Applied Computational Finance

Lesson 5 - The Longstaff-Schartz Algorithm (LSM)

Giovanni Della Lunga

Laurea Magistrale in Quantitative Finance

Bologna - April-May 2020

Outline

Introduction

- As we have seen Monte Carlo simulation is a flexible and powerful numerical method to value financial derivatives of any kind.
- However being a forward evolving technique, it is per se not suited to address the valuation of American or Bermudan options which are valued in general by backwards induction.
- Longstaff and Schwartz provide a numerically efficient method to resolve this problem by what they call Least-Squares Monte Carlo.
- The problem with Monte Carlo is that the decision to exercise an American option or not is dependent on the continuation value.

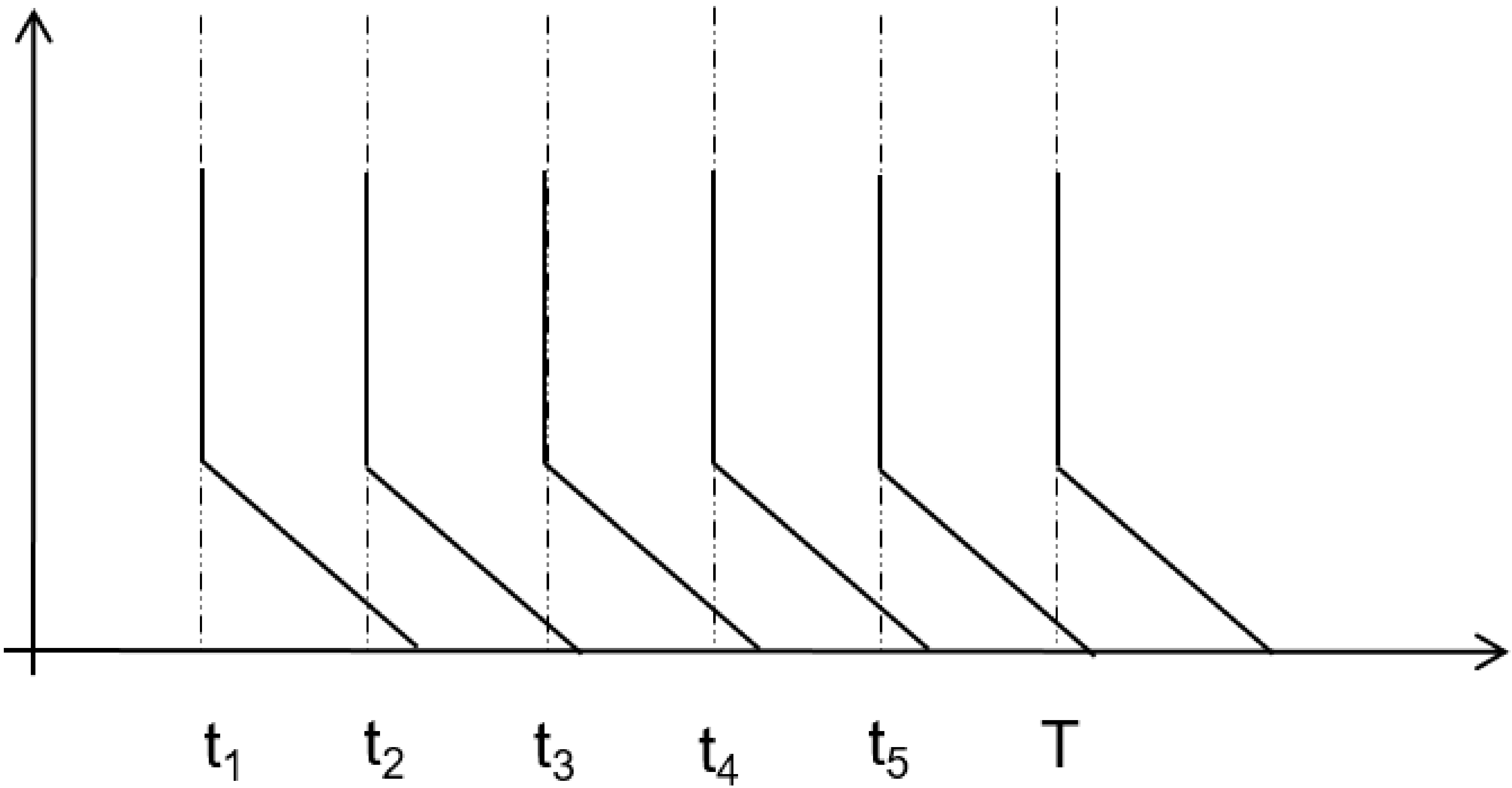
Introduction

- Consider a simulation with $M + 1$ points in time and I paths.
- Given a simulated index level $S_{t,i}, t \in \{0, ..., T\}, i \in \{1, ..., I\}$, what is the continuation value $C_{t,i}(S_{t,i})$, i.e. the expected payoff of not exercising the option?
- The approach of Longstaff-Schwartz approximates continuation values for American options in the backwards steps by an ordinary least-squares regression.
- Equipped with such approximations, the option is exercised if the approximate continuation value is lower than the value of immediate exercise. Otherwise it is not exercised.

