An Efficient Dual Monte Carlo Upper Bound for Bermudan Style Derivatives

A. Kolodko *   J. Schoenmakers †

January 15, 2004

Abstract

Based on a duality approach for Monte Carlo construction of upper bounds for American/Bermudan derivatives (Rogers, Haugh & Kogan), we present a new algorithm for computing dual upper bounds in a more efficient way. The method is applied to Bermudan swaptions in the context of a LIBOR market model, where the dual upper bound is constructed from the maximum of still alive swaptions. We give a numerical comparison with Andersen's lower bound method and its dual considered by Andersen & Broadie.

2000 Mathematical Subject Classification: 65C05, 91B28.

Keywords: Bermudan options, Monte Carlo, duality approach, LIBOR models

1 Introduction

Evaluation of American style derivatives on a high dimensional system of underlyings is considered a perennial problem for the last decades. On the one hand such high dimensional options are difficult, if not impossible, to compute by PDE methods for free boundary value problems. On the other hand Monte Carlo simulation, which is for high dimensional European options an almost canonical alternative to PDE solving, is for American options highly non-trivial since the (optimal) exercise boundary is usually unknown. In the past literature, many approaches for Monte Carlo simulation of American options are developed. With respect to Bermudan derivatives, which are in fact American options with a finite number of exercise

*Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin ; E-mail: kolodko@wias-berlin.de. Supported by the DFG Research Center "Mathematics for key technologies" (FZT 86) in Berlin.

†Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin ; E-mail: schoenma@wias-berlin.de

dates, there is, for example, the stochastic mesh method of Broadie & Glasserman (1997,2000), a cross-sectional regression approach by Longstaff & Schwartz (2001), and for Bermudan swaptions a method by Andersen (1999). In general, the price of an American option can be represented as a supremum over a set of stopping times. As a remarkable result Rogers (2001) (and independently Haugh & Kogan (2001) for Bermudan style instruments) showed that this supremum representation can be converted into a 'dual' infimum representation, where the infimum is taken over a set of (super-)martingales. In Andersen & Broadie (2001) this dual approach is carried out and tested with respect to Andersen's (1999) method for Bermudan swaptions. Further Joshi & Theis (2002) use the dual approach for finding Bermudan swaption prices via a minimization procedure. For a more detailed overview on Monte Carlo methods for American options we refer to Glasserman (2003) and the references therein.

In the papers of Anderson & Broadie (2001) and Haugh & Kogan (2001) upper bounds of Bermudan options are constructed by applying the duality approach to the (Doob-Meyer) martingale part of an approximative process. For instance, in Andersen & Broadie (2001) these upper bounds are constructed to investigate the quality of an approximative lower bound process obtained by suboptimal stopping, without particular focus on the efficiency of the upper bound computation however. The central theme in this paper is the construction of a Monte Carlo estimator for an upper bound for a Bermudan derivative which is computationally more efficient. Our upper bound construction will be based on duality via the martingale part of an approximative processes as well. But, as main contribution, we will enclose the 'theoretical' upper bound by approximating from above and below by using a new lower estimator for the theoretical upper bound. Then, by taking a convex combination of the lower and upper estimator we obtain a family of combined estimators for the target upper bound with usually higher computational efficiency. This efficiency gain will be demonstrated by upper bound computation of Bermudan swaptions.

The paper is organised as follows. In Section 2 we give a concise recap of the Bermudan pricing problem and in Section 3 we outline the duality approach. Then, in Section 4 we present new Monte Carlo estimators for constructing a target upper bound and in Section 5 we propose two canonical approximative processes to which our method could be applied. Finally, in Section 6 we apply our method to computation of upper bounds of Bermudan swaptions in a LIBOR market model. This application is based on the maximum of still alive swaptions, one of the canonical candidates in Section 5 in fact, and we give a numerical comparison with the results obtained by Andersen (1999) and Andersen & Broadie (2001).

2 The Bermudan Pricing Problem

We consider general Bermudan style derivatives with respect to an underlying process $L(t)$, over some finite time interval $[0, T]$ with time horizon $T < \infty$. The process L is assumed to be Markovian with state space $\mathbb{R}^D$. For example, L can be a system of asset prices, but also a not explicitly tradable object such as the term structure of interest rates, or a system of LIBOR rates. Consider a set of future dates $\mathbb{T} := \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_k\}$ with $0 < \mathcal{T}_1 < \mathcal{T}_2 < \cdots < \mathcal{T}_k \leq T$. The dates are denoted with calligraphic letters to distinguish in the case where L is a LIBOR rate process, if necessary, from a particular LIBOR tenor structure usually denoted by T_j's.

An option issued at time $t = 0$, to exercise a cashflow $C_{\mathcal{T}_\tau} := C(\mathcal{T}_\tau, L(\tau))$ at a future time $\mathcal{T}_\tau \in \mathbb{T}$ is called a Bermudan style derivative. Without restriction we assume for technical reasons that the option cannot be exercised at $t = 0$. With respect to a pricing measure P connected with some pricing numeraire B, the value of the Bermudan derivative at time $t = 0$ is given by

$$V_0 = B(0) \sup_{\tau \in \{1,\ldots,k\}} E^{\mathcal{F}_0} \frac{C_{\mathcal{T}_\tau}}{B(\mathcal{T}_\tau)}. \tag{1}$$

The fact that (1) can be considered as the fair price for the Bermudan derivative is due to general no-arbitrage principles, e.g. see Duffie (2001). For example, if L is a LIBOR process, P in (1) could be the spot LIBOR measure P^* induced by the spot measure numeraire B^* or a bond measure $P^{(m)}$ induced by some zero bond B_m maturing at tenor T_m, where $\mathcal{T}_k < T_m$. The supremum in (1) is taken over all integer valued $\mathbb{F}$-stopping times τ with values in the set $\{1, ..., k\}$, where $\mathbb{F} := \{\mathcal{F}_t,\ 0 \leq t \leq T\}$ denotes the usual filtration generated by the process L. At a future time point t, when the option is not exercised before t, the Bermudan option value is given by

$$V_t = B(t) \sup_{\tau \in \{\kappa(t),\ldots,k\}} E^{\mathcal{F}_t} \frac{C_{\mathcal{T}_\tau}}{B(\mathcal{T}_\tau)}$$

with $\kappa(t) := \min\{m : \mathcal{T}_m \geq t\}$. Note that V_t can also be seen as the price of a Bermudan option newly issued at time t, with exercise opportunities $\mathcal{T}_{\kappa(t)}, \ldots, \mathcal{T}_k$.

The process

$$Y_t := \frac{V_t}{B(t)},$$

called the *Snell envelope* process, is a supermartingale. This can be seen as follows. Let $s < t$ and τ_t^* be an optimal stopping index at time t (which exists by general arguments), then it holds

$$E^{\mathcal{F}_s} Y_t = E^{\mathcal{F}_s} E^{\mathcal{F}_t} \frac{C_{\mathcal{T}_{\tau_t^*}}}{B(\mathcal{T}_{\tau_t^*})} = E^{\mathcal{F}_s} \frac{C_{\mathcal{T}_{\tau_t^*}}}{B(\mathcal{T}_{\tau_t^*})} \leq \sup_{\tau \in \{\kappa(s),\ldots,k\}} E^{\mathcal{F}_s} \frac{C_{\mathcal{T}_\tau}}{B(\mathcal{T}_\tau)} = Y_s.$$

3 Upper bounds by a Duality approach

We introduce the discrete filtration $\left(\mathcal{F}^{(j)}\right)_{j=0,...,k}$ with $\mathcal{F}^{(j)} := \mathcal{F}_{\mathcal{T}_j}$, $1 \le j \le k$, $\mathcal{F}^{(0)} := \mathcal{F}_0$, and consider with respect to this filtration a discrete martingale $(M_j)_{j=0,...,k}$ with $M_0 = 0$. Following Rogers (2001) we observe that

$$\begin{aligned} Y_0 &= \sup_{\tau \in \{1,...,k\}} E^{\mathcal{F}_0} \frac{C_{\mathcal{T}_\tau}}{B(\mathcal{T}_\tau)} = \sup_{\tau \in \{1,,...,k\}} E^{\mathcal{F}_0} \left[\frac{C_{\mathcal{T}_\tau}}{B(\mathcal{T}_\tau)} - M_\tau \right] \\ &\le E^{\mathcal{F}_0} \max_{1 \le j \le k} \left[\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - M_j \right]. \end{aligned} \tag{2}$$

Hence the right-hand-side of (2) provides an upper bound for the Bermudan price Y_0. Moreover, due to the next theorem of Rogers (2001) and independently Haugh & Kogan (2001), there exists a particular martingale M^Y, such that (2) holds with equality.

