Foresight Bias and Suboptimality Correction in Monte-Carlo Pricing of Options with Early Exercise: Classification, Calculation & Removal

Christian P. Fries
email@christian-fries.de

November 29, 2005

(Version 1.1.4)

Abstract

In this paper we investigate the so called *foresight bias* that may appear in the Monte-Carlo pricing of Bermudan and compound options if the exercise criteria is calculated by the same Monte-Carlo simulation as the exercise values. The standard approach to remove the foresight bias is to use two independent Monte-Carlo simulations: One simulation is used to estimate the exercise criteria (as a function of some state variable), the other is used to calculate the exercise price based on this exercise criteria. We shall call this the *numerical removal of the foresight bias*.

In this paper we give an exact definition of the *foresight bias* in closed form and show how to apply an analytical correction for the foresight bias.

Our numerical results show that the analytical removal of the foresight bias gives similar results as the standard numerical removal of the foresight bias. The analytical correction allows for a simpler coding and faster pricing, compared to a numerical removal of the foresight bias.

Our analysis may also be used as an indication of when to neglect the foresight bias removal altogether. While this is sometimes possible, neglecting foresight bias will break the possibility of parallelization of Monte-Carlo simulation and may be inadequate for Bermudan options with many exercise dates (for which the foresight bias may become a Bermudan option on the Monte-Carlo error) or for portfolios of Bermudan options (for which the foresight bias grows faster than the Monte-Carlo error).

In addition to an analytical removal of the foresight bias we derive an analytical correction for the suboptimal exercise due to the uncertainty induced by the Monte-Carlo error. The combined correction for foresight bias (biased high) and suboptimal exercise (biased low) removed the systematic bias even for Monte-Carlo simulations with very small number of paths.

Acknowledgment

I would like to thank Markus Meister for his interest in an early version of this paper and a very stimulating discussion as well as for his constructive comments to this and my other works. I thank Oliver Dauben for his continuous advisory with respect to Java^{TM} technology.[1] I am grateful to Rosemarie Philippi for her support.

[1] Java^{TM} is a registered trademark of Sun Microsystems, Inc.

Contents

[2] The reader familiar with Bermudan option pricing in Monte-Carlo may skip Sections 2 to 4.

1 Introduction

Estimating conditional expectation in a Monte-Carlo simulation is a frequent challenge in the pricing of complex derivatives, e.g. options on options or Bermudan options. Here the payoff is given by the optimal choice between an underlying[3] and the value of a future option, the latter given by a conditional expectation.

The problem of pricing Bermudan or American style options by Monte Carlo is tackled by numerous approaches, like binning (aka. state space partitioning), parameter optimization methods, regression methods, dual methods (optimal stopping), see e.g. [2, 6, 7, 9] and references there in.

Here we concentrate on those methods that rely on an estimator for the conditional expectation. If the conditional expectation or a suitable approximation for it is known, the optimal exercise strategy (i.e. the optimal stopping time) may be calculated by the backward algorithm. Then the price of the Bermudan option is the (unconditional) expectation of the values under optimal exercise.

If however the conditional expectation estimator (hence the optimal exercise strategy) and the pricing of the Bermudan share the same Monte-Carlo simulation, a systematic positive bias will occur: the foresight bias. The exercise strategy may be super optimal by exercising optimal on the common Monte-Carlo error. In other words:

The foresight bias is the value of the option on the Monte Carlo error.

It is straight forward to eliminate the foresight bias by introducing two independent Monte-Carlo simulations.[4] However this makes the pricing slow and the implementation a bit cumbersome. We show that a simple calculation allows us to

a) give a very exact estimate of the foresight bias and

b) apply an correction that removes the foresight bias (without the need of a second independent Monte Carlo simulation).

In fact, we simple calculate the price of the option on the Monte Carlo error. This allows us to analytically correct the foresight bias making pricing faster and coding leaner.

1.1 Plan of the Paper

We start with a small introduction to the pricing of Bermudan options in Monte Carlo. For a more detailed introduction to the pricing of Bermudan options in Monte Carlo and a review of the literature see e.g. [6, 9] and references therein. In Section 2 we give a fairly general definition of an Bermudan option and define the optimal exercise time and the optimal exercise value. In Section 3 we present the well known backward algorithm by which the optimal exercise value and thus the Bermudan option value may be calculated. The main ingredient to the backward algorithm is the exercise criteria and here an estimator for the conditional expectation. The estimation of conditional expectation in Monte-Carlo is shortly reviewed in Section 4.

In Section 5 we will then present an estimator for the *foresight bias* and discuss the (analytic) removal of the foresight bias by a small additional term in the backward algorithm. We conclude with some numerical results in Section 6.

[3] We use a rather general definition of Bermudan option, where the underlying may be different at each exercise date and where it may be a constant (like a strike for a compound option) or stochastic.

[4] The exercise criteria will still be influence by the Monte Carlo error one of the simulations, but exercising will give the (independent) Monte Carlo error of the other simulation.

2 Bermudan Options: Notation

We now give a fairly general definition of an Bermudan option and fix notation. Let $\{T_i\}_{i=1,\ldots,n}$ denote a set of exercise dates and $\{V_{\mathrm{underl},i}\}_{i=1,\ldots,n}$ a corresponding set of underlyings. The Bermudan option is the right to receive at one and only one time T_i the corresponding underlying $V_{\mathrm{underl},i}$ (with $i = 1,\ldots,n$) or receive nothing.

At each exercise date T_i, the optimal strategy compares the value of the product upon exercise with the value of the product upon non-exercise and chooses the larger one. Thus the value of the Bermudan is given recursively

$$V_{\mathrm{berm}}(T_i,\ldots,T_n;T_i) \;:=\; \max\big(V_{\mathrm{berm}}(T_{i+1},\ldots,T_n;T_i)\;,\;V_{\mathrm{underl},i}(T_i)\big), \qquad (1)$$

where $V_{\mathrm{berm}}(T_n;T_n) := 0$ and $V_{\mathrm{underl},i}(T_i)$ denotes the value of the underlying $V_{\mathrm{underl},i}$ at exercise date T_i.

2.1 Relative Prices

Let $N(t)$ denote the time t value of a chosen Numéraire and $\mathbb{Q}^N$ the corresponding pricing measure, see [1]. Since the conditional expectation (w.r.t. the pricing measure) of a Numéraire relative price is a Numéraire relative price the presentation will be simplified by considering the Numéraire relative quantities. We will then define:

$$\tilde{V}_{\mathrm{underl},i}(T_j) := \frac{V_{\mathrm{underl},i}(T_j)}{N(T_j)} \qquad \text{und} \qquad \tilde{V}_{\mathrm{berm},i}(T_j) := \frac{V_{\mathrm{berm}}(T_i,\ldots,T_n;T_j)}{N(T_j)},$$

thus we have

$$\begin{aligned}\tilde{V}_{\mathrm{berm},n} &\equiv 0,\\ \tilde{V}_{\mathrm{berm},i+1}(T_i) &= \mathrm{E}^{\mathbb{Q}^N}\big(\tilde{V}_{\mathrm{berm},i+1}(T_{i+1}) \mid \mathcal{F}_{T_i}\big),\\ \tilde{V}_{\mathrm{berm},i}(T_i) &= \max\big(\tilde{V}_{\mathrm{berm},i+1}(T_i)\;,\;\tilde{V}_{\mathrm{underl},i}(T_i)\big),\end{aligned}$$

where $\{\mathcal{F}_t\}$ denotes the filtration. The relative prices are marked by a tilde. The Bermudan pricing consists of finding the relative value of the longest Bermudan $\tilde{V}_{\mathrm{berm},1}$ as seen in T_0 (today). We write shortly $\tilde{V}_{\mathrm{berm}}(T_0) := \tilde{V}_{\mathrm{berm},1}(T_0)$.

2.2 Bermudan Option as Optimal Exercise Problem

A Bermudan option consists of the right to receive one (and only one) of the underlyings $V_{\mathrm{underl},i}$ at the corresponding exercise date T_i. The recursive definition (1) represents the optimal exercise strategy in each exercise time. We formalize this optimal exercise strategy:

For a given path $\omega \in \Omega$ let

$$T(\omega) := \min\{T_i : V_{\mathrm{berm},i+1}(T_i,\omega) < V_{\mathrm{underl},i}(T_i,\omega)\}.$$

The definitions of T gives a description of the exercise strategy: $T(\omega)$ is the optimal exercise time on a given path ω. It should be noted that $\{T \leq T_k\} \subset \mathcal{F}_{T_k}$ (i.e. T is a stopping time).

2.3 Bermudan Option Value as single (unconditioned) Expectation: The Optimal Exercise Value

With the definition of the optimal exercise strategy T it is possible to define a random variable which allows to express the Bermudan option value as a single (unconditioned) expectation. With

$$\tilde{U}(T_i) := \tilde{V}_{\mathrm{underl},i}(T_i) \quad i = 1,\ldots,n$$

denoting the relative price of the i-th underlying upon its exercise date T_i we have for the Bermudan value

$$\tilde{V}_{\mathrm{berm}}(T_0) = \mathrm{E}^{\mathbb{Q}}\big(\tilde{U}(T) \mid \mathcal{F}_{T_0}\big).$$

The random variable $\tilde{U}(T)$ may be calculated directly using the *Backward Algorithm*. We will consider this in the next section and conclude by giving $\tilde{U}(T)$ a name:

Definition 1 (Option Value upon Optimal Exercise):
Let $\tilde{U}$ be the stochastic process who's time t value $\tilde{U}(t)$ is the (Numéraire relative) option value received upon exercise in t. Let T be the optimal exercise strategy. Then the random variable $\tilde{U}(T)$, where

$$\tilde{U}(T)[\omega] := \tilde{U}(T(\omega), \omega),$$

is the (Numéraire relative) *option value received upon optimal exercise*. The (Numéraire relative) Bermudan option value is given by $\mathrm{E}^{\mathbb{Q}}\big(\tilde{U}(T) \mid \mathcal{F}_{T_0}\big)$.

Thus the value of $V_{\mathrm{berm}}(T_1, \ldots, T_n)$ may be expressed through a single expectation conditioned to T_0 and does not need any calculation of a conditional expectation at later times, *if* we have the optimal exercise date $T(\omega)$ for any path ω.

