Empirical pricing of American Options

By
Anastasiia Zalashko

Submitted to
Central European University
Department of Mathematics and its Applications

In partial fulfilment of the requirements for the degree of
Master of Science

Supervisor: Dr. László Györfi

Budapest, Hungary
2013

Table of Contents

Introduction

This thesis is devoted to the problem of pricing American options. An option is a contract that gives his holder the right to buy or sell the underlying security at an agreed moment of time (maturity date) at an agreed amount of money (exercise price). There are two basic types of options, European and American. The European option can be exercised at an exact moment of time. As for American option, it can be exercised at any moment of time up to maturity date. The price of the option is the maximal expected payoff the holder can get from exercising this option. For European options the closed-form solution exists, but for American ones the solution does not exist. Since trading the options is an everyday occurrence, this problem is important and highly investigated by many researches.

More specifically, the problem of pricing the options is the optimal stopping problem. We have to find that moment of time when we should stop the price process in order to receive the maximal payoff. That exact moment of time will be the optimal stopping time. Unfortunately, the future prices of the asset are unknown in advance. Therefore, the price of the option cannot be calculated straightforwardly. We can use the simulation methods together with advanced regression techniques to get the approximation of the solution.

When trying to solve this problem, we introduce the continuation value function. It gives the maximal expected payoff if we do not exercise the option at time t knowing the price at this time. The form of the optimal stopping time is given by the theorem in Shiryaev (2007). It is exactly the first moment of time when the payoff from immediate exercising the option is higher than the payoff from continuation. If the continuation values were known, we would be able to determine the optimal stopping time, and hence, the price of the option. Therefore, the problem of pricing is reduced to the problem of estimation of the continuation values. Continuation function is a true regression function.

Three different regression representations for the continuation function were introduced in Longstaff and Schwartz (2001) and Tsitsiklis and Van Roy (1999). These representations allow direct or recursive computation of the continuation values by computing conditional expectations. Longstaff and Schwartz (2001) suggested to use the Least Squares Monte Carlo algorithm. This method gives the estimations of the asset prices using known parameters. After, the final estimate of the option price is given using the received estimation of the prices. Nevertheless, a false price estimate can be received if the parameter estimates are biased. The representation given by Tsitsiklis and Van Roy (1999) gives us the backward recursive scheme for the computation of the continuation values.

For option pricing, Carriér (1996) was the first who suggested to apply methods of non-parametric regression estimation for the continuation function. These methods do not assume parameter estimation. Therefore, they do not have errors occurring at this step. The first article where the theoretical examination of nonparametric methods of regression estimation was conducted was Egloff (2005). However, the estimates that were defined there, were hard to compute in practice. The truncation of the estimator was firstly suggested in Egloff et al. (2007).

The article of Györfi et al. (2012) serves as a starting point for my thesis. In that article the comparison of performance of the various algorithms (Longstaff-Schwartz, parametric, nonparametric) was made. Furthermore, the authors made the comparison of the Markov and memoryless models. The main conclusion of those experiments is that, although the mean of the prices is almost the same, the variance of the prices is the smallest if we work under assumptions of the memoryless model. Notwithstanding, the estimator for the continuation function that was received was not correct. When the authors tested the results on the real data, they used more data than was available at the exact moment of time, as if we knew the data some time in advance. We present the simulation results for the changed estimator in which we use the data only available up to exact moment of time. In addition, we prove that the continuation values function in the memoryless model is non-increasing. This proves that the optimal stopping time should be exactly the first moment when the immediate payoff is higher than the payoff from continuation.

The structure of this thesis is the following. Chapter 1 contains the theory that enables us to introduce the problem of pricing further in the thesis. Chapter 2 gives the statement of the option pricing problem and the form of the optimal stopping times. Afterwards, it leads the reader to the idea of the empirical pricing methods. Chapter 3 gives the general setup for the regression problem and discusses the methods of local averaging estimates of the true regression function. Finally, chapter 4 shows two basic models that were investigated and demonstrates the empirical results that were received in this thesis.

Chapter 1

Theoretical background

This chapter contains the theoretical part that is necessary to introduce the problem of pricing the options. Here we are following Shiryaev (2007). As we will see in the next chapters, in order to find the price of the option, we need to find the optimal stopping time. Section 1.1 gives us the definition and the main properties of the stopping times.

1.1 Markov times

Let $(\Omega, \mathcal{F})$ be a measure space, let $T = [0, \infty)$, and let $F = \{\mathcal{F}_t\}$, $t \in T$ be a non-decreasing sequence of sub-σ-algebras, i.e., $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $s \leq t$.

Definition 1.1.1 *The random variable $\tau = \tau(\omega)$ with values in $\bar{T} = [0, \infty]$ is said to be a Markov time with respect to the system $F = \{\mathcal{F}_t\}$, $t \in T$ if for each $t \in T$*

$$\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t.$$

Markov times can be interpreted as random variables independent of the future.

Definition 1.1.2 *The Markov time $\tau = \tau(\omega)$ defined in a probability space $(\Omega, \mathcal{F}, P)$ is said to be a stopping time or a finite Markov time if*

$$P\{\tau(\omega) < \infty\} = 1.$$

With each Markov time $\tau = \tau(\omega)$ we may associate σ-algebra $\mathcal{F}_\tau$ of τ-past. This σ-algebra consists of the sets $A \in \mathcal{F}$ such that $A \cap \{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$ for all $t \in T$.

Claim 1.1 *$\mathcal{F}_\tau$ is σ-algebra.*

Proof. We will prove this fact by checking the axioms of σ-algebra.

(a) $\Omega \in \mathcal{F}_\tau$, since

$$\{\Omega \cap \{\omega : \tau(\omega) \leq t\}\} = \{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t.$$

The last set is in the σ-algebra $\mathcal{F}_t$ because τ is Markov time. Hence, the set Ω is in the σ-algebra $\mathcal{F}_\tau$.

(b) If $A_i \in \mathcal{F}_\tau$, then $\{A_i \cap \{\omega : \tau(\omega) \le t\}\} \in \mathcal{F}_\tau$. It follows

$$\bigcup_i (A_i \cap \{\omega : \tau(\omega) \le t\}) = \left(\bigcup_i A_i\right) \cap \{\omega : \tau(\omega) \le t\}.$$

Hence, $\bigcup_i A_i \in \mathcal{F}_\tau$

(c) If $A \in \mathcal{F}_\tau$, then $\bar{A} \in \mathcal{F}_\tau$, because

$$\bar{A} \cap \{\omega : \tau(\omega) \le t\} = \{\omega : \tau(\omega) \le t\} \setminus (A \cap \{\omega : \tau(\omega) \le t\}).$$

The first event $\{\omega : \tau(\omega) \le t\} \in \mathcal{F}_t$, because τ is Markov time. The second set $(A \cap \{\omega : \tau(\omega) \le t\}) \in \mathcal{F}_t$, because the set $A \in \mathcal{F}_\tau$.

We see that all the axioms are held. That is why $\mathcal{F}_\tau$ is indeed the σ-algebra. ■

The σ-algebra $\mathcal{F}_t$ can be interpreted as the totality of events related to some physical process and observed before time t. Therefore one can interpret the σ-algebra $\mathcal{F}_\tau$ as the totality of events that can be observed over the random time τ.

Definition 1.1.3 *The system of σ-algebras $F = \{\mathcal{F}_t\}, t \in T$, is said to be a right continuous system if for each $t \in T$*

$$\mathcal{F}_t = \mathcal{F}_{t+},$$

where $\mathcal{F}_{t+} = \cap_{s>t} \mathcal{F}_s$

Lemma 1.2 *Let τ be a Markov time. Then the events $\{\tau < t\}$ and $\{\tau = t\}$ belong to $\mathcal{F}_t$ for each $t \in T$.*

Proof.

Event $\{\tau < t\}$ can be written as:

$$\{\tau < t\} = \bigcup_{k=1}^{\infty} \left(\tau \le t - \frac{1}{k}\right),$$

where each of the events

$$\left(\tau \le t - \frac{1}{k}\right) \in \mathcal{F}_{t-1/k} \subseteq \mathcal{F}_t.$$

Thus, $\{\tau < t\} \in \mathcal{F}_t$, as a union of events from $\mathcal{F}_t$. The same reasoning applies to $\{\tau = t\} = \{\tau \le t\} \setminus \{\tau < t\} \in \mathcal{F}_t$. ■

Lemma 1.3 *If τ and ρ are Markov times, then*

(a) $\tau \wedge \rho = \min(\tau, \rho)$,

(b) $\tau \vee \rho = \max(\tau, \rho)$,

(c) $(\tau + \rho)$

are also Markov times.

Proof. If τ and ρ are Markov times, then it follows that $\{\tau \leq t\} \in \mathcal{F}_t$ and $\{\rho \leq t\} \in \mathcal{F}_t$, then the following properties can be received:

(a)
$$\{\tau \wedge \rho \leq t\} = \{\tau \leq t\} \cup \{\rho \leq t\} \in \mathcal{F}_t,$$

(b)
$$\{\tau \vee \rho \leq t\} = \{\tau \leq t\} \cap \{\rho \leq t\} \in \mathcal{F}_t,$$

(c) To prove that $(\tau + \rho) \in \mathcal{F}_t$, we start from the complement of this event:

$$\begin{aligned}\{\tau + \rho > t\} &= \{\tau = 0, \tau + \rho > t\} \cup \{0 < \tau < t, \tau + \rho > t\} \cup \{\tau \geq t, \tau + \rho > t\} \\ &= \{\tau = 0\} \cap \{\rho > t\} + \{0 < \tau < t, \tau + \rho > t\} + \{\tau > t\} \cap \{\rho = 0\} \\ &+ \{\tau \geq t\} \cap \{\rho > 0\},\end{aligned}$$

where
$$\begin{gathered}\{\tau = 0\} \cap \{\rho > t\} \in \mathcal{F}_t, \\ \{\tau > t\} \cap \{\rho = 0\} \in \mathcal{F}_t, \\ \{\tau \geq t\} \cap \{\rho > 0\} = \{\tau \geq t\} \cap \{\rho = 0\}^c \in \mathcal{F}_t, \\ \{0 < \tau < t, \tau + \rho > t\} = \bigcup_{r \in (0,1) \cap \mathcal{Q}} (\{r < \tau < t\} \cap \{\rho > t - r\}) \in \mathcal{F}_t.\end{gathered}$$

Consequently, we have shown that the complement of the set $\{\tau + \rho \leq t\}$ is in $\mathcal{F}_t$, therefore the original set is also in $\mathcal{F}_t$.

■

1.2 Conditional Expectations

Conditional expectations and their properties are used extensively throughout the thesis. Hence, in this section we collect all the facts that we will need further. The theory of conditional expectations can be found in most advanced probability books, but we are following the ideas of Shreve (2004) to give basic definitions and properties.

Let $(\Omega, \mathcal{F}, P)$ be the probability space, and X is the random variable defined on this space. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. If random variable X is independent of $\mathcal{G}$, we cannot say anything about its value. If X is measurable with respect to $\mathcal{G}$, we can determine its value. In the intermediate case we can only estimate the value of X based on the information available at $\mathcal{G}$. Such an estimate is the conditional expectation of X given $\mathcal{G}$.