The Idea

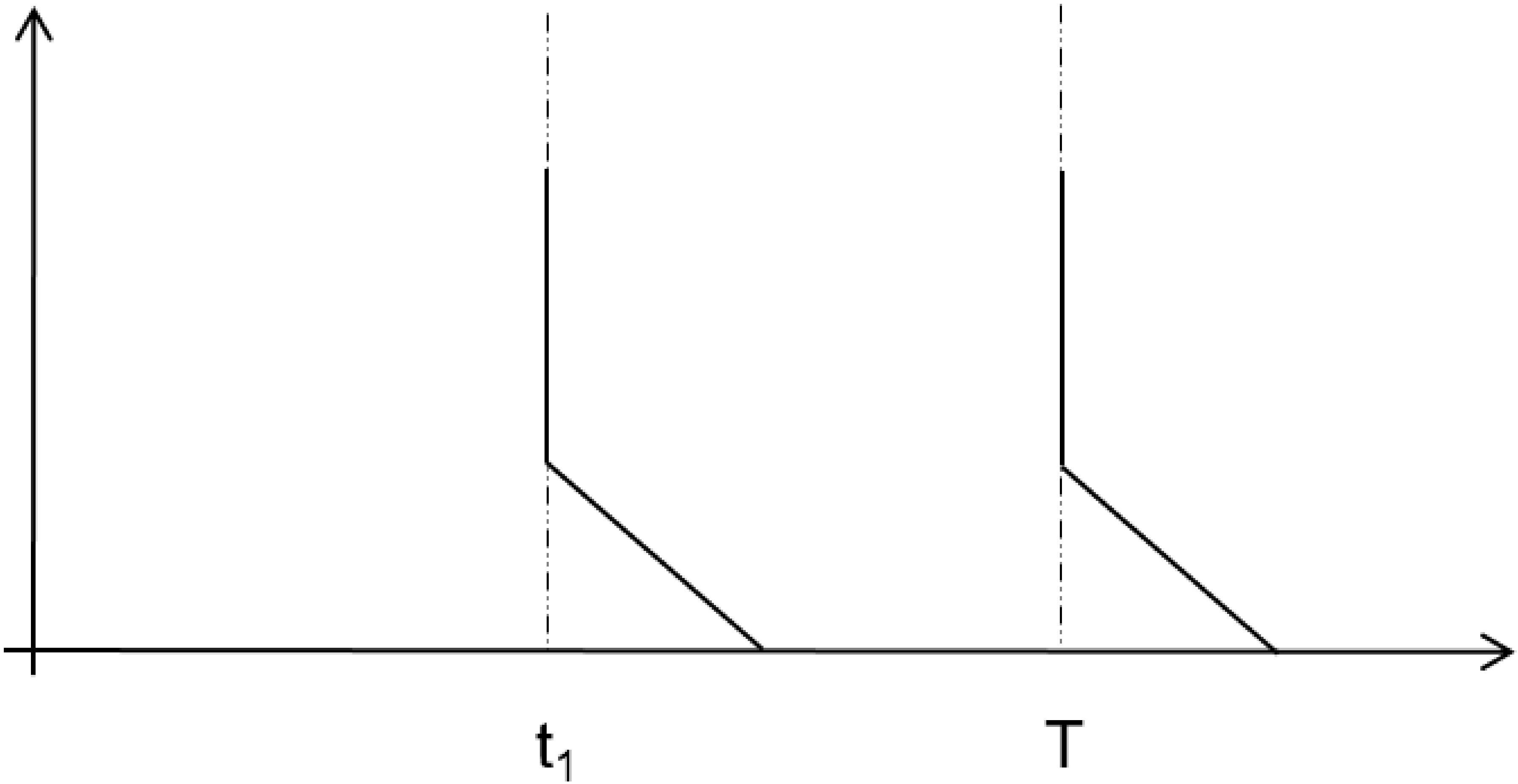
- In order to explain the metodology, let's start from a simpler problem.
- Consider a bermudan option which is similar to an american option, except that it can be early exercised once only on a specific set of dates.
- In the next figure, we can represent the schedule of a put bermudan option with strike K and maturity in 6 years. Each year you can choose whether to exercise or not ...

The Idea



The Idea

Let's consider a simpler example: a put option which can be exercised early only once ...



