

Pricing Options with Early Exercise by Monte Carlo Simulation: An Introduction

Christian P. Fries
email@christian-fries.de

December 5, 2005

(Version 0.3)

Abstract

In this paper we review some of the most common methods used in pricing Bermudan (or American) type options by Monte Carlo simulation. We assume basic knowledge of risk neutral pricing theory¹ and Monte Carlo simulation².

Note

This version contains only the first chapters on *lower bound methods*. If you are interested in a discussion of *upper bound methods* check the papers home page

<http://www.christian-fries.de/finmath/earlyexercise>

for an updated version.

¹ For an introduction to risk neutral pricing see, e.g., [3, 11].

² For an introduction to Monte Carlo simulation see, e.g., [4, 5, 11, 12, 14].

Contents

1	Introduction	3
2	Bermudan Options: Notation	3
2.1	Bermudan Callable	4
2.2	Relative Prices	4
3	Bermudan Option as Optimal Exercise Problem	5
3.1	Bermudan Option Value as single (unconditioned) Expectation: The Optimal Exercise Value	5
4	Bermudan Option Pricing - The Backward Algorithm	6
5	Re-simulation	7
6	Perfect Foresight	7
7	Conditional Expectation as Functional Dependence	8
8	Binning	9
8.1	Binning as a Least-Square Regression	10
9	Foresight Bias	11
10	Regression Methods - Least Square Monte Carlo	12
10.1	Least Square Approximation of the Conditional Expectation	12
10.2	Example: Evaluation of an <i>Bermudan</i> Option on a Stock (Backward Algorithm with Conditional Expectation Estimator)	13
10.3	Example: Evaluation of an Bermudan Callable	14
10.4	Implementation	17
10.5	Binning as linear Least-Square Regression	17
11	Optimization Methods	19
11.1	Andersen Algorithm for Bermudan Swaptions	19
11.2	Review of the Threshold Optimization Method	20
11.2.1	Fitting the exercise strategy to the product	20
11.2.2	Disturbance of the Optimizer through Monte Carlo Error	21
11.3	Optimization of Exercise Strategy: A more general Formulation	21
11.4	Comparisson of Optimization Method and Regression Method	22
12	Duality Method: Upper Bound for Bermudan Option Prices	23
	References	24
	Notes	25

1 Introduction

Let us first consider the simple case of a Bermudan option $V_{\text{berm}}(T_1, T_2)$ with two exercise dates only: The option holder has the right to receive an underlying $V_{\text{underl},1}$ in T_1 or wait and retain the right to either receive an underlying $V_{\text{underl},2}$ in T_2 or receive nothing. Put differently, the option holder has the choice of receiving the underlying value $V_{\text{underl},1}(T_1)$ and the value of an option $V_{\text{option}}(T_1)$ on $V_{\text{underl},2}(T_2)$. The Bermudan may be interpreted as an option on an option. In T_1 the optimal exercise is given by choosing the maximum value

$$\max(V_{\text{underl}}(T_1), V_{\text{option}}(T_1)), \quad (1)$$

where (having chosen a Numéraire N)

$$\begin{aligned} V_{\text{option}}(T_1) &= N(T_1) E^{\mathbb{Q}^N} \left(\frac{V_{\text{option}}(T_2)}{N(T_2)} \mid \mathcal{F}_{T_1} \right) \\ &= N(T_1) E^{\mathbb{Q}^N} \left(\frac{\max(V_{\text{underl},2}(T_2), 0)}{N(T_2)} \mid \mathcal{F}_{T_1} \right). \end{aligned} \quad (2)$$

is the value of the option with exercise in T_2 , evaluated in T_1 .

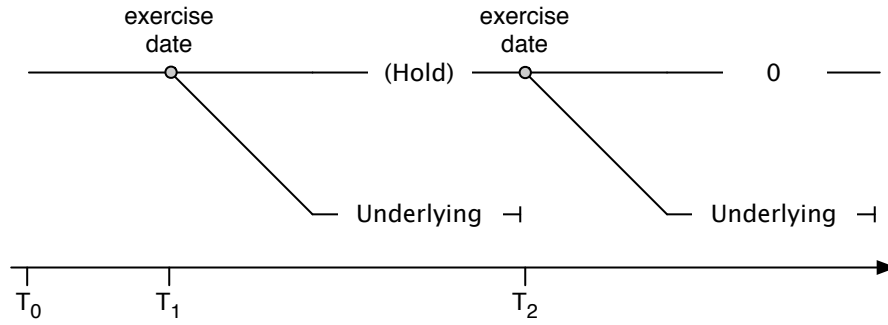


Figure 1: A simple Bermudan option with two exercise dates.

Thus, to evaluate the exercise criteria (1) it is in general necessary to calculate a conditional expectation. The calculation of a conditional expectation within a Monte-Carlo simulation is a non trivial problem. The two main issues are *complexity* and *foresight bias*, which we will illustrate in Section 5 and 6. In the then following Section we will present methods to efficiently estimate conditional expectations and/or the Bermudan exercise criteria within Monte-Carlo simulation. The application to the pricing of Bermudan option will always be present, but apart from the description of the backward algorithm in Section 4 the methods presented are not limited to Bermudan option pricing.

2 Bermudan Options: Notation

We now give a fairly general definition of an Bermudan option and fix notation. Let $\{T_i\}_{i=1,\dots,n}$ denote a set of exercise dates and $\{V_{\text{underl},i}\}_{i=1,\dots,n}$ a corresponding set of underlyings. The Bermudan option is the right to receive at one and only one time T_i the corresponding underlying $V_{\text{underl},i}$ (with $i = 1, \dots, n$) 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

$$V_{\text{berm}}(T_i, \dots, T_n; T_i) := \max(V_{\text{berm}}(T_{i+1}, \dots, T_n; T_i), V_{\text{underl},i}(T_i)), \quad (3)$$

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

2.1 Bermudan Callable

The most common Bermudan option is the Bermudan Callable. For a Bermudan Calable the underlyings consist of periodic payments X_k and differ only in the beginning of the periodic payments. The value of the underlying then becomes

$$V_{\text{underl},i}(T_i) = V_{\text{underl}}(T_i, \dots, T_n; T_i) = N(T_i) \mathbb{E}^{\mathbb{Q}^N} \left(\sum_{k=i}^{n-1} \frac{X_k}{N(T_{k+1})} \middle| \mathcal{F}_{T_i} \right)$$

Here X_k denotes a payment fixed in T_k (i.e. \mathcal{F}_{T_k} measurable) and paid in T_{k+1} . This is the common setup for interest rate bermudan callable. Other payment dates are a minor modification, they simply change the time argument of the Numéraire. For the value upon non exercise we have as before

$$V_{\text{berm}}(T_{i+1}, \dots, T_n; T_i) = N(T_i) \mathbb{E}^{\mathbb{Q}^N} \left(\frac{V_{\text{berm}}(T_{i+1}, \dots, T_n; T_{i+1})}{N(T_{i+1})} \middle| \mathcal{F}_{T_i} \right).$$

If the value of the underlying may not be expressed by means of an analytical formula *two* conditional expectations have to be evaluated to calculate the exercise strategy (3).

2.2 Relative Prices

Since the conditional expectation of a Numéraire relative price is a Numéraire relative price the presentation will be simplified by considering the Numéraire relative quantities. We will therefore define:

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

thus we have

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

and in the case of a Bermudan callable

$$\tilde{V}_{\text{underl},i}(T_j) = \mathbb{E}^{\mathbb{Q}^N} \left(\sum_{k=i}^{n-1} \tilde{X}_k(T_{k+1}) \mid \mathcal{F}_{T_j} \right) \quad \text{where} \quad \tilde{X}_i(T_{i+1}) := \frac{X_i}{N(T_i)}.$$

The relative prices are marked by a tilde.

