Nonparametric estimation of American options' exercise boundaries and call prices

Mark Broadie[a,*], Jérôme Detemple[b], Eric Ghysels[c], Olivier Torrès[d]

[a]*Graduate School of Business, Columbia University, New York, NY 10027, USA*
[b]*Boston University, School of Management, Boston, MA 02215, USA*
[c]*Department of Economics, Penn State University, University Park, PA, 16802, USA*
[d]*UFR de Mathematiques, Sciences Economiques et Sociales, Université de Lille 3, Villeneuve d' Asq, France*

Abstract

Unlike European-type derivative securities, there are no simple analytic valuation formulas for finite-lived American options, even when the underlying asset price has constant volatility. The early exercise feature considerably complicates the valuation of American contracts. The strategy taken in this paper is to rely on nonparametric statistical methods using market data to estimate the call prices and the exercise boundaries. A comparison is made with parametric constant volatility model-based prices and exercise boundaries. The paper focuses on assessing the adequacy of conventional formulas by comparing them to nonparametric estimates. We use daily market option prices and exercise data on the S&P100 contract, the most actively traded American option contract. We find large discrepancies between the parametric and nonparametric call prices and exercise boundaries. We also find remarkable similarities of the nonparametric estimates before and after the crash of October 1987.

1. Introduction

American option contracts figure prominently among the wide range of derivative securities which are traded. An American call option not only

---

*Corresponding author. Tel.: + 1-212-854-4103; fax: + 1-212-316-9180.

E-mail addresses: mnb2@columbia.edu (M. Broadie), detemple@bu.edu (J. Detemple), eghysels@psu.edu (E. Ghysels), torres@univ-lille3.fr (O. Torrès).

provides the possibility of buying the underlying asset at a particular strike price, but it also allows the owner to exercise his right at any point in time before maturity. This early exercise feature of the contract considerably complicates its evaluation. Indeed, the option price critically depends on the *optimal* exercise policy which must be determined in the evaluation process. The earliest analysis of the subject by McKean (1965) formulates the valuation of the derivative security as a free boundary problem. Additional insights about the properties of the optimal exercise boundary are provided by Van Moerbeke (1976) and more recently by Barles et al. (1995), Büttler and Waldvogel (1996) and Kuske and Keller (1998). Bensoussan (1984) and Karatzas (1988) provide a formal financial argument for the valuation of an American contingent claim in the context of a general market model, in which the price of the underlying asset follows an Itô process. It should not come as a surprise that the distributional properties of the underlying asset price determine those of the exercise boundary. However, in such a general context, analytical closed-form solutions are typically not available and the computations of the optimal exercise boundary and the contract price can be achieved only via numerical methods. The standard approach consists of specifying a process for the underlying asset price, generally a geometric Brownian motion process (GBM), and uses a numerically efficient algorithm to compute the price and the exercise boundary. A whole range of numerical procedures have been proposed, including binomial or lattice methods, methods based on solving partial differential equations, integral equations, or variational inequalities, and other approximation and extrapolation schemes.[1]

The first purpose of our paper is to suggest a new and different strategy for dealing with the pricing of American options and the characterization of the exercise boundary. The paper does *not* come up with a new twist that boosts numerical efficiency or a major innovation in algorithm design. Instead, it suggests a different approach which consists of using *market* data, both on exercise decisions and option prices, and relies on *nonparametric* statistical techniques. The idea of applying nonparametric methods to option pricing has been suggested recently in a number of paper, e.g., Abadir and Rockinger (1997), Abken et al. (1996), Aït-Sahalia (1996), Aït-Sahalia and Lo (1998), Broadie et al. (2000), Gouriéroux et al. (1994), Hutchinson et al. (1994), Jackwerth and Rubinstein (1996), Stutzer (1996) and several others. As there are a multitude of nonparametric methods it is no surprise that the aforementioned papers use different methods. Moreover, they do not address the same topics either. Indeed,

---

[1] A partial list of contributions to this area includes Brennan and Schwartz (1977), Cox et al. (1979), Geske (1979), Whaley (1981), Geske and Johnson (1984), Barone-Adesi and Whaley (1987), Selby and Hodges (1987), Boyle (1988), Breen (1991), Yu (1993), Broadie and Detemple (1996) and Carr and Faguet (1994), among others. For a review and comparison of these procedures, see Broadie and Detemple (1996).

some aim for nonparametric corrections of a standard (say Black–Scholes) option pricing formula, others estimate risk-neutral densities, etc. So far this literature has focused exclusively on European type options. By studying American options, our paper models both pricing and exercise strategies via nonparametric methods.[2] It is worth noting that the approach taken in this paper is somewhat similar to that of Hutchinson et al. (1994), except that we use kernel-based estimation methods instead of neural networks. The approach we suggest can handle a fairly rich class of processes with multiple state variables, including stochastic volatility, stochastic dividends and interest rates (see Broadie et al., 2000).

The second purpose of our paper is to use the nonparametric models to assess the adequacy of conventional American asset pricing models. Let us illustrate this intuitively using the case of the exercise boundary. Suppose that we have observations on the exercise decisions of investors who own American options, along with the features of the contracts being exercised. Such data are available for instance for the S&P100 Index option or OEX contract, as they are collected by the Option Clearing Corporation (OCC).[3] The idea is that with enough data, such as ten years of daily observations, we should be able to gather information about how market participants perceive the exercise boundary. Our approach can be seen as a way to characterize the exercise boundary for American options using observations on exercises.[4] This empirical boundary can then be compared to the boundary computed via the usual algorithms. This idea can also be applied to the pricing of the option, again assuming that we have data on call and put contracts and their attributes. Unlike exercise data, option price data are quite common and figure prominently in several financial data bases. The empirical application reported in the paper involves three types of data, namely: (1) time-series data on the asset or index underlying the option contract, (2) data on call and put prices obtained from the CBOE, and (3) data on exercise decisions recorded by the OCC.

Section 2 is devoted to a brief review of the literature on American option pricing. Section 3 covers parametric and nonparametric estimation of exercise

---

[2] Bossaerts (1988) and de Matos (1994) are to our knowledge the only papers discussing some of the theoretical issues of estimating American option exercise boundaries. We do not know of any empirical work attempting to estimate such boundaries.

[3] Option exercise data have been used in a number of studies, including Ingersoll (1977), Bodurtha and Courtadon (1986), Overdahl (1988), Dunn and Eades (1989), Gay et al. (1989), Zivney (1991), French and Maberly (1992) and Diz and Finucane (1993).

[4] Questions as to whether market participants exercise 'optimally', regardless of what the model or assumptions might be, will not be the main focus of our paper although several procedures that we suggest would create a natural framework to address some of these issues. For the most recent work on testing market rationality using option exercise data and for a review of the related literature, see Diz and Finucane (1993).

boundaries while Section 4 handles estimation of option prices using similar methods. In both cases a comparison between model-based and data-based approaches is presented. Finally, Section 5 concludes the paper.

2. American option pricing: a brief review

Let us consider an American call option on an underlying asset whose price S follows an Itô process. The option is issued at $t_0 = 0$ and matures at date $T > 0$ with strike price $K > 0$. We adopt the standard specification in the literature and assume that the dividend rate is constant and proportional to the stock price. Consider the policy of exercising the option at time $\tau \in [0, T]$. A call option with automatic exercise at time τ has a payoff $(S_\tau - K)^+$. In the absence of arbitrage opportunities, the price at time $t \in [0, \tau)$ of this contingent claim, $V_t(\tau)$, is given by the expected value of the discounted payoff, where the expectation is taken relative to the equivalent martingale probability measure Q, i.e.

$$V_t(\tau) = E^Q \left[\exp \left(- \int_t^\tau r_s \, ds \right) (S_\tau - K)^+ | \mathscr{F}_t \right], \tag{2.1}$$

where r_s denotes the time s risk-free interest rate in the economy, E^Q denotes the expectation taken with respect to Q (see Harrison and Kreps, 1979) and $\mathscr{F}_{(\cdot)} \equiv \{ \mathscr{F}_t : t \geq 0 \}$ is a filtration on $(\Omega, \mathscr{F})$ the measurable space on which the price process S is defined. Since an American option can be exercised at any time in the interval $[0, T]$, an option holder will choose the policy (i.e. the exercise time) which maximizes the value of the claim in (2.1). This stopping time solves

$$\max_{\tau \in \mathscr{T}_{[0,T]}} V_0(\tau) \tag{2.2}$$

and at any date t the price of the American call is given by

$$C_t = \sup_{\tau \in \mathscr{T}_{[t,T]}} E^Q \left[\exp \left(- \int_t^\tau r_s \, ds \right) (S_\tau - K)^+ | \mathscr{F}_t \right], \tag{2.3}$$

where $\mathscr{T}_{[u,v]}$ is the set of stopping times (w.r.t. $\mathscr{F}_{(\cdot)}$) with values in $[u, v]$. The existence of a τ^* solving (2.2) has been proved by Karatzas (1988) under some regularity conditions on S. Furthermore, the optimal exercise time is the first time at which the option price equals the exercise payoff, i.e.,

$$\tau^* \equiv \inf \{ t \in [0, T] : C_t = (S_t - K)^+ \}. \tag{2.4}$$

This characterization, however, is of limited interest from an empirical point of view since the option price, which determines the optimal exercise policy, is an unknown endogenous function.

