Monte Carlo Pricing of American Options Using Nonparametric Regression

C. Pizzi *
Dept. of Statistics
University of Venice
S. Polo, 2347
30123 Venice
Italy

P. Pellizzari
Dept. of Applied Mathematics
University of Venice
Dorsoduro 3825/E
30123 Venice
Italy

First draft: July 2001
This draft: August 2002

Abstract

This paper provides an introduction to Monte Carlo algorithms for pricing American options written on multiple assets, with special emphasis on methods that can be applied in a multi-dimensional setting. Simulated paths can be used to estimate by nonparametric regression the continuation value of the option or the optimal exercise policy and the value functions can then be computed by backward induction. The flexibility of nonparametric regression allows to obtain accurate price estimates with remarkable speed. For illustrative purpose we price one- and two-dimensional American options.

1 Introduction

This paper deals with the problem of pricing American-style derivatives on multiple assets. The American feature offers to the holder the right to early exercise the option, cashing the payoff immediately. If rare cases are excluded, one being an American call option on one stock that does not pay dividends, there are not close evaluation formulae for such derivatives. One of the main difficulties lies in the computation of the optimal exercise time, as a stochastic control problem is to be solved. In particular, this fact has made till recently the application of Monte Carlo simulation difficult in such a context, because the standard *backward* arguments required by the

*We thank two referees and the participants to "IV Workshop di Finanza Quantitativa", Torino, 2003, for useful comments that improved the final version of the paper.

optimality principle do not allow straightforward use of implicitly *forward* Monte Carlo tools.

The literature that tries to overcome the lack of analytical pricing formulae for American options written on one asset is impressive and dates back to the binomial model of [Cox et al., 1979], that can be used checking at each node if immediate exercise is profitable. A selection of other important papers on the subject includes (but is in no way limited to) [Barone Adesi and Whaley, 1987], [Geske and Johnson, 1984], [Carr, 1998], [Bunch and Johnson, 2000], [Grant et al., 1997] and [Huang et al., 1996]. Some of these contributions require advanced mathematical tools, like Richardson extrapolation or recursive integration schemes. Binomial trees with tens of thousand nodes are still routinely used to get a quick approximate answer with little coding effort.

Some researchers focused in the last decade on methods that can in principle be employed to price options written on multiple assets. Pioneering work in [Bossaerts, 1989] and [Tilley, 1993] explicitly models the early exercise region or provides a bundling technique that groups simulated paths in bunches indexed by price. A strenuous effort to use simulation-based methods despite the slow convergence is motivated by the ease of implementation and by linear computational complexity in the number of risk sources (assets). However, Tilley's and many other algorithms give fairly inaccurate results and require exponential time in the number of assets.

The Stratified State Aggregation along Price (SSAP) method described in [Barraquand and Martineau, 1995] was a first important development of Tilley's ideas but nevertheless the burden of dimensionality is not solved and [Coyle and Yang, 1999] show that the method is non convergent. The paper of [Raymar and Zwecher, 1997] extends the SSAP methods by using a two dimensional stratification that can in some cases provide convergence when the SSAP method is failing. [Boyle et al., 1997] presents another method but its success is again limited by exponential complexity. In other cases, like the stochastic mesh method, [Broadie and Glasserman, 1997], this drawback is avoided at a price of dramatically slow $O(N^{-1/4})$ convergence, where N is the number of simulations.

A major breakthrough toward a more efficient solution is provided by [Longstaff and Schwartz, 2001], that show how to estimate by regression the continuation value conditionally on optimal exercise strategy. This paper ignited a strong interest in the use of regression to solve the American pricing problems, that can now be tackled in cases of practical importance, see [Haugh and Kogan, 2001, Andersen and Broadie, 2001]. The early work by [Carriere, 1996], that somehow did not get at first the deserved attention, is now recognized to contain many ideas developed in the following years. In order to use regression, the points of the design matrix are randomly sampled simulating lognormal stock paths and the continuation value is computed backward for each time-slice, allowing to evaluate the discounted

average of sample payoffs conditional on an approximate optimal timing of the exercise feature. In the Longstaff–Schwartz approach the sample paths are used to estimate weather immediate exercise is greater than the expected risk-neutral payoff, but only realized payoffs are used to compute the price.

The method is appealing even if there are details in its application that might require some effort. For example, the continuation value is computed using a parametric regression on a number of variables that are polynomial transformations of the normalized asset values. Though the authors claim that the results are quite insensitive to the chosen regressors, a subset selection is implicitly in order to ensure that convergence has occurred. The regressions in the examples provided in the Longstaff–Schwartz paper use for example a constant and 3 nonlinearly transformed regressors (based on Laguerre polynomials) to price a univariate put option. The number of regressors increases to 22 in pricing more complex derivatives and some numerical care is needed to get accurate results, as reported in Section 8 of the paper.

