Efficient Control Variates and Strategies for Bermudan Swaptions in a Libor Market Model

Malene Shin Jensen  
Department of Management  
University of Aarhus  
e-mail: msjensen@econ.au.dk

Mikkel Svenstrup  
Department of Finance  
Aarhus School of Business  
e-mail: msv@asb.dk

28-October 2002  
(First version October 2001)

Abstract

In this paper we apply various improvement techniques on a Monte Carlo based valuation algorithm for Bermudan swaptions in a Libor market model. We suggest a simplification of the Andersen (2000) exercise strategy and find it to much more efficient. Furthermore, we test a range of control variates for Bermudan swaptions using a control variate technique for American options proposed in Rasmussen (2002). Application of these efficiency improvements in the Primal-Dual simulation algorithm of Andersen & Broadie (2001), improves both upper and lower bounds for the price estimates. For the Primal-Dual simulation algorithm we examine the variance-bias tradeoff between the number of outer and inner paths. Finally, we examine the effect of stochastic volatility on the potential losses of following the simplest exercise strategy.

JEL Codes: G12; G13; E43;

Keywords: Bermudan Swaptions; Control Variates; Exercise Strategy; Primal-Dual Algorithm; Stochastic Volatility;

1 Introduction

For a long time valuation of options with early exercise features or other free boundary problems by simulation was considered impossible. Researchers and practitioners have been focusing on lattice methods such as binomial trees and finite difference methods whenever these American style contingent claims were encountered. It is well known that lattice methods suffer from a curse of dimensionality. This means that valuation of assets with payoffs depending on more than two, or no more than three, state variables is considered to be unfeasible.
This is the case for options on several assets, interest rate models with a large number of state variables and models driven by several factors.

A key contribution to the solution to this problem was given in Broadie & Glasserman (1997), where a simulation algorithm providing asymptotically unbiased upper and lower bounds on the option value was presented. The problem with this method is that it is difficult to generalize and requires that there are few exercise decisions. However, recently new algorithms have been developed.

Haugh & Kogan (2001) present the value of an American option as the solution to a dual minimization problem over all super martingales, so that any given super martingale will result in an upper bound for the option. Andersen & Broadie (2001) recognize that the use of a martingale instead of a submartingale in the dual problem expressed in Haugh & Kogan (2001) will result in tighter upper bounds and they propose an improved simulation algorithm. A primal solution algorithm that provides an exercise strategy and hence an estimate of the lower bound is input to this new Primal-Dual simulation algorithm. Any primal algorithm can be applied, for example Least-Square Monte-Carlo (Longstaff & Schwartz (2001)), nonparametric methods (Andersen (2000)), low-dimensional lattice methods (Longstaff, Santa-Clara & Schwartz (2001b)), and others.

Andersen & Broadie (2001) demonstrate the simplicity and efficiency of the simulation algorithm in several examples, including multi-asset equity options and Bermudan swaptions. However, only little emphasis was placed on numerical efficiency and variance reduction techniques.

For the purpose of pricing American style contingent claims in a Monte Carlo framework, Rasmussen (2002) develops an extension of traditional control variate techniques, in which sampling of the controls occurs at the time of exercise of the American option and demonstrates the efficiency in the multi-asset equity option case.

The primary contributions of our paper are the following. First, we demonstrate that a simplification of the exercise strategies proposed in Andersen (2000) proves much more efficient in the Bermudan swaption case. Due to the structure of Bermudan swaptions, the most valuable core European swaption will almost always be the first to mature, and we show that ignoring the rest is computationally more efficient but results in the same prices. Secondly, we test the efficiency of a range of control variates with respect to Bermudan swaptions using the Rasmussen (2002) sampling algorithm. We illustrate how to apply it using dividend paying assets. Thirdly, we consider the rate of convergence of the Primal-Dual algorithm with an emphasis of the variance-bias trade-off due to nested simulations. When combining the simplified strategies, the control variates with a better variance-bias trade-off we experience significant improvements, and more specifically we get lower duality gaps, lower standard deviations in less time.

Finally, we implement a version of the Libor market model with stochastic volatility in order to examine the effect on Bermudan swaptions. As expected we find that the duality gaps are increasing in variance of the volatility and decreasing in mean reversion. Our work is also related to Bjerregaard Pedersen (1999) who applies the Broadie & Glasserman (1997) method to Bermudan
Swaptions.

2 The Optimal Stopping Problem

The general model setup, notation and problem statement are laid out in this section. Furthermore, we briefly summarize theoretical results required in the rest of the paper.

We assume that we have a dynamically complete financial market in which all uncertainty is described by a standard filtered probability space \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq T}, P)\).

The information set \(\mathcal{F}_t\) is defined as the natural filtration generated by a multi-dimensional Wiener process which drives all asset prices until some fixed time \(T\) augmented with the usual null sets. We assume the existence of a \(P\) equivalent probability measure \(Q\) under which asset prices discounted with the numeraire asset \(B(t)\) are martingales. Let \(E^Q_t(\cdot) = E^Q_t(\cdot | \mathcal{F}_t)\) denote the expectation conditional on information available at time \(t\). Hence, the time \(t\) value of any asset \(Q(t)\) is given by

\[
Q(t) = E^Q_t \left( \frac{Q(s)}{B(s)} \right), \quad t \leq s.
\]

American options are options with an early exercise feature and are characterized by the payoff \(h(t)\) paid upon exercise and the set of dates \(T\) on which the holder is allowed to exercise. American options, which can only be exercised on a discrete set of dates, are often referred to as Bermudan options.

The problem of valuing an American option is usually posed as an optimal stopping problem. Following Andersen & Broadie (2001) and Haugh & Kogan (2001) we will denote the following the primal problem

\[
\frac{Q(t)}{B(t)} = \sup_{\tau \in \Gamma(t)} E^Q_t \left( \frac{h(\tau)}{B(\tau)} \right),
\]

where \(\Gamma(t)\) denotes the set of optional stopping times \(\tau\) taking values in \(T(t) = T \cap [t, T]\), \(T\) being the maturity of the option. Notice that any optional stopping time belonging to \(\Gamma(t)\) will generate a lower bound on the true option price.

Andersen & Broadie (2001) define the dual problem to (1) by first noting that a valid upper bound can be found by using any adapted martingale \(\pi_t\)

\[
\frac{Q(0)}{B(0)} = \sup_{\tau \in \Gamma(0)} E^Q_0 \left( \frac{h(\tau)}{B(\tau)} + \pi_\tau - \pi_\tau \right)
\]

\[
= \pi_0 + \sup_{\tau \in \Gamma(0)} E^Q_0 \left( \frac{h(\tau)}{B(\tau)} - \pi_\tau \right)
\]

\[
\leq \pi_0 + E^Q_0 \left( \max_{t \in T(0)} \left( \frac{h(t)}{B(t)} - \pi_t \right) \right).
\]

\footnote{We refer to stochastic calculus text books (e.g. Protter (1995)) for a definition of optional stopping times and we assume that all stoppings considered are optional for the given processes in question.}
Where the second equality follows from the Optional Sampling Theorem and the martingale property of $\pi$. As (3) holds for any martingale, it follows that

$$\frac{Q(0)}{B(0)} \leq \inf_{\pi} \left( \pi_0 + E_0^Q \left[ \max_{t \in T(0)} \left( \frac{h(t)}{B(t)} - \pi_t \right) \right] \right).$$

(4)

This definition of the upper bound is slightly different to the one proposed in Haugh & Kogan (2001). They define the dual problem over all super martingales which, of course, includes all martingales. This generates more conservative upper bounds, as also noted in Footnote 1 of Andersen & Broadie (2001). This can easily be seen as the optional sampling theorem for super martingales adds an extra inequality in the derivation of the upper bound in equation (2) as $\pi_0 \geq E_0^Q(\pi_\tau)$ when $\pi$ is a supermartingale.

It can be proved that, when choosing $\pi$ in equation (2) as the martingale part of the Doob-Meyer decomposition of the deflated price process $Q(t)/B(t)$, equation (2) holds with equality. As a result, Andersen & Broadie (2001) propose to use the martingale part of a discounted lower bound price $L(t)$ as a proxy for the true value process. The lower bound price is defined by

$$L(t) = E_t^Q \left( \frac{\pi(t)}{B(t)} \right),$$

where $\tau_t$ is an optional stopping time given by some exercise strategy used from $t$ and onward. Having defined the lower bound process we use the following $\pi$ martingale process

$$\pi(t_1) = \frac{L(t_1)}{B(t_1)}$$

and for exercise dates $t_2, ..., t_d$

$$\pi_k = \pi_{k-1} + \frac{L(t_k)}{B(t_k)} - \frac{L(t_{k-1})}{B(t_{k-1})} \left[ I((t_{k-1}) \frac{L(t_k)}{B(t_k)} - \frac{L(t_{k-1})}{B(t_{k-1})} \right].$$

(5)

With this choice of $\pi$ process, an upper bound is seen to be the lower bound plus the value of a non-standard lookback option $D_0$, which is denoted the duality gap

$$\frac{Q(0)}{B(0)} \leq \frac{L(0)}{B(0)} + E_0^Q \left[ \max_{t \in T(0)} \left( \frac{h(t)}{B(t)} - \pi_t \right) \right] = \frac{L(0)}{B(0)} + D_0.$$

3 The Primal-Dual Simulation Algorithm

We apply the simulation algorithm laid out in Andersen & Broadie (2001) and refer to their paper for a thorough discussion of the entire algorithm. However, a short outline is appropriate. The algorithm generates estimates of the duality gap $D_0$ by simulating the $\pi$ process and the discounted payoff process $h(t)/B(t)$. The main difficulty is determining the $L(t)/B(t)$ terms in the $\pi$ process given by equation (5.). This is effectively done by running nested simulations replacing
\[
L(t_k)/B(t_k) \text{ and } E_k^Q \left( \frac{L(t_{k+1})}{B(t_{k+1})} \right) \text{ with Monte Carlo estimates based on } m \text{ nested simulations } \frac{L(t_k)}{B(t_k)} + \varepsilon_k \text{ and } E_k^Q \left( \frac{L(t_{k+1})}{B(t_{k+1})} \right) + \varepsilon'_k \text{ respectively. } \varepsilon_k, \varepsilon'_k \text{ are the mean zero simulation errors. By summing noise terms in } \pi \text{ it follows that the Monte Carlo estimate } \hat{\pi} \text{ of the exact } \pi \text{ process is }
\[
\hat{\pi}_k = \pi_k + \tilde{\varepsilon}_k,
\]
where the noise term \( \tilde{\varepsilon} \) is a sum of the mean zero noise terms. It can easily be proved that the noisy estimates of the \( \pi \) process make the estimate of the duality gap \( D_0 \) upward biased. Hence, the Monte Carlo estimator of the duality gap \( \hat{D}_0 \) using \( n \) "outer" simulations and \( m \) inner simulations, is
\[
\hat{D}_0(n,m) = \frac{1}{n} \sum_{i=1}^{n} \max_{1 \leq k \leq d} \left( \frac{h^i(t_k)}{B^i(t_k)} - \hat{\pi}_k \right).
\]
Using both the upper and lower bound we can construct a somewhat conservative confidence interval for the price estimate based on an \( n \times m \) simulation trial, as
\[
\left[ \hat{L}_0(n) - z_{1-\frac{\alpha}{2}} \hat{s}_L(n) \sqrt{\frac{n}{m}}, \hat{L}_0(n) + \hat{D}_0(n,m) + z_{1-\frac{\alpha}{2}} \frac{\hat{s}_L(n) + \hat{s}_D(n,m)}{\sqrt{n}} \right]
\]
where \( z_x \) denotes the \( x \)th percentile of the standard normal distribution and \( \hat{s}_L \) and \( \hat{s}_D \) are the sample standard deviations.