A more precise characterization of the optimal exercise policy is obtained if we restrict our attention to the Black–Scholes economy. In this model, the underlying asset price follows the geometric Brownian motion (GBM) process,[5]

$$dS_t = S_t[(r - \delta)\,dt + \sigma\,dW_t^*], \quad t \in [0, T],\ S_0 \text{ given}, \tag{2.5}$$

where δ is the constant dividend rate, r the constant interest rate, σ the constant volatility of the underlying asset price and W^* is a Brownian motion on $(\Omega, \mathscr{F}, \mathscr{F}_{(\cdot)}, Q)$. The price process (2.5) is expressed in its risk-neutral form, i.e., in terms of the equivalent martingale measure. Under these assumptions, the American call option value is given by

$$C_t(S_t, B) = C_t^{\mathrm{E}}(S_t) + \int_t^T [\delta S_t \mathrm{e}^{-\delta(s-t)} \Phi[d_1(S_t, B_s, s - t)]$$
$$- rK\mathrm{e}^{-r(s-t)} \Phi[d_2(S_t, B_s, s - t)]]\,ds, \tag{2.6}$$

where $C_t^{\mathrm{E}}(S_t)$ denotes the price of the corresponding European option, Φ is the cumulative standard normal distribution function and

$$d_1(S_t, B_s, s - t) \equiv (\sigma\sqrt{s - t})^{-1} \times [\log(S_t/B_s) + (r - \delta + \sigma^2/2)(s - t)],$$
$$d_2(S_t, B_s, s - t) \equiv d_1(S_t, B_s, s - t) - \sigma\sqrt{s - t}.$$

In (2.6) the exercise boundary B solves the recursive integral equation,

$$B_t - K = C_t(B_t, B), \quad t \in [0, T), \tag{2.7}$$

$$\lim_{t\uparrow T} B_t = \max\left\{K, \frac{r}{\delta}K\right\}. \tag{2.8}$$

This characterization of the option value and its associated exercise boundary is the early exercise premium representation of the option. It was originally demonstrated by Kim (1990), Jacka (1991) and Carr et al. (1992).[6] The early exercise representation (2.6)–(2.8) of the American call option price is useful since it can be used as a starting point for the design of computational algorithms. In this paper, we implement a fast and accurate procedure proposed by Broadie

---

[5] These assumptions, combined with the possibility of continuous trading, imply that the market is complete. Moreover, in this economy there is absence of arbitrage opportunities. This is the setting underlying the analysis of Kim (1990), Jacka (1991) and Myneni (1992).

[6] This representation is in fact the Riesz decomposition of the value function which arises in stopping time problems. The Riesz decomposition was initially proved by El Karoui and Karatzas (1991) for a class of stopping time problems involving Brownian motion processes. This decomposition is also applied to American put options by Myneni (1992); it has been extended by Rutkowski (1994) to more general payoff processes.

and Detemple (1996), henceforth BD, for the parametric pricing of American options and the estimation of the optimal exercise boundary.[7]

We now turn our attention to the parametric and nonparametric analysis of exercise boundaries. Thereafter, option prices will be covered in the same way.

3. Parametric and nonparametric analysis of exercise boundaries

Probably the only study that addresses the issue of finding an estimate of the optimal exercise boundary is the work by de Matos (1994), which is an extension of Bossaerts (1988). It proposes an estimation procedure which is based on orthogonality conditions which characterize the optimal exercise time for the contract. However, although no particular dynamic equation for S is postulated, de Matos (1994) assumes that the optimal exercise boundary is deterministic and continuous, and approximates it by a finite order polynomial in time, whose parameters are estimated from the moment conditions. In this paper, we use nonparametric cubic splines and kernel-based estimators to extract an exercise boundary from the data.[8] Our approach readily extends to more general models with additional state variables such as models with random dividend payments or with stochastic volatility (see Broadie et al., 1998). The procedure of de Matos is more restrictive and does not generalize easily.

We describe the exercise data for the S&P100 Stock Index American option contract in Section 3.1 and report summary statistics, plots and finally the nonparametric estimates of the exercise boundary using market data. Next in Section 3.2 we use S&P100 Stock Index data to estimate the GBM diffusion and invoke the BD algorithm to produce a parametric boundary. Finally, in Section 3.3 we discuss comparisons of the parametric and nonparametric boundaries.

3.1. Description of the exercise data and nonparametric boundary estimates

The data on the characteristics of S&P100 Index American contracts (maturity, strike price, number of exercises) is the same as in Diz and Finucane (1993) and we refer to their paper for a description of the sources. These are end-of-the-trading-day daily data on S&P100 Index American put and call contracts which are traded on the Chicago Board Options Exchange. The contract is

---

[7] For the boundary estimation, the BD algorithm provides a lower bound on the boundary. We checked that the difference between the bound and the true boundary was small by comparing the results with the recursive integral procedure (using a fine discretization) suggested in Kim (1990) and detailed in Huang et al. (1996).

[8] For a discussion on the relationship between cubic splines and kernel methods, see the appendix to the paper.
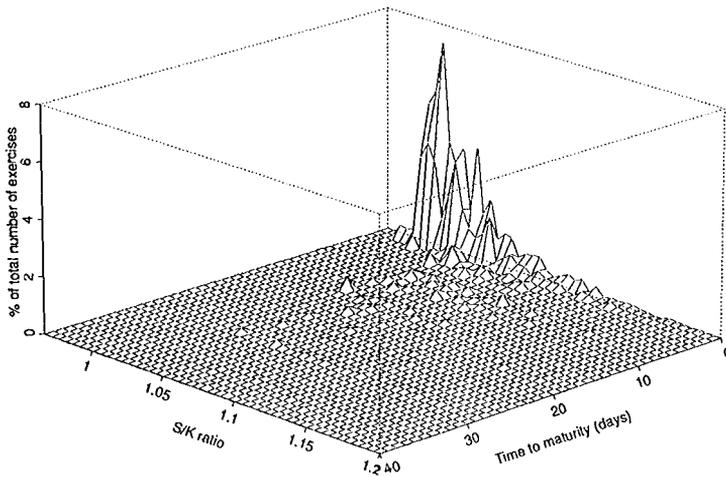
Fig. 1 . Distribution of the number of call contracts exercised, conditional on S/K and τ.

described in OEX-S&P100 Index Options (1995). To these data we added the corresponding series of observed S&P100 Index daily closure prices obtained from Standard and Poors.[9] The sample we consider runs from 3 January 1984 to 30 March 1990. Fig. 1 shows the sample distribution of the number of exercises of call contracts, conditional on the current S&P100 Index to strike price ratio (S/K), and on the time to maturity (τ).[10]

Table 1 provides summary statistics of the data. N_{call} is the number of exercises of call options, S is the S&P100 Index and δ is the dividend rate on S. The latter is derived from the S&P100 Index dividend series constructed by Harvey and Whaley (1992).[11] $\bar{X}$ denotes the sample mean of the series and $x\%$ represents the $x\%$ quantile of the empirical distributions, i.e., the observed value X_0 such that $x\%$ of the observations are less than or equal to X_0. More

---

[9] The wildcard feature of the OEX contract, described in detail in Diz and Finucane (1993) for instance, results in some nonsynchronous effects in the exercise and index data which will be ignored (at least explicitly). Any systematic nonsynchronous effect will (implicitly) be captured, however, in our nonparametric analysis.

[10] Fig. 1 presents truncated data, since we left aside observations corresponding to high values of τ. The purpose was to obtain a better visualization of what happens when the number of exercises is significant. However, the complete sample was used at the estimation stage.

[11] The series derived by Harvey and Whaley (1992) gives D_t, the amount in $ of the dividend paid on the S&P100 Index at date t. In order to be consistent with Eq. (2.5), we need *annualized* dividend *rates*. The series δ whose empirical mean and quantiles are reported in Table 1 has been computed as $\delta_t \equiv D_t/S_t/\mathrm{d}t$, where S_t is the S&P100 Index and $\mathrm{d}t = \frac{1}{360}$.

Table 1
Summary statistics of exercise data

Var.	$\bar{X}$	min	5%	25%	50%	75%	95%	max
N_{call}	2697	1	3	24	202	1357	16946	72590
S	236.9	146.5	154.495	181.8	238.5	280.9	322.838	336.1
δ	0.05254	0	0	0	0.00788	0.06046	0.26044	0.79783

formally, this quantile is defined as the value $X_0 \in \{X_i: i = 1, 2, \ldots, n_X\}$ such that $n_X^{-1} \sum_{i=1}^{n_X} \mathbb{1}_{(-\infty, X_0]}(X_i) = x/100$, where n_X is the number of observations for the variable X and $\mathbb{1}_A$ is the indicator function of the set A.

Fig. 1 shows that most of the exercises occur during the last week prior to expiration. Except for a period of one or two days to maturity, exercise decisions are taken when the ratio S/K is close to one. During this period, the ratio is never below one. However, in the last days before maturity, although most exercise decisions take place at S/K close to one, the dispersion of the observed ratio is highly increased towards values close one.[12]

The objective here is to estimate a boundary by fitting a curve through a scatterplot in the space $(\tau, S/K)$. We proceed as follows. Over the entire observation period, consider the set of observed values for the time to maturity variable $\mathcal{T} \equiv \{0, 1, \ldots, \tau_{\max}\}$. Over the same period, we observe a total of N call options indexed by $i \in \mathcal{I} \equiv \{1, 2, \ldots, N\}$. Each of these options is characterized by the date of its issue, t_0^i, the date at which it matures, $t_0^i + T^i$, and its strike price, K^i. In addition to these variables, for $\tau \in \mathcal{T}$, we observe $S_\tau^i \equiv S_{t_0^i + T^i - \tau}$ and $n_\tau^i \equiv n_{t_0^i + T^i - \tau}^i$ which are respectively the price of the S&P100 Index and the number of exercises of option i at date $t_0^i + T^i - \tau$, $i \in \mathcal{I}_\tau \equiv \{j \in \mathcal{I}: n_\tau^j \neq 0\}$.