Definition 1.2.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ and let X be a random variable that is either nonnegative or integrable. The conditional expectation of X given $\mathcal{G}$, denoted $\mathbb{E}\{X \mid \mathcal{G}\}$, is any random variable that satisfies*

1. *Measurability*

$$\mathbb{E}\{X \mid \mathcal{G}\} \text{ is } \mathcal{G} - \text{measurable},$$

2. *Partial averaging*

$$\int_A \mathbb{E}\{X \mid \mathcal{G}\}(\omega) d\mathcal{P}(\omega) = \int_A X(\omega) d\mathcal{P}(\omega) \text{ for all } A \in \mathcal{G}$$

The partial averaging property can be rewritten in the following way:

$$\mathbb{E}\{\mathbb{E}\{X \mid \mathcal{G}\}\, I_A\} = \mathbb{E}\{X\, I_A\} \tag{1.1}$$

Note that $\mathbb{E}\{X \mid \mathcal{G}\}$ is random variable. It can be seen from the first property that the value of this variable can be determined from the information that is contained in the σ-algebra $\mathcal{G}$. The second property tell us that the conditional expectation $\mathbb{E}\{X \mid \mathcal{G}\}$ is indeed an estimate of the random variable X, because it gives the same average as X over all subsets of $\mathcal{G}$.

The conditional expectation exists, and it is unique. This fact is nontrivial, and its proof can be found for example in Klenke (2008).

Theorem 1.4 *Let $(\Omega, \mathcal{F}, P)$ be the probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.*

(a) **(Linearity of conditional expectations)** *If X and Y are integrable random variables and c_1 and c_2 are constants, then*

$$\mathbb{E}\{c_1 X + c_2 Y \mid \mathcal{G}\} = c_1 \mathbb{E}\{X \mid \mathcal{G}\} + c_2 \mathbb{E}\{Y \mid \mathcal{G}\}. \tag{1.2}$$

This equality holds if we assume that X and Y are nonnegative (rather than integrable) and c_1 and c_2 are positive, although both sides may be $+\infty$.

(b) **(Taking out what is known)** *If X and Y are integrable random variables, Y and XY are integrable, and X is $\mathcal{G}$-measurable, then*

$$\mathbb{E}\{XY \mid \mathcal{G}\} = X \mathbb{E}\{Y \mid \mathcal{G}\}. \tag{1.3}$$

This equality holds if we assume that X is positive and Y is nonnegative (rather than integrable), although both sides may be $+\infty$.

(c) **(Iterated conditioning)** *If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$ and X is an integrable random variable, then*

$$\mathbb{E}\{\mathbb{E}\{X \mid \mathcal{G}\} \mid \mathcal{H}\} = \mathbb{E}\{X \mid \mathcal{H}\}. \tag{1.4}$$

This equality holds if we assume that X is nonnegative (rather than integrable), although both sides may be $+\infty$.

(d) **(Independence)** *If X is integrable and independent of $\mathcal{G}$, then*

$$\mathbb{E}\{X \mid \mathcal{G}\} = \mathbb{E}\{X\}. \tag{1.5}$$

This equality holds if we assume that X is nonnegative (rather than integrable), although both sides may be $+\infty$.

The proof of this theorem is a standard machinery, and it can be found, for example in Klenke (2008).

Chapter 2

Pricing of American Put Options

2.1 The problem

We want to find the price of American put option. An option is a contract on an underlying security which allows its holder to trade in a fixed number of shares of a specified stock at an agreed amount of money at any time on or before the specified day. The act of making this transaction is known as exercising the option. The last day when the option can be traded is called expiration date or maturity date. The fixed price at which the option is traded is called the strike price. A put option is such an option that gives its holder the right to sell the asset. American option can be exercised at any time up to maturity date. It is well-known that in complete and arbitrage-free markets the price of the option can be expressed as the expected value of the payoff with respect to the equivalent martingale measure, see for instance Shreve (2004). So that, let us formalize the problem of option pricing.

Let

- $X_0, X_1, \ldots, X_T$ be a $\mathbb{R}$ -valued Markov process containing the information about the prices of the underlying asset,
- $\mathcal{F}_t$ be the σ-algebra generated by asset prices,
- K be the strike price,
- r be the discount rate,
- T be the maturity date,
- $\mathcal{T}[0, \ldots, T]$ be the set of all possible stopping times with values in $[0, T]$,
- $\tau \in \mathcal{T}[0, \ldots, T]$ be a stopping time.

The payoff function f_t of American put option , with discount factor e^{-rt} is

$$f_t(X_t) = e^{-rt}(K - X_t)^+.$$

The task of pricing the option is to determine the value

$$V_0 = \sup_{\tau \in \mathcal{T}[0, \ldots, T]} \mathbb{E}\{f_\tau(X_\tau)\}.$$

The computation of this value can be done by choosing the optimal stopping time $\tau^* \in \mathcal{T}[0,\ldots,T]$ satisfying

$$V_0 = \sup_{\tau \in \mathcal{T}[0,\ldots,T]} \mathbb{E}\{f_\tau(X_\tau)\} = \mathbb{E}\{f_{\tau^*}(X_{\tau^*})\}.$$

Unfortunately, the future prices of the asset are not known in advance. Therefore, the price can not be calculated straightforwardly and so analytical solution can not be found for this problem. However, what we can do, is to use the simulation methods together with advanced regression techniques to get the approximation of the solution.

First, it is necessary to reformulate this problem in a more convenient form to work with.

For $0 \le t < T$ let

$$q_t(x) = \sup_{\tau \in \mathcal{T}[t+1,\ldots,T]} \mathbb{E}\{f_\tau(X_\tau) \mid X_t = x\} \tag{2.1}$$

be the so-called continuation value. This is the payoff the holder will get if he continues to keep the option alive knowing that at time t the price of the asset is x.

For $t = T$ we define the corresponding continuation value as

$$q_T(x) = 0,$$

because T is the expiration date of the option, and we will not get any money if we sell it after this time.

As we will se further, the optimal stopping time has the following form:

$$\tau^* = \inf\{s \in \{0,\ldots,T\} : q_s(X_s) \le f_s(X_s)\}. \tag{2.2}$$

The right-hand side of this equation is well defined, since $q_T(x) = 0$ and $f_T(x) \ge 0$. There will always be some moment of time s for which $q_s(X_s) \le f_s(X_s)$ holds.

We should understand the optimal stopping time in the following way. The holder of the option optimally compares the payoff from continuation with payoff from immediately exercising the option, and he exercises the option if the immediate payoff is higher.

2.2 Optimal Stopping Time

This chapter contains the proof that τ^* defined in (2.2) is indeed the optimal stopping time, when the holder will get the maximal payoff if he stops the price process. Moreover, the theorem shows the important role of the continuation values. We will see that it is enough to find them in order to determine the optimal stopping time. Further the techniques from the general theory of optimal stopping from Shiryaev (2007) and Chow et al. (1971) will be used.

Theorem 2.1 *Kohler (2010) Let $\mathcal{T}(t, t+1, \ldots, T)$ be the subset of $\mathcal{T}(0,\ldots,T)$ consisting of all stopping times which take on values only in $\{t, t+1, \ldots, T\}$ and let*

$$V_t(x) = \sup_{\tau \in \mathcal{T}(t,t+1,\ldots,T)} \mathbb{E}\{f_\tau(X_\tau) | X_t = x\} \tag{2.3}$$

be the value function that describes the expected value we get if we sell the option in the optimal way after time $t-1$ given that $X_t = x$.

For $t \in \{-1, 0, \dots, T-1\}$ define

$$\tau_t^* = \inf\{s \geq t+1 : q_s(X_s) \leq f_s(X_s)\} \tag{2.4}$$

Under the above assumptions we have that for any $t \in \{0, \dots, T-1\}$ and all $x \in \mathbb{R}^d$:

$$V_t(x) = \mathbb{E}\{f_{\tau_{t-1}^*}(X_{\tau_{t-1}^*})|X_t = x\}. \tag{2.5}$$

and

$$V_0 = \mathbb{E}\{f_{\tau^*}(X_{\tau^*})\}. \tag{2.6}$$

In this theorem we want to show that

$$\sup_{\tau \in \mathcal{T}(t,t+1,\dots,T)} \mathbb{E}\{f_\tau(X_\tau)|X_t = x\} = \mathbb{E}\{f_{\tau_{t-1}^*}(X_{\tau_{t-1}^*})|X_t = x\}.$$

This formula shows that the maximal expected payoff will be reached at time $\tau_{t-1}^* \in \mathcal{T}(t, t+1, \dots, T)$, at the first moment of time when our expected payoff will be greater than the payoff from the continuation. To prove this formula, we will use the next logic:

$$\text{if} \quad a = c \quad \text{and} \quad b = c \quad \Rightarrow \quad a = b,$$

where

$$a = \sup_{\tau \in \mathcal{T}(t,t+1,\dots,T)} \mathbb{E}\{f_\tau(X_\tau)|X_t = x\},$$
$$b = \mathbb{E}\{f_{\tau_{t-1}^*}(X_{\tau_{t-1}^*})|X_t = x\},$$
$$c = \max\{f_t(X_t), q_t(X_t)\}.$$

The whole proof goes by induction on time t. In the proof we use the definition of the stopping time, the fact the the asset price process (X_t) is Markov process and such properties of conditional expectations as "taking out what is known" and iterated conditioning. **Proof.**

- **First step**. In this step we want to show that for $t = T$ the formula 2.5 holds. From (2.4) we get:

$$\tau_{T-1}^* = \{\inf s \geq T : q_s(X_s) \leq f_s(X_s)\} = T. \tag{2.7}$$

  We take infinum over all possible stopping times, but in this case there is only one possible choice left, which is T, otherwise we will get 0 for sure.

$$\tau \in \mathcal{T}(t, \dots, T) = \mathcal{T}(T) \Rightarrow \tau = T.$$

  Using (2.3) in the first equality, the fact that $t = T$, and (2.7) in the last equality, we get:

$$V_T(x) = \sup_{\tau \in \mathcal{T}(T)} \mathbb{E}\{f_\tau(X_\tau) \mid X_T = x\}$$
$$= \mathbb{E}\{f_T(X_T) \mid X_T = x\}$$
$$= \mathbb{E}\{f_{\tau_{T-1}^*}(X_{\tau_{T-1}^*})|X_T = x\}.$$

Next we assume $V_s(x) = \mathbb{E}\{f_{\tau^*_{s-1}}(X_{\tau^*_{s-1}})|X_s = x\}$ holds for any $t < s \leq T$ and $t \in 0, \ldots, T$. In order to prove that (2.5) holds for any t, we have to make several additional steps.