Moreover, the use of linear regression on nonlinearly transformed variables suggests that more sophisticated regression techniques could simplify some parts of the algorithm, avoiding experimentation and selection of different sets of variables or transformations. This idea is exploited in [Haugh and Kogan, 2001] where neural networks are used in view of their universal approximation properties and accurate bounds for the price are given. The use of nonparametric regression encompasses both the need of nonlinear transformations and variable subset selection. In fact, being the continuation value estimated by a flexible functional form (for example, locally linear), there is no purpose in estimating a multiple linear regression and it often suffices to use just the underlying stocks or risk sources.

This paper aims to provide a simple and introductory description of a pricing algorithm that makes use of nonparametric techniques, in the spirit of [Carriere, 1996]. The plan of the paper is the following: in the next section we describe how to work backward the value functions using Monte Carlo generated data. In Section 3 we briefly describe some nonparametric techniques and point to some relevant references. This section is only intended to suggest some introductory readings on nonparametric regression that might not be routinely known by the finance community. In Section 4 we present two applications by pricing an American 'standard' put option and an American put option written on the minimum of two assets. The results we obtain are compared to, respectively, [Longstaff and Schwartz, 2001] and [Boyle, 1988]. We then give some conclusive remarks.

2 The pricing framework

We describe in this section the valuation framework for an american option and propose a general pricing procedure that uses nonparametric regression. Denote by $\mathbf{X}_t = (S_{1t}, \ldots, S_{nt})$ the vector of state variables (prices) at time t and let $f(\mathbf{X}_t)$ be the payoff of the option. We discretize the lifespan of the option $[0, T]$ assuming that the early exercise is possible only in a set of $K-1$ discrete intermediate dates $0 < t_1 < \ldots < t_{K-1} < T$. For convenience we define $t_0 = 0$ and $t_K = T$. This Bermudan option can approximate its American counterpart if K is large. Assume for simplicity a constant interest rate r and let $F_t(\mathbf{X}_t)$ denote the value of the option (conditional on optimal exercise) at time t. It is well known that

$$F_{t_j}(\mathbf{X}_{t_j}) = \max\left(f(\mathbf{X}_{t_j}), \exp(-r(t_{j+1} - t_j))\mathbf{E}_{t_j}\left[F_{t_{j+1}}(\mathbf{X}_{t_{j+1}})\right]\right), \qquad (1)$$

where the $E_{t_j}[\cdot]$ denotes expected value conditional on the information at t_j, along with the boundary condition

$$F_T(\mathbf{X}_T) = f(\mathbf{X}_T).$$

Conditional on optimal exercise in $[t_{j+1}, T]$, equation (1) says that the value of the option at time t_j is the maximum between the gain of immediate exercise (which is a known quantity) and the discounted expected value of holding the right of future exercise. Note that taking the maximum between two operands closely matches the structure of an American option, that allows at any time only two strategies, but (1) can be adapted to more general situations where various different decisions can be taken. In this case, the maximum among the various outcomes is to be selected (for an example, think to the classical oil well that can be opened, temporarily put in idle state, or abandoned forever depending in the simplest case on the oil price at time t).

The price of the option is then simply $E_0[F_0(\mathbf{X}_0)]$, that can be estimated as

$$\frac{1}{N}\sum_{i=1}^{N}\max\left(f(\mathbf{X}_0), \exp(-rt_1)\hat{F}_{t_1}(\mathbf{X}_{t_1}^{(i)})\right),$$

where $\hat{F}_{t_1}(\mathbf{X}_{t_1})$ is a (sample) estimate of $E_0[F_{t_1}(\mathbf{X}_{t_1})]$, $\mathbf{X}_{t_1}^{(i)}$ is the $i-$th realization of state variable $\mathbf{X}$ at time t_1 and N is the number of simulations.

The problematic part of backward evaluation of (1) lies in the computation of the expected value

$$\mathbf{E}_{t_j}\left[F_{t_{j+1}}(\mathbf{X}_{t_{j+1}})\right] = \mathbf{E}\left[F_{t_{j+1}}(\mathbf{X}_{t_{j+1}})|\mathbf{X}_{t_j}\right]. \qquad (2)$$

A number of approaches has been used to approximate such conditional mean. [Raymar and Zwecher, 1997] estimates using a bundling approach

the transition probabilities from state $\mathbf{X}_{t_j}$ to $\mathbf{X}_{t_{j+1}}$ that allow to compute the mean. A more direct and early approach is advocated in [Rebonato and Cooper, 1998] that estimates the mean using Fast Fourier methods. A somewhat similar idea is presented and developed in [Longstaff and Schwartz, 2001] that again estimate directly the conditional mean using a least square linear regression of $F(\mathbf{X}_{t_{j+1}})$'s on the (transformed) sample $\mathbf{X}_{t_j}$'s.