The Idea

- Can we price this product by means of a Monte Carlo? Yes we can! Let's see how.
- Let's implement a MC which actually simulates, besides the evolution of the market, what an investor holding this option would do (clearly an investor who lives in the risk neutral world). In the following example we will assume the following data, $S(T) =$, $K =$, $r =$, $\sigma =$, $t_1 = 1y$, $T = 2y$.
- We simulate that 1y has passed, computing the new value of the asset and the new value of the money market account

$$S(t_1 = 1y) = S(t_0)e^{(r-\frac{1}{2}\sigma^2)(t_1-t_0)+\sigma\sqrt{t_1-t_0}N(0,1)}$$

$$B(t_1 = 1y) = B(t_0)e^{r(t_1-t_0)}$$

The Idea

- At this point the investor could exercise. How does he know if it is convenient?
- In case of exercise he knows exactly the payoff he's getting.
- In case he continues, he knows that it is the same of having a European Put Option.
- So, in mathematical terms we have the following payoff in t_1

$$\max\left[K - S(t_1), P(t_1, T; S(t_1), K)\right]$$

  where $P(t_1, T; S(t_1), K)$ is the price of a Put which we compute analytically!
- In the jargon of american products, P is called the continuation value, i.e. the value of holding the option instead of early exercising it.

The Idea

- So the premium of the option is the average of this discounted payoff calculated in each iteration of the Monte Carlo procedure.

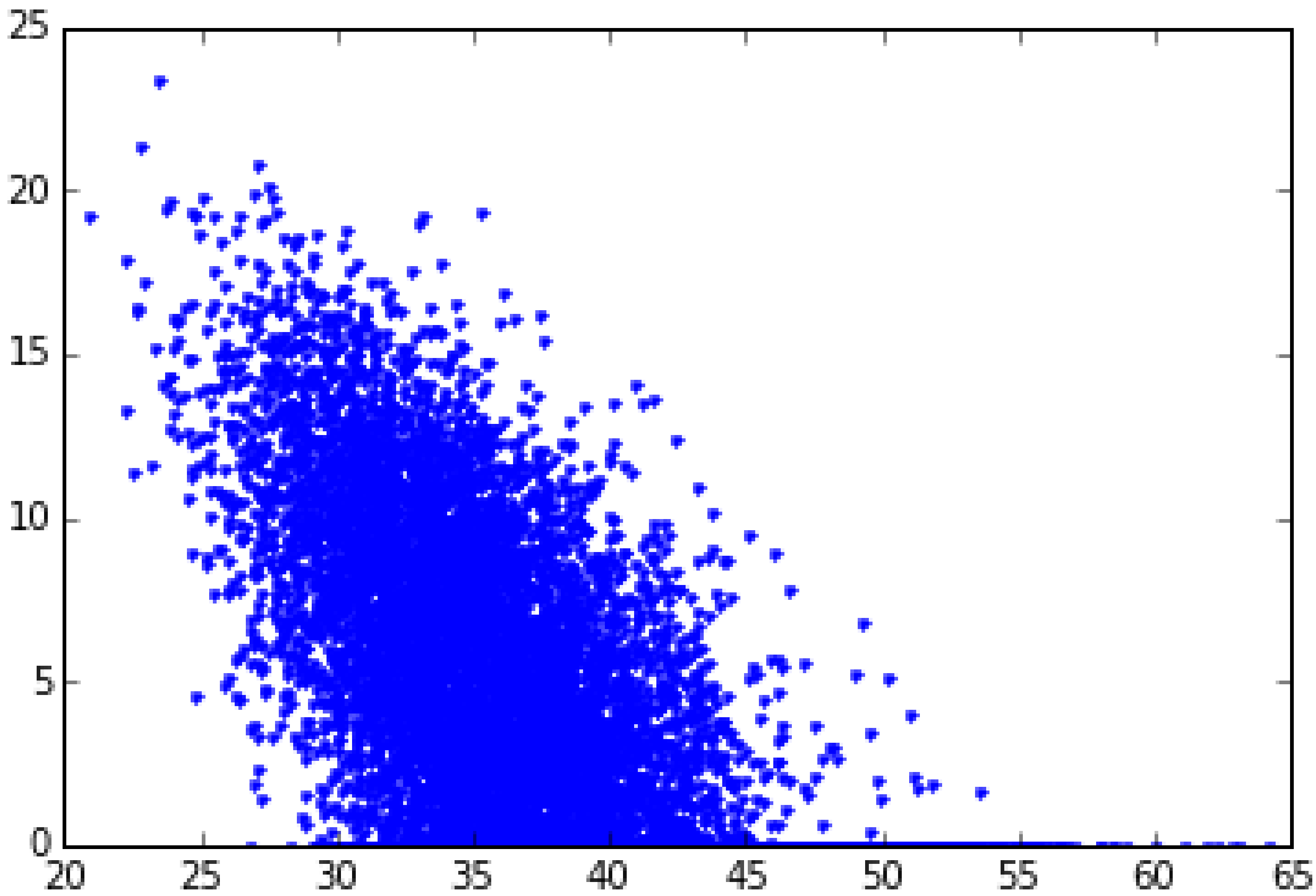
$$\frac{1}{N}\sum_i \max\left[K - S_i(t_1), P(t_1, T; S_i(t_1), K)\right]$$

- Some considerations are in order.
- We could have priced this product because we have an analytical pricing formula for the put. What if we didn't have it?

