// Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Regression methods in pricing American and Bermudan options using consumption processes

Denis Belomestny, [1] Grigori N. Milstein [2] and Vladimir Spokoiny [3]

submitted: 22nd June 2006

[1] Weierstrass Institute
for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117
Berlin, Germany
E-Mail: belomest@wias-berlin.de

[2] Department of Mathematics,
Ural State University,
Lenin Str. 51, 620083
Ekaterinburg, Russia
E-Mail: Grigori.Milstein@usu.ru

[3] Weierstrass Institute
for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117
Berlin, Germany
E-Mail: spokoiny@wias-berlin.de

No. 1145
Berlin 2006

2000 *Mathematics Subject Classification.* 60H30, 65C05, 91B28 .

Key words and phrases. American and Bermudan options, low and upper bounds, Monte Carlo method, consumption process, regression methods, optimal stopping times.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

ABSTRACT. Here we develop methods for efficient pricing multidimensional discrete-time American and Bermudan options by using regression based algorithms together with a new approach towards constructing upper bounds for the price of the option. Applying the sample space with payoffs at the optimal stopping times, we propose sequential estimates for continuation values, values of the consumption process, and stopping times on the sample paths. The approach admits constructing both low and upper bounds for the price by Monte Carlo simulations. The methods are illustrated by pricing Bermudan swaptions and snowballs in the Libor market model.

1. Introduction

Valuation of high-dimensional American and Bermudan options is one of the most difficult numerical problems in financial engineering. Besides its practical relevance, investigations in this field are of great theoretical importance because pricing of the American style options is an archetype for high-dimensional optimal stopping problems. Several approaches have been proposed recently for pricing such options using Monte Carlo simulation technique (see, e.g. [1]-[12], [14]-[17], [21, 22, 24] and references therein). With simulation approaches it is often an open question whether or not an obtained numerical result is sufficiently accurate. As a rule, during the realization of a numerical procedure there arise many errors of different kind which are difficult to take into account. That is why in a number of works (see, e.g. [3, 4, 14, 15, 16, 17, 21, 22]), different procedures are proposed that are able to produce lower and upper bounds for the true price. The knowledge of lower and upper bounds makes possible to evaluate the accuracy of price estimates. Our aim is to construct effective numerical methods providing with both lower and upper bounds for the price of American and Bermudan options.

In [3] we develop an approach for pricing American options both in the case of discrete-time and continuous-time financial models. The approach is based on the fact that an American option is equivalent to a European one with a consumption process involved (the so called Earlier Exercise Premium representation). It allows us, in principle, to construct iteratively a sequence v^1, V^1, v^2, V^2, v^3, ..., where v^1, v^2, v^3, ..., is an increasing sequence of lower bounds and V^1, V^2, ..., is a decreasing sequence of upper bounds. Unfortunately, the construction of the above sequence of bounds requires very laborious calculations. Even V^2 is, as a rule, too expensive. In [4] we propose to use an increasing sequence of low bounds for constructing both upper bound and low bound at initial position (t_0, X_0). It is assumed that the sequence is not too expensive from computational point of view. This can be achieved by using local low bounds which take into account a small number of steps ahead. The method of [4] is suitable for getting rough estimates. However, for obtaining more accurate results one needs rather expensive calculations.

Let us consider a discrete-time financial model

$$(B_{t_i}, X_{t_i}) = (B_{t_i}, X^1_{t_i}, ..., X^d_{t_i}),\ i = 0, 1, ..., \mathcal{I},$$

where B_{t_i} is price of a scalar riskless asset (we assume that B_{t_i} is deterministic and $B_{t_0} = 1$) and $X_{t_i} = (X^1_{t_i}, ..., X^d_{t_i})$ is price vector of risky assets. Along with index t_i we shall use below the index i, writing (t_i, X_i) instead of (t_i, X_{t_i}). Let $f_i(x)$ be a payoff at time t_i provided that $X_{t_i} = X_i = x$, $x \in \mathbf{X} \subset \mathbf{R}^d$, where $\mathbf{X}$ is a state space (e.g., $\mathbf{X} = \mathbf{R}^d$, $\mathbf{X} = \mathbf{R}^d_+$).

We assume that the modelling is based on the filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_i)_{0 \le i \le \mathcal{I}}, P)$, where the probability measure P is the risk-neutral pricing measure for the problem under consideration, and X_i is a Markov chain with respect to the filtration $(\mathcal{F}_i)_{0 \le i \le \mathcal{I}}$.

With respect to the probability measure P the discounted process X_i/B_i is a martingale and the price $u_i(X_i)$ of the American option is given by

$$u_i(x) = \sup_{\tau \in \mathcal{T}_{i,\mathcal{I}}} B_i E\left(\frac{f_\tau(X^{t_i,x}_\tau)}{B_\tau}\right). \tag{1.1}$$

In (1.1) $X^{t_i,x}_{t_j}$ is the value of Markov chain at instant $t_j \ge t_i$ starting at t_i from x, $\mathcal{T}_{i,\mathcal{I}}$ is set of stopping times τ taking values in $\{i, i+1, ..., \mathcal{I}\}$.

The value process u_i (Snell envelope) can be determined by induction as follows:

$$\begin{aligned} &u_{\mathcal{I}}(x) = f_{\mathcal{I}}(x), \\ &u_i(x) = \max\left\{ f_i(x), B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x\right)\right\},\ i = \mathcal{I}-1, ..., 0. \end{aligned} \tag{1.2}$$

We see that theoretically the problem of evaluating $u_0(X_0)$, the price of the discrete-time American option at the initial position (t_0, X_0), is easily solved using iteration procedure (1.2). However, if X is high dimensional and $\mathcal{I}$ is large, the iteration procedure is not practical.

In order to use regression methods for sequential evaluation of u_i, one can consider the $(d+1)$-dimensional sample

$$({}_mX_i, \frac{B_i}{B_{i+1}} u_{i+1}({}_mX_{i+1})),\ m = 1, ..., M,\ i = 0, ..., \mathcal{I}-1, \tag{1.3}$$

from $(X_i, u_{i+1}(X_{i+1}))$, where $(t_i,\ {}_mX_i)$ are M independent trajectories all starting from the point (t_0, X_0) (see, e.g., [24] and [12]).

The samples using optimal stopping times $\tau^{t_i,x} = \tau^{i,x}$ were first introduced in [19] (see [9] and [12] as well). They are from $(X_i,\ f_\tau(X^{t_{i+1},X_{i+1}}_\tau)) = (X_i,\ f_\tau(X^{t_i,X_i}_\tau))$, with $\tau = \tau^{t_{i+1},X_{i+1}}$ and have the form

$$({}_mX_i,\ \frac{B_i}{B_\tau} f_\tau({}_mX^{t_{i+1},\ {}_mX_{i+1}}_\tau)) = ({}_mX_i,\ \frac{B_i}{B_\tau} f_\tau({}_mX^{t_i,\ {}_mX_i}_\tau)),\ \tau = \tau^{t_{i+1},\ {}_mX_{i+1}},\ m = 1, ..., M. \tag{1.4}$$

Applying (1.3), we use some estimate $\hat{u}_{i+1}(X_{i+1})$ instead of $u_{i+1}(X_{i+1})$ while applying (1.4), we can employ an estimate $\hat{\tau} = \hat{\tau}^{t_{i+1},\ X_{i+1}}$ for $\tau^{t_{i+1},\ X_{i+1}}$. This makes possible to construct a low bound for continuation value (low continuation value) and an upper bound for consumption process (upper consumption process). If the payoff at $(t_i,\ {}_mX_i)$ is less or equal to a low continuation value, then first, the position $(t_i,\ {}_mX_i)$ belongs to the continuation region (consequently, it is natural to take $\hat{\tau}^{t_i,\ {}_mX_i} = \hat{\tau}^{t_{i+1},\ {}_mX_{i+1}}$) and second the consumption process at $(t_i,\ {}_mX_i)$ is equal to zero. Otherwise the position $(t_i,\ {}_mX_i)$ can belong either to the exercise region or to the continuation region. In the latter case we compute the upper consumption process at $(t_i,\ {}_mX_i)$ as a difference between the payoff and the low continuation value and set $\hat{\tau}^{t_i,\ {}_mX_i} = t_i$. As a result all the positions $(t_i,\ {}_mX_i)$ are equipped with stopping times and consumption processes. Due to this it becomes possible to find the low and upper bounds for the price of the option under consideration at the initial position (t_0, X_0).

In Section 2, we recall the approach (see [3], [4]) to pricing American and Bermudan options using consumption processes in the form suitable for our purposes. Furthermore, we give here a comparison with the dual approach (see [21], [14]) for the first time. In Section 3, we propose a number of algorithms for subsequent estimating optimal stopping times and continuation values using different regression methods. Special attention is paid to linear regression methods (see [19] and [9]). In contrast to other works using the regression approach in pricing American and Bermudan options, we construct together with an estimate of continuation value an upper consumption process. Section 4 gives formulas for the Monte Carlo calculation of low and upper bound at the initial position (t_0, X_0). Section 5 is devoted to simulations: the results of numerical experiments for Bermudan swaptions and cancellable snowballs in a full factor Libor market model confirm efficiency of the proposed algorithms. Finally, in Appendix we show that using of procedure (1.2) and sample (1.3) for sequential evaluating $u_i(X_i)$ together with modern methods of multidimensional approximation (see e.g., [10], [25] and references therein) can give effective algorithms for pricing high-dimensional American and Bermudan options.

2. The approach based on consumption processes

To be self-contained, let us briefly recall the approach to pricing American and Bermudan options using consumption processes [3].

2.1. The continuation value, the continuation and exercise regions. For the considered American option, let us introduce the continuation value

$$C_i(x) = B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x\right),\ i = 0, ..., \mathcal{I}-1;\ C_{\mathcal{I}}(x) = f_{\mathcal{I}}(x), \tag{2.1}$$

the continuation region $\mathcal{C}$ and the exercise (stopping) region $\mathcal{E}$:

$$\begin{aligned} \mathcal{C} &= \{(t_i, x) : f_i(x) < C_i(x)\}, \\ \mathcal{E} &= \{(t_i, x) : f_i(x) \geq C_i(x)\}. \end{aligned} \tag{2.2}$$

Clearly, $(t_{\mathcal{I}}, x) \in \mathcal{E}$ for any x.

Let $X_j^{i,x}$, $j = i, i+1, ..., \mathcal{I}$, be the Markov chain starting at the step i from the point x : $X_i^{i,x} = x$, and ${}_mX_j^{i,x}$, $m = 1, ..., M$, be independent trajectories of the Markov chain. The Monte Carlo estimator $\hat{u}_i(x)$ for $u_i(x)$ (in the case when $\mathcal{E}$ is known) has the form

$$\hat{u}_i(x) = \frac{1}{M}\sum_{m=1}^{M}\frac{B_i}{B_\tau}f({}_mX_\tau^{i,x}), \tag{2.3}$$

where τ is the first time at which $X_j^{i,x}$ gets into $\mathcal{E}$ (of course, τ in (2.3) depends on i, x, and m : $\tau =_m \tau^{i,x}$). Thus, for estimating $u_i(x)$, it is sufficient to examine sequentially the position $(t_j,\ {}_mX_j^{i,x})$ for $j = i, i+1, ..., \mathcal{I}$ whether it belongs to $\mathcal{E}$ or not.

Let us give a simple sufficient condition for moving along the trajectory using a low bound v. Introduce the set

$$\mathcal{C}_v = \left\{(t_k, x) : f_k(x) < B_k E\left(\frac{v_{k+1}(X_{k+1})}{B_{k+1}}|X_k = x\right)\right\}.$$

Since $\mathcal{C}_v \subset \mathcal{C}$, the sufficient condition consists in fulfilment of the inclusion $(t_j,\ {}_mX_j^{i,x}) \in \mathcal{C}_v$.

Clearly, if $v_i^1, ..., v_i^l$ are some lower bounds, then the function $v_i(x) = \max_{1\leq k\leq l} v_i^k(x)$ is a lower bound as well. Besides, $f_i(x)$ is also a lower bound. Henceforth we consider lower bounds satisfying the inequality $v_i(x) \geq f_i(x)$.