3 The Backward Algorithm

The random variable $\tilde{U}(T)$ may be derived in a Monte-Carlo simulation through the Backward algorithm, *given* the exercise criteria (1), i.e. the conditional expectation. The algorithm consists of the application of the recursive definition of the Bermudan value in (1) with a slight modification. Let:

Induction start:

$$\tilde{U}_{n+1} \equiv 0$$

Induction step $i+1 \rightarrow i$ for $i = n, \ldots, 1$:

$$\tilde{U}_i = \begin{cases} \tilde{U}_{i+1} & \text{if } \tilde{V}_{\mathrm{underl},i}(T_i) < \mathrm{E}^{\mathbb{Q}}(\tilde{U}_{i+1} | \mathcal{F}_{T_i}) \\ \tilde{V}_{\mathrm{underl},i}(T_i) & \text{else.} \end{cases}$$

From the tower law we have by induction $\mathrm{E}^{\mathbb{Q}}(\tilde{U}_{i+1} | \mathcal{F}_{T_i}) = \mathrm{E}^{\mathbb{Q}}(\tilde{V}_{\mathrm{berm},i+1}(T_i) | \mathcal{F}_{T_i})$ and thus

$$\tilde{V}_{\mathrm{berm}}(T_1, \ldots, T_n, T_0) = \mathrm{E}^{\mathbb{Q}}(\tilde{U}_1 | \mathcal{F}_{T_0}) \tag{2}$$

and $\tilde{U}_1 = \tilde{U}(T)$ with the notation from the previous section.

The recursive definition of $\tilde{U}_i$ differs from the recursive definition of $\tilde{V}_{\mathrm{berm},i}(T_i)$. We have

$$\tilde{U}_i = \begin{cases} \tilde{U}_{i+1} & \text{if } \tilde{V}_{\mathrm{underl},i}(T_i) < \mathrm{E}^{\mathbb{Q}}(\tilde{U}_{i+1} | \mathcal{F}_{T_i}) \\ \tilde{V}_{\mathrm{underl},i}(T_i) & \text{else,} \end{cases}$$

and

$$\tilde{V}_{\mathrm{berm},i}(T_i) = \begin{cases} \mathrm{E}^{\mathbb{Q}}(\tilde{V}_{\mathrm{berm},i+1}(T_{i+1}) | \mathcal{F}_{T_i}) & \text{if } \tilde{V}_{\mathrm{underl},i}(T_i) < \mathrm{E}^{\mathbb{Q}}(\tilde{U}_{i+1} | \mathcal{F}_{T_i}) \\ \tilde{V}_{\mathrm{underl},i}(T_i) & \text{else.} \end{cases}$$

This is a subtle but crucial difference. While both definitions give the Bermudan option value (through application of (2)), we have that the definition of $\tilde{U}_i$ requires the conditional expectation operator only to calculate the exercise criteria.

4 Conditional Expectation Estimators

We concentrate on the problem of calculating the conditional expectation of a Numéraire relative value process $\frac{V}{N}$.

4.1 Conditional Expectation as Functional Dependence

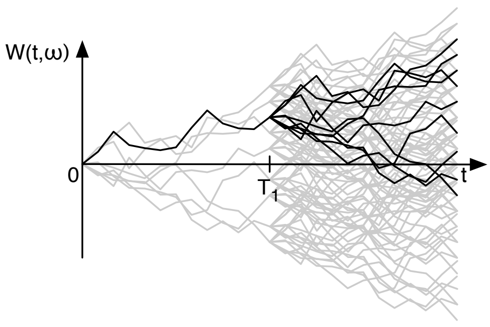


Figure 1: *Brute force calculation of the conditional expectation by pathwise resimulation – not feasible since the number of paths grows exponentially with the number of exercise dates.*

Let us reconsider the calculation of the conditional expectation through brute force re-simulation as depicted in Figure 1. On each path of the original simulation a re-simulation has to be created. These re-simulations differ in their initial conditions (e.g. the value $S(T_1)$ in a simulation of a stock price following a Black-Scholes Modell, or the values $L_i(T_1)$ in a simulation of forward rates following a LIBOR Market Model). These initial conditions are $\mathcal{F}_{T_1}$ measurable random varialbes (known as of T_1). Thus the conditional expectation is a function of these initial conditions (and possibly other model parameters known in T_1). If it is known that the conditional expectation is a function of a $\mathcal{F}_{T_1}$ measurable random variable Z (we assume here that $Z : \Omega \rightarrow \mathbb{R}^d$ with some d) we have

$$\mathrm{E}^{\mathbb{Q}^N}\left(\frac{V(T_2)}{N(T_2)} \mid \mathcal{F}_{T_1}\right) = \mathrm{E}^{\mathbb{Q}^N}\left(\frac{V(T_2)}{N(T_2)} \mid Z\right). \tag{3}$$

4.2 Perfect Foresight

In a path simulation the approximation of $\mathrm{E}^{\mathbb{Q}}\left(\frac{V(T_2)}{N(T_2)} \mid Z\right)$ will be given by averaging over all paths for which Z attains the same value. However in general the situation will be such that there are no two or more paths for which Z attains the same value - apart from the construction of the unfeasible resimulation. In other word one would use the crude approximation

$$\mathrm{E}^{\mathbb{Q}}\left(\frac{V(T_2)}{N(T_2)} \mid Z\right)[\omega] \approx \frac{V(T_2,\omega)}{N(T_2,\omega)}$$

This approximation is called perfect foresight.

4.3 Binning

An improvement is given by a *binning*, where the averaging will be done over those paths for which Z lies in a neighborhood (*bin*). If the quantities are continuous we have:

$$\mathrm{E}^{\mathbb{Q}}\left(\frac{V(T_2)}{N(T_2)} \mid Z\right)[\omega] \approx \mathrm{E}^{\mathbb{Q}}\left(\frac{V(T_2)}{N(T_2)} \mid Z \in U_\epsilon(Z(\omega))\right),$$

where $U_\epsilon(Z(\omega)) := \{z \mid ||Z(\omega) - z|| < \epsilon\}$.

Instead of defining a bin $U_\epsilon(Z(\omega))$ for each path ω it is more efficient to start with a partition of $Z(\Omega)$ into a finite set of disjoint bins $U_i \subset Z(\Omega)$. The approximation of the conditional expectation

$$\mathrm{E}^{\mathbb{Q}}\left(\frac{V(T_2)}{N(T_2)} \mid Z(\omega)\right)$$

will then be given by

$$H_i := \mathrm{E}^{\mathbb{Q}}\left(\frac{V(T_2)}{N(T_2)} \mid Z \in U_i\right)$$

where U_i denote the set with $Z(\omega) \in U_i$.

Example: Pricing of a simple Bermudan Option on a Stock

We illustrate the method in a simple Black-Scholes model for a stock S. In T_1 we wish to evaluate the option of receiving $N_1 \cdot (S(T_1) - K_1)$ in T_1 or to receive $N_2 \cdot \max(S(T_2) - K_2, 0)$ at later time T_2 (where N_1, N_2 (notional), K_1, K_2 (strike) are given). The optimal exercise in T_1 compares the exercise value with the value of the T_2 option, i.e.

$$\mathrm{E}^{\mathbb{Q}}\left(\frac{N_2 \cdot \max(S(T_2) - K_2, 0)}{N(T_2)} \mid \mathcal{F}_{T_1}\right).$$

From the model specification, e.g. here a Black-Scholes model

$$\mathrm{d}S(t) = r \cdot S(t)\mathrm{d}t \;+\; \sigma S(t)\mathrm{d}W^{\mathbb{Q}}(t), \qquad N(t) = \exp(rt)$$

it is obvious that the price of the T_2 option seen in T_1 is a given function $S(T_1)$ and the given model parameters (r, σ). Thus it is sufficient to calculate

$$\mathrm{E}^{\mathbb{Q}}\left(\frac{N_2 \cdot \max(S(T_2) - K_2, 0)}{N(T_2)} \mid S(T_1)\right).$$

In this example the functional dependence is known analytically. It is given by the Black-Scholes formula. Nevertheless we use the binning to calculate an approximation to the conditional expectation. If we plot

$$\underbrace{\frac{N_2 \cdot \max(S(T_2, \omega_i) - K_2, 0)}{N(T_2)}}_{\text{Continuation Value}} \quad \text{as a function of} \quad \underbrace{S(T_1, \omega_i)}_{\text{Underlying}}$$

we obtain the scatter plot in Figure 2. For a given $S(T_1)$ none or very few values of the *continuation values* exists. An estimate is not possible or exhibits a foresight bias. For an interval $[S_1 - \epsilon, S_1 + \epsilon]$ with sufficiently large ϵ we have enough values to calculate an estimate of

$$\mathrm{E}^{\mathbb{Q}}\left(\frac{N_2 \cdot \max(S(T_2) - K_2, 0)}{N(T_2)} \mid S(T_1) \in [S_1 - \epsilon, S_1 + \epsilon]\right)$$

which in turn may be used as estimate of

$$\mathrm{E}^{\mathbb{Q}}\left(\frac{N_2 \cdot \max(S(T_2) - K_2, 0)}{N(T_2)} \mid S(T_1) = S_1\right).$$

4.4 Regression Methods - Least Square Approximation of the Conditional Expectation

Let us start with a fairly general definition of the *least square approximation* of the conditional expectation of random variable U.

Definition 2 (Least Square Approximation of the Conditional Expectation): ⌝
Let $(\Omega, \mathcal{F}, \mathbb{Q}, \{\mathcal{F}_t\})$ be a filtered probability space and V a $\mathcal{F}_{T_1}$ measurable random variable defined as the conditional expectation of U

$$V = \mathrm{E}^{\mathbb{Q}}(U \mid \mathcal{F}_{T_1}),$$

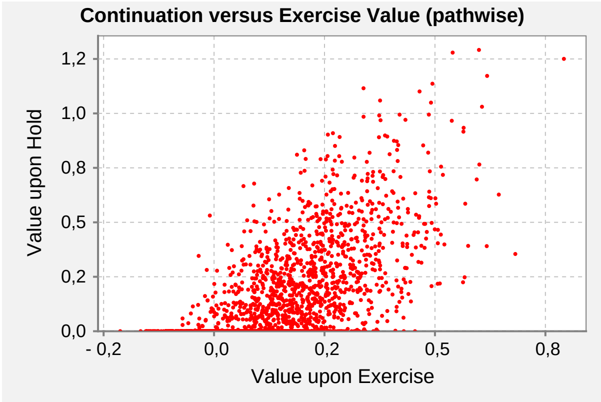


Figure 2: *The value realized on a path upon non exercise as a function of the value realized upon exercise.*

where U is at least $\mathcal{F}$ measurable. Furthermore let $Y := (Y_1, \ldots, Y_p)$ be a given $\mathcal{F}_{T_1}$ measurable random variable and $f : \mathbb{R}^p \times \mathbb{R}^q$ a given function. Let $\Omega^* = \{\omega_1, \ldots \omega_N\}$ a drawing from Ω (e.g. a Monte-Carlo simulation corresponding to $\mathbb{Q}$) and $\alpha^* := (\alpha_1, \ldots, \alpha_q)$ such that

$$||U - f(Y, \alpha^*)||_{L_2(\Omega^*)} = \min_{\alpha} ||U - f(Y, \alpha)||_{L_2(\Omega^*)}$$

where $||U - f(Y, \alpha^*)||^2_{L_2(\Omega^*)} = \sum_{j=1}^{N} (U(\omega_j) - f(Y(\omega_j), \alpha^*))^2$. We set

$$V^{\mathrm{LS}} := f(Y, \alpha^*).$$

The random variable V^{LS} is $\mathcal{F}_{T_1}$ measurable. It is defined over Ω and a *least square* approximation of V on Ω^*. ⌟

The Method of Longstaff and Schwartz[5] uses a function f with $q = p$ and

$$f(y_1, \ldots, y_p, \alpha_1, \ldots, \alpha_p) := \sum_{i=1}^{p} \alpha_i \cdot y_i,$$

such that α^* may be calculated analytically as a linear regression.

