, T.G., T. Bollerslev and F.X. Diebold (2010) Parametric and Nonparametric Volatility Measurement, in Y. Aït Sahalia and L. Hansen (eds.), *Handbook of Financial Econometrics*, Amsterdam: North-Holland. In press.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996) Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 74, 3–30.

Bakshi, G. and D.B. Madan (1999) Crash Discovery in Stock and Options Markets, Working Paper, University of Maryland.

Bakshi, G. and D. Madan (2000) Spanning and Derivative-Security Valuation, *Journal of Financial Economics*, 55, 205–238.

Bakshi, G., C. Cao and Z. Chen (1997) Empirical Performance of Alternative Option Pricing Models, *Journal of Finance*, 94, 1–2, 277–318.

Bakshi, G., N. Kapadia and D. Madan (2003) Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options, *Review of Financial Studies*, 16, 101–143.

Bandi, F. and B. Perron (2006) Long Memory and the Relation Between Implied and Realized Volatility, *Journal of Financial Econometrics*, 4, 636–670.

Bansal, R. and S. Viswanathan (1993) No Arbitrage and Arbitrage Pricing: A New Approach, *Journal of Finance*, 4, 1231–1262.

Banz, R. and M. Miller (1978) Prices for State-Contingent Claims: Some Estimates and Applications, *Journal of Business*, 51, 653–672.

Barndorff-Nielsen, O.E. (1998) Processes of Normal Inverse Gaussian Type, *Finance and Stochastics*, 2, 41–68.

Barndorff-Nielsen, O.E. (2001) Superposition of Ornstein-Uhlenbeck Type Processes, *Theory of Probability and Its Applications*, 45(2), 175–194.

Barndorff-Nielsen, O.E. and N. Shephard (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.

Bates, D. (1991) The Crash of '87: Was it Expected? *Journal of Finance*, 46, 1009–1044.

Bates, D. (1996a) Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options, *Review of Financial Studies*, 9, 69–107.

Bates, D. (1996b) Testing Option Pricing Models, in G.S. Maddala and C.R. Rao (eds.), *Statistical Methods in Finance (Handbook of Statistics, vol. 14)*, Amsterdam: Elsevier.

Bates, D. (2000) Post-'87 Crash Fears in the S&P 500 Futures Option, *Journal of Econometrics*, 94, 181–238.

Bates, D. (2003) Empirical Option Pricing: A Retrospection, *Journal of Econometrics*, 116, 387–404.

Bates, D. (2007) The Market for Crash Risk, *Journal of Economic Dynamics and Control*, 32, 2291–2321.

Benzoni, L. (1998) Pricing Options Under Stochastic Volatility: An Econometric Analysis, Unpublished Manuscript, Kellog Graduate School of Management.

Bertsimas, D., L. Kogan and A. Lo (2000) When is Time Continuous? *Journal of Financial Economics*, 55, 173–204.

Birke, M. and K. Pilz (2009) Nonparametric Option Pricing with No-Arbitrage Constraints, *Journal of Financial Econometrics*, 7, 53–76.

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.

Black, F. and M. Scholes (1972) The Valuation of Option Contracts and a Test of Market Efficiency, *Journal of Finance*, 27, 399–417.

Black, F. and M. Scholes (1973) The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81, 637–659.

Bliss, R.R. and N. Panigirtzoglou (2004) Option-implied Risk Aversion Estimates, *Journal of Finance*, 59, 407–446.

Bollerslev, T. and H.O. Mikkelsen (1996) Modelling and Pricing Long Memory in Stock Market Volatility, *Journal of Econometrics*, 73, 151–184.

Bollerslev, T. and H.O. Mikkelsen (1999) Long-Term Equity Anticipation Securities and Stock Market Volatility Dynamics, *Journal of Econometrics*, 92, 75–99.

Bondarenko, O. (2003) Estimation of Risk-neutral Densities Using Positive Convolution Approximation, *Journal of Econometrics*, 116, 85–112.