2.2. **Equivalence of American options to European ones with consumption processes involved.** For $0 \leq i \leq \mathcal{I}-1$ equation (1.2) can be rewritten in the form

$$u_i(x) = B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right) + \left[f_i(x) - B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right)\right]^+. \tag{2.4}$$

Introduce the functions

$$\gamma_i(x) = \left[f_i(x) - B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right)\right]^+,\ i = \mathcal{I}-1, ..., 0. \tag{2.5}$$

Due to (2.4), we have

$$u_{\mathcal{I}-1}(X_{\mathcal{I}-1}) = B_{\mathcal{I}-1}E\left(\frac{f_{\mathcal{I}}(X_{\mathcal{I}})}{B_{\mathcal{I}}}|\mathcal{F}_{\mathcal{I}-1}\right) + \gamma_{\mathcal{I}-1}(X_{\mathcal{I}-1}),$$

$$u_{\mathcal{I}-2}(X_{\mathcal{I}-2}) = B_{\mathcal{I}-2}E\left(\frac{u_{\mathcal{I}-1}(X_{\mathcal{I}-1})}{B_{\mathcal{I}-1}}|\mathcal{F}_{\mathcal{I}-2}\right) + \gamma_{\mathcal{I}-2}(X_{\mathcal{I}-2})$$

$$= B_{\mathcal{I}-2}E\left(\frac{f_{\mathcal{I}}(X_{\mathcal{I}})}{B_{\mathcal{I}}}|\mathcal{F}_{\mathcal{I}-2}\right) + B_{\mathcal{I}-2}E\left(\frac{\gamma_{\mathcal{I}-1}(X_{\mathcal{I}-1})}{B_{\mathcal{I}-1}}|\mathcal{F}_{\mathcal{I}-2}\right) + \gamma_{\mathcal{I}-2}(X_{\mathcal{I}-2}).$$

Doing in just the same way further, we get

$$u_i(X_i) = B_i E\left(\frac{f_{\mathcal{I}}(X_{\mathcal{I}})}{B_{\mathcal{I}}}|\mathcal{F}_i\right) + B_i\sum_{k=1}^{\mathcal{I}-(i+1)} E\left(\frac{\gamma_{\mathcal{I}-k}(X_{\mathcal{I}-k})}{B_{\mathcal{I}-k}}|\mathcal{F}_i\right) + \gamma_i(X_i),\ i = 0, ..., \mathcal{I}-1. \tag{2.6}$$

Putting $X_0 = x$ and recalling that $B_0 = 1$, we obtain

$$u_0(x) = E\left(\frac{f_{\mathcal{I}}(X_{\mathcal{I}})}{B_{\mathcal{I}}}\right) + \gamma_0(x) + \sum_{i=1}^{\mathcal{I}-1} E\left(\frac{\gamma_i(X_i)}{B_i}\right). \tag{2.7}$$

Formula (2.7) gives the value of the European option with the payoff function $f_i(x)$ and with the consumption process γ_i defined by (2.5).

2.3. **Upper and low bounds using consumption processes.** The obtained result about equivalence of the discrete-time American option to the European option with the consumption process cannot be used directly because $u_i(x)$ and consequently $\gamma_i(x)$ are unknown. We take advantage of the discovered connection in the following way (see [3]).

Let $v_i(x)$ be a lower bound on the true option price $u_i(x)$. We introduce the function (upper consumption process)

$$\gamma_{i,v}(x) = \left[f_i(x) - B_i E\left(\frac{v_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right)\right]^+, \ i = 0, ..., \mathcal{I} - 1. \tag{2.8}$$

Clearly,

$$\gamma_{i,v}(x) \geq \gamma_i(x).$$

Hence the price $V_i(x)$ of the European option with the payoff function $f_i(x)$ and with the upper consumption process $\gamma_{i,v}(x)$ is an upper bound: $V_i(x) \geq u_i(x)$.

Conversely, if $V_i(x)$ is an upper bound on the true option price $u_i(x)$ and

$$\gamma_{i,V}(x) = \left[f_i(x) - B_i E\left(\frac{V_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right)\right]^+, \ i = 0, ..., \mathcal{I} - 1, \tag{2.9}$$

then

$$\gamma_{i,V}(x) \leq \gamma_i(x).$$

and the price $v_i(x)$ of the European option with the low consumption process $\gamma_{i,V}(x)$ is a lower bound: $v_i(x) \leq u_i(x)$.

Thus, starting from a lower bound $v_i^1(x)$, one can construct the upper bound $V_i^1(x)$ as the European option with the consumption process $\gamma_{i,v^1}(x)$ and so on. This procedure gives us the sequences $v_i^1(x) \leq v_i^2(x) \leq v_i^3(x) \leq ... \leq u_i(x)$, and $V_i^1(x) \geq V_i^2(x) \geq ... \geq u_i(x)$. All the bounds v^k and V^k can in principle be evaluated by the Monte Carlo simulations. However each further step of the procedure requires labor-consuming calculations and in practice it is possible to realize only a few steps of this procedure. In this connection, much attention is given to variance reduction technique and some constructive methods reducing statistical errors are proposed (see [3]).

2.4. **Comparison with the dual approach.** Without loss of generality we assume in this section that $B_i \equiv 1$. The dual approach, developed in [21] and [14] is based on the following observation. For any $0 \le i \le \mathcal{I}$ and any supermartingale $(S_j)_{i \le j \le \mathcal{I}}$ with $S_i = 0$ we have that

$$
\begin{aligned}
u_i(X_i) = \sup_{\tau \in \mathcal{T}_{i,\mathcal{I}}} E\left(f_\tau(X_\tau)|\mathcal{F}_i\right) &\le \sup_{\tau \in \mathcal{T}_{i,\mathcal{I}}} E\left(f_\tau(X_\tau) - S_\tau|\mathcal{F}_i\right) \\
&\le E\left[\max_{i \le j \le \mathcal{I}} \left(f_j(X_j) - S_j\right)|\mathcal{F}_i\right],
\end{aligned}
\tag{2.10}
$$

hence the right-hand side provides an upper bound for $u_i(X_i)$. It can be shown that the equality in (2.10) is attained at the martingale part of the Doob-Meyer decomposition of the price process u_i:

$$
M_i = 0, \quad M_j = \sum_{l=i+1}^{j} \left(u_l(X_l) - E\left(u_l(X_l)|\mathcal{F}_{l-1}\right)\right), \quad i < j \le \mathcal{I}.
$$

The duality representation provides a simple way to estimate the Snell envelope from above, using a lower approximation process $\{v_i(X_i)\}$. Let M^v be the martingale

$$
\begin{aligned}
M_0^v &= 0; \\
M_j^v &= M_{j-1}^v + v_j(X_j) - E\left(v_j(X_j)|\mathcal{F}_{j-1}\right) \\
&= \sum_{l=1}^{j} v_l(X_l) - \sum_{l=1}^{j} E\left(v_l(X_l)|\mathcal{F}_{l-1}\right), \quad 1 \le j \le \mathcal{I}.
\end{aligned}
\tag{2.11}
$$

Then, for any $0 \le i \le \mathcal{I}$ the process $\widetilde{M}_{ij} = M_j^v - M_i^v$, $j = i, \ldots, \mathcal{I}$, is a martingale with $\widetilde{M}_{ii} = 0$ and according to (2.10)

$$
V_i^D(X_i) := E\left[\max_{i \le j \le \mathcal{I}} \left(f_j(X_j) - \widetilde{M}_{ij}\right)|\mathcal{F}_i\right] \ge u_i(X_i).
$$

In particular, for $i = 0$

$$
\begin{aligned}
V_0^D(X_0) &= v_0(X_0) \\
&+ E\left[\max_{0 \le j \le \mathcal{I}} \left(f_j(X_j) - v_j(X_j) + \sum_{l=0}^{j-1} \left(E\left(v_{l+1}(X_{l+1})|\mathcal{F}_l\right) - v_l(X_l)\right)\right)\right].
\end{aligned}
\tag{2.12}
$$

The upper bound $V_0(X_0)$ obtained in section 2.3 can be transformed to

$$
\begin{aligned}
V_0(X_0) &= E\left(f_{\mathcal{I}}(X_{\mathcal{I}})\right) + E\sum_{i=0}^{\mathcal{I}-1} \left[f_i(X_i) - E\left(v_{i+1}(X_{i+1})|\mathcal{F}_i\right)\right]^+ \\
&= v_0(X_0) + E\sum_{i=0}^{\mathcal{I}-1} \left(\max\left\{f_i(X_i), E\left(v_{i+1}(X_{i+1})|\mathcal{F}_i\right)\right\} - v_i(X_i)\right),
\end{aligned}
\tag{2.13}
$$

where it is assumed that

$$
f_i(X_i) \le v_i(X_i), \quad i = 0, \ldots, \mathcal{I} - 1, \quad v_{\mathcal{I}}(X_{\mathcal{I}}) = f_{\mathcal{I}}(X_{\mathcal{I}}).
$$

It is interesting to compare V_0 and V_0^D starting from the same low bound v_i. A comprehensive comparison of $V_0(X_0)$ and $V_0^D(X_0)$ seems to be difficult and we restrict ourselves to some examples. First, we construct examples where $V_0(X_0) \le V_0^D(X_0)$. Let us define

$$\tau := \min\left\{0 \le i \le \mathcal{I}-1 : f_i(X_i) \ge E\left(v_{i+1}|\mathcal{F}_i\right)\right\},$$

and $\tau = \mathcal{I}$ if $f_i(X_i) < E\left(v_{i+1}|\mathcal{F}_i\right)$ for all i. We see that if $\tau = \mathcal{I}$ or

$$f_i(X_i) \ge E\left(v_{i+1}(X_{i+1})|\mathcal{F}_i\right), \quad i \ge \tau,$$

with probability 1, then

$$V_0(X_0) = v_0(X_0) + E\sum_{i=0}^{\tau-1}\left(E\left(v_{i+1}(X_{i+1})|\mathcal{F}_i\right) - v_i(X_i)\right)$$
$$+ E(f_\tau(X_\tau) - v_\tau(X_\tau)) + E\sum_{j=\tau+1}^{\mathcal{I}-1}(f_j(X_j) - v_j(X_j)) \le V_0^D(X_0).$$

The strict inequality $V_0 < V_0^D$ is achieved in the following simple example with $\mathcal{I} = 3$. Due to (2.12), the dual price at time 0 can be computed via the formula

$$\begin{aligned} V_0^D &= E\max\{f_0, f_1 - v_1 + Ev_1, \max\{f_2, E(u_3|\mathcal{F}_2)\} + Ev_1 + E(v_2|\mathcal{F}_1) - v_1 - v_2\} \\ &= E\max\{f_0, f_1 - v_1 + Ev_1, E(v_2|\mathcal{F}_1) + u_2 - v_2 - v_1 + Ev_1\} \\ &= E\max\{f_0, \max\{f_1, E(v_2|\mathcal{F}_1) + u_2 - v_2\} - v_1 + Ev_1\}, \end{aligned} \tag{2.14}$$

where we use the equality $u_2 = \max\{f_2, E(u_3|\mathcal{F}_2)\}$ and the dependence of quantities involved on the underlying process X_i is not shown explicitly for the sake of simplicity. Formula (2.13) gives

$$\begin{aligned} V_0 = E\max\{f_0, Ev_1\} &+ E(\max\{f_1, E(v_2|\mathcal{F}_1)\} - v_1) \\ &+ E(\max\{f_2, E(v_3|\mathcal{F}_2)\} - v_2). \end{aligned} \tag{2.15}$$

Let us take constant payoffs satisfying

$$f_0 < f_1 < f_2 < f_3, \quad f_1 + f_2 < f_0 + f_3.$$

Clearly, $u_i = f_3$, $i = 0, \ldots, 3$ and any low bound v_i satisfies

$$f_0 \le v_0 \le f_3, \quad f_1 \le v_1 \le f_3, \quad f_2 \le v_2 \le f_3, \quad v_3 = f_3.$$