We propose to estimate the value function $F_{t_j}(\mathbf{x})$ by

$$\begin{aligned} \hat{F}_{t_j}(\mathbf{X}_{t_j}) &= \max(f(\mathbf{X}_{t_j}), e^{-r(t_{j+1}-t_j)}\hat{m}_{t_j}(\mathbf{X}_{t_j})), j = K-1,\ldots,1 \quad (3) \\ \hat{F}_T(\mathbf{X}_T) &= f(\mathbf{X}_T), \end{aligned}$$

where $\hat{m}_{t_j}(\mathbf{X}_{t_j})$ is approximated regressing nonparametrically on $\mathbf{x}$

$$\hat{m}_{t_j}(\mathbf{x}) = \mathbf{E}\left[\hat{F}_{t_{j+1}}(\mathbf{X}_{t_{j+1}})|\mathbf{X}_{t_j} = \mathbf{x}\right]. \quad (4)$$

The expectation $\hat{m}(\mathbf{x})$ can be obtained nonparametrically by various algorithms like k-nearest-neighbors, kernel regression, LOESS smoothing or other techniques briefly reviewed in the next section.

In practice the functions m's are estimated using N Monte Carlo paths of prices $\{\mathbf{X}_{t_j}^{(i)}, j = 1,\ldots,K\}$ where $i = 1,\ldots,N$ scans the number of simulations and $j = 1,\ldots,K$ runs along the time variable. At time t_j a (nonparametric) regression is then performed on the dependent variable $F(\mathbf{X}_{t_{j+1}})$ using as independent variable(s) $\mathbf{X}_{t_j}$. Knowing the estimated value of continuation permits then to choose for each time slice and each path if immediate exercise is preferable. We name the combined use of Monte Carlo and nonparametric regression as Nonparametric Monte Carlo (NPMC) algorithm.

Among the advantages of this approach with respect to linear regression there are greater flexibility in choosing a nonlinear response function $\hat{m}(\mathbf{x})$ and no need to select a proper subset of transformed regressors. Moreover, we have not to restrict the design matrix, as in [Longstaff and Schwartz,2001], to the paths that have positive immediate payoffs and obtain the price by discounting in backward fashion the estimated value functions and not the realized payoffs. From this specific point of view the method we propose is similar to [Rebonato and Cooper, 1998]. In [Carriere, 1996] both estimators are considered.

A drawback of multivariate nonparametric over linear regression is the well known curse of dimensionality that might restrict the application of NPMC to options written on few underlying assets, due to exponential increase in the required sample size. As argued for example in [Fan and Gijbels, 1996], if a neighborhood contains b points along each of d dimensions then there are b^d data points in the corresponding $d-$dimensional neighborhood. This means that much larger datasets are needed, with heavy consequences on the computational requirements of the approach even if d is moderate. The curse

of dimensionality problem is discussed among others in [Silverman, 1986] and [Härdle, 1990].

3 A primer in nonparametric regression

In this section we present, in an informal rather than rigorous way, some well known nonparametric models to estimate a conditional expected value. For the sake of simplicity, we will assume in the remaining part of the section the existence of a unique risk source. As discussed previously, the estimation of the conditional mean (4) is of primary interest in American option pricing.

Suppose we observe $(X_i, Y_i), i = 1, \ldots, N$ and we want to study the functional relationship that links the two variables. That is we suppose $Y_i = m(X_i) + \epsilon_i$, where ϵ_i are observation errors and we are interested in the function m.

The scatter plot of the data points is a useful device to obtain a preliminary qualitative idea about the relation. However, care must be exercised in order to avoid mistakes: the eye may fall for the concentration of the points in some region, or by contrast it could be distracted by extreme points. Figure 1 depicts 1000 simulated data points. The values of the explanatory variable X and the noise ϵ are drawn from a standard gaussian random variable and the response variable Y is the following nonlinear transformation of the X variable:

$$Y_i = \sin(X_i) + \frac{1}{2}\epsilon_i, \quad i = 1, \ldots, 1000$$

It might be not unreasonable at first glance to assume that a linear model is appropriate even though some doubts about inclusion of additional powers of X in the analysis could arise.

We may specify the model following a parametric or a nonparametric approach. The first assumes we know the functional form of the regression curve m and that it can be approximated by a parametric model, i.e. a functional form depending on a finite set of parameters. A polynomial regression equation is an example of a parametric model. One of the simplest case is the straight line and we must estimate the intercept and the slope. The solution of the minimization problem $\min_{a,b} \sum_{i=1}^{n} (Y_i - a - bX_i))^2$ gives us an estimate of the parameters. This method is known as Ordinary Least Square (OLS) and if we assume the gaussian distribution for the errors ϵ_i it is equivalent to the Maximum Likelihood estimation method.