The idea underlying the estimation procedure is that observed S_τ^i/K^i ratios result from an exercise policy and can therefore be considered as realizations of the boundary which, besides τ, is a function of the parameter vector $\theta = (r, \delta, \sigma)'$ defined in Eq. (2.5). Accordingly, $B(\theta, \tau)$ stands for the value of the optimal exercise boundary when the vector of parameters is equal to θ and the time to maturity is τ. With such an interpretation of the data, to each τ corresponds only one optimal exercise policy, and we should observe only one S_τ^i/K^i ratio. We observe several realizations of S_τ^i/K^i, however, for a single τ.[13] A natural way to

---

[12] These stylized facts do not contradict the predictions of the option pricing model when the underlying asset price is assumed to be a log-normal diffusion. As shown in Kim (1990, Proposition 2, p. 558) $\lim_{\tau \downarrow 0} B_\tau/K = \max\{r/\delta, 1\}$, for call contracts, while for puts $\lim_{\tau \downarrow 0} B_\tau/K = \min\{r/\delta, 1\}$. Here τ denotes time to maturity.

[13] The fact that we observe a dispersion in exercise decisions may be viewed as sufficient evidence to reject the parametric model in Eq. (2.5) and suggests more complex models (see e.g., Broadie et al., 2000).

summarize the information is to give more weight to S_τ^i/K^i ratios associated with high numbers of exercises n_τ^i. In other words, we consider the weighted averages

$$\left(\frac{S}{K}\right)_\tau \equiv \frac{1}{\sum_{i\in\mathscr{I}_\tau} n_\tau^i} \sum_{i\in\mathscr{I}_\tau} n_\tau^i \frac{S_\tau^i}{K^i} \tag{3.9}$$

as our realizations of $B(\theta, \tau)$.[14] A nonparametric estimator of B is a cubic spline estimator for the model

$$\left(\frac{S}{K}\right)_\tau = g(\tau) + \varepsilon. \tag{3.10}$$

For the details of this estimation procedure, see Eubank (1988, pp. 200–207, and Section 5.3.2). Intuitively, a curve is fitted to the points $(\tau, (S/K)_\tau)$, $\tau \in \mathscr{T}$. It involves a *smoothing parameter λ* which is selected by generalized cross validation (GCV).[15] This is the default procedure of the function `smooth.spline` in the *S-Plus* statistical package. The value of λ computed from observations of the S/K ratio is $\hat{\lambda} = 9.058884 \times 10^{-3}$, which gives a GCV criterion $\text{GCV}(\hat{\lambda}) = 5.911787 \times 10^{-4}$. Details of the choice of the smoothing parameter are discussed in the appendix; see also Eubank (1988, pp. 225–227) and Wahba (1990, Sections 4.4 and 4.9).

3.2. Parametric estimation of the exercise boundary

We now exploit the information provided by the dynamics of the underlying asset price and consistently estimate its trend and volatility parameters. Up to this point, we did not explicitly introduce the distinction between the process which generates the data on S, i.e., the probability distribution P from which the observations are 'drawn', referred to as the 'objective' probability, and the risk-neutral representation of the process described by (2.5). The data generating process (DGP) which is to be estimated is

$$dS_t = S_t[\mu\,dt + \sigma\,dW_t], \quad t \geq 0, \tag{3.11}$$

where W is a standard Brownian motion on $(\Omega, \mathscr{F}, \mathscr{F}_{(\cdot)}, P)$.

Quite a few well-known procedures exist for estimating the parameters of a general diffusion. Most of them are based on simulations of the DGP; examples are the simulated method of moments [see Duffie and Singleton, 1993], the simulated (pseudo) maximum likelihood [see Gouriéroux and

---

[14] Hastie and Tibshirani (1990, p. 74) give a justification to the intuitive solution of averaging the response variable when we observe ties in the predictor.

[15] For a definition of the GCV cirterion, see the appendix, especially Eq. (4).

Monfort, 1995] and indirect inference or moment matching [see Gouriéroux et al., 1993; Gallant and Tauchen, 1996]. Recently, another approach has been proposed by Pedersen (1995a,b) based on a convergent approximation to the likelihood function. In the case of a simple geometric Brownian motion, however, we take advantage of the existence of an exact discretization. Application of Itô's lemma to (3.11) gives

$$\ln S_t = \ln S_0 + \int_0^t \left(\mu - \frac{1}{2}\sigma^2 \right) ds + \int_0^t \sigma \, dW_s, \quad t \geq 0.$$

Therefore the process $\ln S$ has an AR(1) representation:

$$\ln S_t = \ln S_{t-1} + (\mu - \tfrac{1}{2}\sigma^2) + \sigma\varepsilon_t, \quad t \geq 1, \tag{3.12}$$

where $\{\varepsilon_t \equiv W_t - W_{t-1}\} \overset{\text{ind}}{\sim} N(0, 1)$. The vector $\beta \equiv (\mu, \sigma)$ can be estimated by maximum likelihood (ML). The ML estimator (MLE) of β is the solution of

$$\min_{\beta \in \mathcal{B}} \frac{T}{2} \ln \sigma^2 + \frac{1}{2\sigma^2} \sum_{t=2}^{T} \left(\ln \frac{S_t}{S_{t-1}} - \mu + \frac{\sigma^2}{2} \right)^2$$

and denoted by $\hat{\beta}_T$. Here T is the sample size and $\mathcal{B}$ is the set of admissible values for β.

However, what is required for the implementation of the BD algorithm are values for r, δ and σ. Obviously $\hat{\sigma}_T$, the MLE of σ, will be selected as the required value for the volatility parameter. But the estimation of (3.11) does not provide us with values for the risk-free interest rate and the dividend rate. These parameters are extracted from historical series. An estimate of the nominal interest rate is the average daily rate of return on 1 month T-bills, expressed in % *per annum*, and the constant dividend rate on the S&P100 Index is the sample average of the dividend series described in Section 3.1 (see also Table 1). A parametric estimate of the exercise boundary is then derived by implementing the BD algorithm with $\hat{\theta}_T \equiv (\hat{r}_T, \hat{\delta}_T, \hat{\sigma}_T)'$ as the true parameter value, where $\hat{r}_T = 0.05915$, $\hat{\delta}_T = 0.05254$ and $\hat{\sigma}_T = 0.01244$. Note that $\hat{\sigma}_T$ is obtained from daily returns on the S&P100 Index. An estimate of the annual volatility is then $\sqrt{250}\hat{\sigma}_T = 0.1967$.

The parametric and nonparametric estimates of the exercise boundary are shown in Fig. 2. For the moment we will focus exclusively on results which are based on the full sample of data. Later we will also examine subsamples consisting of data before and after the October 1987 crash.

As can be seen, the two estimated exercise boundaries appear quite different in shape. First, the parametric estimate of the boundary lies well above the nonparametric one. In particular, we note that the exercise boundary at the maturity date, which equals $\hat{r}_T/\hat{\delta}_T = 1.12325$, is considerably above the
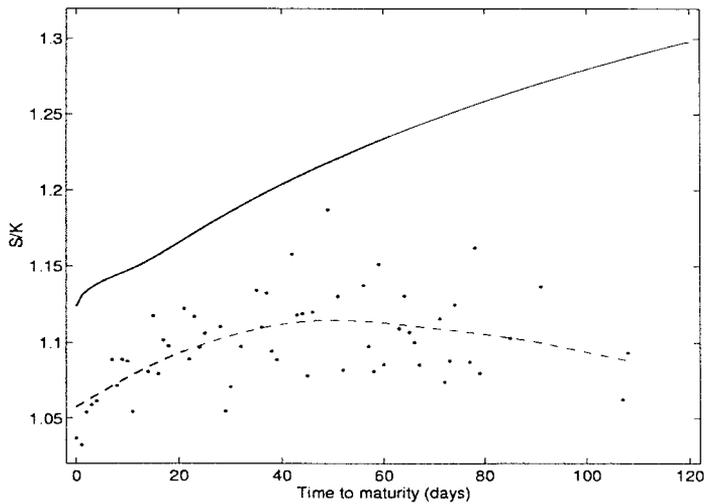
Fig. 2. Parametric (——) and nonparametric (- -) estimates of the exercise boundary.

nonparametric estimate at $\tau = 0$. Second, it is interesting to note that the parametric estimate of the exercise boundary lies *above* the pairs $(\tau, S/K)$ obtained from the exercise data. It is clear that the two estimates predict very different exercise strategies in the few days before expiration, where most of the exercise decisions take place (see Fig. 1). Ideally, we would like to make a formal statistical comparison between the two curves appearing in Fig. 2. Unfortunately, there are several reasons, explained in the next section, why such a comparison is not straightforward.

3.3. Nonparametric and parametric boundaries

So far, we engaged only in casual comparison of the two estimated exercise boundaries drawn in Fig. 2. On the parametric side, there is uncertainty about the position of the curve because the parameters fed into the BD algorithm are estimates of unknown parameters. Likewise, there is uncertainty regarding the position of the nonparametric curve as well. Indeed, the S/K ratios obtained via (3.10) may not directly reflect the exercise boundary because: (1) there is, in fact, a dispersion of exercise decisions which was summarized by a single ratio per time to maturity (see Fig. 1) and (2) the index S in the ratio is the index at the closure which may not exactly coincide with the value of the index when the exercise decision was actually made. Even if we ignore these effects, it is clear that the kernel estimation is also subject to sampling error which we can characterize at least asymptotically.