Boyle, P.P. (1988) A Lattice Framework for Option Pricing with Two State Variables, *Journal of Financial and Quantitative Analysis*, 23, 1–12.

Breeden, D. and R. Litzenberger (1978) Prices of State Contingent Claims Implicit in Option Prices, *Journal of Business*, 51, 621–651.

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.

Brennan, M.J. (1979) The Pricing of Contingent Claims in Discrete-Time Models, *Journal of Finance*, 34, 53–68.

Broadie, M. and Ö. Kaya (2006) Exact Simulation of Stochastic Volatility and Other Affine Jump Diffusion Processes, *Operations Research*, 54, 217–231.

Broadie, M., J. Detemple, E. Ghysels and O. Torrès (2000a) American Options with Stochastic Dividends and Volatility: A Nonparametric Investigation, *Journal of Econometrics*, 94, 53–92.

Broadie, M., J. Detemple, E. Ghysels and O. Torrès (2000b) Nonparametric Estimation of American Options Exercise Boundaries and Call Prices, *Journal of Economic Dynamics and Control*, 24, 1829–1857.

Brockwell, P. (2009) Lévy-Driven Continuous-Time ARMA Processes, in Andersen et al. (eds.), *Handbook of Financial Time Series*, Springer Verlag, 457–479.

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.

Buraschi, A. and J. Jackwerth (2001) The Price of a Smile: Hedging and Spanning in Option Markets, *Review of Financial Studies*, 14, 495–527.

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.

Carmona, P. and L. Coutin (1999) Stochastic Integrated with Respect to Fractional Brownian Motion, Preprint.

Carr, P. and D. Madan (1999) Option Pricing and the Fast Fourier Transform, *Journal of Computational Finance*, 2(4), 61–73.

Carr, P. and L. Wu (2004) Time-changed Lévy Processes and Option Pricing, *Journal of Financial Economics*, 71, 113–142.

Carr, P., H. Geman, D.B. Madan and M. Yor (2003) Stochastic Volatility for Lévy Processes, *Mathematical Finance*, 13(3), 345–382.

Chabi-Yo, F., R. Garcia and E. Renault (2008) State Dependence Can Explain the Risk Aversion Puzzle, *Review of Financial Studies*, 21, 973–1011.

Chacko, G. and L. Viceira (1999) Spectral GMM Estimation of Continuous-Time Processes, Working Paper, Graduate School of Business, Harvard University.

Chamberlain, G. and M. Rothschild (1983) Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica*, 51, 1281–1304.

Chen, R-R. and L. Scott (1992) Pricing Interest-Rate Options in a Two Factor Cox–Ingersoll–Ross Model of the Term Structure, *Review of Financial Studies*, 5, 613–636.

Chernov, M. (2007) On the Role of Risk Premia in Volatility Forecasting, *Journal of Business and Economic Statistics*, 25, 411–426.

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, 407–458.

Chernov, M., R. Gallant, E. Ghysels and G. Tauchen (2000) A New Class of Stochastic Volatility Models with Jumps: Theory and Estimation, Discussion Paper.

Chernov, M., R. Gallant, E. Ghysels and G. Tauchen (2003) Alternative Models for Stock Price Dynamics, *Journal of Econometrics*, 116, 225–257.

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.

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

Cochrane, J. (2001) *Asset Pricing*, Princeton, NJ: Princeton University Press.

Comte, F. and E. Renault (1996) Non-Causality in Continuous Time Models, *Econometric Theory*, 12, 215–256.

Comte, F. and E. Renault (1998) Long Memory Continuous Time Stochastic Volatility Models, *Mathematical Finance*, 8, 291–323.

Comte, F., L. Coutin and E. Renault (2001) Affine Fractional Volatility Models, Discussion Paper, Université de Montréal.

