Nonparametric Methods and Option Pricing*

Eric Ghysels†

Valentin Patilea†

Éric Renault‡

Olivier Torrès§

This version: May 27, 1997

(First draft: June 6, 1996)

Abstract

In this paper, we survey some of the recent nonparametric estimation methods which were developed to price derivative contracts. We focus on equity options and start with a so-called model-free approach which involves very little financial theory. Next we discuss nonparametric and semi-parametric methods of option pricing and illustrate the different approaches.

*Part of this work was funded by the Fonds FCAR of Québec, the SSHRC of Canada under Strategic Grant 804-96-0027, the TMR Work Programme of the European Commission under Grant Nr ERB4001GT950641 and the contract "Projet d'Actions de Recherche Concertées" (PARC No. 93/98-164) of the Belgium Government.

†Pennsylvania State University and CIRANO, *e-mail*: eghysels@psu.edu.

†Institut de Statistique, Université catholique de Louvain, *e-mail*: patilea@stat.ucl.ac.be.

‡GREMAQ and IDEI, Université des Sciences Sociales de Toulouse and Institut Universitaire de France.

§GREMARS, Université de Lille 3, *e-mail*: torres@univ-lille3.fr.

1 Introduction

Derivative securities are widely traded financial instruments which inherit their statistical properties from those of the underlying assets and the features of the contract. The now famous Black and Scholes (1973) formula is one of the few cases where a call option is priced according to an analytical formula and applies to European-type contracts written on a stock which follows a Geometric Brownian Motion. The formula can be derived via a dynamic hedging argument involving a portfolio of a riskless bond and the underlying stock (see *e.g.*, Duffie (1996) for more details). Strictly speaking the restrictive assumptions underlying the so-called Black-Scholes economy are rarely met. Indeed, among the violations one typically encounters and cites are: (1) volatility is time-varying, (2) trading is not costless or faces liquidity constraints, (3) interest rates may be stochastic, (4) the stock has dividend payments, *etc.* It should also be noted that many derivative securities traded on exchanges and over-the-counter are not "plain vanilla" but feature deviations from the basic European call contract design. In particular, some contracts feature early exercise privileges, *i.e.* so-called American-type options, some involve caps or floors, some involve multiple securities like swaps or quanto options, while others are written on fixed income securities instead of stocks, *etc.* Obviously, there are many extensions of the Black-Scholes model which take into account some of the deviations one encounters in practice either in terms of assumptions or contract design. In most cases, however, there is no longer an elegant analytical formula and the contract must by priced via numerical methods. Very often there are limitations to these numerical methods of approximation as well since they remain very specific and inherit many of the aforementioned restrictions which apply to the Black-Scholes model. These are some of the motivating reasons why statistical nonparametric methods are applied to option pricing. Indeed, these methods are appealing for the following reasons: (1) the formula is known but too complex to calculate numerically or (2) the pricing formula is unknown and (3) there is an abundance of data which makes the application of nonparametric methods attractive. The first applies for instance to the case of American-type options with stochastic volatility or barriers, to interest rate derivative security pricing, to "look back" options, *etc.* The second applies because of incompleteness of markets, trading frictions or else the desire to leave unspecified either the stochastic properties of the underlying asset and/or the attitudes of agents towards risk.

There are a multitude of nonparametric methods, see *e.g.*, Silverman (1986) for an introduction to the statistical literature on the subject. In addition there are many ways to tackle the pricing of options via nonparametric methods. Moreover, there are many different types of option contracts, some of which require discussion of special features like early exercise decisions in the case of American options. Given the large number of possibilities and the multitude of methods we have to be selective in our survey of methods and applications. The literature is also rapidly growing. Recent papers include Aït-Sahalia (1993, 1996), Aït-Sahalia and Lo (1995), Baum and Barkoulas (1996), Bossaerts, Hafner and Härdle (1995), Broadie, Detemple, Ghysels and Torres (1995,1996), Elsheimer et al. (1995), Ghysels and Ng (1996), Gouriéroux, Monfort and Tenreiro (1994, 1995), Gouriéroux and Scaillet (1995), Hutchinson, Lo and Poggio (1994), Stutzer (1995), among others. To focus the survey we will restrict our attention only to options on equity.

We may be tempted to exclude any *a priori* economic knowledge from our econometric analysis and solely rely on the brute force of nonparametric techniques. This is the so-called *model-free* approach. These nonparametric methods and their use in the context of option pricing will be presented in section 2. They mainly consist of estimating the relation between the dependent variable (usually the option price) and explanatory variables using nonparametric regression techniques. The function characterizing the relationship to be estimated is chosen in a family of loosely defined functions according to an appropriate selection criterion. However, the application of these standard methods in the context of option pricing raises difficulties, some which are not easy to overcome. This is one of several reasons why we may prefer to introduce some restrictions imposed by economic theory. These restrictions are quite often very mild and appear to be sensible, and they can be of great help to eliminate some of the difficulties met in the model-free approach. From a statistical point of view, these restrictions also call for other types of nonparametric methods, such as the nonparametric specification and estimation of equivalent martingale densities. This will be discussed in section 3.

Despite the fact that the Black-Scholes (henceforth BS) framework fails to describe the behavior of call prices or the exercise policy of investors when contracts are of American type, it still is the most widely used formula among practitioners. For instance, although it is believed that underlying stock prices are not represented by a log-normal diffusion because of time-varying volatility, the BS formula is used to measure this instantaneous variance, even though a necessary condition for the BS formula to be valid is that volatility is constant. The primary appeal of the BS formula is its simplicity and the believe among practitioners that it captures the variables relevant to price option contracts. Therefore, following the practitioner, an econometrician may find it convenient to use the BS formula as a benchmark for his analysis. In this context, nonparametric statistical techniques provide a way of filling the gap between the Black-Scholes and the real world. As we will show in section 4, these methods can be called upon to "correct" the BS formula so that it can adequately describe the behavior of observed series.

2 Nonparametric Model-free Option Pricing

In this section we present the simplest of all possible methods, which is probably also the purest in a statistical sense as it involves very little financial theory. Suppose we have a large data set with option prices and the features of the contracts such as strike, time to expiration, *etc.* and data of the underlying security. Such data sets are now commonly found and distributed to the academic and financial communities. Formally, a call is priced via:

$$\Pi_t = f_1(S_t, K, T, t, X_t), \tag{2.1}$$

where Π_t is the option price and S_t the price of the underlying asset both at time t, K the exercise price of the contract and T its expiration date and finally X_t is a vector of variables affecting the price of the option contract. The latter may include underlying asset prices prior to t, as for instance in non-Markovian settings and/or latent variables which appear in stochastic volatility models. For the moment we ignore the fact that several contracts are listed on a daily basis which in principle should require a panel structure instead of a single time series. In section 3 we will say more about panel structures. The pricing functional f_1 is assumed unknown, only its arguments are suggested by the setup of the contract.[1] The purpose of applying nonparametric statistical estimation is to recover f_1 from the data. Obviously, this can only be justified if the estimation is applied to a situation where the regularity conditions for such techniques are satisfied. To discuss this let us briefly review the context of nonparametric estimation. In general, it deals with the estimation of relations such as

$$Y_i = g(Z_i) + u_i\,, \quad i = 1, \ldots, n, \tag{2.2}$$

where, in the simplest case, the pair $((Y_i, Z_i), i = 1, \ldots, n)$ is a family of i.i.d. random variables, and $E(u|Z) = 0$, so that $g(z) = E(Y|Z = z)$. The error terms $u_i, i = 1, \ldots, n$, are assumed to be independently distributed, while g is a function with certain smoothness properties. Several estimation techniques exist, including kernel-based methods, smoothing splines, orthogonal series estimators such as Fourier series, Hermite polynomials and neural networks, among many others. We will focus here on the kernel-based methods for the purpose of exposition. Kernel smoothers produce an estimate of g at $Z = z$ by giving more weight to observations (Y_i, Z_i) with Z_i "close" to z. More precisely, the technique relies on a *kernel function*, K, which acts as a weighting scheme (it is usually a probability density function, see Silverman (1986, p. 38)) and a *smoothing parameter* λ which defines the degree of "closeness" or neighborhood. The most widely used kernel estimator

[1] Nonparametric techniques for selecting the arguments of a nonparametric regression function have recently been proposed by Aït-Sahalia, Bickel and Stoker (1995), Gouriéroux, Monfort and Tenreiro (1994), Lavergne and Vuong (1996) among others.

of g in (2.2) is the Nadaraya-Watson estimator defined by

$$\hat{g}_\lambda(z) = \frac{\sum_{i=1}^n K\left(\frac{Z_i - z}{\lambda}\right) Y_i}{\sum_{i=1}^n K\left(\frac{Z_i - z}{\lambda}\right)}, \tag{2.3}$$

so that $(\hat{g}_\lambda(Z_1), \ldots, \hat{g}_\lambda(Z_n))' = W_n^K(\lambda)Y$, where $Y = (Y_1, \ldots, Y_n)'$ and W_n^K is a $n \times n$ matrix with its (i,j)-th element equal to $K\left(\frac{Z_j - Z_i}{\lambda}\right) / \sum_{k=1}^n K\left(\frac{Z_k - Z_i}{\lambda}\right)$. W_n^K is called the *influence matrix* associated with the kernel K.