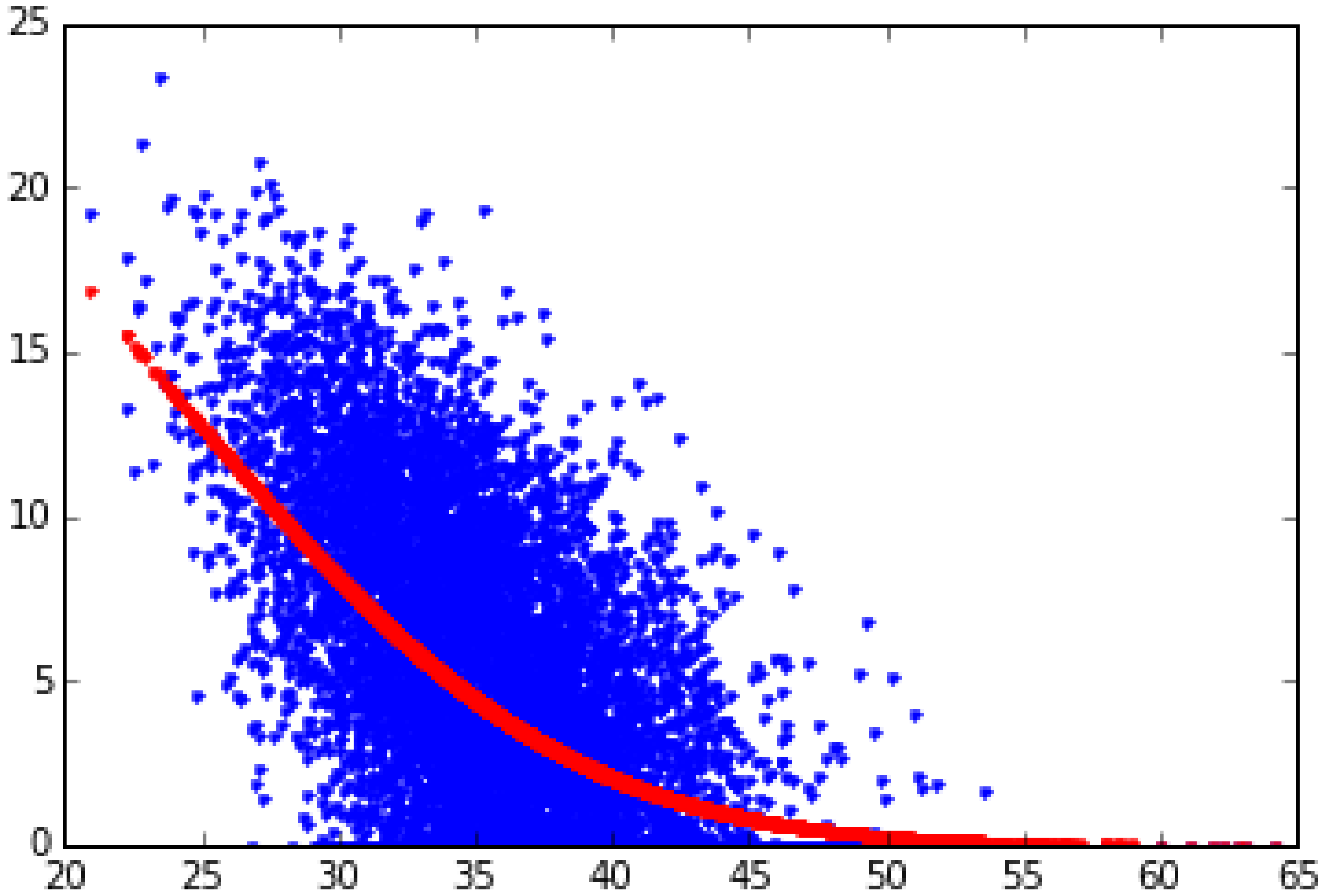
The Idea

- Brute force solution: for each realization of $S(t_1)$ we run another Monte Carlo to price the put.
- This method (called Nested Monte Carlo) is very time consuming. For this very simple case it's time of execution grows as N^2, which becomes prohibitive when you deal with more than one exercise date!
- Let's search for a finer solution analyzing the relationship between the continuation value (in this very simple example) and the simulated realization of S at step t_1.
- let's plot the discounted payoff at maturity, P_i, versus $S_i(t_1)$ ...