Lemma 3 (Linear Regression): Let $\Omega^* = \{\omega_1, \ldots, \omega_n\}$ be a given sample space, $V : \Omega^* \to \mathbb{R}$ and $Y := (Y_1, \ldots, Y_p) : \Omega^* \to \mathbb{R}^p$ given random variables. Furthermore let

$$f(y_1, \ldots, y_p, \alpha_1, \ldots, \alpha_p) := \sum \alpha_i y_i.$$

Then we have for any α^* with $X^{\mathsf{T}} X \alpha^* = X^{\mathsf{T}} v$

$$||V - f(Y, \alpha^*)||_{L_2(\Omega^*)} = \min_{\alpha} ||V - f(Y, \alpha)||_{L_2(\Omega^*)},$$

where

$$X := \begin{pmatrix} Y_1(\omega_1) & \ldots & Y_p(\omega_1) \\ \vdots & & \vdots \\ Y_1(\omega_n) & \ldots & Y_p(\omega_n) \end{pmatrix}, \qquad v := \begin{pmatrix} V(\omega_1) \\ \vdots \\ V(\omega_n) \end{pmatrix}.$$

If $(X^{\mathsf{T}} X)^{-1}$ exists then $\alpha^* := (X^{\mathsf{T}} X)^{-1} X^{\mathsf{T}} v$.

[5] See [8].

Definition 4 (Basis Functions):
The random variables $Y_1, \ldots, Y_p$ of Lemma 3 are called *Basis Functions* (*explanatory variables*).

4.4.1 Example: Evaluation of an Bermudan Option on a Stock (Backward Algorithm with Conditional Expectation Estimator)

We consider a Bermudan option on a Stock. The Bermudan should allow exercise at times $T_1 < T_2 < \ldots T_n$. Upon exercise in T_i the holder of the option will receive

$$N_i \cdot (S(T_i) - K_i)$$

once. If no exercise is made he will receive nothing.

We will apply the *backward algorithm* to derive the optimal exercise strategy. All payments will be considered in their Numéraire relative form. Thus the exercise criteria given by a comparison of the conditional expectation of the payments received upon non-exercise with the payments recieved upon exercise.

Induction start: $t > T_n$ Beyond the last exercise we have:

- The value of the (future) payments is $\tilde{U}_{n+1} = 0$.

Induction step: $t = T_i$, $i = n, n-1, n-2, \ldots 1$ In T_i we have:

- In the case of exercise is in T_i the value is

$$\tilde{V}_{\mathrm{underl,i}}(T_i) := \frac{N_i(S(T_i) - K_i)}{N(T_i)}. \tag{4}$$

- In the case of non-exercise in T_i the value is $\tilde{V}_{\mathrm{hold,i}}(T_i) = \mathrm{E}^{\mathbb{Q}}(\tilde{U}_{i+1} \mid \mathcal{F}_{T_i})$. This value is estimated through a regression for given paths $\omega_1, \ldots, \omega_m$:
  - Let Y_j be given ($\mathcal{F}_{T_i}$ measurable) basis functions.[6] Let the matrix X consist of the column vectors $Y_j(\omega_k)$, $k = 1, \ldots, m$. Then we have

$$\begin{pmatrix} \tilde{V}_{\mathrm{hold,i}}(T_i, \omega_1) \\ \vdots \\ \tilde{V}_{\mathrm{hold,i}}(T_i, \omega_m) \end{pmatrix} \approx X \cdot (X^{\mathsf{T}} \cdot X)^{-1} \cdot X^{\mathsf{T}} \cdot \begin{pmatrix} \tilde{U}_{i+1}(\omega_1) \\ \vdots \\ \tilde{U}_{i+1}(\omega_m) \end{pmatrix}. \tag{5}$$

- The value of the payments of the product in T_i under optimal exercise is given by

$$\tilde{U}_i := \begin{cases} \tilde{V}_{\mathrm{underl,i}}(T_i) & \text{if } \tilde{V}_{\mathrm{hold,i}}(T_i) < \tilde{V}_{\mathrm{underl,i}}(T_i) \\ \tilde{U}_{i+1} & \text{else.} \end{cases}$$

Remark 5 (Backward Algorithm): Our example is of course just the backward algorithm with an explicit specification of an underlying (4) and an explicit specification of an exercise criteria, here given by the estimator of the conditional expectation (5).

Remark 6 (Binning as Linear Regression): In [6] it is shown that binning may be understood as a special case of the linear regression: Binning is a linear regression where the basis functions are the indicator functions of the bins. See [6] for the (simple) proof.

[6] Suitable basis functions for this example are 1 (constant), $S(T_i)$, $S(T_i)^2$, $S(T_i)^3$, etc., such that the regression function f will be a polynomial in $S(T_i)$.

5 Foresight Bias: Classification, Calculation & Removal

The foresight bias is an option on the Monte Carlo error of the conditional expectation estimator. The standard deviation of the Monte Carlo error is the volatility of that option and the foresight bias is always non-negative.

Consider the optimal exercise value $\max(K, \mathrm{E}(\tilde{V} \mid Z))$ where the conditional expectation estimator has a Monte Carlo error which we denote by ϵ. Then the foresight bias is given by:

$$\mathrm{E}\big(\max(K, \mathrm{E}(\tilde{V} \mid Z) + \epsilon) \bigm| Z \big) \ = \ \max(K, \mathrm{E}(\tilde{V} \mid Z)) + \text{foresightbias}.$$

Remark 7 (Notation): Here and in the following we will consider the exercise criteria $\max(K, \mathrm{E}(\tilde{V} \mid Z))$, i.e. with the notation used in the previous section $\tilde{V}$ stands for $\tilde{U}_{i+1}$ and K stands for $\tilde{V}_{\mathrm{underl,i}}(T_i)$ for some i. The conditional expectation estimator (e.g. binning, regression) will be denoted by $\mathrm{E}^{\mathrm{est}}$ in place of E, i.e.

$$\mathrm{E}^{\mathrm{est}}(\tilde{V} \mid Z) = \mathrm{E}(\tilde{V} \mid Z) + \epsilon.$$

5.1 Numerical Removal of the Foresight Bias

The standard approach to remove the foresight bias is to use two independent Monte-Carlo simulations. One will be used to estimate the exercise criteria (as a functional dependence on some state variable), the other will be used to calculate the payouts.

5.2 Motivation for an Analytical Estimate and Removal of Foresight Bias

The numerical removal of the foresight bias has two disadvantages:

- Numerical removal of the foresight bias slows down the pricing. Two independent Monte-Carlo simulations of the stochastic processes have to be generated. For some models (e.g. high dimensional interest rate models like the LIBOR Market Model) the generation of the Monte-Carlo paths is relatively time consuming.
- Numerical removal of the foresight bias makes the code of the implementation cumbersome. It is a desired design pattern to separate the stochastic process model and the generation of the Monte-Carlo paths from product pricing. The structure of the code will likely become less clear if a second independent simulation has to be created.

An alternative to the numerical removal of the foresight bias is to not remove the foresight bias at all. This approach may be justified by the fact that the foresight bias will tend to zero as the number of paths tends to infinity. In addition the foresight bias is rather small, usually it is within Monte-Carlo errors. We will give an estimate of the foresight bias in the Section 5.3.

However neglecting foresight bias may create larger relative errors when considering multiple exercise dates or a book of multiple options with foresight bias. The discussion of whether it is feasible to neglect the foresight bias will be given in Section 5.7.

5.3 Estimation of the Foresight Bias

We want to asses the foresight bias induced by a Monte-Carlo error ϵ of the conditional expectation estimator $\mathrm{E}(\tilde{V} \mid Z)$, i.e. we consider the optimal exercise criteria

$$\max(K, \mathrm{E}(\tilde{V} \mid Z) + \epsilon).$$

Conditioned on a given $Z = z^*$ we assume that ϵ has normal distribution with mean 0 and standard deviation σ for fixed $\mathrm{E}(\tilde{V} \mid Z)$, i.e. we assume independence of $\tilde{V}|Z$. Then we have the following result for the foresight bias:

Lemma 8 (Estimation of Foresight Bias): Given a conditional expectation estimator of $\mathrm{E}(\tilde{V}|Z)$ with (conditional) Monte-Carlo error ϵ having normal distribution with mean 0 and standard deviation σ will result in a bias of the conditional mean of $\max(K, \mathrm{E}(\tilde{V}|Z)+\epsilon)$ given by

$$\underbrace{\underbrace{\sigma \cdot \phi(-\frac{\mu - K}{\sigma})}_{\text{foresight bias}}}_{\text{biased high}} \quad + \quad \underbrace{\underbrace{(\mu - K) \cdot \big(1 - \Phi(-\frac{\mu - K}{\sigma})\big) + K}_{\text{smoothed payout}} - \underbrace{\max(K, \mathrm{E}(\tilde{V}|Z))}_{\text{true payout}}}_{\text{diffusive part, biased low}} \quad , \tag{6}$$

where $\mu := \mathrm{E}(\tilde{V}|Z)$, $\phi(x) := \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$ and $\Phi(x) := \int_{-\infty}^{x} \phi(\xi) \, \mathrm{d}\xi$.