There are at least two ways to tackle the comparison between the parametric and nonparametric boundaries.[16] Given what we know about the statistical properties of the nonparametric boundary, we could entertain the possibility of formulating a confidence region which, if it does not contain the entire parametric boundary, suggests rejecting the model. Such (uniform) confidence regions were discussed in Härdle (1990), Horowitz (1993) and Aït-Sahalia (1993, 1996). The former two deal exclusively with i.i.d. data, while the latter considers temporally dependent data. Only the latter would be appropriate since the exercise data described in Section 3.1 are not i.i.d. There are essentially two approaches to compute confidence regions with temporally depend data: (1) using asymptotic distribution theory combined with the so-called delta method applied to distribution functions of the data (see Aït-Sahalia, 1993) or (2) applying bootstrap techniques. The former can be implemented provided that the derivatives of the distribution functions are not too complicated to compute. Since this is typically not the case it is more common to rely on bootstrap techniques. Since the data are temporally dependent one applies bootstrapping by blocks (see for instance Künsch, 1989). Unfortunately, our data are not straightforwardly interpretable as time series since the exercise boundary is obtained from observations at fixed time to maturity. This scheme does not amount to a simple sequential temporal sampling procedure. Moreover at each point in time one records exercise decisions on different contracts simultaneously which have very different coordinates in the time-to-maturity and boundary two-dimensional plane. The conditions on the temporal dependence in calendar time (such as the usual mixing conditions) do not easily translate into dependence conditions in the relevant plane where the empirical nonparametric boundary is defined (see the appendix for more details). Because of these unresolved complications, we opted for another strategy similar to the one just described, but concentrated instead on the parametric specification. Under the assumption that the parametric model is correctly specified, we can use the asymptotic distribution, namely

$$\sqrt{T}(\hat{\theta}_T - \theta) \overset{A}{\sim} N(0, \Omega) \Rightarrow \sqrt{T}(B(\hat{\theta}_T, \tau) - B(\theta, \tau))$$

$$\overset{A}{\sim} N[0, (\partial B / \partial \theta') \Omega (\partial B / \partial \theta)],$$

---

[16] One possibility which we do not consider is to calculate implied volatilities by inverting the BD boundary which best fits the exercise data and test whether this volatility is compatible with the estimates of the underlying process. The reason why we do not pursue this is that it would be difficult to compute standard errors for the implied volatilities. Another strategy one could pursue is to use a Jackknife approach, though one would somehow have to deal with the dependence in the data.

(which holds under standard regularity assumptions, e.g., see Lehmann (1983, Theorem 1.9, p. 344).[17] The estimate of θ is denoted $\hat{\theta}_T$ and $B(\theta, \tau)$ stands for the value of the optimal exercise boundary when the vector of parameters is equal to θ and the time to maturity is τ. However, in our situation the vector $\hat{\theta}_T$ is obtained by stacking estimates of its components, namely $\hat{r}_T, \hat{\delta}_T$ and $\hat{\sigma}_T$, which were computed from separate series, with unknown joint distribution. Hence, the asymptotic normality of $\hat{\theta}$ may be questionable, while the covariance matrix Ω would remain unknown.

Clearly, we need to make some compromises to be able to assess the effect of parameter uncertainty on the boundary. We should note first and foremost that $\hat{r}_T$ and $\hat{\delta}_T$ play a role different from $\hat{\sigma}_T$. The former two are estimates which determine the drift under the risk neutral measure. They are sample averages of observed series and computed from a relatively large number of observations. In contrast, $\hat{\sigma}_T$ is estimated from a GBM specification. It is typically more difficult to estimate, yet at the same time plays a much more important and key role in the pricing (and exercising) of options. Indeed, $\hat{r}_T$ and $\hat{\delta}_T$ primarily determine the intercept (see footnote 12), while $\hat{\sigma}_T$ affects essentially the curvature of the exercise boundary. For these reasons, we will ignore for the moment uncertainty regarding $\hat{r}_T$ and $\hat{\delta}_T$, and focus exclusively on the role played by $\hat{\sigma}_T$ on the location of $B(\hat{\theta}_T, \tau)$. The confidence bounds appearing in Fig. 3 were obtained through a Monte Carlo simulation of the GBM volatility parameter empirical distribution and its impact on that of $B(\hat{\theta}_T, \tau)$.

For the reasons explained above, the simulations are performed considering $\hat{r}_T$ and $\hat{\delta}_T$ as fixed. The parametric estimator of the exercise boundary at maturity τ is denoted by $B(\hat{\sigma}_{\text{MLE}}, \tau)$ since the volatility coefficient is obtained by maximum likelihood estimation. We simulate $R = 10{,}000$ samples $((S_t^\gamma, t = 1, \ldots, T), \gamma = 1, \ldots, R)$ of the S&P100 Index using $\hat{\beta}_T$ in (3.12). We then estimate β by $\hat{\beta}_T^\gamma \equiv (\hat{\mu}_T^\gamma, \hat{\sigma}_T^\gamma)'$, its MLE computed from the γth sample, and derive $B^\gamma(\hat{\sigma}_{\text{MLE}}, \tau)$ using the BD algorithm with $\theta = \hat{\theta}_T^\gamma \equiv (\hat{r}_T, \hat{\delta}_T, \hat{\sigma}_T^\gamma)'$, $\tau \in \mathcal{T}$. For $\hat{r}_T$ and $\hat{\delta}_T$ fixed and R large, the sample variance

$$\hat{V}^R(\hat{r}_T, \hat{\delta}_T, \hat{\beta}_T, \tau) = \frac{1}{R} \sum_{\gamma=1}^R \left[B^\gamma(\hat{\sigma}_{\text{MLE}}, \tau) - \frac{1}{R} \sum_{\zeta=1}^R B^\zeta(\hat{\sigma}_{\text{MLE}}, \tau) \right]^2, \quad \tau \in \mathcal{T}$$

is close to $V_{\hat{\sigma}_T}[B(\hat{\sigma}_{\text{MLE}}, \tau)]$, the variance of $B(\hat{\sigma}_{\text{MLE}}, \tau)$ when $\hat{r}_T$ and $\hat{\delta}_T$ are fixed and $\hat{\sigma}_T$ is assumed to be the true value of the volatility coefficient. When T is large, this can be expected to be a good approximation of $V_{\sigma_0}[B(\hat{\sigma}_{\text{MLE}}, \tau)]$, where σ_0 denotes the true value of σ.

---

[17] Please note that the asymptotic distribution could be derived under the assumption of model misspecification as well, resulting in a QMLE interpretation. Under suitable regularity conditions this would amount to modifying the covariance matrix of the asymptotic distribution.
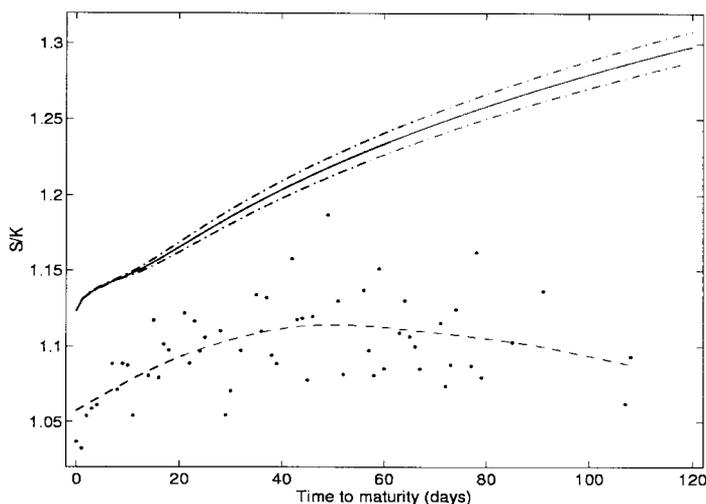
Fig. 3. Parametric (——) and nonparametric (- -) estimates of the exercise boundary with 95% confidence bounds on the parametric boundary (- · -).

If we further assume that for each $\tau \in \mathcal{T}$ $B(\hat{\sigma}_{\mathrm{MLE}}, \tau)$ is approximately normally distributed (recall that $\hat{\sigma}$ is a MLE), we can build a confidence interval for $B(\theta, \tau)$ at level $1 - \alpha$, whose limits are given by $B(\hat{\theta}_T, \tau) \pm c_\alpha \hat{V}^R(\hat{r}_T, \hat{\delta}_T, \hat{\beta}_T, \tau)^{1/2}$, $\tau \in \mathcal{T}$, where c_α satisfies $\Phi(c_\alpha) = 1 - \alpha/2$, Φ being the cumulative distribution function (cdf) of $N(0, 1)$.

The confidence bands obtained in this way show clearly that, provided that $\hat{r}_T$ and $\hat{\delta}_T$ are not too far from their true values and that the normal approximation is good enough, the two boundaries are significantly different from each other. Indeed, the nonparametric curve *and* the data points appearing in Fig. 3 lie outside the parametric curve confidence region. Before we turn to call price estimation, it is worth noting that the uncertainty on the volatility parameter is of less importance for the exercise policy when the contract approaches its maturity. This is expected since the volatility of the underlying asset becomes less important in the decision of exercising the call contract, or in other words $\partial B(\theta, \tau)/\partial \sigma \approx 0$, for $\tau \approx 0$, and for any θ.