Observe that a parametric approach has been adopted in the American option pricing algorithm of Longstaff and Schwartz but, due to the nonlinear regression function, many linear terms are needed to obtain reasonable precision.

The nonparametric approach departs from the global paradigm of the parametric approach and considers flexible functional forms. The errors dis-

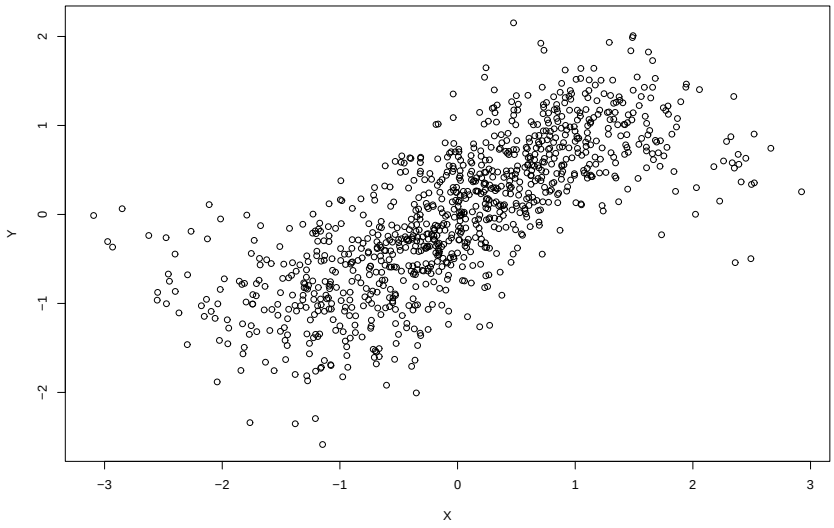

Figure 1: The simulated data: $X_i, \epsilon_i \sim N(0,1)$, $Y_i = sin(X_i) + \frac{1}{2}\epsilon_i$, N=1000.

tribution and the functional form of the relationship are not specified in advance and hence we say it is a distribution-free methods. In the nonparametric approach the regression curve is represented in terms of a function that is "built" point by point. The value of $m(x)$ at each fitted point is estimated using a subset of the sample that is changing together with the value of the point x where we fit.

Some natural questions at this stage are: how many data points do we consider to estimate the value of the regression function? What is the relevance of each data point in the fitting processes? How we summarize the sample information to estimate the conditional expected value? The combination of different answers to these questions has produced many different local models able to capture the features of the regression curve.

The *k-nearest-neighbors* (k-NN) method, introduced in [Loftsgaarden and Quesenberry, 1965], obtains $\hat{m}_k(x)$ at point x using a constant number k of observations to compute the weighted mean

$$\hat{m}_k(x) = \sum_{i=1}^{n} W_{k,i}(x) Y_i,$$

where the weights $W_{k,i} = 1/k$ if X_i is one of the k data points closest to the estimation point x, otherwise $W_{k,i} = 0$. The smoothing parameter k plays a key role controlling the smoothness of the fitted curve. On one hand, if k is of the same order of N the local behavior of $\hat{m}_k(x)$ disappears and degenerates to the non conditional expected value. The symmetric case, when k is too small, leads to the loss of smoothness of the curve.

The above idea of local model is that the nearest data points are more informative about the behavior in a neighborhood of x. If the distance from the fit point is important, than the weights $W_{k,i}$ should be a decreasing function of this distance. So we can define the weights as

$$W_{k,i}(x) = \frac{K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)},$$

where $K_h(u) = h^{-1}K(u/h)$ is called a *kernel function*. The weights depend on the kernel function and on the smoothing parameter h, called *bandwidth*. The kernel estimator, first introduced in [Nadaraya, 1964] and [Watson, 1964] for the regression curve is

$$\hat{m}_h(x) = \frac{\sum_{i=1}^{n} K_h(x - X_i)Y_i}{\sum_{i=1}^{n} K_h(x - X_i)}.$$

Like in the k-NN models, when the smoothness parameter $h \to \infty$ the estimated regression curve degenerates to the non conditional expected value of Y, while $h \to 0$ will produce essentially interpolation.

Another way to estimate the regression function replaces the weighted average by polynomial in x. The polynomial parameters are estimated by solving the following minimization problem

$$\min_{\mathbf{a}} \sum_{i=1}^{n} K_h(x - X_i)(Y_i - \hat{m}_h(x; \mathbf{a}))^2,$$

where $\mathbf{a}$ is a local vector of polynomial coefficients. The weights are computed, as before, using a kernel function.