Constantinides, G. (1992) A Theory of Nominal Term Structure of Interest Rates, *Review of Financial Studies*, 5, 531–552.

Cont, R. and P. Tankov (2004) Non-parametric Calibration of Jump-Diffusion Option Pricing Models, *Journal of Computational Finance*, 7, 1–49.

Cox, J.C., S.A. Ross and M. Rubinstein (1979) Option Pricing: A Simplified Approach, *Journal of Financial Economics*, 7, 229–263.

Daglish, T. (2003) A Pricing and Hedging Comparison of Parametric and Nonparametric Approaches for American Index Options, *Journal of Financial Econometrics*, 1, 327–364.

Dai, Q. and K. Singleton (2000) Specification Analysis of Affine Term Structure Models, *Journal of Finance*, 55, 1943–1978.

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.

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.

De Jong, F. (2000) Time Series and Cross Section Information in Affine Term Structure Models, *Journal of Business and Economic Statistics*, 18(3), 300–314.

Derman, E. and I. Kani (1994) Riding on the Smile, *Risk*, 7, 32–39.

Ding, Z., C.W.J. Granger and R.F. Engle (1993) A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance*, 1, 83–106.

Duan, J.C. (1995) The GARCH Option Pricing Model, *Mathematical Finance*, 5, 13–32.

Duan, J.C. (2001) Risk Premium of Derivatives in Complete Markets, Discussion Paper, Rotman School of Management, University of Toronto.

Duan, J.C. and J.G. Simonato (1999) Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter, *Review of Quantitative Finance and Accounting*, 13, 111–135.

Duffie, D. (2001) *Dynamics Asset Pricing*, Princeton, NJ: Princeton University Press.

Duffie, D. and R. Kan (1996) A Yield-Factor Model of Interest Rates, *Mathematical Finance*, 6, 379–406.

Duffie, D., J. Pan and K. Singleton (2000) Transform Analysis and Asset Pricing for Affine Jump-Diffusions, *Econometrica*, 68, 1343–1376.

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*, 13.

Dumas, B., J. Fleming and R.E. Whaley (1998) Implied Volatility Functions: Empirical Tests, *Journal of Finance*, 53(6), 2059–2106.

Dupire, B. (1994) Pricing with a Smile, *Risk*, 7, 18–20.

Eberlein, E. (2009) Jump-Type Lévy Processes, in Andersen et al. (eds.), *Handbook of Financial Time Series*, Springer Verlag, 439–454.

Eberlein, E., U. Keller and K. Prause (1998) New Insights into Smile, Mispricing, and Value at Risk: The Hyperbolic Model, *The Journal of Business*, 71, 371–405.

Engle, R.F. and G.G.J. Lee (1999) A Permanent and Transitory Component Model of Stock Retrun 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.

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.

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.

Eraker, B. (2004) Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices, *Journal of Finance*, 58, 1367–1403.

Eraker, B., M.S. Johannes and N.G. Polson (2003) The Impact of Jumps in Returns and Volatility, *Journal of Finance*, 53, 1269–1300.

Erikkson, A., E. Ghysels and F. Wang (2009) Normal Inverse Gaussian Risk Neutral Probability Measures and the Pricing of Derivatives, *Journal of Derivatives*, 16, 23–37.

Fan, J. and L. Mancini (2008) Option Pricing with Model-guided Nonparametric Methods, Working Paper, Bendheim Center for Finance, Princeton University.

Figlewski, S. (2009) Estimating the Implied Risk Neutral Density for the U.S. Market Portfolio, in T. Bollerslev, J. Russell and M. Watson (eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, Oxford, UK: Oxford University Press.

Föllmer, H. and A. Schied (2004) *Stochastic Finance: An Introduction in Discrete Time*, Berlin and New York: Walter de Gruyter.

Fouque, J.-P., G. Papanicolaou and K.R. Sircar (2000) *Derivatives in Financial Markets with Stochastic Volatility*, Princeton University Press.