The parameter λ controls the level of neighboring in the following way. For a given kernel function K and a fixed z, observations (Y_i, Z_i) with Z_i far from z are given more weight as λ increases; this implies that the larger we choose λ, the less $\hat{g}_\lambda(z)$ is changing with z. In other words, the degree of smoothness of $\hat{g}_\lambda$ increases with λ. As in parametric estimation techniques, the issue here is to choose K and λ in order to obtain the best possible fit. A natural measure of the goodness of fit at $Z = z$ is the mean squared error ($\mathrm{MSE}(\lambda, z) = E\left[\left(\hat{g}_\lambda(z) - g(z)\right)^2\right]$), which has a bias/variance decomposition similar to parametric estimation. Of course both K and λ have an effect on $\mathrm{MSE}(\lambda, z)$, but it is generally found in the literature that the most important issue is the choice of the smoothing parameter.[2] Indeed, λ controls the relative contribution of bias and variance to the mean squared error; high λs produce smooth estimates with a low variance but a high bias, and conversely. It is then crucial to have a good rule for selecting λ. Several criteria have been proposed, and most of them are transformations of $\mathrm{MSE}(\lambda, z)$. We may simply consider $\mathrm{MSE}(\lambda, z)$, but this criterion is local in the sense that it concentrates on the properties of the estimate at point z. We would generally prefer a global measure such as the *mean integrated squared error* defined by $\mathrm{MISE}(\lambda) = E\left[\int \left(\hat{g}_\lambda(z) - g(z)\right)^2 \mathrm{d}z\right]$, or the *sup mean squared error* equal to $\sup_z \mathrm{MSE}(\lambda, z)$, *etc...* The most frequently used measure of deviation is the sample mean squared error $M_n(\lambda) = (1/n)\sum_{i=1}^n \left[\hat{g}_\lambda(Z_i) - g(Z_i)\right]^2 \omega(Z_i)$, where $\omega(\cdot)$ is some known weighting function. This criterion only considers the distances between the fit and the actual function g at the sample points Z_i. Obviously, choosing $\lambda = \tilde{\lambda}_n \equiv \underset{\lambda}{\mathrm{argmin}}\ M_n(\lambda)$ is impossible to implement since g is unknown. The strategy consists of finding some function $m_n(\cdot)$ of λ (and of $\left((Y_i, Z_i), i = 1, \ldots, n\right)$) whose argmin is denoted $\hat{\lambda}_n$, such that $|\tilde{\lambda}_n - \hat{\lambda}_n| \longrightarrow 0$ a.s. as $n \to \infty$. For a review of such functions m_n, see Härdle and Linton (1994, section 4.2).[3] The most widely used m_n function is the *cross-validation* function

$$m_n(\lambda) = CV_n(\lambda) \equiv \frac{1}{n}\sum_{i=1}^n \left[Y_i - \hat{g}_\lambda^{(-i)}(Z_i)\right]^2,$$

where $\hat{g}_\lambda^{(-i)}(z)$ is a Nadaraya-Watson estimate of $g(z)$ obtained according to (2.3) but with the i-th observation left aside. Craven and Wahba (1979) proposed the *generalized cross-validation*

[2] For a given λ, the most commonly used kernel functions produce more or less the same fit. Some measures of relative efficiency of these kernel functions have been proposed and derived, see Härdle and Linton (1994, p. 2303) and Silverman (1986, section 3.3.2).

[3] See also Silverman (1986, section 3.4), Andrews (1991) and Wand and Jones (1995).

function with

$$m_n(\lambda) = GCV_n(\lambda) \equiv \frac{n^{-1}\sum_{i=1}^n [Y_i - \hat{g}_\lambda(Z_i)]^2}{\left[1 - n^{-1}\ \mathrm{tr}\big(W_n(\lambda)\big)\right]^2}$$

where W_n is the influence matrix.[4]

Another important issue is the convergence of the estimator $\hat{g}_{\hat{\lambda}_n}(z)$. Concerning the Nadaraya-Watson estimate (2.3), Schuster (1972) proved that under some regularity conditions, $\hat{g}_{\hat{\lambda}_n}(z)$ is a consistent estimator of $g(z)$ and is asymptotically normally distributed.[5] Therefore when the argmin $\hat{\lambda}_n$ of $m_n(\lambda)$ is found in the set Λ_n (see footnote 5), we obtain a consistent and asymptotically normal kernel estimator $\hat{g}_{\hat{\lambda}_n}(z)$ of $g(z)$, which is optimal in the class of the consistent and asymptotically Gaussian kernel estimators for the criterion $M_n(\lambda)$.[6]

When the errors are not spherical, the kernel estimator remains consistent and asymptotically normal. The asymptotic variance is affected, however, by the correlation of the error terms. Moreover, the objective functions for selecting λ such as CV_n or GCV_n do not provide optimal choices for the smoothing parameters. It is still not clear what should be done in this case to avoid over- or undersmoothing.[7] One solution that has suggested consists in modifying the selection criterion (CV_n or GCV_n) in order to derive a constant estimate of M_n. An alternative strategy tries to orthogonalize the error term and apply the usual selection rules for λ. When the autocorrelation function of u is unknown, one has to make the transformation from sample estimates obtained from a first step smooth. In that view, the second alternative seems to be more tractable. Altman (1987, 1990) presents some simulation results which show that in some situations, the pre-whitening method seems to work relatively well. However there is no general result on the efficiency of the procedure. See also Härdle and Linton (1994, section 5.2) and Andrews (1991, section 6).

When the observations (Y, Z) are drawn from a stationary dynamic bivariate process, Robinson (1983) provides conditions under which kernel estimators of regression functions are consistent. He also gives some central limit theorems which ensure the asymptotic normality of the estimators. The conditions under which these results are obtained have been weakened by Singh and Ullah (1985). These are mixing conditions on the bivariate process (Y, Z). For a detailed treatment, see Györfy *et al.* (1989). This reference (chap. 6) also discusses the choice of the smoothing parameter in the context of nonparametric estimation from time series observations. In particular, if the error terms are independent, and when $\hat{\lambda}_n = \underset{\lambda\in\Lambda_n}{\operatorname{argmin}}\ CV_n(\lambda)$, then under certain regularity

[4] This criterion generalizes CV_n since GCV_n can be written as $n^{-1}\sum_{i=1}^n \left[Y_i - \hat{g}_\lambda^{(-i)}(Z_i)\right]^2 a_{ii}$, where the a_{ii}s are weights related to the influence matrix. Moreover, GCV_n is invariant to orthogonal transformations of the observations.

[5] The regularity conditions bear on the smoothness and continuity of g, the properties of the kernel function K, the conditional distribution of Y given Z, the marginal distribution of Z, and the limiting behavior of $\hat{\lambda}_n$. The class of $\hat{\lambda}_n$s which satisfy these regularity conditions is denoted Λ_n.

[6] By definition, the choice $\lambda = \lambda_n^*$ is optimal for the criterion $D(\lambda)$ if $D(\lambda_n^*)/\inf_{\lambda\in\Lambda_n} D(\lambda) \xrightarrow[n\to\infty]{\text{a.s.}} 1$.

[7] Altman (1990) shows that when the sum of the autocorrelations of the error term is negative (positive), then the functions CV_n and GCV_n tend to produce values for λ that are too large (small), yielding oversmoothing (undersmoothing).

conditions $\hat{\lambda}_n$ is an optimal choice for λ according to the *integrated squared error*, $\mathrm{ISE}(\lambda) = \int [\hat{g}_\lambda(z) - g(z)]^2 \,\mathrm{d}z$ (see Györfy *et al.* (1989, corollary 6.3.1)). Although the function $CV_n(\lambda)$ can produce an optimal choice of λ for the criterion $M_n(\lambda)$ in some particular cases (such as the pure autoregression, see Härdle and Vieu (1992) and Kim and Cox (1996)), there is no general result for criterions such as $\mathrm{MISE}(\lambda)$ or $M_n(\lambda)$. For studies of the performance of various criteria for selecting λ in the context of dependent data, see Cao *et al* . (1993).