The Locally Weighted Regression Smoother (LOESS) ([Cleveland, 1979] and [Cleveland and Devlin, 1988]) and the Local Polynomial Models (LPM) are two models that (locally) use the polynomial approximation, see [Fan and Gijbels, 1996] for a complete treatment of the subject. Other developments of the same idea are based on the use of non-polynomial local approximations like spline functions, see [de Boor, 1978] or [Green and Silverman, 1994], that have nice theoretical and computational features.

The LOESS is in some respect similar to k-NN model as it uses a constant fraction, often named *span*, $0 \leq d \leq 1$ of data points in a two-step estimation procedure. In the first step, the polynomial parameters are estimated weighting each data points by a kernel function. In the following step, the weights are updated to take into account the residuals obtained by the previous fit: the points that have large residuals are downweighted so that robustness with respect to outliers is obtained.

In the case of LPM, the parameters of the local polynomial are estimated using the points whose distance from the fit point is smaller than the bandwidth h. Hence the estimation of the regression curve considers a variable number of data points, weighted by a kernel function.

Observe that there is no need to specify parametrically a regression function. Hence, no trial-and-error effort is required to introduce and validate additional terms in the regression model. The potential advantages in the estimation of the continuation value are apparent in the option pricing framework. On the other hand, some smoothing parameter (d and h for LOESS and LPM, respectively) must nevertheless be chosen with care to get reliable results from any nonparametric method. The book [Fan and Gijbels, 1996] discusses the convergence of nonparametric estimators that is asymptotically obtained, under suitable regularity conditions, if $h \to 0, N \to \infty$ and $Nh \to \infty$.

Figure 2 shows a comparison of different regression models, both linear and nonlinear, together with the true sinusoidal transformation. It is manifest that nonparametric models track closely the true regression function in the whole domain, while the linear model is somehow misleading and should be re-estimated with additional regressors.

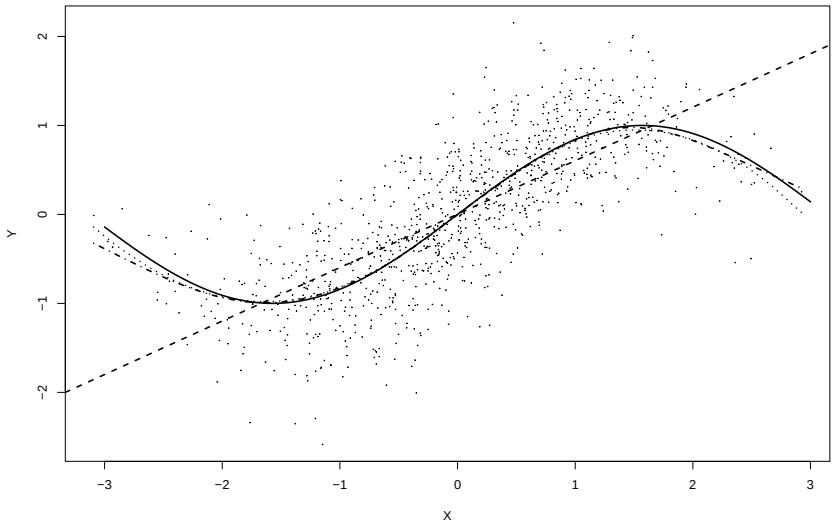

Figure 2: Comparison of different models: true curve (solid), linear model (dashed), LOESS (dotted) and spline regression (dotdashed).

4 Some applications of NPMC for option pricing

We provide in this section two applications to American option pricing, comparing our results with the ones given in [Longstaff and Schwartz, 2001] and [Boyle, 1988] on univariate and bivariate derivatives.

Consider first an American put option written on an asset following the

equation

$$dS = rSdt + \sigma SdZ,$$

where r, σ are constants and Z is a standard Brownian motion. The strike price is fixed at 40 and $r = 0.06$. Figure 3 depicts the estimated value functions $F_{t_j}, j = 1, \ldots, K$, that are computed recursively starting from the lower left to the upper right corner. The figure shows that the unknown value functions are smoothly estimated using nonparametric regression. Table 1 reports the prices obtained by the NPMC method for a variety of initial prices S_0, volatilities σ and maturities T. All the results are obtained using $N = 500$ antithetic simulations (hence 1000 random numbers are used for the computation of each price) assuming that the option is exercisable 10 times per year.