4 Variance Reduction

Andersen & Broadie (2001) do not consider variance reduction techniques. In this paper we test the effect of antithetic variates (AS) and control variates (CV) see e.g. Glynn (1994). A more elaborate discussion of variance reduction techniques applied in a finance setting can be found in e.g. Boyle, Brodie & Glasserman (1997).

4.1 Antithetic Variates

The method of antithetic variates is widely used and is based on the simple observation that if \( \varepsilon \) has a standard normal distribution so does \( -\varepsilon \). The idea is that random inputs obtained from the collection of antithetic pairs \( \{(\varepsilon_i, -\varepsilon_i)\}_{i=1}^L \) are more regularly distributed than a collection of \( 2L \) independent samples. However antithetic variates only work when the discounted payoffs estimated from \( \varepsilon_i \) and \( -\varepsilon_i \) are negatively correlated - increasing efficiency with correlation. As a result antithetic variates work very well for linear integrands and fails in the case of symmetric integrands. For simple options, determining the efficiency of antithetics beforehand can sometimes be done. However, when dealing with path-dependent American options, this is not possible.
4.2 Control Variates

The method of control variates is based on the principle ‘use what you know’. The most straightforward implementation of control variates replaces the evaluation of an unknown expectation with the evaluation of the difference between the unknown quantity and another expectation whose value is known. Suppose we at time $t$ know the expectation $E^Q_t[\mathcal{Y}]$ of an $M$-dimensional stochastic variable $\mathcal{Y}$. Assuming, that we can sample $I$ realizations of a scalar variable $Z$ exactly, an unbiased estimator $Z^CV_t$ of $E^Q_t[Z]$ is given by

$$Z^CV_t(\beta) = \frac{1}{I} \sum_{i=1}^{I} \left( Z^i - \beta^i \left( Y^i - E^Q_t[\mathcal{Y}] \right) \right)$$  \hspace{1cm} (7)

with variance

$$Var (Z^CV_t(\beta)) = \frac{1}{I} \left( \sigma_Z^2 - 2\beta^i \Sigma_{YZ} + \beta^i \Sigma_{YZ} \Sigma_{Y} \beta \right)$$  \hspace{1cm} (8)

for some appropriately chosen vector $\beta \in \mathbb{R}^M$. Here $\sigma_Z^2$ is the variance of $Z$, $\Sigma_Y$ denotes the covariance matrix of the controls and $\Sigma_{YZ}$ is the vector of covariances between $Z$ and $Y$.

The variance minimizing choice of $\beta$ is given by

$$\beta^* = \Sigma_Y^{-1} \Sigma_{YZ}.$$  \hspace{1cm} (9)

Inserting (9) into (8) we have at optimality

$$Var (Z^CV_t(\beta^*)) = \frac{1}{I} \left( 1 - R^2 \right) \sigma_Z^2,$$

where

$$R^2 = \frac{\Sigma_{YZ} \Sigma_Y^{-1} \Sigma_{YZ}}{\sigma_Z^2}.$$  

Thus effectiveness is determined by the size of the coefficient of multiple correlation $R$ between $Z$ and the control variates $Y$. In addition we notice that since $R^2 \in (0, 1)$ using the variance-minimizing coefficient $\beta^*$ we are guaranteed not to increase variance.

4.2.1 Control Variates for American options

It is not immediately clear how control variates may be applied to American style options. A naive guess would be to sample the controls at fixed times in the exercise period. In a recent paper, Rasmussen (2002) illustrates that the control sampled at the time of exercise of the Bermudan option, has much higher correlation with the discounted payoff from the option than with the control sampled at e.g. expiry of the option. It is shown in particular, that for any given martingale process $\mathcal{Y}$ and optional stopping times $\tau, \sigma$ for which $t \leq \tau \leq \sigma \leq T$ a.s. then

$$corr_t (Z^{\tau}, Y^\tau)^2 \geq corr_t (Z^\tau, Y^\sigma)^2.$$
Hence, we choose to sample controls at the exercise time, that is
\[ Y(\tau) = E^Q_\tau (Y_T) , \tau \leq T. \]

5 Bermudan Swaptions in Libor Market Models

5.1 Libor Market Models

Since the seminal papers of Miltersen, Sandmann & Sondermann (1997), Brace & Musiela (1997) and Jamshidian (1997), Libor market models have become increasingly popular in the practitioners’ world, as the models are reasonably easy to calibrate and allow closed form solutions for caps and swaptions (though not simultaneously). Several extensions have been proposed. Jamshidian (1999) develops a general theory for Libor market models driven by semimartingales. Along this line Glasserman & Kou (1999) developed a version with jumps, driven by a marked point process, which have closed form solutions for certain derivatives. We follow the approach taken by Andersen & Brotherton-Ratcliffe (2001) who developed an extended Libor Market model with a continuous stochastic volatility process independent of the forward rates themselves.

The Libor market model is defined on an increasing tenor structure \(0 = T_0 < T_1 < ... < T_{K+1}\). Let \(\eta(t)\) denote the right continuous mapping function returning the index of the next tenor time \(\eta(t) = \{j : T_{j-1} < t \leq T_j\}\). The simple forward rate \(F_k(t)\) for the period \(T_k\) to \(T_{k+1}\) is defined by
\[
F_k(t) = \frac{1}{\delta_k} \left( \frac{P(t, T_k)}{P(t, T_{k+1})} - 1 \right), \quad \delta_k = T_{k+1} - T_k,
\]
where \(P(t, \cdot)\) denotes the time \(t\) discount function. We work under the "spot Libor" measure \(Q\) (see e.g. Jamshidian (1997)) under which all assets discounted by the "Bank" account \(B(t)\) are martingales. \(B(t)\) is the value of an initial $1 investment rolled, over at the spot Libor rate at each tenor date
\[
B(t) = P(t, T_{\eta(t)}) \prod_{k=0}^{\eta(t)-1} (1 + \delta_k F_k(T_k)).
\]

Andersen & Brotherton-Ratcliffe (2001) model stochastic volatility with a variance process \(V(t)\) used to scale the diffusion term of all forward rates \(\varphi(F_k(t)) \lambda_k(t), k = 1, ..., K\).

Using no-arbitrage arguments (see e.g. Jamshidian (1999)) it can be proven that the full dynamics under the spot Libor measure \(Q\) of the \((K + 1)\)–Markov system of forward Libor rates is given by
\[
\begin{align*}
\text{d}F_k(t) &= \varphi(F_k(t)) \sqrt{V(t)} \lambda_k(t) \left( \mu_k(t) \text{d}t + \text{d}W^Q_t \right), \quad k = 1, ..., K \\
\text{d}V(t) &= \kappa_V (\theta_V - V(t)) \text{d}t + \varepsilon_V \kappa_V (\theta_V - V(t)) \text{d}t + \varepsilon_V \psi(V(t)) \text{d}Z^Q_t
\end{align*}
\]
Where \(\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}^+\) is a one dimensional function satisfying certain regularity conditions, \(\psi : \mathbb{R}^+ \rightarrow \mathbb{R}^+\) is well-behaved, \(\lambda_k(t)\) is a bounded deterministic
function taking values in $\mathbb{R}^{1 \times m}$, $\kappa_V, \theta_V, \varepsilon_V$ are positive constants and $W^Q$ and $Z^Q$ are $m$-dimensional and one dimensional Brownian motions under $Q$, respectively. The drift function for the $k$-th Libor rate is given by

$$
\mu_k(t) = \sum_{j=\eta(t)}^{k} \frac{\delta_j \varphi (F_j(t))}{1 + \delta_j F_j(t)} \lambda_j(t)^T.
$$

(12)

It is important that the Wiener processes driving the variance process and the Libor rates are uncorrelated, for the approximation formulae for European options derived by Andersen & Brotherton-Ratliffe (2001) to hold. We do not consider calibration issues in this paper but keep the assumption anyway. Furthermore, it is natural to scale the variance process such that $\theta = 1$, meaning that $V(t) - 1$ represents a percentage deviation from the long term mean of the variance.

5.2 Exercise strategies

An exercise strategy is basically all that we need to get the Primal-Dual algorithm running. Several possibilities exist regarding an appropriate exercise strategy including exercise strategies from lower dimensional models which can be solved using lattice methods, least-square Monte Carlo (LS) (see Longstaff & Schwartz (2001)), neural networks, or trigger strategies like the ones explored in Andersen (2000). We have chosen the latter, because of its simplicity and speed. Bjerregaard Pedersen (1999) finds the Longstaff & Schwartz (2001) and Andersen (2000) approaches to be mutually consistent for Bermudan swaptions.

The idea in trigger strategies is to reduce the dimensionality of the exercise decision. We consider exercise strategies of the following form. First, we let $X(t_i)$ denote the option payoff if exercised at time $t_i$. The exercise rules specify exercise if the payoff is larger than some parameterized function $f$ of the state vector $Z(t_i)$.

$$
I(t_i) = \begin{cases} 
1 & \text{if } X(t_i) > f(Z(t_i); p(t_i)) \\
0 & \text{otherwise}
\end{cases}
$$

Among the exercise strategies we consider for Bermudan swaptions are the ones proposed in Andersen (2000). First, we let $M(t_i) = \max_{j=1, \ldots, e} (EO_{j,e}(t_i))$ denote the most valuable of the still alive core European options $EO_{j,e}(t_i)$.

Strategy 1:

$$
I^1(t_i) = \begin{cases} 
1 & \text{if } X(t_i) > H_1(t_i) \\
0 & \text{otherwise}
\end{cases}
$$

\[\text{This is also in agreement with theory and empirical evidence regarding unspanned stochastic volatility documented in Collin-Dufresne & Goldstein (2002). Nonzero correlation between volatility variance and Libor rates would (in principle) make it possible to hedge all assets using only bonds.}\]
Strategy 2:

\[ I^2(t_i) = \begin{cases} 
  1 & \text{if } X(t_i) > H_2(t_i) \text{ and } X(t_i) > M(t_i) \\
  0 & \text{otherwise}
\end{cases} \]

Strategy 3:

\[ I^3(t_i) = \begin{cases} 
  1 & \text{if } X(t_i) > H_3(t_i) + M(t_i) \\
  0 & \text{otherwise}
\end{cases} \]

Strategy 4:

\[ I^4(t_i) = \begin{cases} 
  1 & \text{if } X(t_i) > H_2(t_i) \text{ and } X(t_i) > EO_{t_i+1,e}(t_i) \\
  0 & \text{otherwise}
\end{cases} \]

Strategy 5:

\[ I^5(t_i) = \begin{cases} 
  1 & \text{if } X(t_i) > H_3(t_i) + EO_{t_i+1,e}(t_i) \\
  0 & \text{otherwise}
\end{cases} \]

Here \( H_i(t) \) denotes a deterministic function from \( R^+ \to R \). In strategy 2, the payoff should be larger than the barrier and the maximal value of the remaining European options. Finally, in strategy 3, the payoff should exceed the sum of the barrier and the maximal value of remaining European options. Andersen (2000) finds that, in certain scenarios the more advanced strategies pick up additional value compared to the first strategy, but they also increase the computational burden - especially for options with several exercise times. When computing the maximal value of the remaining European options required for strategy 3, the worst case number of option valuations for each path is in order \( n^2 \) where \( n \) is the number of exercise times. In worst case, an option with 40 exercise times would require as much as \( 40(40 + 1)/2 = 820 \) European option valuations for each path. The rationale for these strategies is of course that the holder of the American option could always sell it to the value of the most valuable European option.