Remark 1 (Notation): The processes $t \mapsto \tilde{V}_{\text{underl},i}(t)$ and $t \mapsto \tilde{V}_{\text{berm},i}(t)$ are \mathcal{F}_t conditional expectations of $\tilde{V}_{\text{underl},i}(T_i)$ and $\tilde{V}_{\text{berm},i}(T_i)$ respectively and thus martingales by definition. The time-discrete processes $i \mapsto \tilde{V}_{\text{underl},i}(T_i)$, $i \mapsto \tilde{V}_{\text{berm},i}(T_i)$ consist of different products at different times and are thus not time-discrete martingales in general.

3 Bermudan Option as Optimal Exercise Problem

A Bermudan option consists of the right to receive one (and only one) of the underlyings $V_{\text{underl},i}$ at the corresponding exercise date T_i . The recursive definition (3) represents the optimal exercise strategy in each exercise time. We formalize this optimal exercise strategy:

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

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

and

$$\eta : \{1, \dots, n-1\} \times \Omega \rightarrow \{0, 1\}, \quad \eta(i, \omega) := \begin{cases} 1 & \text{falls } T(\omega) \geq T_i \\ 0 & \text{sonst.} \end{cases}$$

The definitions of T and η give equivalent descriptions of the exercise strategy: $T(\omega)$ is the optimal exercise time on a given path ω , $\eta(\cdot, \omega)$ is an indicator function which changes from 0 to 1 at the time index i corresponding to $T_i = T(\omega)$. The boundary $\partial\{\eta = 1\}$ of the set $\{\eta = 1\}$ is the so called exercise boundary. It should be noted that $\eta(k)$ is \mathcal{F}_{T_k} measurable.

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

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

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

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

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

For the Bermudan Callable we may alternatively write

$$\tilde{V}_{\text{berm}}(T_0) = \mathbb{E}^{\mathbb{Q}}\left(\sum_{k=1}^{n-1} \tilde{X}_k(T_{k+1}) \cdot \eta(k) \mid \mathcal{F}_{T_0}\right).$$

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 2 (Option Value upon Optimal Exercise): ▮

Let \tilde{U} be the stochastic process whose time t value $U(t)$ is the (Numéraire relative) option value received upon exercise in t . Let T be the optimal exercise strategy. The random variable $\tilde{U}(T)$, where

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

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

Thus the value of $\tilde{V}_{\text{berm}}(T_1, \dots, T_n)$ 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 $T(\omega)$ (and thus $\eta(\cdot, \omega)$) for any path ω .

Remark 3 (Stopped Prozess): The random variable \tilde{U} is a so called *stopped process*. \tilde{U} is a stochastic process and T is a random variable with the interpretation of a (stochastic) time. Furthermore T is a stopping time, see Definition ???. Here the stochastic process \tilde{U}

is the family of underlyings received upon exercise, parametrized by exercise time, and T is the optimal exercise time. Thus $\tilde{U}(T)$ is the underlying received upon optimal exercise. All quantities are stochastic, i.e. depend on the path.

4 Bermudan Option Pricing - 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 (3), i.e. the conditional expectation. The algorithm consists of the application of the recursive definition of the Bermudan value in (3) with a slight modification. Let:

Induction start:

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

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

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

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

$$\tilde{V}_{\text{berm}}(T_1, \dots, T_n, T_0) = \mathbb{E}^{\mathbb{Q}}(\tilde{U}_1 | \mathcal{F}_{T_0}) \quad (4)$$

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

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

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

and

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

This is a subtle but crucial difference. While both definitions give the Bermudan option value (through application of (4)), we have that the definition of \tilde{U}_i requires the conditional expectation operator only to calculate the exercise criteria.

Since a Monte-Carlo simulation requires advanced methods to obtain an (often not very accurate) estimate for the conditional expectation it is important to reduce their use.

Note that $\tilde{V}_{\text{berm},i}(T_i)$ is \mathcal{F}_{T_i} measurable by definition as a \mathcal{F}_{T_i} conditional expectation, while all \tilde{U}_i are at most \mathcal{F}_{T_n} measurable since they are defined pathwise from \mathcal{F}_{T_k} measurable random variables $\tilde{V}_{\text{underl},k}(T_k)$ for $i \leq k \leq n$.

The pricing of a Bermudan option may thus be reduced to either the calculation of conditional expectations or to the calculation of the optimal exercise strategy T .

As motivation we consider in Sections 5 and 6 two methods which are not suitable to calculate conditional expectations.

5 Re-simulation

Let us consider the simplified example of a Bermudan option as given in Section 1. If no analytic calculation of the conditional expectation (2) is possible and if Monte-Carlo is the numerical tool to calculate expectations the straight forward way to calculate the conditional expectation is to create in T_1 a new Monte-Carlo simulation (conditioned) on each path - see Figure 2. This leads to a much higher number of total simulation paths needed. Especially

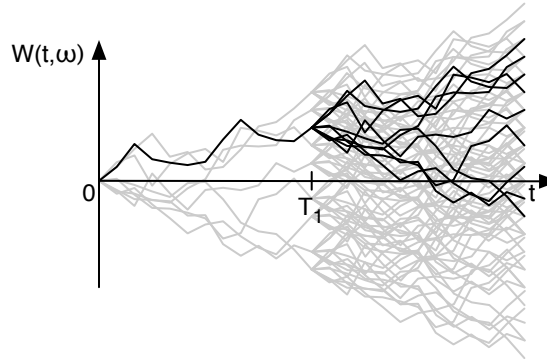


Figure 2: Brute-Force Berechnung der bedingten Erwartung in einer Pfad-Simulation durch (pfadweise) Resimulation

if one considers more than one exercise date (option on option on option...) this method becomes impractical. The required number of paths, i.e. the complexity of the algorithm and thus the calculation time, grows exponentially with the number of exercise dates. This creates the need for efficient alternatives.

Interpretation: The calculation of conditional expectation in a path simulation requires further measures since the path simulation does not offer a suitable discretization of the filtration.

6 Perfect Foresight

If one refuses to use a full resimulation and sticks to the paths generated in the original simulation then one effectively estimates the conditional expectation by a single path, namely by

$$E^Q \left(\frac{V(T_2)}{N(T_2)} \mid \mathcal{F}_{T_1} \right) \approx \frac{V(T_2)}{N(T_2)}.$$

Put differently this is a limit case of the resimulation where each resimulation consists of a single path only, namely the one of the original simulation. If this estimate is used in the exercise criteria the exercise will be *super optimal* since it is based on future information that would be unknown otherwise.

The exercise criteria at time T_1 may only depend on information available in T_1 , i.e. on \mathcal{F}_{T_1} measurable random variables. The estimate $\frac{V(T_2)}{N(T_2)}$ is *not* \mathcal{F}_{T_1} measurable.

