The least squares method for option pricing revisited

Maciej Klimek* and Marcin Pitera†

First Circulated: April 07, 2014

Abstract

It is shown that the the popular least squares method of option pricing converges even if the underlying is non-Markovian, the pay-offs are path dependent and with a very flexible setup for approximation of conditional expectations. The main benefit is the increase of freedom in creating specific implementations of the method, but depending on the extent of adopted generality and complexity, the method may become very demanding computationally. It is argued, however, that in many practical applications even modest but computationally viable extensions of standard linear regression may produce satisfactory results from the empirical point of view. This claim is illustrated with several empirical examples.

Key words: least squares option pricing, Snell envelopes, optimal stopping, approximation of conditional expectation, American options, basket options, Monte Carlo simulation, LIBOR market model, Heston-Nandi model.

MSC 2010: Primary - 91G20, 91G60, 93E24; Secondary - 60G40, 62J02.

1 Introduction

For over a decade several variants of the so called least-squares method of American option pricing have been widely used by financial practitioners and at the same time studied by researchers. The origins of the method can be found in the work of Carriere [5], Tsitsiklis, Van Roy [22] (see also [21]), Longstaff, Schwartz [16] and Clément, Lamberton, Protter [6]. Basically the method seeks a way of approximating conditional expectations needed in the valuation process either directly as in [16] and [6], or indirectly through the value function as in [22]. A modification of the algorithm from [16] was studied in [6] from the point of

*Department of Mathematics, P.O.Box 480, Uppsala University, 751 06 Uppsala, Sweden; *e-mail:* Maciej.Klimek@math.uu.se

†Department of Mathematics, Jagiellonian University, ul. Łojasiewicza 6, 30-348 Kraków. Poland; *e-mail:* Marcin.Pitera@im.uj.edu.pl

view of the convergence of the method. Subsequently, several papers on this subject have been published — we will mention just a few of them related to the present article.

Glasserman and Yu [10] investigated in 2004 the convergence of the least-squares like methods, where — basically — the necessary conditional expectations are approximated by finite linear combinations of approximating functions. More specifically they look into the problem of accuracy of estimations when the number of approximating functions and the number of simulated trajectories increase. They assume that the underlying is a multidimensional Markov process. The rather pessimistic outcome, from the practical point of view, is that for polynomials as the approximating functions and for conventional (resp. geometric) Brownian motion as the underlying, the number of required paths may grow exponentially in the degree (resp. the square of the degree) of the polynomials. Glasserman and Yu remark that similar property may hold also for more general approximating functions (with the number of approximating functions replacing the maximal degree).

Also in 2004 Stentoft [20] analyzed and extended the convergence results presented in [6]. In particular he has considered the problem of choosing the optimal number of regressors in relation to the number of simulated trajectories.

In 2005 Egloff [8] proposed an extension to the original Longstaff-Schwartz [16] as well as Tsitsiklis – Van Roy ([21], [22]) algorithms by treating the optimal stopping problem for multidimensional discrete time Markov processes as a generalized statistical learning problem. His results also improve those from [6]. Egloff comments that despite very good performance of least-squares algorithms in some practical calculations, precise estimates of the statistical quantities involved in these procedures may be difficult, leading to some less impressive performance in other cases.

Zanger [23] proposed in 2009 another extension to the least-squares method by considering fairly arbitrary subsets of information spaces as the approximating sets. He has also produced some new and interesting convergence results showing in particular that sometimes the exponential dependence on the number of time steps can be avoided.

Two features seem to be common to the articles mentioned above. Firstly, the underlying is assumed to be Markovian. Secondly, the convergence rates of the method, in all its incarnations, are not encouraging from the computational point of view. In the present paper, we extend the Clément, Lamberton, Protter approach [6] to show that the method converges even if the underlying is not a Markov process and if the pay-offs are path-dependent, with a fairly general setting for the regression approximating conditional expectations. Obviously by giving up the Markov property and aiming at better approximation of conditional expectation, the potential computational complexity increases considerably. However, the main advantage of relaxation of the assumptions is the increase in freedom to customize the method. Moreover, we would like to argue that the least-squares methods should be seen as a general framework leading to a variety of specific implementations. The main reason is essentially the fact that the information space for conditional expectation, or in other words its range, is in many interesting cases infinite dimensional. Inevitably, in these cases any approximation of conditional expectations, or value functions depending on conditional expectations, has to involve significantly restrictive extrinsic assumptions to make practical computations possible. While general convergence results are necessary to motivate the overall approach and some computational complexity may be addressed along the lines of [18], it is most likely that the future developments will evolve closer to simplified time-series models. It is quite conceivable that an alternative source of realism and numerical efficiency could exploit the advances in both time-series analysis and frame theory (see e.g. [14]). The empirical basis for such speculations comes from the fact, that in many real problems even taking only a few non-linear regressors, and sometimes ignoring lack of the Markov property, often leads to satisfactory results from the practical point of view. There seem to be much anecdotal evidence coming from the financial industry supporting the last statement and in this paper we provide further corroborating evidence in the form of three empirical examples.

It should be mentioned that the least squares approach can be also seen as part of the stochastic mesh framework proposed by Broadie and Glasserman ([3], [4]; see also [15] and [9]).

The material is organized as follows. The introduction is followed by a short review of Snell envelopes and consequences of the classic Dobrushin-Minlos theorem, which can lead to viable numerical approximations of conditional expectations. Then we show that the methods proposed by Clément, Lamberton and Protter [6] can be extended to cover the case of American style options with path dependent pay-offs, with a non-Markovian multidimensional underlying and with a very general approach to regression. This is followed by three computational examples illustrating the viability of the method under rather restrictive assumptions. First we present pricing of a one year Eurodollar American put and call options with different strike prices. Then we use the least-squares approach to price a 1.5 month American put option, whose payoff function depends on two market indices, namely DAX and EUROSTOXX50. Finally, we use the least-squares algorithm to price two 1.5 month American put options, whose payoff function is based on a single market index under the assumption that the underlyings can be described by the Heston-Nandi GARCH(1,1) model [11]. Again, we will use EUROSTOXX50 and DAX indices as the respective underlying instruments.

2 Snell envelopes and information spaces

It is well known that Snell envelopes are useful in valuation of American put options in discrete time models (see e.g. [17], p.127). They also furnish the main theoretical ingredient of the least squares option pricing algorithm which is the main topic of this paper. The standard use of Snell envelopes can be easily extended to provide pricing algorithms for more general American style options, that is options that allow execution at any time prior to maturity, but with a wide variety of pay-off patterns. Before justifying this statement we will recall without proofs some basic properties of Snell envelopes.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(\mathcal{F}_t)_{t=0}^T$ be a filtration, where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_T = \mathcal{F}$. Assume that an adapted stochastic process $(Z_t)_{t=0}^T$ is non-negative and integrable.

Let the symbol $\mathcal{C}_i^T$ denote the set of all stopping times with values in the set $\{i, i+1, \ldots, T\}$. A stopping time $\tau \in \mathcal{C}_0^T$ is said to be *optimal for* (Z_t) if

$$\mathrm{E}Z_\tau = \sup_{\nu \in \mathcal{C}_0^T} \mathrm{E}Z_\nu.$$

The *Snell envelope of* (Z_t), is defined as the adapted process

$$\begin{aligned} U_T &= Z_T, \\ U_t &= \max\left(Z_t, \mathrm{E}[U_{t+1}|\mathcal{F}_t]\right), \qquad t \in \{0, \ldots, T-1\}. \end{aligned}$$

Since (Z_t) is assumed to be integrable, (U_t) is also integrable due to L^1-continuity of the conditional expectation operator. The following theorem collects the standard properties of Snell envelopes:

Theorem 2.1. *Snell envelopes have the following properties:*

1. *(U_t) is the smallest supermartingale dominating (Z_t).*

2. *$U_t = \text{ess sup}\{\mathrm{E}[Z_\tau|\mathcal{F}_t] \ : \ \tau \in \mathcal{C}_t^T\}$.*

3. *Let*

   $$\tau_t = \min(s \geq t \mid U_s = Z_s).$$

   Then $\tau_t \in \mathcal{C}_t^T$ and

   $$\begin{aligned} \tau_T &= T, \\ \tau_t &= t\,\mathbf{1}_{\{Z_t \geq \mathrm{E}[U_{t+1}|\mathcal{F}_t]\}} + \tau_{t+1}\mathbf{1}_{\{Z_t < \mathrm{E}[U_{t+1}|\mathcal{F}_t]\}}, \qquad t \in \{0, \ldots, T-1\}. \end{aligned}$$

4. *$U_t = \mathrm{E}[Z_{\tau_t} \mid \mathcal{F}_t]$.*

5. *$\mathrm{E}[U_{t+1}|\mathcal{F}_t] = \mathrm{E}[Z_{\tau_{t+1}}|\mathcal{F}_t]$.*

6. *τ_0 is optimal for (Z_t). In particular, for any optimal stopping time σ,*

   $$U_0 = \mathrm{E}[Z_{\tau_0}] = \mathrm{E}[Z_\sigma].$$

7. *τ_t can also be defined recursively:*

   $$\begin{aligned} \tau_T &= T, \\ \tau_t &= t\,\mathbf{1}_{\{Z_t \geq \mathrm{E}[Z_{\tau_{t+1}}|\mathcal{F}_t]\}} + \tau_{t+1}\mathbf{1}_{\{Z_t < \mathrm{E}[Z_{\tau_{t+1}}|\mathcal{F}_t]\}}, \qquad t \in \{0, \ldots, T-1\}. \end{aligned}$$

8. $\tau \in \mathcal{C}_0^T$ *is optimal for* (Z_t) *if and only of*

   (a) $Z_\tau = U_\tau$,

   (b) $U_{t\wedge\tau}$ *is a martingale (where* $t \wedge \tau = \min(t, \tau)$*).*

9. $U_0 = \max(Z_0, \mathrm{E}[Z_{\tau_1}])$.

10. *The random variable* τ_0 *is the smallest optimal stopping time.*

11. *Let* $U_t = M_t - A_t$ *be the Doob decomposition of the Snell envelope into a martingale* M_t *and a non-decreasing predictable process* A_t *starting at* 0. *If* $K = \{t : \; A_{t+1} > 0\} \subset \{0, 1, \ldots, T-1\}$, *then*

$$\varrho = \begin{cases} T & \text{if } K = \emptyset, \\ \min K & \text{if } K \neq \emptyset, \end{cases}$$

   is the largest optimal stopping time.

(Proofs can be found e.g. in [13].)

In order to clarify our statement from the beginning of this section, consider a discrete time market model with a risky asset and a bank account. The risky asset price process $(S_t)_{t=0}^T$, which can be vector-valued, generates a filtration $(\mathcal{F}_t)_{t=0}^T$. The bank account process $(B_t)_{t=0}^T$ is assumed to be adapted to this filtration, positive and non-decreasing, with $B_0 = 1$. Lack of arbitrage implies the existence of a risk-neutral probability measure $\mathbb{Q}$ with respect to which the discounted price $\bar{S}_t = S_t/B_t$ is a martingale. Consider an American style option written on this stock. Strictly speaking, because early exercise is restricted to the dates from our discrete time scale, we are dealing here with a Bermudan style option. On the other hand the model allows making the time scale arbitrarily fine, so this terminological distinction is not particularly important. Let Z_t denote the intrinsic value process, that is the value of executing the option at time t. We will assume that $Z_t = f_t(S_0, \ldots, S_t)$, where f_t is a deterministic non-negative Borel function for each t. For instance, for an American put option $f_t(s_0, \ldots, s_t) = \max(K - s_t, 0)$, for some constant $K > 0$ representing the strike price. Let $\bar{Z}_t$ be its discounted version of Z_t i.e. $\bar{Z}_t = Z_t/B_t$. It is not difficult to notice that the value of the option at time $t-1$ is

$$U_{t-1} = \max\left(Z_{t-1}, \mathrm{E}\left[\frac{B_{t-1}}{B_t} U_t \,\middle|\, \mathcal{F}_{t-1}\right]\right), \qquad t = 1, \ldots, T,$$

with $U_T = Z_T$. In other words $\bar{U}_t = U_t/B_t$ is simply the Snell envelope of $\bar{Z}_t$. Moreover $U_0 = \bar{U}_0$. Any optimal stopping time provides a recipe for exercising the option at the right moment.