In applications involving option price data we have correlated as well as *nonstationary* data. Indeed S_t, which is one of the arguments of f_1 in (2.1), is usually not a stationary process. Likewise, variables entering X_t may be nonstationary as well. Moreover, characterizing the correlation in the data may also be problematic as well. Indeed, the relevant time scale for the estimation of f_1 is not calendar time, as in a standard time series context, but rather the time to expiration of the contracts which are sampled sequentially through the cycle of emissions. It becomes even more difficult once it is realized that at each time t several contracts are listed simultaneously and trading may take place only in a subset of contracts. Some of these technical issues can be resolved. For instance, while S is nonstationary the variable (S/K) is found to be stationary as exercise prices bracket the underlying asset price process. This suggests an alternative formulation of (2.1) as $\Pi_t = f_2(S_t/K, K, T, t, X_t)$. Moreover, under mild regularity conditions f_1 is homogenous of degree one in (S, K) (see Broadie, Detemple, Ghysels and Torrès (1996) or Garcia and Renault (1995)). Under such conditions we have:

$$\Pi_t/K = f_3(S_t/K, T, t, X_t). \tag{2.4}$$

A more difficult issue to deal with is the correlation in the data. Indeed, while it easy to capture the serial correlation in calendar time it is tedious to translate and characterize such dependence in a time to maturity scale (see Broadie, Detemple, Ghysels and Torrès (1996) for further discussion). Furthermore, the panel data structure of option contracts even worsen the dependence characteristics. Finally, we also face the so-called curse of dimensionality problem. Nonparametric kernel estimators of regression functions $Y = g(Z)$, where Z is a vector of dimension d, as f_3 in equation (2.4), are *local* smoothers in the sense that the estimate of g at some point z depends only on the observations (Z_i, Y_i) with Z_i in a neighborhood $\mathcal{N}(z)$ of z. The so-called curse of dimensionality relates to the fact that, if we measure the degree of localness of a smoother by the proportion of observations (Z_i, Y_i) for which Z_i is in $\mathcal{N}(z)$, then the smoother becomes less local when d increases, in the sense that for a fixed degree of localness $\mathcal{N}(z)$ increases in size as the dimension of Z increases. Consequently, the precision of the estimate deteriorates as we add regressors in g, unless the sample size increases drastically.[8] This problem arises in our context as X_t may contain

[8] For more details on the curse of dimensionality and how to deal with it, see Hastie and Tibshirani (1990), Scott (1992, chap. 7) and Silverman (1986, p. 91 – 94).

many variables. A good example is the case of nonparametric estimation of option pricing models with stochastic volatility, a latent variable which requires filtering from past squared returns of the underlying asset (see below for further discussion).

The homogeneity property of option prices helps to reduce the dimension of the pricing function f_1 by eliminating one of its arguments. Moreover, in most pricing models, the expiration date T and the calendar date t affect Π_t through their difference $\tau \equiv T - t$, a variable called time to expiration or alternatively time to maturity. Therefore the nonparametric regression to be estimated becomes:

$$\Pi_t/K = f_3(S_t/K, \tau, X_t). \tag{2.5}$$

Let us elaborate further now on the specification of the vector of variables X_t affecting the option price. Examples of variables that might enter X_t are series such as random dividends or random volatility. Dividend series are observable while volatility is a latent process. This raises a number of issues we need to discuss here as there are fundamental difference between the two cases. In principle, one can filter the latent volatility process from the data, using series on the underlying asset. Obviously we need a parametric model if we were to do this in an explicit and optimal way. This would be incompatible however with a nonparametric approach. Hence, we need to proceed somehow without making specific parametric assumptions. In principle, one could consider a nonparametric fit between $\left(S/K\right)_t$ and past squared returns $(\log S_{t-j} - \log S_{t-j-1})^2, j = 1, \ldots, L$, for some finite lag L, resulting in the following $L+1$-dimensional nonparametric fit:

$$\Pi_t/K = f_3(S_t/K, T, t, (\log S_{t-j} - \log S_{t-j-1})^2, j = 1, 2, \ldots, L), \tag{2.6}$$

It is clear that this approach is rather unappealing as it would typically require a large number of lags, say $L = 20$ with daily observations. Hence, we face the curse of dimensionality problem discussed before. A more appealing way to proceed is to summarize the information contained in past squared returns (possibly the infinite past). Broadie *et al.* (1996) consider three different strategies using: (a) historical volatilities, (b) EGARCH volatilities and (c) implied volatilities. Each approach raises technical issues, some of which are relatively straightforward to deal with, while others are more tedious. For example, using GARCH or EGARCH models raises several issues: (1) are equations of EGARCH models compatible with the unspecified asset return processes generating the data? (2) how does parameter estimation of the GARCH process affect nonparametric inference? and (3) how do the weak convergence results also affect the nonparametric estimation? Broadie *et al.* (1996) discuss the details of these issues and illustrate with an empirical example that the three aforementioned approaches yield the same results.

The nonparametric regression f_3 discussed so far does not rely, at least directly, on a theoretical financial model. Yet, it is possible to use the nonparametric estimates to address certain questions regarding the specification of theoretical models, namely questions which can be formulated as inclusion or exclusion of variables in the nonparametric option pricing regression. We may illustrate this with an example drawn from Broadie *et al.* (1996). For European type options there has been considerable interest in formulating models with stochastic volatility (see *e.g.*, Hull and White (1987) among many others) while there has been relatively little attention paid to cases involving stochastic dividends. It is quite the opposite with American type options. Indeed, the widely traded S&P100 Index option or OEX contract has been extensively studied, see in particular Harvey and Whaley (1992), with exclusive emphasis on stochastic dividends (with fixed volatility). This prompted Broadie *et al.* (1996) to test the specification of OEX option pricing using the nonparametric methods described in this section combined with tests described in Aït-Sahalia, Bickel and Stoker (1995). Hence, they tested the relevant specification of the vector X_t whether it should include dividends and/or volatility (where the latter is measured via one of the three aforementioned proxies). They found, in the case of the OEX contract that both stochastic volatility and dividends mattered. It implies that either ignoring volatility or dividends results in pricing errors, which can be significant as Broadie *et al.* show. This is an important illustration on how to use this so-called model-free approach to address specification of option pricing without much financial theory content.

To conclude we should mention that there are several applications of the techniques discussed here which can be found in Aït-Sahalia and Lo (1995) as well as Broadie, Detemple, Ghysels and Torrès (1995,1996). The former study the European option on the S&P500 contract, while the latter study the American contract on the S&P100. By using slightly different techniques, Hutchinson, Lo and Poggio (1994) achieve the same objective.

3 Nonparametric specification of equivalent martingale measures

An obvious difficulty for the model-free option pricing setup in the previous section is the so-called panel structure of option prices data. Namely, one typically observes several simultaneously traded contracted (with various exercise prices and maturity dates) so that the option pricing formula of interest must involve two indexes:

$$\Pi_{jt} = f(S_t, K_j, T_j, t, X_{jt}) \tag{3.1}$$

where $j = 1, 2, \ldots, J_t$ describes the (possibly large) set of simultaneously quoted derivative contracts at time t written on the same asset (with price S_t at time t).[9] In such a case, nonparametric

[9]For instance Dumas, Fleming and Whaley (1996) consider S&P500 Index option prices traded on the Chicago Board Options Exchange (CBOE) during the period June 1988 through December 1993. After applying three

model free option pricing becomes quickly infeasible since it is not able to capture a large set of crucial restrictions implied by arbitrage. Indeed, as stressed by Merton (1973), any option pricing research must start from deducing a set of restrictions which are necessary conditions for a formula to be consistent with a rational pricing theory. Fitting option pricing formula using purely (model-free) statistical methodologies therefore forgoes imposing an important feature of derivative asset markets, namely: If a security A is dominant over a security B (that is the return on A will be at least as large as on B in all states of the world and exceed the return on B for some states), then any investor willing to purchase security B would prefer to purchase A. A first example of restrictions stressed by Merton (1973) for European call options prices are:

$$K_2 > K_1 \Rightarrow f(S_t, K_2, T, t, X_{jt}) \leq f(S_t, K_1, T, t, X_{jt})$$

and, if no payouts (*e.g.*, dividends) are made to the underlying asset (*e.g.*, a stock) over the life of the option

$$f(S_t, K, T, t, X_t) \geq \max[0, S_t - K \; B(t, T)],$$

where $B(t, T)$ is the price of a riskless pure discount bond which pays one dollar $T - t$ periods from now (that is date t).