To illustrate the super optimality consider a simulation consisting of two path, see Figure 3. Both path are identical on $[0, T_1]$, i.e. $\mathcal{F}_{T_1} = \{\emptyset, \Omega\} = \{\emptyset, \{\omega_1, \omega_2\}\}$. We consider the option V to receive either $S(T_1) = 2$ or $S(T_2) \in \{1, 4\}$ at later time T_2 . The random variable $\eta : \Omega \rightarrow \{0, 1\}$ denotes the exercise strategy for T_1 : It is 1 on paths that exercise in T_1 , else 0. With a *perfect foresight* the super optimal exercise strategy is $T(\omega_1) = T_2$, $T(\omega_2) = T_1$, i.e. $\eta(\omega_1) = 0$, $\eta(\omega_2) = 1$ and an average value of $V(T_0) = \frac{1}{2}(4 + 2) = \frac{6}{2}$ will be received. Note that then η is not \mathcal{F}_{T_1} measurable. The exercise decision is made in T_1 with knowledge

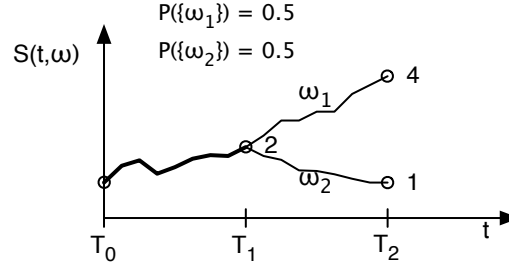


Figure 3: Illustration of perfect foresight.

of the future outcome. If we restrict the exercise strategy to the set of \mathcal{F}_{T_1} measurable random variables we either get $\frac{1}{2}(4 + 1) = \frac{5}{2}$ using $\eta \equiv 0$ or $\frac{1}{2}(2 + 2) = \frac{4}{2}$ using $\eta \equiv 1$. Thus the optimal, \mathcal{F}_{T_1} measurable (and thus admissible) exercise strategy is $\eta(\omega_1) = \eta(\omega_2) = 0$.

Perfect foresight is not a suitable method to estimate conditional expectation and through this the exercise criteria.

7 Conditional Expectation as Functional Dependence

Let us reconsider the calculation of the conditional expectation through brute force re-simulation as described in Section 5 and depicted in Figure 2. 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). The initial conditions are \mathcal{F}_{T_1} measurable random variables (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

$$\mathbb{E}^{\mathbb{Q}^N} \left(\frac{V(T_2)}{N(T_2)} \mid \mathcal{F}_{T_1} \right) = \mathbb{E}^{\mathbb{Q}^N} \left(\frac{V(T_2)}{N(T_2)} \mid Z \right). \quad (5)$$

Interpretation: If the random variable Z is such that \mathcal{F}_{T_1} is the smallest σ field with respect to which Z is measurable (i.e. we have $Z^{-1}(\mathcal{B}(\mathbb{R}^d)) = \mathcal{F}_{T_1}$), then equation (5) is merely the definition of an expectation conditioned to a random variable. If however the conditional expectation on the left hand side (i.e. $\frac{V(T_2)}{N(T_2)}$) is known to be measurable with respect to a smaller σ field (e.g. because its functional dependence relies on a smaller set of random variable), then it might be advantageous to use the right hand side representation. This representation is also useful to derive approximation, e.g. if the functional dependence with respect to one component of Z is known to be weak and thus neglectable.

Example: Consider a LIBOR Market Model with stochastic processes for the forward rates L_1, L_2, \dots, L_n . In T_1 we wish to calculate the conditional expectation of a derivative with a Numéraire relative payoff that depends on L_2, \dots, L_k only (e.g. on a swap rate). While the filtration \mathcal{F}_{T_1} is generated by the full set of forward rates $L_1(T_1), L_2(T_1), \dots, L_n(T_1)$ it is sufficient to know $L_2(T_1), \dots, L_k(T_1)$ to describe the conditional expectation (i.e. the conditional value of the product).

We will now describe methods that derive the functional dependance of the conditional expectation from a given set of random variables.

8 Binning

In a path simulation the approximation of $E^{\mathbb{Q}} \left(\frac{V(T_2)}{N(T_2)} \mid Z \right)$ will be given by averaging over all paths for which Z attains the same value. For the simple example in Figure 3 this would remove the *perfect foresight* since there $S(T_1)^{-1}(2) = \Omega$. 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 unfeasable resimulation. Thus this approximation will show a perfect foresight.

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:

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

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

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

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

will then be given by

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

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.

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

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

$$dS(t) = r \cdot S(t)dt + \sigma S(t)dW^{\mathbb{Q}}(t), \quad N(t) = \exp(rt)$$

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

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

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 4, left. 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

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

which in turn may be used as estimate of

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

In Figure 4, right, we calculate this estimate for $S_1 = 1$ and $\epsilon = 0.05$.

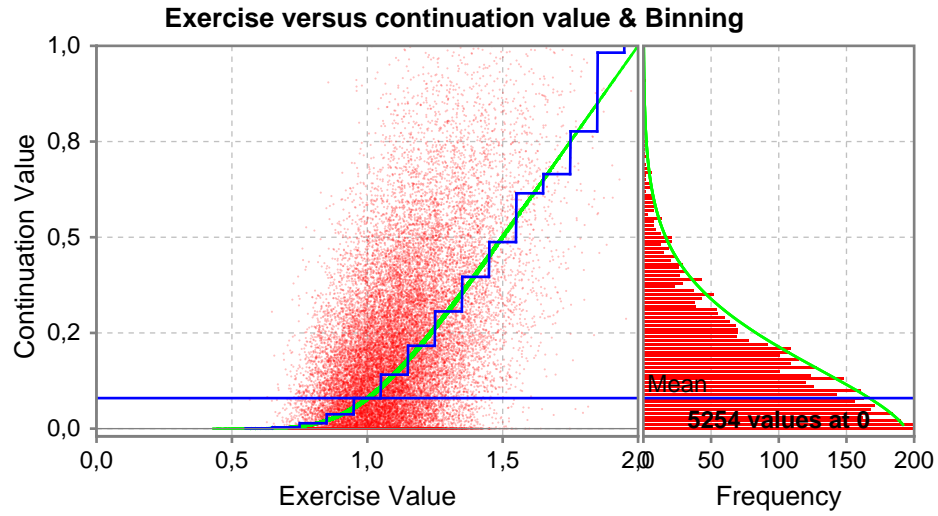


Figure 4: The Continuation Value as a function of the Underlying (Spot Value) and the calculation of the conditional expectation through a binning

8.1 Binning as a Least-Square Regression

Consider again the binning: As estimate for the conditional expectation $\mathbb{E}^{\mathbb{Q}} \left(\frac{V(T_2)}{N(T_2)} \mid Z(\omega) \right)$ we calculated the conditional expectation

$$H_i := \mathbb{E}^{\mathbb{Q}} \left(\frac{V(T_2)}{N(T_2)} \mid Z \in U_i \right) \quad (6)$$

given a *bin* U_i with $Z(\omega) \in U_i$.

For the expectation operator $\mathbb{E}^{\mathbb{Q}}$ an alternative characterization may be used:

Lemma 7 (Characterization of the Expectation as Least-Square Approximation): The expectation of a random variable X is the number h for which $X - h$ has the smallest variance (i.e. $L_2(\Omega)$ norm).