Theorem 3.1 *Let us consider the Snell envelope process* Y *at the discrete time set* $\{0, \mathcal{T}_1, ..., \mathcal{T}_k\}$, *and define* $Y^{(j)} := Y(\mathcal{T}_j)$, $1 \le j \le k$, $Y^{(0)} := Y_0$. *Let further* M^Y *be the (unique) Doob-Meyer martingale part of* $\left(Y^{(j)}\right)_{0 \le j \le k}$, *i.e.* M^Y *is an* $\left(\mathcal{F}^{(j)}\right)$-*martingale which satisfies*

$$Y^{(j)} = Y_0 + M_j^Y - F_j^Y, \quad j = 0, ..., k,$$

with $M_0^Y := F_0^Y := 0$ *and* F^Y *being such that* F_j^Y *is* $\mathcal{F}^{(j-1)}$ *measurable for* $j = 1, ..., k$. *Then we have*

$$Y_0 = E^{\mathcal{F}_0} \max_{1 \le j \le k} \left[\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - M_j^Y \right].$$

Proof. Note that always $Y_j \ge C_{T_j}/B(T_j)$ and that F_j^Y is nondecreasing since $(Y^{(j)})$ is an $\left(\mathcal{F}^{(j)}\right)$-supermartingale. So, (2) applied to M^Y yields

$$\begin{aligned} Y_0 &\le E^{\mathcal{F}_0} \sup_{1 \le j \le k} \left[\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - M_j^Y \right] = E^{\mathcal{F}_0} \left[Y_0 + \sup_{1 \le j \le k} \left[\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - Y^{(j)} - F_j^Y \right] \right] \\ &\le E^{\mathcal{F}_0} \left[Y_0 + \sup_{1 \le j \le k} \left[-F_j^Y \right] \right] = Y_0 - F_1^Y = Y_0, \end{aligned}$$

where $F_1^Y = 0$ because of $Y_0 = E^{\mathcal{F}_0} Y^{(1)} = Y_0 - F_1^Y$. ■

4 Efficient Monte Carlo construction of upper bounds

Consider some approximative process $\widetilde{V}_t$ for the price of a Bermudan style option issued at time t. As an example, for any exercise strategy, i.e. a family of integer valued stopping times $\{\tau_t \in \{\kappa(t), ..., k\} : t \ge 0\}$, the process

$$\widetilde{V}_t := B(t) E^{\mathcal{F}_t} \frac{C_{\mathcal{T}_{\tau_t}}}{B(\mathcal{T}_{\tau_t})}, \tag{3}$$

is a lower approximation, $\widetilde{V}_t \le V_t$. The discounted process $\widetilde{Y} := \widetilde{V}/B$ is the with $\widetilde{V}$ associated approximation of the Snell envelope process. Similar as in Section 3 we introduce the discrete processes $\widetilde{Y}^{(j)}$ and $\widetilde{V}^{(j)}$, adapted to $\mathcal{F}^{(j)}$ for $j = 0, \ldots, k$. Let $\widetilde{M}$ be the martingale part of the Doob-Meyer decomposition of $(\widetilde{Y}^{(j)})$. Hence

$$\widetilde{Y}^{(j)} = \widetilde{Y}_0 + \widetilde{M}_j - \widetilde{F}_j, \quad j = 0, \ldots, k, \tag{4}$$

with $\widetilde{M}_0 = \widetilde{F}_0 = 0$ and $\widetilde{F}_j$ being $\mathcal{F}^{(j-1)}$ measurable for $j = 1, \ldots, k$. By taking the conditional expectation with respect to $\mathcal{F}^{(j-1)}$ at both sides of (4), it follows that

$$\begin{aligned} \widetilde{M}_j &= \widetilde{M}_{j-1} + \widetilde{Y}^{(j)} - E^{\mathcal{F}^{(j-1)}} \widetilde{Y}^{(j)} \\ &= \sum_{i=1}^{j} \widetilde{Y}^{(i)} - \sum_{i=1}^{j} E^{\mathcal{F}^{(i-1)}} \widetilde{Y}^{(i)}, \quad 1 \le j \le k. \end{aligned}$$

So, by Theorem 3.1 we obtain an upper bound for the Bermudan option via

$$\begin{aligned} Y_0 = \frac{V_0}{B(0)} &\le E \sup_{1 \le j \le k} [\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \sum_{i=1}^{j} \widetilde{Y}^{(i)} + \sum_{i=1}^{j} E^{\mathcal{F}^{(i-1)}} \widetilde{Y}^{(i)}] \\ &= \widetilde{Y}_0 + E \sup_{1 \le j \le k} \left[\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \widetilde{Y}^{(j)} + \sum_{i=1}^{j} E^{\mathcal{F}^{(i-1)}} [\widetilde{Y}^{(i)} - \widetilde{Y}^{(i-1)}] \right] \\ &=: \widetilde{Y}_0 + \Delta =: \frac{V_0^{up}}{B(0)}. \end{aligned}$$

Let us assume that $(\widetilde{V}^{(j)})$ satisfies $\widetilde{V}^{(j)} \ge C_{\mathcal{T}_j}$, hence, the approximative price process is never below the cash flow by exercising. This is no restriction in fact, since otherwise we might take $\widetilde{\widetilde{V}}^{(j)} := \max(\widetilde{V}^{(j)}, C_{\mathcal{T}_j})$ instead. We then have the following estimate,

$$\begin{aligned} \Delta &\le E \sup_{1 \le j \le k} \sum_{i=1}^{j} [E^{\mathcal{F}^{(i-1)}} \widetilde{Y}^{(i)} - \widetilde{Y}^{(i-1)}] \\ &\le E \sup_{1 \le j \le k} \sum_{i=1}^{j} \max(E^{\mathcal{F}^{(i-1)}} \widetilde{Y}^{(i)} - \widetilde{Y}^{(i-1)}, 0) \\ &\le E \sum_{i=1}^{k} \max(E^{\mathcal{F}^{(i-1)}} \widetilde{Y}^{(i)} - \widetilde{Y}^{(i-1)}, 0). \end{aligned} \tag{5}$$

When $\widetilde{Y}$ coincides with the Snell envelope process Y we have $\Delta = 0$ by Theorem 3.1 and then, due to the supermartingale property of the Snell envelope, $E^{\mathcal{F}^{(i-1)}} Y^{(i)} \le Y^{(i-1)}$, so the right-hand-side estimate vanishes as well. The estimation (5) indicates that the distance Δ between Y and $\widetilde{Y}$ is due to those exercise dates $\mathcal{T}_i$, where $E^{\mathcal{F}^{(i-1)}} \widetilde{Y}^{(i)} \ge \widetilde{Y}^{(i-1)}$, hence where $\widetilde{Y}$ doesn't meet the supermartingale property.

Because the process L is assumed to be Markovian in the state space $\mathbb{R}^D$, a conditional probability given $\mathcal{F}^{(j)}$ for $j = 0, \ldots, k$, can be seen as a function of

$L^{(j)} := L(\mathcal{T}_j)$, with $L^{(0)} := L(0)$, and by general arguments (see for instance Ikeda & Watanabe (1981)) there exist a regular conditional probability measure $P(L^{(j)}, \bullet)$, such that for any $\mathcal{F}_T$-measurable random variable Z,

$$[E^{\mathcal{F}^{(j)}} Z](\omega) =: \int P(L^{(j)}, d\widetilde{\omega}) Z(\widetilde{\omega}) \qquad a.s., \qquad j = 0, \ldots, k.$$

We now consider for each j, $j = 1, ..., k$, a sequence of random variables $\left(\xi_i^{(j)}\right)_{i \in \mathbb{N}}$, where for $i \in \mathbb{N}$, $\xi_i^{(j)}$ are i.i.d. copies of $\widetilde{Y}^{(j)}$ under the conditional measure $P(L^{(j-1)}, \bullet)$, independent of the sigma-algebra $\sigma\{L^{(i)} : i = j, \ldots, k\}$. Hence,

$$E^{\mathcal{F}^{(j-1)}} \widetilde{Y}^{(j)} = \int P(L^{(j-1)}, d\widetilde{\omega}) \widetilde{Y}^{(j)}(\widetilde{\omega}) = \int P(L^{(j-1)}, d\widetilde{\omega}) \xi_i^{(j)}(\widetilde{\omega}), \qquad i \in \mathbb{N}.$$

For a fixed but arbitrary $K \in \mathbb{N}$ we consider a discrete process $\widetilde{M}^{(K)}$ defined by $\widetilde{M}_0^{(K)} = 0$ and then, recursively,

$$\begin{aligned} \widetilde{M}_j^{(K)} &:= \widetilde{M}_{j-1}^{(K)} + \widetilde{Y}^{(j)} - \frac{1}{K} \sum_{i=1}^{K} \xi_i^{(j)} \\ &= \sum_{q=1}^{j} \widetilde{Y}^{(q)} - \sum_{q=1}^{j} \frac{1}{K} \sum_{i=1}^{K} \xi_i^{(q)}, \qquad j = 1, \ldots, k. \end{aligned}$$

The process $\widetilde{M}^{(K)}$ is thus defined on an extended probability space $\Omega \times \prod$ with $\prod := \prod_{j=1}^{k} \mathbb{R}^K$. So a generic sample element in this space is $(\omega, (\xi^{(j)})_{1 \le j \le k})$, with $\omega \in \Omega$ being a realisation of the process L and $\xi^{(j)} := (\xi_i^{(j)})_{i=1,\ldots,K} \in \mathbb{R}^K$, for $j = 1, \ldots, k$.