Proof (of Lemma 8): Let ϵ have Normal distribution with mean 0 and standard deviation σ. For $a, b \in \mathbb{R}$ we have with $\mu^* := b - a$

$$\begin{aligned}
\mathrm{E}(\max(a, b+\epsilon)) &= \mathrm{E}(\max(0, b-a+\epsilon)) + a = \mathrm{E}(\max(0, \mu^*+\epsilon)) + a \\
&= \frac{1}{\sigma}\int_0^\infty x \cdot \phi(\frac{x-\mu^*}{\sigma}) \, \mathrm{d}x + a = \frac{1}{\sigma}\int_{-\mu^*}^\infty (x+\mu^*) \cdot \phi(\frac{x}{\sigma}) \, \mathrm{d}x + a \\
&= \int_{-\frac{\mu^*}{\sigma}}^\infty (\sigma \cdot x + \mu^*) \cdot \phi(x) \, \mathrm{d}x + a \\
&= \sigma \cdot \phi(\frac{\mu^*}{\sigma}) + \mu^* \cdot \big(1 - \Phi(-\frac{\mu^*}{\sigma})\big) + a,
\end{aligned}$$

where we used $\int x\phi(x) \, \mathrm{d}x = \phi(x)$.

The result follows with $b = \mathrm{E}(\tilde{V}|Z)$, $a := K$, i.e. $\mu^* = \mu - K$. □|

Remark 9 (Interpretation): The bias induced by the Monte-Carlo error of the conditional expectation estimator consists of two parts: The first part in (6) consists of the systematic one sided bias resulting from the non linearity of the $\max(a, b+x)$ function. The second part is a diffusion of the original payoff function. The Monte-Carlo error smears out the original payoff. The first part should be attributed to super-optimal exercise due to foresight, the second part to sub-optimal exercise due to Monte-Carlo uncertainty.

In Figure 3 we graph the two parts, namely the function $x \mapsto \sigma \cdot \phi(\frac{x}{\sigma})$ (foresight bias, red) and $x \mapsto x \cdot (1 - \Phi(-\frac{x}{\sigma}))$ (smoothed out payout, blue).

Since the payout with foresight bias (i.e. the sum of the red and the blue curve in Figure 3) is always greater than the payout without foresight bias (the green curve in Figure 3), we have that the first part in (6) is always a dominant part. Usually (as in Figure 3) it is much larger than the negative bias from the diffusive part. See also Figure 4.

The smeared out payout (blue curve) lies below the true payout (green curve) since after having removed the foresight bias part we are left with a suboptimal exercise strategy, where the suboptimality stems from the disturbance induced by the Monte-Carlo error. This effect is also visible when removing the foresight bias numerically: for a lower number of path the price will be more biased low.

We define the first term in (6) as the foresight bias correction.

Definition 10 (Foresight Bias Correction): ⌝
With the notation as in Lemma 8 we define

$$\beta := \sigma \cdot \phi(-\frac{\mu - K}{\sigma})$$

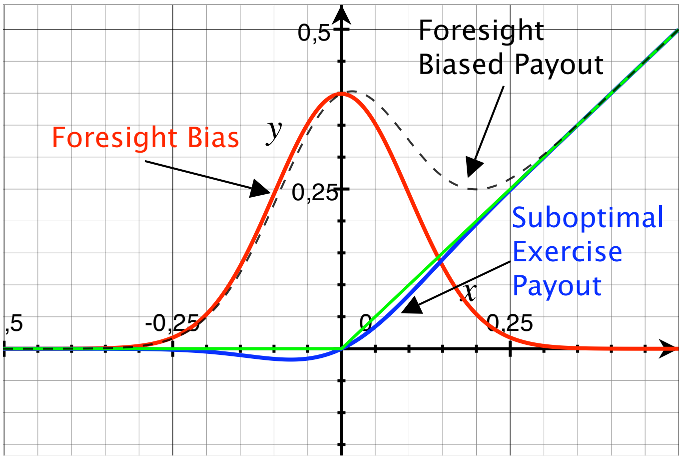


Figure 3: *Components of a foresight biased payout (1). The figure shows the plot of the foresight bias part (red) and the smeared out payout (blue) for $\sigma = 0.1$ and $K = 0$ as a function of μ (the distance from the exercise boundary).*

as the *foresight bias correction* of the optimal exercise criteria

$$\max(K, \mathrm{E}(\tilde{V}|Z)),$$

where $\mu := \mathrm{E}(\tilde{V}|Z)$ and σ^2 is the variance of the Monte-Carlo error ϵ of the estimator μ. ⌟

We also give a name to the second part in (6).

Definition 11 (Suboptimality Correction): ⌝
With the notation as in Lemma 8 we define

$$\gamma := (\mu - K) \cdot \big(1 - \Phi(-\frac{\mu - K}{\sigma})\big) - \max(0, \mu - K)$$

as the *suboptimal exercise correction* of the optimal exercise criteria

$$\max(K, \mathrm{E}(\tilde{V}|Z)),$$

where $\mu := \mathrm{E}(\tilde{V}|Z)$ and σ^2 is the variance of the Monte-Carlo error ϵ of the estimator μ. ⌟

5.4 Analytical Removal of Foresight Bias

We correct for the foresight bias induced for the optimal exercise

$$\max(K, \mathrm{E}(\tilde{V} \mid Z))$$

by subtracting the term

$$\beta^{\text{est}} := \sigma^{\text{est}} \cdot \phi(\frac{\mu^{\text{est}} - K}{\sigma^{\text{est}}}) \tag{7}$$

from the payout on each path, where $\mu^{\text{est}} := \mathrm{E}^{\text{est}}(\tilde{V} \mid Z)$ and σ^{est} is some estimator for the Monte-Carlo error ϵ (see below). With the notation in Section 3, we correct for the foresight bias by modifying the update rule of the backward algorithm towards

$$\tilde{U}_i := -\beta^{\text{est}} + \begin{cases} \tilde{V}_{\text{underl}}(T_i) & \text{if } \tilde{V}_{\text{underl}}(T_i) > \mathrm{E}^{\text{est}}(\tilde{V} \mid Z) \\ \tilde{U}_{i+1} & \text{else.} \end{cases}$$

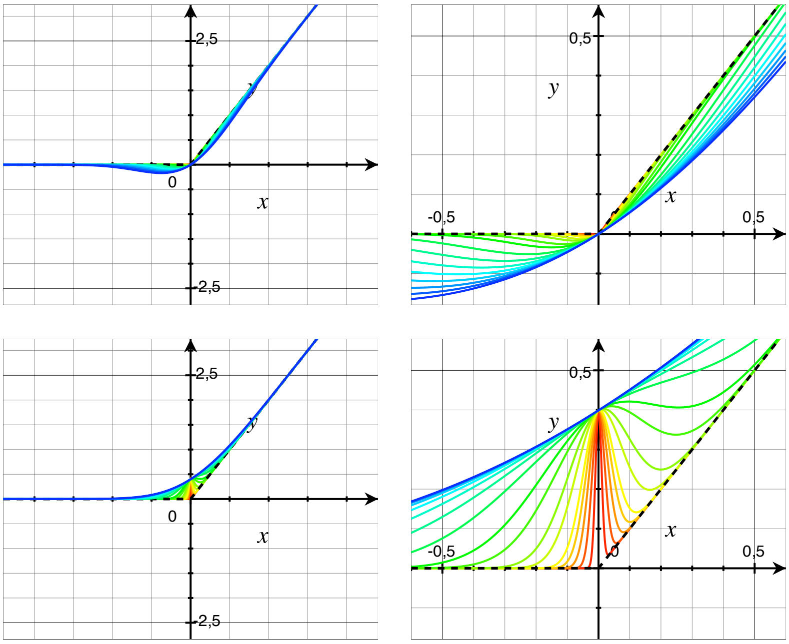


Figure 4: *Components of a foresight biased payout (2). The two graphs above show the function* $x \mapsto x \cdot (1 - \Phi(-\frac{x}{\sigma}))$. *Compared to the true payout it is biased low. The two graphs below show the function* $x \mapsto \sigma \cdot \phi(\frac{x}{\sigma}) + x \cdot (1 - \Phi(-\frac{x}{\sigma}))$. *Obviously the term* $\sigma \cdot \phi(\frac{x}{\sigma})$ *dominates and compared to the true payout the result is biased high. The figure shows the functions for from* $\sigma = 0.01$ *(red) to* $\sigma = 0.1$ *(green) to* $\sigma = 1.0$ *(blue).*

Note that β is stochastic since $\mu := \mathrm{E}(\tilde{V} \mid Z)$ and σ are stochastic. They are conditional means and conditional standard deviations. However we have for a given $Z = z^*$ that $\beta|\{Z = z^*\}$ is not stochastic. Conditioned on $\{Z = z^*\}$ the foresight bias is thus removed. Integrating over z^* we have that subtracting β removes the foresight bias globally.

5.5 Analytical Removal of Foresight Bias and Suboptimal Exercise

In addition to a correction for the foresight bias we may also correct for the additional term in (6) which represents the subobtimality induced by the Monte-Carlo noise. To do so we not only subtract β^{est} but also subtract

$$\gamma^{\mathrm{est}} := (\mu^{\mathrm{est}} - K) \cdot \big(1 - \Phi(-\frac{\mu^{\mathrm{est}} - K}{\sigma^{\mathrm{est}}})\big) - \max(0, \mu^{\mathrm{est}} - K) \tag{8}$$

from the payout on each path, where $\mu^{\mathrm{est}} := \mathrm{E}^{\mathrm{est}}(\tilde{V} \mid Z)$ and σ^{est} is some estimator for the Monte-Carlo error ϵ (see below). With the notation in Section 3, we correct for the foresight bias by modifying the update rule of the backward algorithm towards

$$\tilde{U}_i := -\beta^{\mathrm{est}} - \gamma^{\mathrm{est}} + \begin{cases} \tilde{V}_{\mathrm{underl}}(T_i) & \text{if } \tilde{V}_{\mathrm{underl}}(T_i) > \mathrm{E}^{\mathrm{est}}(\tilde{V} \mid Z) \\ \tilde{U}_{i+1} & \text{else.} \end{cases}$$

Since γ^{est} is negative (it is the amount lost due to suboptimal exercise due to Monte-Carlo noise) this correction will increase the price. However The absolute value of γ^{est} is much smaller than β^{est}. The effect of this correction is only visible when using a low number of path.

We will present numerical results of the full correction $\beta + \gamma$ in Section 6.2.