An alternative to the risk-neutral valuation, is to adopt an "actuarial" or "empirical" approach. Namely, suppose that one can identify a "risk-aware" adapted process D_t such that

$D_0 = 1$ and the expected value of U_t at time $t-1$ is $\mathrm{E}\left[D_t U_t \,\middle|\, \mathcal{F}_{t-1}\right]$, with respect to the objective probability measure. Then the above conclusion still holds with a new process $B_t = (D_0 D_1 \cdot \ldots \cdot D_t)^{-1}$.

The key element in any numerical implementation of Snell envelopes, is the ability to approximate the conditional expectation operator. Except for the finite case, one has to deal with infinite-dimensional spaces of random variables. Some elucidation seems to be in order here.

Since we will be dealing only with random variables of finite variance, we can rely on the Hilbert space geometry in addressing the issues of interest (see [19]). A closed subspace $S \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$ is said to be *probabilistic* if it contains constants and is closed with respect to taking the maximum of two of its elements, i.e. if $X, Y \in S$, then $X \vee Y \in S$. For any non-empty set $\mathbf{X} \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$, its *lattice envelope* $\mathrm{Latt}(\mathbf{X})$ is defined as the smallest probabilistic subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ containing $\mathbf{X}$. Obviously, even if $\mathbf{X}$ consists of just one random variable, $\mathrm{Latt}(\mathbf{X})$ can be infinite-dimensional. Moreover, if $\mathbf{X} = \{X_1, \ldots, X_n\}$ and $\mathcal{B}_n$ denotes the σ-algebra of Borel sets in $\mathbb{R}^n$, then it is not difficult to prove (see e.g. [13]) that

$$\mathrm{Latt}(\mathbf{X}) = L^2(\Omega, \sigma(\mathbf{X}), \mathbb{P}) = L^2(\Omega, (X_1, \ldots, X_n)^{-1}(\mathcal{B}_n), \mathbb{P}).$$

The latter will be referred to as the *information space generated by* $X_1, \ldots, X_n$. Since this is also the range of the orthogonal projection $\mathrm{E}[\cdot \mid X_1, \ldots, X_n]$, it would be desirable from the numerical standpoint to be able to approximate such projections, with projections onto smaller finite-dimensional vector spaces using available least-squares algorithms.

To this end one could use the following theorem, which is a slight reformulation of a result of Dobrushin and Minlos [7].

Theorem 2.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\alpha > 0$. Let $\mathcal{P}_n$ denote the space of all polynomials of n real variables. If $X_1, \ldots, X_n$ are random variables such that $e^{|X_j|} \in L^\alpha(\Omega, \mathcal{F}, \mathbb{P})$ for $j = 1, \ldots, n$, then:*

(a) *$P(X_1, \ldots, X_n) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for any polynomial $P \in \mathcal{P}_n$ and $p \in [1, \infty)$;*

(b) *the vector space $\{P(X_1, \ldots, X_n) \,:\, P \in \mathcal{P}_n\}$ is dense in $L^p(\Omega, \mathcal{F}, \mathbb{P})$ for every $p \in [1, \infty)$.*

It should be noted that the converse to part (a) is false as shown in the following example.

Example 2.3. Define a probability measure $\mathbb{P}$ on the real axis via its density

$$f(x) = \frac{\sum_{m=1}^{\infty} \frac{\delta(x-m)}{m^{\ln m}}}{\sum_{m=1}^{\infty} \frac{1}{m^{\ln m}}},$$

where $\delta = \mathbf{1}_{\{0\}}$. If $q \geq 1$, then

$$\sum_{m=1}^{\infty} \frac{m^q}{m^{\ln m}} < \infty.$$

On the other hand

$$\sum_{m=1}^{\infty} \frac{e^{\alpha m}}{m^{\ln m}} = \infty$$

for any $\alpha > 0$. ■

If the probability measure $\mathbb{P}$ has a bounded support, in $\mathbb{R}^n$, then the assumption of the Dobrushin-Minlos theorem is trivially satisfied. In fact, in this special case the conclusion of the theorem follows directly from the Stone-Weierstrass Theorem. It is also easy to see that if X is Gaussian, then $e^{|X|} \in L^1$. However, if X is lognormal, then its moment generating function does not exist in the interval $(0, \infty)$ and hence $e^{\alpha|X|} \notin L^\alpha$ for all $\alpha > 0$.

In concrete applications, the condition $e^{|X|} \in L^\alpha$ can sometimes be achieved by changing the probability distribution of "very large" values of $|X|$. For instance, this can be accomplished by truncation of probability distribution or some direct attenuation of the random variable X. Another possibility is the use of suitable weight functions. In this context the Dubrushin-Minlos theorem can be used to justify the density part in the construction of several classic polynomial bases in spaces of square integrable functions, associated with the names of Jacobi, Gagenbauer, Legendre, Chebyshev, Laguerre and Hermite (see e.g. [13]).

Let V be an information space generated by random variables $X_1, \ldots, X_n$. Suppose that one can furnish a sequence of Borel functions $q_m : \mathbb{R}^n \longrightarrow \mathbb{R}$, with $m \in \mathbb{N}$, such that the set $\{q_m(X_1, \ldots, X_n) : m \in \mathbb{N}\}$ is linearly dense in V (e.g. with the help of the Dobrushin-Minlos theorem). Then the conditional expectation operator $\mathrm{E}[\cdot \,|\, X_1, \ldots, X_n]$ is the pointwise limit of the sequence of projections onto linear spaces $V^m = \{q_k(X_1, \ldots, X_n) : 1 \leq k \leq m\}$ as $m \nearrow \infty$. This observation leads to an auxiliary concept of admissible projection systems.

Given a discrete time filtration $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_T \subset \mathcal{F}$ in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we define an *admissible projection system* as a family of orthogonal projections

$$\left\{ P_t^m \,:\, L^2(\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow L^2(\Omega, \mathcal{F}, \mathbb{P}) \right\}_{\substack{t = 1, \ldots, T \\ m \in \mathbb{N}}}$$

with ranges $V_t^m = P_t^m \left(L^2(\Omega, \mathcal{F}, \mathbb{P}) \right)$, such that for all $t = 1, \ldots, T$ and $m \in \mathbb{N}$ we have

$$V_t^m \subset V_t^{m+1}$$

and

$$\overline{\bigcup_{k \in \mathbb{N}} V_t^k} = L^2(\Omega, \mathcal{F}_t, \mathbb{P}).$$

Note that for any such system and for any fixed t, we get pointwise convergence of the projections P_t^m to $\mathrm{E}[\cdot \,|\, \mathcal{F}_t]$. However, this is not a norm convergence unless the underlying sequence of subspaces becomes constant after finitely many steps.

Suppose that (Z_t) is a stochastic process adapted to the filtration $(\mathcal{F}_t)$. Let us fix $m \in \mathbb{N}$. Given an admissible projection system we define the stopping times $\tau_t^{[m]}$ by recursion:

$$
\begin{aligned}
\tau_T^{[m]} &= T, \\
\tau_t^{[m]} &= t\mathbf{1}_{\{Z_t \geq P_t^m(Z_{\tau_{t+1}^{[m]}})\}} + \tau_{t+1}^{[m]}\mathbf{1}_{\{Z_t < P_t^m(Z_{\tau_{t+1}^{[m]}})\}}, \qquad t = 1, \dots, T-1.
\end{aligned}
$$

Recall, that in our discussion of properties of Snell envelopes we defined a somewhat similar stopping time:

$$
\begin{aligned}
\tau_T &= T, \\
\tau_t &= t\mathbf{1}_{\{Z_t \geq E[Z_{\tau_{t+1}} \mid \mathcal{F}_t]\}} + \tau_{t+1}\mathbf{1}_{\{Z_t < E[Z_{\tau_{t+1}} \mid \mathcal{F}_t]\}}, \qquad t = 1, \dots, T-1.
\end{aligned}
$$

The following theorem generalizes a result due to Clément, Lamberton and Protter (see Theorem 3.1 in [6]):

Theorem 2.4. *If (P_t^m) is an admissible projection system, then*

$$
\lim_{m \to \infty} \mathrm{E}\left[Z_{\tau_t^{[m]}} \mid \mathcal{F}_t\right] = \mathrm{E}[Z_{\tau_t} \mid \mathcal{F}_t]
$$

for $t = 1, \dots, T$, where the convergence is in L^2. In particular

$$
\lim_{m \to \infty} \mathrm{E}\left[Z_{\tau_t^{[m]}}\right] = \mathrm{E}[Z_{\tau_t}]
$$

in L^2.

Proof: Despite a much more general setting we have adopted here and slightly different notation, we can proceed as in [6]. Since the case $t = T$ is obvious, we can use induction on t. Assume that the formula is true for $t+1$. Let $\mathrm{E}_t[\,\cdot\,] = \mathrm{E}[\,\cdot \mid \mathcal{F}_t]$. Define five subsets of Ω as collections of points satisfying the following inequalities:

$$
\begin{aligned}
C_1 &= \{Z_t \geq P_t^m(Z_{\tau_{t+1}^{[m]}})\} \quad , \quad C_2 = \Omega \setminus C_1, \\
C_3 &= \{\ Z_t \geq \mathrm{E}_t[Z_{\tau_{t+1}}]\ \} \quad , \quad C_4 = \Omega \setminus C_3, \\
C_5 &= \left\{ \left|Z_t - \mathrm{E}_t[Z_{\tau_{t+1}}]\right| \ \leq \ \left|\mathrm{E}_t[Z_{\tau_{t+1}}] - P_t^m(Z_{\tau_{t+1}^{[m]}})\right| \right\}.
\end{aligned}
$$

Obviously, for $t < T$ we we have the formulas

$$
\begin{aligned}
\tau_t^{[m]} &= t\mathbf{1}_{C_1} + \tau_{t+1}^{[m]}\mathbf{1}_{C_2}, \\
\tau_t &= t\mathbf{1}_{C_3} + \tau_{t+1}\mathbf{1}_{C_4}.
\end{aligned}
$$

Observe that

$$
\begin{aligned}
\mathrm{E}_t\left[Z_{\tau_t^{[m]}} - Z_{\tau_t}\right] &= \mathrm{E}_t\left[Z_t\mathbf{1}_{C_1} + Z_{\tau_{t+1}^{[m]}}\mathbf{1}_{C_2} - Z_t\mathbf{1}_{C_3} - Z_{\tau_{t+1}}\mathbf{1}_{C_4}\right] \\
&= Z_t(\mathbf{1}_{C_1} - \mathbf{1}_{C_3}) + \mathrm{E}_t\left[Z_{\tau_{t+1}^{[m]}}\right]\mathbf{1}_{C_2} - \mathrm{E}_t\left[Z_{\tau_{t+1}}\right]\mathbf{1}_{C_4} \\
&= Z_t(\mathbf{1}_{C_1} - \mathbf{1}_{C_3}) + \mathrm{E}_t\left[Z_{\tau_{t+1}^{[m]}}\right]\mathbf{1}_{C_2} - \mathrm{E}_t[Z_{\tau_{t+1}}](\mathbf{1}_{C_1} + \mathbf{1}_{C_2} - \mathbf{1}_{C_3}) \\
&= \mathrm{E}_t\left[Z_{\tau_{t+1}^{[m]}} - Z_{\tau_{t+1}}\right]\mathbf{1}_{C_2} + \left(Z_t - \mathrm{E}_t\left[Z_{\tau_{t+1}}\right]\right)(\mathbf{1}_{C_1} - \mathbf{1}_{C_3}) \\
&= \mathrm{E}_t\left[Z_{\tau_{t+1}^{[m]}} - Z_{\tau_{t+1}}\right]\mathbf{1}_{C_2} + L_t^m.
\end{aligned}
$$