The only way for a pricing scheme to take into account the necessary conditions for a formula to be consistent with a rational pricing theory is to incorporate at a convenient stage the requirement that the derivative asset price $f(S_t, K_j, T_j, t, X_{jt})$ has to be related to its terminal payoff, for instance $(S_T - K)^+ := \max[0, S_T - K]$ in case of an European call option. Fortunately, modern derivative asset pricing theory provides us a versatile tool to do this using equivalent martingale measures. Roughly speaking, the Harrison and Kreps (1979) theory ensures the equivalence between the absence of arbitrage and the existence of a probability measure Q with the property that the discounted price processes are martingales under Q. Hence, for a European option with strike K and maturity T we have:

$$f(S_t, K, T, t, X_t) = B(t, T) E^Q \left[(S_T - K)^+ | X_t, S_t \right], \tag{3.2}$$

where the expectation operator E^Q is defined with respect to the pricing probability measure Q.[10] The vector of variable X_t assumes here again the role it played in the previous section, namely a set of state variables relevant to the pricing of the option. The fundamental difference between the nonparametric methods described in the previous section and those which rely on the Harrison and

exclusionary criteria to avoid undesirable heterogeneity, they find quotes for an average of 44 option series during the last half-hour each Wednesday.

[10] For further details see *e.g.*, Duffie (1996). We do not discuss here: (1) the assumptions of frictionless markets which ensure the equivalence between the absence of arbitrage and the existence of a such a pricing measure Q, and (2) the interpretation of Q which is unique in case of complete markets and often called a "risk neutral probability" when there is no interest rate risk.

Kreps theory is that the nonparametric statistical inference focuses on the conditional expectation operator $E^Q[\,\cdot\,|X_t, S_t]$ instead of the pricing function $f(S_t, K, T, t, X_t)$. It is important to note of course that in general the Q-conditional probability distribution of S_T given X_t, S_t coincides with the Data Generating Process (DGP characterized by a probability measure denoted P hereafter). Early contributions assumed that

$$E^Q[\,\cdot\,|X_t, S_t] = E^P[\,\cdot\,|X_t, S_t] \tag{3.3}$$

(see for instance Engle and Mustafa (1992) or Renault and Touzi (1996)) but recently several attempts were made to estimate $E^Q[\,\cdot\,|X_t, S_t]$ without assuming that Q coincides with P (see e.g. Rubinstein (1994), Abken et al. (1996) and Aït-Sahalia and Lo (1995)). For instance, it is now well-known (see *e.g.*, Breeden and Lutzenberger (1978) and Huang and Litzenberger (1988) page 140) that there is a one-to-one relationship between an European call option pricing function $f(S_t, \cdot, T, t, X_t)$ as a function of the strike price K and the pricing probability measure $Q|X_t, S_t$ via

$$Q\left[\frac{S_T}{S_t} \geq \frac{K}{S_t}|X_t, S_t\right] = -\frac{1}{B(t,T)}\frac{\partial f}{\partial K}(S_t, K, T, t, X_t). \tag{3.4}$$

Hence observing European call option prices for any strike K, it is possible to recover the pricing probability measure $Q \mid X_t, S_t$ or the corresponding pricing operator $E^Q[\cdot|X_t, S_t]$. This forms the basis for consistent nonparametric estimation of this operator without the restrictive assumption (3.3) and using options price data, not only time series ($t = 1, 2, \ldots, T$) but also cross-sections ($j = 1, 2, \ldots, J_t$).

Such inference may be developed within two paradigms: Bayesian (covered in a first subsection 3.1) or classical (subsection 3.2 below). In the former we treat the probability distribution as a random variable, hence the reference to Bayesian analysis. Obviously one needs to find a flexible class that covers a large set of probability measures. In section 3.1 we will present such a class, which enables one to characterize the martingale restrictions of option pricing as well as their panel structure. In section 3.2 we present another approach which builds on the estimation of option pricing formulae presented in the previous section. Both approaches have advantages as well as drawbacks which we will discuss.

3.1 Nonparametric Bayesian specification of equivalent martingale measures

The basic ideas presented in this section were introduced by Clément, Gouriéroux and Monfort (1993) and extended more recently by Renault (1996) and Patilea and Renault (1995).[11] Let us

[11] In this section we follow closely the analysis developed in Renault (1996) and Patilea and Renault (1995).

reconsider a European option with price $\Pi_t(K)$ written as

$$\begin{aligned} \Pi_t(K) &= B(t,T)E^Q\left[(S_T-K)^+|S_t,X_t\right] \\ &= B(t,T)\int_{\mathcal{S}}(s_T-K)^+Q_t(\mathrm{d}s_T) \end{aligned} \tag{3.5}$$

where $\mathcal{S}$ is the set of all possible values of S_T and the conditional distribution of S_T given X_t, S_t is now denoted by Q_t. First, we should note that even when markets are complete, one may not observe the full set of securities which complete the market and therefore one is not able to determine unambiguously the pricing probability measure Q_t. A nonparametric Bayesian methodology views this measure as a random variable (also denoted by Q_t) taking values in the set $\mathcal{P}(\mathcal{S})$ of all probability distributions on $(\mathcal{S},\mathcal{B}(\mathcal{S}))$. If we denote $Q_t(\,\cdot\,,\omega)$ as a realization of this random variable, then the option pricing formula (3.5) becomes

$$\Pi_t(K,\omega) = B(t,T)\int_{\mathcal{S}}(s_T-K)^+Q_t(\mathrm{d}s_T,\omega). \tag{3.6}$$

A good class of distributions to characterize the random probability Q_t is the Dirichlet process, which is a distribution on $(\mathcal{P}(\mathcal{S}),\mathcal{B}(\mathcal{P}(\mathcal{S})))$, introduced by Ferguson (1973). More specifically: a random probability π on $(\mathcal{S},\mathcal{B}(\mathcal{S}))$ is called a Dirichlet process with parameter λQ_t^0 where $\lambda > 0$ and $Q_t^0 \in \mathcal{P}(\mathcal{S})$ if, for any measurable partition $B_1, B_2, \ldots, B_L$ of $\mathcal{S}$, the random vector $(\pi(B_\ell), 1 \le \ell \le L)$ has a Dirichlet distribution with parameters $(\lambda Q_t^0(B_\ell), 1 \le \ell \le L)$. If π is a Dirichlet process with parameter λQ_t^0, we write hereafter $\pi \rightsquigarrow \mathcal{D}i(\lambda Q_t^0)$.[12] We may give an interpretation to Q_t^0 as a mean value of the process and to λ as a precision of the process around this mean value (if necessary, λ may also depend on t). Large values of λ make the realizations of π more concentrated around Q_t^0. For $\lambda = \infty$ the realizations of the Dirichlet process are, with probability one, equal to Q_t^0. To elaborate on the moment properties and the asymptotic behavior of Dirichlet processes let us consider a real-valued function f defined on $\mathcal{S}$ and integrable w.r.t. Q_t^0. If $\pi \rightsquigarrow \mathcal{D}i(\lambda Q_t^0)$ we define the random variable:

$$\Pi(f) = \int_{\mathcal{S}} f(s)\pi(\mathrm{d}s,\omega)$$

It can be shown (see *e.g.*, Rolin (1992), Clément, Gouriéroux and Monfort (1993)) that:

$$\begin{aligned} E(\Pi(f)) &= E^{Q_t^0}f \\ Var(\Pi(f)) &= (1+\lambda)^{-1}\,Var^{Q_t^0}(f) \\ Cov(\Pi(f_1),\Pi(f_2)) &= (1+\lambda)^{-1}\,Cov^{Q_t^0}(f_1,f_2) \end{aligned}$$

[12] It is worth recalling that the Dirichlet distribution is a multivariate extension of the Beta distribution on [0,1]. More precisely, the Dirichlet distribution on the simplex $\left\{(p_1,\ p_2,\ldots,p_L), p_\ell \ge 0, \sum p_\ell = 1\right\}$ is characterized by the pdf

$$g\left(p_1,\ p_2,\ldots,p_L\right) = \frac{\Gamma(\eta_1+\eta_2+\ldots+\eta_L)}{\Gamma(\eta_1)\Gamma(\eta_2)\ldots\Gamma(\eta_L)}\,p_1^{\eta_1-1}...p_L^{\eta_L-1},$$

where $(\eta_1,\eta_2,...,\eta_L)$ are given nonnegative parameters.