- **Second step**. Here we have to show that

$$\mathbb{E}\{f_\tau(X_\tau)\} \leq \max\{f_t(X_t), \mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t\}\}$$

holds for arbitrary $\tau \in \mathcal{T}(t, \ldots, T)$. Observe that

$$\begin{aligned} f_\tau(X_\tau) &= f_\tau(X_\tau) \cdot I_{\{\tau=t\}} + f_\tau(X_\tau) \cdot I_{\{\tau>t\}} \\ &= f_t(X_t) \cdot I_{\{\tau=t\}} + f_{\max\{\tau,t+1\}}(X_{\max\{\tau,t+1\}}) \cdot I_{\{\tau>t\}}. \end{aligned} \tag{2.8}$$

Using the representation (2.8) and the fact that $(X_t)_{0 \leq t \leq T}$ is a Markov process we have:

$$\begin{aligned} &\mathbb{E}\{f_\tau(X_\tau) \mid X_t\} = \\ &= \mathbb{E}\{f_t(X_t) \cdot I_{\{\tau=t\}} \mid X_t\} + \mathbb{E}\{f_{\max\{\tau,t+1\}}(X_{\max\{\tau,t+1\}}) \cdot I_{\{\tau>t\}} \mid X_t\} \\ &= \mathbb{E}\{f_t(X_t) \cdot I_{\{\tau=t\}} \mid X_0, \ldots, X_t\} \\ &+ \mathbb{E}\{f_{\max\{\tau,t+1\}}(X_{\max\{\tau,t+1\}}) \cdot I_{\{\tau>t\}} \mid X_0, \ldots, X_t\}. \end{aligned} \tag{2.9}$$

Since τ is a stopping time, then $\{\tau = t\}$ and $\{\tau > t\} = \Omega \setminus \{\tau \leq t\}$ are measurable with respect to $X_0, \ldots, X_t$. It follows that $I_{\{\tau=t\}}$ and $I_{\{\tau>t\}}$ are also measurable with respect to $X_0, \ldots, X_t$. Consequently, we can take the measurable variables out of the conditional expectation:

$$\mathbb{E}\{f_t(X_t) \cdot I_{\{\tau=t\}} \mid X_0, \ldots, X_t\} = f_t(X_t) \cdot I_{\{\tau=t\}}. \tag{2.10}$$

The same arguments work for the second expectation:

$$\begin{aligned} &\mathbb{E}\{f_{\{\max \tau,t+1\}}(X_{\{\max \tau,t+1\}}) \cdot I_{\{\tau>t\}} \mid X_0, \ldots, X_t\} \\ &= I_{\{\tau>t\}}\mathbb{E}\{f_{\{\max \tau,t+1\}}(X_{\{\max \tau,t+1\}}) \mid X_0, \ldots, X_t\} \\ &= I_{\{\tau>t\}}\mathbb{E}\{f_{\{\max \tau,t+1\}}(X_{\{\max \tau,t+1\}}) \mid X_t\} \\ &= I_{\{\tau>t\}}\mathbb{E}\{\mathbb{E}\{f_{\{\max \tau,t+1\}}(X_{\{\max \tau,t+1\}}) \mid X_{t+1}\} \mid X_t\}, \end{aligned} \tag{2.11}$$

where in the second step we used Markov property, and in the third one the iterated conditioning property.

As $\max\{\tau, t+1\} \in \mathcal{T}(t+1, \ldots, T)$, the value

$$\mathbb{E}\{f_{\{\max\{\tau,t+1\}\}}(X_{\{\max\{\tau,t+1\}\}}) \mid X_{t+1}\} \leq V_{t+1}(X_{t+1})$$

because V_{t+1} is supremum over all possible stopping times by its definition in (2.3).

Hence from (2.11) we get:

$$I_{\{\tau>t\}}\mathbb{E}\{\mathbb{E}\{f_{\{\max \tau,t+1\}}(X_{\{\max \tau,t+1\}}) \mid X_{t+1}\} \mid X_t\} \leq I_{\{\tau>t\}}\mathbb{E}\{V_{t+1}(X_{t+1})|X_t\}. \tag{2.12}$$

Collecting the results we got in 2.10 and in 2.12, we can rewrite (2.9) in the following way:

$$\begin{aligned}\mathbb{E}\{f_\tau(X_\tau) \mid X_t\} &\le f_t(X_t)\cdot I_{\{\tau=t\}} + I_{\{\tau>t\}}\mathbb{E}\{V_{t+1}(X_{t+1})|X_t\} \\ &= \begin{cases} 0, & \text{if } \{\tau<t\} \\ f_t(X_t), & \text{if } \{\tau=t\} \\ \mathbb{E}\{V_{t+1}(X_{t+1})|X_t\}, & \text{if } \{\tau>t\} \end{cases} \\ &\le \max\{f_t(X_t), \mathbb{E}\{V_{t+1}(X_{t+1})|X_t\}\}. \end{aligned} \tag{2.13}$$

- **Third step.** In this step we want to receive the representation of the value function at time $\tau=\tau^*_{t-1}$. Substituting $\tau=\tau^*_{t-1}$ in 2.9 we have:

$$\begin{aligned}\mathbb{E}\{f_{\tau^*_{t-1}}(X_{\tau^*_{t-1}}) \mid X_t\} &= \mathbb{E}\{f_t(X_t)\cdot I_{\{\tau^*_{t-1}=t\}} \mid X_t\} \\ &+ \mathbb{E}\{f_{\max\{\tau^*_{t-1},t+1\}}(X_{\max\{\tau^*_{t-1},t+1\}})\cdot I_{\{\tau^*_{t-1}>t\}} \mid X_t\}. \end{aligned}\tag{2.14}$$

Using Markov property for $(X_t)_{0\le t\le T}$, we can continue (2.14) in the following way:

$$\begin{aligned}&\mathbb{E}\{f_{\tau^*_{t-1}}(X_{\tau^*_{t-1}}) \mid X_t\} = \mathbb{E}\{f_t(X_t)\cdot I_{\{\tau^*_{t-1}=t\}} \mid X_0,\dots,X_t\} \\ &+ \mathbb{E}\{f_{\max\{\tau^*_{t-1},t+1\}}(X_{\max\{\tau^*_{t-1},t+1\}})\cdot I_{\{\tau^*_{t-1}>t\}} \mid X_0,\dots,X_t\} \\ &= f_t(X_t)\cdot I_{\{\tau^*_{t-1}=t\}} + I_{\{\tau^*_{t-1}>t\}}\mathbb{E}\{f_{\max\{\tau^*_{t-1},t+1\}}(X_{\max\{\tau^*_{t-1},t+1\}}) \mid X_t\}. \end{aligned}\tag{2.15}$$

In the last step of (2.15) we have used that $I_{\{\tau^*_{t-1}=t\}}$ and $I_{\{\tau^*_{t-1}>t\}}$ are measurable with respect to $X_0,\dots,X_t$.

If $\{\tau^*_{t-1}>t\}$, that means that we did not stop the process $(X_t)_{0\le t\le T}$ before and at time t. Hence, the possible stopping time will start from $t+1$. And in (2.4) we have denoted the optimal stopping time as τ^*_t. Therefore, $\max\{\tau^*_{t-1},t+1\}=\tau^*_t$ on the set $\{\tau^*_{t-1}>t\}$. We can rewrite (2.15) in the next way:

$$\begin{aligned}&\mathbb{E}\{f_{\tau^*_{t-1}}(X_{\tau^*_{t-1}}) \mid X_t\} = \\ &= f_t(X_t)\cdot I_{\{\tau^*_{t-1}=t\}} + I_{\{\tau^*_{t-1}>t\}}\mathbb{E}\{f_{\tau^*_t}(X_{\tau^*_t}) \mid X_t\} \\ &= f_t(X_t)\cdot I_{\{\tau^*_{t-1}=t\}} + I_{\{\tau^*_{t-1}>t\}}\mathbb{E}\{\mathbb{E}\{f_{\tau^*_t}(X_{\tau^*_t}) \mid X_{t+1}\} \mid X_t\} \\ &= f_t(X_t)\cdot I_{\{\tau^*_{t-1}=t\}} + I_{\{\tau^*_{t-1}>t\}}\mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t\}. \end{aligned}\tag{2.16}$$

- **Fourth step.** Here we want to show that

$$\mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t\} = q_t(X_t). \tag{2.17}$$

Since $\tau^*_t \in \mathcal{T}(t+1,\dots,T)$

$$\begin{aligned}\mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t\} &= \mathbb{E}\{\mathbb{E}\{f_{\tau^*_t}(X_{\tau^*_t}) \mid X_{t+1}\} \mid X_t\} \\ &= \mathbb{E}\{f_{\tau^*_t}(X_{\tau^*_t}) \mid X_t\} \\ &\le \sup_{\tau\in\mathcal{T}(t+1,\dots,T)} \mathbb{E}\{f_\tau(X_\tau) \mid X_t\} = q_t(X_t). \end{aligned}\tag{2.18}$$

On the other hand, we have

$$\begin{aligned}\mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t\} &= \mathbb{E}\{\sup_{\tau\in\mathcal{T}(t+1,\dots,T)} \mathbb{E}\{f_\tau(X_\tau) \mid X_{t+1}\} \mid X_t\} \\ &\geq \sup_{\tau\in\mathcal{T}(t+1,\dots,T)} \mathbb{E}\{\mathbb{E}\{f_\tau(X_\tau) \mid X_{t+1}\} \mid X_t\} \\ &= \sup_{\tau\in\mathcal{T}(t+1,\dots,T)} \mathbb{E}\{f_\tau(X_\tau) \mid X_t\} = q_t(X_t). \end{aligned} \tag{2.19}$$

Combining (2.18) and (2.19) together, we get (2.17). Using this proved fact, we can rewrite (2.16) in the following way:

$$\begin{aligned}&\mathbb{E}\{f_{\tau^*_{t-1}}(X_{\tau^*_{t-1}}) \mid X_t\} = \\ &= f_t(X_t)\cdot I_{\{\tau^*_{t-1}=t\}} + I_{\{\tau^*_{t-1}>t\}}\mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t\} \\ &= f_t(X_t)\cdot I_{\{\tau^*_{t-1}=t\}} + I_{\{\tau^*_{t-1}>t\}}q_t(X_t). \end{aligned} \tag{2.20}$$

We know that τ^*_{t-1} is a stopping time. By equivalent formulation, that is the first time t when $f_t(X_t) \geq q_t(X_t)$. For all other $t < \tau^*_{t-1}$ we have $f_t(X_t) < q_t(X_t)$. Therefore, from (2.20) we have:

$$\begin{aligned}&\mathbb{E}\{f_{\tau^*_{t-1}}(X_{\tau^*_{t-1}}) \mid X_t\} = \\ &= f_t(X_t)\cdot I_{\{\tau^*_{t-1}=t\}} + I_{\{\tau^*_{t-1}>t\}}q_t(X_t) \\ &= \max\{f_t(X_t), q_t(X_t)\}. \end{aligned} \tag{2.21}$$

- **Fifth step.** From (2.13),(2.17) and (2.21) we get:

$$\begin{aligned}\mathbb{E}\{f_\tau(X_\tau) \mid X_t = x\} &\leq \max\{f_t(X_t), \mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t = x\}\} \\ &= \max\{f_t(X_t), q_t(X_t)\} \\ &= \mathbb{E}\{f_{\tau^*_{t-1}}(X_{\tau^*_{t-1}}) \mid X_t = x\}, \end{aligned} \tag{2.22}$$

where $\tau \in \mathcal{T}(t,\dots,T)$ and $\tau^*_{t-1} \in \mathcal{T}(t,\dots,T)$. And finally we can conclude that

$$V_t(x) = \sup_{\tau\in\mathcal{T}(t,\dots,T)} \mathbb{E}\{f_\tau(X_\tau) \mid X_t = x\} = \mathbb{E}\{f_{\tau^*_{t-1}}(X_{\tau^*_{t-1}}) \mid X_t = x\}. \tag{2.23}$$