Proof: Let X be a real valued random variable. Then we have for any $h \in \mathbb{R}$

$$E((X - h)^2) = E(X^2) - 2 \cdot E(X) \cdot h + h^2 =: f(h).$$

Since $f' = 0 \Leftrightarrow h = E(X)$ and $f'' = E(X^2) \geq 0$ we have that f attains its minimum in $h = E(X)$. For vector valued random variables this follows componentwise. The same results holds for conditional expectations. \square

Using Lemma 7 we may write (6) as minimization problem:

$$\mathbb{E}^{\mathbb{Q}} \left(\left(\frac{V(T_2)}{N(T_2)} - H_i \right)^2 \mid Z \in U_i \right) = \min_{G \in \mathbb{R}} \mathbb{E}^{\mathbb{Q}} \left(\left(\frac{V(T_2)}{N(T_2)} - G \right)^2 \mid Z \in U_i \right).$$

For disjoint bins U_i this may be written in a single minimization problem for the vector $(H_i)_{i=1, \dots, \dots}$:

$$\sum_i \mathbb{E}^{\mathbb{Q}} \left(\left(\frac{V(T_2)}{N(T_2)} - H_i \right)^2 \mid Z \in U_i \right) = \min_{G_i \in \mathbb{R}} \sum_i \mathbb{E}^{\mathbb{Q}} \left(\left(\frac{V(T_2)}{N(T_2)} - G_i \right)^2 \mid Z \in U_i \right). \quad (7)$$

This condition admits an alternative interpretation: H_i represents the piecewise constant function (constant on U_i) with the minimal distance to $\frac{V(T_2)}{N(T_2)}$ in the least square sense.

Let \mathcal{H} be the space of functions $H : \Omega \rightarrow \mathbb{R}$ being constant on the bins $Z^{-1}(U_i)$.³ Let $H \in \mathcal{H}$ with $H(\omega) := H_i$ for $\omega \in Z^{-1}(U_i)$. Then (7) is equivalent to

$$\mathbb{E}^{\mathbb{Q}} \left(\left(\frac{V(T_2)}{N(T_2)} - H \right)^2 \mid Z \right) = \min_{G \in \mathcal{H}} \mathbb{E}^{\mathbb{Q}} \left(\left(\frac{V(T_2)}{N(T_2)} - G \right)^2 \mid Z \right). \quad (8)$$

Equation (8) is the definition of a regression: Find the function H from a function space \mathcal{H} with minimum distance to $\frac{V(T_2)}{N(T_2)}$ in the L_2 norm. Binning is just a special choice of functional space:

Lemma 8 (Binning as L_2 Regression): Binning is an L_2 regression on the space of functions piecewise constant on U_i .

9 Foresight Bias

Definition 9 (Foresight Bias (Definition 1)): \lrcorner

A *foresight bias* is a super optimal exercise strategy. \lrcorner

A foresight bias arises due to a violation of the measurability requirements: If the exercise decision in T_i is based on random variable which is not \mathcal{F}_{T_i} measurable the exercise may be super optimal, i.e. better than if based on the information theoretically available (\mathcal{F}_{T_i}). If we use the same Monte-Carlo simulation to first estimate the exercise criteria and then use this criteria to price the derivative we most certain generate a foresight bias. In this case the foresight bias is created by the Monte-Carlo error of the estimate, which is in general not \mathcal{F}_{T_i} measurable. The existence of this problem becomes obvious if we consider a limit case of binning where each bin contains a single path only. Here we would have a [perfect foresight](#).

If our exercise criteria at time T_i uses only \mathcal{F}_{T_i} measurable random variables then there is - in theory - no foresight bias. If however the exercise criteria is calculated within a Monte Carlo simulation the Monte-Carlo error of the calculation represents a non \mathcal{F}_{T_i} measurable random variable, thus it induces a foresight bias. In this case we may give an alternative definition for the foresight bias:

Definition 10 (Foresight Bias (Definition 2)): \lrcorner

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

³ Note that the bins U_i were defined as subsets of $Z(\Omega)$, here we consider H as a function on Ω .

With increasing number of paths the foresight bias introduced by binning tends to zero since the Monte-Carlo error with respect to a bin tends to zero.

A general solution of the problem of a foresight bias is given by using two independent Monte-Carlo simulations: One to estimate the exercise criteria (for binning this is given by the H_i corresponding to the U_i 's), the other to apply the criteria in pricing. This is a numerical removal of the foresight bias. In [10] an analytic formula for the (Monte Carlo error induced) foresight bias is derived that may be used to correct the foresight bias analytically.

10 Regression Methods - Least Square Monte Carlo

Motivation (Disadvantage of Binning): The partition of the state space $Z(\Omega)$ into a finite number of bins results in a piecewise constant approximation of the conditional expectation. An obvious improvement would be to give the conditional expectation as a smooth function of the state variable Z .

The considerations in Section 8.1 suggest a simple yet powerful improvement to the binning: The function giving our estimate for the conditional expectation is defined by a least square approximation (regression).

10.1 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 12 (Least Square Approximation of the Conditional Expectation): \lrcorner

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

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

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

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

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

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

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^* . \lrcorner

The method of Carriere, Longstaff and Schwartz⁴ uses a function f with $q = p$ and

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

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

Lemma 13 (Linear Regression): Let $\Omega^* = \{\omega_1, \dots, \omega_n\}$ be a given sample space, $V : \Omega^* \rightarrow \mathbb{R}$ and $Y := (Y_1, \dots, Y_p) : \Omega^* \rightarrow \mathbb{R}^p$ given random variables. Furthermore let

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

⁴ See [6, 15, 16].

Then we have for any α^* with $X^\top X \alpha^* = X^\top v$

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

where

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

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

Definition 14 (Basis Functions):

The random variables Y_1, \dots, Y_p of Lemma 13 are called *Basis Functions* (explanatory variables). ⌋

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

We consider a simple Bermudan option on a Stock. The Bermudan should allow exercise at times $T_1 < T_2 < \dots < T_n$. 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 received 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: $t = T_i, i = n, n-1, n-2, \dots, 1$ In T_i we have:

- In the case of exercise in T_i the value is

$$\tilde{V}_{\text{underl},i}(T_i) := \frac{N_i(S(T_i) - K_i)}{N(T_i)}. \quad (9)$$

- In the case of non-exercise in T_i the value is $\tilde{V}_{\text{hold},i}(T_i) = E^{\mathbb{Q}}(\tilde{U}_{i+1} \mid \mathcal{F}_{T_i})$. This value is estimated through a regression for given paths $\omega_1, \dots, \omega_m$:

- Let B_j be given (\mathcal{F}_{T_i} measurable) basis functions.⁵ Let the matrix X consist of the column vectors $B_j(\omega_k), k = 1, \dots, m$. Then we have

$$\begin{pmatrix} \tilde{V}_{\text{hold},i}(T_i, \omega_1) \\ \vdots \\ \tilde{V}_{\text{hold},i}(T_i, \omega_m) \end{pmatrix} \approx X \cdot (X^\top \cdot X)^{-1} \cdot X^\top \cdot \begin{pmatrix} \tilde{U}_{i+1}(\omega_1) \\ \vdots \\ \tilde{U}_{i+1}(\omega_m) \end{pmatrix}. \quad (10)$$

⁵ 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 $S(T_i)$.

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

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

Remark 15: Our example is of course just the backward algorithm with an explicit specification of an underlying (9) and an explicit specification of an exercise criteria, here given by the estimator of the conditional expectation (10).

10.3 Example: Evaluation of an Bermudan Callable

We consider a Bermudan Callable. The Bermudan should allow exercise at times $T_1 < T_2 < \dots < T_n$. Upon exercise in T_i the holder of the option will receive a payment of X_i in T_{i+1} , i.e. the relative value $\tilde{X}_i(T_{i+1}) := \frac{X_i}{N(T_{i+1})}$.

We will apply the *backward algorithm* to derive the optimal exercise strategy. All payments will be considered in their Numéraire relative form.