We propose a small simplification of the strategies using the maximal of the European values, namely to use only a subset of the core European options as these are likely to be correlated. In the Bermudan swaption case considered below the most valuable core swaption is almost always the first to mature (the underlying swap is the longest). So we test two simplifications of the strategies 2 and 3, in which we replace the maximal of the remaining swaptions with the first to mature European swaption \( ES(t_i) \). These are denoted strategy 4 and strategy 5 respectively and we return to their performances in section 6.4.

The critical barriers \( H_i(t) \) are found using an optimization procedure described in Andersen (2000) consisting in a presimulation of paths and a series of one dimensional optimizations at each exercise time.
5.3 Bermudan Swaptions

The Bermudan swaption (payer) is basically an option to enter into a (payer) swap contract with a coupon of $\kappa$. The standard product denoted $BS_{s,x,e}$ is characterized by three dates: the lockout date $T_s$, the last exercise date $T_x$ and the final swap maturity $T_e$. We will consider the fixed maturity case where $T_s < T_x < T_e$, allowing early exercise on dates in the set $T = \{T_s, T_{s+1}, \ldots, T_x\}$.

If the option is exercised at time $T_k$, the holder of the option receives the payoff from the corresponding European swaption with $e - k$ periods and coupon $\kappa$. If exercised at time $T_k \in T$ the payoff is

$$BS_{s,x,e}(T_k, \kappa) = (\text{Swap}(T_k, T_k, T_e, \kappa))^+$$

$$= \left(1 - P(T_k, T_e) - \kappa \sum_{i=k}^{e-1} \delta_i P(T_k, T_{i+1})\right)^+.$$

From (1) we know that the value at time $t \leq T_{e-1}$ of the Bermudan payer swaption $BS$ is given as the solution to the optimal stopping problem

$$BS_{s,x,e}(t, \kappa) = \sup_{\tau \in \Gamma(t)} E_t^Q \left[ \frac{B(t)}{B(\tau)} (\text{Swap}(\tau, \tau, T_e, \kappa))^+ \right].$$

5.4 Possible control variates

In this section we consider several assets which we test as control variates. However, it is important to notice that when applying the Rasmussen (2002) technique of sampling at the exercise time, we have to be careful with assets that have several payoff times, as the ex-dividend value of an asset is a supermartingale. If a control asset $X(t)$ is generating a dividend process $\{\gamma(T_i)\}_{1 \leq i \leq K+1}$, we know that the time $t$ value of the remaining discounted dividends is a martingale and given by

$$\frac{X(t)}{B(t)} = E_t^Q \left[ \sum_{i=\eta(t)}^{K+1} \frac{\gamma(T_i)}{B(T_i)} \right]$$

$$= E_t^Q \left[ E_{\tau}^Q \left( \sum_{i=\eta(t)}^{K+1} \frac{\gamma(T_i)}{B(T_i)} \right) \right]$$

$$= E_t^Q \left[ \sum_{i=1}^{\tau} \frac{\gamma(T_i)}{B(T_i)} + E_{\tau}^Q \left( \sum_{i=\tau+1}^{K+1} \frac{\gamma(T_i)}{B(T_i)} \right) \right]$$

$$= E_t^Q \left[ \sum_{i=1}^{\tau} \frac{\gamma(T_i)}{B(T_i)} + \frac{X(\tau)}{B(\tau)} \right].$$

for a given stopping time $\tau \in \Gamma(t)$. Where we have used the Law of Iterated Expectations for the second equality.
Equation (14) states that, as the initial value includes all future dividends received from the asset, we have to include all dividends reinvested in the numeraire asset\(^3\) in the sampling value \(Y\) of the asset \(X\). Hence we sample

\[
Y(\tau) = \sum_{k=1}^{s} \frac{\gamma(T_k)}{B(T_k)} + \frac{X(T_\tau)}{B(T_\tau)}. \tag{15}
\]

Notice that assets maturing before the option expires can still be used (e.g. zero coupon bonds).

5.4.1 European Swaptions

From (13) an obvious choice of controls would be to use the discounted payoff from the core European swaptions \(ES_{k,e}(\cdot,e) / B(\cdot)\) for \(T_k = T_s, T_{s+1}, \ldots, T_x\) which are \(Q\) martingales\(^4\). However, only approximative closed form solutions for swaptions have been derived (see e.g. Brace & Musiela (1997) or Andersen & Andreasen (2000)), which in effect means that European swaptions are not applicable as controls in a Libor market model. Another possibility is to create a self-financing portfolio replicating as closely as possible the European swaptions by using the hedge ratios from the approximative solutions, as any discounted self-financing strategy has to be a martingale in a no-arbitrage setting. This approach is described below in section 5.4.4.

5.4.2 Caps and Caplets

A more simple solution is to consider caps and caplets, as these have closed form solutions in a standard Libor market model (see e.g. Brace & Musiela (1997) or Andersen & Andreasen (2000)). We note that a caplet is a one-period swaption and hence the price of the cap will always be higher than that of a swaption with the same coupon, start date and end date, and thus make an upper bound of the price of the swaption. The question is now which caps or caplets should be used as controls. In theory, including all possible control variates will never decrease efficiency, but in practice we could experience problems with multicollinearity.

Hence, a first choice of including all caplets with starting dates in \(\mathcal{T}(t)\) with a strike equal to the coupon of the Bermudan swaption would mean sampling all \(x-s\) controls\(^5\). If we let \(\wedge\) denote the minimum operator the sampled values are

\[
Y_k(\tau) = \frac{\text{Caplet}(\tau \wedge T_k, T_k, \kappa)}{B(\tau)}.
\]

We have just argued that a cap with same strike rate, start- and maturity date as a Bermudan swaption only constitutes an upper bound, but could be

\(^{3}\)All assets give the same return under the spot libor measure. Hence anyone be used.

\(^{4}\)We cannot use the discounted forward swap rate as a control as this is not a martingale under the measure \(Q\). Each forward swap rate \(\omega_{T_k}(\cdot,e)\) is a martingale under the forward swap measure \(Q^{A,e}\) induced by the numeraire \(B^{A,e}(\cdot)\).

\(^{5}\)For upper bound calculations we will use ATM caplets. Thus when calculating \(E_t(\cdot)\) we fix the strike at time \(t\) such that the caplet is ATM.
a reasonable control. In principle it is just a portfolio of caplets and one could suspect that we might gain more by including all constituent caplets as separate controls. This is generally not the case as we avoid a lot of multicollinearity using the cap. By using the cap as we essentially estimate a common $\beta$ for all caplets. Referring to (15) the sampling value $Y(\tau)$ is

$$Y(\tau) = \sum_{k=1}^{\tau} \frac{(F_{k-1}(T_{k-1}) - \kappa)^+}{B(T_k)} + \frac{\text{Cap}(T_\tau, T_{e-1}, \kappa)}{B(T_\tau)}.$$ 

5.4.3 Swaps and Zero Coupons

Standard fixed income securities can be used as control variates in the Libor market models as the yield curve is known. We consider using the swap underlying the Bermudan as control variate. Thus we sample

$$Y(\tau) = \sum_{k=s}^{\tau} P(T_k, T_{k+1}) \frac{\delta_k (F_k(T_k) - \kappa)}{B(T_k)} + \frac{\text{Swap}(\tau, \tau+1, T_e, \kappa)}{B(\tau)}.$$ 

We also test a series of zero coupon bonds with maturity dates equal to the Bermudan option’s set of exercise dates $T$ as a controls.

$$Y_k(\tau) = \frac{P(\tau \land T_k, T_k)}{B(\tau \land T_k)}, \quad k \in T.$$ 

Notice that the swap is a portfolio of zero coupon bonds and hence the difference between the two control types is basically that we restrict the $\beta$ parameter when using a swap as control.

5.4.4 Approximate Delta-Hedge

The value of any self-financing portfolio discounted with the pricing numeraire must be a martingale. Hence, we know that the expected value of such a portfolio is the initial value of the portfolio. As such we can use the payoff from a portfolio that approximately replicates the option we are trying to value. (see e.g. Clewlow & Carverhill (1994) for more simple examples).

As a quite general control variate we test a hedge portfolio. More specifically we form a portfolio delta hedging the first of the remaining core European swaptions to expire using the delta implied by the approximate swaption formula given in Andersen & Andreasen (1998). Hence

$$\Delta_{s,e}(t) = \frac{\partial ES_{s,e}(t, \kappa)}{\partial S_{s,e}(t)} = B_{s,e}(t)\Phi(d_+)$$

where we denote the forward swap rate and the accrual factor by

$$S_{s,e}(t) = \frac{P(t, T_s) - P(t, T_e)}{\sum_{k=s+1}^{e} \delta_{k-1} P(t, T_k)}$$
\[ B_{s,e}(t) = \sum_{k=s+1}^{e} \delta_{k-1} P(t, T_k), \]

respectively. \( \Phi \) denotes the standard normal cumulative density function and

\[ d_+ (t, T_s) = \log \left( \frac{\omega_{T_s}(t,e-s)}{\xi(t,T_s)} + \frac{1}{2} \xi^2(t,T_s) \right) \]

where \( \xi(t,T_s) \) is the approximative integrated volatility. That is, at time \( t_j \) we buy

\[ \Delta_{s,e}(t_j) \cdot S_{s,e}(t_j), \quad t_j < T_s \]

\[ \Delta_{j+1,e}(t_j) \cdot S_{j+1,e}(t_j), \quad T_s \leq t_j \leq T_e \]

financing it with a short position in the bank account. At the following tenor date we liquidate the swap position and enter a new, putting the profits or losses in the bank account. This is done until exercise or maturity.

The drawback of this general approach is that it requires the hedge portfolio to be updated all along each path, which is computationally expensive. Furthermore, as the hedging errors are decreasing in the number of resettings and we only reset at the tenor dates, we do not expect a correlation of the payoff very close to 1. Numerical tests will show how well this strategy performs.

A more precise hedging strategy would be to take into account the model dynamics of the swap rate and try to hedge each factor driving the yield curve. Though, increasing precision should be measured against increases in computation time.

## 6 Numerical Results

Simulation of Libor market models cannot be simulated exactly. However, the resulting discretization biases are manageable and small compared to the variance as shown in e.g. Andersen & Andreasen (1998) and is ignored in the following analysis. The exact simulation schemes have been included in the appendix for completeness.