Formula (2.15) gives $V_0 = f_3$ and (2.14) implies

$$V_0^D = E\max\{f_0, E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1\}.$$

Clearly,

$$V_0^D \ge E[E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1] = f_3.$$

If v_1 and v_2 are such that the inequality

$$E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1 \ge f_0$$

is fulfilled with probability 1, then $V_0^D = f_3$. However, if

$$E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1 < f_0 \tag{2.16}$$

with positive probability, then

$$\max\{f_0, E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1\} > E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1$$

with the same probability and consequently $V_0^D > V_0$. The inequality (2.16) is achieved, for example, if Ev_1 is close to f_1, $E(v_2|\mathcal{F}_1)$ is close to f_2 and v_1 and v_2 are equal to f_3 with positive probability.
At the same time it is possible to construct examples when $V_0^D \leq V_0$. Indeed, let us take $v_i(X_i) = f_i(X_i)$ for all $i = 0, \ldots, \mathcal{I} - 1$, then according to (2.12)

$$V_0^D = f_0 + E\left[\max_{0\leq j\leq \mathcal{I}} \sum_{l=0}^{j-1} \left(E\left(f_{l+1}|\mathcal{F}_l\right) - f_l\right)\right]$$

and due to (2.13)

$$V_0 = f_0 + \sum_{i=0}^{\mathcal{I}-1} \left(E\left(f_{i+1}|\mathcal{F}_i\right) - f_i\right)^+ \geq V_0^D.$$

However, the method based on the representation (2.6) has some advantages over dual approach. First, $V_0(X_0)$ depends on v_i monotonically that is if we have two low bounds v and $\tilde{v}$ such that $v_i(X_i) \leq \tilde{v}_i(X_i)$ for all i, then $V_0(X_0) \geq \tilde{V}_0(X_0)$. This immediately follows from the first line in (2.13). For the dual method this is not always the case. Indeed, with three exercises ($\mathcal{I} = 2$) formula (2.12) gives

$$V_0^D = E\max\{f_0, E(v_1|\mathcal{F}_0) + u_1 - v_1\}.$$

Consider the case when the probability of event $A := \{Ev_1 - u_1 - v_1 \geq f_0\}$ is positive and $v_1 < u_1 - \theta$ with some constant $\theta > 0$. Then taking $\tilde{v}_1 = v_1 + \theta/2$ on A and $\tilde{v}_1 = v_1 + \theta$ outside A we obtain

$$\tilde{V}_0^D := E\max\{f_0, E(\tilde{v}_1|\mathcal{F}_0) + u_1 - \tilde{v}_1\} > V_0^D,$$

though $\tilde{v}_1 > v_1$. Second, adaptive local low bounds of the form

$$v_i(x) = \max_{1\leq k\leq l} v_i^k(x), \quad i = 0, \ldots, \mathcal{I} - 1,$$

where $v_1(x), \ldots, v_l(x)$ are low bounds at x ordered according to their complexity and l may depend on x, can be used to construct $V_0(X_0)$ (see [4]). Third, $V_0(X_0)$ is computationally less expensive than $V_0(X_0)$. It is also worthwhile mentioning that our approach allows us to construct low bounds using upper ones.

2.5. **Bermudan options.** As before we consider the discrete-time model

$$(B_i, X_i) = (B_i, X_i^1, ..., X_i^d), \quad i = 0, 1, ..., \mathcal{I}.$$

However, now an investor can exercise his right only at time belonging to the set of stopping times $S = \{s_1, ..., s_l\}$ within $\{0, 1, ..., \mathcal{I}\}$ where $s_l = \mathcal{I}$. The price $u_i(X_i)$ of the Bermudan option is given by

$$u_i(X_i) = \sup_{\tau\in\mathcal{T}_{S\cap[i,\mathcal{I}]}} B_i E\left(\frac{f_\tau(X_\tau)}{B_\tau}|\mathcal{F}_i\right),$$

where $\mathcal{T}_{S\cap[i,\mathcal{I}]}$ is the set of stopping times τ taking values in $\{s_1, ..., s_l\} \cap \{i, i+1, ..., \mathcal{I}\}$.

The value process u_i is determined as follows:

$$u_{\mathcal{I}}(x) = f_{\mathcal{I}}(x),$$
$$u_i(x) = \begin{cases} \max\left\{ f_i(x),\ B_i E\left(\dfrac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x\right)\right\},\ i \in S, \\ B_i E\left(\dfrac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x\right),\ i \notin S. \end{cases}$$

Thus, we obtain that the Bermudan option is equivalent to the European option with the payoff function $f_i(x)$ and with the consumption process γ_i defined by

$$\gamma_i(x) = \begin{cases} \left[f_i(x) - B_i E\left(\dfrac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x\right)\right]^+,\ i \in S, \\ 0,\ i \notin S. \end{cases}$$

From here all the results for discrete-time American options obtained in this section can be carried over to the Bermudan options. For example, if $v_i(x)$ is a lower bound of the true option price $u_i(x)$, the price $V_i(x)$ of the European option with the payoff function $f_{\mathcal{I}}(x)$ and with the consumption process

$$\gamma_{i,v}(x) = \begin{cases} \left[f_i(x) - B_i E\left(\dfrac{v_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x\right)\right]^+,\ i \in S, \\ 0,\ i \notin S, \end{cases}$$

is an upper bound: $V_i(x) \geq u_i(x)$.

3. Optimal stopping times and algorithms with low continuation values

The samples with optimal stopping times are introduced first in [19] (see [9] as well).

3.1. Basic relations for optimal stopping times.

The optimal stopping time $\tau^{i,x} = \tau^{t_i,x}$ depends on the initial position (t_i, x). It is defined recurrently by the dynamic programming principle in the following way. We set

$$\begin{aligned} \tau^{\mathcal{I},x} &= \tau^{T,x} = T, \\ \tau^{i,x} &= t_i \chi_{\{C_i(x) \leq f_i(x)\}} + \tau^{i+1, X^{i,x}_{i+1}} \chi_{\{C_i(x) > f_i(x)\}} \\ &= t_i \chi_{\{u_i(x) = f_i(x)\}} + \tau^{i+1, X^{i,x}_{i+1}} \chi_{\{u_i(x) > f_i(x)\}}, \\ i &= \mathcal{I} - 1, ..., 0. \end{aligned} \tag{3.1}$$

Thus, for any position (t_i, x), the optimal stopping time $\tau^{i,x}$ is either equal to t_i : $\tau^{i,x} = t_i$, or $\tau^{i,x} > t_i$. It is also clear that (t_i, x) is a stopping point (i.e., $\tau^{i,x} = t_i$) iff $(t_i, x) \in \mathcal{E}$ (i.e., (t_i, x) belongs to the exercise region). The instant $\tau^{i,x}$ is the first one at which the trajectory $(t_j, X^{i,x}_j)$ either gets into $\mathcal{E}$ during $i \leq j \leq \mathcal{I} - 1$ or $\tau^{i,x} = \mathcal{I}$. So, $(\tau^{i,x}, X^{i,x}_{\tau^{i,x}}) \in \mathcal{E}$ (see (2.2). Let us give some recurrence relations for

$u_i(x)$ and $C_i(x)$:

$$(3.2a) \qquad u_i(X_i) = \max\{f_i(X_i), C_i(X_i)\}, \ u_{\mathcal{I}}(x) = f(x),$$

$$(3.2b) \qquad C_i(X_i) = \frac{B_i}{B_{i+1}} E(u_{i+1}(X_{i+1})|X_i), \ C_{\mathcal{I}}(x) = f(x),$$

$$(3.2c) \qquad C_i(X_i) = \frac{B_i}{B_{i+1}} E(\max\{f_{i+1}(X_{i+1}), C_{i+1}(X_{i+1})\}|X_i),$$

$$(3.2d) \qquad u_i(X_i) = \max\{f_i(X_i), \frac{B_i}{B_{i+1}} E(u_{i+1}(X_{i+1})|X_i)\}.$$

We note that

$$(3.3) \qquad u_{i+1}(X_{i+1}) = B_{i+1} E\left(\frac{f_\tau(X_\tau^{t_{i+1},X_{i+1}})}{B_\tau}|X_{i+1}\right),$$

$$(3.4) \qquad E(u_{i+1}(X_{i+1})|X_i) = E\left(B_{i+1} E\left(\frac{f_\tau(X_\tau^{t_{i+1},X_{i+1}})}{B_\tau}|\mathcal{F}_{i+1}\right)|\mathcal{F}_i\right) = B_{i+1} E\left(\frac{f_\tau(X_\tau^{t_{i+1},X_{i+1}})}{B_\tau}|X_i\right),$$

where

$$\tau = \tau^{t_{i+1},X_{i+1}}.$$

Hence due to (3.2b),

$$(3.5) \qquad C_i(X_i) = B_i E\left(\frac{f_\tau(X_\tau^{t_{i+1},X_{i+1}})}{B_\tau}|X_i\right).$$

We emphasize that for any stopping time $\tilde{\tau} \geq t_{i+1}$ the function

$$(3.6) \qquad v_{i+1}(x) = B_{i+1} E\left(\frac{f_{\tilde{\tau}}(X_{\tilde{\tau}}^{t_{i+1},x})}{B_{\tilde{\tau}}}\right)$$

is a low bound for $u_{i+1}(x)$.
Since

$$(3.7) \quad C_i(x) = \sup_{\tau \in \mathcal{T}_{i+1,\mathcal{I}}} B_i E\left(\frac{f_\tau(X_\tau^{t_{i+1},X_{i+1}})}{B_\tau}|X_i = x\right) = \sup_{\tau \in \mathcal{T}_{i+1,\mathcal{I}}} B_i E\left(\frac{f_\tau(X_\tau^{t_i,x})}{B_\tau}\right),$$

the function

$$(3.8) \qquad c_i(x) = B_i E\left(\frac{f_{\tilde{\tau}}(X_{\tilde{\tau}}^{t_i,x})}{B_{\tilde{\tau}}}\right)$$

is a low continuation value for any stopping time $\tilde{\tau} \geq t_{i+1}$.

3.2. **Subsequent estimating optimal stopping times.** Considering $C_i(x)$ as a regression function (see (3.5)), it is natural to introduce after [19] and [9] the sample

$$(3.9)\qquad ({}_mX_i,\ \frac{B_i}{B_\tau}f_\tau({}_mX_\tau^{t_{i+1},\ {}_mX_{i+1}})) = ({}_mX_i,\ \frac{B_i}{B_\tau}f_\tau({}_mX_\tau^{t_i,\ {}_mX_i})),$$
$$\tau = \tau^{t_{i+1},\ {}_mX_{i+1}},\ m = 1, ..., M,$$

from $(X_i,\ \frac{B_i}{B_\tau}f_\tau(X_\tau^{t_{i+1},X_{i+1}})) = (X_i,\ \frac{B_i}{B_\tau}f_\tau(X_\tau^{t_i,X_i}))$, where $\tau = \tau^{t_{i+1},X_{i+1}}$.