Gagliardini P., C. Gouriéroux and E. Renault (2008) Efficient Derivative Pricing and the Extended Method of Moments, Mimeo.

Gallant, A.R. and G. Tauchen (1996) Which Moments to Match? *Econometric Theory*, 12, 657–681.

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.

Gallant, A.R. and G. Tauchen (2010) Simulated Score Methods and Indirect Inference for Continuous-Time Models, in Aït Sahalia and L. Hansen (eds.), *Handbook of Financial Econometrics*, Amsterdam: North-Holland. In press.

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.

Gallant, A.R. and H. White (1992) On Learning the Derivatives of an Unknown Mapping with Multilayer Feedforward Networks, *Neural Networks*, 5, 129–138.

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, 617–631.

Garcia, R. and E. Renault (1998a) A Note on GARCH Option Pricing and Hedging, *Mathematical Finance*, 8(2), 153–161.

Garcia, R. and E. Renault (1998b) Risk Aversion, Intertemporal Substitution and Option Pricing, Working Paper 98s-02, CIRANO.

Garcia, R. and E. Renault (2001) State Variables Models for Stochastic Discount Factors, in J. Cvitanic, E. Jouini and M. Musiela (eds.), *Handbooks of Mathematical Finance*: Topics in Option Pricing, Interest Rates and Risk Management, Cambridge University Press.

Garcia, R. and R. Gençay (2000) Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint, *Journal of Econometrics*, 94, 93–115.

Garcia, R., M.A. Lewis, S. Pastorello and E. Renault (2009) Estimation of Objective and Risk-Neutral Distributions Based on Moments of Integrated Volatility, *Journal of Econometrics* (Forthcoming).

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.

Garcia, R., R. Luger and E. Renault (2005) Viewpoint: Option Prices, Preferences and State Variables, *Canadian Journal of Economics*, 38, 1–27.

Gençay, R. and M. Qi (2001) Pricing and Hedging Derivative Securities with Neural Networks: Bayesian Regularization, Early Stopping, and Bagging, *IEEE Transactions on Neural Networks*, 12, 726–734.

Ghysels, E. and F. Wang (2009) Some Useful Densities for Risk Management and their Properties, Discussion Paper, UNC.

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.

Ghysels, E., A. Harvey and E. Renault (1996) Stochastic Volatility, in G.S. Maddala (eds.), *Statistical Methods in Finance (Handbook of Statistics, vol. 14)*, Amsterdam: North-Holland, 119–191.

Ghysels, E., C. Gouriéroux and J. Jasiak (1997) Market Time and Asset Price Movements: Theory and Estimation, in D. Hand and S. Jacka (eds.), *Statistics in Finance*, Chapter 15, 307–332. London: Edward Arnold.

Ghysels, E., V. Patilea, E. Renault and O. Torrès (1996) Nonparametric Methods and Option Pricing, in D. Hand and S. Jacka (eds.), *Statistics in Finance*, 261–282. London: Edward Arnold.

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.

Gouriéroux, C., A. Monfort and C. Tenreiro (1994) Kernel M-Estimators: Nonparametric Diagnostics for Structural Models, Working Paper 9405, CEPREMAP, Paris.

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

Gouriéroux, C. and A. Monfort (2001) Pricing with Splines, Working Paper, Crest-Insee (Paris).

Gray, P. and S. Newman (2005) Canonical Valuation of Options in the Presence of Stochastic Volatility, *The Journal of Futures Markets*, 25, 1–19.

Haley, R. and T. Walker (2007) Alternative Tilts for Nonparametric Option Pricing, mimeo, Department of Economics, Indiana University.

Hamilton, J.D. (1989) A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, 57, 357–84.

Hansen, L. (1982) Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029–1054.

Hansen, L.P. and R. Jagannathan (1997) Assessing Specification Errors in Stochastic Discount Factor Models, *Journal of Finance*, 52, 557–590.