Clearly, $\widetilde{M}^{(K)}$ is a martingale w.r.t. the filtration $(\widetilde{\mathcal{F}}^{(j)})_{j=0,\ldots,k}$, defined by $\widetilde{\mathcal{F}}^{(0)} := \mathcal{F}_0$ and $\widetilde{\mathcal{F}}^{(j)} := \sigma\{F \times H : \Omega \supset F \in \mathcal{F}^{(j)}, \prod \supset H \in \sigma\{\xi^{(1)}, \ldots, \xi^{(j)}\}\}$, for $j = 1, \ldots, k$, and we observe that

$$\begin{aligned} E \sup_{1 \le j \le k} [\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \widetilde{M}_j^{(K)}] &= E E^{\mathcal{F}^{(k)}} \sup_{1 \le j \le k} [\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \sum_{q=1}^{j} \widetilde{Y}^{(q)} + \sum_{q=1}^{j} \frac{1}{K} \sum_{i=1}^{K} \xi_i^{(q)}] \\ &\ge E \sup_{1 \le j \le k} [\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \sum_{q=1}^{j} \widetilde{Y}^{(q)} + \sum_{q=1}^{j} \frac{1}{K} \sum_{i=1}^{K} E^{\mathcal{F}^{(k)}} \xi_i^{(q)}] \\ &= E \sup_{1 \le j \le k} [\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \sum_{q=1}^{j} \widetilde{Y}^{(q)} + \sum_{q=1}^{j} \frac{1}{K} \sum_{i=1}^{K} E^{\mathcal{F}^{(q-1)}} \xi_i^{(q)}] \\ &= E \sup_{1 \le j \le k} [\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \sum_{q=1}^{j} \widetilde{Y}^{(q)} + \sum_{q=1}^{j} E^{\mathcal{F}^{(q-1)}} \widetilde{Y}^{(q)}] \\ &= \frac{V_0^{up}}{B(0)} \ge \frac{V_0}{B(0)}, \end{aligned}$$

where $E^{\mathcal{F}^{(k)}}\xi_i^{(q)} = E^{\mathcal{F}^{(q-1)}}\xi_i^{(q)}$ holds because $\xi_i^{(q)}$ is independent of $L^{(q)}, \ldots, L^{(k)}$. Via the martingale $\widetilde{M}^{(K)}$ we have thus obtained a new upper bound

$$V_0^{up^{up},K} := B(0)E \sup_{1\le j\le k}[\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \widetilde{M}_j^{(K)}], \tag{6}$$

which is larger than our target upper bound V_0^{up}. It is natural to expect, however, that $V_0^{up^{up},K}$ will be already close to V_0^{up} for numbers K which are much smaller than the number of Monte Carlo trajectories needed for low variance estimation of the mathematical expectation in (6).

We now proceed with a second approach, which gives a lower bound for our target upper bound V_0^{up}. Consider an $(\mathcal{F}^{(k)})$-measurable random index $j_{\max}$ which satisfies

$$\sup_{1\le j\le k}[\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \sum_{q=1}^{j}\widetilde{Y}^{(q)} + \sum_{q=1}^{j}E^{\mathcal{F}^{(q-1)}}\widetilde{Y}^{(q)}] = \frac{C_{\mathcal{T}_{j_{\max}}}}{B(\mathcal{T}_{j_{\max}})} - \sum_{q=1}^{j_{\max}}\widetilde{Y}^{(q)} + \sum_{q=1}^{j_{\max}}E^{\mathcal{F}^{(q-1)}}\widetilde{Y}^{(q)}.$$

Then, for any integer $K > 0$,

$$\begin{aligned}
\frac{V_0^{up}}{B(0)} &= E[\frac{C_{\mathcal{T}_{j_{\max}}}}{B(\mathcal{T}_{j_{\max}})} - \sum_{q=1}^{j_{\max}}\widetilde{Y}^{(q)} + \sum_{q=1}^{j_{\max}}E^{\mathcal{F}^{(q-1)}}\widetilde{Y}^{(q)}] \\
&= E[\frac{C_{\mathcal{T}_{j_{\max}}}}{B(\mathcal{T}_{j_{\max}})} - \sum_{q=1}^{j_{\max}}\widetilde{Y}^{(q)} + \sum_{q=1}^{j_{\max}}\frac{1}{K}\sum_{i=1}^{K}E^{\mathcal{F}^{(q-1)}}\xi_i^{(q)}] \\
&= E[\frac{C_{\mathcal{T}_{j_{\max}}}}{B(\mathcal{T}_{j_{\max}})} - \sum_{q=1}^{j_{\max}}\widetilde{Y}^{(q)} + \sum_{q=1}^{j_{\max}}\frac{1}{K}\sum_{i=1}^{K}\xi_i^{(q)}],
\end{aligned}$$

where we have used again the fact that $E^{\mathcal{F}^{(k)}}\xi_i^{(q)} = E^{\mathcal{F}^{(q-1)}}\xi_i^{(q)} = E^{\mathcal{F}^{(q-1)}}\widetilde{Y}^{(q)}$. This brings us to the idea of localizing $j_{\max}$ for each particular simulation of the process L. To this aim, we carry out the following procedure. We consider on the extended probability space $\Omega \times \prod$ the random index $\widehat{j}_{\max}$ which satisfies,

$$\frac{C_{\mathcal{T}_{\widehat{j}_{\max}}}}{B(\mathcal{T}_{\widehat{j}_{\max}})} - \sum_{q=1}^{\widehat{j}_{\max}}\widetilde{Y}^{(q)} + \sum_{q=1}^{\widehat{j}_{\max}}\frac{1}{K}\sum_{i=1}^{K}\xi_i^{(q)} = \sup_{1\le j\le k}[\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \sum_{q=1}^{j}\widetilde{Y}^{(q)} + \sum_{q=1}^{j}\frac{1}{K}\sum_{i=1}^{K}\xi_i^{(q)}].$$

Next, we extend the probability space once again to $\Omega \times \prod \times \prod$ and simulate *independent copies* $\widehat{\xi}^{(j)} := (\widehat{\xi}_i^{(j)})_{i=1,\ldots,K} \in \mathbb{R}^K$, of $\xi^{(j)} \in \mathbb{R}^K$, for $j = 1, \ldots, k$. We then consider on $\Omega \times \prod \times \prod$ the random variable,

$$\frac{C_{\mathcal{T}_{\widehat{j}_{\max}}}}{B(\mathcal{T}_{\widehat{j}_{\max}})} - \sum_{q=1}^{\widehat{j}_{\max}}\widetilde{Y}^{(q)} + \sum_{q=1}^{\widehat{j}_{\max}}\frac{1}{K}\sum_{i=1}^{K}\widehat{\xi}_i^{(q)}$$

with expectation

$$
\begin{aligned}
\frac{V_0^{up_{low},K}}{B(0)} &:= E\Big[\frac{C_{\mathcal{T}_{\widehat{j}_{\max}}}}{B(\mathcal{T}_{\widehat{j}_{\max}})} - \sum_{q=1}^{\widehat{j}_{\max}} \widetilde{Y}^{(q)} + \sum_{q=1}^{\widehat{j}_{\max}} \frac{1}{K}\sum_{i=1}^{K} \widehat{\xi}_i^{(q)}\Big] \\
&= E\Big[\frac{C_{\mathcal{T}_{\widehat{j}_{\max}}}}{B(\mathcal{T}_{\widehat{j}_{\max}})} - \sum_{q=1}^{\widehat{j}_{\max}} \widetilde{Y}^{(q)} + \sum_{q=1}^{\widehat{j}_{\max}} \frac{1}{K}\sum_{i=1}^{K} E^{\widetilde{\mathcal{F}}^{(k)}} \widehat{\xi}_i^{(q)}\Big] \qquad (7)\\
&= E\Big[\frac{C_{\mathcal{T}_{\widehat{j}_{\max}}}}{B(\mathcal{T}_{\widehat{j}_{\max}})} - \sum_{q=1}^{\widehat{j}_{\max}} \widetilde{Y}^{(q)} + \sum_{q=1}^{\widehat{j}_{\max}} E^{\mathcal{F}^{(q-1)}} \widetilde{Y}^{(q)}\Big] \\
&\leq E\Big[\frac{C_{\mathcal{T}_{j_{\max}}}}{B(\mathcal{T}_{j_{\max}})} - \sum_{q=1}^{j_{\max}} \widetilde{Y}^{(q)} + \sum_{q=1}^{j_{\max}} E^{\mathcal{F}^{(q-1)}} \widetilde{Y}^{(q)}\Big] = \frac{V_0^{up}}{B(0)},
\end{aligned}
$$

where, most importantly, (7) holds while the $\widehat{\xi}^q$ are re-sampled *independent* of the determination of $\widehat{j}_{\max}$ and then we have $E^{\widetilde{\mathcal{F}}^{(k)}} \widehat{\xi}_i^{(q)} = E^{\mathcal{F}^{(k)}} \widehat{\xi}_i^{(q)} = E^{\mathcal{F}^{(q-1)}} \widehat{\xi}_i^{(q)} = E^{\mathcal{F}^{(q-1)}} \widetilde{Y}^{(q)}$.