We are about to use (3.10) for subsequent constructing an estimate $\hat{\tau}^{t_i,\ {}_mX_i}$ for optimal stopping time $\tau^{t_i,\ {}_mX_i}$. Clearly, $\tau^{\mathcal{I},\ {}_mX_\mathcal{I}} = \hat{\tau}^{\mathcal{I},\ {}_mX_\mathcal{I}} = \mathcal{I}$. Let $\tau^{t_{i+1},\ {}_mX_{i+1}}$, $i = \mathcal{I}-1, ..., 1$, (in reality $\hat{\tau}^{t_{i+1},\ {}_mX_{i+1}}$) be known. Using the sample (3.10) at the step t_i, we evaluate $C_i({}_mX_i)$ as a regression due to (3.5). Let $\hat{C}_i({}_mX_i)$ be an estimate of $C_i({}_mX_i)$ (we recall that knowledge of $\hat{C}_i({}_mX_i)$ gives $\hat{u}_i({}_mX_i)$ due to (3.2a)). If $f_i({}_mX_i) \geq \hat{C}_i({}_mX_i)$ then $\hat{\tau}^{t_i,\ {}_mX_i} = t_i$, otherwise $\hat{\tau}^{t_i,\ {}_mX_i} = \hat{\tau}^{t_{i+1},\ {}_mX_{i+1}}$ (see (3.1)). As a result we obtain the sample like (3.10) at the step t_{i-1}:

$$(3.10)\quad ({}_mX_{i-1},\ \frac{B_{i-1}}{B_\tau}f_\tau({}_mX_\tau^{t_i,\ {}_mX_i})) = ({}_mX_{i-1},\ \frac{B_{i-1}}{B_\tau}f_\tau({}_mX_\tau^{t_{i-1},\ {}_mX_{i-1}})),$$
$$\tau = \tau^{t_i,\ {}_mX_i},\ m = 1, ..., M.$$

Coming to $\tau^{t_1,\ {}_mX_1}$, we can evaluate $u_0(X_0)$. Indeed, since X_0 is a nonrandom vector, we have (see (3.2d) and (3.4)

(3.11)

$$u_0(X_0) = \max\{f_0(X_0), \frac{1}{B_1}E(u_1(X_1^{t_0,X_0})\} = \max\left\{f_0(X_0), E\left(\frac{f_\tau(X_\tau^{t_1,X_1})}{B_\tau}\right)\right\},\ \tau = \tau^{t_1,X_1}.$$

So, our main problem is to evaluate the continuation value $C_i({}_mX_i)$ using sample (3.10). There are a lot of nonparametric regression methods to attain this objective (see, e.g., [13]). In the next subsection we propose some algorithms basing both on local modelling and least squares estimation. In contrast to other works using the regression approach in pricing American options, we construct together with the estimate $\hat{C}_i({}_mX_i)$ an upper consumption process.

The most appropriate are methods for which the estimate $\hat{C}_i({}_mX_i)$ is a low continuation value. Then we are able to construct both a low and an upper bounds.

3.3. **Algorithms with the local Monte Carlo approach.** For every position $(t_i,\ {}_mX_i)$, $m = 1, ...M$, let us construct $N = N_{i,m}$ additional independent trajectories on $[t_i, t_{i+1}]$, i.e., the trajectories with the length of one step. To the instant t_{i+1} we obtain $N+1$ points ${}_nX_{t_{i+1}}^{t_i,\ {}_mX_i}$, $n = 0, 1, ..., N$, where we put ${}_0X_{t_{i+1}}^{t_i,\ {}_mX_i} =_m X_{i+1}$. Introduce the notation ${}_{m,n}X_{i+1} :=_n X_{t_{i+1}}^{t_i,\ {}_mX_i}$. Let $\tau_{m,n} := \tau^{t_{i+1},\ {}_{m,n}X_{i+1}}$. Due to (3.5) and the Monte Carlo approach (let us note that $\tau_{m,n} = \tau^{t_{i+1},\ {}_{m,n}X_{i+1}}$ is equal to $\tau^{t_i,\ {}_mX_i}$ provided $\tau^{t_i,\ {}_mX_i} \geq t_{i+1}$; see also (3.7)) , we have

(3.12)

$$C_i({}_mX_i) = B_iE\left(\frac{f_\tau(X_\tau^{t_{i+1},X_{i+1}})}{B_\tau}|X_i =_m X_i\right) \simeq \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\tau_{m,n}}(X_{\tau_{m,n}}^{t_{i+1},\ {}_{m,n}X_{i+1}})}{B_{\tau_{m,n}}}.$$

For every point ${}_{m,n}X_{i+1} =_n X_{t_{i+1}}^{t_i,\ {}_mX_i}$ we find the nearest one among ${}_kX_{i+1}$, $k = 1, ...M$, let it be ${}_{k(m,n)}X_{i+1}$. For the position $(t_{i+1},\ {}_{k(m,n)}X_{i+1})$, it is known the estimate $\hat{\tau}_{k(m,n)}$ of the optimal stopping time $\tau^{t_{i+1},\ {}_{k(m,n)}X_{i+1}}$. To avoid confusion, let us emphasize that the points ${}_{m,n}X_{i+1}$ lie on the trajectories starting from the same position $(t_i,\ {}_mX_i)$ while the points ${}_{k(m,n)}X_{i+1}$ lie on the trajectories which have different starting positions $(t_i,\ {}_{k(m,n)}X_i)$. For any point $X_{i+1} = X_{t_{i+1}}^{t_i,\ {}_mX_i}$ one can define the stopping $\tilde{\tau} = \tilde{\tau}(X_{i+1}) \geq t_{i+1}$ analogously to $\hat{\tau}_{k(m,n)}$, i.e., first, you find the nearest point to X_{i+1} among ${}_kX_{i+1}$, $k = 1, ...M$, say ${}_{\tilde{k}}X_{i+1}$, and second, for the position $(t_{i+1},\ {}_{\tilde{k}}X_{i+1})$ you know the estimate $\hat{\tau}_{\tilde{k}}$ of the optimal stopping time $\tau^{t_{i+1},\ {}_{\tilde{k}}X_{i+1}}$ which you take as $\tilde{\tau} : \tilde{\tau} = \tilde{\tau}(X_{i+1}) = \hat{\tau}_{\tilde{k}}$. Clearly, for the points ${}_{m,n}X_{i+1}$ this stopping time $\tilde{\tau} = \tilde{\tau}({}_{m,n}X_{i+1}) := \tilde{\tau}_{m,n}$ coincides with $\hat{\tau}_{k(m,n)}$. Introduce

$$\tilde{C}_i(x) = B_i E\left(\frac{f_{\tilde{\tau}}(X_{\tilde{\tau}}^{t_{i+1},X_{i+1}})}{B_{\tilde{\tau}}} | X_i = x\right).$$

From (3.7) and (3.8) it follows

$$C_i(x) = \tilde{C}_i(x) + r_i(x), \tag{3.13}$$

where $r_i(x) \geq 0$, i.e. $\tilde{C}_i(x)$ is a low continuation value at the position $(t_i,\ x)$. Analogously to (3.12) we have

$$\begin{aligned} \tilde{C}_i({}_mX_i) &= \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\tilde{\tau}_{m,n}}(X_{\tilde{\tau}_{m,n}}^{t_{i+1},\ {}_{m,n}X_{i+1}})}{B_{\tilde{\tau}_{m,n}}} + \alpha_i({}_mX_i) \\ &= \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\hat{\tau}_{k(m,n)}}(X_{\hat{\tau}_{k(m,n)}}^{t_{i+1},\ {}_{m,n}X_{i+1}})}{B_{\hat{\tau}_{k(m,n)}}} + \alpha_i({}_mX_i), \end{aligned} \tag{3.14}$$

where $\alpha_i({}_mX_i)$ is the Monte Carlo error which becomes small with increasing N. Let us pay attention that in general the points $X_{\tilde{\tau}_{m,n}}^{t_{i+1},\ {}_{m,n}X_{i+1}}$ do not belong to the considered sample of M independent trajectories all starting from the initial point (t_0, X_0). That is why the sum in (3.14) cannot be taken as an estimate for the continuation value $C_i({}_mX_i)$.

For the continuation value, it is natural to introduce the estimate

$$\hat{C}_i({}_mX_i) = \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\hat{\tau}_{k(m,n)}}(X_{\hat{\tau}_{k(m,n)}}^{t_{i+1},\ {}_{k(m,n)}X_{i+1}})}{B_{\hat{\tau}_{k(m,n)}}}. \tag{3.15}$$

Let us note that in (3.15) and in (3.14) we consider the trajectories $X_s^{t_{i+1},\ {}_{k(m,n)}X_{i+1}}$ and $X_s^{t_{i+1},\ {}_{m,n}X_{i+1}}$ starting from different positions $(t_{i+1},\ {}_{k(m,n)}X_{i+1})$ and $(t_{i+1},\ {}_{m,n}X_{i+1})$ but with the same sources of randomness. If M is large, the points ${}_{m,n}X_{i+1}$ and ${}_{k(m,n)}X_{i+1}$ are at a short distance and we get

$$\begin{aligned} \hat{C}_i({}_mX_i) &= \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\hat{\tau}_{k(m,n)}}(X_{\hat{\tau}_{k(m,n)}}^{t_{i+1},\ {}_{m,n}X_{i+1}})}{B_{\hat{\tau}_{k(m,n)}}} - \beta_i({}_mX_i) \\ &= \tilde{C}_i({}_mX_i) - \alpha_i({}_mX_i) - \beta_i({}_mX_i), \end{aligned} \tag{3.16}$$

where the approximation error β_i is small.

From (3.13) we obtain

$$\hat{C}_i({}_mX_i) = C_i({}_mX_i) + \rho_i({}_mX_i) - r_i({}_mX_i), \tag{3.17}$$

where $\rho_i = -\alpha_i - \beta_i$.

We can claim that the estimate $\hat{C}_i({}_mX_i)$ is a low continuation value at the position $(t_i,\ {}_mX_i)$ within the accuracy depending on N and M, because ρ_i becomes small with increasing M and N and $r_i \geq 0$. It should be noted that r_i essentially depends on a procedure of subsequent estimating optimal stopping times and can be comparatively large (i.e. $r_i \gg 0$) if the procedure is unsuccessful. Thus the following theorem is justified.

Theorem 3.1. *The estimate $\hat{C}_i({}_mX_i)$ is a low continuation value within the accuracy depending on N (the accuracy determined by the Monte Carlo error) and M (the accuracy determined by the approximation error).*

Corollary 3.2. *Consider the consumption*

$$\hat{\gamma}_i({}_mX_i) = [f_i({}_mX_i) - \hat{C}_i({}_mX_i)]^+. \tag{3.18}$$

Because $\hat{\gamma}_i({}_mX_i) = [f_i({}_mX_i) - C_i({}_mX_i) + r_i({}_mX_i) - \rho_i({}_mX_i)]^+$, we have

$$\gamma_i({}_mX_i) \leq \hat{\gamma}_i({}_mX_i),\ \text{ if } r_i \geq \rho_i, \tag{3.19}$$

$$[\gamma_i({}_mX_i) - \rho_i({}_mX_i) + r_i({}_mX_i)]^+ \leq \hat{\gamma}_i({}_mX_i) \leq \gamma_i({}_mX_i),\ \text{ if } \rho_i > r_i.$$

We see that $\hat{\gamma}_i({}_mX_i)$ is an upper consumption in the most typical case $r_i \geq \rho_i$, otherwise it can be not an upper bound however in such a case $\hat{\gamma}_i({}_mX_i)$ is insignificantly distinguished from $\gamma_i({}_mX_i)$, i.e., $\hat{\gamma}_i({}_mX_i)$ is an upper consumption within the accuracy depending on M and N.

3.4. **Algorithms with the local Monte Carlo approach, continuation.** For the estimate (3.15) we use one nearest point ${}_{k(m,n)}X_{i+1}$ among ${}_mX_{i+1},\ m = 1, \ldots, M$, to every point ${}_{m,n}X_{i+1}$. Now let us for every point ${}_{m,n}X_{i+1} =_n X_{t_{i+1}}^{t_i,\ {}_mX_i}$ find a few (say $K_{m,n}$) nearest ones among ${}_mX_{i+1},\ m = 1, \ldots M$. Let us denote them by ${}_{k[m,n]}X_{i+1},\ k = 1, \ldots, K_{m,n}$ (in contrast to $k(m,n)$, the function $k[m,n]$ is a multifunction). The estimates $\hat{\tau}_{k[m,n]}$ of the optimal stopping times $\tau_{k[m,n]} := \tau^{t_{i+1},\ {}_{k[m,n]}X_{i+1}}$ are known. Then the following expression

$$v_{i+1}({}_nX_{t_{i+1}}^{t_i,\ {}_mX_i}) = \frac{B_{i+1}}{K_{m,n}} \sum_{k=1}^{K_{m,n}} \frac{f(X_{\hat{\tau}_{k[m,n]}}^{t_{i+1},\ {}_{k[m,n]}X_{i+1}})}{B_{\hat{\tau}_{k[m,n]}}} \tag{3.20}$$

is a low bound for $u_{i+1}(x)$ at the position $(t_{i+1},\ {}_nX_{t_{i+1}}^{t_i,\ {}_mX_i})$ (of course, within the accuracy of approximation).