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: $t = T_i, i = n, n-1, n-2, \dots, 1$ In T_i we have:

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

$$\tilde{V}_{\text{underl},i}(T_i) := E^{\mathbb{Q}^N} \left(\sum_{k=i}^{n-1} \frac{X_k}{N(T_{k+1})} \middle| \mathcal{F}_{T_i} \right).$$

This value is estimated through a regression for given paths $\omega_1, \dots, \omega_m$:

Let B_j^1 be given (\mathcal{F}_{T_i} measurable) basis functions. Let the matrix X^1 consist of the column vectors $B_j^1(\omega_k), k = 1, \dots, m$. Then we have

$$\begin{pmatrix} \tilde{V}_{\text{underl},i}(T_i, \omega_1) \\ \vdots \\ \tilde{V}_{\text{underl},i}(T_i, \omega_m) \end{pmatrix} \approx X^1 \cdot (X^{1,\top} \cdot X^1)^{-1} \cdot X^{1,\top} \cdot \begin{pmatrix} \sum_{k=i}^{n-1} \frac{X_k(\omega_1)}{N(T_{k+1}, \omega_1)} \\ \vdots \\ \sum_{k=i}^{n-1} \frac{X_k(\omega_m)}{N(T_{k+1}, \omega_m)} \end{pmatrix}. \quad (11)$$

- In the case of non-exercise in T_i the value is $\tilde{V}_{\text{hold},i}(T_i) = E^{\mathbb{Q}}(\tilde{U}_{i+1} \mid \mathcal{F}_{T_i})$. This value is estimated through a regression for given paths $\omega_1, \dots, \omega_m$:

Let B_j^0 be given (\mathcal{F}_{T_i} measurable) basis functions. Let the matrix X^0 consist of the column vectors $B_j^0(\omega_k), k = 1, \dots, m$. Then we have

$$\begin{pmatrix} \tilde{V}_{\text{hold},i}(T_i, \omega_1) \\ \vdots \\ \tilde{V}_{\text{hold},i}(T_i, \omega_m) \end{pmatrix} \approx X^0 \cdot (X^{0,\top} \cdot X^0)^{-1} \cdot X^{0,\top} \cdot \begin{pmatrix} \tilde{U}_{i+1}(\omega_1) \\ \vdots \\ \tilde{U}_{i+1}(\omega_m) \end{pmatrix}.$$

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

$$\tilde{U}_i := \begin{cases} \sum_{k=i}^{n-1} \frac{X_k}{N(T_{k+1})} & \text{if } \tilde{V}_{\text{hold},i}(T_i) < \tilde{V}_{\text{underl},i}(T_i) \\ \tilde{U}_{i+1} & \text{else.} \end{cases}$$

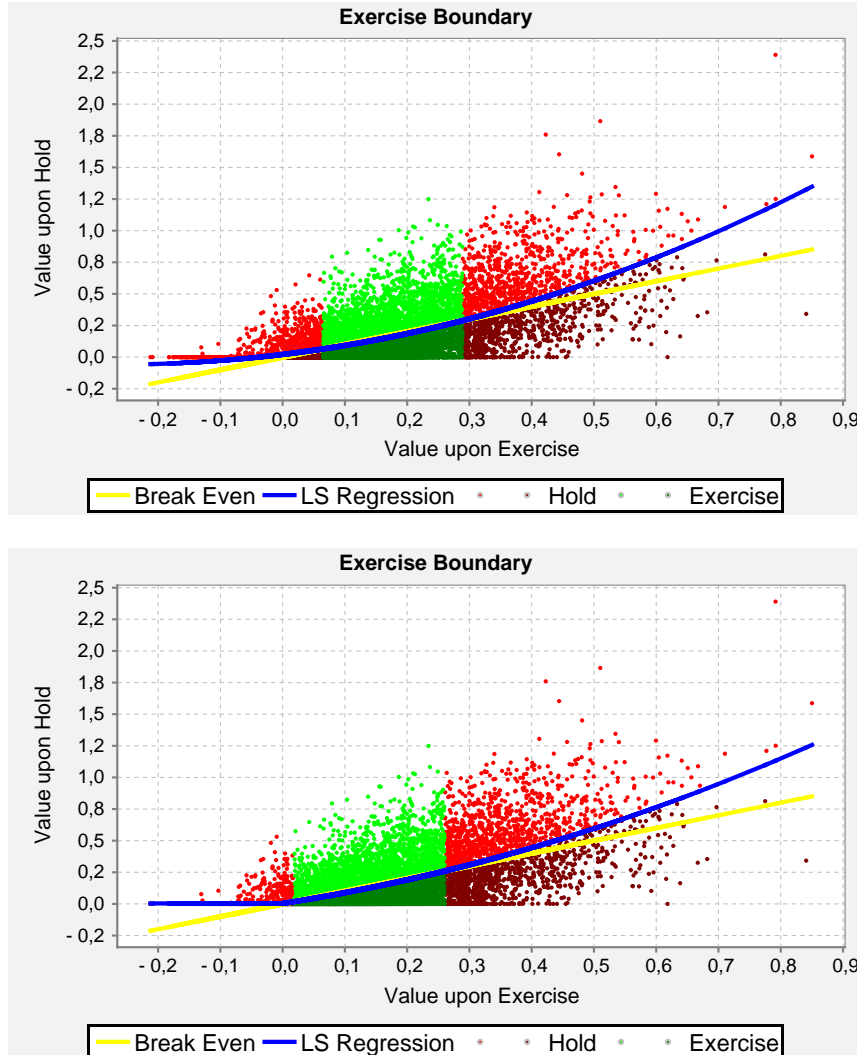


Figure 5: Regression of the conditional expectation estimator with and without restriction of the regression domain: We consider a Bermudan option with two exercise dates $T_1 = 1.0$, $T_1 = 2.0$. Notional and strike are as follows: $N_1 = 0.7$, $N_2 = 1.0$, $K_1 = 0.82$, $K_2 = 1.0$. The model for the underlying S is a Black-Scholes model with $r = 0.05$ and $\sigma = 20\%$. The plot shows the values received upon exercise depending on the values received upon non-exercise in T_1 . Each dot corresponds to a path. The regression polynomial gives the estimator for the expectation of the value upon non-exercise. It is optimal to exercise if this estimate lies above the value received upon exercise. In the upper diagram the regression polynomial is a second order polynomial in $S(T_1)$. In the lower diagram it is a second order polynomial in $\max(S(T_1) - K_1, 0)$. By this values where $S(T_1) - K_1 \leq 0$ are aggregated into a single point. For the product under consideration this is advantageous since for $S(T_1) - K_1 \leq 0$ exercise is not optimal with probability 1. This restriction of the regression domain increases the regression accuracy over the remaining regression domain.