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.

Hansen, L.P. and J. Scheinkman (2009) Long Term Risk: An Operator Approach, *Econometrica*, 77, 177–234.

Hansen, L. and K. Singleton (1982) Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models, *Econometrica*, 50, 1269–1285.

Härdle, W. and A. Yatchew (2001) Dynamic Nonparametric State Price Density Estimation, Working Paper, University of Toronto.

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.

Harrison, J.M. and D. Kreps (1979) Martingales and Arbitrage in Multiperiod Securities Markets, *Journal of Economic Theory*, 20, 381–408.

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.

Heston, S. and S. Nandi (2000) A Closed-Form GARCH Option Valuation Model, *Review of Financial Studies*, 13, 585–625.

Huang, C. and R.H. Litzenberger (1998) *Foundations for Financial Economics*, Amsterdam: North-Holland.

Hull, J. (2005) *Options, Futures and Other Derivatives*, 6th Edition, Prentice Hall.

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

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.

Imbens, G., R. Spady and P. Johnson (1998) Information Theoretic Approaches to Inference in Moment Conditions Models, *Econometrica*, 66, 333–357.

Jackwerth, J. (2000) Recovering Risk Aversion from Option Prices and Realized Returns, *Review of Financial Studies*, 13, 433–451.

Jackwerth, J. and M. Rubinstein (1996) Recovering Probabilities Distributions from Option Prices, *Journal of Finance*, 51, 1611–1631.

Johannes and Polson (2010) MCMC Methods for Continuous-Time Financial Econometrics, in Y. Aït Sahalia and L. Hansen (eds.), *Handbook of Financial Econometrics*, Amsterdam: North-Holland. In press.

Jones, C. (2003) The Dynamics of Stochastic Volatility: Evidence From Underlying and Options Markets, *Journal of Econometrics*, 116, 181–224.

Kallsen, J. and M.S. Taqqu (1998) Option Pricing in ARCH-type Models, *Mathematical Finance*, 8, 13–26.

Kamrad, B. and P. Ritchken (1991) Multinomial Approximating Models for Options with *k* State Variables, *Management Science*, 37, 1640–1652.

Kitamura, Y. (2007) Empirical Likelihood Methods in Econometrics: Theory and Practice, in R. Blundell, W. Newey and T. Persson (eds.), Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, Vol. 3, Econometric Society Monograph Series, 174–237. Cambridge University Press.

Kitamura, Y. and M. Stutzer (1997) An Information Theoretic Alternative to Generalized Method of Moments Estimation, *Econometrica*, 65, 861–874.

Kitamura, Y., G. Tripathi and H. Ahn (2004) Empirical Likelihood Based Inference in Conditional Moment Restriction Models, *Econometrica*, 72, 1667–1714.

Lewis, A. (2000) *Option Valuation Under Stochastic Volatility with Mathematica Code*, Finance Press.

Longstaff, F.A. (1995) Option Pricing and the Martingale Restriction, *Review of Financial Studies*, 8, 1091–1124.

Madan, D.B. and E. Seneta (1990) The Variance Gamma (VG) Model for Share Market Returns, *Journal of Business*, 63, 511–524.

Madan, D.B. and F. Milne (1991) Option Pricing with VG Martingale Components, *Mathematical Finance*, 1, 39–56.

Manaster, S. and R.J. Rendleman, Jr. (1982) Option Prices as Predictors of Equilibrium Stock Prices, *Journal of Finance*, 37, 1043–1057.

Merton, R.C. (1973) Theory of Rational Option Pricing, *Bell Journal of Economics and Management Science*, 4, 141–183.

Merton, R.C. (1992) *Continuous-Time Finance*, Cambridge: Basil Blackwell.

Nicolato, E. and E. Venardos (2003) Option Pricing in Stochastic volatility Models of the Ornstein-Uhlenbeck Type, *Mathematical Finance*, 13, 445–466.