So we come up with two different Monte Carlo estimators for the target upper bound V_0^{up}.

Lower estimate for V_0^{up}:

$$
\widehat{V}_0^{up_{low},K,M} := \frac{B(0)}{M} \sum_{m=1}^{M} \Big[\frac{C_{\mathcal{T}_{\widehat{j}^{(m)}_{\max}}}}{B(\mathcal{T}_{\widehat{j}^{(m)}_{\max}})} - \sum_{q=1}^{\widehat{j}^{(m)}_{\max}} \widetilde{Y}^{(q;m)} + \sum_{q=1}^{\widehat{j}^{(m)}_{\max}} \frac{1}{K} \sum_{i=1}^{K} \widehat{\xi}_i^{(q;m)}\Big] \qquad (8)
$$

Upper estimate for V_0^{up}:

$$
\widehat{V}_0^{up^{up},K,M} := \frac{B(0)}{M} \sum_{m=1}^{M} \sup_{1\leq j \leq k} \Big[\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \sum_{q=1}^{j} \widetilde{Y}^{(q;m)} + \sum_{q=1}^{j} \frac{1}{K} \sum_{i=1}^{K} \xi_i^{(q;m)}\Big] \qquad (9)
$$

In (8), (9), $\widehat{j}^{(m)}_{\max}$ and $\widetilde{Y}^{(q;m)}$ denote the m-th independent sample of $\widehat{j}_{\max}$ and $\widetilde{Y}^{(q)}$, respectively.

It is not difficult to show that

$$
V_0^{up^{up},K} \downarrow V_0^{up} \quad \text{and} \quad V_0^{up_{low},K} \uparrow V_0^{up} \quad \text{for} \quad K \to \infty, \qquad (10)
$$

for a proof see the Appendix.

As a third alternative, in view of (10), the estimators (8) and (9) can be combined into a convex family of new estimators,

$$
\widehat{V}_0^{\alpha,K_u,K_l,M} := \alpha \widehat{V}_0^{up^{up},K_u,M} + (1-\alpha) \widehat{V}_0^{up_{low},K_l,M}, \qquad (11)
$$

for $0 \leq \alpha \leq 1$, and suitably chosen simulation numbers K_u, K_l, M. In Section 6 we will demonstrate in practical examples that the combined estimator may have a much higher efficiency than either $\widehat{V}_0^{up^{up},K,M}$ or $\widehat{V}_0^{up_{low},K,M}$.

We here note that, essentially, the estimator (9) can also be found in Andersen & Broadie (2001) and Haugh & Kogan (2001).

Heuristic motivation of the combined estimator

In view of the Appendix we suppose that for some $\beta_u, \beta_l > 0$ the following expansions hold,

$$\begin{aligned} V_0^{up^{up},K} &= V_0^{up} + \frac{c_u}{K^{\beta_u}} + o(\frac{1}{K^{\beta_u}}), \quad c_u > 0 \qquad \text{and} \\ V_0^{up_{low},K} &= V_0^{up} - \frac{c_l}{K^{\beta_l}} + o(\frac{1}{K^{\beta_l}}), \quad c_l > 0. \end{aligned} \tag{12}$$

Let α, $0 < \alpha < 1$, be such that $\alpha c_u - (1-\alpha)c_l = 0$ and let κ_u, $\kappa_l > 0$ be such, that $\kappa_u \beta_u = \kappa_l \beta_l$. Consider for some integer K,

$$\widehat{V}_0^{\alpha,[K^{\kappa_u}],[K^{\kappa_l}]} := \alpha V_0^{up^{up},[K^{\kappa_u}]} + (1-\alpha) V_0^{up_{low},[K^{\kappa_l}]} = V_0^{up} + o(\frac{1}{K^{\kappa_u \beta_u}}), \tag{13}$$

with brackets denoting the Entier function. We then consider the complexity of the two estimators $\mathcal{U} := \widehat{V}_0^{up^{up},K,M}$ and $\mathcal{A} := \widehat{V}_0^{\alpha,[K^{\kappa_u}],[K^{\kappa_l}],M}$. As usual, the accuracy ε of an estimator $\widehat{s}$ for a target value p is defined via

$$\varepsilon^2 := E(\widehat{s} - p)^2 = Var(\widehat{s}) + (E\widehat{s} - p)^2$$

and so we may write by (12),(13),

$$\begin{aligned} \varepsilon_{\mathcal{U}}^2 &: = \frac{1}{M} Var(\widehat{V}_0^{up^{up},K,1}) + \frac{c_u^2}{K^{2\beta_u}} + o(\frac{1}{K^{2\beta_u}}), \\ \varepsilon_{\mathcal{A}}^2 &: = \frac{\alpha^2}{M} Var(\widehat{V}_0^{up^{up},[K^{\kappa_u}],1}) + \frac{(1-\alpha)^2}{M} Var(\widehat{V}_0^{up_{low},[K^{\kappa_l}],1}) + o(\frac{1}{K^{2\kappa_u\beta_u}}), \end{aligned}$$

where the simulation of the up-up and up-low estimator is assumed to be done independently.

Remark 4.1 In practice it is more efficient to localize $\widehat{\jmath}_{\max}$ using the samples for $V_0^{up^{up},K}$. Then, the up-up and up-low estimator are dependent in general, so

$$\varepsilon_{\mathcal{A}}^2 : = \frac{1}{M} Var\left(\alpha \widehat{V}_0^{up^{up},[K^{\kappa_u}],1} + (1-\alpha)\widehat{V}_0^{up_{low},[K^{\kappa_l}],1}\right) + o(\frac{1}{K^{2\kappa_u\beta_u}}).$$

Since $Var(\widehat{V}_0^{up^{up},K,1})$ and $Var\left(\alpha \widehat{V}_0^{up^{up},[K^{\kappa_u}],1} + (1-\alpha)\widehat{V}_0^{up_{low},[K^{\kappa_l}],1}\right)$ are uniformly bounded in K, we can deduce in the spirit of Schoenmakers & Heemink (1997) and Duffy & Glynn (1995) an asymptotically optimal tradeoff between bias and statistical error of the estimators $\mathcal{U}$ and $\mathcal{A}$. In fact, their bias and statistical error should be of comparable magnitude. For the up-up estimator $\mathcal{U}$ we thus take $K \propto M^{1/(2\beta_u)}$ yielding $\varepsilon_{\mathcal{U}}^2 \propto M^{-1}$, with $\propto$ denoting asymptotic equivalence, and so for the required computational costs to achieve an accuracy ε we have

$$\text{Cost}_{\mathcal{U}}(\varepsilon) \propto MK \propto M^{1+\frac{1}{2\beta_u}} \propto \frac{1}{\varepsilon^{2+1/\beta_u}}.$$

For a suitable choice of K_u, K_l for the combined estimator $\mathcal{A}$ we need to know a bit more about the bias term $o(K^{-\kappa_u\beta_u})$ in (13). Suppose we can identify a

$\gamma > 1$, preferably as large as possible, such that this bias term may be represented as $O(K^{-\gamma\kappa_u\beta_u})$. Then, by choosing $K \propto M^{1/(2\kappa_u\beta_u\gamma)}$ we obtain in a similar way $\varepsilon^2_{\mathcal{A}} = O(M^{-1})$ and

$$\mathrm{Cost}_{\mathcal{A}}(\varepsilon) \propto MK^{\max(\kappa_u,\kappa_l)} \propto M^{1+\frac{\max(\kappa_u,\kappa_l)}{2\kappa_u\beta_u\gamma}} \propto \frac{1}{\varepsilon^{2+\max(1/\beta_u,1/\beta_l)/\gamma}}.$$

So, under the assumptions above,

$$\frac{\mathrm{Cost}_{\mathcal{U}}(\varepsilon)}{\mathrm{Cost}_{\mathcal{A}}(\varepsilon)} \longrightarrow \infty \quad \text{as} \quad \varepsilon \downarrow 0, \qquad \text{if} \quad \frac{\beta_u}{\beta_l} < \gamma. \tag{14}$$

Remark 4.2 We see, that the complexity of the combined estimator $\mathcal{A}$ does only depend on the ratio $\kappa_l/\kappa_u = \beta_u/\beta_l$ and thus may take $\kappa_u = \min(1, \beta_l/\beta_u)$ and $\kappa_l = \min(1, \beta_u/\beta_l)$, such that $O(K)$ is always the order of the number of inner simulations.

The above analysis, which is build on some additional assumptions however, indicates why the combined estimator may be superior in several applications, see Section 6.

5 Two canonical approximative processes

In this section we consider two approximative processes for the general Bermudan style derivative which arise from two canonical exercise strategies.