The Idea



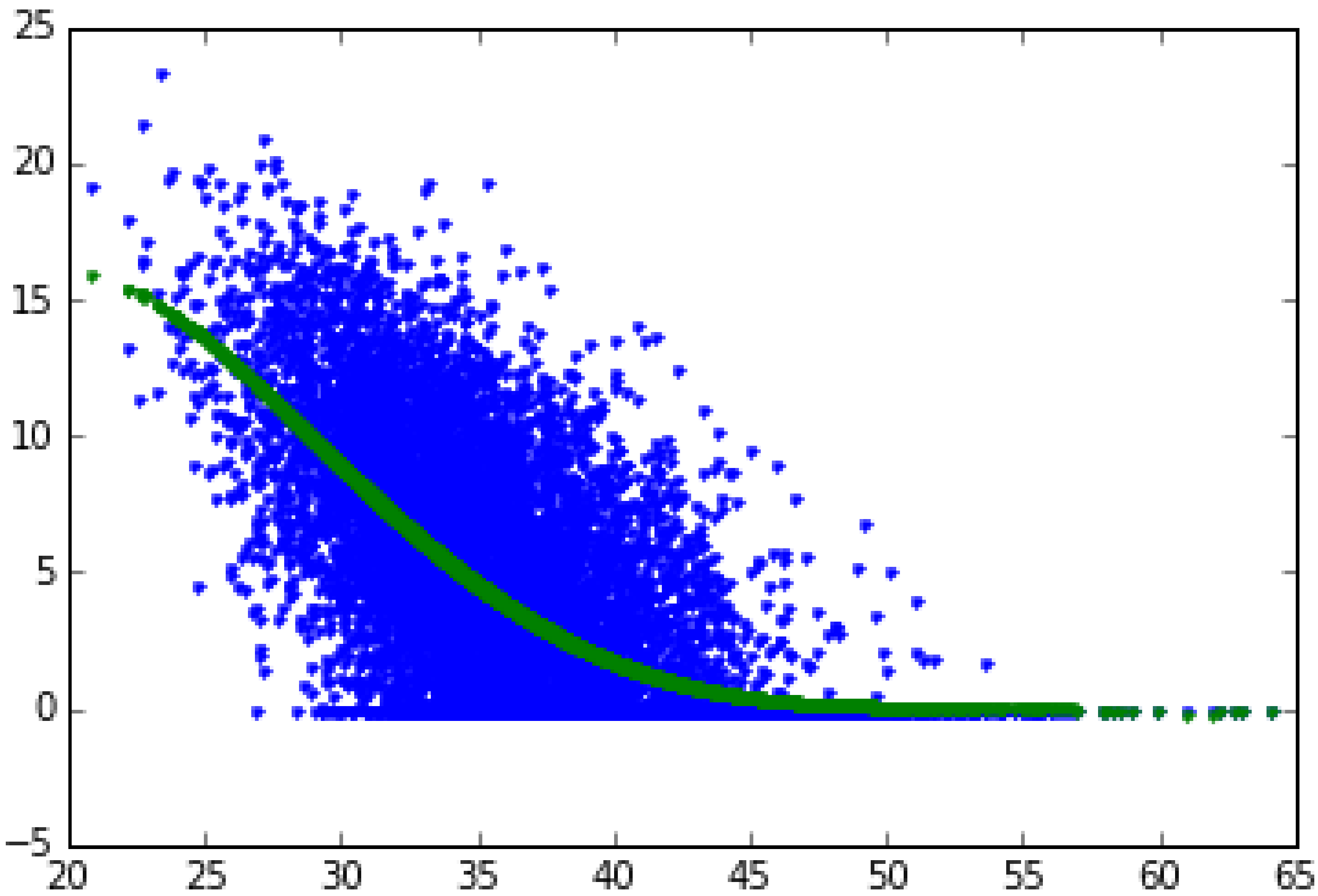
The Idea



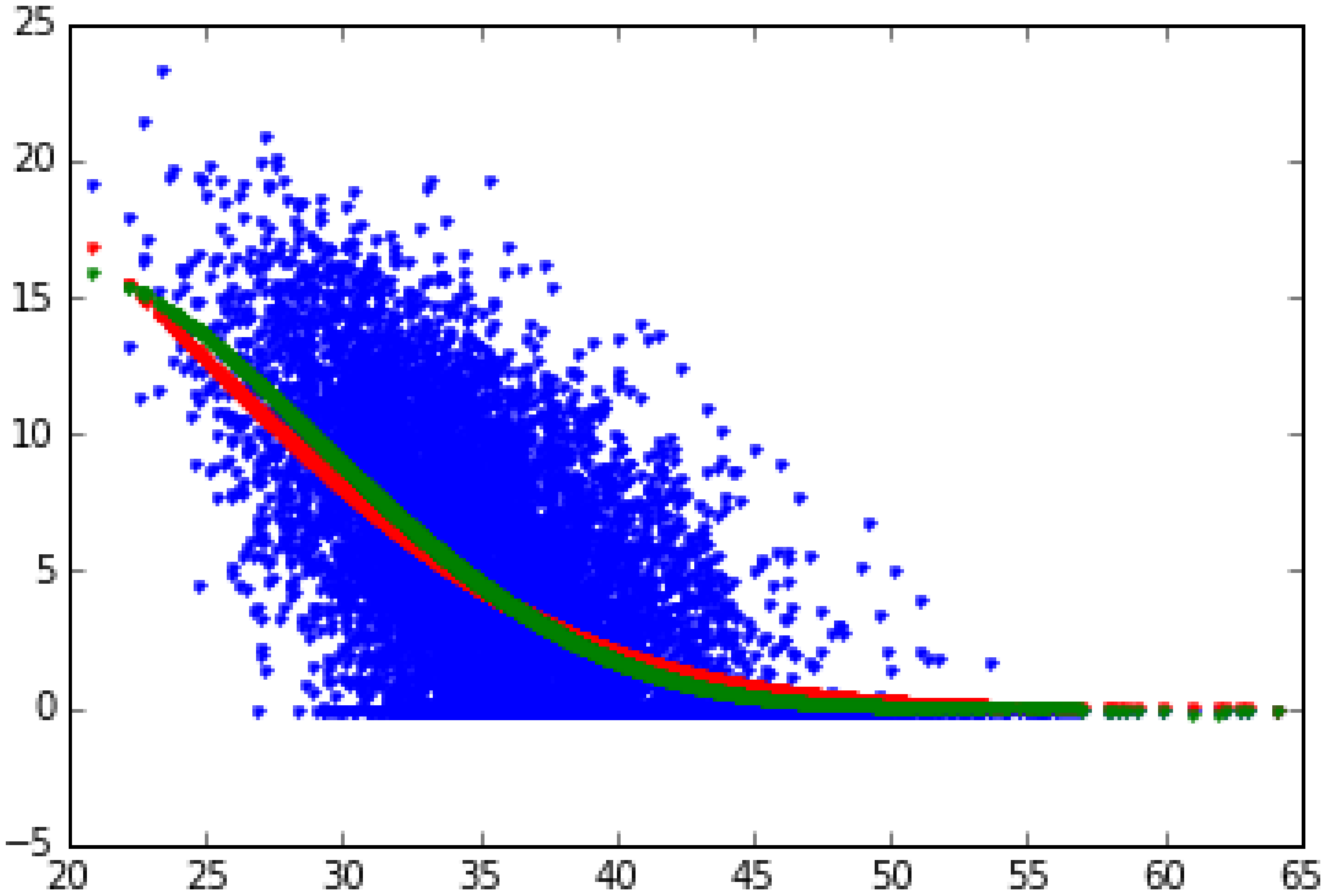
The Idea

- As you can see, the analytical price of the put is a curve which kinds of interpolate the cloud of Monte Carlo points.
- This suggest us that the price at time t_1 can be computed by means of an average on all discounted payoff (i.e. the barycentre of the cloud made of discounted payoff)
- So maybe... the future value of an option can be seen as the problem of finding the curve that best fits the cloud of discounted payoff (up to date of interest)!!!
- In the next slide, for example, there is a curve found by means of a linear regression on a polynomial of 5th order...

The Idea



The Idea



The Idea

- We now have an empirical pricing formula for the put to be used in my MCS

$$P(t_1, T, S(t_1), K) = c_0 + c_1 S(t_1) + c_2 S(t_1)^2 + c_3 S(t_1)^3 + c_4 S(t_1)^4 + c_5 S(t_1)^5$$

- The formula is obviously fast, the cost of the algorithm being the best fit.
- Please note that we could have used any form for the curve (not only a polynomial).
- This method has the advantage that it can be solved as a linear regression, which is fast.