In order to show (2.6) we use previous results and receive:

$$\begin{aligned}V_0 &= \sup_{\tau\in\mathcal{T}(0,\dots,T)} \mathbb{E}\{f_\tau(X_\tau)\} \\ &= \sup_{\tau\in\mathcal{T}(0,\dots,T)} \mathbb{E}\{f_0(X_0)I_{\{\tau=0\}} + f_{\max\{\tau,1\}}(X_{\max\{\tau,1\}})I_{\{\tau>0\}}\} \\ &= \mathbb{E}\{f_0(X_0)I_{\{f_0(X_0)\geq q_0(X_0)\}} + f_{\tau^*_0}(X_{\tau^*_0})I_{\{f_0(X_0)<q_0(X_0)\}}\} \\ &= \mathbb{E}\{f_0(X_0)I_{\{f_0(X_0)\geq q_0(X_0)\}} + \mathbb{E}\{V_1(X_1) \mid X_0\}I_{\{f_0(X_0)<q_0(X_0)\}}\} \\ &= \mathbb{E}\{f_0(X_0)I_{\{f_0(X_0)\geq q_0(X_0)\}} + q_0(X_0)I_{\{f_0(X_0)<q_0(X_0)\}}\} \\ &= \mathbb{E}\{\max\{f_0(X_0), q_0(X_0)\}\} \\ &= \mathbb{E}\{f^*_\tau(X^*_\tau)\}. \end{aligned}$$

■

We have received that in order to find the price of the option, we have to determine the optimal stopping time, for which task we need to find the continuation value function.

The theorem also showed us that the continuation values and the values of the value function are closely related. We have proved in (2.17) and (2.21):

$$q_t(x) = \mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t = x\}$$

and

$$V_t(x) = \max\{f_t(x), q_t(x)\}.$$

This remark shows that $q_t(X_t) \leq f_t(X_t)$ is equivalent to $V_t(X_t) \leq f_t(X_t)$. Therefore, the optimal stopping time can be expressed as

$$\tau^* = \inf\{s \in \{0, \ldots, T\} : V_s(X_s) \leq f_s(X_s)\} \tag{2.24}$$

Though it can be expressed using V_t, we will work with q_t functions. For explaining the reasons of this decision, we have to prove some more facts.

Theorem 2.2 *Kohler (2010) Under the above assumptions, we have that for any $t \in \{0, \ldots, T-1\}$ and all $x \in \mathbb{R}^d$:*

(a)

$$q_t(x) = \mathbb{E}\{f_{\tau^*}(X_{\tau^*}) \mid X_t = x\}$$

(b)

$$q_t(x) = \mathbb{E}\{\max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1}) \mid X_t = x\}$$

Proof.

(a)

$$\begin{aligned} q_t(X_t) &= \mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t\} \quad \text{(from (2.17))} \\ &= \mathbb{E}\{\mathbb{E}\{f_{\tau^*}(X_{\tau^*}) \mid X_{t+1}\} \mid X_t\} \quad \text{(from (2.21))} \\ &= \mathbb{E}\{\mathbb{E}\{f_{\tau^*}(X_{\tau^*}) \mid X_0, \ldots, X_{t+1}\} \mid X_0, \ldots, X_t\} \quad \text{(Markov property)} \\ &= \mathbb{E}\{f_{\tau^*}(X_{\tau^*}) \mid X_0, \ldots, X_t\} \quad \text{(Iterated conditioning)} \\ &= \mathbb{E}\{f_{\tau^*}(X_{\tau^*}) \mid X_t\} \quad \text{(Markov property).} \end{aligned}$$

(b) We can receive $V_{t+1}(X_{t+1})$ from (2.21):

$$V_{t+1}(X_{t+1}) = \max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\}.$$

From (2.17) we have:

$$q_t(x) = \mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t = x\} = \mathbb{E}\{\max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} \mid X_t = x\}.$$

■

Now we can compare the regression representations for the continuation values

$$q_t(x) = \mathbb{E}\{\max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} \mid X_t = x\}$$

with the regression representation for the value function

$$V_t(X_t) = \max\{f_t(X_t), \mathbb{E}\{V_{t+1}(X_{t+1}) \mid X_t = x\}\}.$$

In the last expression the maximum is outside of the expectation, therefore the value function can be in general not differentiable. Comparing the representation for the continuation value, we see that the maximum is inside of the expectation, so the function is smooth. Therefore, we will concentrate on the continuation value function because it is easier to get regression estimates for smooth functions.

2.3 Backward recursive scheme

In the previous section the final formula for calculation of the continuation value function was received.

$$q_t(x) = \mathbb{E}\{\max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} \mid X_t = x\}.$$

Substituting the payoff function for American Put option and doing several transformations, we receive:

$$\begin{aligned}
q_t(x) &= \mathbb{E}\{\max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} \mid X_t = x\} \\
&= \mathbb{E}\{\max\{e^{-r(t+1)}(K - X_{t+1})^+, q_{t+1}(X_{t+1})\} \mid X_t = x\} \\
&= \mathbb{E}\left\{\max\left\{e^{-r(t+1)}\left(K - \frac{X_{t+1}}{X_t}X_t\right)^+, q_{t+1}\left(\frac{X_{t+1}}{X_t}X_t\right)\right\} \mid X_t = x\right\} \\
&= \mathbb{E}\left\{\max\left\{e^{-r(t+1)}\left(K - \frac{X_{t+1}}{X_t}x\right)^+, q_{t+1}\left(\frac{X_{t+1}}{X_t}x\right)\right\} \mid X_t = x\right\}. \qquad (2.25)
\end{aligned}$$

Therefore we see, that there is a backward recursive scheme that can be used to calculate the values of the continuation function. We start with $q_T(x) = 0$, and using the formula (2.25) we can find all the previous values.

2.4 Upper and lower bounds of the continuation function

As we already know that the analytical solution does not exist, we need to estimate continuation values and receive the estimated price of the American Put option. We can do the estimation using the backward recursive scheme, but at first we have to answer the following question: if we have an approximate solution, how far is it from the true one? Using the backward recursion, the estimation errors will be accumulated, therefore, we need to control them. For the evaluation of the approximate solution we introduce upper and lower bounds for $q_t(x)$ as in Györfi et al. (2012).

First we introduce an upper bound. For $\tau \in \mathcal{T}(t+1,\dots,T)$, we have that

$$f_\tau(X_\tau) \le \max_{s\in\{t+1,\dots,T\}} f_s(X_s).$$

This upper bound is rather trivial. We assume that we know all the data in advance, and we just choose that moment of time when the payoff will be maximal.

Therefore, using the definition of $q_t(x)$ given in (2.1), we have

$$q_t(x) = \sup_{\tau\in\mathcal{T}(t+1,\dots,T)} \mathbb{E}\left\{f_\tau(X_\tau) \mid X_t = x\right\} \le \mathbb{E}\left\{\max_{s\in\{t+1,\dots,T\}} f_s(X_s) \mid X_t = x\right\}.$$

We introduce the notation

$$q_t^{(u)}(x) := \mathbb{E}\left\{\max_{s\in\{t+1,\dots,T\}} f_s(X_s) \mid X_t = x\right\}, \tag{2.26}$$

and this value will be an upper bound.

Now we introduce a lower bound of $q_t(x)$:

$$q_t^{(l)}(x) = \max_{s\in\{t+1,\dots,T\}} \mathbb{E}\left\{f_s(X_s) | X_t = x\right\}. \tag{2.27}$$

In the last equation, maximum is taken over deterministic stopping times $\{t+1,\dots,T\}$, which is the subset of $\mathcal{T}(t+1,\dots,T)$. That is why it is a lower bound

$$q_t^{(l)}(x) \le q_t(x).$$

To summarize, we have received the next combination of the lower an upper bounds:

$$\max_{s\in\{t+1,\dots,T\}} \mathbb{E}\left\{f_s(X_s) | X_t = x\right\} \le q_t(x) \le \mathbb{E}\left\{\max_{s\in\{t+1,\dots,T\}} f_s(X_s) \mid X_t = x\right\}. \tag{2.28}$$

2.5 Empirical pricing and optimal exercising of American options

In a real life problem we have a single historical data sequence $X_1,\dots,X_N$. We have to generate sample paths in order to be able to perform regression estimation. We need i.i.d copies of the historical data sequence, i.e.:

$$X_{i,1},\dots,X_{i,T}, \quad i = 1,...n. \tag{2.29}$$

There exist several ways to receive sample paths, for example, the Monte Carlo method, disjoint, sliding or bootstrap sampling. In Györfi et al. (2012) a comparison of the behaviour of these sampling schemes was made. The range of the option price is smaller for sampling schemes mentioned above than for the Monte Carlo method. The behaviour of the sliding sample is almost the same compared to disjoint and bootstrap, but it is easier in calculations. Therefore, later on we will use sliding sample scheme.

For sliding sampling the sample path can be found as

$$X_{i,t} := \frac{X_{i+t}}{X_i}, \tag{2.30}$$

$i = 1, \ldots, n = N - T$.

If the continuation values $q_t(x)$, $t = 1, \ldots T$ were known, then the optimal stopping time τ_i for path $X_{i,1}, \ldots, X_{i,T}$ can be calculated:

$$\tau_i = \min \left\{ 1 \leq s \leq T : q_s\left(X_{i,s}\right) \leq f_s\left(X_{i,s}\right) \right\}.$$

Therefore, the price V_0 can be estimated by the average of the payoff received at stopping times for each path:

$$\frac{1}{n} \sum_{i=1}^{n} f_{\tau_i}\left(X_{\tau_i}\right). \tag{2.31}$$

Unfortunately, the continuation values $q_t(x)$, $t = 1, \ldots T$ are unknown, therefore, one has to generate estimates $q_{t,n}(x)$, $t = 1, \ldots T$.

As the continuation values are regression functions, we are going to apply nonparametric methods of estimation to receive the result. The empirical methods use only the available data (known prices from the past) and do not assume parameter estimation.

Chapter 3

Nonparametric regression estimation

This chapter contains the introduction to general regression problem (section 3.1). In further sections it is shown how to estimate the regression function empirically, i.e. from the data we observe. The regression problem and its solution is defined as in Györfi et al. (2002).

3.1 The regression problem

Let us consider a random vector (X, Y), where X is $\mathbb{R}^d$-valued and Y is $\mathbb{R}$-valued. We are interested how the values of the response variable Y depend on the value of the observation vector X. This means that we have to find a function $f : \mathbb{R}^d \to \mathbb{R}$, such that $f(X)$ is a good approximation of Y, $f(X)$ should be close to Y. This implies making $|f(X) - Y|$ as small as possible. This value is random because X and Y are random vectors. Hence, we require the mean squared error of f

$$\mathbb{E}\,|f(X) - Y|^2$$

to be as small as possible.