Asymptotic normality has been established by Lo (1987).[13] In particular if $\pi_n \rightsquigarrow \mathcal{D}i(\lambda_n Q_t^0)$ and $\lambda_n \to \infty$,

$$\sqrt{\lambda_n}\big(\Pi_n(f) - E^{Q_t^0}(f)\big) \xrightarrow{d} \mathcal{N}\big(0,\, Var^{Q_t^0}(f)\big).$$

It proves that for λ sufficiently large, that is for random errors around the basic pricing model (defined by Q_t^0) relatively small, one can characterize their distribution by the first two moments as Clément, Gouriéroux and Monfort (1993) did. Within this framework, we can introduce random error terms *around* an option pricing model defined by Q_t^0 which provides option prices:

$$labeleq37\widetilde{\Pi}_t(K) = B(t,T)\int_{\mathcal{S}} (s_T - K)^+ Q_t^0(ds_T) \tag{3.7}$$

Let us assume for the moment that λ and the parameters definingQ_t^0 are included in the statistician's information set I_t^S at time t while the trader's information set includes (for completeness) both I_t^S and ω. Then, if the pricing probability measure Q_t is described by a Dirichlet model around Q_t^0, *i.e.*, $Q_t \mid I_t^S \rightsquigarrow \mathcal{D}i(\lambda Q_t^0)$, then using the properties of Dirichlet processes we know that the expectation of $\pi_t(\omega)$ *with respect to the draw* of ω is

$$\widetilde{\Pi}_t(K) = E^{\omega}\Pi_t(K,\omega) = B(t,T)E^{Q_t^0}(S_T - K)^+.$$

Hence, for a set of strike prices $K_j, j = 1, 2, \ldots, J_t$, the random variables:

$$u_t(K_j,\omega) = \Pi_t(K_j,\omega) - \widetilde{\Pi}_t(K_j)$$

are zero-mean error terms whose joint probability distribution $\big(u_t(K_j,\omega)\big)_{1\le j\le J_t}$ can be easily deduced from the properties of the Dirichlet process. This can be used to characterize the joint probability distribution of error terms and especially their heteroskedasticity, autocorrelation, skewness, kurtosis, etc., whatever the cross-sectional set of option prices written on the same asset we observe (calls, puts, various strike prices, various maturities ...). For the sake of simplicity of the presentation let us consider only simple one-period models.[14]

First we will randomize the risk-neutral probability around the lognormal distribution of the Black-Scholes model. We suppose that $B(t,T)$ and σ are known. Let us introduce $\tilde{S}_t$ as a latent price of the underlying asset which will appear in the definition of the parameters of the Dirichlet process, more precisely in Q_t^0 (we will justify the use of a latent $\widetilde{S}_t$ later). The resulting model then is:

[13] We thank Professor Rolin for drawing our attention on this result.

[14] For extensions of the models considered in this subsection to more complicated multi-period models see Clément, Gouriéroux and Monfort (1993) and Patilea and Renault (1995).

$$
\begin{aligned}
&Q_t^0 = LN\big(\log \tilde{S}_t - \log B(t,T) - \sigma^2(T-t)/2, \sigma^2(T-t)\big),\\
&S_T \mid \big(Q_t(\,\cdot\,,\omega), I_t^S, \widetilde{S}_t\big) \rightsquigarrow Q_t(\,\cdot\,,\omega),\\
&Q_t \mid (I_t^S \widetilde{S}_t) \rightsquigarrow \mathcal{D}i(\lambda Q_t^0).
\end{aligned}
\tag{3.8}
$$

In this context S_T is what is usually called a sample of size one of the Dirichlet process Q_t, i.e. the conditional distribution of S_T given the realization $Q_t(\,\cdot\,,\omega)$, is $Q_t(\,\cdot\,,\omega)$. It can be shown (see Ferguson (1973)) that the distribution of S_T marginalised with respect to ω is Q_t^0. In this model, given $\widetilde{S}_t$, the price $\Pi_t(K,\omega)$ of an European call option with maturity date T and strike K is defined as in (3.6). In particular the stock price observed at time t is $S_t = S_t(\omega) = \Pi_t(0,\omega)$. With probability one, the observed price will not coincide with the latent price $\widetilde{S}_t$. We introduced $\tilde{S}_t$ not only to make the option pricing formula (3.6) coherent but also because it is an interesting variable to be taken into account.[15] Indeed, if we accept the existence of non-synchronous trading, it is clear that agents willing to buy and sell options have in mind a latent price of the underlying asset. This price $\widetilde{S}_t$ can be viewed as a latent factor, which has various interpretations previously encountered in the option pricing literature. Indeed, Manaster and Rendleman (1982) argue for instance that "just as stock prices may differ, in the short run, from one exchange to another (...), the stock prices implicit in option premia may also differ from the prices observed in the various markets for the stock. In the long run, the trading vehicle that provides the greatest liquidity, the lowest trading costs, and the least restrictions is likely to play the predominant role in the market's determination of the equilibrium values of underlying stocks". Moreover, "investors may regard options as a superior vehicle" for several reasons like trading costs, short sales, margin requirements... Hence, option prices involve *implicit stock prices* that may be viewed as the option market's assessment of equilibrium stock values and may induce a reverse causality relationship from option market to stock market.[16]

Henceforth we will write the models in terms of returns because this setting is better suited for dynamic extensions. We can write (3.8) also as follows:

[15] In their paper Clément, Gouriéroux and Monfort (1993) neglected the incoherence in their option pricing formula. They considered Q_t^0 based on the observed stock price S_t. Thus, at time t, almost surely, they have two stock prices: S_t and $\Pi_t(0,\omega)$.

[16] See also Longstaff (1995) for a related interpretation. While Manaster and Rendleman (1982) and Longstaff (1995) compute implicit stock prices through the BS option pricing formula, Patilea, Ravoteur and Renault (1995) propose an econometric approach in a Hull and White (1987) (HW) setting which is also based on the concept of stock prices implicit in option prices but without choosing between the above theoretical explanations. Indeed, following the state variables methodology set forth by Renault (1996) they argue that if we observe mainly two liquid option contracts at each date: one near the money and another one more speculative (in or out) we need to introduce two unobserved state variables: the first one is stochastic volatility (to apply HW option pricing) and the second one is an "implicit" stock price which is taken into account to apply the HW option pricing formula. They show that even a slight discrepancy between S_t and $\widetilde{S}_t$ (as small as 0.1%, while Longstaff (1995) documents evidence of an average discrepancy of 0.5%) may explain a sensible skewness in the volatility smile.

$$
\begin{aligned}
&Q_t^0 = LN\big(-\log B(t,T) - \sigma^2(T-t)/2, \sigma^2(T-t)\big),\\
&Z_T \overset{\text{def}}{=} \frac{S_T}{\tilde{S}_t} \mid \big(Q_t(\,\cdot\,,\omega), I_t^S, \widetilde{S}_t\big) \sim Z_T \mid Q_t(\,\cdot\,,\omega) \rightsquigarrow Q_t(\,\cdot\,,\omega),\\
&Q_t \mid (I_t^S, \tilde{S}_t) \rightsquigarrow \mathcal{D}i(\lambda Q_t^0).
\end{aligned}
\tag{3.9}
$$

The distribution of Z_T given $\widetilde{S}_t$ is Q_t^0. We should note that this distribution does not depend on $\tilde{S}_t$ and therefore, the option pricing formula will be homogeneous with respect to $(\widetilde{S}_t, K)$, an issue which was deemed important in section 2 to conduct statistical analysis. Indeed, given $\tilde{S}_t$ the price of a call option written in (3.6) becomes

$$
\begin{aligned}
\Pi_t(K,\omega) = \Pi_t(\widetilde{S}_t, K, \omega) &= B(t,T)\int_{\mathcal{S}} (\widetilde{S}_t z_T - K)^+ Q_t(\mathrm{d}z_T, \omega)\\
&= B(t,T)\widetilde{S}_t \int_{\mathcal{S}} (z_T - K/\tilde{S}_t)^+ Q_t(\mathrm{d}z_T, \omega).
\end{aligned}
\tag{3.10}
$$