From now on we will drop the superscript $\cdot^{\text{est}}$ and μ, σ and β denote the corresponding estimates thereof.

5.6 Implementation of the Analytical Removal of Foresight Bias

The foresight correction (7) is independent from the method used to estimate conditional expectation, e.g. binning or polynomial regression (aka. Longstaff & Schwartz, [3, 4, 8]). However some care has to be taken when calculating the estimation of σ used in foresight correction.

5.6.1 Calculation of the Monte-Carlo error

To calculate the foresight correction β^{est} we need to estimate the variance of the Monte-Carlo error ϵ. In other words we are interested in the standard error of the estimated conditional expectation $\mathrm{E}^{\text{est}}(\tilde{V} \mid Z)$.

Binning: Let us consider first the simple case of a binning. Then we have

$$\sigma^2(\omega) \approx \frac{1}{n_\omega} \cdot \mathrm{E}((\tilde{V} - \mu)^2 \mid Z = Z(\omega))$$

where n_ω denotes the number of paths in the bin U with $Z(\omega) \in U$. And the conditional variance $\mathrm{E}((\tilde{V} - \mu)^2 \mid Z)$ is estimated by the *same* binning as

$$\mathrm{E}((\tilde{V} - \mu)^2 \mid Z = Z(\omega)) = \frac{1}{n_\omega} \sum_{\tilde{\omega} \in U(\omega)} (\tilde{V}(\tilde{\omega}) - \mu)^2.$$

Regression: For the general case of a regression the standard error is calculated as

$$\sigma^2 \approx \mathrm{tr}(X \cdot (X^{\mathsf{T}} \cdot X)^{-1} \cdot X^{\mathsf{T}}) \cdot \mathrm{E}((\tilde{V} - \mu)^2 \mid Z),$$

where X is the matrix of basis functions used to estimate $\mathrm{E}(\tilde{V} \mid Z)$. For an in depth discussion and an proof of this result we refer the reader to the standard literature on multiple regression, e.g. [5].

The conditional variance $\mathrm{E}((\tilde{V} - \mu)^2 \mid Z)$ of the residuals may be estimated by the same or a different regression or method. Note however that $\mathrm{E}((\tilde{V} - \mu)^2 \mid Z)$ may become negative when using a numerical approximation to the true conditional variance. This is frequently the case when a polynomial regression is used to estimate the conditional expectation of $(\tilde{V} - \mu)^2$. In our implementation we have fixed this by just taking

$$\max(\ \mathrm{E}((\tilde{V} - \mu)^2 \mid Z)\ ,\ 0\ )$$

instead. For small number of path the adjusted estimate

$$\sigma^2 \ \approx\ \frac{n}{n-q} \mathrm{E}((\tilde{V} - \mu)^2 \mid Z)$$

should be used, where n is the number of path and q the number of basis functions.

Note also that for $\sigma \rightarrow 0$ we have $\beta \rightarrow 0$, which should be used in case of a division by zero exception.

5.7 Is the foresight bias negligible?

Let us consider the analytical removal of β since this will correspond to the numerical removal of the foresight bias and the effect of the additional suboptimality correction γ is only visible for large Monte-Carlo errors, i.e. small number of paths.

Our formula suggests that the foresight bias β is of the order of the Monte-Carlo error. Thus one is tempted to conclude that it is safe to neglect the foresight bias. Indeed, in our test case of a Bermudan option with three exercises (see Section 6.3.1) we found that the foresight bias is within one standard deviation of the Monte-Carlo price.

5.7.1 Aggregating Foresight Biased Options

However, summing up different options - each with a foresight bias and a Monte-Carlo error - may change the picture. If two options differ in strike or maturities their Monte-Carlo errors may become more an more independent.[7] Consider a book of n options (compound or Bermudan). If the n options have independent Monte-Carlo errors with standard deviation σ the Monte-Carlo error for the portfolio will be $\sqrt{n} \cdot \sigma$. But since the foresight bias is a systematic error it will grow linearly in n, i.e. if the n options have a foresight bias β the book will exhibit a foresight bias of $n \cdot \beta$. Assuming that for a family of options β and σ are of the same size we could say that: only if the Monte-Carlo errors of the single product prices are perfectly correlated we would have that the ratio of foresight bias to Monte-Carlo error $\frac{\beta}{\sigma}$ of a portfolio does not grow with the portfolio size.

This is also obvious from the interpretation of the foresight bias as an option on the (individual) Monte-Carlo error. The book will contain n such options.[8] In the end we have that the foresight bias may likely become significant.[9]

For a Bermudan option with many exercise dates the situation is even worse. Here we have a Bermudan option on the foresight biases induced in each maturity. In Section 6.3.2 we calculated the foresight bias for an option with nine exercise dates and compare it to three options each with three of the nine exercise dates. While the option with three exercises showed foresight biases of 0.4 to 0.5 times the Monte-Carlo error the option with all nine exercises showed a foresight bias of 1.3 times the Monte-Carlo error. The calculation in Section 6.3.2 suggest that for a Bermudan option the Monte-Carlo errors average, but the foresight biases add.

The two effects become visible in our numerical experiments, see Table 5.

5.7.2 Parallelization of Pricing with Foresight Bias

Not removing the foresight bias has a severe influence on the parallelization of pricing. For European options Monte-Carlo has the pleasing property that averaging n results from n independent Monte-Carlo simulations reduces the pricing error by a factor of $\frac{1}{\sqrt{n}}$. This is also true for the Monte-Carlo error of a Bermudan, but not for its foresight bias. Aggregating Monte-Carlo prices will increase the size of the foresight bias relative to the Monte-Carlo error thus making the foresight bias a dominant effect.

[7] As example consider the two payouts $\min(\max(S(T), a_1), b_1)$ and $\min(\max(S(T), a_2), b_2)$ (i.e. $S(T)$ capped and floored). If (a_1, b_1) and (a_2, b_2) are disjoint a sampling of $S(T)$ will (in general) generate independent Monte-Carlo errors for the two payouts.

[8] Of course foresight bias may cancel if one averages short options with long options.

[9] Our test case in Section 6.3.1 exhibited a foresight bias 0.5 of the Monte-Carlo error. Pricing a book of 16 options may result in a foresight bias around 2 standard deviations (the 95% quantile).

6 Numerical Results

We will present numerical results for different experiments:

- We show how the foresight bias becomes significant if independent Monte-Carlo simulations are aggregated and how the numerical and analytical method correct for it.
- We show how the foresight bias relates to the Monte-Carlo error for increasing number of paths.
- We show that the ratio of foresight bias to Monte-Carlo error increases for options with more exercise dates or a portfolio of options.

6.1 Benchmark Model

Our Benchmark model is a simple Black-Scholes model for an asset S where S follows

$$\mathrm{d}S = \mu S \mathrm{d}t + \sigma S \mathrm{d}W$$

with $S(0) = 1.0$, $\sigma = 20\%$, and assuming the risk free asset $\mathrm{d}B = rB\mathrm{d}t$ with $r = 5\%$.

6.2 Aggregation of Monte-Carlo Prices

We setup m independent Monte-Carlo simulation with n/m paths and calculate the average price of a Bermudan option price over the set of Monte-Carlo simulations. We vary m from $m = 1$, i.e. a single Monte-Carlo simulation with a huge number of paths, to $m = 2048$, i.e. many small Monte-Carlo simulations.

The aggregated prices have similar Monte-Carlo errors and for an European option the different methods should result in (almost) identical prices. For a Bermudan option the different methods of aggregation are not equivalent. The foresight bias is a systematic error being $O(\sqrt{m/n})$.

Figure 5 and 6 show that the numerical removal of the foresight bias and our analytical removal of the foresight bias give very similar results. The Figure shows the price resulting from the aggregation of m Monte-Carlo simulations, each such that the total number of paths is independent of m. For m large a single Monte-Carlo simulation has a low number of paths, thus a larger foresight-bias. This parallelization has a huge impact on the price accuracy if foresight bias is not removed. If foresight bias is removed the price will slowly become lower. This is due to the fact that the optimal exercise is smeared out by the diffusive term in (6).

Our benchmark product is a simple Bermudan option on S where the holder has the right to receive once

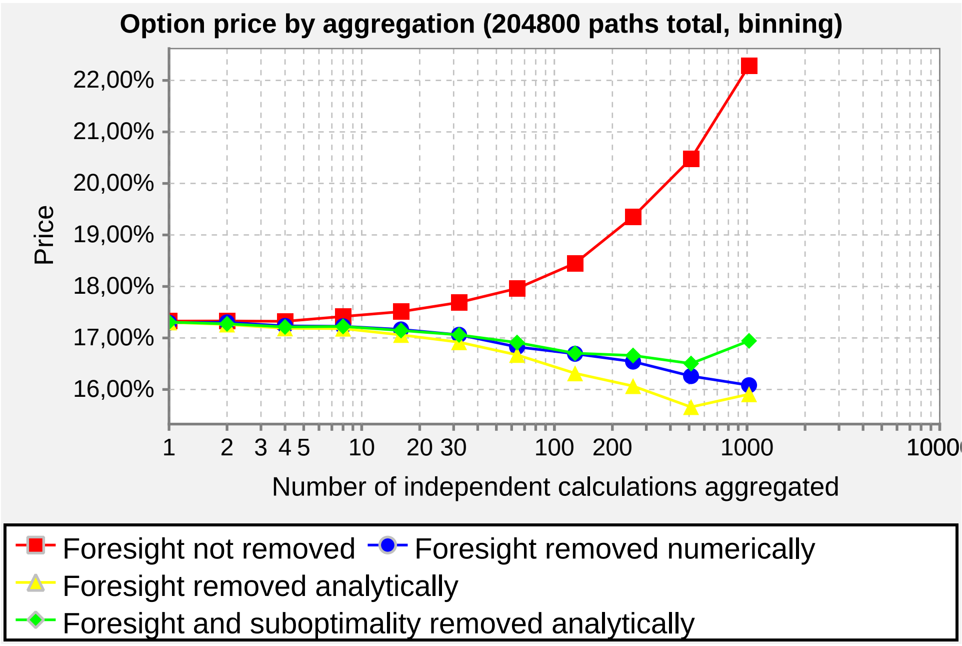
$$N_i \cdot (S(T_i) - K_i) \quad \text{in } T_i$$

or nothing if none of the options is exercised. The calculation shown are for three exercise dates with