Definition 3.1.1 *A function $m^* : \mathbb{R}^d \to \mathbb{R}$ that minimizes the L_2 risk, or*

$$\mathbb{E}\,|m^*(X) - Y|^2 = \min_{f:\mathbb{R}^d \to \mathbb{R}} \mathbb{E}\,|f(X) - Y|^2$$

is called the regression function.

The regression function can be obtained explicitly:

$$m(x) = \mathbb{E}\{Y \mid X = x\}.$$

Let us show that it indeed minimizes the L_2 risk.

For an arbitrary $f : \mathbb{R}^d \to \mathbb{R}$ the mean squared error can be rewritten in the following way:

$$\begin{aligned}\mathbb{E}\,|f(X) - Y|^2 &= \mathbb{E}\,|f(X) - m(X) + m(X) - Y|^2 \\ &= \mathbb{E}\,|f(X) - m(X)|^2 + \mathbb{E}\,|m(X) - Y|^2 \\ &+ \mathbb{E}\{(f(X) - m(X))(m(X) - Y)\}.\end{aligned}$$

The last term of the sum is equal to zero because:

$$\begin{aligned}
&\mathbb{E}\{(f(X)-m(X))(m(X)-Y)\}\\
&=\mathbb{E}\{\mathbb{E}\{(f(X)-m(X))(m(X)-Y)\mid X\}\}\text{(iterated conditioning property)}\\
&=\mathbb{E}\{(f(X)-m(X))\mathbb{E}\{(m(X)-Y)\mid X\}\}\text{ (the term } f(X)-m(X) \text{ is } X \text{ measurable)}\\
&=\mathbb{E}\{(f(X)-m(X))(m(X)-m(X))\}\\
&=0.
\end{aligned}$$

Consequently,

$$\mathbb{E}\,|f(X)-Y|^2=\int\limits_{\mathbb{R}^d}|f(x)-m(x)|^2\,\mu(dx)+\mathbb{E}\,|m(X)-Y|^2\,, \tag{3.1}$$

where μ denotes the distribution of X.

The term $\int\limits_{\mathbb{R}^d}|f(x)-m(x)|^2\,\mu(dx)$ is called the L_2 error of function f. It is always nonnegative as integral of nonnegative function and equals zero when $f(x)=m(x)$. That is why $m^*(x)=m(x)$. The optimal approximation of Y by the function of X is given by $m(X)$.

3.2 Estimation of the regression function

To predict the values of Y we need to build the regression function. For this purpose we need to know the distribution of the random vector (X,Y). However, in practice the distribution is usually unknown. That is why we cannot receive explicitly the regression function, and we are unable to predict the values of Y using this function. The only thing we can do is to estimate the values of this function by numerically observing some data.

There are two approaches for the estimation of the regression function, the parametric and nonparametric methods. In the parametric model one assumes that the structure of the model is known. There are finitely many unknown parameters and there is a need to estimate them using the observed data. The most popular example of parametric regression estimation is a linear regression. The main assumption here is that the regression function is a linear combination of its components, and applying the principle of the least squares it is possible to estimate parameters from data and receive the results. Usually, the advantage of such an approach is that the number of parameters is quite few. This method can be used even for small amounts of data if parametric model is chosen appropriately. The parameters of the model are always easy to interpret.

However, in general, the parametric method behaves poorly in comparison to the nonparametric one. If, for example, the true regression function cannot be approximated by linear function, the error of linear estimate will be big. The next difficulty is how we can deal with multivariate observable variable X. It is not easy to visualize it, therefore it will be hard to find a parametric model that fits well. As a result, the error of the estimate will be big. In contrast to parametric methods, nonparametric estimation methods do not assume any parametric model, meaning that there is no assumption that the regression function can be described using finitely many parameters. This is one of the reasons, why the nonparametric regression plays a crucial role in prediction of time series.

How can we do this estimation? We observe some data.

Let us denote by $(X, Y), (X_1, Y_1), (X_2, Y_2), \ldots$ independent and identically distributed random variables with $\mathbb{E}Y^2 < \infty$. Let $\mathcal{D}_n$ be the set of observed data defined by

$$\mathcal{D}_n = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}.$$

Then we want to use the data from the set $\mathcal{D}_n$ for the construction of estimator $m_n : \mathbb{R}^d \to \mathbb{R}$ of regression function m. Here $m_n(x) = m_n(x, \mathcal{D}_n)$ is a measurable function of x and the data.

The estimators are not unique and they are not equal to the regression function. We need to compare them somehow, which is why we introduce the error criterion, which will help us to distinguish between the quality of the estimators. By quality of the estimator we mean the difference between the regression function and an arbitrary estimate m_n.

There are several different error criteria that are used in the literature:

1. The pointwise error

$$|m_n(x) - m(x)| \text{ for some fixed } x \in \mathbb{R}^d,$$

2. The supremum norm error

$$\|m_n - m\|_\infty = \sup_{x \in C} |m_n(x) - m(x)| \text{ for some fixed set } C \subseteq \mathbb{R}^d,$$

which is mostly used for $d = 1$ where C is a compact subset of $\mathbb{R}$,

3. The L_p error

$$\int_C |m_n(x) - m(x)|^p \, dx,$$

where the integration is with respect to the Lebesgue measure, C is a fixed subset of $\mathbb{R}^d$, and $p \geq 1$ is arbitrary.

Introducing the regression function, our aim is to find function f such that the mean squared error $\mathbb{E}\,|f(X) - Y|^2$ is as small as possible. This fact leads naturally to usage of an L_2 error criterion for measuring the performance of the regression function estimate. The minimal value of this L_2 risk is $\mathbb{E}\,|m(X) - Y|^2$, and it is achieved by the regression function m. The L_2 risk of an estimate m_n can be found as:

$$\mathbb{E}\{|m_n(X) - Y|^2 \mid \mathcal{D}_n\} = \int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \, \mu(dx) + \mathbb{E}\,|m(X) - Y|^2,$$

The proof of this fact is absolutely similar to what we have done before. Thus the L_2 risk of an estimate m_n is close to the optimal value if and only if the L_2 error

$$\int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \, \mu(dx) \tag{3.2}$$

is close to zero. Therefore, it is possible to use L_2 error (3.2) for measuring the quality of an estimate. The L_2 error is random since the estimate m_n depends on the data in the set $\mathcal{D}_n$.

The first and the weakest property the estimator should have is that, as the sample size grows, it should converge to the estimated quantity.

Definition 3.2.1 *The estimator is called consistent if it has the property that the error estimation converges to zero as the sample size grows to infinity.*

We are interested in the convergence of the the expectation of the random variable (3.2) to zero.

Definition 3.2.2 *A sequence of regression function estimates* $\{m_n\}$ *is called weakly consistent for a certain distribution of* (X, Y)*, if*

$$\lim_{n\to\infty} \mathbb{E}\left\{\int (m_n(x) - m(x))^2 \mu(dx)\right\} = 0$$

Definition 3.2.3 *A sequence of regression function estimates* $\{m_n\}$ *is called strongly consistent for a certain distribution of* (X, Y)*, if*

$$\lim_{n\to\infty} \mathbb{E}\left\{\int (m_n(x) - m(x))^2 \mu(dx)\right\} = 0 \quad \textit{with probability one.}$$

The estimator is better if it is consistent for a larger class of distributions.

We use nonparametric estimates mainly when there is no information available about the distribution of the random vector (X, Y). Therefore, the concept of universal consistency is important in the nonparametric regression. It is essential to have estimator that performs well for all distributions because of the lack of information about the true distribution. The requirement of universal goodness can be formulated in the following way:

Definition 3.2.4 *A sequence of regression function estimates* $\{m_n\}$ *is called weakly universally consistent if it is weakly consistent for all distributions of* (X, Y) *with* $\mathbb{E}\{Y^2\} < \infty$.

Definition 3.2.5 *A sequence of regression function estimates* $\{m_n\}$ *is called strongly universally consistent if it is strongly consistent for all distributions of* (X, Y) *with* $\mathbb{E}\{Y^2\} < \infty$.

If an estimate is universally consistent, then, regardless of the true underlying distribution of (X, Y), the L_2 error of the estimate converges to zero if a sample size tends to infinity.

It is desirable that the estimator error tends to infinity as fast as possible. To answer the question what the rate of convergence for this estimator is, we take a look at the expectation of the L_2 error:

$$\mathbb{E} \int |m_n(x) - m(x)|^2 \mu(dx). \tag{3.3}$$

Unfortunately, the estimators for which (3.3) converges to zero at some fixed, nontrivial rate for all distributions do not exist. If one needs to get nontrivial rates, he should restrict the class of distributions. One possible variant of restriction is to assume some smoothness property of the regression function.

Suppose that the data can be written as

$$Y_i = m(X_i) + \epsilon_i, \tag{3.4}$$

where the error term $\epsilon_i = Y_i - m(X_i)$ satisfies $\mathbb{E}\{\epsilon_i \mid X_i\} = 0$. Thus Y_i can be considered as the sum of the value of the regression function at point X_i and some error term with expected value zero. As X_i's are random variables, we call our problem regression estimation with random design. The main properties of this model are:

- The error term ϵ_i depends on X_i, thus it is not independent of X_i and its kind of distribution can be changed with X_i,

- The points in the set $\mathcal{D}_n$ are independent identically distributed,

- The dimension d of random vector X is often much larger than two.

As the data can be written as in (3.4), this motivates the construction of the estimates by local averaging, i.e., estimation of $m(x)$ by the average of those Y_i where X_i is "close" to x. Such an estimate can be written as

$$m_n(x) = \sum_{i=1}^{n} W_{n,i}(x) \cdot Y_i,$$

where the weights $W_{n,i}(x) = W_{n,i}(x, X_1, \ldots, X_n) \in \mathbb{R}$ depend on $X_1, \ldots, X_n$. The weights are nonnegative and $W_{n,i}$ is small if X_i is far from x.

3.3 Local averaging estimate

The first example of the local averaging estimate is the partitioning estimate. To build it, first we need to choose the finite or countably infinite partition $\mathcal{P}_n = \{A_{n,1}, A_{n,2}, \ldots\}$ of $\mathbb{R}^d$, where the sets $A_{n,j} \subseteq \mathbb{R}^d$ are Borel sets. For $x \in A_{n,j}$ the estimate of the regression function is the average of Y_i's with the corresponding X_i's in $A_{n,j}$.

Definition 3.3.1 *The partitioning estimate of the regression function $m(x)$ is defined by*

$$m_n(x) = \frac{\sum_{i=1}^{n} \mathbb{I}_{\{X_i \in A_{n,j}\}} Y_i}{\sum_{i=1}^{n} \mathbb{I}_{\{X_i \in A_{n,j}\}}} \qquad \textit{for } x \in A_{n,j},$$

where $\mathbb{I}_A$ is the indicator of the set A and the weight is

$$W_{n,i}(x) = \frac{\mathbb{I}_{\{X_i \in A_{n,j}\}}}{\sum_{i=1}^{n} \mathbb{I}_{\{X_i \in A_{n,j}\}}} \qquad \text{for } x \in A_{n,j}.$$

Here and further we use $\frac{0}{0} = 0$ for convention.

Under the following assumption the partitioning regression function estimate will be weakly universally consistent:

If for each sphere S centred at the origin

$$\lim_{n \to \infty} \max_{j : A_{n,j} \cap S \neq 0} diam(A_{n,j}) = 0$$

and

$$\lim_{n \to \infty} \frac{|j : A_{n,j} \cap S \neq 0|}{n} = 0.$$

The proof of this fact can be found in Györfi et al. (2002).