Clearly,
(3.21)

$$\hat{C}_i({}_mX_i) = \frac{B_i}{B_{i+1}} \cdot \frac{1}{N+1} \sum_{n=0}^{N} v_{i+1}({}_nX_{t_{i+1}}^{t_i,\ {}_mX_i}) = \frac{B_i}{N+1} \sum_{n=0}^{N} \frac{1}{K_{m,n}} \sum_{k=1}^{K_{m,n}} \frac{f(X_{\hat{\tau}_{k[m,n]}}^{t_{i+1},\ {}_{k[m,n]}X_{i+1}})}{B_{\hat{\tau}_{k[m,n]}}}$$

is a low continuation value at $(t_i,\ {}_mX_i)$ (of course, within the accuracy depending on M and N).

The estimate (3.15) is the particular case of (3.21) when $K_{m,n} = 1$.

Remark 3.3. *For estimate* (3.21), *analogs* of *Theorem 3.1 and Corollary 3.2 are true as well.*

3.5. **Algorithms with k-NN estimates.** In the previous algorithms we construct $N_{i,m}$ additional trajectories for every point ${}_mX_i$, $m = 1,...M$. Let us consider $N = N_{i,m}$ nearest points ${}_{m,1}X_i, ..., {}_{m,N}X_i$ to the point ${}_mX_i$ instead of constructing the additional trajectories. All the points ${}_{m,1}X_i, ..., {}_{m,N}X_i$ belong to the set $\{ {}_mX_i,\ m = 1,...M\}$. We have ${}_{m,n}X_{i+1}^{(t_i,\ {}_{m,n}X_i)} = {}_{m,n}\ X_{i+1}$, $n = 0, 1, ..., N$, ${}_{m,0}X_i = {}_m X_i$, ${}_{m,0}X_{i+1} = {}_m\ X_{i+1}$, with known $\hat{\tau}_{m,n} = \hat{\tau}^{t_{i+1},\ {}_{m,n}X_{i+1}}$ and $f(X_{\hat{\tau}^{t_{i+1},\ {}_{m,n}X_{i+1}}}^{(t_{i+1},\ {}_{m,n}X_{i+1})})$ (let us note that we use another notation in this subsection and, in particular, we emphasize that the points ${}_{m,n}X_{i+1}$ belong to the set $\{ {}_mX_{i+1},\ m = 1,...M\}$). Then analogously to (3.15), we evaluate:

$$\hat{C}_i({}_mX_i) = \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\hat{\tau}_{m,n}}(X_{\hat{\tau}_{m,n}}^{t_{i+1},\ {}_{m,n}X_{i+1}})}{B_{\hat{\tau}_{m,n}}}. \tag{3.22}$$

This estimate is an analog of (3.15). To get an analog of (3.21) let us find for every point ${}_{m,n}X_{i+1} = {}_{m,n}\ X_{i+1}^{(t_i,\ {}_{m,n}X_i)}$ a few (say $K_{m,n}$) nearest ones among ${}_mX_{i+1}$, $m = 1, ...M$. Denote them by ${}_{m,n,k}X_{i+1}$, $k = 1, ..., K_{m,n}$. Then

$$\hat{C}_i({}_mX_i) = B_i \cdot \frac{1}{N+1}\sum_{n=0}^{N}\frac{1}{K_{m,n}}\sum_{k=1}^{K_{m,n}}\frac{f(X_{\hat{\tau}_{m,n,k}}^{t_{i+1},\ {}_{m,n,k}X_{i+1}})}{B_{\hat{\tau}_{m,n.k}}}, \tag{3.23}$$

where $\hat{\tau}_{m,n,k}$ are known estimates of the optimal stopping times $\tau_{m,n,k} := \tau^{t_{i+1},\ {}_{m,n,k}X_{i+1}}$. We note that ${}_{m,n,k}X_{i+1}$ in (3.23) are distinguished from ${}_{m,n,k}X_{i+1}$ in (3.21).

Remark 3.4. *For estimate* (3.23) *analogs* of *Theorem 3.1 and Corollary 3.2 are true as well.*

Remark 3.5. k-NN *estimates belong to the class of local averaging estimates (see* [13]*). One can use other estimates of this class, for example, kernel estimates and local polynomial kernel estimates. Note, that the latter type of estimates can be helpful for estimating deltas (see* (6.8) *and* (6.9)*).*

3.6. **Linear regression.** Regression-based methods approximate the continuation value using a basis function expansion:

$$C_i(x) \approx \sum_{r=1}^{K}\beta_{ir}\psi_r(x), \quad i = 0, 1, \ldots, \mathcal{I}-1,$$

where $\{\psi_r(x)\}_{r=1}^{K}$ is a set of basis functions each mapping $\mathbf{X}$ to $\mathbf{R}$. In the notations

$$C_i(x) \approx \beta_i^{\top}\psi(x)$$

with

$$\beta_i^\top = (\beta_{i1}, \dots, \beta_{iK}), \quad \psi(x) = (\psi_1(x), \dots, \psi_K(x))^\top.$$

Vector β_i can be estimated using the sample

$$({}_mX_i,\ \frac{B_i}{B_{\hat\tau_m}} f_{\hat\tau_m}({}_mX_{\hat\tau_m}^{t_{i+1},\ {}_mX_{i+1}})), \quad \hat\tau_m = \hat\tau^{t_{i+1},\ {}_mX_{i+1}}, \quad m = 1, \dots, M,$$

as

$$\hat\beta_i = \hat A_\psi^{-1} \hat\alpha_{\psi V}.$$

Here $\hat A_\psi$ is the $K \times K$ matrix with qr entry

$$\frac{1}{M} \sum_{m=1}^M \psi_q({}_mX_i)\psi_r({}_mX_i)$$

and $\hat\alpha_{\psi V}$ is the K-vector with rth entry

$$\frac{1}{M} \sum_{m=1}^M \psi_r({}_mX_i) \frac{B_i f_{\hat\tau_m}(X_{\hat\tau_m}^{t_{i+1},\ {}_mX_{i+1}})}{B_{\hat\tau_m}}.$$

The estimate $\hat\beta_i$ then defines an estimate

$$\hat C_i(x) = \hat\beta_i^\top \psi(x)$$

of the continuation value at an arbitrary point x in the state space $\mathbf{X}$. Now, if $f_i({}_mX_i) \geq \hat C_i({}_mX_i)$ then $\hat\tau^{t_i,\ {}_mX_i} = t_i$, otherwise $\hat\tau^{t_i,\ {}_mX_i} = \hat\tau^{t_{i+1},\ {}_mX_{i+1}}$ (see (3.1)). As a result we obtain at the step t_{i-1} the sample :

$$\begin{aligned}({}_mX_{i-1},\ \frac{B_{i-1}}{B_{\hat\tau_m}} f_{\hat\tau_m}({}_mX_{\hat\tau_m}^{t_i,\ {}_mX_i})) \quad &= \quad ({}_mX_{i-1},\ \frac{B_{i-1}}{B_{\hat\tau_m}} f_{\hat\tau_m}({}_mX_{\hat\tau_m}^{t_{i-1},\ {}_mX_{i-1}})),\\ &\qquad \hat\tau_m = \hat\tau^{t_i,\ {}_mX_i},\ m = 1, ..., M.\end{aligned}$$

Theorem 3.6. *The estimate*

$$\hat C_i({}_mX_i) = B_i \hat\beta_i^\top \psi({}_mX_i) \tag{3.24}$$

is a low continuation value within the accuracy depending on K and M.

Proof. Having $\hat C_j(x),\ x \in \mathbf{X},\ j = 0, ..., \mathcal{I} - 1$, one can define a stopping time $\tilde\tau$ for every trajectory $X_{t_j}^{t_i,\ x},\ j = i, ..., \mathcal{I}$, in the following way. If $\hat C_i(x) \leq f_i(x)$, then we put $\hat\tau^{t_i,\ x} = t_i$. If $\hat C_i(x) > f_i(x)$, then we put $\hat\tau^{t_i,\ x} > t_i$. Further, if $\hat C_{i+1}(X_{t_{i+1}}^{t_i,\ x}) \leq f_{i+1}(X_{t_{i+1}}^{t_i,\ x})$, then we put $\hat\tau^{t_i,\ x} = t_{i+1}$, and so on. If $\hat C_j(X_{t_j}^{t_i,\ x}) > f_j(X_{t_j}^{t_i,\ x})$ for all $j = i, ..., \mathcal{I} - 1$, then we put $\hat\tau^{t_i,\ x} = \mathcal{I}$. Clearly, $\tilde\tau^{t_i,\ {}_mX_i} = \hat\tau^{t_i,\ {}_mX_i}$, $m = 1, ..., M$, i.e., $\tilde\tau$ is an extension of $\hat\tau$. Let us introduce the value

$$\tilde C_i(x) = B_i E\left(\frac{f_{\tilde\tau}(X_{\tilde\tau}^{t_{i+1}, X_{i+1}})}{B_{\tilde\tau}} | X_i = x\right),\ \tilde\tau = \tilde\tau^{t_{i+1},\ X_{i+1}}. \tag{3.25}$$

Due to (3.7) and (3.8), $\tilde C_i(x)$ is a low continuation value, i.e.,

$$\tilde C_i(x) = C_i(x) - r_i(x), \tag{3.26}$$

where $r_i(x) \geq 0$. But for the conditional expectation (3.25), $\hat{C}_i(x)$ can be considered as an estimate by the linear regression method. Therefore

$$\tilde{C}_i(x) = \hat{C}_i(x) + \alpha_i(x), \tag{3.27}$$

where $\alpha_i(x)$ is the regression error which depends on K and M. From (3.26) and (3.27) we obtain

$$\hat{C}_i({}_mX_i) = C_i({}_mX_i) - \alpha_i({}_mX_i) - r_i({}_mX_i). \tag{3.28}$$

Theorem 3.6 is proved.

Remark 3.7. *Formally, the theorem is true even if the error $\alpha_i(x)$ is large. But its significance manifests itself when $\alpha_i(x)$ is rather small (this can be reached due to successful choice of $\psi_1(x), \ldots, \psi_K(x)$ and sufficiently large M). Then $\hat{C}_i({}_mX_i)$ is really (not only within the accuracy depending on K and M) a low continuation value.*

4. Global low and upper bounds

Aiming to estimate the price of the American option at a fixed position (t_0, x_0), we simulate the independent trajectories ${}_mX_i,\ i = 1, ..., \mathcal{I},\ m = 1, ..., M$, of the process X_i, starting at the instant $t = t_0$ from $x_0 :\ X_0 = x_0$.

For constructing the global low bound we use formula (3.11). Indeed (3.11) gives the following estimate

$$\hat{u}_0(X_0) = \max\left\{ f_0(X_0), \frac{1}{M}\sum_{m=1}^{M} \frac{f_{\hat{\tau}_m}(X_{\hat{\tau}_m}^{t_1,\ {}_mX_1})}{B_{\hat{\tau}_m}} \right\},\ \hat{\tau}_m = \hat{\tau}^{t_1,\ {}_mX_1}. \tag{4.1}$$

We note that (4.1) always is a low bound for $u_0(X_0)$ even if $\hat{\tau}_m$ is not equal to optimal stopping time $\tau^{t_1,\ {}_mX_1}$.