The second last term goes to zero by the induction hypothesis and the fact that $\mathrm{E}_t\mathrm{E}_{t+1} = \mathrm{E}_t$. We need to estimate the last term. To this end note that

$$
|\mathbf{1}_{C_1} - \mathbf{1}_{C_3}| \le |\mathbf{1}_{C_1\cap C_4} - \mathbf{1}_{C_2\cap C_3}| \le \mathbf{1}_{C_5},
$$

because $(C_1 \cap C_4) \cup (C_2 \cap C_3) \subset C_5$. Hence

$$
\begin{aligned}
L_t^m &\le \left|Z_t - \mathrm{E}_t\left[Z_{\tau_{t+1}}\right]\right|\mathbf{1}_{C_5} \\
&\le \left|\mathrm{E}_t[Z_{\tau_{t+1}}] - P_t^m(Z_{\tau_{t+1}^{[m]}})\right|, \text{ by the definition of } C_5, \\
&\le \left|\mathrm{E}_t[Z_{\tau_{t+1}}] - P_t^m\left(\mathrm{E}_t[Z_{\tau_{t+1}}]\right)\right| + \left|P_t^m\left(\mathrm{E}_t[Z_{\tau_{t+1}}]\right) - P_t^m(Z_{\tau_{t+1}^{[m]}})\right| \\
&= \left|\mathrm{E}_t[Z_{\tau_{t+1}}] - P_t^m\left(\mathrm{E}_t[Z_{\tau_{t+1}}]\right)\right| + \left|P_t^m\left(\mathrm{E}_t[Z_{\tau_{t+1}}]\right) - P_t^m(\mathrm{E}_t[Z_{\tau_{t+1}^{[m]}}])\right| \\
&\le \left|\mathrm{E}_t[Z_{\tau_{t+1}}] - P_t^m\left(\mathrm{E}_t[Z_{\tau_{t+1}}]\right)\right| + \left|\mathrm{E}_t\left[Z_{\tau_{t+1}} - Z_{\tau_{t+1}^{[m]}}\right]\right|,
\end{aligned}
$$

because of the tower property of projections and the fact that the norm of a projection is at most one. The last term goes to zero by the induction hypothesis. The second last one because of the L^2 density of the union of ranges of the projection forming the admissible projection system. ■

Obviously the above considerations remain valid for vector valued stochastic processes.

3 The general case of the least squares method of option pricing

In what follows we will denote the set of all real $(m \times n)$-matrices by $\mathbb{R}^{m\times n}$ with the convention that $\mathbb{R}^m = \mathbb{R}^{1\times m}$. Throughout the section we will use notation and methods similar to those introduced in [6] but adapted to our less restrictive assumptions.

Suppose that $(X_t)_{t=0}^T$ is a discrete time d-dimensional stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with X_0 being a constant. This process is meant to represent the prices of the underlying assets for an American style option we wish to valuate.

Let

$$X = (X_1, \ldots, X_T) : \Omega \longrightarrow \mathbb{R}^{d\times T}$$

and let $\mathcal{F}_t = \sigma(X_0, \ldots, X_t) = \sigma(X_1, \ldots, X_t)$ for $t = 1, \ldots, T$. Given a family of Borel functions

$$f_t : \mathbb{R}^{d\times(t+1)} \longrightarrow \mathbb{R}_+, \qquad t = 0, \ldots, T,$$

we define

$$Z_t = f_t(X_0, \ldots, X_t), \qquad t = 0, \ldots, T.$$

This sequence represents suitably discounted intrinsic prices of the option we want to consider. Such a general choice of functions f_t, expands the potential applicability well beyond American put options. The aim is to calculate U_0, where U_t is the Snell envelope of Z_t and since $U_0 = \max(Z_0, \mathrm{E}[Z_{\tau_1}])$, we basically want to approximate numerically $\mathrm{E}[Z_{\tau_1}]$.

We need to chose an admissible projection system for the filtration associated with X. This is equivalent to choosing for each $t \in \{1, \ldots, T\}$ a suitable sequence of Borel functions

$$q_t^k : \mathbb{R}^{d\times T} \longrightarrow \mathbb{R}, \qquad k \in \mathbb{N},$$

which depend only on the first t column variables, and are such that the sequence $\{q_t^k(X)\}_{k\in\mathbb{N}}$ is linearly dense and linearly independent in the space $L^2(\Omega, \sigma(X_1, \ldots, X_t), \mathbb{P})$. Then, we can select an increasing sequence of integers $(k_m)_{m\in\mathbb{N}}$, such that the spaces

$$V_t^m = \mathrm{Lin}\{q_t^k(X) \,:\, k = 1, \ldots, k_m\},$$

and the orthogonal projections $P_t^m : L^2(\Omega, \sigma(X), \mathbb{P}) \longrightarrow V_t^m$ have all the right properties. The symbol "Lin" denotes the linear envelope of the given set of vectors.

If the stopping times $\tau^{[m]}$ are defined as in the previous section, then for some $\alpha_t^m \in \mathbb{R}^{k_m\times 1}$ we have

$$P_t^m\left(Z_{\tau_{t+1}^{[m]}}\right) = e_t^m(X)\, \alpha_t^m,$$

where the mapping e_t^m is given by the formula

$$e_t^m : \mathbb{R}^{d\times T} \ni x \mapsto \left(q_t^1(x), \ldots, q_t^{k_m}(x)\right) \in \mathbb{R}^{k_m}.$$

In view of our assumptions, the Gram matrix of the components of $e_t^m(X)$ (with respect to the inner product $(Y_1, Y_2) \mapsto \mathrm{E}[Y_1 Y_2]$), that is the matrix

$$A_t^m = \left[\mathrm{E}\left[q_t^i(X) q_t^j(X)\right]\right]_{1\le i,j\le k_m} \in \mathbb{R}^{k_m\times k_m},$$

is invertible and hence

$$\alpha_t^m = (A_t^m)^{-1} \begin{bmatrix} \mathrm{E}\left[Z_{\tau_{t+1}^{[m]}} \, q_t^1(X) \right] \\ \vdots \\ \mathrm{E}\left[Z_{\tau_{t+1}^{[m]}} \, q_t^{k_m}(X) \right] \end{bmatrix}.$$

Given a number N, the next step it to use Monte-Carlo simulation to generate independent trajectories

$$X^{(n)} = \left(X_1^{(n)}, \ldots, X_T^{(n)} \right) \in \mathbb{R}^{d \times T}$$

of the process X, for $n = 1, 2, \ldots, N$. Each simulation has the fixed starting point $X_0^{(n)} = X_0 \in \mathbb{R}^{d \times 1}$.

Define

$$Z_t^{(n)} := f_t \left(X_0^{(n)}, \ldots, X_t^{(n)} \right)$$

and let

$$\widehat{Z}_t = \begin{bmatrix} Z_t^{(1)} \\ \vdots \\ Z_t^{(N)} \end{bmatrix} \in \mathbb{R}^{N \times 1}.$$

This column vector consists simply of the values at time t of all simulated trajectories of the process Z.

Define also

$$V_t^{(m,N)} = \mathrm{Lin} \left\{ \begin{bmatrix} q_t^k(X^{(1)}) \\ \vdots \\ q_t^k(X^{(N)}) \end{bmatrix} : k = 1, \ldots, k_m \right\} \subset \mathbb{R}^{N \times 1}$$

and

$$P_t^{(m,N)} = \mathrm{Proj}_{V_t^{(m,N)}} : \mathbb{R}^{N \times 1} \longrightarrow \mathbb{R}^{N \times 1}$$

with respect to the inner product $\frac{\langle x, y \rangle}{N}$, where $\langle x, y \rangle$ denotes the standard scalar product.

Note that

$$\begin{bmatrix} e_t^m(X^{(1)}) \\ \vdots \\ e_t^m(X^{(N)}) \end{bmatrix} \in \mathbb{R}^{N \times k_m}$$

and

$$V_t^{(m,N)} = \mathrm{Lin} \left\{ \text{the columns of } \begin{bmatrix} e_t^m(X^{(1)}) \\ \vdots \\ e_t^m(X^{(N)}) \end{bmatrix} \right\} \subset \mathbb{R}^{N \times 1}.$$

If we define the stopping times $\tau_t^{[m]}$ by the formula

$$\begin{aligned}
\tau_T^{[m]} &= T, \\
\tau_t^{[m]} &= t\mathbf{1}_{\{Z_t \geq P_t^m(Z_{\tau_{t+1}^{[m]}})\}} + \tau_{t+1}^{[m]}\mathbf{1}_{\{Z_t < P_t^m(Z_{\tau_{t+1}^{[m]}})\}}, \qquad t = 1, \ldots, T-1,
\end{aligned}$$

then for some $\alpha_t^m \in \mathbb{R}^{k_m \times 1}$ we have

$$P_t^m\left(Z_{\tau_{t+1}^{[m]}}\right) = e_t^m(X)\alpha_t^m.$$

Similarly, if we define the approximative stopping times $\tau_t^{n,m,N}$ by the formula

$$\begin{aligned}
\tau_T^{n,m,N} &= T, \\
\tau_t^{n,m,N} &= t\mathbf{1}_{\left\{Z_t^{(n)} \geq \pi_n\left[P_t^{(m,N)}(\widehat{Z}_{\tau_{t+1}^{n,m,N}})\right]\right\}} + \tau_{t+1}^{n,m,N}\mathbf{1}_{\left\{Z_t^{(n)} < \pi_n\left[P_t^{(m,N)}(\widehat{Z}_{\tau_{t+1}^{n,m,N}})\right]\right\}}, \\
&\quad \text{for } t = 1, \ldots, T-1,
\end{aligned}$$

where

$$\pi_n : \mathbb{R}^{N\times 1} \ni \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \mapsto x_n \in \mathbb{R},$$

then for some $\alpha_t^{(m,N)} \in \mathbb{R}^{k_m \times 1}$ we have

$$P_t^{(m,N)}\left(\begin{bmatrix} Z^{(1)}_{\tau_{t+1}^{1,m,N}} \\ \vdots \\ Z^{(N)}_{\tau_{t+1}^{N,m,N}} \end{bmatrix}\right) = \begin{bmatrix} e_t^m(X^{(1)}) \\ \vdots \\ e_t^m(X^{(N)}) \end{bmatrix} \alpha_t^{(m,N)}.$$

Let $A_t^{(m,N)}$ be the $(k_m \times k_m)$-Gram matrix associated with the columns of the matrix

$$\begin{bmatrix} e_t^m(X^{(1)}) \\ \vdots \\ e_t^m(X^{(N)}) \end{bmatrix},$$

that is

$$A_t^{(m,N)} = \frac{1}{N}\begin{bmatrix} e_t^m(X^{(1)}) \\ \vdots \\ e_t^m(X^{(N)}) \end{bmatrix}^* \begin{bmatrix} e_t^m(X^{(1)}) \\ \vdots \\ e_t^m(X^{(N)}) \end{bmatrix}.$$

This is simply the Gram matrix estimator for the given sample.

Then $\alpha_t^{(m,N)}$ is a solution of the equation

$$A_t^{(m,N)}\alpha_t^{(m,N)} = \frac{1}{N}\begin{bmatrix} e_t^m(X^{(1)}) \\ \vdots \\ e_t^m(X^{(N)}) \end{bmatrix}^* \begin{bmatrix} Z^{(1)}_{\tau^{1,m,N}_{t+1}} \\ \vdots \\ Z^{(N)}_{\tau^{N,m,N}_{t+1}} \end{bmatrix}.$$

By the Law of Large Numbers $A_t^{(m,N)} \xrightarrow{a.s.} A_t^m$ as $N \to \infty$, and hence for sufficiently large N the matrix $A_t^{(m,N)}$ is invertible (almost surely). In this case

$$\alpha_t^{(m,N)} = \frac{1}{N}\left(A_t^{(m,N)}\right)^{-1}\begin{bmatrix} e_t^m(X^{(1)}) \\ \vdots \\ e_t^m(X^{(N)}) \end{bmatrix}^* \begin{bmatrix} Z^{(1)}_{\tau^{1,m,N}_{t+1}} \\ \vdots \\ Z^{(N)}_{\tau^{N,m,N}_{t+1}} \end{bmatrix}.$$

For convenience we will write

$$\alpha^m = \left(\alpha_1^m, \ldots, \alpha_{T-1}^m\right), \qquad \alpha^{(m,N)} = \left(\alpha_1^{(m,N)}, \ldots, \alpha_{T-1}^{(m,N)}\right).$$

Both objects are $k_m \times (T-1)$-matrices.