The Longstaff-Schwartz Algorithm

- Regression based methods posit an expression for the continuation value of the form

$$C_i(x) = \mathbb{E}\left[V_{i+1}\left(X_{i+1}\right)|X_i = x\right] = \sum_{r=1}^{M} \beta_{ir}\psi_r(x) \qquad (1)$$

  for some basis functions $\psi_r : \mathbb{R}^b \to \mathbb{R}^b$ and constants $\beta_{ir}\,, r = 1, \ldots, M$.
- We may equivalently write

$$C_i(x) = \bar{\beta}_i^T \bar{\psi}(x) \qquad (2)$$

The Longstaff-Schwartz Algorithm

- Assuming a relation of the form (1) holds and remembering that the general form of the least square estimate (or estimator, in the context of a random sample), β is given by

$$\beta = (X^T X)^{-1} X^T \boldsymbol{y}$$

we can write

$$\begin{aligned} \beta_i &= \left(\mathbb{E}\left[\psi\left(X_i\right)\psi\left(X_i\right)^T\right]\right)^{-1}\mathbb{E}\left[\psi\left(X_i\right)V_{i+1}\left(X_{i+1}\right)\right] \\ &\equiv B_\psi^{-1} B_{\psi V} \end{aligned} \tag{3}$$

since we are working with a number b of simulated path, we have to take the expectation values of the matrix products involved.

- This is why this method is called Least Squares Montecarlo.