Bermudan option with three exercises

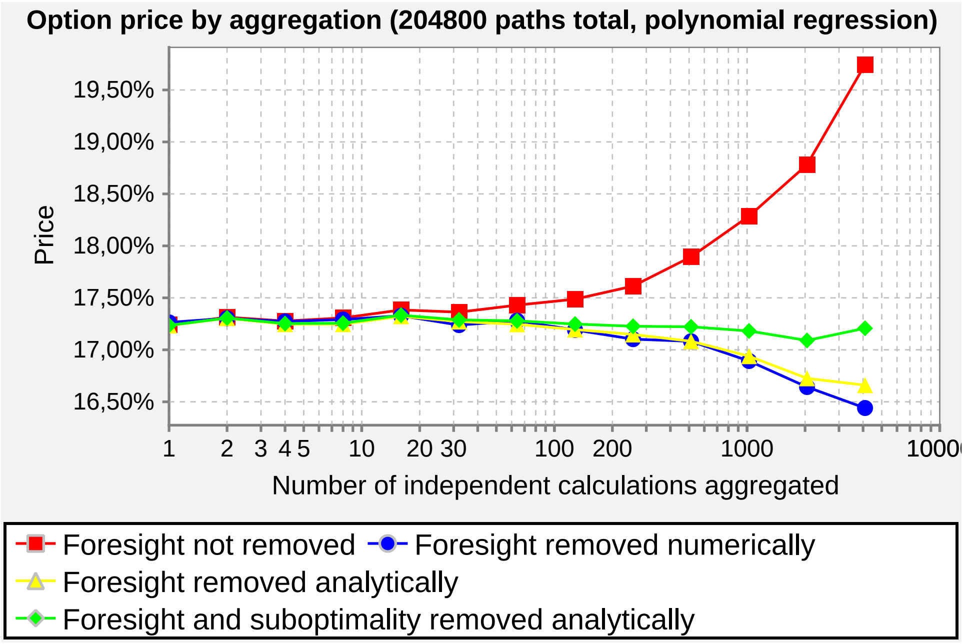
Exercise date T_i	1.00	2.00	3.00
Notional N_i	1.00	1.00	1.00
Strike K_i	0.95	1.00	1.10

Table 1: *Specification of the product used in Figure 5 and 6.*



Number of Independent Simulations	Number of Paths per Simulation	Aggregated Price: Without Foresight Removal	Numerical Foresight Removal	Analytical Foresight only	Analytical Foresight and Suboptimality
1	204800	17,328%	17,295%	17,302%	17,308%
2	102400	17,330%	17,303%	17,262%	17,274%
4	51200	17,323%	17,234%	17,187%	17,214%
8	25600	17,417%	17,226%	17,182%	17,224%
16	12800	17,510%	17,166%	17,059%	17,145%
32	6400	17,689%	17,061%	16,918%	17,060%
64	3200	17,961%	16,826%	16,671%	16,911%
128	1600	18,450%	16,692%	16,314%	16,704%
256	800	19,353%	16,541%	16,066%	16,660%
512	400	20,481%	16,259%	15,659%	16,503%
1024	200	22,287%	16,083%	15,908%	16,944%

Figure 5: *Aggregation of Monte-Carlo Prices with or without Removal of Foresight. The conditional expectation is estimated by a binning with 100 bins. Note that since we use 100 bins the last scenario (using 200 paths per individual Monte Carlo simulations) calculated the conditional expectation using two paths. Thus we almost have a perfect foresight. Correcting the foresight by estimating the Monte Carlo error from the two paths gives very good results in average. Note that a numerical removal of the foresight bias may still result a (small) bias high since it is not guaranteed that the two simulation used are independent, especially when aggregating many subsequent runs.*
For small number of paths the additional correction for the suboptimal exercise (green curve) gives even better results.



Number of Independent Simulations	Number of Paths per Simulation	Aggregated Price: Without Foresight Removal	Numerical Foresight Removal	Analytical Foresight only	Analytical Foresight and Suboptimality
1	204800	17,240%	17,264%	17,234%	17,235%
2	102400	17,314%	17,303%	17,304%	17,306%
4	51200	17,277%	17,273%	17,247%	17,253%
8	25600	17,307%	17,291%	17,247%	17,256%
16	12800	17,384%	17,328%	17,323%	17,333%
32	6400	17,363%	17,238%	17,272%	17,288%
64	3200	17,430%	17,280%	17,246%	17,279%
128	1600	17,487%	17,191%	17,195%	17,248%
256	800	17,613%	17,103%	17,148%	17,227%
512	400	17,895%	17,082%	17,083%	17,222%
1024	200	18,286%	16,893%	16,939%	17,181%
2048	100	18,781%	16,642%	16,726%	17,090%
4096	50	19,744%	16,440%	16,660%	17,207%

Figure 6: *Aggregation of Monte-Carlo Prices with or without Removal of Foresight. The conditional expectation is estimated by a regression using a polynomial of order 5.*
For small number of paths the additional correction for the suboptimal exercise (green curve) gives even better results.

6.3 Relation of Foresight Bias to Monte-Carlo Error

We calculate Monte-Carlo prices for an Bermudan option with three exercise dates. Foresight bias in either not removed, removed numerically or removed analytically. The pricing is performed with different number of paths. The pricing is repeated 15000 times with different sets of random numbers (seeds). We graph the distribution of Monte-Carlo prices and access the Monte-Carlo error. In the following examples we consider only the foresight bias correction β.

6.3.1 Example 1: Bermudan Option with three Exercise Dates

We consider the Bermudan option with three exercises as given in table 2.

Product B1: Option with three exercises

Exercise date T_i	1.00	2.00	3.00
Notional N_i	1.00	1.00	1.00
Strike K_i	1.00	1.06	1.12

Table 2: *Product B1: Option with nine exercises.*

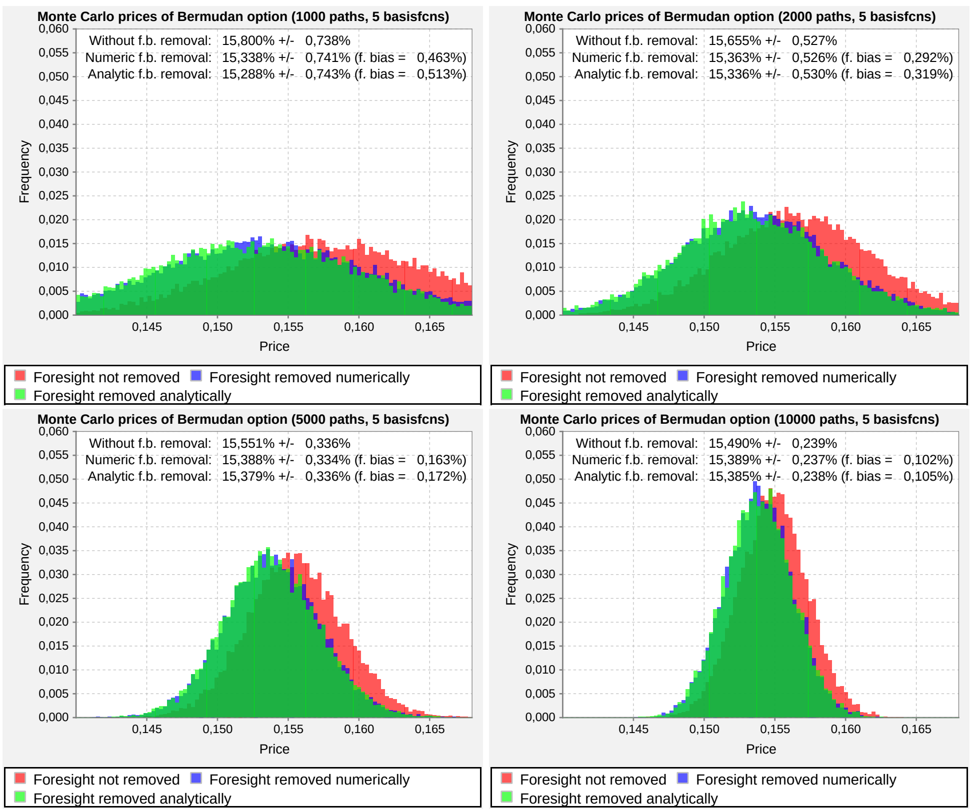


Figure 7: *Distribution of Monte-Carlo prices without (red) and with (blue and green) removal of foresight bias for an Bermudan option with three exercise dates. The blue distribution is almost hidden behind the green distribution since the analytical and numerical removal give very similar results.*

6.3.2 Example 2: Bermudan Option with nine Exercise Dates

We calculate Monte-Carlo prices for an Bermudan option with nine exercise dates. Foresight bias in either not removed, numerically removed or analytically removed. The pricing is performed with different number of paths. The pricing is repeated 20000 times with different sets of random numbers (seeds) to access the Monte-Carlo error. The results are shown in Figure 8 and 10.

We consider the Bermudan option with nine exercises as given in table 3.

Product B4: Option with nine exercises

Exercise date T_i	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00
Notional N_i	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Strike K_i	1.00	1.06	1.12	1.19	1.26	1.34	1.42	1.50	1.59

Table 3: *Product B4: Option with nine exercises.*

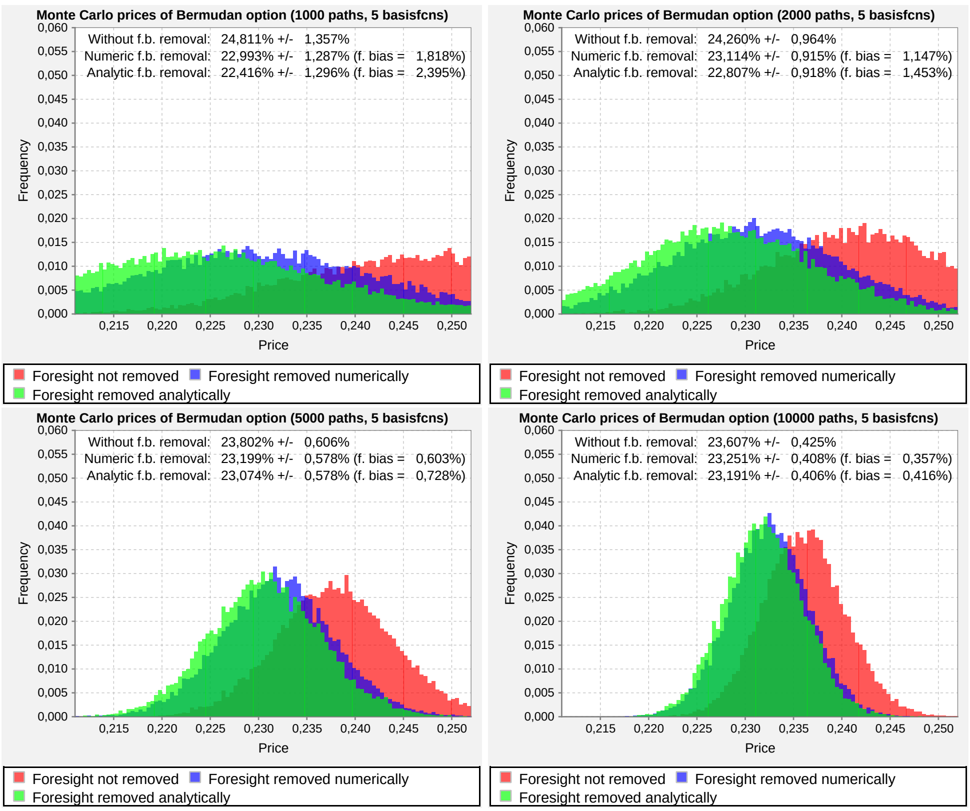


Figure 8: *Distribution of Monte-Carlo prices without (red) and with (blue and green) removal of foresight bias for an Bermudan option with nine exercise dates.*