### 6.1 Benchmarking

To be able to compare with the results found in Andersen & Broadie (2001) we use some of the same test scenario. We start out with a standard log-normal Libor market model with quarterly tenor structure and deterministic volatility corresponding to letting \( \varphi(x) = x, \varepsilon = 0 \) and \( V(0) = 1 \) in equation (10). The initial yield curve is flat 10% and we investigate a two-factor version with the following deterministic volatility structure.

\[ \lambda_k(t) = (0.15, 0.15 - \sqrt{0.009 \cdot (T_k - t)})^T \]

To benchmark our model implementation we include Table 1 which replicates a part of the table presented in Andersen & Broadie (2001).
Table 1: Benchmark Scenarios

<table>
<thead>
<tr>
<th>$T_s$</th>
<th>$T_e$</th>
<th>Strike</th>
<th>Lower</th>
<th>Std Low</th>
<th>$\hat{D}_0$</th>
<th>Std $\hat{D}_0$</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>3.00</td>
<td>0.08</td>
<td>339.41</td>
<td>0.24</td>
<td>0.34</td>
<td>0.05</td>
<td>00:01:32</td>
</tr>
<tr>
<td>1.00</td>
<td>3.00</td>
<td>0.10</td>
<td>124.82</td>
<td>0.34</td>
<td>0.55</td>
<td>0.07</td>
<td>00:02:29</td>
</tr>
<tr>
<td>1.00</td>
<td>3.00</td>
<td>0.12</td>
<td>35.89</td>
<td>0.24</td>
<td>0.44</td>
<td>0.07</td>
<td>00:03:07</td>
</tr>
<tr>
<td>1.00</td>
<td>6.00</td>
<td>0.08</td>
<td>749.59</td>
<td>0.55</td>
<td>3.09</td>
<td>0.26</td>
<td>00:11:20</td>
</tr>
<tr>
<td>1.00</td>
<td>6.00</td>
<td>0.10</td>
<td>317.10</td>
<td>0.68</td>
<td>4.75</td>
<td>0.32</td>
<td>00:20:29</td>
</tr>
<tr>
<td>1.00</td>
<td>6.00</td>
<td>0.12</td>
<td>126.29</td>
<td>0.60</td>
<td>2.52</td>
<td>0.26</td>
<td>00:26:38</td>
</tr>
<tr>
<td>1.00</td>
<td>11.00</td>
<td>0.08</td>
<td>1249.53</td>
<td>1.24</td>
<td>18.87</td>
<td>1.31</td>
<td>01:10:11</td>
</tr>
<tr>
<td>1.00</td>
<td>11.00</td>
<td>0.10</td>
<td>620.62</td>
<td>1.19</td>
<td>19.99</td>
<td>1.09</td>
<td>01:57:07</td>
</tr>
<tr>
<td>1.00</td>
<td>11.00</td>
<td>0.12</td>
<td>329.89</td>
<td>1.17</td>
<td>14.11</td>
<td>0.97</td>
<td>02:27:57</td>
</tr>
</tbody>
</table>

Benchmarking of the lower bound estimates and the dual Bermudan swaption. $T_s$ denotes the lock-out period of the Bermudan swaption and $T_e$ is the final swap maturity. Strike denote the coupon of the underlying swap. Lower and Std Low are the lower bound point estimate and its standard error found using 50,000 antithetic paths and strategy 1. $\hat{D}_0$ is the point estimate of the duality gap and $\text{Std}\hat{D}_0$ its standard deviation based on $n = 750$ antithetic paths and $m = 300$ antithetic paths used for "simulation within the simulation". Time denotes the computation time for the duality gap in hours:minutes:seconds.

Table 1 contains estimates of the lower bound and the duality gap for a set of Bermudan swaptions. $T_s$ denotes the lock-out period of the Bermudan swaption and $T_e$ is the final swap maturity. Strike denotes the coupon of the underlying swap. Lower and Std Low are the lower bound point estimate and standard deviation of the lower bound found using 50,000 antithetic paths and strategy 1. $\hat{D}_0$ is the point estimate of the duality gap and $\text{Std}\hat{D}_0$ the standard deviation based on 750 antithetic paths and 300 antithetic paths used for "simulation within the simulation".

Even though we have estimated our own exercise barriers, we cannot distinguish our estimates from the ones presented in Andersen & Broadie (2001) as the confidence intervals are overlapping. This goes for both the lower and upper bounds.

6.2 Efficiency Improvement

6.2.1 Caps and caplets as control variates

Figure 1 illustrates the effect of the number of caplets used as control variates on the standard deviation. In this setup we have sampled the first and the last possible control and sampled remaining controls equally spaced. We see that for the 11 nc 1 payer Bermudan swaption, about seven caplet controls, equally spaced over the exercise period, should be sufficient and that sampling all 40 caplets will not improve the estimation result significantly.

If we were only interested in reducing the standard deviation of the lower
bound estimates we would include all caplets. However, as we show in the following section we have to take computation time into consideration too.

6.2.2 Testing controls variates

Comparing the efficiency of variance reduction techniques is always difficult, as they depend on the implementation. However, it is still relevant to compare particular implementations. In Glynn & Whitt (1992) it is proved that if we compare two estimators with variances $\sigma_1^2$ and $\sigma_2^2$ requiring an expected work per run of $Ew_1$ and $Ew_2$, respectively, we should favor the estimator with the lowest variance scaled with the workload $\sigma_i^2 \cdot Ew_i$. We denote this product the efficiency factor $EF_i$.

In Table 2 we test the efficiency on lower bound estimates of different types of control variates on 1 into 6 and 1 into 11 Bermudan swaptions of various strikes. The table shows the standard deviations and the variances (in basis points) scaled with the average time per path. We have tested two caplet setups: the case of 7 caplets and all caplets, as we pointed out in section 6.2.1, using all caplets might not always be the best choice for cases of long maturity options.

For both the 1 into 6 and 1 into 11 Bermudan swaption we see that both the cap and the series of caplets perform almost equally well for all strikes, the series of caplets doing slightly better than the cap for the OTM options.
Especially, we note that the more the option is in the money, the less efficient it is to use all caplets compared to only using 7. However, the differences in the efficiency factor between the caplet types and the cap control variates are negligible compared to the other types of controls which perform quite poorly in comparison.

We note that using the cap compared to a series of caplets is easier to apply in the sense that we do not have to decide how many caplets to use - including too many or too few caplets may cause the caplet control to be less efficient than the single cap.

From Table 2 we see that using the self-financing delta hedge strategy results in a reasonably low standard deviation - yet still not better than the caplets or single cap - but when taking the computational effort into account it does not constitute a good control variate. Especially not for out of the money options.

The variance reduction obtained by using the zero coupon bond and the swap is of similar size in all cases. But as the swap uses more computer time and hence it is better to use the zero coupon bond as a control.

As usual the antithetic sampling works better for in the money options as the payoff function is close to being linear. In Table 3 we present results using antithetic variates and control variates. We have used 25,000 antithetic paths - i.e. a total of 50,000 paths - in order to be able to compare table 2 and table 3. The results are similar to the previous results. The single cap and the caplets still outperform the other types of controls, and the zero coupon bonds are more efficient than the swap.

Comparing table 2 with table 3 we notice that the standard deviations and efficiency factors given in table 3 are lower than those of table 2 except for the case of the delta hedge control variate.

We also test the effect of using multiple controls simultaneously. As before we run tests in a two factor Libor model using strategy 1 and 50,000 paths (Crude MC). In Table 4 we have listed the 15 most efficient types of control variates for a 1 into 11 Bermudan with varying strikes.

We see that the caplet control variate still performs reasonably well compared to using multiple control types, especially for out of the money options. However, by using a series of caplets or single cap together with the zero coupon bond or swap is more efficient than using only the caplets; except for the out of the money Bermudan swaption, where it is more efficient to use only the caplet instead of both the caplet and the swap. As expected the zero coupon bond, together with the cap/caplets, is more efficient than the swap together with the cap/caplets. The swap, cap and zero coupon also does quite well for in-the and at-the money Bermudan. Finally we point out that the delta hedge control is far too time consuming compared to the other types of controls.

6.3 The Bias of the Upper Bound

Using Jensen’s inequality it is easily shown that the estimator for the duality gap is upward biased. Andersen & Broadie (2001) do not examine the size and behavior of the upper bound. This is particularly interesting as we want to
Table 2: Efficiency of control variates

<table>
<thead>
<tr>
<th>Coupon</th>
<th>CV</th>
<th>6 nc 1 StdLow</th>
<th>6 nc 1 EF</th>
<th>11 nc 1 StdLow</th>
<th>11 nc 1 EF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08</td>
<td>Antithetic</td>
<td>0.5582</td>
<td>0.0804</td>
<td>1.1291</td>
<td>0.8570</td>
</tr>
<tr>
<td>0.08</td>
<td>Cap</td>
<td>0.1856</td>
<td>0.0061</td>
<td>0.5182</td>
<td>0.1115</td>
</tr>
<tr>
<td>0.08</td>
<td>Caplets</td>
<td>0.1843</td>
<td>0.0057</td>
<td>0.5357</td>
<td>0.1072</td>
</tr>
<tr>
<td>0.08</td>
<td>Caplets All</td>
<td>0.1761</td>
<td>0.0063</td>
<td>0.5046</td>
<td>0.1267</td>
</tr>
<tr>
<td>0.08</td>
<td>DeltaHedge</td>
<td>0.3240</td>
<td>0.0319</td>
<td>0.7485</td>
<td>0.4055</td>
</tr>
<tr>
<td>0.08</td>
<td>None</td>
<td>2.0469</td>
<td>0.6218</td>
<td>3.2819</td>
<td>3.7633</td>
</tr>
<tr>
<td>0.08</td>
<td>Swap</td>
<td>0.5254</td>
<td>0.0525</td>
<td>1.1693</td>
<td>0.6760</td>
</tr>
<tr>
<td>0.08</td>
<td>Z-Bond</td>
<td>0.5259</td>
<td>0.0453</td>
<td>1.1599</td>
<td>0.5527</td>
</tr>
<tr>
<td>0.10</td>
<td>Antithetic</td>
<td>0.6781</td>
<td>0.2146</td>
<td>1.1372</td>
<td>1.7359</td>
</tr>
<tr>
<td>0.10</td>
<td>Cap</td>
<td>0.2620</td>
<td>0.0186</td>
<td>0.6314</td>
<td>0.3013</td>
</tr>
<tr>
<td>0.10</td>
<td>Caplets</td>
<td>0.2675</td>
<td>0.0191</td>
<td>0.6153</td>
<td>0.2740</td>
</tr>
<tr>
<td>0.10</td>
<td>Caplets All</td>
<td>0.2522</td>
<td>0.0189</td>
<td>0.5728</td>
<td>0.2639</td>
</tr>
<tr>
<td>0.10</td>
<td>DeltaHedge</td>
<td>0.4982</td>
<td>0.1257</td>
<td>1.0507</td>
<td>1.4600</td>
</tr>
<tr>
<td>0.10</td>
<td>None</td>
<td>1.5379</td>
<td>0.5974</td>
<td>2.7992</td>
<td>5.3533</td>
</tr>
<tr>
<td>0.10</td>
<td>Swap</td>
<td>0.7379</td>
<td>0.1627</td>
<td>1.4140</td>
<td>1.6782</td>
</tr>
<tr>
<td>0.10</td>
<td>Z-Bond</td>
<td>0.7368</td>
<td>0.1467</td>
<td>1.4164</td>
<td>1.4921</td>
</tr>
<tr>
<td>0.12</td>
<td>Antithetic</td>
<td>0.6005</td>
<td>0.2111</td>
<td>1.1716</td>
<td>2.7579</td>
</tr>
<tr>
<td>0.12</td>
<td>Cap</td>
<td>0.2901</td>
<td>0.0130</td>
<td>0.5102</td>
<td>0.2314</td>
</tr>
<tr>
<td>0.12</td>
<td>Caplets</td>
<td>0.1977</td>
<td>0.0128</td>
<td>0.4875</td>
<td>0.2078</td>
</tr>
<tr>
<td>0.12</td>
<td>Caplets All</td>
<td>0.1858</td>
<td>0.0121</td>
<td>0.4382</td>
<td>0.1890</td>
</tr>
<tr>
<td>0.12</td>
<td>DeltaHedge</td>
<td>0.5185</td>
<td>0.1695</td>
<td>1.2019</td>
<td>2.3707</td>
</tr>
<tr>
<td>0.12</td>
<td>None</td>
<td>1.0172</td>
<td>0.3274</td>
<td>2.1852</td>
<td>4.1335</td>
</tr>
<tr>
<td>0.12</td>
<td>Swap</td>
<td>0.6499</td>
<td>0.1531</td>
<td>1.3210</td>
<td>1.7828</td>
</tr>
<tr>
<td>0.12</td>
<td>Z-Bond</td>
<td>0.6418</td>
<td>0.1371</td>
<td>1.3167</td>
<td>1.6017</td>
</tr>
</tbody>
</table>