Maximum of still alive European options

Suppose the option holder has arrived at a certain exercise date $\mathcal{T}_j$, $1 \le j \le k$, and looks which remaining underlying European instrument has the largest value. More precisely, he considers the index defined by

$$\widetilde{\tau}^{(j)} := \inf \left\{ m \ge j \;\middle|\; E^{\mathcal{F}^{(j)}}\left[\frac{C_{\mathcal{T}_m}}{B(\mathcal{T}_m)}\right] = \max_{j \le i \le k} E^{\mathcal{F}^{(j)}}\left[\frac{C_{\mathcal{T}_i}}{B(\mathcal{T}_i)}\right] \right\}. \tag{15}$$

This index is clearly $\mathcal{F}^{(j)}$-measurable and the option holder has the right to pin down his exercise policy at $\mathcal{T}_j$ for whatever reason, by deciding at $\mathcal{T}_j$ to exercise at $\mathcal{T}_{\widetilde{\tau}(j)}$. In fact, this is the same as selling the Bermudan at $\mathcal{T}_j$ as a European option with exercise date $\mathcal{T}_{\widetilde{\tau}(j)}$, thus receiving a cash amount of $\widetilde{Y}^{(j)}B(\mathcal{T}_j)$, with

$$\widetilde{Y}^{(j)} := \max_{j \le i \le k} E^{\mathcal{F}^{(j)}}\left[\frac{C_{\mathcal{T}_i}}{B(\mathcal{T}_i)}\right] = E^{\mathcal{F}^{(j)}}\left[\frac{C_{\mathcal{T}_{\widetilde{\tau}(j)}}}{B(\mathcal{T}_{\widetilde{\tau}(j)})}\right] \le Y^{(j)}. \tag{16}$$

The process $\widetilde{Y}$ in (16) is a lower estimation of the Snell envelope Y since the policy (15) is suboptimal. For instance, because the optimal policy is not $\mathcal{F}^{(j)}$-measurable.

Exercise when cash flow equals maximum of still alive European options

It is clear that exercising a Bermudan at a time where the cash flow is below the maximum price of the remaining underlying European options is never optimal. This suggests an alternative exercise strategy defined by the following stopping time,

$$\widehat{\tau}^{(j)} := \inf\left\{ m \geq j \;\middle|\; \frac{C_{\mathcal{T}_m}}{B(\mathcal{T}_m)} = \max_{m\leq i\leq k} E^{\mathcal{F}^{(m)}}\left[\frac{C_{\mathcal{T}_i}}{B(\mathcal{T}_i)}\right]\right\},$$

yielding a lower approximation of the Snell envelope,

$$\widehat{Y}^{(j)} := E^{\mathcal{F}^{(j)}}\left[\frac{C_{\mathcal{T}_{\widehat{\tau}^{(j)}}}}{B(\mathcal{T}_{\widehat{\tau}^{(j)}})}\right] \leq Y^{(j)}. \tag{17}$$

In fact, for the Bermudan swaption (see Section 6) the process $\widehat{Y}$ coincides with the lower estimation of Andersen (1999) obtained by Andersen's Strategy 2 with $H = 0$.

The exercise policy $\widehat{\tau}$ is better than $\widetilde{\tau}$, due to the following proposition.

Proposition 5.1 *For each $j = 0, \ldots, k$ it holds,*

$$\widetilde{Y}^{(j)} \leq \widehat{Y}^{(j)} \leq Y^{(j)}.$$

Proof. We only need to show the first inequality, which we will proof by induction. When $j = k-1$, we clearly have the equality

$$\widetilde{Y}^{(k-1)} = \widehat{Y}^{(k-1)}.$$

Suppose the inequality holds for some j. Then, it follows that

$$\begin{aligned}
\widehat{Y}^{(j-1)} &= E^{\mathcal{F}^{(j-1)}}\left[\frac{C_{\mathcal{T}_{\widehat{\tau}^{(j-1)}}}}{B(\mathcal{T}_{\widehat{\tau}^{(j-1)}})}\right] \\
&= E^{\mathcal{F}^{(j-1)}}\left[\frac{C_{\mathcal{T}_{j-1}}}{B(\mathcal{T}_{j-1})}\cdot 1_{\widehat{\tau}^{(j-1)}=\mathcal{T}_{j-1}} + \frac{C_{\mathcal{T}_{\widehat{\tau}^{(j)}}}}{B(\mathcal{T}_{\widehat{\tau}^{(j)}})}\cdot 1_{\widehat{\tau}^{(j-1)}>\mathcal{T}_{j-1}}\right] \\
&= \frac{C_{\mathcal{T}_{j-1}}}{B(\mathcal{T}_{j-1})}\cdot 1_{\widehat{\tau}^{(j-1)}=\mathcal{T}_{j-1}} + E^{\mathcal{F}^{(j-1)}} E^{\mathcal{F}^{(j)}}\left[\frac{C_{\mathcal{T}_{\widehat{\tau}^{(j)}}}}{B(\mathcal{T}_{\widehat{\tau}^{(j)}})}\right]\cdot 1_{\widehat{\tau}^{(j-1)}>\mathcal{T}_{j-1}} \\
&\geq \max_{j-1\leq i\leq k} E^{\mathcal{F}^{(j-1)}}\left[\frac{C_{\mathcal{T}_i}}{B(\mathcal{T}_i)}\right]\cdot 1_{\widehat{\tau}^{(j-1)}=\mathcal{T}_{j-1}} \\
&\quad + E^{\mathcal{F}^{(j-1)}} \max_{j\leq i\leq k} E^{\mathcal{F}^{(j)}}\left[\frac{C_{\mathcal{T}_i}}{B(\mathcal{T}_i)}\right]\cdot 1_{\widehat{\tau}^{(j-1)}>\mathcal{T}_{j-1}} \\
&\geq \max_{j-1\leq i\leq k} E^{\mathcal{F}^{(j-1)}}\left[\frac{C_{\mathcal{T}_i}}{B(\mathcal{T}_i)}\right]\cdot 1_{\widehat{\tau}^{(j-1)}=\mathcal{T}_{j-1}} \\
&\quad + \max_{j-1\leq i\leq k} E^{\mathcal{F}^{(j-1)}}\left[\frac{C_{\mathcal{T}_i}}{B(\mathcal{T}_i)}\right]\cdot 1_{\widehat{\tau}^{(j-1)}>\mathcal{T}_{j-1}} \\
&= \widetilde{Y}^{(j-1)}.
\end{aligned}$$

■

Remark 5.2 In the above derivation we have used a crucial property of $\widehat{\tau}$, namely, it holds $\widehat{\tau}^{(j-1)} \neq \mathcal{T}_{j-1} \Longrightarrow \widehat{\tau}^{(j-1)} = \widehat{\tau}^{(j)}$. Without proof we note that this property does not hold for $\widetilde{\tau}$.

6 Application: Bermudan swaptions in the LIBOR market model

We consider the LIBOR Market Model with respect to a tenor structure $0 < T_1 < T_2 < \ldots < T_n$ in the *spot LIBOR measure* P^*, induced by the numeraire

$$B^*(t) := \frac{B_{m(t)}(t)}{B_1(0)} \prod_{i=0}^{m(t)-1} (1 + \delta_i L_i(T_i))$$

with $m(t) := \min\{m : T_m \geq t\}$ denoting the next reset date at time t. The dynamics of the forward LIBOR $L_i(t)$, defined in the interval $[0, T_i]$ for $1 \leq i < n$, is governed by the following system of SDE's (Jamshidian 1997),

$$dL_i = \sum_{j=m(t)}^{i} \frac{\delta_j L_i L_j \gamma_i \cdot \gamma_j}{1 + \delta_j L_j} dt + L_i \gamma_i \cdot dW^*. \tag{18}$$

Here $\delta_i = T_{i+1} - T_i$ are day count fractions, and

$$t \to \gamma_i(t) = (\gamma_{i,1}(t), \ldots, \gamma_{i,d}(t))$$

are deterministic volatility vector functions defined in $[0, T_i]$, called factor loadings. In (18), $(W^*(t) \mid 0 \leq t \leq T_{n-1})$ is a standard d-dimensional Wiener process under the measure P^* with d, $1 \leq d < n$, being the number of driving factors.

For our experiments we take the following volatility structure:

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

is a parametric volatility function proposed by Rebonato (1999), and e_i are d-dimensional unit vectors, decomposing some input correlation matrix of rank d. For generating LIBOR models with different numbers of factors d, we take as a basis a correlation structure of the form

$$\rho_{ij} = \exp(-\varphi|i-j|); \qquad i, j = 1, \ldots, n-1 \tag{19}$$

which has full-rank for $\varphi > 0$, and then for a particular choice of d we deduce from ρ a rank-d correlation matrix ρ^d with decomposition $\rho^d_{ij} = e_i \cdot e_j$, $1 \leq i, j < n$, by principal component analysis. We note that instead of (19) it is possible to use more general and economically more realistic correlation structures. For instance the parametric structures of Schoenmakers & Coffey (2003).

We will take a flat 10% initial LIBOR curve over a 40 period quarterly tenor structure and choose values of the parameters c, a, b, g_∞, φ, such that the involved correlation structure and scalar volatilities can be regarded as typical for a Euro or GBP market. We take

$$n = 41,\ \delta_i = 0.25,\ c = 0.2,\ a = 1.5,\ b = 3.5,\ g_\infty = 0.5,\ \varphi = 0.0413. \tag{20}$$

For a "practically exact" numerical integration of the SDE (18), we used the log-Euler scheme with $\Delta t = \delta/5$ (e.g., see also Kurbanmuradov, Sabelfeld and Schoenmakers 2002).