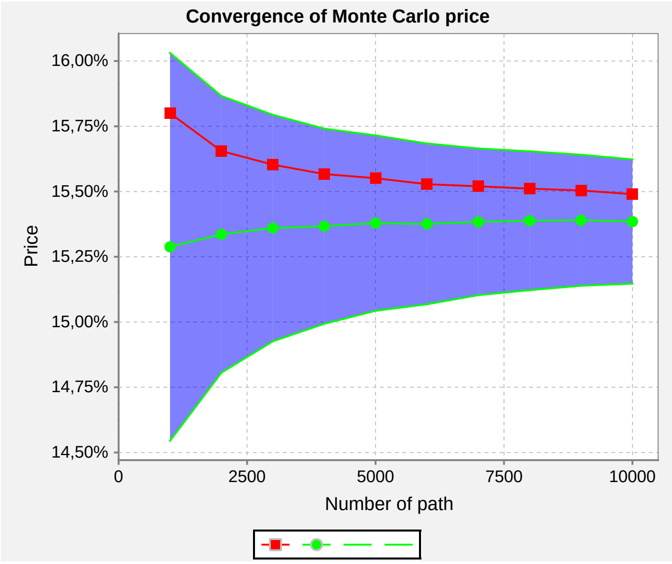


Figure 9: *Product B1 (Bermudan with nine exercise dates): Convergence of Monte-Carlo prices without foresight removal (red) and with analytic foresight removal (green) and Monte-Carlo error (blue corridor = one standard deviation).*

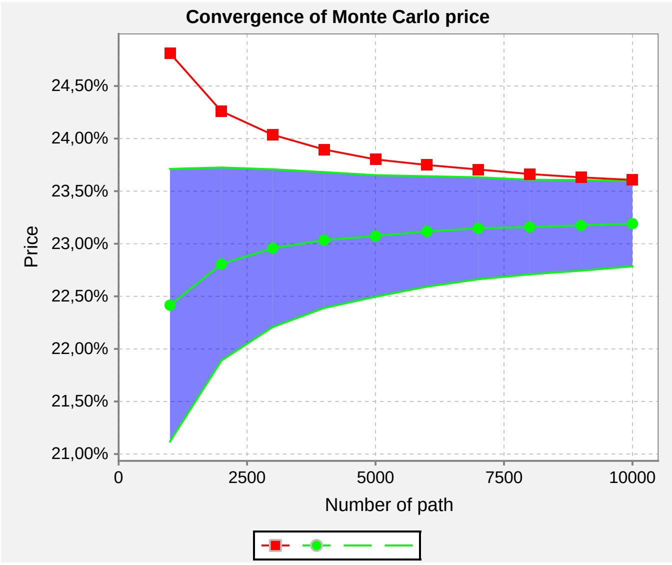


Figure 10: *Product B4 (Bermudan with nine exercise dates): Convergence of Monte-Carlo prices without foresight removal (red) and with analytic foresight removal (green) and Monte-Carlo error (blue corridor = one standard deviation).*

6.3.3 Example 3: A Portfolio of Bermudan Options

In addition to the products B1 and B4 we consider the two Bermudan option with three exercises as given in Table 4. Note that the product B4 is just a Bermudan on B1, B2, B3.

Product B2: Option with three exercises

Exercise date T_i	4.00	5.00	6.00
Notional N_i	1.00	1.00	1.00
Strike K_i	1.19	1.26	1.34

Product B3: Option with three exercises

Exercise date T_i	7.00	8.00	9.00
Notional N_i	1.00	1.00	1.00
Strike K_i	1.42	1.50	1.59

Table 4: *Product B2 and B3: Option with nine exercises.*

We calculate foresight bias (β) and Monte Carlo error (σ) for the prices of the Bermudan options B1, B2, B3 and B4 as well as for the two portfolios $\{\frac{1}{3}\cdot\mathrm{B1}, \frac{1}{3}\cdot\mathrm{B2}, \frac{1}{3}\cdot\mathrm{B3}\}$ $\{\frac{1}{2}\cdot\mathrm{B2}, \frac{1}{2}\cdot\mathrm{B3}\}$. Since the Bermudan option B4 is just the union of the underlyings of B1, B2 and B3 we denote it by $\{\mathrm{B1} \vee \mathrm{B2} \vee \mathrm{B3}\}$.

The results are given in Table 5. Each pricing used 5000 paths an was repeated 20000 times with different sets of random numbers.

Option / Portfolio	foresight bias β	Monte Carlo error σ	Ratio $\frac{\beta}{\sigma}$
B1	0.172	0.336	0.51
B2	0.198	0.491	0.40
B3	0.271	0.656	0.41
$\mathrm{B4} = \{\mathrm{B1} \vee \mathrm{B2} \vee \mathrm{B3}\}$	0.728	0.578	1.26
$\{\frac{1}{2}\cdot\mathrm{B2} + \frac{1}{2}\cdot\mathrm{B3}\}$	0.234	0.518	0.45
$\{\frac{1}{3}\cdot\mathrm{B1} + \frac{1}{3}\cdot\mathrm{B2} + \frac{1}{3}\cdot\mathrm{B3}\}$	0.214	0.428	0.50
Theoretical values assuming perfectly uncorrelated Monte-Carlo errors			
$\{\frac{1}{2}\cdot\mathrm{B2} + \frac{1}{2}\cdot\mathrm{B3}\}$	0.235	0.410	0.57
$\{\frac{1}{3}\cdot\mathrm{B1} + \frac{1}{3}\cdot\mathrm{B2} + \frac{1}{3}\cdot\mathrm{B3}\}$	0.214	0.295	0.72
Theoretical values assuming perfectly correlated Monte-Carlo errors			
$\{\frac{1}{2}\cdot\mathrm{B2} + \frac{1}{2}\cdot\mathrm{B3}\}$	0.235	0.574	0.41
$\{\frac{1}{3}\cdot\mathrm{B1} + \frac{1}{3}\cdot\mathrm{B2} + \frac{1}{3}\cdot\mathrm{B3}\}$	0.214	0.496	0.43

Table 5: *Comparrison of foresight bias (β) and Monte Carlo error (σ) for single options and portfolios.*

We see in Table 5 the behavior which we have discussed qualitatively in Section 5.7: The foresight bias of a portfolio is just the sum of the foresight bias of the foresight biases of its components. Compared to the Monte-Carlo error of the portfolio the foresight bias ratio grows. For example we have

$$\beta_{\frac{1}{2}\cdot\mathrm{B2}+\frac{1}{2}\cdot\mathrm{B3}} = 0.234 \approx 0.235 = \frac{1}{2}\cdot\beta_{\mathrm{B2}} + \frac{1}{2}\cdot\beta_{\mathrm{B3}}$$
$$\sigma_{\frac{1}{2}\cdot\mathrm{B2}+\frac{1}{2}\cdot\mathrm{B3}} = 0.518 < 0.574 = \frac{1}{2}\cdot\sigma_{\mathrm{B2}} + \frac{1}{2}\cdot\sigma_{\mathrm{B3}}$$

and likewise

$$\beta_{\frac{1}{3}\cdot\mathrm{B1}+\frac{1}{3}\cdot\mathrm{B2}+\frac{1}{3}\cdot\mathrm{B3}} = 0.214 \approx 0.214 = \frac{1}{3}\cdot\beta_{\mathrm{B1}} + \frac{1}{3}\cdot\beta_{\mathrm{B2}} + \frac{1}{3}\cdot\beta_{\mathrm{B3}}$$
$$\sigma_{\frac{1}{3}\cdot\mathrm{B1}+\frac{1}{3}\cdot\mathrm{B2}+\frac{1}{3}\cdot\mathrm{B3}} = 0.428 < 0.494 = \frac{1}{3}\cdot\sigma_{\mathrm{B1}} + \frac{1}{3}\cdot\sigma_{\mathrm{B2}} + \frac{1}{3}\cdot\sigma_{\mathrm{B3}}.$$

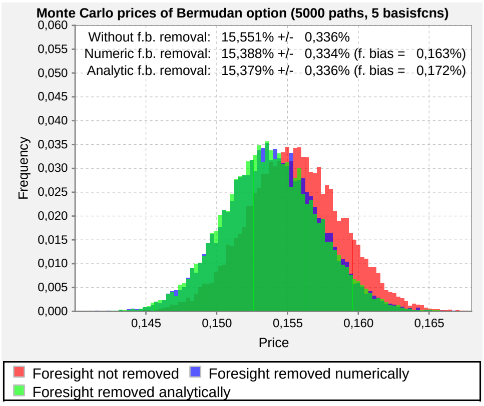
So for both cases we have that the ratio of foresight bias to Monte-Carlo error increases:

$$\frac{\beta_{\frac{1}{2}\cdot\mathrm{B2}+\frac{1}{2}\cdot\mathrm{B3}}}{\sigma_{\frac{1}{2}\cdot\mathrm{B2}+\frac{1}{2}\cdot\mathrm{B3}}} = 0.45 > 0.41 = \frac{\frac{1}{2}\cdot\beta_{\mathrm{B2}}+\frac{1}{2}\cdot\beta_{\mathrm{B3}}}{\frac{1}{2}\cdot\sigma_{\mathrm{B2}}+\frac{1}{2}\cdot\sigma_{\mathrm{B3}}}$$

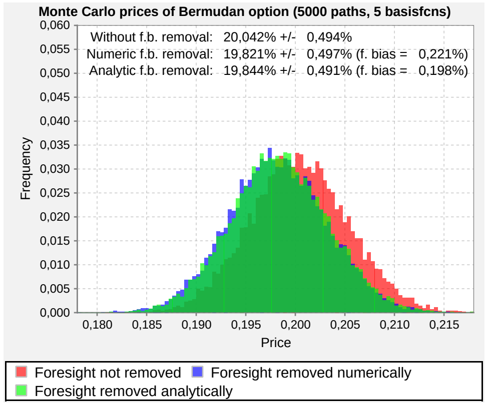
$$\frac{\beta_{\frac{1}{3}\cdot\mathrm{B1}+\frac{1}{3}\cdot\mathrm{B2}+\frac{1}{3}\cdot\mathrm{B3}}}{\sigma_{\frac{1}{3}\cdot\mathrm{B1}+\frac{1}{3}\cdot\mathrm{B2}+\frac{1}{3}\cdot\mathrm{B3}}} = 0.50 > 0.43 = \frac{\frac{1}{3}\cdot\beta_{\mathrm{B1}}+\frac{1}{3}\cdot\beta_{\mathrm{B2}}+\frac{1}{3}\cdot\beta_{\mathrm{B3}}}{\frac{1}{3}\cdot\sigma_{\mathrm{B1}}+\frac{1}{3}\cdot\sigma_{\mathrm{B2}}+\frac{1}{3}\cdot\sigma_{\mathrm{B3}}}.$$

Thus the relative foresight bias increases.