This table contains a summary of efficiency rates for the control variates considered in this paper applied on the 1x6 and 1x11 Bermudan swaptions in the two-factor deterministic volatility scenario. Coupon is the coupon rate of the underlying swap. CV is the control variate and std Low denotes the standard deviation of the lower bound using strategy 1. EF denotes the variance in (bps) scaled with the average time per path. Low values of EF are preferred. The numbers were generated using 50,000 paths.
Table 3: Efficiency of Control Variates with Antithetic Sampling.

<table>
<thead>
<tr>
<th>Antithetic</th>
<th>Coupon</th>
<th>CV</th>
<th>StdLow</th>
<th>EF</th>
<th>StdLow</th>
<th>EF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08</td>
<td>Antithetic</td>
<td>0.7910</td>
<td>0.0826</td>
<td>1.6126</td>
<td>0.8883</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>Cap</td>
<td>0.1755</td>
<td>0.0050</td>
<td>0.5101</td>
<td>0.1053</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>Caplets</td>
<td>0.1544</td>
<td>0.0035</td>
<td>0.4946</td>
<td>0.0874</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>Caplets All</td>
<td>0.1441</td>
<td>0.0038</td>
<td>0.4664</td>
<td>0.1001</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>DeltaHedge</td>
<td>0.3823</td>
<td>0.0427</td>
<td>0.8422</td>
<td>0.5088</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>Swap</td>
<td>0.3681</td>
<td>0.0230</td>
<td>0.9041</td>
<td>0.3736</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>Z-Bond</td>
<td>0.3637</td>
<td>0.0188</td>
<td>0.9097</td>
<td>0.3091</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>Antithetic</td>
<td>0.9576</td>
<td>0.2166</td>
<td>1.6121</td>
<td>1.7755</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>Cap</td>
<td>0.2028</td>
<td>0.0105</td>
<td>0.5427</td>
<td>0.2112</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>Caplets</td>
<td>0.2298</td>
<td>0.0131</td>
<td>0.5861</td>
<td>0.2381</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>Caplets All</td>
<td>0.2043</td>
<td>0.0115</td>
<td>0.5129</td>
<td>0.2072</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>DeltaHedge</td>
<td>0.5256</td>
<td>0.1350</td>
<td>0.9195</td>
<td>1.1137</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>Swap</td>
<td>0.7033</td>
<td>0.1363</td>
<td>1.3863</td>
<td>1.5601</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>Z-Bond</td>
<td>0.7032</td>
<td>0.1218</td>
<td>1.3981</td>
<td>1.4042</td>
<td></td>
</tr>
<tr>
<td>0.12</td>
<td>Antithetic</td>
<td>0.8456</td>
<td>0.2121</td>
<td>1.6562</td>
<td>2.3656</td>
<td></td>
</tr>
<tr>
<td>0.12</td>
<td>Cap</td>
<td>0.1892</td>
<td>0.0109</td>
<td>0.4684</td>
<td>0.1974</td>
<td></td>
</tr>
<tr>
<td>0.12</td>
<td>Caplets</td>
<td>0.1853</td>
<td>0.0105</td>
<td>0.4489</td>
<td>0.1762</td>
<td></td>
</tr>
<tr>
<td>0.12</td>
<td>Caplets All</td>
<td>0.1685</td>
<td>0.0092</td>
<td>0.3979</td>
<td>0.1529</td>
<td></td>
</tr>
<tr>
<td>0.12</td>
<td>DeltaHedge</td>
<td>0.4017</td>
<td>0.0968</td>
<td>0.7565</td>
<td>0.9380</td>
<td></td>
</tr>
<tr>
<td>0.12</td>
<td>Swap</td>
<td>0.6183</td>
<td>0.1281</td>
<td>1.3616</td>
<td>1.8555</td>
<td></td>
</tr>
<tr>
<td>0.12</td>
<td>Z-Bond</td>
<td>0.6182</td>
<td>0.1176</td>
<td>1.3447</td>
<td>1.6409</td>
<td></td>
</tr>
</tbody>
</table>

This table contains a summary of efficiency rates for the control variates considered in this paper applied on the 1x6 and 1x11 Bermudan swaptions in the two-factor deterministic volatility scenario. Coupon is the coupon rate of the underlying swap. CV is the control variate and std Low denotes the standard deviation of the lower bound using strategy 1. EF denotes the efficiency factor i.e. the variance in (bps) scaled with the average time per path. Low values of EF are preferred to high. The numbers were generated using 25,000 (AS) paths.
<table>
<thead>
<tr>
<th>Rank</th>
<th>EF</th>
<th>CV</th>
<th>EF</th>
<th>CV</th>
<th>EF</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0767</td>
<td>ZB+Cap</td>
<td>0.2173</td>
<td>ZB+Cap</td>
<td>0.1821</td>
<td>ZB+Cap</td>
</tr>
<tr>
<td>2</td>
<td>0.0833</td>
<td>ZB+CL</td>
<td>0.2378</td>
<td>SW+Cap</td>
<td>0.1885</td>
<td>ZB+CLA</td>
</tr>
<tr>
<td>3</td>
<td>0.0898</td>
<td>ZB+SW+Cap</td>
<td>0.2428</td>
<td>ZB+CL</td>
<td>0.1890</td>
<td>CLA</td>
</tr>
<tr>
<td>4</td>
<td>0.0902</td>
<td>SW+Cap</td>
<td>0.2523</td>
<td>ZB+SW+Cap</td>
<td>0.1914</td>
<td>ZB+CL</td>
</tr>
<tr>
<td>5</td>
<td>0.0962</td>
<td>SW+CL</td>
<td>0.2536</td>
<td>ZB+CLA</td>
<td>0.2033</td>
<td>SW+CLA</td>
</tr>
<tr>
<td>6</td>
<td>0.1038</td>
<td>ZB+CLA</td>
<td>0.2639</td>
<td>CLA</td>
<td>0.2078</td>
<td>CL</td>
</tr>
<tr>
<td>7</td>
<td>0.1072</td>
<td>CL</td>
<td>0.2678</td>
<td>SW+CL</td>
<td>0.2083</td>
<td>SW+Cap</td>
</tr>
<tr>
<td>8</td>
<td>0.1092</td>
<td>SW+CLA</td>
<td>0.2740</td>
<td>CL</td>
<td>0.2116</td>
<td>SW+CL</td>
</tr>
<tr>
<td>9</td>
<td>0.1115</td>
<td>Cap</td>
<td>0.2797</td>
<td>SW+CLA</td>
<td>0.2314</td>
<td>Cap</td>
</tr>
<tr>
<td>10</td>
<td>0.1169</td>
<td>ZB+SW+CL</td>
<td>0.3013</td>
<td>Cap</td>
<td>0.2443</td>
<td>ZB+SW+Cap</td>
</tr>
<tr>
<td>11</td>
<td>0.1267</td>
<td>CLA</td>
<td>0.3336</td>
<td>ZB+Cap+DH</td>
<td>0.2641</td>
<td>CL+DH</td>
</tr>
<tr>
<td>12</td>
<td>0.1334</td>
<td>ZB+Cap+DH</td>
<td>0.3339</td>
<td>ZB+SW+CLA</td>
<td>0.2715</td>
<td>ZB+CLA+DH</td>
</tr>
<tr>
<td>13</td>
<td>0.1379</td>
<td>ZB+CL+DH</td>
<td>0.3405</td>
<td>ZB+SW+CL</td>
<td>0.2756</td>
<td>ZB+SW+CL</td>
</tr>
<tr>
<td>14</td>
<td>0.1396</td>
<td>SW+Cap+DH</td>
<td>0.3599</td>
<td>SW+Cap+DH</td>
<td>0.2838</td>
<td>CLA+DH</td>
</tr>
<tr>
<td>15</td>
<td>0.1464</td>
<td>ZB+CLA+DH</td>
<td>0.3633</td>
<td>ZB+CLA+DH</td>
<td>0.2840</td>
<td>SW+CLA+DH</td>
</tr>
</tbody>
</table>

**Table 4:** Efficiency of combinations of control variates.

This table contains the results of test runs using various control variates simultaneously when pricing a 11 no-call 1 Bermudan swaption for various coupons of the underlying. To ease the exposition we have ranked the combinations based on their performance. EF denotes the efficiency factor i.e. the product of variance and average time per path. CV columns contain the combinations of control variates. We have used the following abbreviations: ZB=Zero Coupon, SW=Swap, CL=Subsample of Caplets, CLA=Caplets All and DH=Delta Hedge.
Figure 2: The graph shows the results from our estimation of the rate $\beta$ at which the bias decreases. $\text{eps}$ denotes the estimate of the bias and $m$ the number of inner simulation paths. These results are for the 6-11 years Bermudan swaption for various coupon rates of the underlying swap in the two-factor deterministic volatility scenario. The strategy used was strategy 1. It also contains simple OLS-regression results for minus the $\beta$ parameters. For practical applications a $\beta = 1$ seems to be a reasonable value.

make an efficient choice of the number of inner $m$ and outer paths $n$, which is basically a variance-bias trade-off where increases in $m$ decrease bias and increasing $n$ decreases variance. We balance these using the mean square error $\text{MSE}$ given by

$$\text{MSE} = \varepsilon^2 + \text{Var} \left( \hat{D} \right).$$

In appendix B we propose an allocation rule of the form that states how to increase the number of inner paths as we increase outer paths

$$m(n) \propto n^{\beta},$$

where $\beta$ is the rate of which the bias decreases. We also propose a procedure that gives an estimate of $\beta$. Basically we run a regression corresponding to equation (16) in appendix. The output of this procedure is shown in figure 2.