Let us now briefly recall the definition of a (payer) swaption over a period $[T_i, T_n]$, $1 \le i \le k$. A *swaption contract* with maturity T_i and strike θ with principal \$1 gives the right to contract at T_i for paying a fixed coupon θ and receiving floating LIBOR at the settlement dates $T_{i+1}, \ldots, T_n$. So by this definition, its cashflow at maturity is

$$S_{i,n}(T_i) := \left(\sum_{j=i}^{n-1} B_{j+1}(T_i)\delta_j \left(L_j(T_i) - \theta\right) \right)^+ .$$

In this section we consider Bermudan swaptions for which we assume for simplicity that the exercise dates coincide with the LIBOR tenor structure. I.e. $k = n$ and $\mathcal{T}_i = T_i$, for $1 \le i \le n$.

A *Bermudan swaption*, issued at $t = 0$, gives the right to exercise a cashflow

$$C_{T_\tau} := S_{\tau,n}(T_\tau)$$

at an exercise date $T_\tau \in \{T_1, \ldots, T_n\}$ to be decided by the option holder (see also Section 2). The value of the Bermudan swaption, issued at $t = 0$, is given by (1).

We now investigate the bias of the upper bound estimators (8) and (9) for different Bermudan swaptions in the LIBOR market model (18) with flat 10% initial yield curve and model parameters given by (20). As lower approximation of the Snell envelope process we take the maximum of still alive swaption process $\widetilde{Y}_{\max}$. Hence, we have (16), where the European option is now a European swaption,

$$\widetilde{Y}_{\max}^{(j)} = \max_{j \le i \le k} \frac{S_{i,n}(T_j)}{B^*(T_j)} \quad \text{with} \quad S_{i,n}(T_j) = B^*(T_j) E^{\mathcal{F}^{(j)}} \left[\frac{S_{i,n}(T_i)}{B^*(T_i)} \right]. \tag{21}$$

We further assume that expansion (12) holds true, hence,

$$\begin{aligned} V_0^{up^{up},K} - V_0^{up} &= \frac{c_u}{K^{\beta_u}} + o(\frac{1}{K^{\beta_u}}), \\ V_0^{up} - V_0^{up_{low},K} &= \frac{c_l}{K^{\beta_l}} + o(\frac{1}{K^{\beta_l}}), \qquad \beta_u, \beta_l, c_u, c_l > 0, \end{aligned} \tag{22}$$

and aim to identify the parameters β_u, β_l, c_u, c_l in particular cases.

We compute $V_0^{up_{low},K}$ and $V_0^{up^{up},K}$ by estimators (8) and (9), respectively, with $K = 2^2, 2^3, \ldots, 2^7$) and $M = 30000$, for the examples in Table 1. For $M = 30000$ the standard deviations of both estimators are less than 1.5% relative, for all considered K. For $K = 128$, the relative distance between $\widehat{V}_0^{up_{low},K,30000}$ and $\widehat{V}_0^{up^{up},K,30000}$ turns out to be within 1.5%, hence the relative standard deviation of both estimators. So we conclude that within a relative accuracy of 1.5% in this sense, both estimators $\widehat{V}_0^{up_{low},128,30000}$ and $\widehat{V}_0^{up^{up},128,30000}$ give a good approximation of the target upper bound V_0^{up}. Therefore, we take their average $\widehat{V}_0^{1/2,128,128,30000}$ as an approximation of V_0^{up}.

With regard to (22) we next determine the coefficients $\beta_u, \beta_l, c_u, c_l$ by linear regression, hence the following minimizations,

$$RMS_u^{rel} = \tag{23}$$
$$\sqrt{\sum_{i=2}^{6}\left(\frac{\log(\widehat{V}_0^{up^{up},2^i,30000}-\widehat{V}_0^{1/2,128,128,30000})-(\log c_u-\beta_u\log 2^i)}{\log(\widehat{V}_0^{up^{up},2^i,30000}-\widehat{V}_0^{1/2,128,128,30000})}\right)^2}\longrightarrow \min_{\beta_u,c_u}$$

and

$$RMS_l^{rel} = \tag{24}$$
$$\sqrt{\sum_{i=2}^{6}\left(\frac{\log(\widehat{V}_0^{1/2,128,128,30000}-\widehat{V}_0^{up_{low},2^i,30000})-(\log c_l-\beta_l\log 2^i)}{\log(\widehat{V}_0^{1/2,128,128,30000}-\widehat{V}_0^{up_{low},2^i,30000})}\right)^2}\longrightarrow \min_{\beta_l,c_l},$$

by straightforward differentiating. We note that in the linear regressions (23) and (24) we exclude the terms due to $i = 7$, since for $i = 7$ the denominators in(23) and (24) are basically zero within the considered accuracy. The values of β_u, c_u, β_l, c_l, obtained for different types of swaptions and different number of factors d, are given in Table 1. We also show in Table 1 the "optimal" $\alpha = c_l/(c_u + c_l)$ and ratios $\beta_l/\beta_u = \kappa_u/\kappa_l$.

Conclusion 6.1 (Table 1) According to Table 1, the function $\log(V_0^{up^{up},K} - V_0^{up})$ and $\log(V_0^{up_{low},K} - V_0^{up})$ can be approximated rather close by $\log c_u - \beta_u \log K$ and $\log c_l - \beta_l \log K$, respectively, within errors which do not exceed 3.0%. Hence plotting $\log K \to \log(V_0^{up^{up},K} - V_0^{up})$ and $\log K \to \log(V_0^{up^{low},K} - V_0^{up})$ gives approximately straight lines. See Figs. 1–2 for $d = 40$ (full factor model) and out-of-the-money swaptions with strike $\theta = 12\%$. The values of β_u turn out to be roughly equal to one whereas over all β_l seem to be significantly smaller than one. It would be interesting to see an explanation for this. Then, it is remarkable that the optimal value of α for different strikes and number of factors does not vary too much. The same applies for β_u and β_l and we so propose for all examples the combined upper bound estimator

$$\widehat{V}_0^{0.4,[K^{0.87}],K,M} = 0.4\widehat{V}_0^{up^{up},[K^{0.87}],M} + 0.6\widehat{V}_0^{up_{low},K,M}, \tag{25}$$

where $\alpha = 0.4$ is roughly the average value in Table 1, and $\kappa_u = 0.87$ and $\kappa_l = 1$ are based on the average of β_l/β_u and taking into account Remark 4.2. In Fig. 3

we show for a particular example, strike $\theta = 0.12$ (OTM) and $d = 40$, a plot of the estimator (25) together with $\widehat{V}_0^{up^{up},[K^{0.87}],30000}$ and $\widehat{V}_0^{up_{low},K,30000}$ for different values of K. Note that even for any K the bias of the combined estimator is negligible within the given accuracy in this example. Later (in Table 2) we will see that the bias of the estimator (25) for the particular choices of $\alpha, \kappa_u, \kappa_l$, is negligible also for all other examples in Table 1, when $K \geq 4$.

We now compare the combined estimator (25) with the up-up estimator (9) for different strikes and different number of factors d. We consider $\widehat{V}_0^{0.4,4,5,90000}$ and $\widehat{V}_0^{up^{up},100,30000}$, where the respective choices of K and M are determined by experiment, such that both the estimations and the (absolute) standard deviations of the estimators are close for different strikes and different number of factors. The results are given in Table 2, columns 5,6. It is easily seen that the combined estimator $\widehat{V}_0^{0.4,4,5,90000}$ is almost 4 times faster than the up-up estimator $\widehat{V}_0^{up^{up},100,30000}$.

Remark 6.2 In general, depending on the quality of the Snell-envelope approximation, higher accuracies for dual upper bound estimations may be required and then the efficiency gain of the combined up-low estimator (25) with respect to up-up estimator (9) can become tremendous in view of (14).

Now we are going to compare the up-up estimations $V_0^{up^{up},K}$, based on the maximum of still alive swaption process (21), with up-up estimations considered by Andersen & Broadie (2001), denoted by $V_{0,AB}^{up^{up},K}$. The latter estimations are due to an approximative lower bound process $\widetilde{Y}_A$, obtained via a particular exercise boundary which is constructed by strategy 1 of the Andersen method. The process $\widetilde{Y}_A$ has the following form,

$$\widetilde{Y}_A^{(j)} := E^{\mathcal{F}^{(j)}} \left[\frac{S_{\tau_A^{(j)},n}(T_{\tau_A^{(j)}})}{B^*(T_{\tau_A^{(j)}})} \right], \qquad \text{with} \quad \tau_A^{(j)} := \inf \left\{ m \geq j \mid \frac{S_{m,n}(T_m)}{B^*(T_m)} > H_m \right\}.$$

The sequence of constants H_m is pre-computed by the method of Andersen using strategy 1, see Andersen (1999). We compute $\widehat{V}_{0,AB}^{up^{up},100,10000}$ for different strikes and number of factors, and the results are given in Table 2, column 4. As we can see, the values of $\widehat{V}_{0,AB}^{up^{up},100,10000}$ and $\widehat{V}_0^{up^{up},100,30000}$ are rather close. In fact, except for the ATM strikes in the 1 and 2 factor model, the differences do not exceed 1% relative. For a full factor model and a particular OTM strike we also compare the estimators $\widehat{V}_{0,AB}^{up^{up},K,10000}$ and $\widehat{V}_0^{up^{up},K,30000}$ for different numbers of inner simulations, $K = 1, \ldots, 100$, and conclude that both estimators coincide within the considered accuracy, see Fig. 2.