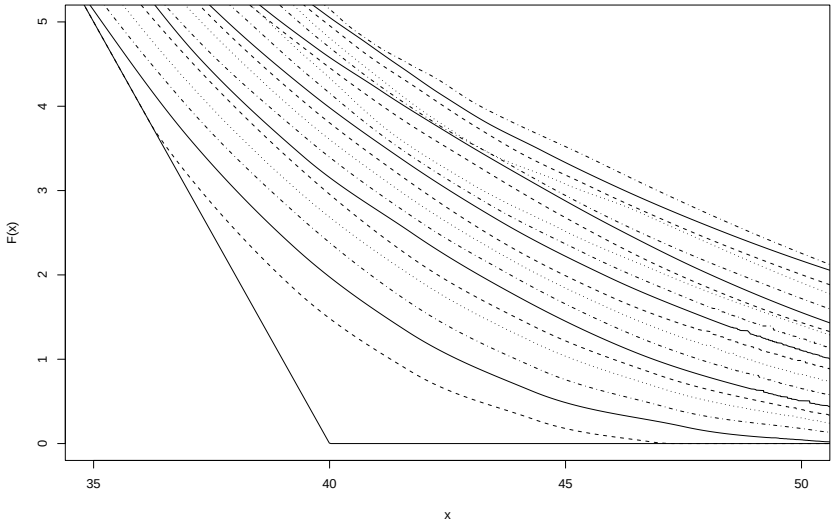

Figure 3: Estimated functions F_{tj} with $K = 20, \sigma = 0.4, N = 1000$. For greater visibility the functions are subsequently plotted with solid, dashed, dot-dashed and dotted lines. Estimation proceeds from the payoff profile at $t = T$ in the lower left corner and ends estimating F_{t_1}, in the upper right part of the plot.

Table 1 shows that NPMC slightly underestimates the price of the option. The error ranges from less than 1 cent to roughly 1/8 (which is often the tick in option markets) in some unfavorable cases. The fact that the obtained prices are slightly smaller than the correct ones is expected as only 10 exercise date are possible in one year. Indeed this numerical procedure approximates an American option with a Bermudan counterpart that is obviously less worthy. More details on the reason of this underpricing and on

Table 1: Price of an American put option, obtained using LOESS and Spline regression ($N = 500$ and $N = 2000$, respectively) and 10 exercise times per year. The reported prices are averaged over 50 repetitions and the standard deviation (SD) is provided. The column labeled "FD" (for finite difference) is taken from [Longstaff and Schwartz, 2001] and contains an accurate estimate of the price.

				LOESS ($N = 500$)		Splines ($N = 2000$)	
S_0	σ	T	FD	Price	SD	Price	SD
36	.2	1	4.478	4.439	.040	4.452	.018
		2	4.840	4.784	.034	4.826	.017
	.4	1	7.101	7.067	.062	7.097	.027
		2	8.508	8.365	.055	8.522	.033
40	.2	1	2.314	2.294	.042	2.309	.031
		2	2.885	2.867	.049	2.888	.028
	.4	1	5.312	5.278	.072	5.317	.047
		2	6.920	6.814	.077	6.941	.032
44	.2	1	1.110	1.106	.046	1.111	.028
		2	1.690	1.715	.055	1.716	.028
	.4	1	3.948	3.929	.104	3.942	.032
		2	5.647	5.564	.089	5.695	.052

the ways to overcome it are provided if we analyze more closely the behavior of NPMC. For brevity, we focus on the middle part of Table 1 (when $S_0 = 40$) and examine only the option of the 5-th row of Table 1. To investigate the dependence of the price on N and K, Table 2 shows some pricing results when the number of simulations and the exercise dates are increased.

Table 2: Average price (50 repetitions) of the put option with $S_0 = 40, \sigma = 0.2, T = 1$ for different exercise dates K and number of simulations N (standard deviations in brackets). The correct price is 2.314, as reported in Table 1.

	N (Simulations)		
K	500	1000	2000
10	2.300 (.037)	2.310 (.039)	2.305 (.021)
20	2.310 (.043)	2.314 (.030)	2.319 (.020)
50	2.324 (.038)	2.325 (.024)	2.330 (.020)

In general, increasing K produces more accurate prices. This is graphically depicted in Figure 4 where boxplots of 50 computed prices are depicted for $K = 10, 20, 50$ (the first column of Table 2 is 'graphed'). For comparison, [Longstaff and Schwartz, 2001] get 2.313 setting $K = 50, N = 50000$ (of course their standard deviation is considerably smaller than ours). There