Using the properties of the Dirichlet process functionals we can compute the moments (conditionally on $\widetilde{S}_t$) of $\Pi_t(K,\omega)$:

$$
\begin{aligned}
&E\big(\Pi_t(K,\omega)\big) = BS(\widetilde{S}_t, K, \sigma),\\
&Var\big(\Pi_t(K,\omega)\big) = (1+\lambda)^{-1}\widetilde{S}_t^2 B(t,T)^2 \left[E^{Q_t^0}(f^2) - \big(E^{Q_t^0}(f)\big)^2\right],\\
&Cov\big(\Pi_t(K_1,\omega), \Pi_t(K_2,\omega)\big) = (1+\lambda)^{-1}\widetilde{S}_t^2 B(t,T)^2 \left[E^{Q_t^0}(f_1 f_2) - E^{Q_t^0}(f_1)E^{Q_t^0}(f_2)\right],
\end{aligned}
$$

where $f(z) = \left[z - \dfrac{K}{\tilde{S}_t}\right]^+$ and $f_j(z) = \left[z - \dfrac{K_j}{\tilde{S}_t}\right]^+, j = 1, 2.$

This shows, as noted before, that the heteroskedasticity and the autocorrelation structure of error terms around the BS price depend in a highly complicated nonlinear way on the underlying characteristics of the options: strike prices, times to maturity, etc.[17] We also obtain the price of the stock at time t as

$$
S_t = S_t(\omega) = B(t,T)\widetilde{S}_t \int_{\mathcal{S}} z_T \Pi(\mathrm{d}z_T, \omega) \overset{\text{def}}{=} \widetilde{S}_t m_t(\omega).
$$

The error term $m_t(\omega)$ is a functional of a Dirichlet process of parameter $\lambda LN\big(-\sigma^2(T-t)/2, \sigma^2(T-t)\big)$ and:

$$
E\big(m_t(\omega)\big) = 1, \quad Var\big(m_t(\omega)\big) = (\lambda+1)^{-1}\left(e^{\sigma^2(T-t)} - 1\right)
$$

Moreover, for large values of λ, the observed price S_t, given $\widetilde{S}_t$, is approximatively normally distributed with mean equal to $\widetilde{S}_t$ and variance $\widetilde{S}_t^2\, Var\big(m_t(\omega)\big)$. Hence, one may choose λ in such way that the variance of S_t does not depend on $T-t$.

A first extension of (3.8) and (3.9) may be obtained by considering a Dirichlet process around the risk-neutral probability of Merton's (1973) model. For this consider:

$$
V_t^T = \int_t^T \sigma_u^2 \mathrm{d}u.
$$

[17] Clément, Gouriéroux and Monfort (1993) provide in detail the explicit formulas and suggest a simulation-based methodology.

Conditionally on V_t^T, one can draw Dirichlet realizations around the lognormal distribution $LN\big(-\log B(t,T) - V_t^T/2, V_t^T\big)$ and therefore:

$$
\begin{aligned}
Q_t^0 &= Q_t^0(V_t^T) = LN\big(-\log B(t,T) - V_t^T/2, V_t^T\big),\\
Z_T &= \frac{S_T}{\widetilde{S}_t} \mid Q_t(\,\cdot\,,\omega), I_t^S, \sigma_t, V_t^T, \widetilde{S}_t \sim V_T \mid Q_t(\,\cdot\,,\omega) \rightsquigarrow Q_t(\,\cdot\,,\omega),\\
Q_t &\mid \big(I_t^S, \sigma_t, V_t^T, \widetilde{S}_t\big) \rightsquigarrow \mathcal{D}i(\lambda Q_t^0).
\end{aligned}
$$

One can draw first V_t^T from a conditional distribution, given σ_t (to be specified). As a result we obtain that $Q_t \mid \sigma_t$ is a mixture of Dirichlet processes.[18] The call option formula, given $\widetilde{S}_t, V_t^T$ and σ_t, does not depend on V_t^T and σ_t and is exactly as in (3.10):

$$
\Pi_t(K,\omega) = \Pi_t(\widetilde{S}_t, K, \omega) = B(t,T)\widetilde{S}_t \int_{\mathcal{S}} (z_t - K/\widetilde{S}_t)^+ Q_t(\mathrm{d}z_T, \omega). \tag{3.11}
$$

The mean of $\Pi_t(K,\omega)$, conditionally on $\widetilde{S}_t$ and σ_t, is

$$
E\big(\Pi_t(K,\omega)\big) = E_{\sigma_t}\big(BS(\tilde{S}_t, K, V_t^T)\big)
$$

Since (3.11) is very similar to several well-known extensions of the BS option pricing model, including the Merton (1976) jump diffusion and Hull and White (1987) stochastic volatility models, we may conclude that the class of mixtures of Dirichlet processes allows us to introduce error terms around any extension of the BS model where unobserved heterogeneity (like stochastic volatility) has been introduced.

To conclude, this analysis suggests that we should distinguish two types of option pricing errors: (1) errors due to a limited set of unobserved state variables like stochastic volatility, stochastic interest rate, $\widetilde{S}_t$, and (2) errors to make the model consistent with any data set. In the first approach state variables are introduced as instruments to define mixtures of Dirichlet processes around the "structural" option pricing model. Therefore, the suggested approach is not fully model-free since the model is built *around* a structural model defined by Q_t^0, with a parameter λ which controls the level of neighboring around this model.[19] This is the price to pay to take into account arbitrage restrictions. We have therefore only two solutions. Either we adopt a semiparametric approach by introducing a nonparametric disturbance around a given probability measure Q_t^0. While this was done above in a Bayesian way, it will be done in section 4 in a classical way through the concept of functional residual plots. Alternatively we consider a genuine

[18] See Antoniak (1974) for the definition and the properties of the mixtures of Dirichlet processes. The fact that a mixture of Dirichlet processes is conditionally a Dirichlet process allows one to carry over many properties of Dirichlet processes to mixtures. Moreover, one can show that any random probability on $\big(\mathcal{S}, \mathcal{B}(\mathcal{S})\big)$ can be approximated arbitrarily closely in the sense of the weak convergence for distributions, by a mixture of Dirichlet processes. Hence, the richness of the class of mixtures of Dirichlet processes suggests that it enables us to build general models.

[19] This parameter λ must not be confused with the smoothing parameter λ in section 2, which was also devised to control the level of neighboring, but with a very different concept of "closeness" or "neighborhood".

nonparametric estimation of the equivalent martingale measure. This is the issue addressed in the following subsection 3.2.

3.2 Nonparametric estimation of state-price densities implicit in financial asset prices

We observed in (3.4) that the risk neutral probability distribution or "state price density" can be recovered from taking derivatives of European calls with respect to their strike price. Kernel estimation techniques provide an estimate of the pricing function f. Provided that they exist, it is straightforward to recover estimates of the derivatives of f from $\hat{f}$. This is particularly important for option valuation in the context of complete markets with no arbitrage opportunities. In this situation, several of the partial derivatives of the pricing function are of special interest. One of them is the "*delta*" of the option defined as $\Delta \equiv \partial \Pi / \partial S$.[20] Another derivative of special interest is $\partial^2 \Pi / \partial K^2$ since it is related by (3.4) to the *state price density*.

The use of kernel methods for deriving estimates of Δ and the state price density is due to Aït-Sahalia and Lo (1995). Many of the issues raised in section 2 regarding kernel estimation and the nonstationarity of the data, the dependence of the data etc. apply here as well of course. In addition, some additional issues should be raised as well. Indeed, we noted in section 2 that kernel smoothing is based on a certain approximation criterion. This approximation criterion applies to the estimation of the function f but not necessarily its derivatives. The bandwidth selection affects the smoothness of the estimate $\hat{f}$ and therefore indirectly its derivatives. Since the ultimate objective is to estimate the derivatives of the function rather than the function itself it is clear that the choice of objective function and approximation criterion of standard kernel estimation are not appropriate. It is a drawback of this approach that still needs to be investigated in greater detail.

Up to now, we presented two nonparametric approaches to the option valuation problem. The first one, the pure nonparametric pricing, makes very little use of the economic or financial dimensions of the problem and relies almost exclusively on the statistical exploitation of *market* data. The second one incorporates elements of a rational option pricing theory: it exploits the equivalence between the absence of arbitrage assumption and the existence of a risk neutral probability measure to derive the pricing formula. From this relation, it appears that the parameter to be estimated is the risk neutral density. The next section presents a third way which can be seen as a blend of the previous approaches.