Define

$$B_t = \{(a^m, z, x) \ : \ z_t < e_t^m(x)a_t^m\} \subset \mathbb{R}^{k_m \times (T-1)} \times \mathbb{R}^T \times \mathbb{R}^{d\times T}$$

for $t = 1, \ldots, T-1$, where $a^m = (a_1^m, \ldots, a_{T-1}^m)$, $z = (z_1, \ldots, z_T)$, and $x = (x_1, \ldots, x_T)$. By B_t^c we will denote the complement of B_t. We define an auxiliary function

$$F_t : \mathbb{R}^{k_m\times(T-1)} \times \mathbb{R}^T \times \mathbb{R}^{d\times T} \longrightarrow \mathbb{R},$$

by recursion:

$$\begin{aligned} F_T(a^m, z, x) &= z_T, \\ F_t(a^m, z, x) &= z_t \mathbf{1}_{B_t^c} + F_{t+1}(a^m, z, x)\mathbf{1}_{B_t}, \qquad t = 1, \ldots, T-1. \end{aligned}$$

Since $\mathbf{1}_{C\cap D} = \mathbf{1}_C \mathbf{1}_D$ for any two sets C and D, it is easy to see that

$$F_t(a^m, z, x) = z_t \mathbf{1}_{B_t^c} + \sum_{s=t+1}^{T-1} z_s \mathbf{1}_{B_t\cap\ldots\cap B_{s-1}\cap B_s^c} + z_T \mathbf{1}_{B_t\cap\ldots\cap B_{T-1}}$$

for $t = 1, \ldots, T-1$. Moreover

$$\begin{aligned} F_t(a^m, z, x) &\quad \text{is} \quad \text{independent of } a_1^m, \ldots, a_{t-1}^m; \\ F_t(\alpha^m, Z, X) &= Z_{\tau_t^{[m]}}; \\ F_t(\alpha^{(m,N)}, Z^{(n)}, X^{(n)}) &= Z^{(n)}_{\tau_t^{n,m,N}}. \end{aligned}$$

For $t = 2, \ldots, T$ define also three other auxiliary functions:

$$\begin{aligned} G_t(a^m, z, x) &= F_t(a^m, z, x)e^m_{t-1}(x); \\ \phi_t(a^m) &= \mathrm{E}[F_t(a^m, Z, X)]; \\ \psi_t(a^m) &= \mathrm{E}[G_t(a^m, Z, X)]. \end{aligned}$$

Using this notation one can see that for $t = 1, \ldots, T-1$:

$$\alpha^m_t = (A^m_t)^{-1}\psi_{t+1}(\alpha^m); \tag{1}$$

$$\alpha^{(m,N)}_t = (A^{(m,N)}_t)^{-1}\frac{1}{N}\sum_{n=1}^{N} G_{t+1}(\alpha^{(m,N)}, Z^{(n)}, X^{(n)}). \tag{2}$$

The following estimate is a higher-dimensional counterpart of Lemma 3.1 in [6].

$$|F_t(a, z, x) - F_t(\tilde{a}, z, x)| \le \sum_{s=t}^{T} |z_s| \left[\sum_{s=t}^{T-1} \mathbf{1}_{\{|z_s - e^m_s(x)\tilde{a}_s| \le |e^m_s(x)| \|\tilde{a}_s - a_s\|\}}\right], \tag{3}$$

where $1 \le t \le T-1$, $a = (a_1, \ldots, a_{T-1}) \in \mathbb{R}^{k_m \times (T-1)}$, $\tilde{a} = (\tilde{a}_1, \ldots, \tilde{a}_{T-1}) \in \mathbb{R}^{k_m \times (T-1)}$, $z \in \mathbb{R}^T$ and $x \in \mathbb{R}^{d \times T}$.

The above estimate can be easily justified. Let $\tilde{B}_t = \{z_t < e^m_t(x)\tilde{a}_t\}$. Note first that

$$\begin{aligned} |\mathbf{1}_{B_t} - \mathbf{1}_{\tilde{B}_t}| &= \mathbf{1}_{B^c_t \cap \tilde{B}_t} + \mathbf{1}_{B_t \cap \tilde{B}^c_t} \\ &\le \mathbf{1}_{\{|z_t - e^m_t(x)\tilde{a}_t| \le |e^m_t(x)| \|\tilde{a}_t - a_t\|\}}. \end{aligned}$$

Moreover

$$\begin{aligned} |\mathbf{1}_{A_1 \cap A_2} - \mathbf{1}_{C_1 \cap C_2}| &= |\mathbf{1}_{A_1}\mathbf{1}_{A_2} - \mathbf{1}_{C_1}\mathbf{1}_{C_2}| \\ &= |\mathbf{1}_{A_1}\mathbf{1}_{A_2} - \mathbf{1}_{A_1}\mathbf{1}_{C_2} + \mathbf{1}_{A_1}\mathbf{1}_{C_2} - \mathbf{1}_{C_1}\mathbf{1}_{C_2}| \\ &\le \mathbf{1}_{A_1}|\mathbf{1}_{A_2} - \mathbf{1}_{C_2}| + \mathbf{1}_{C_2}|\mathbf{1}_{A_1} - \mathbf{1}_{C_1}| \\ &\le |\mathbf{1}_{A_1} - \mathbf{1}_{C_1}| + |\mathbf{1}_{A_2} - \mathbf{1}_{C_2}|, \end{aligned}$$

for any A_1, A_2, C_1, C_2. Consequently

$$\begin{aligned} |\mathbf{1}_{B_t \cap \ldots \cap B_{s-1} \cap B^c_s} - \mathbf{1}_{\tilde{B}_t \cap \ldots \cap \tilde{B}_{s-1} \cap \tilde{B}^c_s}| &\le \sum_{u=t}^{s-1} |\mathbf{1}_{B_u} - \mathbf{1}_{\tilde{B}_u}| + |\mathbf{1}_{B^c_s} - \mathbf{1}_{\tilde{B}^c_s}| \\ &= \sum_{u=t}^{s} |\mathbf{1}_{B_u} - \mathbf{1}_{\tilde{B}_u}|, \end{aligned}$$

because $|\mathbf{1}_{B^c_s} - \mathbf{1}_{\tilde{B}^c_s}| = \mathbf{1}_{B^c_s \Delta \tilde{B}^c_s} = \mathbf{1}_{B_s \Delta \tilde{B}_s} = |\mathbf{1}_{B_u} - \mathbf{1}_{\tilde{B}_u}|$, where Δ denotes the symmetric difference of sets. Similarly

$$|\mathbf{1}_{B_t \cap \ldots \cap B_{T-1}} - \mathbf{1}_{\tilde{B}_t \cap \ldots \cap \tilde{B}_{T-1}}| \le \sum_{u=t}^{T-1} |\mathbf{1}_{B_u} - \mathbf{1}_{\tilde{B}_u}|,$$

Therefore

$$
\begin{aligned}
|F_t(a,z,x) - F_t(\tilde{a},z,x)| &= \Big| z_t(\mathbf{1}_{B_t} - \mathbf{1}_{\tilde{B}_t}) \\
&\quad + \sum_{s=t+1}^{T-1} z_s(\mathbf{1}_{B_t\cap\ldots\cap B_{s-1}\cap B_s^c} - \mathbf{1}_{\tilde{B}_t\cap\ldots\cap\tilde{B}_{s-1}\cap\tilde{B}_s^c}) \\
&\quad + z_T(\mathbf{1}_{B_t\cap\ldots\cap B_{T-1}} - \mathbf{1}_{\tilde{B}_t\cap\ldots\cap\tilde{B}_{T-1}}) \Big| \\
&\leq \left(\sum_{s=t}^{T} |z_s|\right)\left(\sum_{s=t}^{T-1} |\mathbf{1}_{B_s} - \mathbf{1}_{\tilde{B}_s}|\right),
\end{aligned}
$$

as needed.

The next theorem is a direct extension of Theorem 3.2, Lemma 3.2 and their proofs from [6].

Theorem 3.1. *With the above notation and assumptions*

$$\frac{1}{N}\sum_{n=1}^{N} Z^{(n)}_{\tau_t^{n,m,N}} \xrightarrow{a.s.} \mathrm{E}\left[Z_{\tau_t^{[m]}}\right], \text{ as } N \to \infty,$$

for $t = 1, \ldots, T$, provided that

$$\mathbb{P}(e_t^m(X)\alpha_t^m = Z_t) = 0.$$

Proof: First we prove the following:

Claim 1: $\alpha_t^{(m,N)} \xrightarrow{a.s.} \alpha_t^m$ *as* $N \to \infty$ *for* $t = 1, \ldots, T-1$.

We know that $A_t^{(m,N)} \xrightarrow{a.s.} A_t^{(m)}$ because of the Law of Large Numbers. Hence, in view of (1) and (2), we need to prove that:

$$\frac{1}{N}\sum_{n=1}^{N} G_t(\alpha^{(m,N)}, Z^{(n)}, X^{(n)}) \xrightarrow{a.s.} \psi_t(\alpha^{(m)}).$$

We use induction on t starting at $T-1$. For $t = T-1$, we have $G_{t+1}(a^m, z, x) = z_T e_{T-1}^m(x)$, so the statement is true as the Law of Large Numbers implies that

$$\frac{1}{N}\sum_{n=1}^{N} Z_T^{(n)} e_t^m(X^{(n)}) \xrightarrow{a.s.} \mathrm{E}[Z_T e_T^m(X)],$$

which is what we need. Assume that the statement is true for t. The Law of Large Numbers implies that

$$\frac{1}{N}\sum_{n=1}^{N} G_t(\alpha^m, Z^{(n)}, X^{(n)}) \xrightarrow{a.s.} \psi_t(\alpha^m),$$

so it suffices to prove that $\lim_{N\to\infty} G_N = 0$, where

$$G_N = \frac{1}{N}\sum_{n=1}^{N}\left(G_t(\alpha^{(m,N)}, Z^{(n)}, X^{(n)}) - G_t(\alpha^m, Z^{(n)}, X^{(n)})\right).$$

We have

$$\begin{aligned} |G_N| &\le \frac{1}{N}\sum_{n=1}^{N} |e_{t-1}^m(X^{(n)})||F_t(\alpha^{(m,N)}, Z^{(n)}, X^{(n)}) - F_t(\alpha^m, Z^{(n)}, X^{(n)})| \\ &\le \frac{1}{N}\sum_{n=1}^{N} |e_{t-1}^m(X^{(n)})| \left(\sum_{s=t}^{T} |Z_s^{(n)}|\right)\left(\sum_{s=t}^{T-1} \mathbf{1}_{W_I(s,N)}\right), \end{aligned}$$

where

$$W_I(s,N) = \{|Z_s^{(n)} - \alpha_s^m e_s^m(X^{(n)})| \le |\alpha_s^{(m,N)} - \alpha_s^m||e_s^m(X^{(n)})|\}.$$

For $s = t, \ldots, T-1$

$$\alpha_s^{(m,N)} \xrightarrow{a.s.} \alpha_s^m, \qquad N \to \infty.$$

Let

$$\begin{aligned} W_{II}(s,\epsilon) &= \{\{|Z_s^{(n)} - \alpha_s^m e_s^m(X^{(n)})| \le \epsilon|e_s^m(X^{(n)})|\}\}, \\ W_{III}(s,\epsilon) &= \{|Z_s - \alpha_s^m e_s^m(X)| \le \epsilon|e_s^m(X)|\}. \end{aligned}$$