To construct the global upper bound we use Subsection 2.3. Let $v_i(x)$ be a low bound and $(t_i,\ {}_mX_i)$ be the position on the m-th trajectory at the time instant t_i. We calculate the low continuation value

$$c_{i,v}({}_mX_i) = B_i E\left(\frac{v_{i+1}({}_mX_{i+1})}{B_{i+1}} | \mathcal{F}_i\right) \tag{4.2}$$

at the position $(t_i,\ {}_mX_i)$. If

$$f_i({}_mX_i) < c_{i,v}({}_mX_i), \tag{4.3}$$

then $(t_i,\ {}_mX_i) \in \mathcal{C}$ (see (2.2)) and we move one step ahead along the trajectory to the next position $(t_{i+1},\ {}_mX_{i+1})$. Otherwise if

$$f_i({}_mX_i) \geq c_{i,v}({}_mX_i), \tag{4.4}$$

then we cannot say definitely whether the position $(t_i,\ {}_mX_i)$ belongs to $\mathcal{C}$ or to $\mathcal{E}$. In spite of this fact we do one step ahead in this case as well. Let us recall that the true consumption at (t_i, x) is equal to

$$\gamma_i(x) = [f_i(x) - C_i(x)]^+ \tag{4.5}$$

(see (2.5) and (2.1)). Thus, it is natural to define the upper consumption $\gamma_{i,v}$ at any position $(t_i,\ {}_mX_i)$ by the formula

$$\gamma_{i,v}({}_mX_i) = [f_i({}_mX_i) - c_{i,v}({}_mX_i)]^+. \tag{4.6}$$

Obviously, $c_{i,v} \le C_i$ and hence $\gamma_{i,v} \ge \gamma_i$. Therefore, the price $V_i(x)$ of the European option with payoff function $f_i(x)$ and upper consumption process $\gamma_{i,v}$ is an upper bound on the price $u_i(x)$ of the original American option. In the case (4.3) $\gamma_{i,v}({}_mX_i) = \gamma_i({}_mX_i) = 0$ and we do not get any error. If (4.4) holds and besides $c_{i,v}({}_mX_i) < C_i({}_mX_i)$, we get an error. If $\gamma_{i,v}({}_mX_i)$ is large, then it is in general impossible to estimate this error, but if $\gamma_{i,v}({}_mX_i)$ is small, the error is small as well.

Having found $\gamma_{i,v}$, we can construct an estimate $\hat{V}_0(x_0)$ of the upper bound $V_0(x_0)$ for $u_0(x_0)$ by the formula

$$\hat{V}_0(x_0) = \frac{1}{M}\sum_{m=1}^{M}\frac{f_{\mathcal{I}}({}_mX_{\mathcal{I}})}{B_{\mathcal{I}}} + \frac{1}{M}\sum_{i=0}^{\mathcal{I}-1}\sum_{m=1}^{M}\frac{\gamma_{i,v}({}_mX_i)}{B_i}. \tag{4.7}$$

Note that for the construction of an upper bound V_0 one can use different local low bounds depending on a position. This opens various opportunities for adaptive procedures (see [4]). For instance, if $\gamma_{i,v}({}_mX_i)$ is large, then it is reasonable to use a more powerful local instrument at the position $(t_i,\ {}_mX_i)$.

Instead of using a low bound for constructing a global upper one, one can use low continuation values, in particular, those from Section 3. So, let $\hat{C}_i({}_mX_i)$ be a low continuation value. Then (compare with (4.6))

$$\hat{\gamma}_i({}_mX_i) = [f_i({}_mX_i) - \hat{C}_i({}_mX_i)]^+ \tag{4.8}$$

is an upper consumption value and the corresponding global upper bound is given by the formula

$$\hat{V}_0(x_0) = \frac{1}{M}\sum_{m=1}^{M}\frac{f_{\mathcal{I}}({}_mX_{\mathcal{I}})}{B_{\mathcal{I}}} + \frac{1}{M}\sum_{i=0}^{\mathcal{I}-1}\sum_{m=1}^{M}\frac{\hat{\gamma}_i({}_mX_i)}{B_i}. \tag{4.9}$$

Remark 4.1. In reality (see (3.19)) the global upper bound is equal to $\hat{V}_0(x_0) + \Delta$, where $\Delta \to 0$ when $M, N \to \infty$. Therefore we have $\hat{u}_0(X_0) \le u_0(X_0) \le \hat{V}_0(x_0) + \Delta$, i.e. the accuracy is evaluated by the difference $\hat{V}_0(x_0) + \Delta - \hat{u}_0(X_0)$ (not by $\hat{V}_0(x_0) - \hat{u}_0(X_0)$). In practice, it may be happened that $\hat{V}_0(x_0) \le \hat{u}_0(X_0)$. Clearly, in such a case the accuracy is evaluated by Δ.

5. **Simulations**

5.1. **Bermudan max calls on d assets.** This is a benchmark example studied in [7], [14] and [21] among others. Specifically, the model with d identical assets is considered where each underlying has dividend yield δ. The risk-neutral dynamic of assets is given by

$$\frac{dX_t^k}{X_t^k} = (r-\delta)dt + \sigma dW_t^k, \quad k = 1, ..., d,$$

where W_t^k, $k = 1,...,d$, are independent one dimensional Brownian motions and r, δ, σ are constants. At any time $t \in \{t_0, ..., t_{\mathcal{I}}\}$ the holder of the option may exercise it and receive the payoff

$$f(X_t) = (\max(X_t^1, ..., X_t^d) - K)^+.$$

We take $t_i = iT/\mathcal{I}$, $i = 0, ..., \mathcal{I}$, with $T = 3$, $\mathcal{I} = 9$ and apply the local Monte Carlo method described in the section 3.3. The number of outer Monte Carlo simulations $M = 10000$ and the number of inner Monte Carlo simulations $N = 100$. The results are presented in Table 5.1 in dependence on x_0 with $X_0 = (X_0^1, \ldots, X_0^d)^T$, $X_0^1 = ... = X_0^d = x_0$. Monte-Carlo error is computed using M outer trajectories. The true values are quoted from [12]. The good quality of low bound $\widehat{u}_0(X_0)$ comparatively to the

TABLE 5.1. Bounds (with 95% confidence intervals) for Bermudan max call with parameters $K = 100$, $r = 0.05$, $\sigma = 0.2$, $\delta = 0.1$ and different d and x_0

d	x_0	Lower Bound $\widehat{u}_0(X_0)$	Upper Bound $\widehat{V}_0(X_0)$	True Value
	90	7.965±0.239	8.417±0.082	8.08
2	100	13.644±0.300	14.493±0.113	13.90
	110	20.875±0.370	22.014±0.165	21.34
	90	16.795±0.315	19.0126±0.153	16.71
5	100	26.265±0.379	29.340±0.183	26.21
	110	36.790±0.437	40.630±0.208	36.84

upper bound $\widehat{V}_0(X_0)$ can be attributed to the fact that $\widehat{V}_0(X_0)$ uses local estimates of continuation values in an additive form while $\widehat{u}_0(X_0)$ is based on suboptimal stopping family which depends only on the sign of difference between the payoff and continuation value. Also note, that values of upper bound lie outside 95% confidence interval around the true value. This is again due to the local estimation error and can be cured by increasing the number of inner simulations N.

5.2. **Bermudan swaptions in the Libor market model.** Let us consider the Libor market model with respect to a tenor structure $0 = T_0 < T_1 < \ldots < T_{\mathcal{I}}$ in the spot Libor measure P^*. The dynamics of the forward Libor $L_i(t)$, $0 \le t \le T_i$, $i = 1, \ldots, \mathcal{I} - 1$, is governed by the SDE

$$dL_i = \sum_{j=\eta(t)}^{i} \frac{\delta_j L_i L_j \gamma_i^\top \gamma_j}{1 + \delta_j L_j} dt + L_i \gamma_i^\top dW^*, \quad L_i(0) = L_i^0, \quad t \in [0, T_i], \tag{5.1}$$

where $\delta_j = T_{j+1} - T_j$ are day count factors, $t \mapsto \gamma_i(t) = (\gamma_{i,1}(t), \ldots, \gamma_{i,d}(t))$ are deterministic volatility vector functions defined in $[0, T_i]$ (called factor loadings), and $\eta(t) := \min\{m : T_m > t\}$ denotes the next reset date at time t. In (5.1) $W^*(t)$, $0 \le t \le T_{\mathcal{I}-1}$, is a standard d-dimensional Wiener process under the measure P^* with d, $1 \le d < \mathcal{I}$, being the number of driving factors. The spot Libor measure

P^* is induced by the numeraire

$$(5.2) \qquad B^*(t) := B_{\eta(t)}(t) \prod_{i=0}^{\eta(t)-1} (1 + \delta_i L_i(T_i)),$$

where $B_i(t)$, $i = 0, \ldots, \mathcal{I}$, is the value of a zero coupon bond with face value 1 at T_i. At a tenor date T_i, $i = 1, ..., n-1$, we have (see [12])

$$(5.3) \qquad B_n(T_i) = \prod_{j=i}^{n-1} \frac{1}{1 + \delta_j L_j(T_i)}, \quad n = 1, \ldots, \mathcal{I}.$$

Note, that in (5.2) and (5.3) we set by definition $\prod_k^l = 1$ for $k > l$ and $L_0(T_0) = L_0^0$ is a constant. It is also worth mentioning that $B_n(t)$, $n = 1, \ldots, \mathcal{I} - 1$, are uniquely defined by Libors on the tenor grid only (fortunately, we need values of $B^*(t)$ only there as well).

A European swaption with maturity T_i and strike θ gives the right to contract at T_i for paying a fixed coupon θ and receiving floating Libor at the settlement dates $T_{i+1}, \ldots, T_{\mathcal{I}}$. The corresponding payoff at maturity T_i is given by

$$f_i(L_i(T_i), \ldots, L_{\mathcal{I}-1}(T_i)) := \left(\sum_{j=i}^{\mathcal{I}-1} B_{j+1}(T_i) \delta_j (L_j(T_i) - \theta) \right)^+ .$$

Note, that by setting $L_j(t) = L_j(T_j)$, $t > T_j$, for $j = 0, \ldots, \mathcal{I} - 1$, we can define f_i as a function of the whole Libors vector $(L_0(T_i), \ldots, L_{\mathcal{I}-1}(T_i))$.
A Bermudan swaption issued at $t = 0$ gives the right to obtain

$$f_i(L_i(T_i), \ldots, L_{\mathcal{I}-1}(T_i))$$

at an exercise date $i \in \{s_1, \ldots, s_l = \mathcal{I} - 1\} \subset \{1, \ldots, \mathcal{I} - 1\}$, to be decided by the option holder. Its risk-neutral price is given by

$$u_0(L_0(0), \ldots, L_{\mathcal{I}-1}(0)) = \sup_{\tau \in \mathcal{T}_S} E \left(\left. \frac{f_\tau(L_\tau(T_\tau), \ldots, L_{\mathcal{I}-1}(T_\tau))}{B^*(T_\tau)} \right| \mathcal{F}_0 \right),$$

where $\mathcal{T}_S$ is the set of stopping times τ taking values in $\{s_1, ..., s_l\}$.
For our simulation study we use the Libor volatility structure

$$(5.4) \qquad \gamma_i(t) = c_i g(T_i - t) e_i, \quad \text{where} \quad g(s) = g_\infty + (1 - g_\infty + as) e^{-bs},$$

with e_i being d-dimensional unit vectors, decomposing an input correlation matrix of rank d and $g_\infty \geq 0$, $a \geq 0$, $b \geq 0$, $c_i > 0$ being the constants (see [22]). For generating Libor models with different numbers of factors d, we take as a basis a correlation structure of the form

$$\rho_{ij} = \exp(-\phi|i - j|), \quad i, j = 1, \ldots, \mathcal{I} - 1,$$

which has full rank for $\phi > 0$, and then for a particular choice of d we deduce from ρ a rank-d correlation matrix $\rho^{(d)}$ with decomposition $\rho_{ij}^{(d)} = e_i^\top e_j$, $1 \leq i, j < \mathcal{I}$, by principal component analysis. We take as model parameters a flat 10% initial Libor curve (i.e. $L_i^0 = 0.1$ for $i = 0, 1, \ldots, \mathcal{I} - 1$) over a 40 period quarterly tenor structure, and the parameters

$$\mathcal{I} = 41, \ \delta_i = 0.25, \ c_i \equiv 0.2, \ a = 1.5, \ b = 3.5, \ g_\infty = 0.5, \ \phi = 0.0413.$$