There are several issues that emerge from the results we obtained so far. Clearly, the differences between exercise boundaries drawn in Fig. 2 cannot be attributed to the uncertainty in the estimation of σ. Hence, on the parametric side, there is mostly uncertainty about $\hat{r}_T$ and $\hat{\delta}_T$ which can be the source of discrepancy between the parametric and nonparametric curves. These parameters, as noted before, determine the risk-neutral density. One may wonder therefore whether the main reason for the difference between the two approaches is due to a misspecification of the transformation to a risk-neutral
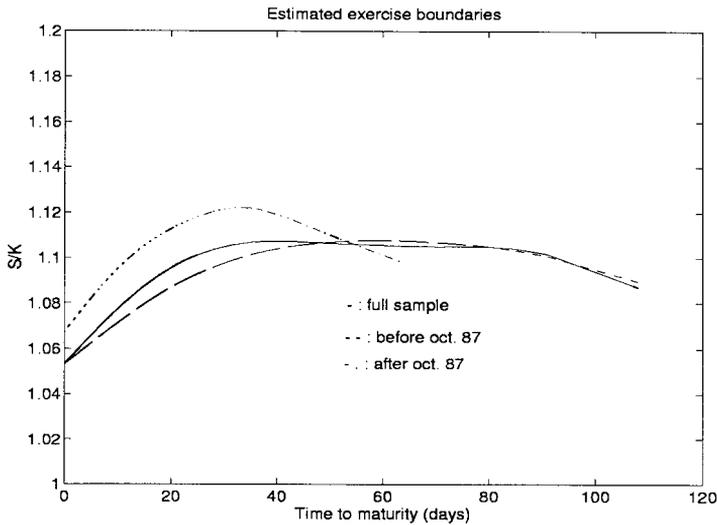
Fig. 4 . Nonparametric estimates of the exercise boundary before and after the October 1987 crash.

representation? Before we come to such a conclusion let us revisit the non-parametric estimates and appraise their robustness. Roughly in the middle of our sample is the October 1987 crash. This event is without any doubt impor-tant and often characterized as a breakpoint in some of the stylized facts regarding derivative securities (see, among others, Bates, 1991). In Fig. 4 we display nonparametric estimates of the exercise boundaries before and after the crash superimposed on the nonparametric curve obtained from the whole sample. Before discussing the figure we should first and foremost express some reservations about this comparison. The nonparametric methods require large data sets. The full sample considered in the previous section is in fact relatively small compared with many applications of nonparametric methods. Moreover, the fact that we apply these methods in a time series context adds even more strains on data requirements. It is therefore not a trivial exercise to split up the sample and apply nonparametric methods to the subsamples. With these reservations in mind we observe that in Fig. 4 the exercise boundaries look quite similar before and after the crash, particularly keeping in mind the scale of the plot compared to the scale in Fig. 2.[18]

When we bring all the evidence together presented thus far, we must conclude that the parametric model seems to yield very biased estimates of the call

---

[18] Much has been written about the changing behavior of puts since the crash of 1987. The results in Fig. 4 suggest that the exercise behavior of calls has not been significantly affected by the crash.

exercise boundary, primarily because the transformation to the risk neutral density appears to be misspecified.[19]

4. Parametric and nonparametric analysis of call prices

We now turn to the estimation of call option prices. As in Sections 3.1 and 3.2, we consider two types of estimators: (1) a nonparametric estimator entirely based on the data and (2) a model-based (or parametric) estimator. Along the same lines, we first describe the data and then present the estimation results.

4.1. The data

We now use data sets (1) and (2) mentioned in the introduction. The period of observation and the data on the S&P100 Stock Index are the same as for the boundary estimation (see Section 3.1). For the same period, we observed the characteristics (price, strike price and time to maturity) of the call option contracts on the S&P100 Index, described in OEX (1995). They represent daily closing prices obtained from the CBOE.

4.2. Estimation of call option prices

Since the call option price C depends on the underlying stock price S, we may have some problems in estimating C, due to the possible nonstationarity of S. To avoid this, we use the homogeneity of degree one of the pricing formula with respect to the pair (S, K) [see Eqs. (2.6)–(2.8)] and focus on the ratio $C(S, K, \tau)/K = C(S/K, 1, \tau)$, which expresses the normalized call option price as a function of the moneyness and time to maturity.[20] Fig. 5 shows the pairs $((C/K), (S/K))$ observed at different times to maturity, $\tau = 7, 28, 56, 84$ days.

Again, we consider two types of estimators depending on our assumptions about the underlying economic model. The first estimator is entirely based on the Black–Scholes specification of the economy introduced in Section 2. It is derived in two steps. First, we estimate the parameters of (3.12) by maximum

---

[19] Similar findings, though in a very different context, namely that of Heston's stochastic volatility model, were also reported in Chernov and Ghysels (1998). The techniques used in the latter are very different from those in the current paper. Instead of comparing nonparametric and parametric estimates these authors estimate the risk neutral and objective measure representation simultaneously using bivariate models of returns and options contracts. These joint estimates allow the authors to test restrictions on the Radon–Nikodym derivative between the two measures.

[20] The homogeneity property holds for the GBM as well as for a large class of other processes featuring stochastic volatility. See Samuelson (1965) and Merton (1973). See also Broadie et al. (1998) and Garcia and Renault (1996) for further discussion.
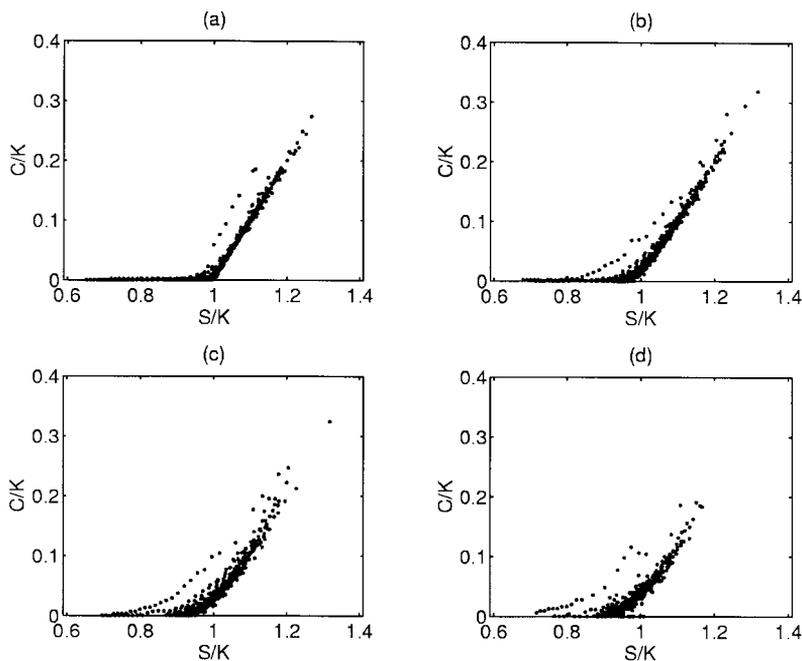
Fig. 5. Observed couples $((C/K)_t, (S/K)_t)$ at different times to maturity. $\tau = 7$ days: (a), $\tau = 28$ days: (b), $\tau = 56$ days: (c), $\tau = 84$ days: (d).

likelihood (see Section 3.2), and second, we use these estimates in the BD routine to compute $C(S/K, 1, \tau)$. We implemented the BD algorithm with $\theta = \hat{\theta}_T$, with S/K running from 0.6 to 1.4 and τ from 0 to 120 days. These values match the range of the observed values of S/K and τ. The resulting surface is shown in Fig. 6.

Similarly, we derive a second estimator of the same surface. This estimator requires no particular assumptions about the economy. We simply express the normalized call price as a function of time to maturity and of the moneyness ratio S/K:

$$C/K = C(S/K, 1, \tau) + \varepsilon = V(S/K, \tau) + \varepsilon, \tag{4.13}$$

where ε is an error term. The unknown function V is estimated by fitting a surface through the observations $((C/K)_t, (S/K)_t, \tau_t)$ using kernel smoothing.[21] The surface appears in Fig. 7.

---

[21] For a discussion on multivariate nonparametric estimation, we refer to Hastie and Tibshirani (1990) and Scott (1992). We used here a product of Gaussian kernels with bandwidths $h_\tau = 5$ in the time to maturity direction and $h_{SK} = 0.4$ in the S/K direction.
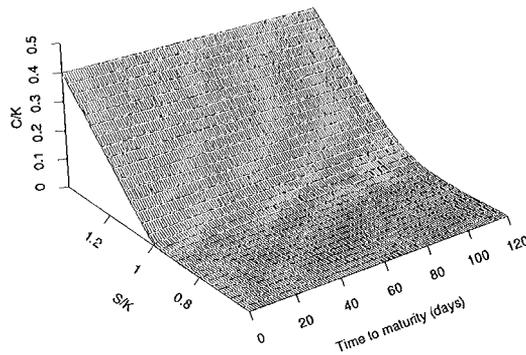
Fig. 6. Parametric estimate of call option price surfaces.
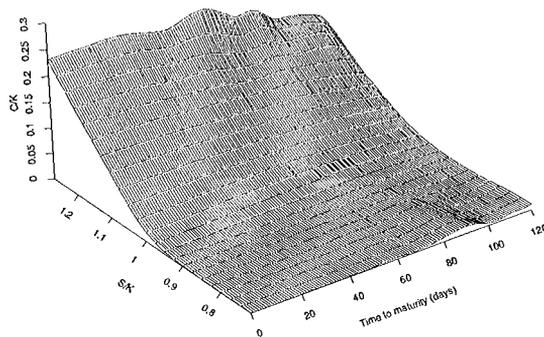
Fig. 7. Nonparametric estimate of call option price surfaces.

The parametric and nonparametric estimates of the relationship between C/K and $(S/K, \tau)$ are very similar in shape, and it is not easy to appraise the differences that may exist between the two estimates by a direct comparison of Figs. 6 and 7. Instead, we could select different times to maturity ($\tau = 7$, 28, 56 and 84 days) and extract the relationship between C/K and S/K from the estimated surfaces in Figs. 6 and 7 for these given τ. It is more appropriate however to re-estimate the relation between C/K and S/K only, for $\tau \in \{7, 28, 56, 84\}$. Obviously, this will produce no change in the parametric estimate. However, for nonparametric estimation, we avoid some difficulties inherent in multivariate kernel estimation [see Silverman, 1986; Hastie and Tibshirani, 1990; Scott, 1992]. We used a smoothing spline where the smoothing parameter is chosen according to the GCV criterion.[22] The resulting difference $e(S/K, \tau) \equiv \hat{C}_{NP}(S/K, 1, \tau) - \hat{C}_P(S/K, 1, \tau)$, ($\tau = 7$, 28, 56, 84 days), between the nonparametric and parametric fits of the call price is shown in Fig. 8.