The second example of local averaging estimate is the Nadaraya-Watson kernel estimate. Let $K : \mathbb{R}^d \to \mathbb{R}_+$ be a kernel function, let $h > 0$ be a bandwidth.

Definition 3.3.2 *The kernel estimate of the regression function $m(x)$ is defined by*

$$m_n(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}$$

The weight is

$$W_{n,i}(x) = \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}.$$

One can use naive kernel estimate (or window kernel) where the kernel function is $K(x) = \mathbb{I}_{\{\|x\|\leq 1\}}$, and we can rewrite the regression estimator as

$$m_n(x) = \frac{\sum_{i=1}^n \mathbb{I}_{\{\|x-X_i\|\leq h\}} Y_i}{\sum_{i=1}^n \mathbb{I}_{\{\|x-X_i\|\leq h\}}} \tag{3.5}$$

Here we estimate $m(x)$ taking the average of those Y_i's for which the distance between the X_i and x is not greater than h. This concept can be generalized in the next way. One can use the weighted average of the Y_i, where the weight depends on the distance between X_i and x.

Assume that there are balls $S_{0,r}$ of radius r and balls $S_{0,R}$ of radius R centred at the origin $(0 < r < R)$, and constant $b > 0$ such that

$$\mathbb{I}_{\{x\in S_{0,R}\}} \geq K(x) \geq b\mathbb{I}_{\{x\in S_{0,r}\}}.$$

If $h_n \to 0$ and $nh_n^d \to \infty$, then the kernel estimate is consistent. The proof of this fact can be found in Györfi et al. (2002).

The last example of local averaging estimate is the k-nearest neighbour estimate. Generally speaking, we determine the k nearest X_i's to x in terms of distance $\|x - X_i\|$ and estimate $m(x)$ by the average of the corresponding Y_i's.

For $x \in \mathbb{R}^d$, let

$$\{(X_{(1)}(x), Y_{(1)}(x)), (X_{(2)}(x), Y_{(2)}(x)), \ldots, (X_{(n)(x)}, Y_{(n)}(x))\}$$

be a permutation of

$$\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$$

such that

$$\left\|x - X_{(1)}\right\| \leq \left\|x - X_{(2)}\right\| \leq \cdots \leq \left\|x - X_{(n)}\right\|.$$

Definition 3.3.3 *The k-nearest neighbour estimate of the regression function is defined by*

$$m_n(x) = \frac{1}{k}\sum_{i=1}^k Y_{(i)}(x).$$

Here the weight $W_{n,i}(x)$ equals

$$W_{n,i}(x) = \begin{cases} \frac{1}{k}, & \text{if } X_i \text{ is among the } k\text{-nearest neighbours of } x \\ 0, & \text{otherwise} \end{cases}$$

If $k_n \to \infty$, $\frac{k_n}{n} \to 0$, then the k_n nearest neighbour regression function estimate is weakly consistent for all distributions of (X, Y) where ties occur with probability zero and $\mathbb{E}Y^2 < \infty$. The proof of this fact can be found in Györfi et al. (2002).

The rate of convergence $\mathbb{E}\{\|m_n - m\|^2\}$ of all showed estimates is

$$\mathbb{E}\{\|m_n - m\|^2\} = \mathcal{O}(n^{-\frac{2}{d+2}}).$$

The kernel estimate is relatively easy and has less computational complexity compared to other methods. For this reason, we will apply it further in our models.

Chapter 4

Empirical results

In the second chapter we have received that for determining the option price it is enough to work with the continuation values. As we can not receive the prices explicitly, we want to estimate them using nonparametric methods. In the third chapter we discussed how the estimation can be done. Now we are going to consider two models, making different assumptions on the asset price process, and see the results.

4.1 Markov modelling

One of the models that appears naturally is the Markov model, where we assume that the price process follows the Markov process. Here we can apply the kernel estimate and get

$$q_{t,n}(x) = \frac{\sum_{i=1}^{n} \max\left\{f_{t+1}(X_{i,t+1}), q_{t+1,n}(X_{i,t+1})\right\} \mathbb{I}_{\{|X_{i,t}-x|\leq H/2\}}}{\sum_{i=1}^{n} \mathbb{I}_{\{|X_{i,t}-x|\leq H/2\}}}, \tag{4.1}$$

where $\mathbb{I}$ denotes the indicator, and $0/0 = 0$ by definition. All the estimators are used as in Györfi et al. (2012).

Recall from section 2.5, that $X_{i,1}, \ldots, X_{i,T}, \quad i = 1, \ldots, n$ is the sample path prices generated from the historical data sequence $X_1, \ldots, X_N$ using the sliding sampling:

$$X_{i,t} = \frac{X_{i+t}}{X_i}, \quad i = 1, \ldots, N - T.$$

In section (2.4) we received the combination of lower and upper bounds (2.28), which is

$$\max_{s\in\{t+1,\ldots,T\}} \mathbb{E}\left\{f_s(X_s) \mid X_t = x\right\} \leq q_t(x) \leq \mathbb{E}\left\{\max_{s\in\{t+1,\ldots,T\}} f_s(X_s) \mid X_t = x\right\}.$$

Both of these bounds are true regression functions.

The lower bound can be estimated in the following way

$$q_{t,n}^{(l)}(x) = \max_{s\in\{t+1,\ldots,T\}} \frac{\sum_{i=1}^{n} f_s(X_{i,s})\mathbb{I}_{\{|X_{i,t}-x|\leq H/2\}}}{\sum_{i=1}^{n} \mathbb{I}_{\{|X_{i,t}-x|\leq H/2\}}}.$$

And an estimate of the upper bound is

$$q_{t,n}^{(u)}(x) = \frac{\sum_{i=1}^{n} \max_{s\in\{t+1,\dots,T\}} f_s(X_{i,s})\mathbb{I}_{\{|X_{i,t}-x|\le H/2\}}}{\sum_{i=1}^{n} \mathbb{I}_{\{|X_{i,t}-x|\le H/2\}}}.$$

The empirical results for this model were received in Kohler (2010). The average of the prices, given by Markov model, is almost the same in comparison to other models, but the variance of the prices is higher. Consequently, it makes sense to investigate more closely another model, the memoryless model, which was shown to be better above.

4.2 Memoryless modelling

This section introduces the memoryless model to the reader, giving the detailed analysis of the continuation function that was received under the assumptions of the memoryless model.

Definition 4.2.1 *The process $\{X_t\}$ is called of memoryless multiplicative increments, if $X_1/X_0, X_2/X_1, \dots$ are independent random variables.*

Definition 4.2.2 *The process $\{X_t\}$ is called of stationary multiplicative increments, if the sequence $X_1/X_0 = X_1, X_2/X_1, \dots$ is strictly stationary.*

From now on we will assume that the price process $\{X_t\}$ has memoryless and stationary multiplicative increments. These properties imply that, for $s > t$, $\frac{X_s}{X_t}$ and X_t are independent, and $\frac{X_s}{X_t}$ and $\frac{X_{s-t}}{X_0} = X_{s-t}$ have the same distribution.

From backward recursive scheme (2.25) for $t < T$ we receive

$$\begin{aligned}
q_t(x) &= \mathbb{E}\left\{\max\left\{f_{t+1}\left(X_{t+1}\right), q_{t+1}\left(X_{t+1}\right)\right\} \mid X_t = x\right\} \\
&= \mathbb{E}\left\{\max\left\{e^{-r(t+1)}\left(K - \frac{X_{t+1}}{X_t}x\right)^+, q_{t+1}\left(\frac{X_{t+1}}{X_t}x\right)\right\} \mid X_t = x\right\} \\
&= \mathbb{E}\left\{\max\left\{e^{-r(t+1)}\left(K - \frac{X_{t+1}}{X_t}x\right)^+, q_{t+1}\left(\frac{X_{t+1}}{X_t}x\right)\right\}\right\} \\
&= \mathbb{E}\left\{\max\left\{e^{-r(t+1)}\left(K - X_1 x\right)^+, q_{t+1}\left(X_1 x\right)\right\}\right\},
\end{aligned} \tag{4.2}$$

where

- in the second step we have substituted the payoff function,
- in the third step we used the fact that the increments possess a memoryless property, which is why we can forget about the condition in the expectation,
- in the last step we used that the increments are stationary.

Claim 4.1 *The continuation value function that was derived for the memoryless model is non-increasing.*

Proof.

We are going to use method of mathematical induction. From section 2.1 we know that $q_T(x) = 0$

- **First step** Let us prove that $q_{T-1}(x)$ is non-increasing.

  Suppose that $x_1 < x_2$,

$$q_{T-1}(x_1) = \mathbb{E}\{e^{-rT} (K - X_1 x_1)^+\},$$

$$q_{T-1}(x_2) = \mathbb{E}\{e^{-rT} (K - X_1 x_2)^+\}.$$

  The expressions inside of the expectations are:

$$(K - X_1 x_1)^+ = \begin{cases} K - X_1 x_1, & \text{if} \quad x_1 < \dfrac{K}{X_1} \\ 0, & \text{otherwise} \end{cases},$$

$$(K - X_1 x_2)^+ = \begin{cases} K - X_1 x_2, & \text{if} \quad x_2 < \dfrac{K}{X_1} \\ 0, & \text{otherwise} \end{cases}.$$

  (a) Let $x_1 < x_2 < \frac{K}{X_1}$, then

$$\begin{aligned} q_{T-1}(x_1) - q_{T-1}(x_2) &= \mathbb{E}\{e^{-rT} (K - X_1 x_1)\} - \mathbb{E}\{e^{-rT} (K - X_1 x_2)\} \\ &= e^{-rT} K - x_1 \mathbb{E}\{X_1\} - e^{-rT} K + x_2 \mathbb{E}\{X_1\} \\ &= \mathbb{E}\{X_1\} (x_2 - x_1) > 0. \end{aligned}$$

  (b) Let $x_1 < \frac{K}{X_1} < x_2$, then

$$q_{T-1}(x_1) - q_{T-1}(x_2) = \mathbb{E}\{e^{-rT} (K - X_1 x_1)\} > 0.$$

  (c) Let $\frac{K}{X_1} < x_1 < x_2$, then

$$q_{T-1}(x_1) - q_{T-1}(x_2) = 0.$$

  Thus, we have proved that $q_{T-1}(x)$ is a non-increasing function. Making inductive step, we assume that the functions $q_{T-2}(x), \ldots, q_{T-n}(x)$ are non-increasing.