[20] This quantity is useful as it determines the quantity of the underlying stock an agent must hold in a hedging portfolio that replicates the call option.

4 Extended Black and Scholes models and objective driven inference

Practitioners recognize that the assumptions of constant dividends, interest rates and volatility of the Black and Scholes (BS) model are not realistic. The most revealing evidence of this is the systematic use of the BS formula as a pricing and hedging tool by practitioners through the so-called BS implicit volatility, that is the volatility measure which equates the BS option valuation formula to the observed option price:

$$\Pi_t = BS\left(S_t, \tau, K, r, \sigma_t(S_t, K)\right), \tag{4.1}$$

where Π_t denote the observed call price at time t and $\sigma_t(S_t, K)$ the corresponding BS implicit volatility. This practice can be assimilated in a forecasting rule where the BS formula is used as a *black box* which integrates the time varying and stochastic environment through the volatility parameter. Since the BS formula is inherently misspecified we can think of modifying the underlying model so that it incorporates these new features, such as stochastic volatility. However such attempts lead to very complex models which usually do not admit a unique and closed form solution except for some special cases.

Taking into account that (1) analytic extensions of BS often pose many computational difficulties, (2) the BS formula is not a valid *modeling* tool, (3) the BS formula is used as a *prediction* tool, Gouriéroux, Monfort and Tenreiro (1994, 1995), henceforth GMT, present the statistical foundations of dealing with a misspecified BS model. We first present the GMT approach in its general formulation and then show how it can be applied to model modified BS formula.

4.1 Kernel M-estimators

To discuss the generic setup of kernel M-estimators, let us suppose that we observe a realization of length T of a stationary stochastic process $\{(Z_t, Y_t) : t \in \mathbf{Z}\}$ and that the parameter vector parametrizes the conditional distribution of Y_t given $\mathcal{F}_t = \sigma(Z_s, Y_{s-1}, s \leq t)$. The estimation strategy proposed by GMT tries to approximate a *functional* parameter vector $\theta(\mathcal{F}_t)$ implicitly defined as the solution of:

$$\min_\theta \; E_0\left[\psi(U_t, \theta) | \mathcal{F}_t\right], \tag{4.2}$$

where, $U_t = (Z_t, Y_t, Y_{t-1})$, ψ is an objective function, and $E_0(\,\cdot\,|\mathcal{F}_t)$ denotes the conditional expectation with respect to the unknown true conditional distribution of U_t given $(Z_s, Y_{s-1}, s \leq t)$. Let X be a d-dimensional process such that the solution $\theta(X_t)$ of

$$\min_\theta E_0\left[\psi(U_t, \theta) | X_t\right] \tag{4.3}$$

coincides with that of (4.2). GMT suggest approximating (4.3) by

$$\min_\theta \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_T^d}K\left(\frac{x_t-x}{h_T}\right)\psi(u_t,\theta), \tag{4.4}$$

where $K(\cdot)$ is a kernel function, and h_T is the bandwidth, depending on T, the sample size. Estimators obtained according to (4.4) are denoted $\hat{\theta}_T(s)$ and called kernel M-estimators, since they are derived by *minimizing* an "empirical" criterion, which is a *kernel* approximation of the unknown theoretical criterion to be minimized appearing in (4.3). Under suitable regularity conditions [see Gouriéroux, Monfort and Tenreiro (1995)], $\hat{\theta}_T(x)-\theta(x)$ converges to 0, where $\theta(x)$ denotes the solution of $\min_\theta E_0\left(\psi(U_t,\theta)|X_t=x\right)$.

Gouriéroux, Monfort and Tenreiro (1995) also show how local versions of such estimators may be used to compute *functional residuals* in order to check the hypothesis of a constant function $\theta(\cdot)$. In other words, as is fairly standard in econometrics, error terms (and residual plots) are introduced to examine whether unobserved heterogeneity is hidden in seemingly constant parameters θ. The GMT contribution is to give a nonparametric appraisal of these error terms, which justifies the terminology " *functional* residual plots". GMT show how their functional residual plots are related to some standard testing procedures for the hypothesis of parameter constancy and how they may be introduced as important tools in a modeling strategy. A Bayesian alternative is considered by Jacquier and Jarrow (1995) who suggest to consider draws from the posterior distribution of parameters of an extended model in order to deduce some draws of the residual vector and to perform a Bayesian residual analysis.

4.2 Extended Black-Scholes formulations

The methodology proposed by GMT can be applied to the problem of option pricing by extending the Black-Scholes formulation. The search for such a prediction model is made starting from the BS formula, namely we consider that $E_0(\Pi_t|X_t)$ belongs to a model containing functions of X_t as

$$BS\big(S_t,K,\tau,\theta(X_t)\big), \tag{4.5}$$

where θ is the vector of parameters (r,σ) which enters the standard BS formula and X_t is a vector of state variables which is believed to affect the volatility and the interest rate. For instance, we may decide to include in X_t variables such as $S_{t-1}, K, \ldots$. This approach is very much in line with the idea of computing BS implicit volatilities except that it involves a statistically more rigorous scheme which also serves as a basis for building new prediction tools. GMT propose choosing the objective ψ function

$$\psi(U_t,\theta)=\left[\Pi_t-\ BS\big(S_t,K,\tau,\theta\big)\right]^2,$$

with $U_t = (S_t, \Pi_t)$.[21] If the model (4.5) is correct and $\theta_0(X_t)$ is unique such that

$$E_0(\Pi_t|X_t) = BS\big(S_t, K, \tau, \theta_0(X_t)\big)$$

holds, then $\theta_0(X_t)$ is unique verifying

$$\theta_0(X_t) = \arg\min_\theta E_0\big[\psi(U_t,\theta)|X_t\big].$$

Therefore, if the regularity conditions are satisfied, the convergence result of kernel M-estimators ensures that for T large enough, $\hat\theta_T(x)$ will be close to $\theta_0(x)$. This is a justification to the use of the modified BS formula $BS\big(S_t, K, \tau, \hat\theta_T(X_t)\big)$ as a predictor of option prices.

One of the conditions to be imposed to obtain the convergence of $\hat\theta_T(\,\cdot\,)$ is the stationarity of the process (S_t, Π_t). Such an assumption is hardly sustainable in view of the stylized facts concerning the variables entering this process. To remedy this problem, GMT suggest using the homogeneity of degree one in (S, K) of the BS function together with the measurability of S_t with respect to $\sigma(X_t)$. Indeed, this leads to a new pricing relationship:

$$E_0\left[\frac{\Pi_t}{S_t}|X_t\right] = \frac{1}{S_t} BS\big(S_t, \tau, K, \theta(X_t)\big) = BS^*\big(\tau, k_t, \theta(X_t)\big), \tag{4.6}$$

where $k_t \equiv K/S_t$ is the inverse of the moneyness ratio. In this formulation, all the prices are expressed in terms of the time t underlying asset price. Then a new objective function is now $\psi^*(U_t^*, \theta^*) \equiv \left[\frac{\Pi_t}{S_t} - BS^*\big(\tau, \theta^*\big)\right]^2$, where $U_t^* = \Pi_t/S_t$ and $\theta^* = (r, \sigma, k)$ and the kernel M-estimator, denoted $\hat\theta^*_T(x)$, is the solution of:

$$\min_{\theta^*} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{h_T^d} K\left(\frac{x_t - x}{h_T}\right) \psi^*(U_t^*, \theta^*).$$

An issue that arises when implementing the GMT approach is the choice of the variables to be included in X_t. This problem can be seen as a problem of model choice which arises very often in econometrics. Gouriéroux, Monfort and Tenreiro (1995) propose a modelling approach based on the use of functional residual plots and confidence bands for these residuals. In a first step, the parameter θ is assumed to be constant and is estimated by θ^*_{0T} obtained by minimizing the sample average of $\psi^*(U_t^*, \theta^*)$. Then a state variable X_1 potentially affecting the parameter is introduced. In order to test whether θ^* should be considered as constant or depending on X_1, one computes in a second step *functional residuals*. They are defined as the difference between an approximation of $\hat\theta^*_T(x_1)$ near the constancy hypothesis and θ^*_{0T}. Confidence bands on these residuals help to determine whether they depend on X_1. A plot of these residuals against X_1 is also helpful in choosing a parametric form for expressing the relationship between θ^* and X_1. In

[21] In our case, since we are interested in option pricing or option price prediction, the ψ function is related to a pricing error. However if option hedging is the main goal, the ψ may be chosen accordingly as a tracking error.

case of a rejection of the constancy hypothesis, the parameters of this relationship are considered as the new functional parameters to be estimated. And the procedure goes on by repeating the previous steps with new state variables.