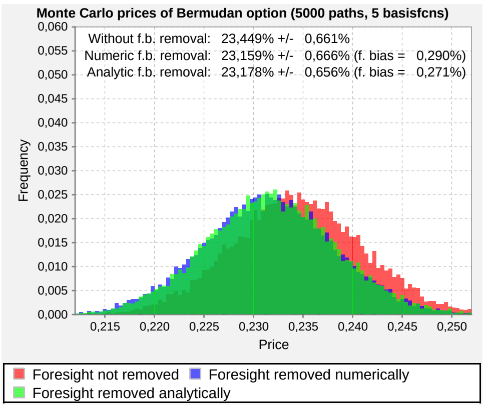
For the Bermudan option B4 $=\{\mathrm{B1}\vee\mathrm{B2}\vee\mathrm{B3}\}$ having nine exercise dates the ratio $\frac{\beta}{\sigma}$ increases to 1.26, so the ratio almost triples. Here we have that the foresight biases almost add, while the Monte-Carlo errors are just averaged. See Figure 11.



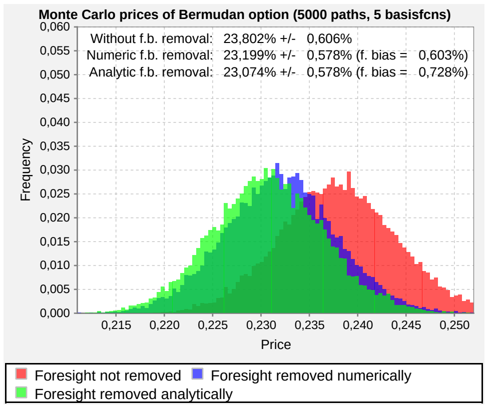
(a) Product B1: Thee exercises $\{1.0, 2.0, 3.0\}$



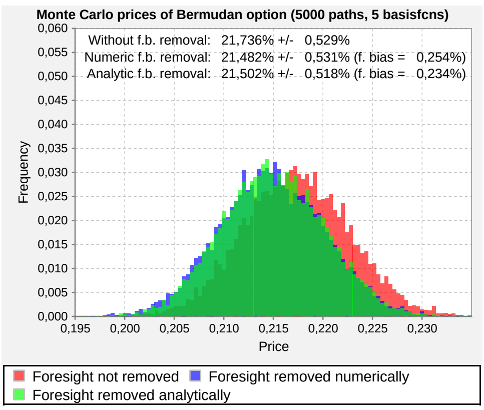
(b) Product B2: Thee exercises $\{4.0, 5.0, 6.0\}$



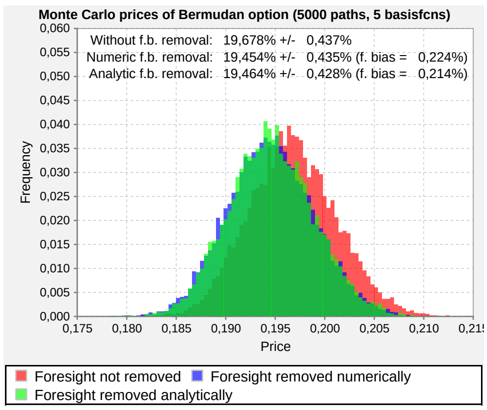
(c) Product B3: Thee exercises $\{7.0, 8.0, 9.0\}$



(d) Product B4: All nine exercises $\{\text{B1} \vee \text{B2} \vee \text{B3}\}$



(e) Protfolio $\frac{1}{2} \cdot \text{B2} + \frac{1}{2} \cdot \text{B3}$



(f) Protfolio $\frac{1}{3} \cdot \text{B1} + \frac{1}{3} \cdot \text{B2} + \frac{1}{3} \cdot \text{B3}$

Figure 11: *Distribution of Monte-Carlo prices without (red) and with (blue and green) removal of foresight bias for B1, B2, B3, the Bermudan B4 and the portfolios $\{B2, B3\}$ and $\{B1, B2, B3\}$. Note how the foresight bias is increased for the Bermudan in 11(d).*

7 Conclusion

The analytical foresight correction gives almost identical result to a numerical foresight correction even for small number of paths. Using the analytical foresight corrections allows to keep the implementation as simple and efficient as without foresight correction. The foresight correction β may also be used as a safe guard indicating whether or not you are allowed to neglect foresight bias.

Our approach is independent of the method used to estimate the conditional expectation and easy to implement. We provide a closed formula for the foresight bias β when a multiple linear regression is used to estimate the conditional expectations (aka. Longstaff-Schwartz).

Our numerical results indicate that foresight bias should not be neglected, at least for Bermudan option with many exercise dates or portfolios of Bermudan options.

The additional correction for the suboptimal exercise (γ) improves the pricing when aggregating Monte-Carlo simulations with small number of path.

Since our approach consisted of the analytic calculation of the error induced by the additional Monte-Carlo error variance it is universal in the sense that it is independet of the model and the of the product considered.

Future work should run tests with high dimensional models like the LIBOR Market Model. Here we expect the effects to be even greater due to higher decorrelation of Monte-Carlo error.

List of Symbols

Symbol	Meaning
$\tilde{V}$	The random variable representing the value received upon non-exercise (*continuation value*) and for which a conditional expectation has to be estimated (i.e. one of the $\tilde{U}_i$ defined in the backward algorithm).
Z	The random variable representing the value on which the expectation of $\tilde{V}$ is conditioned (the predictor). In the pricing of Bermudan stock options this might be the stock ($Z = S$) or some intrinsic value ($Z = \max(S - K, 0)$).
Y	The basis functions ($Y = (Y_1, \ldots, Y_q)$ used to approximate the functional dependance $z \mapsto \mathrm{E}(\tilde{V} \mid Z = z)$. For a polynomial regression the basis functions are often monomials in Z. In the pricing of Bermudan stock options this might be $Y = (1, Z, Z^2, \ldots, Z^{q-1})$.
X	The matrix of Monte Carlo samples of the regression basis functions Y used to estimate $\mathrm{E}(\tilde{V} \mid Z)$. Each row represent the values of the basis functions on the corresponding Monte Carlo path.
q	Number of basis functions.
n	Number of Monte Carlo paths.
ϵ	Monte Carlo error of the conditional expectation estimator of $\tilde{V}$.
σ	Standard deviation of the Monte Carlo error ϵ.
$\mu - K$	Distance of $\mathrm{E}(\tilde{V} \mid Z)$ from the exercise boundary K.
ϕ	The density of the standard normal distribution.
Φ	The cumulative distribution function of the standard normal distribution.
β	The foresight error correction.

References

[1] BINGHAM, NICHOLAS H.; KIESEL, RÜDIGER: Risk-neutral valuation: pricing and hedging of financial derivatives. (Springer Finance). Springer Verlag, London, 1998. ISBN 1-852-33458-4. 296 Seiten.

[2] BROADIE, MARK; GLASSERMAN, PAUL: Pricing American-Style Securities by Simulation. *J. Econom. Dynam. Control*, 1997, Vol. 21, 1323-1352.

[3] CARRIERE, JACQUES F.: Valuation of Early-Exercise Price of Options Using Simulations and Nonparametric Regression. *Insurance: Mathematics and Economics* 19, 19-30, 1996.

[4] CLÉMENT, EMMANUELLE; LAMBERTON, DAMIEN; PROTTER, PHILIP: An analysis of a least squares regression method for American option pricing. *Finance and Stochastics* 6, 449-471, 2002.

[5] DRAPER, NORMAN R.; SMITH, HARRY: Applied Regression Analysis. Third Edition. Wiley-Interscience, 1998. ISBN 0-471-02995-5.

[6] FRIES, CHRISTIAN: Pricing Options with Early Exercise by Monte Carlo Simulation: An Introduction. 2005, http://www.christian-fries.de/finmath/earlyexercise.

[7] GLASSERMAN, PAUL: Monte Carlo Methods in Financial Engineering. 596 Seiten. Springer, 2003. ISBN 0-387-00451-3.

[8] LONGSTAFF, FRANCIS A.; SCHWARTZ EDUARDO S.: Valuing American Options by Simulation: A Simple Least-Square Approach. *Review of Financial Studies* 14:1, 113-147, 2001.

[9] PITERBARG, VLADIMIR V.: A practitioner's guide to pricing and hedging callable LIBOR exotics in forward LIBOR models, *Preprint*. 2003.

Notes

Suggested Citation

Fries, Christian P.: Foresight Bias and Suboptimality Correction in Monte-Carlo Pricing of Options with Early Exercise: Classification, Calculation and Removal. (2005).
http://www.christian-fries.de/finmath/foresightbias.

Classification: MSC-class: 65C05 (Primary), 68U20, 60H35 (Secondary).
ACM-class: G.3; I.6.8.
JEL-class: C15, G13.

Keywords: Early-Exercise Option, Monte-Carlo, Least Square Monte Carlo, Foresight, Perfect Foresight, Suboptimal Exercise Correction

27 pages. 11 figures. 5 tables.