- **Second step** Let us prove that the function $q_{T-n-1}(x)$ is non-increasing. We are going to look at the sign of the difference $q_{T-n-1}(x_1) - q_{T-n-1}(x_2)$, which should always be positive. Suppose that $x_1 < x_2$,

$$q_{T-n-1}(x_1) = \mathbb{E}\{\max\{e^{-r(T-n)} (K - X_1 x_1)^+, q_{T-n}(X_1 x_1)\}\},$$

$$q_{T-n-1}(x_2) = \mathbb{E}\{\max\{e^{-r(T-n)} (K - X_1 x_2)^+, q_{T-n}(X_1 x_2)\}\}.$$

As $x_1 < x_2$, then $X_1x_1 < X_1x_2$, and since the function $q_{T-n}(x)$ is non-increasing, we have

$$q_{T-n}(X_1x_1) > q_{T-n}(X_1x_2). \tag{4.3}$$

The next observation is, as

$$K - X_1x_1 > K - X_1x_2,$$

it follows that

$$(K - X_1x_1)^+ > (K - X_1x_2)^+,$$

and

$$e^{-r(T-n)}(K - X_1x_1)^+ > e^{-r(T-n)}(K - X_1x_2)^+. \tag{4.4}$$

– **First case** Suppose that

$$e^{-r(T-n)}(K - X_1x_1)^+ > q_{T-n}(X_1x_1) \tag{4.5}$$

then

$$q_{T-n-1}(x_1) = \mathbb{E}\{e^{-r(T-n)}(K - X_1x_1)^+\}.$$

We have two possible situations for the value of $q_{T-n-1}(x_2)$: either

$$q_{T-n-1}(x_2) = \mathbb{E}\{e^{-r(T-n)}(K - X_1x_2)^+\} \tag{4.6}$$

or

$$q_{T-n-1}(x_2) = \mathbb{E}\{q_{T-n}(X_1x_2)\}. \tag{4.7}$$

If (4.6) is true, then the difference is

$$\begin{aligned}
&q_{T-n-1}(x_1) - q_{T-n-1}(x_2)\\
&= \mathbb{E}\{e^{-r(T-n)}(K - X_1x_1)^+\} - \mathbb{E}\{e^{-r(T-n)}(K - X_1x_2)^+\}\\
&= \mathbb{E}\{e^{-r(T-n)}(K - X_1x_1)^+ - e^{-r(T-n)}(K - X_1x_2)^+\} > 0.
\end{aligned}$$

The last inequality holds because of (4.4).
If (4.7) is true, then the difference is

$$\begin{aligned}
&q_{T-n-1}(x_1) - q_{T-n-1}(x_2)\\
&= \mathbb{E}\{e^{-r(T-n)}(K - X_1x_1)^+\} - \mathbb{E}\{q_{T-n}(X_1x_2)\}\\
&= \mathbb{E}\{e^{-r(T-n)}(K - X_1x_1)^+ - q_{T-n}(X_1x_2)\} > 0.
\end{aligned}$$

The last difference is positive because of (4.3) and (4.5).

– **Second case** Suppose

$$e^{-r(T-n)}\left(K - X_1x_1\right)^+ \leq q_{T-n}(X_1x_1), \tag{4.8}$$

then

$$q_{T-n-1}(x_1) = \mathbb{E}\{q_{T-n}(X_1x_1)\}.$$

Again, we have the same cases for the function $q_{T-n-1}(x_2)$ as (4.6) and (4.7). If (4.6) is true, then the difference is

$$\begin{aligned} &q_{T-n-1}(x_1) - q_{T-n-1}(x_2) \\ &= \mathbb{E}\{q_{T-n}(X_1x_1)\} - \mathbb{E}\{e^{-r(T-n)}\left(K - X_1x_2\right)^+\} \\ &= \mathbb{E}\{q_{T-n}(X_1x_1) - e^{-r(T-n)}\left(K - X_1x_2\right)^+\} > 0. \end{aligned}$$

The last inequality holds because of (4.4) and (4.8).
If (4.7) is true, then the difference is

$$\begin{aligned} &q_{T-n-1}(x_1) - q_{T-n-1}(x_2) \\ &= \mathbb{E}\{q_{T-n}(X_1x_1)\} - \mathbb{E}\{q_{T-n}(X_1x_2)\} \\ &= \mathbb{E}\{q_{T-n}(X_1x_1) - q_{T-n}(X_1x_2)\} > 0. \end{aligned}$$

The last difference is positive because of (4.3). Therefore, the continuation value function $q_t(x)$ is indeed non-increasing.

■

This claim shows us that the payoff from continuation is getting less with time, and there is no sense in waiting infinitely long. Also, this is another argument for the fact that the optimal stopping time should be exactly the first moment of time where the payoff from immediate exercising the option is higher than the payoff from continuation.

4.3 Estimators in memoryless model

If we are given data $X_1, \ldots, X_N$, $i = 1, \ldots, N$ then, for any fixed t the continuation value function can be estimated by simple average. Let $q_{t+1,N}(x)$ be an estimate of $q_{t+1}(x)$. Estimators of the continuation function, lower bound and upper bound are used as in Györfi et al. (2012). Using the backward recursive representation we have:

$$q_{t,N}(x) = \frac{1}{N}\sum_{i=1}^{N} \max\left\{e^{-r(t+1)}\left(K - xX_i/X_{i-1}\right)^+, q_{t+1,N}(xX_i/X_{i-1})\right\}. \tag{4.9}$$

As we are using the recursion, here the errors are likely to accumulate, and so we need to control this. Therefore, we introduce the estimates of the lower and upper bounds of the continuation values.

For memoryless process, the lower bound (2.27) of $q_t(x)$ has a simple form:

$$\begin{aligned}
q_t^{(l)}(x) &= \max_{s\in\{t+1,\dots,T\}} \mathbb{E}\left\{f_s(X_s)|X_t=x\right\} \\
&= \max_{s\in\{t+1,\dots,T\}} e^{-rs}\mathbb{E}\left\{\left(K-\frac{X_s}{X_t}X_t\right)^+ \mid X_t=x\right\} \\
&= \max_{s\in\{t+1,\dots,T\}} e^{-rs}\mathbb{E}\left\{\left(K-\frac{X_s}{X_t}x\right)^+ \mid X_t=x\right\} \\
&= \max_{s\in\{t+1,\dots,T\}} e^{-rs}\mathbb{E}\left\{\left(K-\frac{X_s}{X_t}x\right)^+\right\} \\
&= \max_{s\in\{t+1,\dots,T\}} e^{-rs}\mathbb{E}\left\{(K-X_{s-t}x)^+\right\},
\end{aligned}$$

where we are using the same arguments as in the formula (4.2) (memoryless and stationary multiplicative increments).

Thus,

$$q_t^{(l)}(x) = \sup_{s\in\{t+1,\dots,T\}} e^{-rs}\mathbb{E}\left\{(K-X_{s-t}x)^+\right\}.$$

If we are given the sample path $X_{i,1},\dots,X_{i,T}$, $i=1,\dots n$ then the estimate of $q_t^{(l)}(x)$ would be

$$q_{t,n}^{(l)}(x) = \max_{s\in\{t+1,\dots,T\}} e^{-rs}\frac{1}{n}\sum_{i=1}^{n}(K-X_{i,s-t}x)^+ .$$

As previously we agreed to use the sliding sample scheme, then in this case $n=N-T$.

For the upper bound we can use the same arguments as before and receive

$$\begin{aligned}
q_t^{(u)}(x) &= \mathbb{E}\left\{\max_{s\in\{t+1,\dots,T\}} f_s(X_s) \mid X_t=x\right\} \\
&= \mathbb{E}\left\{\max_{s\in\{t+1,\dots,T\}} e^{-rs}(K-X_s)^+ \mid X_t=x\right\} \\
&= \mathbb{E}\left\{\max_{s\in\{t+1,\dots,T\}} e^{-rs}\left(K-\frac{X_s}{X_t}X_t\right)^+ \mid X_t=x\right\} \\
&= \mathbb{E}\left\{\max_{s\in\{t+1,\dots,T\}} e^{-rs}\left(K-\frac{X_s}{X_t}x\right)^+ \mid X_t=x\right\} \\
&= \mathbb{E}\left\{\max_{s\in\{t+1,\dots,T\}} e^{-rs}\left(K-\frac{X_s}{X_t}x\right)^+\right\} \\
&= \mathbb{E}\left\{\max_{s\in\{t+1,\dots,T\}} e^{-rs}(K-X_{s-t}x)^+\right\}.
\end{aligned}$$

If we are given sample path $X_{i,1},\dots,X_{i,T}$, $i=1,\dots n$, then the estimate of $q_t^{(u)}(x)$ would be

$$q_{t,n}^{(u)}(x) = \frac{1}{n}\sum_{i=1}^{n}\max_{s\in\{t+1,\dots,T\}} e^{-rs}(K-X_{i,s-t}x)^+ .$$

The combination of the lower and upper bounds is

$$\max_{s\in\{t+1,\ldots,T\}} \mathbb{E}\left\{e^{-rs}\left(K - X_{s-t}x\right)^+\right\} \le q_t(x) \le \mathbb{E}\left\{\max_{s\in\{t+1,\ldots,T\}} e^{-rs}\left(K - X_{s-t}x\right)^+\right\}.$$

Using the estimates of the lower and upper bound, the following truncation of the estimates of the continuation value is suggested in Györfi et al. (2012):

$$\hat{q}_{t,N}(x) = \begin{cases} q^{(u)}_{t,n}(x) & \text{if} \quad q^{(u)}_{t,n}(x) < q_{t,N}(x), \\ q_{t,N}(x) & \text{if} \quad q^{(u)}_{t,n}(x) \ge q_{t,N}(x) \ge q^{(l)}_{t,n}(x), \\ q^{(l)}_{t,n}(x) & \text{if} \quad q_{t,N}(x) < q^{(l)}_{t,n}(x). \end{cases} \tag{4.10}$$

The last formula gives us an improved estimator of the continuation value function.

The problem with the estimator of the continuation function is the following. We are using the whole amount of data we have, the whole historical data sequence, but in real life we do not know all the prices in advance. Therefore, this estimator has to be slightly changed. At each moment of time t we have to use only the prices available up to this exact moment. Suppose that at initial moment of time when we make a contract, we know A prices. Then at the moment of time t, where $t = 1, \ldots, T$ (T is the maturity date of the option), we know $N' = A + t$ prices. Consequently, the changed estimator takes the following form:

$$q_{t,N'}(x) = \frac{1}{N'}\sum_{i=1}^{N'} \max\left\{e^{-r(t+1)}\left(K - xX_i/X_{i-1}\right)^+, q_{t+1,N'}(xX_i/X_{i-1})\right\}, \tag{4.11}$$

and $q_{T,N'}(x) = 0$.

The lower and upper bounds essentially remain the same. The only thing we have to change is the amount of data we use for generating sample paths. Therefore, the maximal number of sample paths will be $n = A + t - T = N' - T$. We will refer to all these values as changed lower bounds, changed upper bounds and changed truncated estimator.

4.4 Simulations

This section contains the simulation results that we have performed using the Python software. The main aim of these simulations is to compare the behaviour of two estimators, with changes and without changes. We used daily NYSE data available at Log-optimal portfolio homepage. We selected the asset *coke*, which contains daily returns during the 44 years period from 1966 to 2006. We fixed the time unit equal to one week. The expiration time is $T = 20$, the strike price is $K = 110$, the initial price is $X_0 = 100$. Figures 1 and 2 illustrate the estimates for the memoryless model for $t = 4$ and $t = 16$, taking into account the changes that we have made. The curves on the pictures are:

- the dark blue curve is $q_{t,n}(x)$ without correction (4.11),
- the yellow curve is $q^{(l)}_{t,n}(x)$ (changed),

- the red curve is $q_{t,n}^{(u)}(x)$ (changed),
- the light blue curve is $q_{t,n}(x)$ with correction (changed).