So $\forall\ \epsilon > 0$

$$\begin{aligned} \limsup |G_N| &\overset{a.s.}{\le} \limsup \frac{1}{N}\sum_{n=1}^{N}\left[|e_{t-1}^m(X^{(n)})|\left(\sum_{s=t}^{T}|Z_s^{(n)}|\right)\left(\sum_{s=t}^{T-1}\mathbf{1}_{W_{II}(s,\epsilon)}\right)\right] \\ &\overset{a.s.}{=} \mathrm{E}\left[|e_{t-1}^m(X)|\left(\sum_{s=t}^{T}|Z_s^{(n)}|\right)\left(\sum_{s=t}^{T-1}\mathbf{1}_{W_{III}(s,\epsilon)}\right)\right], \end{aligned}$$

The last equality follows from the Law of Large Numbers. If $\epsilon \to 0$, we get convergence to zero, because of the probability assumption: if $A, B, Y \ge 0$ and $\mathbb{P}(A = 0) = 0$, then as $\epsilon \searrow 0$

$$\int_{\{A\le\epsilon B\}} Y d\mathbb{P} \searrow \int_{\bigcap_{\epsilon>0}\{A\le\epsilon B\}} Y d\mathbb{P} = \int_{\{A=0\}} Y d\mathbb{P} = 0.$$

Claim 2:

$$\frac{1}{N}\sum_{n=1}^{N} F_t(\alpha^{(m,N)}, Z^{(n)}, X^{(n)}) \xrightarrow{a.s.} \phi_t(\alpha^{(m)}).$$

As before we use induction on t starting at $T-1$. For $t = T-1$, we have $F_{t+1}(a^m, z, x) = z_T$, so the statement is true as the the Law of Large Numbers implies that

$$\frac{1}{N}\sum_{n=1}^{N} Z_T^{(n)} \xrightarrow{a.s.} \mathrm{E}[Z_T],$$

which is what we need. Assume that the statement is true for t. The Law of Large Numbers implies that

$$\frac{1}{N}\sum_{n=1}^{N} F_t(\alpha^m, Z^{(n)}, X^{(n)}) \xrightarrow{a.s.} \phi_t(\alpha^m),$$

so it suffices to prove that $\lim_{N\to\infty} F_N = 0$, where

$$F_N = \frac{1}{N}\sum_{n=1}^{N} \left(F_t(\alpha^{(m,N)}, Z^{(n)}, X^{(n)}) - F_t(\alpha^m, Z^{(n)}, X^{(n)}) \right).$$

We have

$$\begin{aligned} |F_N| &\leq \frac{1}{N}\sum_{n=1}^{N} |F_t(\alpha^{(m,N)}, Z^{(n)}, X^{(n)}) - F_t(\alpha^m, Z^{(n)}, X^{(n)})| \\ &\leq \frac{1}{N}\sum_{n=1}^{N} \left(\sum_{s=t}^{T} |Z_s^{(n)}|\right)\left(\sum_{s=t}^{T-1} \mathbf{1}_{W_I(s,N)}\right). \end{aligned}$$

Now for any $\epsilon > 0$

$$\begin{aligned} \limsup |F_N| &\overset{a.s.}{\leq} \limsup \frac{1}{N}\sum_{n=1}^{N} \left[\left(\sum_{s=t}^{T} |Z_s^{(n)}|\right)\left(\sum_{s=t}^{T-1} \mathbf{1}_{W_{II}(s,\epsilon)}\right)\right] \\ &\overset{a.s.}{=} \mathrm{E}\left[\left(\sum_{s=t}^{T} |Z_s^{(n)}|\right)\left(\sum_{s=t}^{T-1} \mathbf{1}_{W_{III}(s,\epsilon)}\right)\right], \end{aligned}$$

The last equality follows from the Law of Large Numbers. If $\epsilon \to 0$, we get convergence to zero, which is precisely what the conclusion of the theorem asserts. ■

Theorems 2.4 and 3.1 provide a recipe for approximation of $\mathrm{E}[Z_{\tau_1}]$ and hence also $U_0 = \max(Z_0, \mathrm{E}[Z_{\tau_1}])$, as required.

4 Examples

In this section we will show three examples of applications of the above least-squares algorithm. The first example will cover American call and put options for Eurodollar futures assuming dynamics from the Brace-Gątarek-Musiela model [2]. Next we will price basket and dual-strike American put options for EUROSTOXX50 and DAX indices under the standard bivariate Brownian dynamics. Finally we will show how to price univariate American put options both for EUROSTOXX50 and DAX indices assuming that their dynamics could be expressed using the Heston-Nandi GARCH(1,1) model [11].

We have chosen to use standard and classic models for parameter estimation, Monte Carlo simulation, etc. to keep the examples transparent. In particular we will use only the prices of the underlyings for calibration purposes and will not use any Monte Carlo variance reduction techniques (for more advanced models cf. [1] or [9] and references therein).

The implementation of the least squares algorithm was based on the in-the-money realizations to speed up the convergence and reduce the number of polynomials needed to achieve satisfactory accuracy (see [16] for discussion). We have also normalized the prices (again, see [16] for details) to increase the accuracy. We have used weighted Laguerre polynomials in univariate case and weighted polynomial base in bivariate case.

All computations were done using **R 2.15.2** (64-bit). In particular we have used the libraries **fOptions** (for Heston-Nandi parameter calibration, CRR prices and Monte Carlo simulation), **orthopolynom** (for different base functions in L-S algorithm), **timeSeries** (for market data handling) and **Rsge** (for parallel computations).

4.1 Eurodollar options

In this subsection we will use the least squares algorithm to price a one year Eurodollar American put and call options with different strike prices, given the real-market daily prices of the Eurodollar futures. It should be noted, that the standard Black-Scholes model cannot be used when the option price is based on more than one LIBOR rate (e.g. when the option's lifetime is longer than 3 months). This is due to the fact, that forward rates over consecutive time intervals are related to each other and cannot all be log-normal under the same spot risk-neutral measure. Consequently, models of such instruments under the standard risk-neutral measure are based on non-Markovian dynamics. A. Brace, D. Gątarek and M. Musiela [2] proposed a model (BGM Model) based on the forward arbitrage-free risk-neutral measure which could overcome this inconvenience. In the literature it is also referred to as the LIBOR Market Model (LMM). It is worth mentioning that the dynamics of interest rates described in BGM model is very closely related to the Heath-Jarrow-Morton (HJM) Model. To begin with, let us recall some basic information about Eurodollar futures, Eurodollar options, BGM Model and the setup of the least squares algorithm.

Eurodollar futures

Eurodollar futures (also called LIBOR futures), are basically instruments that reflect the 3-month LIBOR rate for $1 million offshore deposit maturing at some point in the future (regulated by the exchange). Those instruments are very similar to the standard forward rate agreement contracts, with the exception that their terms are strictly regulated by the Chicago Mercantile Exchange (CME). A single contract is constructed in such a way that 1 basis point movement results in \$25 payoff. While the Minimum Price Fluctuation is also regulated, we will omit its description because it has no significant impact on the price of the option (see *www.cmegroup.com* for details). The contract expiration dates are March, June, September and December extending out for 10 years plus the four nearest months that are not in the March quarterly cycle. This results in 44 different contracts. The last trading day is the second London bank business day prior to the third Wednesday of the contract month. The final cash settlement is 100 minus the British Bankers' Association survey of 3-month LIBOR rate. We will only use standard Eurodollar futures (i.e. we will not use data from contacts with expiration date other that March quarterly cycle date). The market practice is to quote the rates L on the Eurodollar futures in terms of the price of the future, i.e. $100 \times (1 - L)$, where L is the predicted LIBOR rate (annualized).

Eurodollar options

Eurodollar option is a standard American-style option based on Eurodollar futures. Such options are among the most actively traded interest rate options listed on the exchange (CME). The underlying contracts for such options are usually 1, 2, 3 or 4 year Eurodollar Futures with the expiration date in December (although options with other expiration dates are also issued). The options lifetime ranges typically from 3 months to two years. They are usually quoted within the range $\pm 150bp$ with strike increments of $12.5bp$ (and $25bp$ within the range ± 550; where bp means basis point). For detailed description and other types of options (i.e. Mid-Curve options, Weekly Mid-Curve options) see *www.cmegroup.com*.

The Brace-Gątarek-Musiela Model

The Brace-Gątarek-Musiela model (BGM) is a stochastic model of time-evolution of interest rates. It will be used here to simulate the (Monte Carlo) paths of LIBOR futures. We will now present a simplified BGM model fitted to our framework and give some comment on the estimation procedure. Let $T_0 = 0$ and $T_i = T_{i-1} + \frac{3}{12}$ for $i = 1, 2, 3, 4$. In reality the dates of expiration for the consecutive Eurodollar futures differ slightly from 90 days, which potentially might have an impact on the results, especially when we consider short time options. Nevertheless we will use the theoretical values for simplicity. Let L_0 be a spot LIBOR rate and let $L_i : [0, T_i] \times \Omega \to \mathbb{R}$ be the i-th forward LIBOR rate. Assuming d sources of randomness, the dynamics of the i-th LIBOR rate could be described as:

$$d \log L_i(t) = \left(\sum_{j=i(t)}^{i} \frac{\delta_j L_j(t)}{1 + \delta_j L_j(t)} \sigma_j(t) - \frac{\sigma_i(t)}{2} \right) \sigma_i(t) dt + \sigma_i(t) dW^{\mathbb{Q}_{\mathrm{Spot}}}(t)$$

for $t \in [0, T_i]$, where $\delta_i = T_{i+1} - T_i = 3/12$ is the accrual period of the i-th LIBOR forward rate, $\sigma_i(t) : [0, T_i] \times \Omega \to \mathbb{R}^d$ is the instantaneous volatility of the i-th LIBOR forward rate, $i(t)$ denotes the index of the bond (corresponding to the appropriate Eurodollar future) which is first to expire at time t and finally $W^{\mathbb{Q}_{\text{Spot}}}(t)$ is a standard (d-dimensional) Brownian motion under the spot LIBOR measure $\mathbb{Q}_{\text{Spot}}$. (See [12] for more details.) We are assuming here that the sources of randomness are independent of each other and that the proper dependency structure is modelled with σ_i. For the Monte Carlo simulation we will use a (standard Euler) discretized version of the above SDE with the time step $\Delta t = \frac{1}{360}$, i.e.

$$\Delta \log L_i(t) = \left(\sum_{j=i(t)}^{i} \frac{\delta_j L_j(t)}{1 + \delta_j L_j(t)} \sigma_j(t) - \frac{\sigma_i(t)}{2} \right) \sigma_i(t) \Delta t + \sigma_i(t) \epsilon_t \sqrt{\Delta t} \tag{4}$$

where $\epsilon_t \sim \mathcal{N}(0, \mathbf{I})$ is a d-dimensional standard normally distributed random vector. In our implementation we will use $d = 3$. To calibrate the model we need to define the functions $\sigma_i(t)$, for $i = 1, 2, 3, 4$. We will assume that $\sigma_i(t)$ (for $i = 1, 2, 3, 4$) is time homogeneous i.e. that there exist a function $\lambda : [0, T] \to \mathbb{R}^3$ such that

$$\sigma_i(t) = \lambda(T_i - t), \quad 0 \le t \le T_i, \quad i = 1, 2, 3, 4.$$

In other words we could say that the structure of $\sigma_i(t)$ depends only on time to maturity of L_i. We will provide values of $\lambda(T_i)$ for $i = 1, 2, 3, 4$ and assume that for $t \in [T_{i-1}, T_i]$,