In Table 2, column 3, we give lower bounds of Bermudan prices $B^*(0)\widetilde{Y}_A^{(0)}$, due to the stopping time $\tau_A^{(0)}$. We see that in case of a 1-factor model the distance between the lower and upper bound of the Bermudan swaption price is rather close for OTM, ATM as well as for ITM strikes. This observation is consistent with the results reported in Andersen & Broadie (2001). For more than one factor this distance appears to increase.

In Table 3 we list the required computational time of the up-up upper bound estimators due to Andersen & Broadie and the process given by the maximum of still alive swaptions. For practical relevance, we required an accuracy of 1% and used the Euler scheme with time steps $\Delta t = \delta$. We do not take into account the cost of the pre-computation of the exercise strategy, which is small compared with the cost of the upper estimators. For further details we refer to Andersen (1999). We conclude that for ATM and OTM strikes $\widetilde{Y}_{\max}$ gives rise to a faster method than a method due to $\widetilde{Y}_A$ (the lower bound process of Andersen). This is caused by the fact that for simulating $\widetilde{Y}_A$ one needs to construct a LIBOR trajectory starting at T_j until the exercise condition is fulfilled (for the description of the algorithm see Andersen & Broadie, (2001)).

Regarding the rather high computation times in Table 3, it is clear that an efficiency gain of about a factor 4 (or maybe more), due to application of the in this paper presented combined upper bound estimator, is very desirable in practice. Moreover, for a particular Bermudan product we recommend the following procedure. *Carry out a pre-computation of the optimal β_u, β_l and α for the given structure, based on up-up and up-low estimations with lower accuracy. Next, take the number K of inner simulations as small as possible and then choose the number M of outer simulations according to the accuracy required.* For example see Fig. 3, where the involved parameters β_u, β_l and α are optimal for the example under consideration and where K can be taken equal to one in fact.

We end with two final remarks.

Remark 6.3 Naturally, the numerical analysis based on the (discounted) maximum of still alive swaption process in this section could also be done for the process (17) in Section 5. This process is in fact consistent with strategy 2, $H = 0$ in Andersen (1999). So, on the one hand, this process is dominated from above by a lower bound process due to strategy 2 with an optimized H. On the other hand, however, as Andersen reports and we found out also, strategy 2 with optimized H performs not substantially better than strategy 1 with optimized H. Therefore, it is to expect that the dual upper bound due to process (17) will be more or less comparable with the upper bound due to $\widetilde{Y}_A^{(0)}$ in this section, which in turn is comparable with the upper bound due to (16) for a more than one factor model. Moreover, it is easily seen that the computation of the dual upper bound by the process (17) will be more costly.

Remark 6.4 Recently, Jamshidian (2003) has constructed a new dual method for American/Bermudan upper bounds, which is based on a multiplicative version of the Doob-Meyer decomposition of some approximative process. The computational aspects of Jamshidian's method will be part of our further study. Also the application of our presented method to so called Israeli options, Bermudans which are cancelable from the issuers side, would be interesting (see Kühn & Kyprianou (2003)).

Table 1. Estimated parameters of (12) for different swaptions and numbers of factors.

θ	d	β_u	c_u	RMS_u^{rel}	β_l	c_l	RMS_l^{rel}	β_l/β_u	α
	1	1.021	0.038	0.016	0.850	0.023	0.026	0.83	0.381
0.08	2	0.991	0.032	0.016	0.862	0.022	0.019	0.87	0.404
(ITM)	10	0.940	0.026	0.003	0.893	0.020	0.023	0.95	0.436
	1	0.970	0.025	0.021	0.746	0.013	0.015	0.77	0.335
0.10	2	0.872	0.020	0.009	0.840	0.014	0.029	0.96	0.417
(ATM)	10	0.968	0.021	0.016	0.717	0.010	0.020	0.74	0.317
	1	0.988	0.099	0.015	0.801	0.006	0.017	0.81	0.363
0.12	2	0.946	0.008	0.007	0.872	0.006	0.016	0.93	0.442
(OTM)	10	0.930	0.007	0.009	0.896	0.006	0.013	0.96	0.460
	40	1.035	0.008	0.029	0.900	0.005	0.019	0.87	0.405

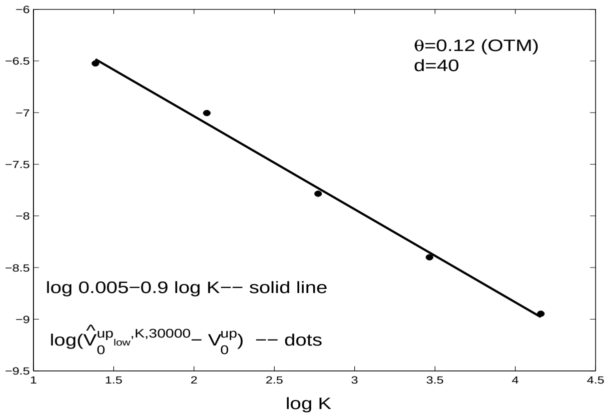

Fig. 1. $\log(\widehat{V}_0^{up_{low},K,30000} - V_0^{up})$ and $\log c_l - \beta_l \log K$ for β_l and c_l minimizing (24).

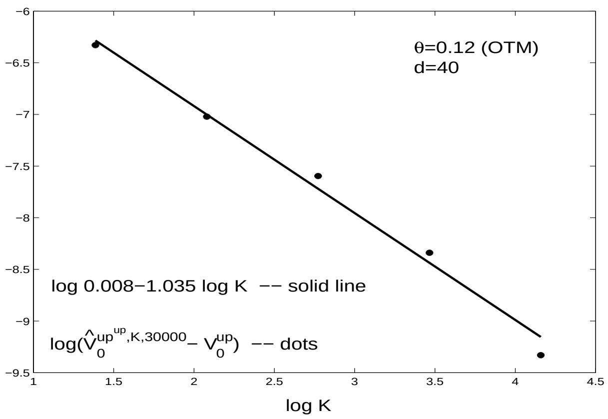

Fig. 2. $\log(\widehat{V}_0^{up^{up},K,30000} - V_0^{up})$ and $\log c_u - \beta_u \log K$ for β_u and c_u minimizing (23).

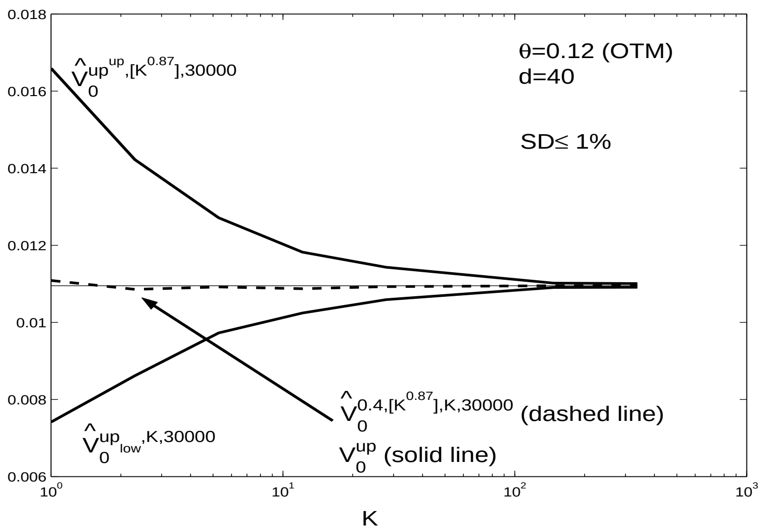

Fig. 3. Different estimators for V_0^{up} due to the approximative process $\widetilde{Y}_{\max}$.

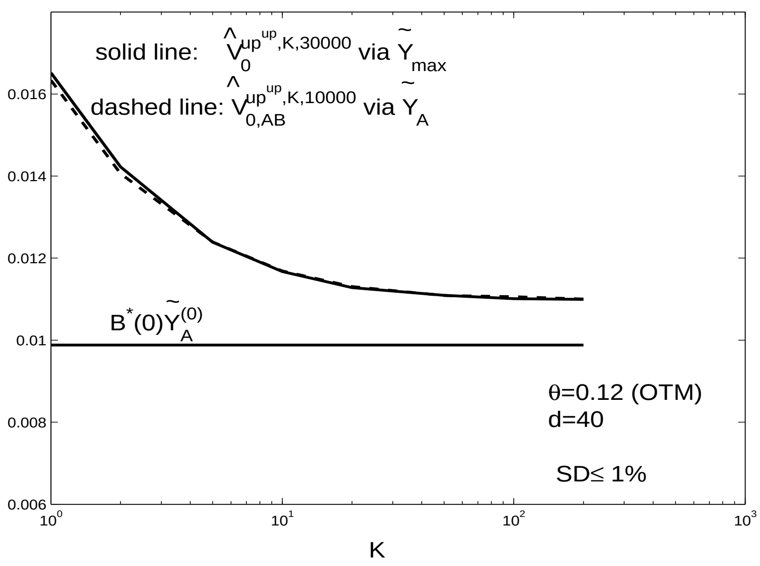

Fig. 4. Up-up estimators due to different approximative processes.