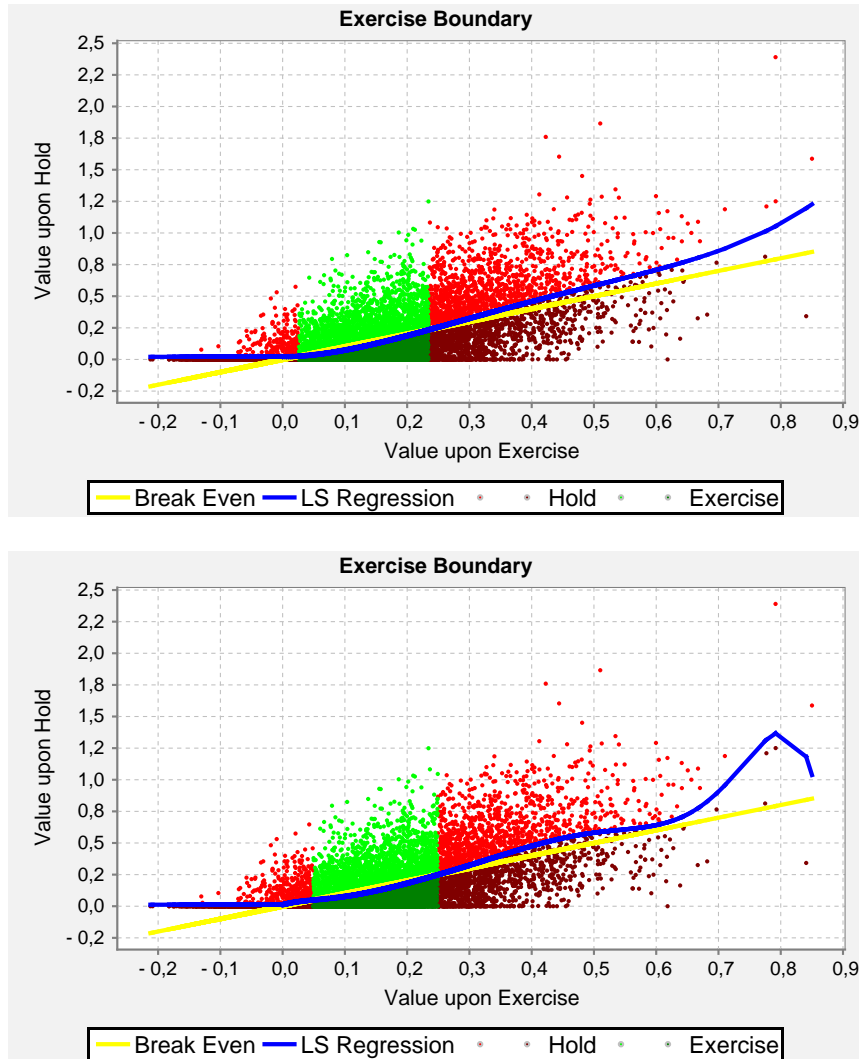


Figure 6: Regression of the conditional expectation estimator - polynomial of fourth (above) and eighth (below) order in $\max(S(T_1) - K_1, 0)$. Parameters as in Figure 5. A polynomial of higher order shows wiggles at the boundary of the regression domain. However only few paths are affected by the wrong estimate. Restricting the regression domain may reduce the errors (compare the left end of the regression domain with the right end).

Remark 16 (Bermudan Callable): The modification to the backward algorithm to price a Bermudan callable consists of the inductions of two conditional expectation estimators: one for the continuation value and (additionally) one for the underlying. As before, the conditional expectation estimators are used only for the exercise criteria (and not for the payment).

Remark 17 (Longstaff-Schwartz):

- The estimator of the conditional expectation is used in the estimation of the exercise strategy only.
- The choice of basis functions is crucial to the quality of the estimate.

Clément, Lamberton and Protter showed convergence of the Longstaff-Schwartz regression method to the exact solution, see [7].

10.4 Implementation

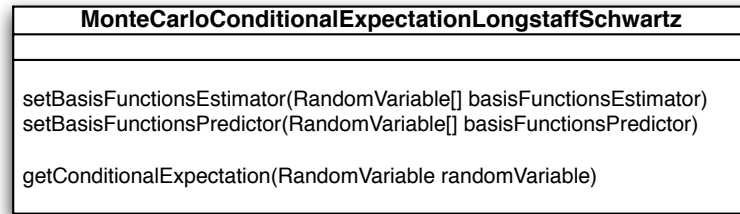


Figure 7: UML Diagram: Conditional Expectation Estimator. The method `setBasisFunctionsEstimator` sets the basis functions which forms the matrix X . The method `setBasisFunctionsPredictor` sets the basis functions which form the matrix X^* . These are the same basis functions as for X , but possibly evaluated in an independent Monte-Carlo simulation (to avoid foresight bias). The method `getConditionalExpectation` calculates the regression parameter $\alpha^* = (X^T X)^{-1} X^T \cdot v$ from a given vector v and returns the conditional expectation estimator $X^* \cdot \alpha^*$ of v . See Lemma 13.

The Longstaff-Schwartz conditional expectation estimator may easily be implemented in a corresponding class, independent from the given model or Monte-Carlo simulation - see Figure 7. This class contains nothing more than a [lineare regression](#) however the methodology may be replaced by more alternative algorithms (e.g. non parametric regressions).

As pointed out in the discussion of the backward algorithm it is in general not necessary to explicitly calculate the exercise strategy in form of T to η . It is sufficient to calculate the random variables \tilde{U}_i in a backward recursion. Since finally only \tilde{U}_1 is needed to calculate the price of the Bermudan option the \tilde{U}_i 's may be stored in the same vector of Monte-Carlo realizations.

10.5 Binning as linear Least-Square Regression

We return once again to the binning. In Section 8.1 it turned out that *Binning* may be interpreted as least square regression with a specific set of basis functions: The indicator variables of the bins U_j which we denote by

$$h_j(\omega) := \begin{cases} 1 & \text{for } \omega \in U_j \\ 0 & \text{else.} \end{cases} \quad (12)$$

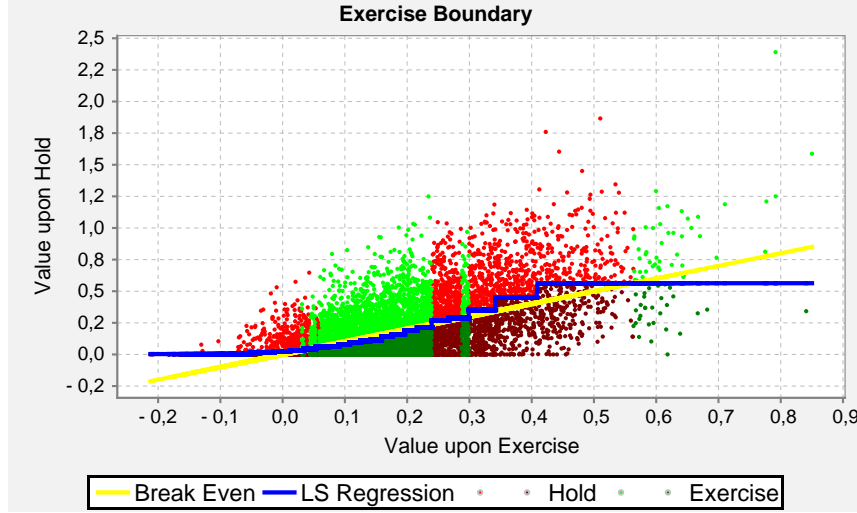


Figure 8: Binning using the linear regression algorithm with piecewise constant basis functions: We use 20 bins (basis functions). Each bin consist of approximately the same number of paths. Model- und product parameters are as in Figure 5.

We now give an explicit calculation using the linear regression algorithm with the bin indicator variables as basis functions.