The Longstaff-Schwartz Algorithm

- The coefficients β_{ir} could be estimated from observations of pairs $(X_{ij}, V_{i+1}(X_{i+1,j}))\ j = 1, \ldots, b$ each consisting of **the state at time** i **and the corrisponding option value at time** $i+1$.
- In practice the least-squares estimate of β_i is given by

$$\hat{\beta}_i = \hat{B}_\psi^{-1} \hat{B}_{\psi V} \tag{4}$$

where $\hat{B}_\psi$ and $\hat{B}_{\psi V}$ are the sample counterparts of B_ψ and $B_{\psi V}$. More explicitly

$$\hat{B}_\psi = \frac{1}{b} \sum_{j=1}^{b} \psi_q\left(X_{ij}\right) \psi_r\left(X_{ij}\right) \tag{5}$$

and

$$\hat{B}_{\psi V} = \frac{1}{b} \sum_{j=1}^{b} \psi_r\left(X_{ij}\right) V_{i+1}\left(X_{i+1,j}\right) \tag{6}$$

The Longstaff-Schwartz Algorithm

- All of these quantities can be calculated from function values at pairs of consecutive nodes $(X_{ij}, X_{i+1,j})$ $j = 1, \ldots, b$.
- In practive V_{i+1} is unknown and must be replaced by estimated values $\hat{V}_{i+1}$.
- The estimate $\hat{\beta}_i$ then defines an estimate

$$\hat{C}_i(x) = \hat{\beta}_i^T \psi(x) \tag{7}$$

of the continuation value at an arbitrary point x in the state space $\mathbb{R}^b$.

The Longstaff-Schwartz Algorithm

- Simulate b independent paths $\{X_{1j}, \ldots, X_{mj}\}$ $j = 1, \ldots, b$ of the Markov chain ;
- For $t = T$ the option value is $\hat{V}_{mj} = h_m(X_{mj})$ $j = 1, \ldots, b$ by arbitrage
- Start iterating backwards: for $i = m - 1, \ldots, 1$:
  - given estimated values $\hat{V}_{i+1,j}$ $j = 1, \ldots, b$, use regression as above to calculate $\hat{\beta}_i = \hat{B}_{\psi}^{-1} \hat{B}_{\psi V}$;
  - set
  $$\hat{V}_{ij} = \max\left[h_i\left(X_{ij}\right), \hat{C}_i\left(X_{ij}\right)\right] \quad j = 1, \ldots, b \tag{8}$$
  with $\hat{C}_i$ given by (7);
- Set $\hat{V}_0 = \left(\hat{V}_{11} + \cdots + \hat{V}_{1b}\right)/b$

The Longstaff-Schwartz Algorithm

- Longstaff and Scwhartz in their original paper followed a slightly different approach replacing (8) with

$$\hat{V}_{ij} = \begin{cases} h_i\left(X_{ij}\right) & \text{if } h_i\left(X_{ij}\right) \geq \hat{C}_i\left(X_{ij}\right) \quad \text{(exercise takes place)} \\ \hat{V}_{i+1,j} & \text{otherwise (no exercise takes place)} \end{cases} \tag{9}$$

- This gives both the regression and a possible value for $\hat{V}_0$.
- However, since we are using the same paths for regression as for evaluation, the estimate for $\hat{V}_0$ **could be** high biased.

Convergence and Biases

- The analysis of the convergence of the Longstaff-Schwartz algorithm is far from trivial.
- In the original paper, Longstaff and Schwartz give few details of the convergence of their algorithm, except in a simple one period case.
- Clement et al. (2002) provide an in depth study of the convergence properties, by proving that convergence actually takes place and by deriving the correspondent rate of convergence.
- Moreover Glasserman and Yu (2004) give an analysis of the relationship between the number of paths required for a particular number of basis functions in order to ensure convergence.
- In particular they found that , in certain cases, the number of paths required for convergence should grow exponentially with the number of basis functions.

Convergence and Biases

One difficulty in analysing the convergence of Longstaff-Schwartz is that there are various sources of error. Without going into the mathematical details of the demonstration of the convergence of the method it is however essential to know what are the different sources of error.

- The first source is that we are approximating the conditional expectation of our payoff by a linear combination of a finite set of $M + 1$ basis functions. As$M \to \infty$, this will become exact.
- In a practical setting, however, it is not possible to use an infinite set of basis functions!
- Using this approximation to the continuation value, we find an approximate (and sub-optimal) stopping rule for the option.

Convergence and Biases

- A further source of error is that the cofficients β_{ir} are being approximated by a Monte-Carlo regression, so that we are using a set of $M+1$ β_{ir}'s whose MSE will be related to the number of paths N that are used in the approximation.
- Once we have this set of regressed β_{ir}'s, we use it to approximate a stopping rule and if we perform (as usual) another Monte-Carlo simulation, introducing more error, to $\hat{V}_0$, the value of our option
- Moreover if we are using Longstaff-Schwartz to approximate the value of American options, there will be another source of error based on the finite number of exercise dates M we use in the algorithm.

Convergence and Biases

- At several points in this lessons we have noted two sources of bias affecting simulation estimates of American option prices.
- High bias resulting from applying backward induction over a finite set of paths.
- Low bias resulting from suboptimal eexercise.
- The development of high and low bias estimator during the study of stochastic tree and stochastic mesh algorithms keeps the two sources of bias well separate.
- With the Longstaff-Schwartz method is relatively easy to find a low bias estimator.
- In order to ensure that it is low biased, we use a new set of N paths for the evaluation according to the following algorithm...