We consider Bermudan swaptions with yearly exercise opportunities, hence (δ_i are equal to a quarter year) $s_i = 4i, \quad i = 1, \dots, 10$. For a practically exactnumerical integration of the SDE, we used the log-Euler scheme with $\Delta t = \delta/5$.
Now, we apply the regression method described in section 3.5, where at each exercise date T_{s_i} the value of the European swaption

$$\mathcal{S}_i(L_{s_i}(T_{s_i}), \dots, L_{n-1}(T_{s_i})) = B^*(T_{s_i}) E\left(\frac{f_{s_{i+1}}(L_{s_{i+1}}(T_{s_{i+1}}), \dots, L_{n-1}(T_{s_{i+1}}))}{B^*(T_{s_{i+1}})} \middle| \mathcal{F}_{s_i} \right)$$

which we can exercise at the next exercise date $T_{s_{i+1}}$ is used as a basis function together with a powers up to second order of the immediate payoff f_{s_i}. Although closed form expressions for European swaptions do not exist in a Libor market model, there do exist very accurate (typically better than 0.3% relative error) formulas (see [22]) which we use for the computation of $\mathcal{S}_i$.
The resulting low bound $\widehat{u}_0$ and upper bound $\widehat{V}_0$ are given in Table 5.2 for different numbers of factors d and different coupons θ. True values (computed with less than 1% relative error) are quoted from [16].

d	θ	$\widehat{u}_0$	$\widehat{V}_0$	True Value
	0.08	1094.8±1.2	1096.1±2.0	1096.1
40	0.10	338.2±1.0	341.2±1.3	339.3
	0.12	96.4±0.5	100.0±0.6	97.2
	0.08	1096.3±1.3	1096.6±2.0	1096.5
10	0.10	344.3±1.0	346.7±1.3	344.7
	0.12	101.7±0.6	104.9±0.7	101.3
	0.08	1108.1±1.5	1110.5±2.4	1109.2
1	0.10	381.7±1.2	384.7±1.6	382.1
	0.12	121.2±0.7	123.1±0.8	121.3

TABLE 5.2. Prices of bermudan swaptions $\times 10^4$

5.3. **Cancellable Snowballs in the Libor market model.** Let us consider a snowball swap contract. According to this contract one has to pay, instead of floating Libor, so called Snowball coupons which follow the following term sheet. One pays on a semi-annual base a constant rate I over the first year and in the forthcoming years (Previous Coupon+A-Libor)$^+$, where A increases as specified in the contract. A cancellable snowball swap is a snowball which may be cancelled (exercised) after the first year. Here we consider this cancellable snowball product in a Libor market model (5.1). The snowballs coupons K_i, settled at T_{i+1}, $i = 0, \dots, \mathcal{I}-1$, are specified by

$$K_i = I, \quad i = 0, 1,$$
$$K_i = (K_{i-1} + A_i - L_i(T_i))^+, \quad i = 2, \dots, \mathcal{I} - 1.$$

We consider the contract where A increases on an annual base in such a way that $A_2 = S$

$$A_{i+1} = A_i + s\,(i \mod 2),$$

with S and s given in the contract. The value u_0 of the cancellable snowball swap at $T_0 = 0$ is given by

$$u_0(L_0(0), \ldots, L_{\mathcal{I}-1}(0)) = \sup_{\tau \in \mathcal{T}_S} E\left(\sum_{j=1}^{\tau} \frac{f_j(L_2(T_2), \ldots, L_{j-1}(T_{j-1})}{B^*(T_j)} \right| \mathcal{F}_0 \Bigg),$$

where $\mathcal{T}_S$ is the set of stopping times τ taking values in $\{2, \ldots, \mathcal{I}\}$ and

$$f_j(L_2(T_2), \ldots, L_{j-1}(T_{j-1})) = \delta_{j-1}(L_{j-1}(T_{j-1}) - K_{j-1}), \quad j = 1, \ldots, \mathcal{I}.$$

Note, that predictable cashflows f_j can take negative values. Since we are going to use linear regression method it is important to find a good basis functions. One possible way would be to include still alive Europeans

$$\max_{j<p\leq\mathcal{I}} E\left(\sum_{q=1}^{p} \frac{f_q(L)}{B^*(T_q)} \right| \mathcal{F}_j \Bigg)$$

at T_j but unfortunately there is no analytical representation for them. However, an approximation can be found (see [5]) using the fact that for any $j + 1 \leq p \leq \mathcal{I}$

$$\begin{aligned} E\left(\sum_{q=j+1}^{p} \frac{f_q(L)}{B^*(T_q)} \right| \mathcal{F}_j \Bigg) &= \frac{1 - B_p(T_j)}{B^*(T_j)} - E\left(\sum_{q=j+1}^{p} \frac{K_{q-1}\delta_{q-1}}{B^*(T_q)} \right| \mathcal{F}_j \Bigg) \\ &= \frac{1 - B_p(T_j)}{B^*(T_j)} - \frac{K_j\delta_j}{B^*(T_{j+1})} - E\left(\sum_{q=j+2}^{p} \frac{K_{q-1}\delta_{q-1}}{B^*(T_q)} \right| \mathcal{F}_j \Bigg). \end{aligned}$$

Replacing in the last summand K_{q-1} by

$$\widetilde{K}_{q-1} = (\alpha K_j + A_{q-1} - L_{q-1}(T_{q-1}))^+, \quad j + 2 \leq q \leq p,$$

where $0 < \alpha < 1$ is a constant which may depend on p and is to be found using optimization, we get a reasonable approximation quality. The value of

$$E\left(\frac{\widetilde{K}_{q-1}\delta_{q-1}}{B^*(T_q)} \right| \mathcal{F}_j \Bigg) = \frac{B_q(T_j)}{B^*(T_j)} E_{B_q}\left((\alpha K_j + A_{q-1} - L_{q-1}(T_{q-1}))^+ \delta_{q-1} \middle| \mathcal{F}_j \right),$$

where E_{B_q} denotes the expectation in respect to T_q forward measure, can be calculated using the Black's formula. Finally, the quadratic polynomials of the spot Libor $L_j(T_j)$ complete the set of basis function at T_j, $j = 2, \ldots, \mathcal{I}$.

As a numerical example let us consider 6yr Snowball with $\delta_i = 0.5$yr ($\mathcal{I} = 12$) and take $I = 0.079$, $S = 0.01$. Further, the volatility structure (5.4) with $a = 0.976$, $b = 2$, $g_\infty = 1.5$ is employed and the correlation matrix is given by

$$\rho_{ij} = \exp\left[\frac{|j - i|}{\mathcal{I} - 2} \log \rho_\infty \right], \quad 1 \leq i, j \leq \mathcal{I} - 1,$$

with $\rho_\infty = 0.663$. The tenor structure, initial Libor curve and factor loadings c_i are shown in Table 5.3. The results in dependence on s are presented in Table 5.4.

Tenors	0.0	0.5	1	1.5	2	2.5
L_0	0.023	0.025	0.027	0.027	0.031	0.031
c_i		0.153	0.143	0.14	0.140	0.139
Tenors	3	3.5	4	4.5	5	5.5
L_0	0.033	0.034	0.036	0.036	0.038	0.039
c_i	0.138	0.137	0.136	0.135	0.134	0.132

TABLE 5.3. Tenor structure, initial Libor curve and factor loadings

s	$\widehat{u}_0$	$\widehat{V}_0$
0.005	64.8±2.4	67.4±2.2
0.004	101.9±2.3	107.3±1.9
0.003	139.8±2.2	143.3±1.7

TABLE 5.4. Prices of cancellable snowballs $\times 10^4$

6. Appendix: Direct Snell envelope by multidimensional approximation and regression

The aim of this section is to show that a modification of procedure (1.2) together with some kind of interpolation can be successfully used in practice. To this end we use $(d+1)$-dimensional sample (see (1.3))

$$({}_mX_i, u_{i+1}({}_mX_{i+1})),\ m = 1, ..., M,\ i = 0, ..., L-1, \tag{6.1}$$

from $(X_i, u_{i+1}(X_{i+1}))$ for sequential evaluating $u_i(X_i),\ i = L-1, ..., 0;\ u_L(X_L) = f_L(X_l)$.

6.1. **Methods based on multivariate interpolation.** Let us suppose that the values $u_{i+1}({}_mX_{i+1})$ of the function $u_{i+1}(x)$ be known. The continuation value $C_i(X_i)$ can be considered as the price of the European option on $[t_i, t_{i+1}]$ at the position (t_i, X_i) with the payoff function $u_{i+1}(x)$. Hence, due to the Monte Carlo approach,

$$C_i({}_mX_i) = \frac{B_i}{B_{i+1}} E(u_{i+1}(X_{i+1}^{(t_i, X_i)}) | X_i =_m X_i) \simeq \frac{B_i}{B_{i+1}} \frac{1}{N} \sum_{n=1}^{N} u_{i+1}({}_nX_{i+1}^{(t_i,\ {}_mX_i)}). \tag{6.2}$$

In (6.2), all the points ${}_nX_{i+1}^{(t_i,\ {}_mX_i)}$, $n = 1, ..., N$, belong to trajectories starting at the instant t_i from ${}_mX_i$ and they are conditionally independent under known ${}_mX_i$. The number N can be chosen depending on i and m : $N = N_{i,m}$. These points differ from ${}_mX_{i+1}$ and therefore we need in an interpolation of $u_{i+1}({}_nX_{i+1}^{(t_i,\ {}_mX_i)})$ through $u_{i+1}({}_jX_{i+1}),\ j = 1, ..., M$. Let us note that at present there are new developments in the theory of multidimensional approximation (see e.g., [10], [25] and references therein) and making use of the theory allows to realize the needed interpolation. Let $\tilde{u}_{i+1}({}_nX_{i+1}^{(t_i,\ {}_mX_i)})$ be an approximation of $u_{i+1}({}_nX_{i+1}^{(t_i,\ {}_mX_i)})$ through $u_{i+1}({}_jX_{i+1}),\ j =$

$1, \dots, M$. Let r_1 be the error of interpolation after substituting $\tilde{u}_{i+1}({}_nX_{i+1}^{(t_i,\ {}_mX_i)})$ instead of $u_{i+1}({}_nX_{i+1}^{(t_i,\ {}_mX_i)})$ in the right-hand side of (6.2). Clearly, r_1 tends to zero if M goes to infinity. Besides, the Monte Carlo error, say r_2, is present in (6.2). Clearly, r_2 tends to zero if N goes to infinity. We have assumed above that the values $u_{i+1}({}_jX_{i+1})$ are known. However, in reality we know their estimates $\hat{u}_{i+1}({}_jX_{i+1})$ only. Interpolation of $u_{i+1}({}_nX_{i+1}^{(t_i,\ {}_mX_i)})$ through $\hat{u}_{i+1}({}_jX_{i+1})$ gives an additional error which increases with decreasing i.

So, we get the following algorithm for evaluating $u_0(x_0)$. We have

$$\hat{u}_L({}_mX_L) = u_L({}_mX_L) = f_L({}_mX_L).$$

The values $\hat{u}_i({}_mX_i)$, $i = L-1, \dots, 0$, are found, if knowing $\hat{u}_{i+1}({}_mX_{i+1})$, in the following way. We simulate $N = N_{i,m}$ trajectories starting from the position $(t_i,\ {}_mX_i)$ on the interval $[t_i, t_{i+1}]$ and obtain the points ${}_nX_{i+1}^{(t_i,\ {}_mX_i)}$, $n = 1, \dots, N$. Then we approximate $\tilde{u}_{i+1}({}_nX_{i+1}^{(t_i,\ {}_mX_i)})$ through $\hat{u}_{i+1}({}_mX_{i+1})$, $m = 1, \dots, M$. We note that the point ${}_mX_{i+1}$ is present among the points ${}_nX_{i+1}^{(t_i,\ {}_mX_i)}$, say ${}_1X_{i+1}^{(t_i,\ {}_mX_i)} ={}_m X_{i+1}$. Clearly, $\tilde{u}_{i+1}({}_1X_{i+1}^{(t_i,\ {}_mX_i)}) = \tilde{u}_{i+1}({}_mX_{i+1}) = \hat{u}_{i+1}({}_mX_{i+1})$. Getting $\tilde{u}_{i+1}({}_nX_{i+1}^{(t_i,\ {}_mX_i)})$, $n = 1, \dots, N$, we evaluate $C_i({}_mX_i)$:

$$\hat{C}_i({}_mX_i) = \frac{B_i}{B_{i+1}} \frac{1}{N} \sum_{n=1}^{N} \tilde{u}_{i+1}({}_nX_{i+1}^{(t_i,\ {}_mX_i)}). \tag{6.3}$$

Then we set

$$\hat{u}_i({}_mX_i) = \max\left\{ f_i({}_mX_i),\ \hat{C}_i({}_mX_i) \right\},\ i = L-1, \dots, 1, \tag{6.4}$$

and finally

$$\hat{u}_0(X_0) = \max\left\{ f_0(X_0),\ \frac{1}{B_1} \frac{1}{M} \sum_{m=1}^{M} \hat{u}_1({}_mX_1) \right\}. \tag{6.5}$$

The number N is chosen taking into account the Monte Carlo error in (6.3) which is surely evaluated during numerical experiments. It is natural that if the one-step errors r_1 and r_2 are sufficiently small, the global error will be small as well. It can be controlled in accordance with the practical rule: if the estimated values do not differ essentially after increasing N and M, then the obtained values are close to the true ones.