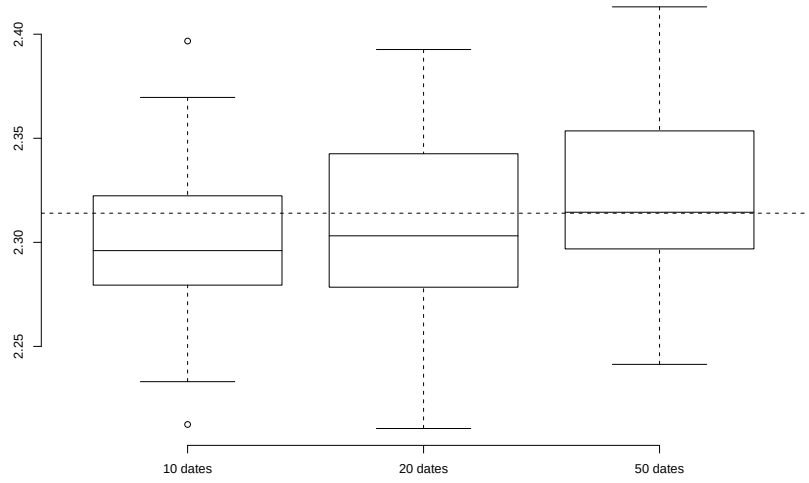

Figure 4: Boxplots of 50 prices of the American put option with $S_0 = 40, \sigma = 0.2$ and $T = 1$ for different K (number of exercise dates). It is apparent that increasing K produce more accurate results, with no visible effect on the standard deviation of the result. The dashed horizontal line corresponds to the correct price, 2.314.

is no surprise in noticing in Table 2 that increasing N decreases the standard deviation of the simulation with little effect on the accuracy of the estimate. As the computation complexity of the LOESS estimator goes with the square of N, it is generally advisable to increase K, that only affects linearly the required computing time, instead of increasing the number of simulations. Very tight bounds on the computed price are obtained in [Haugh and Kogan, 2001] using a useful dual representation of the option price.

We stress that the above results are not specifically due to the nonparametric regression technique used (namely, LOESS). The last two columns of Table 1 shows the pricing results obtained by spline regression with 2000 antithetic simulations, being the other parameters fixed as in the LOESS case. The prices are accurate (the error is at most 5 cents) and the standard deviations in the last column is roughly halved with respect to the LOESS ones, which is in agreement with the expected $O(N^{-1/2})$ law for MC standard deviation. Hence, we argue that the good results can be obtained using different nonparametric methods and we defer to future research the detailed comparison of a host of procedures in order to cast further light on the best methodology in terms of accuracy and speed[1]. These findings back up the

[1]All the pricing algorithms have been coded using `R`, a program that

results obtained in other papers, where different nonlinear regressions are used and good accuracy is obtained regardless of the chosen specific method.

In the remaining part of this section, we price a bivariate American put option on the minimum of two correlated lognormal assets that is evaluated in [Boyle, 1988] using a lattice approach. We consider an option that gives the right to cash at any $0 \leq t \leq T$ the payoff $\max(0, X - \min(S_{1t}, S_{2t}))$, where X is the strike price. In this case the state vector is then $\mathbf{X}_t = (S_{1t}, S_{2t})$ and the assets follow the equations

$$dS_i/S_i = rdt + \sigma_i dZ_i, \ i = 1, 2.$$

Table 3: Average prices (with standard deviations based on 50 repetitions) of an American put option on the minimum of two assets, based on $N = 8000$ simulations and LOESS regression.

Strike Price	[Boyle, 1988]	LOESS	SD
35	1.423	1.429	.029
40	3.892	3.911	.036
45	7.689	7.714	.055

Denoting the correlation of the brownian motions Z_i's by ρ, the parameters we use are the following: $S_{01} = S_{02} = 40, \sigma_1 = 0.2, \sigma_2 = 0.3, \rho = 0.5, r = 0.04879$ continuously compounded corresponding to 5% effective per annum, $T = 7/12$ years and exercise prices $X = 35, 40, 45$. Pricing results provided in Table 3 are based on LOESS regression with $N = 8000$ simulations and span $d = 0.05$ (i.e. only 5% of the observations are on average used in the regression). The results are accurate and the difference from Boyle values is at most 0.025. To give an idea of the required computation effort, the time needed to compute one of the entries of Table 3 (50 prices) is about 9.3 seconds on a Celeron 900 MHz. As the code we employ is not compiled, we have no doubts that much faster pricing can be easily achieved.

5 Conclusions

This papers survey Monte Carlo algorithms for American option pricing, based on nonparametric regression on simulated paths and iterated backward estimation of the continuation value. We aim to provide a simple introductory description of these methods, discussing advantages and pitfalls of linear and nonparametric regression. We show that the estimation of the nonlinear conditional expectation is eased if flexible nonparametric

can be downloaded at `http://cran.r-project.org`. The file containing the program for American options pricing using spline regression is available at `http://www.dma.unive.it/~paolop/long2d_spline.R`

regression is used as, for example, there is no need to perform a subset selection of regressors and the risk sources alone are used as independent variables.

We informally review some nonparametric regression models, providing an illustrative example and pointing to some relevant work in this area.