We have tested this method using strategy 1 in the two-factor deterministic volatility scenario using $n = 10,000$ antithetic paths and $m = 5, 10, 20, 40$ and $80$ in the regression. For these values of $m$, the size of the bias is still large relative to the standard deviation. The estimate of the true $D$ is based on $n^* = 10,000$.

20
and $m^* = 1,280$. Figure 2 also shows that the $\beta$ estimates are close to 1. This means that when we double the number of outer paths we should only increase the number of inner paths with a factor of $\sqrt{2}$.

6.4 Testing Strategies

So far we have only used the simple exercise strategy 1. To test the implications of this we now take a further look at the more complicated strategies 2-5. When using these strategies we have to estimate the value of the remaining core European swaptions at each possible exercise time. This is done using an approximative European swaption formula found in Andersen & Andreasen (1998) (see e.g. Brace & Musiela (1997)).

Andersen (2000) demonstrated that strategy 1 works well for most Bermudan swaptions across several scenarios of the economy. Short options on long swaps in the multifactor model, were the only cases where the more complicated strategies really picked up additional value. Our tests show the same and for that reason we concentrate on the 1 into 11 Bermudan swaption in the two-factor Libor market model with deterministic volatility. Computations are based on 50,000 AV paths and to reduce sample standard deviation we use the forward starting cap, covering the same period as the swaption, as a control variate.

Several interesting observations can be made from Table 5. First of all, the enhanced strategies do pick up additional value with a maximum of 12 basis points for the at-the-money swaption. Secondly, strategy 3 results in increased lower bounds compared to strategy 2, but is also slower to compute. This is due to the fact that all core European swaptions have to be computed in order to determine whether to exercise or not, whereas calculations of the core swaptions in strategy 2 should be skipped as soon as one is more valuable than intrinsic value. This effect is very pronounced in the out of the money case where calculation times are significantly different.

One possible explanation as to why strategy 3 picks up additional value could be that the barrier $H_3(t_i)$ in strategy 3 could be interpreted as the value of the deferred exercise premium at the exercise boundary. Remember that we are searching for the $H_3(t_i)$ that would make intrinsic value equal to the continuation value, $X(t_i) = H_3(t_i) + M(t_i)$. In strategies 1 and 2 the barrier should approximate the sum $H(t_i) + M(t_i)$ instead of just part of it. Furthermore, since the early/deferred exercise value goes to zero when the option is deep in- or out of the money this might be easier to approximate. This could indicate that in cases where we have good approximations for European core options, the parametrization of exercise strategies should be formulated in the early/deferred exercise values rather than just testing on the European value as in strategy 2.

Strategies 4 and 5 result in prices which are not significantly lower than those implied by strategies 2 and 3, respectively, confirming our initial expectations. However, computational savings are significant, especially for the options with long exercise periods which are also the options that would benefit from
This table summarizes the results of a test run of exercise strategies 1 to 5. It contains lower bound estimates, standard deviation and main simulation times for the 11 no call 1 Bermudan swaptions for various coupons. 50,000 antithetic paths and the cap as control variate was used in the simulation. Time is presented as minutes:seconds. The pickup from using the advanced strategies relative to the simple strategy 1 is included as well as the computation time relative to strategy 1. The strategies were estimated using 50,000 antithetic paths.

<table>
<thead>
<tr>
<th>Coupon</th>
<th>Strategy</th>
<th>Low bps</th>
<th>Pickup bps</th>
<th>Time min:sec</th>
<th>Relative Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08</td>
<td>1</td>
<td>1248.3  (0.4)</td>
<td></td>
<td>00:59</td>
<td>1.0</td>
</tr>
<tr>
<td>0.08</td>
<td>2</td>
<td>1251.0  (0.4)</td>
<td>2.7</td>
<td>19:01</td>
<td>19.3</td>
</tr>
<tr>
<td>0.08</td>
<td>3</td>
<td>1255.5  (0.4)</td>
<td>7.2</td>
<td>29:26</td>
<td>29.9</td>
</tr>
<tr>
<td>0.08</td>
<td>4</td>
<td>1251.8  (0.4)</td>
<td>3.5</td>
<td>01:11</td>
<td>1.2</td>
</tr>
<tr>
<td>0.08</td>
<td>5</td>
<td>1254.9  (0.3)</td>
<td>6.6</td>
<td>01:13</td>
<td>1.2</td>
</tr>
<tr>
<td>0.10</td>
<td>1</td>
<td>622.5 (0.4)</td>
<td></td>
<td>01:43</td>
<td>1.0</td>
</tr>
<tr>
<td>0.10</td>
<td>2</td>
<td>628.9 (0.4)</td>
<td>6.4</td>
<td>16:37</td>
<td>9.7</td>
</tr>
<tr>
<td>0.10</td>
<td>3</td>
<td>635.1 (0.3)</td>
<td>12.6</td>
<td>35:14</td>
<td>20.5</td>
</tr>
<tr>
<td>0.10</td>
<td>4</td>
<td>628.7 (0.4)</td>
<td>6.2</td>
<td>01:59</td>
<td>1.2</td>
</tr>
<tr>
<td>0.10</td>
<td>5</td>
<td>635.4 (0.3)</td>
<td>12.9</td>
<td>01:59</td>
<td>1.2</td>
</tr>
<tr>
<td>0.12</td>
<td>1</td>
<td>329.8 (0.3)</td>
<td></td>
<td>02:05</td>
<td>1.0</td>
</tr>
<tr>
<td>0.12</td>
<td>2</td>
<td>334.4 (0.3)</td>
<td>4.6</td>
<td>07:42</td>
<td>3.7</td>
</tr>
<tr>
<td>0.12</td>
<td>3</td>
<td>338.2 (0.3)</td>
<td>8.4</td>
<td>21:52</td>
<td>10.5</td>
</tr>
<tr>
<td>0.12</td>
<td>4</td>
<td>334.9 (0.3)</td>
<td>5.1</td>
<td>02:15</td>
<td>1.1</td>
</tr>
<tr>
<td>0.12</td>
<td>5</td>
<td>338.9 (0.3)</td>
<td>9.1</td>
<td>02:17</td>
<td>1.1</td>
</tr>
</tbody>
</table>
enhancement of the simple strategy. These savings do not only stem from the lower number of swaptions, but also the particular choice of swaptions. Remember that short maturity swaptions are much cheaper to evaluate than long maturity swaptions as we need to integrate the volatility to the maturity of the option.

Strategy 4 and strategy 5 are only about 20% slower than strategy 1 and much faster than strategies 2 and 3. The computation times for strategy 4 range from 6% to 29% of strategy 2 and as little as 4% to 10% for strategy 5 compared to strategy 3. These numbers would be even lower if we included computation time used in the presimulation.

The conclusion is quite clear. For Bermudan swaptions with many exercise times strategy 5 is preferred, as the additional computational cost is low while the lower bounds are as high as the ones from strategy 3.

6.5 Price sensitivity to the estimated barrier

We have tested the sensitivity of the exercise barrier on the upper and lower bound estimates. This is done by scaling the exercise barrier by a constant $\alpha$ and then computing lower and upper bound estimates. Noting here that scaling the barrier only changes the level of the barrier not the shape. The results are plotted in figure 3 for various values of $\alpha$. Interestingly, the lower bound seems to be more sensitive to the barrier scaling factor than the upper bound.

Again, tests have been performed in the two factor deterministic volatility scenario using exercise strategy 1. The Bermudan swaption is a 1 into 6 year with a coupon of 10%.

6.6 Upper Bound Calculations

We are now ready to compute the duality gaps and upper bound estimates. However, as our computational results are virtually identical to the findings in Andersen & Broadie (2001), we will mainly focus on the simplified strategies 4 and 5 as well as the effect of control variates on the estimation of the upper bound. There is, however, only minor discrepancy between our results and Andersen & Broadie (2001)’s results that we must dwell on. They find that application of strategy 3 in the two factor deterministic volatility scenario will generate duality gaps all within 4 bps. We cannot support this finding for the 11nc1 Bermudan swaptions. It is only a minor discrepancy but for completeness we report our results in Table 6. The reductions in the duality gap from applying strategy 3 is not that dramatic. Notice that strategy 5 performs just as well as strategy 3 as we cannot reject that they are the same. Furthermore strategy 5 is much faster.

As the primal-dual simulation algorithm is computationally demanding it calls for the application of efficiency improvement. We have tested the effect of control variates on the estimate of the duality gap. These results are presented in Table 7. It appears that the duality gaps are reduced at a reasonable computational cost.

23
Table 6: Duality gaps in the two-factor scenario.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Coupon</th>
<th>0.08</th>
<th>0.10</th>
<th>0.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18.9 (1.3)</td>
<td>20.0 (1.1)</td>
<td>14.1 (1.0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13.5 (0.8)</td>
<td>13.5 (0.7)</td>
<td>9.7 (0.6)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8.4 (0.5)</td>
<td>6.8 (0.4)</td>
<td>3.4 (0.3)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>13.1 (0.8)</td>
<td>14.0 (0.7)</td>
<td>7.8 (0.5)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8.8 (0.5)</td>
<td>6.5 (0.4)</td>
<td>3.9 (0.3)</td>
<td></td>
</tr>
</tbody>
</table>

Duality gaps for the 11 no-call 1 Bermudan swaption for different strategies and coupons in the two factor deterministic volatility scenario. The numbers have been computed using $n = 750$ and $m = 300$ in the primal-dual simulation algorithm.

Table 7: Effect of control variate on duality gaps.

<table>
<thead>
<tr>
<th>$T_0$</th>
<th>Coupon</th>
<th>None $\hat{D}_0$</th>
<th>Time</th>
<th>Cap $\hat{D}^{Cap}_0$</th>
<th>Time</th>
<th>Reduce</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.08</td>
<td>0.19 (0.02)</td>
<td>00:33</td>
<td>0.09 (0.01)</td>
<td>00:45</td>
<td>53%</td>
<td>39%</td>
</tr>
<tr>
<td>3</td>
<td>0.10</td>
<td>0.24 (0.02)</td>
<td>00:44</td>
<td>0.11 (0.01)</td>
<td>00:53</td>
<td>54%</td>
<td>21%</td>
</tr>
<tr>
<td>3</td>
<td>0.12</td>
<td>0.30 (0.04)</td>
<td>00:48</td>
<td>0.22 (0.03)</td>
<td>00:54</td>
<td>27%</td>
<td>13%</td>
</tr>
<tr>
<td>6</td>
<td>0.08</td>
<td>2.15 (0.13)</td>
<td>03:58</td>
<td>0.88 (0.06)</td>
<td>05:15</td>
<td>59%</td>
<td>32%</td>
</tr>
<tr>
<td>6</td>
<td>0.10</td>
<td>1.99 (0.12)</td>
<td>06:04</td>
<td>0.89 (0.06)</td>
<td>06:54</td>
<td>55%</td>
<td>14%</td>
</tr>
<tr>
<td>6</td>
<td>0.12</td>
<td>0.99 (0.09)</td>
<td>07:09</td>
<td>0.43 (0.05)</td>
<td>07:41</td>
<td>57%</td>
<td>7%</td>
</tr>
<tr>
<td>11</td>
<td>0.08</td>
<td>14.17 (0.52)</td>
<td>23:15</td>
<td>7.63 (0.30)</td>
<td>28:03</td>
<td>46%</td>
<td>21%</td>
</tr>
<tr>
<td>11</td>
<td>0.10</td>
<td>11.96 (0.42)</td>
<td>34:50</td>
<td>6.71 (0.27)</td>
<td>38:40</td>
<td>44%</td>
<td>11%</td>
</tr>
<tr>
<td>11</td>
<td>0.12</td>
<td>6.38 (0.30)</td>
<td>41:50</td>
<td>3.96 (0.18)</td>
<td>44:36</td>
<td>38%</td>
<td>7%</td>
</tr>
</tbody>
</table>

The effect on the duality gaps when we use the cap as control variate and strategy 5. Numbers are from the two factor deterministic volatility scenario computed using $n = 1500$ and $m = 40$ antithetic paths in the primal-dual simulation algorithm. All Bermudan Swaptions have a lockout period of one 1 year.
Figure 3: Illustration of the sensitivity to the exercise strategy. The figure shows the lower and upper bound for an ATM, 6-year Bermudan swaption with a 1-year lockout period as we scale the exercise barrier. The example is the deterministic volatility scenario using exercise strategy 1.