The approach discussed so far can be extended to models other than the BS. For instance, for American type options one may replace the BS formula with formula tailored for such options, typically involving numerical approximations. Like the BS formula, they rest on the restrictive assumptions of constant volatility, interest rates, etc. The M-estimators approach suggested by GMT readily extends to such and other applications.

References

[1] Abken, P.A., D.B. Madan and S. Ramamurtie (1996) "Basis Pricing of Contingent Claims: Application to Eurodollar Futures Options", Discussion Paper, Federal Reserve Board of Atlanta.

[2] Aït-Sahalia, Y. (1993) "Nonparametric Functional Estimation with Applications to Financial Models", *Ph.D. dissertation*, M.I.T.

[3] Aït-Sahalia, Y. (1996) "Nonparametric Pricing of Interest Rate Derivative Securities", *Econometrica* 64, 527 – 560.

[4] Aït-Sahalia, Y., P. Bickel and T. Stoker (1995) "Goodness-of-Fit Tests for Regression Using Kernel Methods", Discussion Paper, University of Chicago.

[5] Aït-Sahalia, Y. and A. W. Lo (1995) "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices", *Discussion Paper*, Sloan School of Management, M.I.T.

[6] Anscombe F.J. (1990), "Residuals" in *New Palgrave - Time Series and Statistics*, 244 – 250.

[7] Antoniak C. (1974), "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems", *Annals of Statistics* 2, 1152 – 1174.

[8] Altman, N. S. (1987) "Smoothing Data with Correlated Errors", *Technical Report 280*, Department of Statistics, Stanford University.

[9] Altman, N. S. (1990) "Kernel Smoothing of Data with Correlated Errors", *Journal of the American Statistical Association* 85, 749 – 759.

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

[11] Baum, C. and J. Barkoulas (1996) "Essential Nonparametric Prediction of U.S. Interest Rates", *Discussion Paper*, Boston College.

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

[13] Bossaerts P. and P. Hillion (1993), "A Test of a General Equilibrium Stock Option Pricing Model", *Mathematical Finance* 3, 311 – 347.

[14] Bossaerts P. and P. Hillion (1994), "Local Parametric Analysis of Hedging in Discrete Time", Discussion Paper, CentER, Tilburg University.

[15] Broadie, M., J. Detemple, E. Ghysels and O. Torrès (1995) "Nonparametric Estimation of American Options Exercise Boundaries and Call Prices", *Discussion Paper*, CIRANO, Montréal.

[16] Broadie, M., J. Detemple, E. Ghysels and O. Torrès (1996) "American Options with Stochastic Volatility and Stochastic Dividends: A Nonparametric Investigation", *Discussion Paper*, CIRANO, Montréal.

[17] Cao, R., A. Quintela-del-Rió and J. M. Vilar-Fernández (1993) "Bandwidth Selection in Nonparametric Density Estimation Under Dependence: A Simulation Study", *Computational Statistics* 8, 313 – 332.

[18] Clément E., C. Gouriéroux and A. Monfort (1993), "Prediction of Contingent Price Measures", Discussion Paper, CREST.

[19] Craven, P. and G. Wahba (1979) "Smoothing Noisy Data with Spline Functions", *Numerical Mathematics* 31, 377 – 403.

[20] Duffie D. (1996), *Dynamic Asset Pricing Theory, 2nd Edition*, Princeton University Press.

[21] Dumas, B., J. Fleming and R.E. Whaley (1995) "Implied Volatility Functions: Empirical Tests", *Discussion Paper*, HEC, Paris.

[22] Garcia, R. and E. Renault (1996) "Risk Aversion, Intertemporal Substitution and Option Pricing", *Discussion Paper*, CIRANO, Montréal and GREMAQ, Université de Toulouse I.

[23] Ghysels, E., A. Harvey and E. Renault (1996) "Stochastic Volatility", in G.S. Maddala and C.R. Rao (ed.) *Handbook of Statistics – Vol. 14, Statistical Methods in Finance*, North Holland, Amsterdam, Chapter 5.

[24] Ghysels, E. and S. Ng (1996) "A Semiparametric Factor Model of Interest Rates", *Discussion Paper*, CIRANO, Montréal.

[25] Ghysels E. and O. Torrès (1996), "Nonparametric Calibration of Option Price Formula", Work in progress.

[26] Gouriéroux, C., A. Monfort and C. Tenreiro (1994), "Kernel M-Estimators: Nonparametric Diagnostics for Structural Models", *Discussion Paper 9405*, CEPREMAP, Paris.

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

[28] Györfi, L., W. Härdle, P. Sarda and P. Vieu (1989) *Nonparametric Curve Estimation from Time Series*, Lecture Notes in Statistics 60, J. Berger *et al.*, eds., Springer-Verlag, Heidelberg.

[29] Härdle, W. (1990) *Applied Nonparametric Regression*, Cambridge University Press.

[30] Härdle, W. and O. Linton (1994) "Applied Nonparametric Methods", in R. F. Engle and D. L. McFadden (ed.) *Handbook of Econometrics*, vol. 4, North Holland, Amsterdam.

[31] Härdle, W. and P. Vieu (1992) "Kernel Regression Smoothing of Time Series", *Journal of Time Series Analysis* 13, 209 – 232.

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

[33] Harvey, C. R. and R. E. Whaley (1992) "Dividends and S&P 100 Index Option Valuation", *Journal of Futures Markets*, 12, 123 –137.

[34] Hastie, T. J. and R. J. Tibshirani (1990) *Generalized Additive Models*, Chapman & Hall.

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

[36] Hutchison, J. M., A. W. Lo and T. Poggio (1994) "A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks", *Journal of Finance*, 49, 851 – 889.

[37] Jacquier E. and R. Jarrow (1995), "Dynamic Evaluation of Contingent Claim Models", *Discussion Paper*, Cornell University.

[38] Kim, T. Y. and D. D. Cox (1996) "Bandwidth Selection in Kernel Smoothing of Time Series", *Journal of Time Series Analysis*, 17, 49 – 63.

[39] Lavergne, L. and Q. Vuong (1996), "Nonparametric Selection of Regressors: The Nonnested Case", *Econometrica* 64, 207-219.

[40] Lo, A.Y. (1987), "A Large Sample Study of the Bayesian Bootstrap", *The Annals of Statistics*, vol. 15, no 1, 360-375.

[41] Merton, R. C. (1973) "Theory of Rational Option Pricing ", *Bell Journal of Economics*, 4, 141 – 183.

[42] Patilea V., Ravoteur M.P. and E. Renault (1995), "Multivariate Time Series Analysis of Option Prices ", Working Paper GREMAQ, Toulouse.

[43] Patilea V. and E. Renault (1995), "Random Probabilities for Option Pricing", Discussion Paper, GREMAQ, Toulouse.

[44] Renault, E. (1996) "Econometric Models of Option Pricing Errors", *Discussion Paper*, GREMAQ, Université de Toulouse I.

[45] Rolin, J-M (1992), "Some Useful Properties of the Dirichlet Process", *Discussion Paper*, Institute of Statistics, Univ. Catholique de Louvain.

[46] Rubinstein, M. (1994), "Implied Binomial Trees", *Journal of Finance* 49, 771-818.

[47] Schuster, E. F. (1972) "Joint Asymptotic Distribution of the Estimated Regression Function at a Finite Number of Distinct Points", *Annals of Mathematical Statistics*, 43, 84 – 88.

[48] Scott, D. W. (1992) *Multivariate Density Estimation: Theory, Practice and Visualization*, John Wiley & Sons Inc., New York.

[49] Silverman, B. W. (1984) "Spline Smoothing: the Equivalent Variable Kernel Method", *Annals of Statistics*, 12, 898 – 916.

[50] Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman & Hall.

[51] Stutzer, M. (1995) "A Simple Nonparametric Approach to Derivative Security Valuation", *Discussion Paper*, Carlson School of Management, University of Minnesota .

[52] Wahba, G. (1990): *Spline Models For Observational Data*, Conference Board of Mathematical Sciences – National Science Foundation (CBMS – NSF) Regional Conference Series, 59, Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

[53] Wand, M. P. and M. C. Jones (1995) *Kernel Smoothing* Chapman & Hall, London.