6.2. **Using the nearest points.** The previous algorithm is rather expensive because, knowing $\hat{u}_{i+1}({}_mX_{i+1})$ and aiming to evaluate $u_i({}_mX_i)$, we construct $N_{i,m}$ additional trajectories for every point ${}_mX_i$, $m = 1, \dots M$, and then interpolate $N_{i,m}$ unknown values of the function $u_{i+1}(x)$. It turns out that in principle it is possible to avoid both the construction of additional trajectories and interpolation. To this end let us consider $K = K_{i,m}$ nearest points ${}_{m_1}X_i, \dots, {}_{m_K}X_i$ to the point ${}_mX_i$. We have ${}_{m_k}X_{i+1}^{(t_i,\ {}_{m_k}X_i)} ={}_{m_k} X_{i+1}$ with known values $\hat{u}_{i+1}({}_{m_k}X_{i+1})$, $k = 1, \dots, K$. Because ${}_{m_k}X_i$ are close to ${}_mX_i$, one can approximately consider all the points ${}_{m_k}X_{i+1}$ as points on the trajectories starting from the same point ${}_mX_i$ at the instant t_i. Therefore it is reasonable to introduce the following estimate of $C_i({}_mX_i)$:

$$\hat{C}_i({}_mX_i) = \frac{B_i}{B_{i+1}}\frac{1}{K+1}\sum_{k=0}^{K}\hat{u}_{i+1}({}_{m_k}X_{i+1}), \tag{6.6}$$

where m_0 is equal to m.

Then we use (6.4)-(6.5). This procedure is not so expensive as the previous one.

6.3. **Estimation of $C_i({}_mX_i)$ as a regression function.** Let us consider a sample (X_m, Y_m), $m = 1, ..., M$, from (X, Y) and recall the local regression approach. Of interest is to estimate the regression function

$$c(x) = E(Y|X = x) \tag{6.7}$$

and the derivatives $c'(x), ..., c^{(p)}(x)$ at a point $x = x_0$. For simplicity in writing we suppose for a while that X and Y are one-dimensional. Assume that there exists the $(p+1)$-th derivative of $c(x)$ at the point x_0. The unknown regression function $c(x)$ can be locally approximated by a polynomial of order p due to the weighted least squares regression problem (see [13])

$$\sum_{m=1}^{M} K_h(X_m - x_0)\cdot\left[Y_m - \sum_{j=0}^{p}\beta_j(X_m - x_0)^j\right]^2 \to \min_{\beta_0,...,\beta_p}, \tag{6.8}$$

where h is a bandwidth, K is a kernel function. Denote by $\hat{\beta}_j$, $j = 0, ..., p$, the solution to the problem (6.8). Then

$$\hat{c}(x_0) = \hat{\beta}_0, ..., \hat{c}^{(j)}(x_0) = j!\hat{\beta}_j,\ j = 0, ..., p. \tag{6.9}$$

One can apply the regression approach to the sample

$$({}_mX_i, \frac{B_i}{B_{i+1}}u_{i+1}({}_mX_{i+1})),\ m = 1, ..., M,\ \text{from}\ (X_i, \frac{B_i}{B_{i+1}}u_{i+1}(X_{i+1})) \tag{6.10}$$

and get the continuation value $C_i(x)$ as the regression function

$$C_i(x) = \frac{B_i}{B_{i+1}}E\left(u_{i+1}(X_{i+1})|X_i = x\right).$$

This approach gives us $\hat{C}_i({}_mX_i)$ as a function of ${}_mX_i$ (m is fixed) and of all the values $u_{i+1}({}_kX_{i+1})$, $k = 1, ..., M$, (of course, in reality we have $\hat{u}_{i+1}({}_kX_{i+1})$ instead of $u_{i+1}({}_kX_{i+1})$, $k = 1, ..., M$). Then we use (6.4)-(6.5).

6.4. **Direct Snell envelope in the case of known transition probabilities for asserts.** Let the transition probabilities

$$P(X_k \in dx|X_j = y) = p_{j,k}(y, x)dx,\ j \le k, \tag{6.11}$$

be known. For example, they are known for the Black-Scholes model

$$dX_t^l = (r - \delta)X_t^l dt + \sigma X_t^l dw_t^l,\ l = 1, ..., d. \tag{6.12}$$

Let the values $u_{i+1}({}_mX_{i+1})$, $m = 1, ..., M$, be known (in reality $\hat{u}_{i+1}({}_mX_{i+1})$ are known). We have

$$(6.13) \qquad C_i({}_kX_i) = \frac{B_i}{B_{i+1}} E\left(u_{i+1}(X_{i+1})|{}_kX_i\right) = \frac{B_i}{B_{i+1}} \int u_{i+1}(x) p_{i,i+1}({}_kX_i, x) dx.$$

The known values ${}_mX_{i+1}$, $m = 1, ..., M$, are distributed as i.i.d. due to the law

$$(6.14) \qquad p_{0,i+1}(X_0, x) := p_{i+1}(X_0, x).$$

We get

$$(6.15) \qquad E\left(u_{i+1}(X_{i+1})|{}_kX_i\right) = \int u_{i+1}(x) p_{i,i+1}({}_kX_i, x) dx$$

$$= \int u_{i+1}(x) \frac{p_{i,i+1}({}_kX_i, x)}{p_{i+1}(X_0, x)} p_{i+1}(X_0, x) dx = E\left(u_{i+1}(X_{i+1}) \frac{p_{i,i+1}({}_kX_i, X_{i+1})}{p_{i+1}(X_0, X_{i+1})} |{}_kX_i\right)$$

$$\simeq \frac{1}{M} \sum_{m=1}^{M} u_{i+1}({}_mX_{i+1}) \frac{p_{i,i+1}({}_kX_i, \ {}_mX_{i+1})}{p_{i+1}(X_0, \ {}_mX_{i+1})} .$$

In (6.15) only one error is present, namely the Monte Carlo error. It will be accumulated because in reality instead of $u_{i+1}({}_mX_{i+1})$ we have the estimate $\hat{u}_{i+1}({}_mX_{i+1})$. As a result we obtain the following estimate for the continuation value

$$(6.16) \qquad \hat{C}_i({}_kX_i) = \frac{B_i}{B_{i+1}} \cdot \frac{1}{M} \sum_{m=1}^{M} \hat{u}_{i+1}({}_mX_{i+1}) \frac{p_{i,i+1}({}_kX_i, \ {}_mX_{i+1})}{p_{i+1}(X_0, \ {}_mX_{i+1})} .$$

Then we use (6.4)-(6.5).

We emphasize that in the case considered we need not both in simulation of additional trajectories and in any interpolation. It is suitable for construction of test examples. Let us note that this case is connected with the mesh method. To be convinced in this it suffices to set

$$(6.17) \qquad W^i_{km} := \frac{p_{i,i+1}({}_kX_i, \ {}_mX_{i+1})}{p_{i+1}(X_0, \ {}_mX_{i+1})} ,$$

see details in [12].

7. Acknowledgment

This work was finished while the second author was a visitor of the Weierstrass-Institute für Angewandte Analysis und Stochastik (WIAS), Berlin, due to the financial supports from this institute and DFG (grant No. 436 RUS 17/137/05) which are gratefully acknowledged.

References

[1] L. Andersen, M. Broadie (2001). A primal-dual simulation algorithm for pricing multidimensional American options. Working paper, Columbia Business School, New York.

[2] V. Bally, G. Pagès, J. Printems (2005). A quantization tree method for pricing and hedging multidimensional American options. Mathematical Finance, **15**, No. 1, 119-168.

[3] D. Belomestny, G.N. Milstein (2006). Monte Carlo evaluation of American options using consumption processes. International Journal of Theoretical and Applied Finance, **9**, No. 4, 1-27.

[4] D. Belomestny, G.N. Milstein (2005). Adaptive simulation algorithms for pricing American and Bermudan options by local analysis of financial market. WIAS-Preprint No. 1022, Berlin.
[5] C. Bender, A. Kolodko, J. Schoenmakers (2005). Iterating Snowballs and related path dependent callables in a multi-factor Libor model. Preprint No. 1061, WIAS, Berlin.
[6] P. Boyle, M. Broadie, P. Glasserman (1997). Monte Carlo methods for security pricing. Journal of Economic Dynamics and Control, **21**, 1267-1321.
[7] M. Broadie, P. Glasserman (1997). Pricing American-style securities using simulation. J. of Economic Dynamics and Control, **21**, 1323-1352.
[8] J. Carriere (1996). Valuation of early-exercise price of options using simulations and nonparametric regression. Insuarance: Mathematics and Economics, **19**, 19-30.
[9] E. Clément, D. Lamberton, P. Protter (2002). An analysis of a least squares regression algorithm for American option pricing. Finance and Stochastics, **6**, 449-471.
[10] R.A. DeVore (1998). Nonlinear approximation. Acta Numer., **7**, 51-150.
[11] D. Egloff (2005). Monte Carlo algorithms for optimal stopping and statistical learning. Ann. Appl. Probab. **15**, no. 2, 1396-1432.
[12] P. Glasserman (2004). Monte Carlo Methods in Financial Engineering. Springer.
[13] L. Györfi, M. Kohler, A. Krzyżak, H. Walk (2002). A Distribution-Free Theory of Nonparametric Regression. Springer.
[14] M. Haugh, L. Kogan (2004). Pricing American options: a duality approach. Opeations Research, **52**, No. 2, 258–270.
[15] F. Jamshidian (2003). Minimax optimality of Bermudan and American claims and their Monte Carlo upper bound approximation. Working paper.
[16] A. Kolodko, J. Schoenmakers (2004). Upper bounds for Bermudan style derivatives. Monte Carlo Methods and Appl., **10**(3-4), 331-343.
[17] A. Kolodko, J. Schoenmakers (2006). Iterative construction of the optimal Bermudan stopping time. Finance and Stochastics, **10**, No. 1, 27-49.
[18] D. Lamberton, B. Lapeyre (1996). Introduction to Stochastic Calculus Applied to Finance. Chapman & Hall.
[19] F.A. Longstaff, E.S. Schwartz (2001). Valuing American options by simulation: a simple least-squares approach. Review of Financial Studies, **14**, 113-147.
[20] G.N. Milstein, M.V. Tretyakov (2005). Numerical Analysis of Monte Carlo evaluation of Greeks by finite differences. J. of Computational Finance, **8**, No. 3, 1-33.
[21] L.C.G. Rogers (2001). Monte Carlo valuation of American options. Mathematical Finance, **12**, 271-286.
[22] J. Schoenmakers (2005). Robust Libor Modelling and Pricing of Derivative Products. Chapman & Hall/CRC.
[23] A.N. Shiryaev (1999). Essentials of Stochastic Finance: Facts, Models, Theory. World Scientific.
[24] J. Tsitsiklis, B. Van Roy (1999). Regression methods for pricing complex American style options. IEEE Trans. Neural. Net., **12**, 694-703.
[25] H. Wendland (2005). Scattered Data Approximation. Cambridge: Cambridge University Press.