---

[22] For a definition of the GCV cirterion, see the appendix, especially Eq. (4).
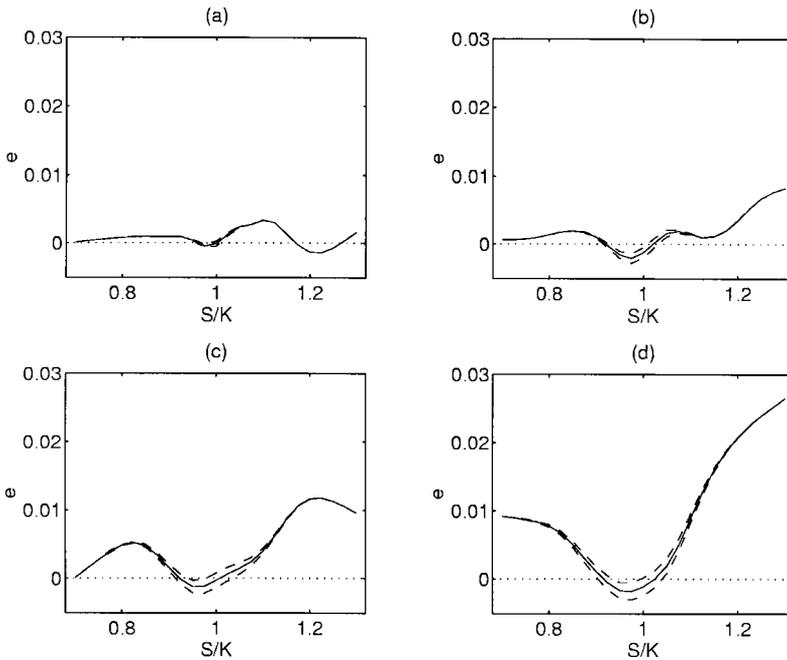
Fig. 8. Difference between the nonparametric and parametric fits of call prices (solid line ——) and 95% confidence bounds (dash lines - - -). $\tau = 7$ days: (a); $\tau = 28$ days: (b); $\tau = 56$ days: (c); $\tau = 84$ days: (d).

Several remarks emerge from these figures. We note from Fig. 7 that the nonparametric estimate captures the dependence of option prices on time to maturity, i.e., as we move away from the maturity date, the normalized call option price increases as time to maturity gets larger, for any fixed moneyness ratio S/K. However, this dependence dampens out as this ratio moves away from unity. The largest differences occur for 'near-the-money' or at-the-money contracts.

The plots of the differences $e(S/K, \tau)$ [see Figs. 8(a)–(d)] reveal some interesting features. First, we see that the parametric estimates tend to underprice the call option contract, when the nonparametric estimated relation is taken to be the true one. Although this holds for all the times to maturity we considered ($\tau = 7, 28, 56, 84$), it is remarkable that the discrepancy between the two price predictors diminishes as we approach to maturity. One possible explanation for this is the following. Suppose the observed underpricing of the parametric estimator can be attributed to a misspecification in the dynamics of the underlying asset price process, S. Then we see that the effects of this misspecification on option pricing disappears as $\tau \downarrow 0$. Indeed, as the option approaches its maturity, the degree of uncertainty on its normalized price C/K vanishes and C/K tends to

be more and more directly related to the observed difference between S/K and 1 (when $\tau = 0$, $C/K = (S/K - 1)^+$). This is always true in option pricing models, irrespective of the specification of the dynamics of S. In particular, this is true for the GBM specification we adpoted here. A second remark about the estimation results is that, for a fixed time to maturity, the two estimates of C/K seem to agree for S/K close to 1. This is in accordance with the stylized facts compiled in the literature on option pricing [see Ghysels et al., 1996; Renault, 1996]. The usual practice of evaluating 'near-the-money' options not far from maturity according to models as simple as those of Section 2 seems to be well founded in the light of the results reported in this section.

To assess statistically the significance of e, we derived some confidence bounds as in Section 3.3, which measure the effect of the uncertainty in the estimation of σ on the parametrically estimated call prices. Considering $\hat{C}_{NP}(S/K, 1, \tau)$ as fixed and equal to the true pricing formula, and for a fixed τ, we say that $e(S/K, \tau)$ is not significantly different from 0 for a given moneyness ratio S/K, if 0 lies in the confidence interval for $e(S/K, \tau)$. These intervals, derived in a similar way as in Section 3.3, are reported in Fig. 8. The results confirm the previous remark that the two estimates of call prices agree only for S/K close to 1. Finally, we should also note that we computed the nonparametric estimates for the calls before and after the crash of 1987, similar to the comparison performed with the excercise boundaries. We do not report the results in a figure as the estimated curves are literally on top of each other. Hence, under the reservations about the use of nonparametric estimates in small samples, we find again stable results, which is compatible with the earlier results regarding excercise boundaries.

5. Conclusion

In this paper we proposed nonparametric estimation procedures to deal with the computational complications typically encountered in American option contracts. We focused on the most active market in terms of trading volume and open interest. It provided us with a wealth of data on the exercise and pricing decisions under different circumstances, i.e. differences in time-to-maturity and strike prices, and enabled us to estimate the functions nonparametrically. In principle our methods apply to any type of contract, as complex as it may be, provided the data are available and the suitable regularity conditions to apply nonparametric methods are applicable. We also reported a comparison of the nonparametric estimates with the nowadays standard parametric model involving a GBM for the underlying asset. While the comparison of the nonparametric and parametric estimated functions raised several unresolved issues, our results suggest large discrepancies between the two. It obviously raises questions about the parametric models. Of course, typically, practitioners will 'calibrate' their

parameters to the market data instead of estimating the unknown parameters via statistical techniques as we did. This may improve the fit, yet it remains limited to the constant volatility framework modified through time-varying (implied) volatilities. The advantages of the framework we propose is that it can be extended to deal with state variables such as random dividends, etc. (see Broadie et al., 2000). One remaining drawback of the approach we suggested in this paper is that it does not lend itself easily to imposing absence of arbitrage conditions. However, the advantages in terms of computing exercise boundaries and call pricing overshadow, at least in the single asset case, this disadvantage.

Appendix on nonparametric estimation

In this appendix, we briefly present the nonparametric estimation techniques used in the paper. We also provide references with more details on the subject.

In this paper, we mainly used two kinds of nonparametric estimators, namely kernel and spline smoothing. Since the issues related to these estimation techniques are similar, we present the kernel estimator first and then digress on the smoothing spline estimator.

Nonparametric estimation deals with the estimation of relations such as

$$Y_i = f(Z_i) + u_i, \quad i = 1, \ldots, n, \tag{A.1}$$

where, in the simplest case, $((Y_i, Z_i), i = 1, \ldots, n)$ is a family of i.i.d. pairs of random variables, and $E(u|Z) = 0$, so that $f(z) = E(Y|Z = z)$. The error terms $u_i, i = 1, \ldots, n$, are also assumed to be independent, while f is a function with smoothness properties which have to be estimated from the data on the pair (Y, Z). Kernel smoothers produce an estimate of f at $Z = z$ by giving more weight to observations (Y_i, Z_i) with Z_i 'close' to z. More precisely, the technique introduces a *kernel function*, K, which acts as a weighing 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 of f in (A.1) is the Nadaraya–Watson estimator defined by

$$\hat{f}_\lambda(z) = \frac{\sum_{i=1}^n K((Z_i - z)/\lambda) Y_i}{\sum_{i=1}^n K((Z_i - z)/\lambda)}, \tag{A.2}$$

so that $(\hat{f}_\lambda(Z_1), \ldots, \hat{f}_\lambda(Z_n))' = W_n^K(\lambda) Y$, 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 $K((Z_j - Z_i)/\lambda)/\sum_{k=1}^n K((Z_k - Z_i/\lambda)$. 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 $\hat{f}_\lambda(z)$ is changing with z. In other words, the degree of smoothness of $\hat{f}_\lambda$ 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 (MSE$(\lambda, z) = E[(\hat{f}_\lambda(z) - f(z))^2]$), which has a bias/variance decomposition similar to parametric estimation. Of course both K and λ have an effect on MSE(λ, z), but it is generally agreed in the literature that the most important issue is the choice of the smoothing parameter.[23] 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 MSE(λ, z). We may simply consider MSE(λ, 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 MISE$(\lambda) = E[\int(\hat{f}_\lambda(z) - f(z))^2 \, dz]$, or the *sup mean squared error* equal to $\sup_z$ MSE(λ, z), etc. The most frequently used measure of deviation is the sample mean squared error $M_n(\lambda) = (1/n)\sum_{i=1}^n [\hat{f}_\lambda(Z_i) - f(Z_i)]^2 \omega(Z_i)$, where $\omega(\cdot)$ is some known weighing function. This criterion only considers the distances between the fit and the actual function f at the sample points Z_i. Obviously, choosing $\lambda = \tilde{\lambda}_n \equiv \operatorname{argmin}_\lambda M_n(\lambda)$ is impossible to implement since f is unknown. The strategy consists of finding some function $m_n(\cdot)$ of λ (and of $((Y_i, Z_i), i = 1, \ldots, n)$) whose argmin is denoted $\hat{\lambda}_n$, such that $|\tilde{\lambda}_n - \hat{\lambda}_n| \to 0$ a.s. as $n \to \infty$. For a review of such functions m_n, see Härdle and Linton (1994, Section 4.2).[24] The most widely used m_n function is the *cross-validation* function

$$m_n(\lambda) = \mathrm{CV}_n(\lambda) \equiv \frac{1}{n}\sum_{i=1}^n [Y_i - \hat{f}_\lambda^{(-i)}(Z_i)]^2, \tag{A.3}$$

where $\hat{f}_\lambda^{(-i)}(z)$ is a Nadaraya–Watson estimate of $f(z)$ obtained according to (A.2) but with the ith observation left aside. Craven and Wahba (1979) proposed the *generalized cross-validation* function with

$$m_n(\lambda) = \mathrm{GCV}_n(\lambda) \equiv \frac{n^{-1}\sum_{i=1}^n [Y_i - \hat{f}_\lambda(Z_i)]^2}{[1 - n^{-1}\mathrm{tr}(W_n(\lambda))]^2}, \tag{A.4}$$

where W_n is the influence matrix.[25]

---

[23] 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).