Some applications to option pricing are described. A standard American put option is first evaluated and the results are compared to the ones in [Longstaff and Schwartz, 2001], showing that accurate price estimation is obtained using a smaller number of simulations and exercise dates. This appears to be true no matter of which regression model is selected (LOESS or spline smoothing, for example, give fairly similar results). We analyze the unbiasedness and standard deviations of the computed prices varying the number of exercise dates K.

We finally apply the method to an American put option on the minimum of two assets, showing that the algorithm can be tailored to multivariate derivatives with little effort (at least when the number of assets is low to avoid the curse of dimensionality).

References

[Andersen and Broadie, 2001] Andersen, L. and Broadie, M. (2001). A prima-dual simulation algorithm for pricing multi-dimensional american options. Technical report, Columbia University, New York.

[Barone Adesi and Whaley, 1987] Barone Adesi, G. and Whaley, R. (1987). Efficient analytic approximation of american option values. *Journal of Finance*, 42:301–320.

[Barraquand and Martineau, 1995] Barraquand, J. and Martineau, D. (1995). Numerical valuation of high dimensional multivariate american securities. *Journal of Financial and Quantitative Analysis*, 30:383–405.

[Bossaerts, 1989] Bossaerts, P. (1989). Simulation estimators of optimally early exercise. Technical report, Graduate School of Industrial Admistration, Carnegie Mellon University.

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

[Boyle et al., 1997] Boyle, P., Broadie, M., and Glasserman, P. (1997). Monte carlo methods for security pricing. *Journal of Economics Dynamics and Control*, 21:1267–1321.

[Broadie and Glasserman, 1997] Broadie, M. and Glasserman, P. (1997). A stochastic mesh method for pricing. high-dimensional american options. Technical report, Columbia University Working Paper.

[Bunch and Johnson, 2000] Bunch, D. and Johnson, H. (2000). The american put option and its critical stock price. *Journal of Finance*, 55:2333–2356.

[Carr, 1998] Carr, P. (1998). Randomization and the american put. *Review of Financial Studies*, 11:597–626.

[Carriere, 1996] Carriere, J. (1996). Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19:19–30.

[Cleveland, 1979] Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Society*, 74:829–836.

[Cleveland and Devlin, 1988] Cleveland, W. and Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Society*, 83:596–610.

[Cox et al., 1979] Cox, J. C., Ross, S. A., and Rubinstein, M. (1979). Option pricing: a simplified approach. *Journal of Financial Economics*, 7:229–263.

[Coyle and Yang, 1999] Coyle, L. and Yang, J. (1999). Analysis of the ssap method for the numerical valuation of high-dimensional multivariate american securities. *Algoritmica*, 25:75–98.

[de Boor, 1978] de Boor, C. (1978). *A Practical Guide to Splines*. Springer-Verlag.

[Fan and Gijbels, 1996] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Application*. Chapman & Hall, London.

[Geske and Johnson, 1984] Geske, R. and Johnson, H. (1984). The american put option valued analytically. *Journal of Finance*, 39:1511–1524.

[Grant et al., 1997] Grant, D., Vora, G., and Weeks, D. (1997). Path-dependent options and early exercise: extending the monte carlo simulation method. *Management Science*, 43:1589–1602.

[Green and Silverman, 1994] Green, P. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. Chapman and Hall.

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

[Haugh and Kogan, 2001] Haugh, M. and Kogan, L. (2001). Pricing american options: a duality approach. Technical report, The Wharton School, University of Pennsylvania.

[Huang et al., 1996] Huang, J., Subrahmanyam, M., and Yu, G. (1996). Pricing and hedging american options: A recursive integration method. *Review of Financial Studies*, 9(1):277–300.

[Loftsgaarden and Quesenberry, 1965] Loftsgaarden, D. and Quesenberry, G. (1965). A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, 36:1049–1051.

[Longstaff and Schwartz, 2001] Longstaff, F. and Schwartz, E. (2001). Valuing american options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14(1):113–147.

[Nadaraya, 1964] Nadaraya, E. (1964). On estimating regression. *Th. Prob. Appl.*, 9:141–142.

[Raymar and Zwecher, 1997] Raymar, S. and Zwecher, M. (1997). A monte carlo valuation of american call options on the maximum of several stocks. *The Journal of Derivatives*, Fall 1997:7–23.

[Rebonato and Cooper, 1998] Rebonato, R. and Cooper, I. (1998). Coupling backward induction with monte carlo simulations: a fast fourier transform (fft) approach. *Applied Mathematical Finance*, 5:131–141.

[Silverman, 1986] Silverman, B. (1986). *Density Estimation for Statistic and Data Analysis*. Chapman & Hall, London.

[Tilley, 1993] Tilley, J. (1993). Valuing american options in a path simulation model. *Transactions of the Society of Actuaries*, 45:83–104.

[Watson, 1964] Watson, G. (1964). Smooth regression analysis. *Sankhya*, 26:359–372.