Notice that we have changed the relative amount of work used in the inner and outer computations relative to the previous calculations. This is due to the analysis in section 6.3. Particularly when combined with the control variate the estimates of the lower bound are not significantly different from the ones presented in e.g. table 6, but the standard deviations are lower and the calculation takes only about one third of the previous computations.

Tests have also been performed using several control variates in the inner loop, however remember that we determine the variance by minimizing $\beta$ from OLS regression, so we have experienced problems due to the small number of inner loops.

6.7 Stochastic Volatility

In this section we investigate the effect of stochastic interest rate volatility on the size of the duality gap. The parameters in our test case have been taken from a scenario in Andersen & Brotherton-Ratcliffe (2001): $\theta_V = 1$, $\kappa_V = 1$, $\varepsilon_V = 1.4$ and $\psi(x) = x^{0.75}$.

We test whether the simple exercise strategy 1 still performs well when the forward curve dynamics exhibit stochastic volatility. This is done by examining the duality gap for various parameters in the variance of volatility process.
Notice that lower bounds are not comparable as the volatility of the forward rates varies as we vary the parameters. What could be a problem for the simple exercise strategy is that it does not distinguish whether the current volatility \( V_t \) is higher or lower than the long term mean. So a priori we would expect that the duality gap increases as the volatility of the variance process increases or when the mean reversion rate decreases. By estimating duality gaps we are able to assess the average present value loss that swaption holders incur from ignoring this feature of the yield curve dynamics.\(^6\)

The results in Tables 8 and 9 confirm these a priori expectations. We have calculated the duality gap for a 6-year contract with 1-year lockout for various coupon rates of the underlying swap. Strategy 1 is reestimated using 50,000 (AS) paths in the presimulation for each value of the volatility parameter \( \varepsilon_V \). Based on our observations in section 6.2.2, we apply the zero coupon bonds as a control variate to reduce the variance, as no closed form solutions for caps and swaptions are available\(^7\), and we use \( n = 1,500 \) and \( m = 300 \) antithetic paths in the main simulation algorithm.

![Figure 4:](image)

As expected, higher volatility of the variance process increases the duality gap as the simple strategy fails to incorporate the volatility aspect in the exercise

\(^6\)Longstaff et al. (2001) argue that capturing the exact dynamics of the yield curve is very important for Bermudan swaptions.

\(^7\)The approximations for European caps and swaptions given by Andersen & Brotherton-Ratcliffe (2001) are extremely precise and would likely be applicable as control variates as the bias would be very low relative to the variance. However, this remains to be investigated.
Table 8: Stochastic volatility - volatility of variance.

<table>
<thead>
<tr>
<th>Coupon</th>
<th>$\varepsilon_V$</th>
<th>Lower</th>
<th>StdLow</th>
<th>$\bar{D}_0$</th>
<th>Std $\bar{D}_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08</td>
<td>0.25</td>
<td>750.03</td>
<td>1.50</td>
<td>3.76</td>
<td>0.22</td>
</tr>
<tr>
<td>0.08</td>
<td>0.50</td>
<td>751.24</td>
<td>1.58</td>
<td>4.16</td>
<td>0.23</td>
</tr>
<tr>
<td>0.08</td>
<td>0.75</td>
<td>748.41</td>
<td>1.56</td>
<td>4.47</td>
<td>0.24</td>
</tr>
<tr>
<td>0.08</td>
<td>1.00</td>
<td>749.87</td>
<td>1.73</td>
<td>5.87</td>
<td>0.35</td>
</tr>
<tr>
<td>0.08</td>
<td>1.40</td>
<td>747.76</td>
<td>1.73</td>
<td>7.15</td>
<td>0.47</td>
</tr>
<tr>
<td>0.08</td>
<td>1.50</td>
<td>747.11</td>
<td>1.74</td>
<td>8.21</td>
<td>0.49</td>
</tr>
<tr>
<td>0.08</td>
<td>1.75</td>
<td>755.15</td>
<td>2.01</td>
<td>8.38</td>
<td>0.45</td>
</tr>
<tr>
<td>0.10</td>
<td>0.25</td>
<td>317.53</td>
<td>2.98</td>
<td>4.82</td>
<td>0.24</td>
</tr>
<tr>
<td>0.10</td>
<td>0.50</td>
<td>313.46</td>
<td>3.02</td>
<td>5.58</td>
<td>0.25</td>
</tr>
<tr>
<td>0.10</td>
<td>0.75</td>
<td>311.03</td>
<td>3.17</td>
<td>6.70</td>
<td>0.31</td>
</tr>
<tr>
<td>0.10</td>
<td>1.00</td>
<td>311.30</td>
<td>3.33</td>
<td>8.45</td>
<td>0.38</td>
</tr>
<tr>
<td>0.10</td>
<td>1.40</td>
<td>314.03</td>
<td>3.42</td>
<td>10.92</td>
<td>0.56</td>
</tr>
<tr>
<td>0.10</td>
<td>1.50</td>
<td>309.74</td>
<td>3.58</td>
<td>11.30</td>
<td>0.50</td>
</tr>
<tr>
<td>0.10</td>
<td>1.75</td>
<td>326.87</td>
<td>3.99</td>
<td>12.99</td>
<td>0.62</td>
</tr>
<tr>
<td>0.12</td>
<td>0.25</td>
<td>124.10</td>
<td>2.43</td>
<td>2.58</td>
<td>0.20</td>
</tr>
<tr>
<td>0.12</td>
<td>0.50</td>
<td>132.74</td>
<td>2.77</td>
<td>2.94</td>
<td>0.19</td>
</tr>
<tr>
<td>0.12</td>
<td>0.75</td>
<td>123.18</td>
<td>2.63</td>
<td>4.10</td>
<td>0.21</td>
</tr>
<tr>
<td>0.12</td>
<td>1.00</td>
<td>124.22</td>
<td>2.89</td>
<td>4.17</td>
<td>0.26</td>
</tr>
<tr>
<td>0.12</td>
<td>1.40</td>
<td>122.64</td>
<td>2.90</td>
<td>5.33</td>
<td>0.31</td>
</tr>
<tr>
<td>0.12</td>
<td>1.50</td>
<td>131.23</td>
<td>3.19</td>
<td>5.82</td>
<td>0.39</td>
</tr>
<tr>
<td>0.12</td>
<td>1.75</td>
<td>143.53</td>
<td>3.42</td>
<td>6.92</td>
<td>0.42</td>
</tr>
</tbody>
</table>

This table illustrates the degree of inoptimality of the strategy 1. The contracts are 6 year contracts with 1 year lock out for three degrees of moneyness. As the volatility of the variance process $V(t)$ increases, the duality cap $\bar{D}_0$ increases as well. The basic scenario is the two-factor libor market model with a variance of the volatility process with the following initial value and parameters: $V(0) = \theta_V = 1$, $\kappa_V = 1$, $\psi(x) = x^{0.75}$. All zero coupon bonds were used as control variates for $n = 1500$ antithetic and $m = 300$. Barriers were estimated using 50,000 (AS) paths.
decision. The same conclusion holds for the speed of mean reversion in the variance process. Higher mean reversion keeps the process closer to the long term mean.

Table 9: Stochastic volatility - mean reversion in variance.

<table>
<thead>
<tr>
<th>Coupon</th>
<th>$\kappa_V$</th>
<th>Lower</th>
<th>StdLow</th>
<th>$\hat{D}_0$</th>
<th>Std$\hat{D}_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08</td>
<td>0.25</td>
<td>747.04</td>
<td>1.86</td>
<td>10.74</td>
<td>0.63</td>
</tr>
<tr>
<td>0.08</td>
<td>0.50</td>
<td>746.35</td>
<td>1.78</td>
<td>8.19</td>
<td>0.47</td>
</tr>
<tr>
<td>0.08</td>
<td>0.75</td>
<td>745.32</td>
<td>1.54</td>
<td>6.79</td>
<td>0.38</td>
</tr>
<tr>
<td>0.08</td>
<td>1.00</td>
<td>748.51</td>
<td>1.65</td>
<td>6.07</td>
<td>0.35</td>
</tr>
<tr>
<td>0.08</td>
<td>1.25</td>
<td>750.67</td>
<td>1.71</td>
<td>4.92</td>
<td>0.28</td>
</tr>
<tr>
<td>0.10</td>
<td>0.25</td>
<td>296.62</td>
<td>3.73</td>
<td>15.19</td>
<td>0.67</td>
</tr>
<tr>
<td>0.10</td>
<td>0.50</td>
<td>300.01</td>
<td>3.39</td>
<td>10.91</td>
<td>0.56</td>
</tr>
<tr>
<td>0.10</td>
<td>0.75</td>
<td>300.93</td>
<td>3.25</td>
<td>8.81</td>
<td>0.39</td>
</tr>
<tr>
<td>0.10</td>
<td>1.00</td>
<td>311.72</td>
<td>3.22</td>
<td>7.21</td>
<td>0.34</td>
</tr>
<tr>
<td>0.10</td>
<td>1.25</td>
<td>314.11</td>
<td>3.24</td>
<td>7.44</td>
<td>0.35</td>
</tr>
<tr>
<td>0.12</td>
<td>0.25</td>
<td>121.17</td>
<td>3.25</td>
<td>7.82</td>
<td>0.48</td>
</tr>
<tr>
<td>0.12</td>
<td>0.50</td>
<td>123.35</td>
<td>3.08</td>
<td>6.92</td>
<td>0.45</td>
</tr>
<tr>
<td>0.12</td>
<td>0.75</td>
<td>117.58</td>
<td>2.87</td>
<td>4.73</td>
<td>0.30</td>
</tr>
<tr>
<td>0.12</td>
<td>1.00</td>
<td>122.46</td>
<td>2.85</td>
<td>4.45</td>
<td>0.30</td>
</tr>
<tr>
<td>0.12</td>
<td>1.25</td>
<td>125.58</td>
<td>2.72</td>
<td>3.50</td>
<td>0.23</td>
</tr>
</tbody>
</table>

This table illustrates the degree of inoptimality of the strategy 1. The contracts are 6-year contracts with 1-year lock out for three degrees of moneyness. As the mean reversion of the variance process $\kappa_V$ increases the duality cap $\hat{D}_0$ decreases.