Pan, J. (2002) The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study, *Journal of Financial Economics*, 63, 3–50.

Pastorello, S., E. Renault and N. Touzi (2000) Statistical Inference for Random Variance Option Pricing, *Journal of Business and Economics Statistics*, 18, 358–367.

Pastorello, S., V. Patilea and E. Renault (2003) Iterative and Recursive Estimation of Structural Non Adaptive Models, *Journal of Business and Economic Statistics*, 21, 449–509.

Pliska, S. (1997) *Introduction to Mathematical Finance: Discrete Time Models*, Basil Blackwell.

Poteshman, A.M. (2000) Estimating a General Stochastic Variance Model from Option Prices, Working Paper, University of Illinois at Urbana-Champaign.

Renault, E. (1997) Econometric Models of Option Pricing Errors, in D. Kreps and K. Wallis (eds.), Advances in *Economics and Econometrics: Theory and Applications*, Vol. 3, Cambridge University Press.

Renault, E. and N. Touzi (1996) Option Hedging and Implied Volatilities in a Stochastic Volatility Model, *Mathematical Finance*, 6, 279–302.

Robinson, P.M. (1991) Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression, *Journal of Econometrics*, 47, 67–84.

Rogers, L.C.G. (1997) Arbitrage with Fractional Brownian Motion, *Mathematical Finance*, 7, 95–105.

Romano, M. and N. Touzi (1997) Contingent Claims and Market Completeness in a Stochastic Volatility Model, *Mathematical Finance*, 7, 399–412.

Ross, S.A. (1976) The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory*, 13, 341–360.

Rosenberg, J.V. (2000) Asset Pricing Puzzles: Evidence from Option Markets, Working Paper, NYU Stern School of Business.

Rosenberg, J.V. and R.F. Engle (2002) Empirical Pricing Kernels, *Journal of Financial Economics*, 64, 341–372.

Rubinstein, M. (1994) Implied Binomial Trees, *Journal of Finance*, 44, 771–818.

Rubinstein, M. (1996) The Valuation of Uncertain Income Streams and the Pricing of Options, *Bell Journal of Economics and Management Science*, 7, 407–425.

Sato, K. (1999) *Lévy Processes and Infinite Divisibility*, Cambridge: Cambridge University Press.

Shephard, N. (1996) Statistical Aspects of ARCH and Stochastic Volatility, in D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielson (eds.), *Time Series Models in Econometrics, Finance and Other Fields*, 1–67. London: Chapman & Hall.

Shimko, D. (1993) Bounds on Probability, *Risk*, 6, 33–37.

Stutzer, M. (1996) A Simple Nonparametric Approach to Derivative Security Valuation, *Journal of Finance*, 51, 1633–1652.

Sundaresan, S. (2000) Continuous-Time Methods in Finance: A Review and an Assessment, *Journal of Finance*, 55, 1569–1622.

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

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

Tenreiro, C. (1995) Estimation fonctionnelle: Applications Aux Tests d'équation et de paramètre Constant, Ph.D. Thesis, University of Paris VI.

Turnbull, S. and F. Milne (1991) A Simple Approach to Interest-Rate Option Pricing, *Review of Financial Studies*, 4, 87–121.

Wiggins, J. (1987) Option Values Under Stochastic Volatility, *Journal of Financial Economics*, 19, 351–372.

Willard, G.A. (1997) Calculating Prices and Sensitivities for Path-independent Derivative Securities in Multifactor Models, *The Journal of Derivatives*, 45–61.

Yatchew, A. and W. Härdle (2005) Nonparametric State Price Density Estimation Using Constrained Least Squares and the Bootstrap, *Journal of Econometrics*, 133, 579–599.

Yoon, J. (2008) Option Pricing with Stochastic Volatility Models, PhD Thesis, UNC at Chapel Hill.