[24] See also Silverman (1986, Section 3.4) and Andrews (1991).

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

Another important issue is the convergence of the estimator $\hat{f}_{\hat{\lambda}_n}(z)$. Concerning the Nadaraya–Watson estimate (A.2), Schuster (1972) proved that under some regularity conditions, $\hat{f}_{\hat{\lambda}_n}(z)$ is a consistent estimator of $f(z)$ and is asymptotically normally distributed.[26] Therefore when the argmin $\hat{\lambda}_n$ of $m_n(\lambda)$ is found in the set Λ_n (see footnote 26), we obtain a consistent and asymptotically normal kernel estimator $\hat{f}_{\hat{\lambda}_n}(z)$ of $f(z)$, which is optimal in the class of the consistent and asymptotically Gaussian kernel estimators for the criterion $M_n(\lambda)$.[27]

While kernel estimators of regression functions (or conditional expectation functions) are based on kernel estimates of density functions (see for instance Härdle and Linton, 1994, Section 3.1), spline estimators are derived from a least-squares approach to the problem. One could think of solving the following problem:

$$\min_{g \in \mathcal{M}} \sum_{i=1}^{n} [Y_i - g(Z_i)]^2, \tag{A.5}$$

where $\mathcal{M}$ is a class of functions satisfying a number of desirable properties (e.g., continuity, smoothness, etc.). Obviously, any $\tilde{g} \in \mathcal{M}$ restricted to satisfy $\tilde{g}(Z_i) = Y_i, i = 1, \ldots, n$, is a candidate solution of the minimization problem, and would merely consists of interpolating the data. Even if we restrict g to have a certain degree of smoothness (by imposing continuity conditions on its derivatives), functions g such that $g(Z_i) = Y_i, i = 1, \ldots, n$, may be too wiggly to be a good approximation of f. To avoid this, the solution of the problem is chosen so that functions not smooth enough are 'penalized'. A criterion to obtain such solutions is

$$\min_{g \in \mathcal{M}} \sum_{i=1}^{n} [Y_i - g(Z_i)]^2 + \lambda \int_I [g^{(2)}(x)]^2 \, dx. \tag{A.6}$$

I is an interval $[a, b]$ such that $a < \min\{Z_i : i = 1, \ldots, n\} \leq \max\{Z_i : i = 1, \ldots, n\} < b$, and $g^{(k)}$ denotes the kth derivative of g. The integral in the second term of (A.6) is a measure of the degree of smoothness of the function g since it can be interpreted as the total variation of the slope of g. Then for high λ, we penalize functions which are too wiggly and we move away from solutions that tend to interpolate the data. If λ becomes too high, we decrease the goodness of the fit. In the limit, if $\lambda \to \infty$, the problem tends to minimizing the second term of (A.6), whose solution is a function that is 'infinitely smooth'.

---

[26] The regularity conditions bear on the smoothness and continuity of f, 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.

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

Such a function is a straight line which has a zero second derivative everywhere. Conversely, if $\lambda \to 0$, the solution of (A.6) tends to the solution of (A.5) which is the interpolant. Therefore, the parameter λ plays exactly the same role as in kernel estimation.

When $\mathscr{M}$ is taken as the class of continuously differentiable functions on I, with square integrable second derivative on I, the solution of (A.6) is unique and is a natural cubic spline, which we denote by $\hat{f}_\lambda$ [see Wahba, 1990, pp. 13–14; Eubank, 1988, pp. 200–207]. By natural cubic spline, it is meant that, given the mesh on I defined by the order statistic $Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(n)}$, $\hat{f}_\lambda$ is a polynomial of order three on $[Z_{(i)}, Z_{(i+1)}]$, $i = 1, \ldots, n-1$, with second derivatives continuous everywhere, and such that $\hat{f}_\lambda^{(2)}(Z_{(1)}) = \hat{f}_\lambda^{(2)}(Z_{(n)}) = 0$. It can be shown [see Härdle, 1990, pp. 58–59] that the spline $\hat{f}_\lambda$ is a linear transformation of the vector of observations Y, i.e.,

$$\hat{f}_\lambda(z) = \sum_{i=1}^{n} w_i^\lambda(z) Y_i. \tag{A.7}$$

A result of Silverman (1984) proves that the weight functions w_i^λ behave asymptotically like kernels. If we write (A.7) for observations points $Z_1, \ldots, Z_n$, we have $(\hat{f}_\lambda(Z_1), \ldots \hat{f}_\lambda(Z_n))' = W_n^S(\lambda) Y$ where the influence matrix $W_n^S(\lambda)$, has its (i, j)th entry equal to $w_j^\lambda(Z_i)$ [see Wahba, 1990, p. 13]. This matrix is explicitly derived in Eubank (1988, Section 5.3.2) and is shown to be symmetric, positive definite.

It appears that, like kernel estimators, spline function estimators are linear estimators involving a smoothing parameter and are asymptotically kernel estimators. Therefore, the criteria for selecting λ described above also apply for spline estimation (see Wahba, 1990, Sections 4.4 and 4.9) and Eubank (1988, pp. 225–227)).

Things are a little bit more complicated when the errors are not spherical. Under general conditions, the kernel and spline estimators remain convergent and asymptotically normal. Only the asymptotic variance is affected by the correlation of the error terms. However, 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 in the literature what should be done in this case to avoid over- or undersmoothing.[28] Two kind of solutions have been proposed. The first one consists of modifying the selection criterion (CV_n or GCV_n) in order to derive a consistent estimate of M_n, and the second one tries to orthogonalize the error term and apply the usual selection rules for λ. When the autocorrela-

---

[28] 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 of λ that are too large (small), yielding oversmoothing (undersmoothing).

tion function of u is unknown, one has to make the transformation from sample estimates obtained from a first step smoothing. 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 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 observed pairs of (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 Györfi et al. (1989). This reference (Chapter 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 $\hat{\lambda}_n = \mathrm{argmin}_{\lambda \in \Lambda_n} \mathrm{CV}_n(\lambda)$, then under regularity conditions $\hat{\lambda}_n$ is an optimal choice for λ according to the *integrated squared error*, $\mathrm{ISE}(\lambda) = \int [\hat{f}_\lambda(z) - f(z)]^2 \, dz$ (see Györfi et al., 1989, Corollary 6.3.1). Although the function $\mathrm{CV}_n(\lambda)$ 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), there is no general result for criteria such as $\mathrm{MISE}(\lambda)$ or $M_n(\lambda)$.

The most general results concerning the convergence of nonparametric kernel estimators of regression functions seem to be found in Aït-Sahalia (1993). In this work, very general regularity conditions which ensure the convergence and the asymptotic normality of functional estimators, whose argument is the cdf which has generated the observed sample, are given (Aït-Sahalia, 1993, Theorem 3, pp. 33–34). This result is derived from a functional CLT for kernel estimators of cdfs (Aït-Sahalia, 1993, Theorem 1, p. 23) combined with a generalization of the delta method to nonparametric estimators. Therefore, provided that the asymptotic variances can be approximated, one can apply usual Wald-type tests or confidence regions to make proper statistical inference. When the asymptotic distribution is too complex, a block bootstrapping technique, specially adapted to resampling from dependent data, can be used (see Künsch, 1989; Liu and Singh, 1992; Aït-Sahalia, 1993). Although this method is very general, a mixing condition is required when dealing with dependent data. Even though this condition allows for many types of serial dependence, application of these results in the context of Sections 3.1 and 4.2 is not straightforward. Indeed, in the case of nonparametric exercise boundary estimation as well as in call price estimation, the data points from which we derive our estimates are not sampled *via* a simple chronological scheme. In the case of exercise boundary estimation, the data points we use are weighted averages of observations of ordinary time series. In the case of call price estimation, the difficulty comes from the panel

structure of option prices and strike prices. In both cases it is not obvious to see how the original dependences characterized in calendar time translate in the dimensions we are looking at. This makes the aforementioned approaches developed for dependent data more difficult to justify and implement.

Acknowledgements

References

Abadir, K.M., Rockinger, M., 1997. Density-embedding functions. Discussion Paper, 97/16, University of York.

Abken, P., Madan, D., Ramamurtie, S., 1996. Estimation of risk-neutral and statistical densities by hermite polynomial approximations: with an application to Eurodollar futures options. Discussion Paper, Federal Reserve Bank of Atlanta.

Aït-Sahalia, Y., 1993. Nonparametric functional estimation with applications to financial models. Ph.D. Dissertation, M.I.T.

Aït-Sahalia, Y., 1996. Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527–560.

Aït-Sahalia, Y., Lo, A.W., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. Journal of Finance 53, 499–547.

Altman, N.S., 1987. Smoothing data with correlated errors. Technical Report 280, Department of Statistics, Stanford University.