Table 2. (all values to be multiplied by 10^{-4})

θ	d	$B^*(0)\widetilde{Y}_A^{(0)}$ (SD)	$\widehat{V}_{0,AB}^{up^{up},100,10000}$ (SD)	$\widehat{V}_0^{up^{up},100,30000}$ (SD)	$\widehat{V}_0^{0.4,4,5,90000}$ (SD)
0.08 (ITM)	1	1116.2(1.6)	1121.4(0.1)	1128.8(0.3)	1124.8(0.4)
	2	1103.2(1.4)	1117.6(0.4)	1121.1(0.3)	1117.6(0.3)
	10	1097.1(1.3)	1111.0(0.4)	1113.7(0.3)	1109.5(0.3)
	40	1093.2(1.3)	1106.9(0.4)	1110.1(0.3)	1106.9(0.3)
0.10 (ATM)	1	403.3(1.2)	408.3(0.1)	416.5(0.5)	416.1(0.5)
	2	372.6(1.1)	394.0(0.4)	397.3(0.5)	397.7(0.4)
	10	347.4(1.0)	373.6(0.5)	375.8(0.4)	375.3(0.4)
	40	341.6(1.0)	367.5(0.5)	368.5(0.4)	369.8(0.4)
0.12 (OTM)	1	133.5(0.7)	135.4(0.1)	136.3(0.4)	135.5(0.3)
	2	119.7(0.7)	127.4(0.3)	127.5(0.3)	126.5(0.3)
	10	102.8(0.6)	113.6(0.3)	114.5(0.3)	113.2(0.3)
	40	98.8(0.5)	110.3(0.3)	109.6(0.3)	108.6(0.3)

Table 3. Computation time[1] (in sec.) of the upper estimators in Table 2.

θ	d	$\widehat{V}_{0,AB}^{up^{up},100,1000}$	$\widehat{V}_0^{up^{up},100,3000}$
0.08 (ITM)	1	59	115
	2	69	134
	10	183	241
	40	468	603
0.10 (ATM)	1	166	113
	2	213	134
	10	510	239
	40	1467	598
0.12 (OTM)	1	229	115
	2	299	145
	10	718	263
	40	2076	625

[1]The simulations are run on a 1 GHz processor

Appendix: Proof of the convergence property

$$
\begin{aligned}
\frac{V_0^{up^{up},K}}{B(0)} &= E \sup_{1\le j\le k} [\frac{C_{\mathcal{T}_j}}{B(\mathcal{T}_j)} - \sum_{q=1}^{j} \widetilde{Y}^{(q)} + \sum_{q=1}^{j} \frac{1}{K} \sum_{i=1}^{K} \xi_i^{(q)}] \\
&= E[\frac{C_{\mathcal{T}_{\widehat{j}_{\max}}}}{B(\mathcal{T}_{\widehat{j}_{\max}})} - \sum_{q=1}^{\widehat{j}_{\max}} \widetilde{Y}^{(q)} + \sum_{q=1}^{\widehat{j}_{\max}} \frac{1}{K} \sum_{i=1}^{K} \xi_i^{(q)}] \\
&= E\,(1 - 1_{[j_{\max}\neq \widehat{j}_{\max}]})[\frac{C_{\mathcal{T}_{j_{\max}}}}{B(\mathcal{T}_{j_{\max}})} - \sum_{q=1}^{j_{\max}} \widetilde{Y}^{(q)} + \sum_{q=1}^{j_{\max}} \frac{1}{K} \sum_{i=1}^{K} \xi_i^{(q)}] \\
&+ \; E\,1_{[j_{\max}\neq \widehat{j}_{\max}]}[\frac{C_{\mathcal{T}_{\widehat{j}_{\max}}}}{B(\mathcal{T}_{\widehat{j}_{\max}})} - \sum_{q=1}^{\widehat{j}_{\max}} \widetilde{Y}^{(q)} + \sum_{q=1}^{\widehat{j}_{\max}} \frac{1}{K} \sum_{i=1}^{K} \xi_i^{(q)}] \\
&= \frac{V_0^{up}}{B(0)} + O((P(j_{\max} \neq \widehat{j}_{\max}))^{1-1/p_1})
\end{aligned}
$$

for any integer p_1, by Hölder's inequality and the fact that for any p_1 the p_1-th moment of both

$$
\begin{aligned}
&\frac{C_{\mathcal{T}_{\widehat{j}_{\max}}}}{B(\mathcal{T}_{\widehat{j}_{\max}})} - \sum_{q=1}^{\widehat{j}_{\max}} \widetilde{Y}^{(q)} + \sum_{q=1}^{\widehat{j}_{\max}} \frac{1}{K} \sum_{i=1}^{K} \xi_i^{(q)} \\
&\qquad\qquad\qquad\text{and} \\
&\frac{C_{\mathcal{T}_{j_{\max}}}}{B(\mathcal{T}_{j_{\max}})} - \sum_{q=1}^{j_{\max}} \widetilde{Y}^{(q)} + \sum_{q=1}^{j_{\max}} \frac{1}{K} \sum_{i=1}^{K} \xi_i^{(q)}
\end{aligned} \tag{26}
$$

exist and are uniformly bounded in K (we omit the proof). Then, since $\lim_{K\to\infty} P(j_{\max} \neq \widehat{j}_{\max}) = 0$, the convergence for $K \to \infty$ of $V_0^{up^{up},K} \to V_0^{up}$ follows.

Similarly, we can show that

$$
V_0^{up_{low},K} = \frac{V_0^{up}}{B(0)} + O((P(j_{\max} \neq \widehat{j}_{\max}))^{1-1/q_1}),
$$

for any integer q_1, hence $V_0^{up_{low},K} \to V_0^{up}$.

References

[1] Andersen L. (1999): A simple approach to the pricing of Bermudan swaptions in the multi-factor LIBOR market model. *Journal of Computational Finance*, **3**, No. 2, 5–32.

[2] Andersen L., Broadie M. (2001): A primal-dual simulation algorithm for pricing multidimensional American options. Working paper.

[3] Broadie M., Glasserman P. (1997): A stochastic mesh method for pricing high-dimensional American options. PaineWebber Working Papers in Money, Economic and Finance #PW9804, Columbia Business School, New York.

[4] Broadie M., Glasserman P., Ha Z. (2000): Pricing American options by simulations using a stochastic mesh with optimized weights, *Probabilistic Constrained Optimization: Methodology and Applications*, S. Uryasev, ed., Kluwer Publishers, Norwell, Mass, 32–50.

[5] Duffie D. (2001): *Dynamic Asset Pricing Theory.* Third Edition, Princeton University Press, Princeton, New Jersey.

[6] Duffie D., Glyn P. (1995): Efficient Monte Carlo simulation of security prices. *The Annals of Applied Probability,* **5**, 897–905.

[7] Glasserman P. (2003): *Monte Carlo Methods in Financial Engineering.* Springer-Verlag, New York Berlin Heidelberg.

[8] Haugh M.B., Kogan L. (2001): Pricing American options: a duality approach. *Operations Research*, to appear.

[9] Ikeda N., Watanabe S. (1981): *Stochastic Differential Equations and Diffusion Processes.* North-Holland/Kodansha.

[10] Jamshidian F. (1997): LIBOR and swap market models and measures. *Finance and Stochastics*, **1**, 293–330.

[11] Jamshidian F. (2003): Minimax Optimality of American and Bermudan Claims and their Monte Carlo Valuation. Working Paper.

[12] Joshi M.S., Theis J. (2002): Bounding Bermudan swaptions in a swap-rate market model. *Quantitative Finance*, **2**, No. 5, 371–361.

[13] Kühn C., Kyprianou A.E. (2003): Pricing Israeli options: a pathwise approach. Working paper

[14] Kurbanmuradov O., Sabelfeld K., Schoenmakers J. (2002): Lognormal approximations to LIBOR market models. *Journal of Computational Finance*, **6**, No. 1, 69–100.

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

[16] Rebonato R. (1999): *Volatility and Correlation.* John Wiley & Sons.

[17] Rogers L.G.G. (2001): Monte Carlo valuation of American options. *Mathematical Finance*, **12**, 271–286.

[18] Schoenmakers J. & Coffey B. (2003): Systematic Generation of Parametric Correlation Structures for the LIBOR Market Model. *International Journal of Theoretical and Applied Finance*, **6**, No. 5, 507-519.

[19] Schoenmakers J.G.M., Heemink A.W. (1997): Fast valuation of financial derivatives. *Journal of Computational Finance*, **1**, No. 1, 47–62.