These figures show us that the lower bounds are really close to the estimates of the true continuation functions, but the upper bounds are loose. The results that were received for the estimates of the continuation function and its bounds in Györfi et al. (2012) were the same. That means that the upper bounds still need to be improved.

Figures 3, 4, 5 and 6 show us the behaviour of the estimator with changes denoted by light-blue curve, and without changes by red curve. Figures 5 and 6 are zoomed marked areas on Figure 3 and 4, respectively. We see in the figures that the two functions do not differ significantly. Hence, the changes we made did not influence much the payoff from continuation, and as a consequence, the price of the option. However, the changes were crucial and enable us to use those estimates for data in a real life.

Figure 4.1: Estimates for memoryless modelling, $t = 4$

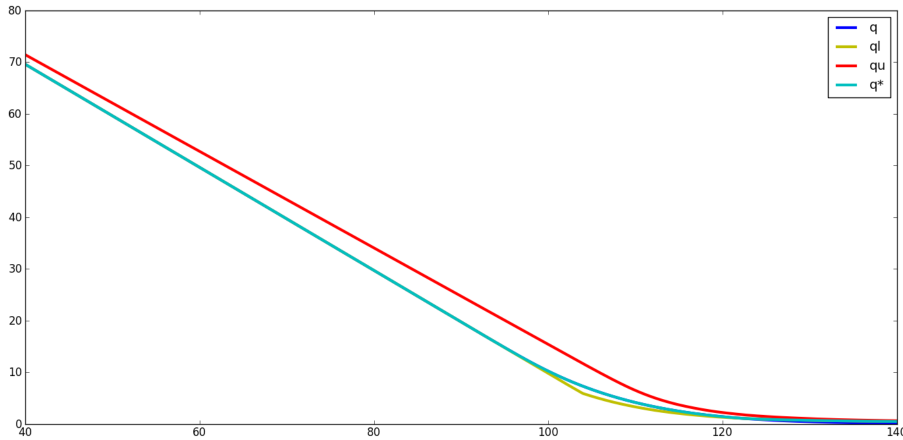


Figure 4.2: Estimates for memoryless modelling, $t = 16$

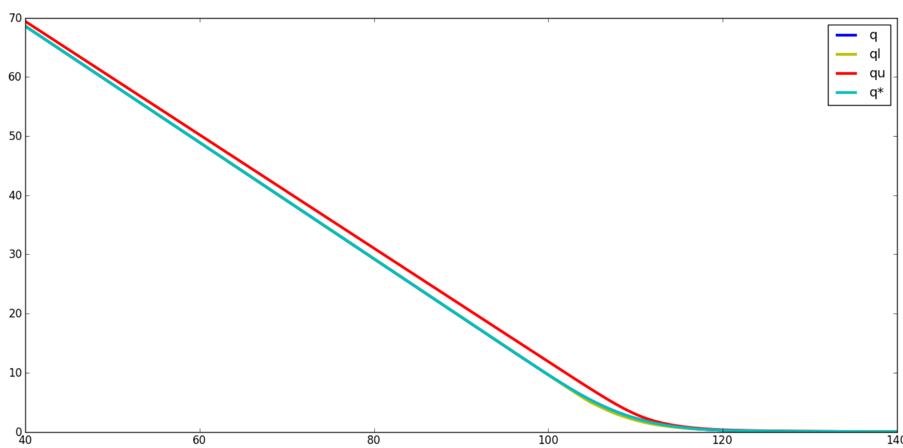


Figure 4.3: Original and changed estimates, t=4

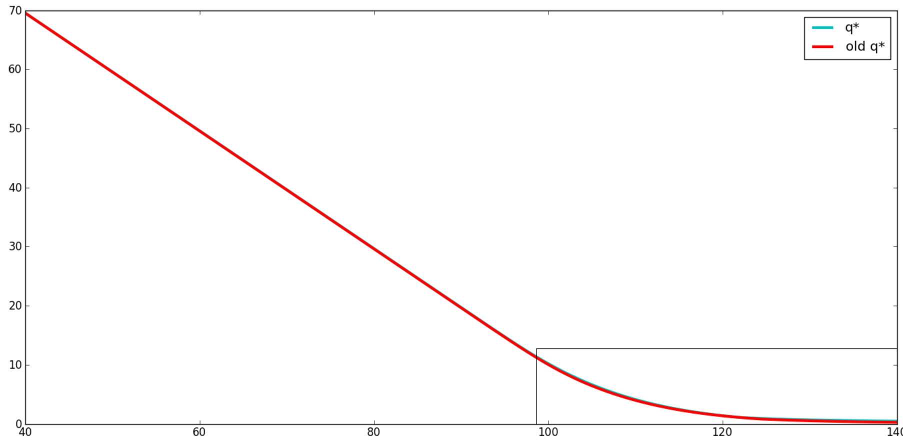


Figure 4.4: Original and changed estimates, $t = 16$

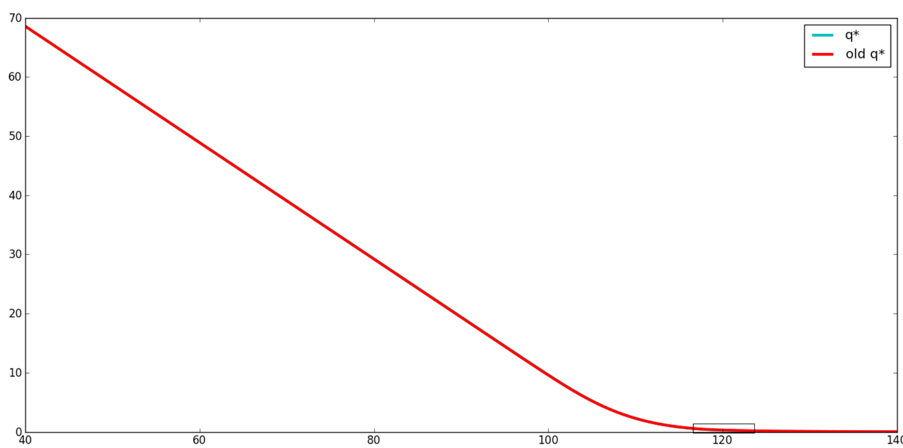


Figure 4.5: Original and changed estimates, zoomed part, $t = 4$

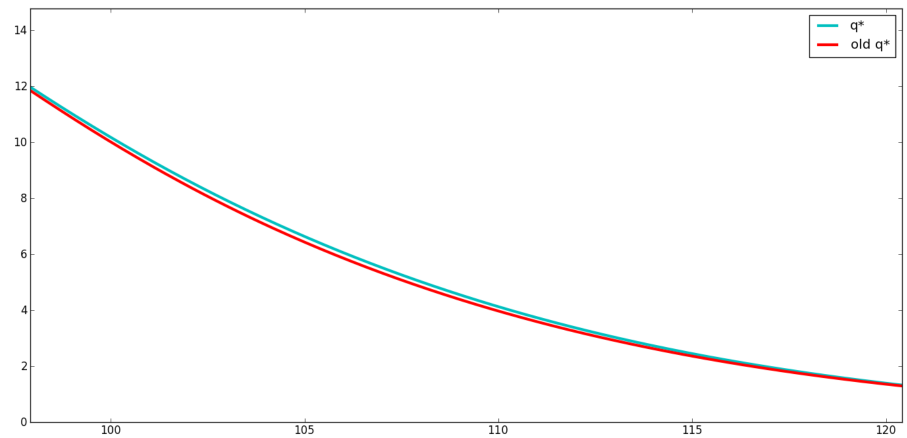
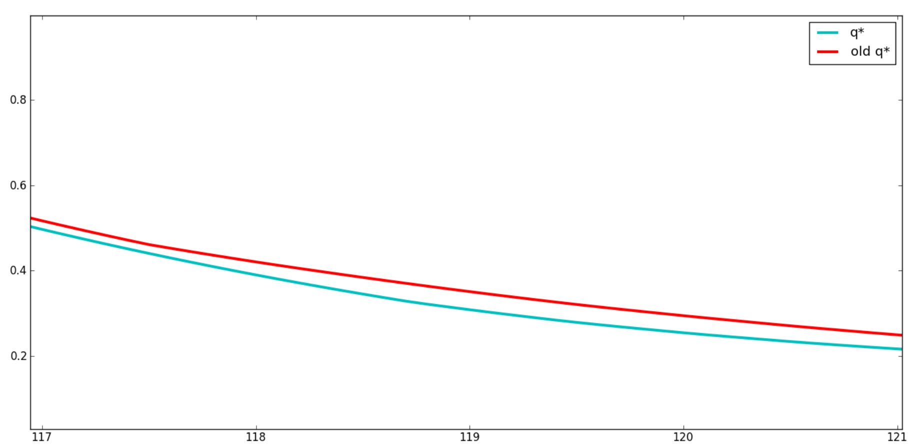


Figure 4.6: Original and changed estimates, zoomed part, $t = 16$



Conclusion

This thesis has built the general setup for the problem of pricing the American options. We have used empirical methods to receive the price of the options. The empirical approach means that we use the known prices from the past to receive the estimate of the future price of the options. We summarized the results developed by different researchers, making full, self-contained survey of this problem.

We have performed the simulations to evaluate the behaviour of the changed estimator that is suited for the real life problems. We saw that the estimates for the continuation value are almost the same as in Györfi et al. (2012). This shows that the previously received results are truthful.

In addition, we proved that the continuation function in memoryless model is non-increasing, and this explains the intuition behind the optimal stopping time.

Bibliography

Log-optimal portfolio homepage. URL `www.szit.bme.hu/~oti/portfolio`.

J. Carriér. Valuation of early-exercise price of options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19:19–30, 1996.

Y. S. Chow, H. Robbins, and D. Siegmund. *Great expectations: The theory of optimal stopping*. Houghton Mifflin Boston, 1971.

D. Egloff. Monte carlo algorithms for optimal stopping and statistical learning. *The Annals of Applied Probability*, 15(2):1396–1432, 2005.

D. Egloff, M. Kohler, and N. Todorovic. A dynamic look-ahead monte carlo algorithm for pricing bermudan options. *The Annals of Applied Probability*, 17(4):1138–1171, 2007.

L. Györfi, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer, 2002.

L. Györfi, D. Semler, and A. Telcs. Comparison of methods for pricing american options, 2012.

A. Klenke. *Probability theory: a comprehensive course*. Springerverlag London, 2008.

M. Kohler. A review on regression-based monte carlo methods for pricing american options. In *Recent Developments in Applied Probability and Statistics*, pages 37–58. Springer, 2010.

F. A. Longstaff and E. S. Schwartz. Valuing american options by simulation: A simple least-squares approach. *Review of Financial studies*, 14(1):113–147, 2001.

A. N. Shiryaev. *Optimal stopping rules*, volume 8. Springer, 2007.

S. E. Shreve. Stochastic calculus for finance ii, continuous time models, 2004.

J. N. Tsitsiklis and B. Van Roy. Optimal stopping of markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. *Automatic Control, IEEE Transactions on*, 44(10):1840–1851, 1999.