Altman, N.S., 1990. Kernel smoothing of data with correlated errors. Journal of the American Statistical Association 85, 749–759.

Andrews, D.K., 1991. Asymptotic optimality of generalized C_L, cross-validation, and generalized cross-validation in regression with heteroskedastic errors. Journal of Econometrics 47, 359–377.

Barles, G., Burdeau, J., Romano, M., Samsœn, N., 1995. Critical stock price near expiration. Mathematical Finance 5, 77–95.

Barone-Adesi, G., Whaley, R.E., 1987. Efficient analytic approximation of American option values. Journal of Finance 42, 301–320.

Bates, D.S., 1991. The crash of '87 – was it expected? The evidence from options markets. Journal of Finance 46, 1009–1044.

Bensoussan, A., 1984. On the theory of option pricing. Acta Applicandae Mathematicae 2, 139–158.

Bodurtha, J.N., Courtadon, G.R., 1986. Efficiency tests of the foreign currency options market. Journal of Finance 42, 151–162.

Bossaerts, P., 1988. Simulation estimators of optimal early exercise. Working paper, Graduate School of Industrial Administration, Carnegie Mellon University.

Boyle, P., 1988. A lattice framework for option pricing with two state variables. Journal of Financial and Quantitative Analysis 23, 1–12.

Breen, R., 1991. The accelerated binomial option pricing model. Journal of Financial and Quantitative Analysis 26, 153–164.

Brennan, M.J., Schwartz, E., 1977. The valuation of American put options. Journal of Finance 32, 449–462.

Broadie, M., Detemple, J., 1996. American option valuation: new bounds, approximations, and a comparison of existing methods. Review of Financial Studies 9, 1211–1250.

Broadie, M., Detemple, J., Ghysels, E., Torrès, O., 1995. American options with stochastic volatility: a nonparametric investigation. Discussion paper, CIRANO, Montréal.

Broadie, M., Detemple, J., Ghysels, E., Torrès, O., 2000. American options with stochastic dividends and volatility: a nonparametric investigation. Journal of Econometrics 94, 53–92.

Büttler, H.-J., Waldvogel, J., 1996. Pricing callable bonds by means of Green's function. Mathematical Finance 6, 53–88.

Carr, P., Jarrow, R., Myneni, R., 1992. Alternative characterizations of American put options. Mathematical Finance 2, 87–106.

Carr, P., Faguet, D., 1994. Fast accurate valuation of American options. Working paper, Cornell University.

Chernov, M., Ghysels, E., 1998. What data should be used to price options? Discussion paper CIRANO, Montreal.

Cox, J.C., Ross, S.A., Rubinstein, M., 1979. Option pricing: a simplified approach. Journal of Financial Economics 7, 229–263.

Craven, P., Wahba, G., 1979. Smoothing noisy data with spline functions. Numerical Mathematics 31, 377–403.

de Matos, J.A., 1994. MSM estimators of American option pricing models. Ph.D. Dissertation, INSEAD, Fontainebleau.

Diz, F., Finucane, T.J., 1993. The rationality of early exercise decisions: evidence from the S&P100 index options markets. Review of Financial Studies 6, 765–797.

Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Dunn, K.B., Eades, K.M., 1989. Voluntary conversion of convertible securities and the optimal call strategy. Journal of Financial Economics 23, 273–302.

El Karoui, N., Karatzas, I., 1991. A new approach to the Skorohod problem and its applications. Stochastics and Stochastics Reports 34, 57–82.

Eubank, R.L., 1988. Spline Smoothing and Nonparametric Regression. Marcel Dekker, New York.

French, D.W., Maberly, E.D., 1992. Early exercise of American index options. Journal of Financial Research 15, 127–137.

Gallant, R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.

Garcia, R., Renault, E., 1996. Risk aversion, intertemporal substitution and option pricing. Discussion paper, CIRANO, Montréal and GREMAQ, Université de Toulouse I.

Gay, G.D., Kolb, R.W., Yung, K., 1989. Trader rationality in the exercise of futures options. Journal of Financial Economics 23, 339–361.

Geske, R., 1979. A note on an analytic formula for unprotected American call options on stocks with known dividends. Journal of Financial Economics 7, 375–380.

Geske, R., Johnson, H.E., 1984. The American put options valued analytically. The Journal of Finance 39, 1511–1524.

Ghysels, E., Harvey, A., Renault, E., 1996. Stochastic volatility. In: Maddala et al. (Eds.), Handbook of Statistics, Statistical Methods in Finance, vol. 14. North-Holland, Amsterdam, forthcoming.

Gouriéroux, C., Monfort, A., 1995. Simulation Based Econometric Methods. CORE Lecture Series. Louvain-la-Neuve.

Gouriéroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, S85–S118.

Gouriéroux, C., Monfort, A., Tenreiro, C., 1994. Kernel M-estimators: nonparametric diagnostics for structural models. Working Paper 9405, CEPREMAP, Paris.

Györfi, L., Härdle, W., Sarda, P., Vieu, P., 1989. In: Berger, J., et al. (Eds.), Nonparametric Curve Estimation from Time Series., Lecture Notes in Statistics, vol. 60. Springer, Heidelberg.

Härdle, W., 1990. Applied Nonparametric Regression. Econometric Society Monographs. Cambridge University Press, Cambridge.

Härdle, W., Linton, O., 1994. Applied nonparametric methods. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. Elsevier, Amsterdam (Chapter 38).

Härdle, W., Vieu, P., 1992. Kernel regression smoothing of time series. Journal of Time Series Analysis 13, 209–232.

Harrison, J.M., Kreps, D.M., 1979. Martingales and arbitrage in multiperiod securities markets. Journal of Economic Theory 20, 381–408.

Harvey, C.R., Whaley, R.E., 1992. Dividends and S&P100 index option valuation. Journal of Futures Markets 12, 123–137.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall, London.

Horowitz, J.L., 1993. Semiparametric estimation of a work-trip mode choice model. Journal of Econometrics 58, 49–70.

Huang, J., Subrahmanyam, M., Yu, G., 1996. Pricing and hedging American options: a recursive integration method. Review of Financial Studies 9, 277–300.

Hutchinson, J.M., Lo, A.W., Poggio, T., 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. Journal of Finance 49, 851–889.

Ingersoll, J., 1977. An examination of corporate call policies on convertible securities. The Journal of Finance 32, 463–478.

Jacka, S.D., 1991. Optimal stopping and the American put. Mathematical Finance 1, 1–14.

Jackwerth, J.C., Rubinstein, M., 1996. Recovering probability distributions from option prices. Journal of Finance 51, 1611–1631.

Karatzas, I., 1988. On the pricing of American options. Applied Mathematics and Optimization 17, 37–60.

Kim, I.J., 1990. The analytic valuation of American options. The Review of Financial Studies 3, 547–572.

Künsch, H.R., 1989. The Jacknife and the bootstrap for general stationary observations. Annals of Statistics 17, 1217–1241.

Kuske, R., Keller, J., 1998. Optimal excercise boundary for American put options. Applied Mathematical Finance 5, 107–116.

Lehmann, E.L., 1983. Theory of Point Estimation. Wiley, New York.

Liu, R.Y., Singh, K., 1992. Moving blocks jacknife and bootstrap capture of weak dependence. In: Lepage, R., Billard, Y. (Eds.), Exploring the Limits of Bootstrap. Wiley, New York.

McKean, H.P., 1965. Appendix: a free boundary problem for the heat equation arising from a problem in Mathematical Economics. Industrial Management Review 6, 32–39.

Merton, R.C., 1973. Theory of rational option pricing. Bell Journal of Economics 4, 141–183.

Myneni, R., 1992. The Pricing of the American Option. Annals of Applied Probability 2, 1–23.

OEX-S&P100 Index Options, 1995. Chicago Board Options Exchange, Inc.

Overdahl, J.A., 1988. The early exercise of options on treasury bond futures. Journal of Financial and Quantitative Analysis 23, 437–449.

Pedersen, A.R., 1995a. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Working paper, Department of Theoretical Statistics, University of Aarhus.

Pedersen, A.R., 1995b. Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. Working paper, Department of Theoretical Statistics, University of Aarhus.

Renault, E., 1996. Econometric models of option pricing errors, Discussion paper, GREMAQ, Université de Toulouse I.

Robinson, P.M., 1983. Nonparametric estimators for time series. Journal of Time Series Analysis 4, 185–207.

Rutkowski, M., 1994. The early exercise premium representation of foreign market American options. Mathematical Finance 4, 313–325.

Samuelson, P.A., 1965. Rational theory of warrant pricing. Industrial Management Review 6, 13–31.

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.

Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice and Visualization. Wiley, New York.

Selby, M.J.P., Hodges, S.D., 1987. On the evaluation of compound options. Management Science 33, 347–355.

Silverman, B.W., 1984. Spline smoothing: the equivalent variable kernel method. Annals of Statistics 12, 898–916.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.

Singh, R.S., Ullah, A., 1985. Nonparametric time-series estimation of joint DGP, conditional DGP, and vector autoregressions. Econometric Theory 1, 27–52.

Stutzer, M., 1996. A simple nonparametric approach to derivative security valuation. Journal of Finance 51, 1633–1652.

Van Moerbeke, P., 1976. On optimal stopping and free boundary problems. Archive for Rational Mechanics and Analysis 60, 101–148.

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

Whaley, R.E., 1981. On the valuation of American call options on stocks with known dividends. Journal of Financial Economics 9, 207–211.

Yu, G., 1993. Essays on the valuation of American options. Ph.D. Dissertation, New York University.

Zivney, T.L., 1991. The value of early exercise in option prices: an empirical investigation. Journal of Financial and Quantitative Analysis 26, 129–138.