$$\lambda(t) = \lambda(T_i).$$

The values of the 3-dimensional vectors $\lambda(T_i)$ could be arranged in a 4×3 matrix Λ whose entries are $\Lambda_{i,j} = \lambda_j(T_i)$. We will apply the Principal Components Analysis (PCA) to Eurodollar futures data to approximate the values of Λ. In other words we base our estimation process on the correlation between the Eurodollar futures. A major problem in calibration of PCA is that Eurodollar futures have fixed maturity dates, and so for a given T we could observe a contract with volatility $\lambda(T)$ only once per three months. To overcome this, we will use linear interpolation of the quoted prices of Eurodollar futures (which is in fact a common market practice). In other words, we assume that given two contracts $L_i(t)$ and $L_{i+1}(t)$ with maturities T_i and T_{i+1}, for any $\alpha \in [0, 1]$, the Eurodollar future with maturity date $(1 - \alpha)T_i + \alpha T_{i+1}$ has the price $(1 - \alpha)L_i(t) + \alpha L_{i+1}(t)$. One should note that we need the L_5 prices to do such interpolation. Using that approach with Eurodollar futures prices we obtain the values of contracts with volatility $\lambda(T_i)$ (i=1,2,3,4) for every trading day t. We also use linear interpolation of forward LIBOR rates for the days when the market is not operating (i.e. we interpolate the contract prices using known quotes from the last trading day before and the next trading day after the date in question.). Because of that assumption, in order to conduct the PCA and to estimate σ_i (for $i = 1, 2, 3, 4$), we will need (for each day) the prices of the five Eurodollar futures closest to delivery. Let us now comment on the PCA estimation process. We assume that

$$\Lambda_{i,j} = \frac{\Theta_i s_j \alpha_{i,j}}{\sqrt{\sum_{k=1}^{d} s_k^2 \alpha_{i,k}^2}}$$

for $i = 1, 2, 3, 4$ and $j = 1, 2, 3$. Here, s_j^2 denotes the variance of the j-th factor computed by PCA ($s_1^2 \geq s_2^2 \geq s_3^2$), $\alpha_{i,j}$ measures the influence of the j-th factor when the time to maturity is in the period $[T_{i-1}, T_i]$ and $\Theta_i := \sum_{j=1}^{3} s_j \alpha_{i,j}$ is total volatility in the i-th period. We also assume that the factors are uncorrelated and that the relative influence of every factor is 1 (i.e. for $j_1, j_2 \in \{1, 2, 3\}$ we have $\sum_i^4 \alpha_{i,j_1} \alpha_{i,j_2} = 0$ if $j_1 \neq j_2$ and $\sum_i^4 \alpha_{i,j_1} \alpha_{i,j_2} = 1$ if $j_1 = j_2$). Note that using (4) and parameters from the PCA we could simulate Eurodollar futures paths.

Setup, data details and the least squares method parameters

We will price the quarterly Eurodollar American call and put options EDZ2 (GEZ2 in Globex notation; it means that the underlying instrument is the December 2012 Eurodollar future). While the values of the American call options could be computed without the use of least squares algorithm, because of coinciding with the European calls, we calculate them anyway to provide more insight about how the parameters are fitted to the market data. In other words we want to check empirically if the differences between the market prices and the computed prices are the result of misfitting of the model parameters or are due to a problem with accuracy of the least squares algorithm.

The first trade day for EDZ2 is December 13, 2010, and the expiration date is December 17, 2012. (It should be noted that the option was a two year option when it was issued but we will be pricing it after one year). We will price several put and call options with different closing prices - ranged from 98.00 to 99.75 - in the period from December 20, 2011 to January 20, 2012.

For the calibration purposes we will use the daily closing prices of Eurodollar futures and the spot LIBOR rate. Given a date t, we will use a period of the same size as the time to maturity of the option (i.e. if the option lifetime is 300 days then we take last 300 days data before time t to calibrate our model).

The least squares algorithm needs several inputs. For the base functions generating the "information about the past" we use a constant and the weighted Laguerre polynomials with the exponential normalization factors up to 3rd degree i.e.

$$f_0(x) = 1, \quad f_1(x) = e^{-x/2},$$
$$f_2(x) = e^{-x/2}(-x + 1), \quad f_3(x) = e^{-x/2}(0.5x^2 - 2x + 2).$$

Our implementation is based on the Monte Carlo simulation of the L_4 values obtained using (4). The algorithm needs also formulas for interest rate (for discounting purposes) in two places. Firstly, to discount the values of options from one period to another (in the recursive step-by-step part). Secondly, to compute the final price of the option (i.e. to discount the optimal prices from every simulation to time $T_0 = 0$). While the second interest rate could be associated with standard spot LIBOR rate, the first one must be based on the evolution of assets (i.e. for every path in the Monte Carlo run, one must estimate separately the spot rate at time t using prices of Eurodollar contracts).

Estimation and numerical results

Let us present detailed estimation results for the date December 20, 2011. Note that a similar procedure must be conducted for all remaining days under consideration. Assuming the BGM dynamics and taking into account the Eurodollar futures closing prices during the period from December 26, 2010 to December 20, 2011 we have conducted PCA and obtained

$$\hat{\Lambda} = \begin{pmatrix} 0.024063776 & 0.033758193 & 0.040538115 & 0.043033555 \\ 0.024267981 & 0.018222734 & 0.007111945 & -0.004846372 \\ 0.007801289 & -0.001039692 & -0.006052515 & -0.004629562 \end{pmatrix}.$$

To price several put and call options with different strike prices and closing date on December 20, 2011 we have simulated a 10,000 strong Monte Carlo sample and using the least-squares algorithm we have obtained prices for different strikes. The procedure was repeated 1000 times. The results are presented in Table 1 (with the sample mean μ and the sample standard deviation σ consistent with the simulated distribution of the prices). The Monte Carlo distributions of the prices of the Eurodollar put and call options with the strike price 99.50 could be seen in Figure 2. In Figure 1 we show examples of 100 Monte Carlo paths together with the actual realization of the process.

We have performed a similar analysis for all days from December 21, 2011 to January 20, 2012. It should be noted that during this period, EDZ2 is the fourth closest to delivery Eurodollar Future, so the estimation procedure is analogous. In Fig. 3 we can see the dynamics of the original put and call option prices, the means from 1000 simulations (each of size 10,000) and the lower and upper 5% quantiles for the put and call options with strike price 99.50. The values of the mean and standard deviation of the simulated prices of the options as well as the corresponding market prices of the options can be seen in Table 2. We have chosen the strike price 99.50 because the mean volume of transactions was highest in the considered period. It is interesting to note that the Monte Carlo price corresponding to this strike price is always higher than the market price (see Table 2), which might be the result of the fact that the option was actively traded (cf. Table 1, where the market price is usually lower than the Monte Carlo mean price). Note also that the value of σ in Table 2 is highest for the strike price equal to 99.50, which might explain the interest in the option with this particular strike price.

4.2 Basket and dual-strike options

In this subsection we will use the least squares algorithm to price 1.5 month basket and dual-strike American put options, whose payoff functions are based on two market indices, namely DAX and EUROSTOXX50 (which we will denote by EUR for brevity). We will assume that the underlying instruments follow the standard bivariate Brownian dynamics. Unfortunately, bivariate options are usually over-the-counter (OTC) instruments, so it is

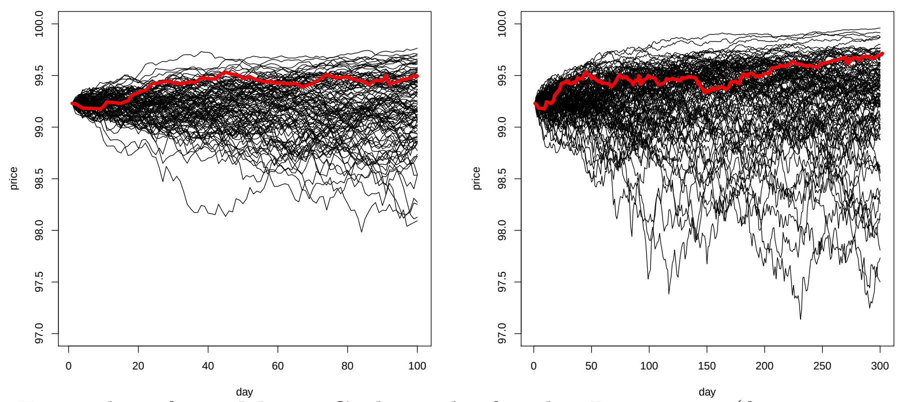

Figure 1: Examples of 100 Monte Carlo paths for the L_4 contact (for 20.12.2011) and the realized path (red) during the first 100 and 300 days.

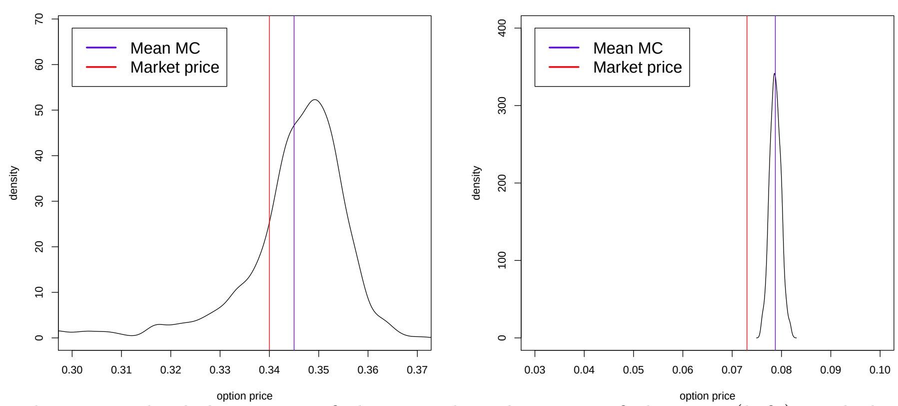

Figure 2: The smoothed densities of the simulated prices of the put (left) and the call (right) option with strike 99.50 on December 20, 2011. The distribution is based on 1000 Monte Carlo runs, each of size 10,000.

difficult to find market data for such options. Nevertheless, we could do a partial comparison with the relevant one-dimensional standard American put options based on DAX and EUR. As in the first example, let us start with some background.

DAX and EUROSTOXX50 Indices

The univariate standard American put options based on DAX and EUR are traded on the Eurex Exchange (see *www.eurexchange.com*). In fact the underlyings are not indices but exchange-traded funds (ETF) which are actively traded on the German stock market (Deutsche Börse Group). The base currency for both options is the Euro. The iShares DAX (DE) ETF Option (EXS1) underlying ISIN is DE00593391 and the iShares EUROSTOXX50

Table 1: Prices of the Eurodollar options according to least squares algorithm (L-S), based on 1000 Monte Carlo simulations (each of size 10,000). Here μ denotes the sample mean of the 1000 prices obtained with MC simulation while σ denotes the sample standard deviation.

Date: 20.11.2011	Put			Call		
Strike	Market price	L-S μ	L-S σ	Market price	L-S μ	L-S σ
98.00	0.070	0.045	0.0038	1.295	1.267	0.0038
98.12	0.078	0.052	0.0043	1.178	1.154	0.0043
98.25	0.085	0.061	0.0044	1.060	1.032	0.0044
98.37	0.095	0.070	0.0048	0.945	0.922	0.0048
98.50	0.105	0.082	0.0050	0.833	0.804	0.0050
98.62	0.120	0.096	0.0055	0.723	0.698	0.0055
98.75	0.138	0.114	0.0058	0.615	0.587	0.0058
98.87	0.155	0.134	0.0061	0.508	0.488	0.0061
99.00	0.175	0.160	0.0067	0.403	0.386	0.0067
99.12	0.203	0.191	0.0078	0.308	0.298	0.0078
99.25	0.238	0.232	0.0088	0.218	0.211	0.0088
99.37	0.280	0.280	0.0097	0.135	0.141	0.0097
99.50	0.340	0.345	0.0110	0.073	0.079	0.0110
99.62	0.425	0.421	0.0094	0.033	0.036	0.0094
99.75	0.525	0.528	0.0044	0.008	0.009	0.0044

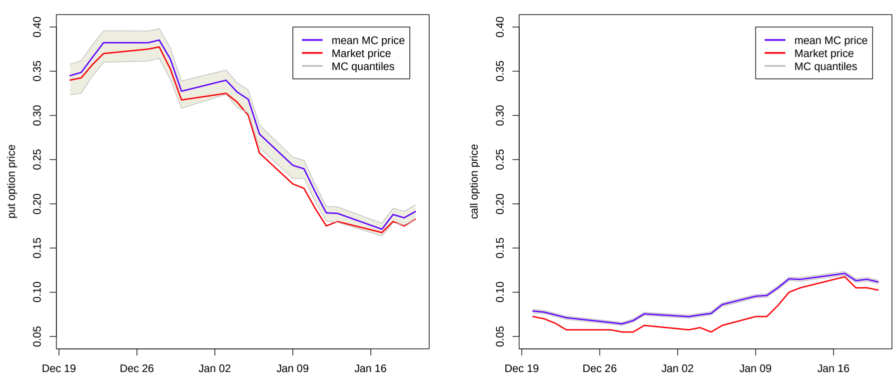

Figure 3: Dynamics of the put (left) and call (right) option prices in the period from December 20, 2011 to January 20, 2012 with the (0.05,0.95) confidence interval based on the Monte Carlo simulation. The strike price is 99.50.