Let ω_k denote the paths of a Monte-Carlo simulation and X the matrix $(h_j(\omega_k))$, j column index, k row index. Since the U_j 's are disjoint we have $X^T X = \text{diag}(n_1, \dots, n_m)$, where n_j is the number of paths for which $h_j(\omega_k) = 1$. Thus we have for the regression parameter

$$\alpha^* = (X^T X)^{-1} X^T \cdot v = \text{diag}\left(\frac{1}{n_1}, \dots, \frac{1}{n_m}\right) \cdot X^T \cdot v.$$

It follows that the regression parameter gives the expectation on the corresponding bin

$$\alpha_j^* = \frac{1}{n_j} \sum_{v_k \in U_j} v_k.$$

11 Optimization Methods

Motivation: In the discussion of the *backward algorithm* it has become obvious that the conditional expectation estimator is needed to derive the optimal exercise strategy only. Since a sub-optimal exercise will lead to a lower Bermudan price the optimal exercise has an alternative characterization: it maximizes the Bermudan value. A solution to the pricing problem of the Bermudan thus consists of maximize the Bermudan value over a suitable, sufficiently large space of admissible⁶ exercise strategy.

11.1 Andersen Algorithm for Bermudan Swaptions

The following method was proposed for the valuation of Bermudan swaptions by L. Andersen in [1]. We thus restrict our presentation to the evaluation of the Bermudan Callable and use the notation of Section 10.3. In [1] the method appears less generic than the Longstaff & Schwartz regression. However one might reformulate the optimization method in a fairly generic way. Since then the optimization is a high dimensional one the method becomes less useful in practice.

The exercise strategy is given by a parametrized function of the underlyings

$$I_i(\lambda) := f(V_{\text{underl},i}(T_i, \omega), \dots, V_{\text{underl},n-1}(T_i, \omega), \lambda)$$

where we replace the optimal exercise

$$\tilde{V}_{\text{underl},i}(T_i) < E^{\mathbb{Q}}(\tilde{U}_{i+1} \mid \mathcal{F}_{T_i})$$

by

$$I_i(\lambda) > 0.$$

Here the function f may represent a variety of exercise criterias, e.g.

$$I_i(\omega, \lambda) = V_{\text{underl},i}(T_i, \omega) - \lambda. \quad (13)$$

We assume that I_i is such that it may be calculated without resimulation, i.e. we assume that the underlyings $V_{\text{underl},j}(T_i)$ are either given by an analytic formula or a suitable approximation. E.g. this is the case for a swap within a LIBOR Market Model. Using the optimization method within the backward algorithm now looks as follows:

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: $t = T_i$, $i = n, n-1, n-2, \dots, 1$ In T_i we have:

- In the case of exercise in T_i the value is $\tilde{V}_{\text{underl},i}(T_i)$
- In the case of non-exercise in T_i the value is $\tilde{V}_{\text{hold},i}(T_i) = E^{\mathbb{Q}}(\tilde{U}_{i+1} \mid \mathcal{F}_{T_i})$. This value is estimated through an optimization for given paths $\omega_1, \dots, \omega_m$:

$$- I_i(\lambda, \omega) = f(V_{\text{underl},i}(T_i, \omega), \dots, V_{\text{underl},n-1}(T_i, \omega), \lambda).$$

⁶ By an admissible exercise strategy we denote one that respects the measurability requirements. As we noted, a violation of measurability requirements, i.e. a foresight bias or even perfect foresight, will result in a super optimal strategy. The Bermudan value with a super optimal strategy is higher than the Bermudan value with the optimal strategy, however super optimal exercise is impossible.

$$\begin{aligned}
- \tilde{U}_i(\lambda, \omega) &:= \begin{cases} \tilde{V}_{\text{underl},i}(T_i) & \text{if } I_i(\lambda, \omega) > 0 \\ \tilde{U}_{i+1} & \text{else.} \end{cases} \\
- \tilde{V}_{\text{berm},i}(T_0, \lambda) &= \mathbb{E}^{\mathbb{Q}}(\tilde{U}_i(\lambda) \mid \mathcal{F}_{T_0}) \approx \frac{1}{m} \sum_k \tilde{U}_i(\lambda, \omega_k) \\
- \lambda^* &= \arg \max_{\lambda} \left(\frac{1}{m} \sum_k \tilde{U}_i(\lambda, \omega_k) \right)
\end{aligned}$$

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

$$\tilde{U}_i(\omega) := \begin{cases} \tilde{V}_{\text{underl},i}(T_i) & \text{if } I_i(\lambda^*, \omega) > 0 \\ \tilde{U}_{i+1} & \text{else.} \end{cases} = \tilde{U}_i(\lambda^*, \omega)$$

The exercise strategy is estimated in T_i by choosing the λ^* for which I_i gives the maximal Bermudan option value. This is done by going backward in time, from exercise date to exercise date.

11.2 Review of the Threshold Optimization Method

11.2.1 Fitting the exercise strategy to the product

We apply the optimization method to the pricing of a simple Bermudan option on a stock following a Black-Scholes model. We show that choosing the exercise strategy too simple will give surprisingly unreliable results.

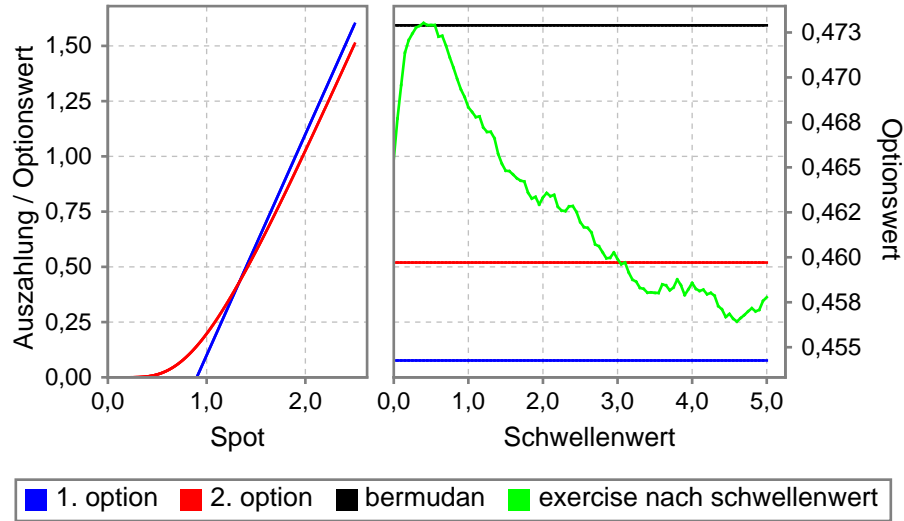


Figure 9: Example of the successful optimization of the exercise criteria (intersection of the two price curves (left)). The graph on the right shows the Bermudan option value as a function of the exercise threshold λ .

The simple strategy (13) fails for the most simple type of a Bermudan option. We consider the option to receive

$$N_1 \cdot (S(T_1) - K_1)$$

in T_1 or receive

$$\max(N_2 \cdot (S(T_2) - K_2), 0) \quad \text{in } T_2$$

in T_2 , where - as before - N_i and K_i denote notional and strike and S follows a Black-Scholes model. This gives us an analytic formula for the option in T_2 and thus the true optimal exercise.

Figure 9 shows an example where the optimization of the simple strategy (13) gives the value of the Bermudan option.