The basic scenario is the two-factor libor market model with a variance of the volatility process with the following initial value and parameters: $V(0) = \theta_V = 1$, $\kappa_V = 1.4$, $\psi(x) = x^{0.75}$. All zero coupon bonds were used as control variates for $n = 1500$ antithetic and $m = 300$. Barriers were estimated using 50,000 (AS) paths.

The sizes of the duality gaps are not very large even for rather extreme values of the parameters in the volatility process, but still economically significant. As already mentioned, Andersen & Brotherton-Ratcliffe (2001) provide very accurate approximations for European caps and swaptions that could be used to enhance strategy 1 to 5. Alternatively, one could use the Least-Square Monte Carlo approach of Longstaff & Schwartz (2001) including the current level of the variance process in the basis functions. As a final test we have implemented the LS-MC method following Bjørregaard Pedersen (1999). He concludes that a simple specification including a constant the first two powers of the intrinsic value and bankbook as well as the cross product in the regressions gave good results. We denote this strategy $LS1$. To account for stochastic volatility in the basis function we furthermore include the level of the variance process $V_t$ in strategy $LS2$. Table 10 illustrates that the Andersen approach performs slightly better than the $LS1$ approach. As expected the $LS2$ strategy is superior to the other two as it picks up additional value using information in the variance level.
Figure 5:

7 Conclusion

This paper has addressed the issue of pricing Bermudan swaptions in a Libor market model. A control variate technique especially tailored for American options, which was recently proposed in Rasmussen (2002), was implemented and tested in the Libor market model. Furthermore, we demonstrated how to handle dividend paying control variates, and a range of controls were tested in the Bermudan swaption case. The results were reasonable showing reductions in standard deviations in the order of 3 to 5. The combination of the zero coupon bonds and caplets underlying the swaption performed well for a range of Bermudan swaptions.

A simplification of the strategies proposed in Andersen (2000) was demonstrated to give equally lower bounds at a significantly smaller computational effort. In particular, for the strategy resulting in the highest lower bounds the computation times were reduced to between 4% and 10% depending on moneyness of the option.

We also demonstrated the effect of the control variate technique on the duality gap from the Andersen & Broadie (2001) Primal-Dual algorithm, by applying the control variate technique in the nested simulations. The results gave significant improvements. However, we still note that the Primal-Dual algorithm is not suited for real time work. However, it is still extremely important due to the generality and reasonable computational effort. In particular it will enable us to determine when strategies are "good enough" so that we can use the lower
<table>
<thead>
<tr>
<th>Strategy</th>
<th>( T_s )</th>
<th>( T_e )</th>
<th>Coupon</th>
<th>( Lower )</th>
<th>( Time )</th>
<th>( Lower )</th>
<th>( Time )</th>
<th>( Lower )</th>
<th>( Time )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 6 0.08</td>
<td>750.4  (0.3)</td>
<td>0:14</td>
<td>746.7 (0.4)</td>
<td>0:16</td>
<td>751.2 (0.4)</td>
<td>0:16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 6 0.10</td>
<td>313.2  (0.6)</td>
<td>0:27</td>
<td>311.3 (0.7)</td>
<td>0:29</td>
<td>315.7 (0.7)</td>
<td>0:29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 6 0.12</td>
<td>130.8  (0.6)</td>
<td>0:33</td>
<td>128.5 (0.6)</td>
<td>0:34</td>
<td>132.1 (0.6)</td>
<td>0:34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 11 0.08</td>
<td>1245.0 (0.8)</td>
<td>0:39</td>
<td>1237.9 (0.9)</td>
<td>0:48</td>
<td>1246.8 (0.9)</td>
<td>0:48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 11 0.10</td>
<td>616.9  (1.2)</td>
<td>1:17</td>
<td>617.9 (1.4)</td>
<td>1:25</td>
<td>624.2 (1.4)</td>
<td>1:23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 11 0.12</td>
<td>335.4  (1.1)</td>
<td>1:36</td>
<td>335.5 (1.2)</td>
<td>1:40</td>
<td>341.0 (1.2)</td>
<td>1:39</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This figure contains a simple comparison of the Andersen and the Least-Square approach. Strategy LS 1 incorporates the intrinsic value of the underlying swap and the bankbook. LS 2 furthermore incorporates the current level of the variance process \( V_t \). Lower denotes the lower bound estimate. Time is calculation time in minutes and seconds. Numbers are based on 50,000 AS paths with zero coupon bonds as controls in a two factor model with stochastic volatility. The strategies were estimated using 5,000 AS paths.

When setting up the primal-dual simulation algorithm one faces a variance-bias trade off. Our tests indicated that one should only increase the number of nested paths (reducing bias) with \( \sqrt{2} \) when doubling the number of outer paths (reducing variance).

Finally, we considered an extended Libor market model with stochastic volatility developed in Andersen & Brotherton-Ratcliffe (2001), and demonstrate that this will increase the duality gap and make the enhanced strategies necessary also for short Bermudan swaptions. The Least-Square MC proposed in Longstaff & Schwartz (2001) was found to be a good substitute in case of stochastic volatility.
References


A Simulation

The simulation of the Libor market model is carried out under the measure $\mathbb{Q}$. We apply simple log Euler Schemes to (10) on a simulation time grid $0 \leq t_0 < t_1 < \cdots < t_{I+1} = T_{K+1}$. It is not necessary that the time grid is and the LIBOR maturity structure $T_0 < T_1 < \cdots < T_{K+1}$ are the same, but we will require that \{T_0, T_1, \ldots, T_{K+1}\} $\subseteq \{t_0, t_1, \ldots, t_{I+1}\}$ i.e. that the tenor dates are among the simulated dates. Thus we simulate the following

$$\hat{F}_k(t_{i+1}) = \hat{F}_k(t_i) \exp \left( \lambda_k(t_i) \frac{\varphi(\hat{F}_k(t_i))}{\hat{F}_k(t_i)} \sqrt{\hat{V}(t_i)} \left( \hat{\mu}_k(t_i) \Delta_i + \sqrt{\Delta_i} \hat{\epsilon}_i \right) \right) \cdot \exp \left( -\frac{1}{2} \lambda_k(t_i) \left( \frac{\varphi(\hat{F}_k(t_i))}{\hat{F}_k(t_i)} \right)^2 \hat{V}(t_i) \lambda_k(t_i)^T \Delta_i \right),$$

where $t_{i+1} - t_i = \Delta_i$. For simplicity we will only use equally spaced tenor space and simulation grid where $\delta = \Delta$. So far our mapping function $n(t)$, used in the drift term (12), has been defined as left continuous.

So far this has not been a problem as in the continuous time setup this is without importance - for the obvious reasons this is not the case for the discrete time setup. Andersen & Andreasen (1998) argue that even though we operate with a left continuous mapping function in continuous time, we should use a right continuous mapping function in discrete time.

The $V$ process in (11) is simulated using a Gaussian Ornstein-Uhlenbeck process proposed in Andersen & Brotherton-Ratcliffe (2001). We run $n_i$ steps within each interval $[t_i, t_{i+1}]$ such that $t_j = t_i + j \cdot (t_{i+1} - t_i) / n_i$, $j = 1, \ldots, n_i$

$$\hat{V}(t_{j+1}) = \theta_V + \left( \hat{V}(t_j) - \theta_V \right) e^{-\kappa_V (t_{j+1} - t_j)} + \sqrt{\kappa_V} \psi(V(t_i)) \sqrt{\frac{1}{2} \kappa_V \left( 1 - e^{-2\kappa_V (t_{j+1} - t_j)} \right)}.$$

B Bias Variance Trade-off

Let $\hat{D}(n, m) = \sum_{i=1}^{n} X_{i}^m$ denote the point estimate for $D$ using $n$ inner simulations and $m$ nested simulations. Assume that the estimator is consistent but biased

$$E \left( \hat{D}_m \right) = D_m \rightarrow D$$

for $m \rightarrow \infty$.

We assume that the bias decreases with rate $\beta$ as a function of $m$ as

$$D_m - D \approx b \cdot \left( \frac{1}{m} \right)^\beta.$$

The expected time $T(m)$ per path using $m$ inner simulations is assumed to be

$$T(m) = c \cdot m^{\eta}.$$
What we are looking for is an allocation rule. To determine \( n(\cdot) \) for a given \( m \) we postulate the following rule

\[
n(\cdot) = aT(\cdot)^\alpha = ac^\alpha \cdot m^{\alpha\eta}
\]

and use the estimator

\[
\hat{D}(m) = \hat{D}(n(m), m).
\]

The variance obtained using this rule

\[
\text{Var}\left(\hat{D}(m)\right) = \frac{\sigma_m^2}{n(m)}
\]

if we assume that \( \sigma_m^2 \to \sigma^2 \) for \( m \to \infty \).

Given all these assumptions we let

\[
\text{MSE}\left(\hat{D}(m)\right) = \left(\hat{D}(m) - D\right)^2 + \text{Var}(D_m)
\]

approximating by

\[
\approx \left( b \cdot \left( \frac{1}{m} \right)^\beta \right)^2 + \frac{\sigma_m^2}{n(m)}
\]

\[
\approx b \cdot \left( \frac{1}{m} \right)^{2\beta} + \frac{\sigma_m^2}{aT(\cdot)\alpha}
\]

\[
\approx b \cdot m^{-2\beta} + \frac{\sigma_m^2}{ac^\alpha} \cdot m^{-\alpha\eta}
\]

So what we need to balance is

\[-2\beta = -\alpha\eta \Leftrightarrow \alpha = 2\frac{\beta}{\eta}.
\]

This means that doubling the number of inner paths \( m \) we have

\[n(\cdot) \propto m^{2\alpha\eta} = m^{2\beta}.
\]

Hence, we have to estimate the rate \( \beta \) at which the bias decreases.

We propose the following procedure. First observe that

\[
\hat{D}(n(m), m) - D = b \cdot m^{-\beta} + \frac{\sigma_m}{\sqrt{n}} z_n, \; z_n \sim N(0, 1).
\]

As we don’t know \( D \), we provide an estimate of this based on relatively high values of \( n \) and \( m \) say \( n^* \) and \( m^* \). Hence, the following equation gives a biased but consistent estimate of the bias using \( m \) inner paths given that \( n, n^* \) and \( m^* \) are high.

\[
\varepsilon_m = \hat{D}(n(m), m) - \hat{D}(n^*(m^*), m^*)
\]

\[
\approx b \cdot m^{-\beta} + \frac{\sigma_m}{\sqrt{n^*}} z_n - b \cdot (m^*)^{-\beta} + \frac{\sigma_m}{\sqrt{n^*}} z_n^*
\]

\[
= b \cdot m^{-\beta} - \varepsilon_n^*
\]

\[
\Leftrightarrow \varepsilon_m + \varepsilon_n^* = b \cdot m^{-\beta}
\]
Assuming that the right side is positive (we can always increase \( n, n^* \) and \( m^* \) to ensure this) we take logarithms and apply a first order Taylor expansion to get

\[
\ln (\varepsilon_m) + \frac{1}{\varepsilon_m} \varepsilon^*_n \approx \ln (b) - \beta \ln(m). \tag{16}
\]

Now we can get a rough estimate of \( \beta \) by regression.