(DE) ETF Option (EUN2) underlying ISIN is IE0008471009. The DAX and EUR indices are highly correlated. The estimated value of Pearson's linear correlation coefficient in the period from October 23, 2012 to January 08, 2013 is equal to 0.920. These indices have some common stocks in their basket, which justifies the high correlation between them. There also might be some contagion between these indices, but in such a short period of time (i.e.

Table 2: Prices of the Eurodollar options with strike price 99.50, according to least squares algorithm (L-S), based on 1000 Monte Carlo simulations (each of size 10,000). As before μ denotes the sample mean of the 1000 prices obtained from the Monte Carlo simulation whereas σ denotes the sample standard deviation.

Date	Put			Call		
	Market price	L-S μ	L-S σ	Market price	L-S μ	L-S σ
20.12.2011	0.340	0.345	0.0110	0.073	0.079	0.0011
21.12.2011	0.342	0.349	0.0110	0.070	0.078	0.0011
22.12.2011	0.358	0.366	0.0112	0.065	0.074	0.0011
23.12.2011	0.370	0.382	0.0111	0.058	0.071	0.0011
27.12.2011	0.375	0.382	0.0103	0.058	0.066	0.0010
28.12.2011	0.378	0.385	0.0102	0.055	0.064	0.0010
29.12.2011	0.352	0.364	0.0104	0.055	0.068	0.0010
30.12.2011	0.318	0.327	0.0093	0.062	0.076	0.0011
03.01.2012	0.325	0.340	0.0088	0.058	0.072	0.0011
04.01.2012	0.315	0.326	0.0086	0.060	0.074	0.0011
05.01.2012	0.300	0.318	0.0084	0.055	0.076	0.0011
06.01.2012	0.258	0.279	0.0079	0.062	0.086	0.0012
09.01.2012	0.222	0.244	0.0072	0.072	0.095	0.0012
10.01.2012	0.218	0.240	0.0071	0.072	0.096	0.0012
11.01.2012	0.195	0.214	0.0069	0.085	0.105	0.0012
12.01.2012	0.175	0.190	0.0054	0.100	0.115	0.0013
13.01.2012	0.180	0.189	0.0054	0.105	0.115	0.0013
17.01.2012	0.168	0.171	0.0046	0.118	0.121	0.0012
18.01.2012	0.180	0.188	0.0053	0.105	0.113	0.0013
19.01.2012	0.175	0.184	0.0053	0.105	0.115	0.0013
20.01.2012	0.182	0.191	0.0054	0.102	0.112	0.0012

1.5 month) it should not present a problem. If necessary, one could adopt here models with different dynamics (e. g. from the multivariate GARCH variety). Such approach would be very closely related to the Heston and Nandi option pricing model [11]. We will describe this type of approach in the next example.

Basket and Dual-strike options

As has been already stated above, basket and dual-strike options are mainly OTC derivatives. In this example we will consider a bivariate American put option. The payoff functions at time t, for a bivariate Basket American put option (1) and Dual-strike American put option (2) is given by

$$p^{(1)}(t) = \max\left(X_1 - S_1(t), X_2 - S_2(t), 0\right),$$

$$p^{(2)}(t) = \max\left(\frac{X_1 + X_2}{2} - \frac{S_1(t) + S_2(t)}{2}, 0\right),$$

where $S_1(t)$ and $S_2(t)$ are the prices of the first and the second underlying at time t, respectively, and X_1, X_2 are the strike prices.

The asset dynamics

We will assume that the stochastic process $(S_1(t), S_2(t))$ is a 2-dimensional geometric Brownian motion. In this case the instantaneous covariance matrix for the processes $\log S_1$ and $\log S_2$ is

$$\mathrm{Cov}[d\log S_1(t), d\log S_2(t)] = H\,dt,$$

where H is a positive definite (2×2)-matrix. Let σ be a (2×2)-matrix such that $\sigma\sigma^* = H$, where σ can be obtained – for example – via the standard Cholesky factorization. Let σ_1 and σ_2 denote the rows of σ. With respect to the risk-neutral probability measure

$$S_i(t_2) = S_i(t_1)\exp\left[\left(r - \frac{\sigma_i\sigma_i^*}{2}\right)(t_2 - t_1) + \sigma_i\left(W(t_2) - W(t_1)\right)\right], \qquad i = 1, 2,$$

where $0 \le t_1 < t_2$, r is a risk-free rate and W is a standard 2-dimensional Wiener process written in a column form. The following is a discretized version of this formula expressed in terms of log-returns:

$$\Delta\log S_i(t) = \left(r - \frac{\sigma_i\sigma_i^*}{2}\right)\Delta t + \sqrt{\Delta t}\sigma_i\epsilon_t, \qquad i = 1, 2,$$

where ϵ_t is a standard bivariate normal random column vector.

Setup, data details and the least squares method parameters

In this example we will construct a bivariate Basket and Dual-Strike American put options based on 1 DAX ETF share and 2.5 EUR ETF shares (to have similar strike prices in both cases). We will price Basket and Dual-strike put options on January 08, 2013 with the expiration date March 16, 2013 (to make it comparable to existing univariate options). The option lifetime will be 49 business days. The strike prices will range from 65 to 75 and from 66 to 76, for the first and the second strike price, respectively.

To estimate the covariance matrix H we will use the last 50 observations of the price of ETF (DE) DAX and ETF (DE) EUROSTOXX50. Choosing of a relatively short time interval for calibration purposes is quite common in practice (e.g. this is the case with the estimation of the VIX volatility index).

As in the previous case, we will need two inputs for the least squares algorithm: an interest rate (for discounting) and appropriate basis functions. Because the option lifetime is short, we will assume that the interest rate is constant and equal to $r = 1.50\%$ (the ECB interest rate on January 08, 2013). Moreover, we will use weighted polynomial base up to the second degree with exponential normalization factors (as in [16]):

$$f_{0,0}(x,y) = e^{\frac{1}{2}}, \qquad f_{1,0}(x,y) = e^{\frac{-x}{2}}x, \qquad f_{0,1}(x,y) = e^{\frac{-y}{2}}y, \qquad f_{1,1}(x,y) = e^{\frac{-(x+y)}{4}}xy,$$

$$f_{1,2}(x,y) = e^{\frac{-(x+y)}{4}}xy^2, \qquad f_{2,1}(x,y) = e^{\frac{-(x+y)}{4}}x^2y, \qquad f_{2,2}(x,y) = e^{\frac{-(x+y)}{4}}x^2y^2.$$

Estimation and numerical results

The estimated (annualized) covariance matrix is equal to

$$\hat{H} := \begin{bmatrix} 0.0178 & 0.0147 \\ 0.0147 & 0.0142 \end{bmatrix},$$

from which we obtain $\rho = 0.920$, $\sigma_1 = 0.133$ and $\sigma_2 = 0.119$. Using the matrix $\hat{H}$, we run 100,000 Monte Carlo simulations (each of size 49). Next, using the least squares algorithm we compute the prices of basket and dual-strike American put options for different strike prices. We also compute the least squares prices for the univariate American put options based on 1 DAX ETF share and 2.5 EUR ETF share. Apart from the market data we also present the theoretical price according to the Cox-Ross-Rubinstein model (CRR) as it is used by the Eurex Exchange to quote option prices when no trading takes place. It should be noted that the volume of transaction of American put options is very low, so unfortunately the market price is just for comparison purposes. Also, the least squares price should be compared with the CRR price, rather than the market price (as it is computed under the compatible assumptions about the asset dynamics).

The prices (obtained using single 100,000 Monte Carlo run) can be seen in Table 3. The columns with names DAX and EUR denote the standard univariate put options (i.e. with 1 DAX ETF and 2.5 EUR ETF share as the underlying respectively). We have also performed multiple Monte Carlo runs (1000), each of size 10,000 for the Basket and Dual-Strike options with strike price $(70, 70)$. The corresponding Monte Carlo density function could be seen in Figure 4 (this could provide some information about the model and/or the Monte Carlo bias).

4.3 The Heston-Nandi model

In the last example we will use the least squares algorithm to price two 1.5 month American put options, whose payoff function is based on a single market index. We will use data from the previous example, i.e. we will price options based on DAX and EUR indices. We will assume that the dynamics of the underlyings could be described with the Heston-Nandi GARCH Model [11].

Table 3: Prices of the options according to historical stock market data, the CRR model and the least squares algorithm (L-S). Here $S_0 = (68.05, 69.72)$, $r = 1.50\%$, $T = 49/252$, $\sigma_1 = 0.133$, $\sigma_2 = 0.119$, $\rho = 0.920$.

Strike price		Market price		CRR price		L-S price			
EUR	DAX	EUR	DAX	EUR	DAX	EUR	DAX	Basket	Dual-Strike
65	66	0.58	0.34	0.45	0.25	0.44	0.24	0.31	0.46
67.5	66	1.50	0.34	1.25	0.25	1.23	0.24	0.59	1.23
70	66	3.10	0.34	2.67	0.25	2.63	0.24	1.00	2.65
72.5	66	5.15	0.34	4.63	0.25	4.62	0.24	1.58	4.62
75	66	7.53	0.34	6.95	0.25	6.95	0.24	2.33	6.95
65	68	0.58	0.85	0.45	0.69	0.44	0.67	0.52	0.72
67.5	68	1.50	0.85	1.25	0.69	1.23	0.67	0.91	1.27
70	68	3.10	0.85	2.67	0.69	2.63	0.67	1.45	2.65
72.5	68	5.15	0.85	4.63	0.69	4.62	0.67	2.17	4.62
75	68	7.53	0.85	6.95	0.69	6.95	0.67	3.04	6.95
65	70	0.58	1.74	0.45	1.52	0.44	1.49	0.82	1.50
67.5	70	1.50	1.74	1.25	1.52	1.23	1.49	1.33	1.64
70	70	3.10	1.74	2.67	1.52	2.63	1.49	2.01	2.67
72.5	70	5.15	1.74	4.63	1.52	4.62	1.49	2.86	4.62
75	70	7.53	1.74	6.95	1.52	6.95	1.49	3.85	6.95
65	72	0.58	3.00	0.45	2.79	0.44	2.75	1.22	2.76
67.5	72	1.50	3.00	1.25	2.79	1.23	2.75	1.86	2.77
70	72	3.10	3.00	2.67	2.79	2.63	2.75	2.67	3.09
72.5	72	5.15	3.00	4.63	2.79	4.62	2.75	3.64	4.63
75	72	7.53	3.00	6.95	2.79	6.95	2.75	4.72	6.95
65	76	0.58	6.43	0.45	6.29	0.44	6.28	2.33	6.28
67.5	76	1.50	6.43	1.25	6.29	1.23	6.28	3.24	6.28
70	76	3.10	6.43	2.67	6.29	2.63	6.28	4.28	6.28
72.5	76	5.15	6.43	4.63	6.29	4.62	6.28	5.41	6.30
75	76	7.53	6.43	6.95	6.29	6.95	6.28	6.62	7.16

Heston-Nandi GARCH Model

As we have mentioned before, sometimes it is better to use GARCH(1,1) dynamics or its modifications rather than the standard discretised geometric Brownian motion to describe price fluctuations. Let S_t denote the price of the underlying. Using Heston-Nandi GARCH dynamics, simplified to our framework, we will assume that the log-returns of the random process S_t could be described by formula

$$\Delta \log S_t = r_{\text{daily}} + \lambda \sigma_t^2 + \sigma_t \epsilon_t,$$