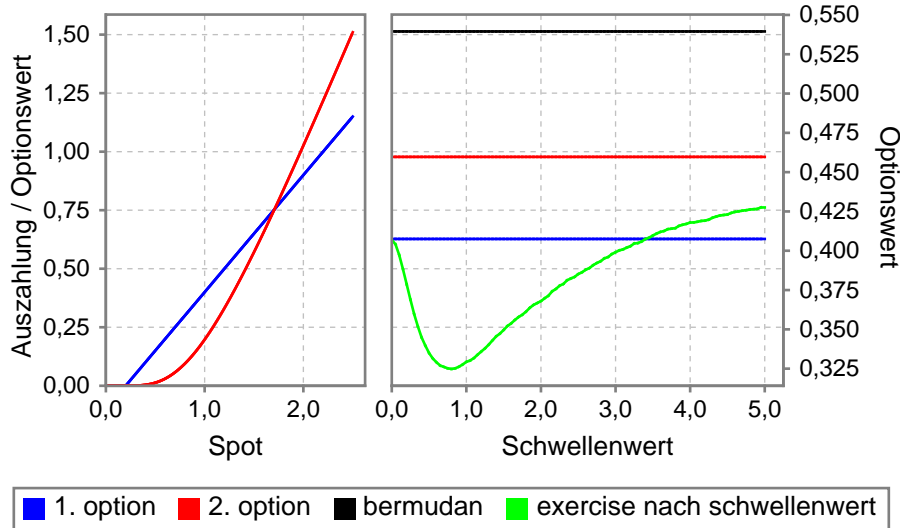


Figure 10: Example of a failing optimization of the exercise criteria (intersection of the two price curves (left)). The graph on the right shows the Bermudan option value as a function of the exercise threshold λ .

A small change of notional N_1 and strike K_1 changes the picture. If both are smaller than N_2 and K_2 respectively we may obtain two intersection points of the exercise and continuation value. In T_1 it is optimal to exercise in between these two intersection point. Our simple exercise criteria cannot render this case. Optimizing the threshold parameter λ shows two maxima: the value of the two european options "exercise never" and "exercise always". Both values are below the true Bermudan option value, see Figure fig:andersenBad, right.

The conclusion of this is example is that the choice of the exercise strategy had to be made carefully in accordance with the product. But this remark applies to any method in varying extends.

11.2.2 Disturbance of the Optimizer through Monte Carlo Error

We are maximizing a Monte-Carlo price. Since the Monte-Carlo price is an average of a finite number of path values it is (in general) a discontinuous function of the parameter λ . The price (as a function of λ) will not only exhibit discontinuities but also small local minima, see Figure 9 (right). These may prevent the minimizer algorithm from finding the global minima. Note that for sufficiently many path the local minima appear only on a small scale. Thus with a robust minimizer one would rarely encounter this problem.

11.3 Optimization of Exercise Strategy: A more general Formulation

There is a trivial generalization of the optimization method considered in Section 11.1:

- The exercise criteria will be given as a function of arbitrary \mathcal{F}_{T_i} measurable random variables.
- The exercise criteria will be given as a function of a parameter vector $\lambda \in \mathbb{R}^k$.

Thus we replace the "true" exercise criteria used in the backward algorithm

$$\tilde{V}_{\text{underl},i}(T_i) < E^Q(\tilde{U}_{i+1} \mid \mathcal{F}_{T_i})$$

by a function

$$I_i(\lambda, \omega) := f(B_{i,1}(\omega), \dots, B_{i,m}(\omega), \lambda),$$

where $B_{i,j}$ is a set of \mathcal{F}_{T_i} measurable random variables and $\lambda \in \mathbb{R}^k$.

11.4 Comparisson of Optimization Method and Regression Method

The difference between the optimization method and the regression method becomes appearend in Figur 5-6. While the regression method requires that the regression functions give a good fit to the conditional expectations across the whole domain of the independent variable, the optimization method only requires that the functional $I_i(\lambda)$ captures the exercise boundary. In Figures 5-6 the conditional expectation estimator is a curve, but the exercise boundary is given by two points only (the intersection of the conditional expectation estimator with the bisector).

Thus the optimization method can cope with far less parameters than the regression. On the other hand, as noted in the example above, it is far more important that the functional is adapted to the Bermudan product under consideration.

It is trivial to choose the map $I_i(\lambda)$ such that the optimal λ^* give the same or a better value than the least-square regression. If $B_{i,j}$ denote the basis functions used in the least-square regression for exercise date T_i we set

$$I_i(\lambda_1, \dots, \lambda_n) = (\lambda_1 B_{i,1} + \dots + \lambda_n B_{i,n}) - \tilde{V}_{\text{underl},i}(T_i).$$

Then we have that for $\lambda = \alpha^*$, where α^* is the regression parameter from the least-square regression, the exercise criteria agrees with the one of the least-square regression. This result however is rarely an advantage of the optimization method since in practice a high dimensional optimization does not represent an alternative.

12 Duality Method: Upper Bound for Bermudan Option Prices

This section is under revision.

It will reappear on or before February 30th, 2006. Please check
<http://www.christian-fries.de/finmath/earlyexercise>
for updates.

References

- [1] ANDERSEN, LEIF: A simple approach to the pricing of Bermudan swaptions in the multi-factor Libor market model. Working paper. General Re Financial Products, 1999.
- [2] ANDERSEN, LEIF; BROADIE, MARK: A Primal-Dual Simulation Algorithm for Pricing Multi-Dimensional American Options. Working paper. General Re Financial Products, 1999.
- [3] 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 pages.
- [4] BOYLE, PHELIM; BOADIE, MARK; GLASSERMAN, PAUL: Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21, 1267-1321 (1997).
- [5] BROADIE, MARK; GLASSERMAN, PAUL: Pricing American-Style Securities by Simulation. *J. Econom. Dynam. Control*, 1997, Vol. 21, 1323-1352.
- [6] CARRIERE, JACQUES F.: Valuation of Early-Exercise Price of Options Using Simulations and Nonparametric Regression. *Insurance: Mathematics and Economics* 19, 19-30, 1996.
- [7] 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.
- [8] DAVIS, MARK; KARATZAS, IOANNIS: A Deterministic Approach to Optimal Stopping, with Applications. In: Whittle, Peter (Ed.): *Probability, Statistics and Optimization: A Tribute to Peter Whittle*, 455-466, 1994. John Wiley & Sons, New York and Chichester, 1994.
- [9] DRAPER, NORMAN R.; SMITH, HARRY: *Applied Regression Analysis*. Third Edition. Wiley-Interscience, 1998. ISBN 0-471-02995-5.
- [10] 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>.
- [11] FRIES, CHRISTIAN: *Finanzmathematik. Theorie, Modellierung, Implementierung. Lecture Notes*. Frankfurt am Main, 2005, <http://www.christian-fries.de/finmath/script>.
- [12] GLASSERMAN, PAUL: *Monte Carlo Methods in Financial Engineering*. 596 pages. Springer, 2003. ISBN 0-387-00451-3.
- [13] HAUGH, MARTIN; KOGAN, LEONIS: Pricing American Options: A Duality Approach, MIT Sloan Working Paper No. 4340-01, 2001.
- [14] JÄCKEL, PETER: *Monte Carlo Methods in Finance*. 238 Seiten. John Wiley and Sons Ltd., 2002. ISBN 0-471-49741-X.
- [15] 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.

- [16] PITERBARG, VLADIMIR V.: A practitioner's guide to pricing and hedging callable LIBOR exotics in forward LIBOR models, *Preprint*. 2003.
- [17] ROGERS, L. C. G.: Monte Carlo valuation of American options, *Preprint*. 2001.

Notes

Suggested Citation

Fries, Christian P.: Pricing Options with Early Exercise by Monte Carlo Simulation: An Introduction. (2005).

<http://www.christian-fries.de/finmath/earlyexercise>.

MSC-class: 65C05 (Primary), 68U20, 60H35 (Secondary). **ACM-class:** G.3; I.6.8. **JEL-class:** C15, G13.
Keywords: Early-Exercise Options, Monte-Carlo, Foresight

25 pages. 10 figures. 0 tables.