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha(\epsilon_{t-1} - \gamma \sigma_{t-1})^2,$$

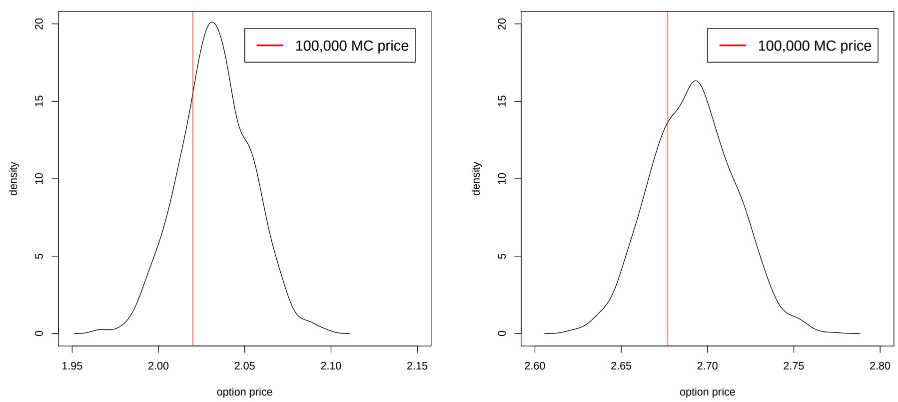

Figure 4: The smoothed densities of the least squares prices of the basket (left) and the dual-strike (right) American put options for the strike price $(70, 70)$. Density estimation is based on multiple Monte Carlo runs (1000) each consisting of 10,000 price paths. The vertical line depicts the least squares price obtained with the single 100,000 Monte Carlo simulation.

where Δ denotes the daily backward difference (i.e. $\Delta \log S_t = \log(S_t/S_{t-1})$, the parameter r_{daily} denotes daily riskless interest rate, $(\lambda, \omega, \beta, \alpha, \gamma)$ are model parameters and ϵ_t is a standard Gaussian white noise. We will additionally assume that there is no asymmetry, i.e. $\gamma = 0$.

If we use the standard Heston-Nandi dynamics (i.e. the objective probability measure) then the discounting part of the least squares algorithm will be path dependent. In order to avoid this complication we will switch to the risk-neutral measure and use the risk-neutral dynamic of the underlying return. The risk neutral process is obtained simply by replacing (previously estimated) parameters λ and γ with (-0.5) and $(\gamma+\lambda+0.5)$ respectively (see [11] for details). Moreover, we will use the long run (expected) standard deviation from Heston Nandi model for comparison purposes. It is equal to [11]

$$\sigma_{HN} = \frac{\omega + \alpha}{1 - \beta - \alpha\gamma^2}. \tag{5}$$

Setup, data details and the least squares method parameters

We will use the data from the previous example, i.e. we will use EUR and DAX as the underlyings. As before, the options expiration date will be March 16, 2013 and we will price them on January 08, 2013 (thus the option lifetime will be 49 business days).

Also, as in the first example, we will use weighted Laguerre polynomials with the exponential normalization factors up to 3rd degree i.e.

$$f_0(x) = 1, \quad f_1(x) = e^{-x/2},$$

$$f_2(x) = e^{-x/2}(-x+1), \quad f_3(x) = e^{-x/2}(0.5x^2 - 2x + 2)$$

as base polynomials for the regression procedure.

Estimation and numerical results

As before, we will assume that (annualised) risk free rate is equal to $r = 1.50\%$ and put $r_{\text{daily}} = r/252$ (as there are 252 trading days each year). Using the last 50 prices of 1 DAX ETF and 2.5 EUR ETF shares we have obtained two sets of parameters:

$$\begin{array}{llll} \lambda_{EUR} = 7.280, & \omega_{EUR} = 2.738\text{E-}05, & \alpha_{EUR} = 5.238\text{E-}05, & \beta_{EUR} = 0.086, \\ \lambda_{DAX} = 16.971, & \omega_{DAX} = 1.954\text{E-}05, & \alpha_{DAX} = 5.404\text{E-}05, & \beta_{DAX} = 4.758\text{E-}28. \end{array}$$

The (annualised) volatilities obtained from (5) are equal to $\sigma_1 = 0.149$ and $\sigma_2 = 0.137$, respectively. The mean sample prices of American put options, obtained from ten simulations (each consisting of 100,000 Monte Carlo paths) can be seen in Tables 4 and 5. We also present the theoretical European put option prices according to Heston-Nandi model [11] as well as the American put options prices and early exercise premiums (i.e. the differences between the prices of American and European put options) according to Cox-Ross-Rubinstein model (CRR), with volatilities obtained from (5). Both models are presented for comparison purposes. Moreover, we perform multiple Monte Carlo runs (1000), each of size 10,000, to calculate prices of the American put options with strike price 70 (both for EUR and DAX). The obtained Monte Carlo density functions could be seen in Figure 5 (this could provide some information about the model and/or the Monte Carlo bias).

Table 4: Prices of the EUR American put options according to the least squares algorithm, compared with the actual market prices, CRR model prices and the Heston-Nandi European put option prices. EA denotes the early exercise premium. Here $S_0 = 68.05$, $r = 1.50\%$, $T = 49/252$, $\sigma_1 = 0.149$.

Strike price	Market price	CRR price	CRR EA	H-N price	L-S price
65.0	0.58	0.57	0.00	0.57	0.57
67.5	1.50	1.41	0.01	1.40	1.40
70.0	3.10	2.79	0.03	2.78	2.79
72.5	5.15	4.67	0.06	4.66	4.71
75.0	7.53	6.88	0.11	6.88	6.98

Table 5: Prices of the DAX American put options according to the least squares algorithm, compared with the actual market prices, the CRR model prices and the prices of the European put options based on the Heston-Nandi model. EA denotes Early Exercise Premium. Here $S_0 = 69.72$, $r = 1.50\%$, $T = 49/252$, $\sigma_1 = 0.137$.

Strike price	Market price	CRR price	CRR EA	H-N price	L-S price
66	0.34	0.38	0.00	0.38	0.38
68	0.85	0.88	0.01	0.87	0.87
70	1.74	1.74	0.01	1.70	1.70
72	3.00	2.98	0.03	2.91	2.92
76	6.43	6.33	0.10	6.22	6.31

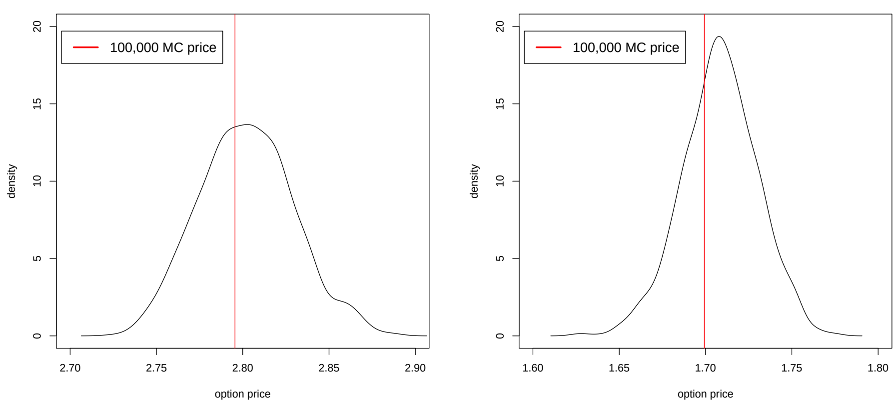

Figure 5: The smoothed distributions of the least squares prices of the EUR (left) and DAX (right) American put options (with strike price 70) with the dynamics of the underlying described by the Heston Nandi GARCH model. The prices were computed using multiple Monte Carlo runs (1000) each generating 10,000 values. The vertical line corresponds to the sample mean least squares price obtained with ten 100,000 Monte Carlo simulations.

5 Concluding remarks

In the paper we have shown that the widely used least squares approach to Monte Carlo based pricing of American options can remain valid under very general and flexible choice of assumptions. In particular, convergence to the theoretical price obtained via Snell envelopes is true even if the pay-offs are path dependent, the underlying is non-Markovian and with a highly adaptable setup for approximation of conditional expectations. On the one hand, the computational cost of liberalization of the assumptions may be potentially very high. On the other hand, we present three examples supporting a growing body of empirical evidence showing how in many practical applications even relatively limited non-linear extensions of standard regression may produce reasonable results. The relaxation of the assumptions of the method should be seen primarily as increase in freedom of choice of settings for a specific implementation of the algorithm, which with careful choices may nevertheless retain computational viability.

Acknowledgments

The second author acknowledges the support by Project operated within the Foundation for Polish Science IPP Programme "Geometry and Topology in Physical Models" co-financed by the EU European Regional Development Fund, Operational Program Innovative Economy 2007-2013.

References

[1] C.J. Beveridge and M.S. Joshi. *Practical policy iteration: generic methods for obtaining rapid and tight bounds for Bermudan exotic derivatives using Monte Carlo simulation.* University of Melbourne, Faculty of Business & Economics, Centre for Actuarial Studies, 2009.

[2] A. Brace, D. Gątarek, and M. Musiela. The market model of interest rate dynamics. *Mathematical Finance*, 7(2):127–155, 1997.

[3] M. Broadie and P. Glasserman. Pricing american-style securities using simulation. *Journal of Economic Dynamics and Control*, 21:1323–1352, 1997.

[4] M. Broadie and P. Glasserman. A stochastic mesh method for pricing high-dimensional american options. Technical report, Columbia Business School, New York, 2002. A revised version of a working paper from 1997.

[5] J.F. Carriere. Valuation of early-exercise price for options using simulations and non-parametric regression. *Insurance: Mathematics and Economics*, 19:19–30, 1996.

[6] E. Clément, D. Lamberton, and P. Protter. An analysis of a least squares regression method for american option pricing. *Finance and Stochastics*, 6:449–471, 2002.

[7] R.L. Dobrushin and R.A. Minlos. Polynomials in linear random functions. *Russian Math. Surveys*, 32:2:71–127, 1977.

[8] D. Egloff. Monte Carlo algorithms for optimal stopping and statistical learning. *Ann. Appl. Probab.*, 15(2):1396–1432, 2005.

[9] P. Glasserman. *Monte Carlo methods in financial engineering.* Springer, New York, 2010.

[10] P. Glasserman and B. Yu. Number of paths versus number of basis functions in American option pricing. *Ann. Appl. Probab.*, 14(4):2090–2119, 2004.

[11] S.L. Heston and S. Nandi. A closed-form garch option valuation model. *Review of Financial Studies*, 13(3):585–625, 2000.

[12] F. Jamshidian. Libor and swap market models and measures. *Finance and Stochastics*, 1(4):293–330, 1997.

[13] M. Klimek. Lecture Notes: Least squares method of American option pricing – a guide to the underlying mathematics. Technical report, Jagiellonian University, Kraków, 2012.

[14] M. Klimek, M. Matsuura, and Y. Okabe. Stochastic flows and finite block frames. *Journal of Mathematical Analysis and Applications*, 342(2):816–829, 2008.

[15] G. Liu and L.J. Hong. Revisit of stochastic mesh method for pricing american options. *Operations Research Letters*, 37(6):411–414, 2009.

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

[17] S. Pliska. *Introduction to Mathematical Finance - Discrete Time Models*. Blackwell Publishers, Oxford, 1999.

[18] J. Rust. Using randomization to break the curse of dimensionality. *Econometrica*, 65(3):487–516, 1997.

[19] C.G. Small and D.L. McLeish. *Hilbert space methods in probability and statistical inference*. Wiley, New York, 1994.

[20] L. Stentoft. Convergence of the least squares Monte Carlo approach to american option valuation. *Management Sci.*, 50(9):1193–1203, 2004.

[21] 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. *IEEE Trans. Automat. Control*, 44(10):1840–1851, 1999.

[22] J. N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex American-style options. *IEEE Trans. on Neutral Networks*, 12(4):694–703, 2001.

[23] D. Z. Zanger. Convergence of a least-squares Monte Carlo algorithm for bounded approximating sets. *Appl. Math. Finance*, 16(1-2):123–150, 2009.