Convergence and Biases

- Generate the set of β_{ir} as previously described ("backward phase");
- Generate N new paths of M timesteps ("forward phase");
- For each path iterate on time step performing the following step:
  - At each time step $t_i = i\,\Delta t$ find

$$\hat{C}_i(X_i) = \sum_{r=0}^{M} \beta_{ir}\psi_i(X_i)$$

  - if $C_i(X_i) < h_i(X_i)$ or $i = M$ set

$$V_i(0) = e^{-i\,r\,\Delta t} h_i(X_i)$$

    otherwise continue to next time step
- Average the V_0 over all N paths.

Convergence and Biases

- Finding a High Bias estimator is a difficult task...
- This is because the LSM estimator has inherently both low- and high-side biases; hence, Glasserman calls it an interleaving estimator.
- As we have seen a standard technique for eliminating look-ahead bias is to calculate the exercise decision by using an additional independent set of Monte Carlo paths, thereby eliminating the correlation between the exercise decision and simulated payoff. While this two-pass approach removes look- ahead bias, it comes at the cost of doubling the computational cost, which is already heavy because the simulation of stochastic processes frequently requires the time-discretized Euler scheme.
- This is still an active field of research (see references Woo J. et al. [2019])

Regression and Mesh Weights

- In the last lesson, we said that the regression-based method corresponds to a stochastic mesh estimator with an implicit choice of mesh weights.
- We now made this explicit
- We can write the estimated continuation value at node X_{ij} as

$$
\begin{aligned}
\hat{C}_i(x) &= \psi\left(X_{ij}\right)^T \hat{\beta}_i \\
&= \psi\left(X_{ij}\right)^T \hat{B}_{\psi}^{-1} \hat{B}_{\psi V} \\
&= \frac{1}{b} \sum_{k=1}^{b} \left(\psi\left(X_{ij}\right)^T \hat{B}_{\psi}^{-1} \psi\left(X_{ij}\right) \right) \hat{V}_{i+1,k}
\end{aligned}
\tag{10}
$$

Regression and Mesh Weights

- Thus the estimated continuation value at node X_{ij} is a weighted average of the estimated option price at step $i+1$ with weights

$$W_{ij}^{k} = \psi\left(X_{ij}\right)^{T} \hat{B}_{\psi}^{-1} \psi\left(X_{ij}\right) \tag{11}$$

- In other words the Longstaff-Schwartz method is a special case of the general mesh approximation

$$\hat{C}_i(x) = \frac{1}{b} \sum_{k=1}^{b} W_{ij}^{k} \hat{V}_{i+1,k} \tag{12}$$

References

- Glasserman P. **Monte Carlo Methods in Financial Engineering** *Springer (2004)*
- Glasserman P. and Yu Bin **Number of Paths versus Number of Basis Functions in American Option Pricing** *The Annals of Applied Probability, 2004, Vol. 14, No. 4, pp. 2090-2119*
- Woo J., Liu C. and Choi J. **Leave-one-out Least Square Monte Carlo Algorithm For Pricing American Options** *arXiv:1810.02071v2 [q-fin.CP] 25 May 2019*
- Clément E., Lamberton D. and Protter P. **An Analysis of the Longstaff Schwartz Algorithm for American Option Pricing** *Finance and Stochastics, vol. 6, no. 4, pp. 449-471.*

Appendix - Derivation of Eq. 3

We want to minimise the expected squared error in our approximation with respect to the coefficients β_{ir}. So we differentiate

$$\mathbb{E}\left[\left(\mathbb{E}\left[V_{i+1}\left(X_{i+1}\right)|X_i\right]-\sum_{r=0}^{M}\beta_{ir}\psi_r\left(X_i\right)\right)^2\right]$$

w.r.t. β_{ir} and set the result equal to zero. This gives us

$$\mathbb{E}\left[\mathbb{E}\left[V_{i+1}\left(X_{i+1}\right)|X_i\right]\psi_r\left(X_i\right)\right]=\sum_{r=0}^{M}\beta_{ir}\mathbb{E}\left[\psi_r\left(X_i\right)\psi_s\left(X_i\right)\right]$$

Appendix - Derivation of Eq. 3

In matrix notation we have

$$\left(B_{\psi}\right)_{rs} = \mathbb{E}\left[\psi_r\left(X_i\right)\psi_s\left(X_i\right)\right]$$

and

$$\left(B_{\psi\,V}\right)_{r} = \mathbb{E}\left[\mathbb{E}\left[V_{i+1}\left(X_{i+1}\right)|X_i\right]\psi_r\left(X_i\right)\right]$$

However $\psi_r\left(X_i\right)$ is measurable with respect to X_i so we can write

$$\begin{aligned}\left(B_{\psi\,V}\right)_{r} &= \mathbb{E}\left[\mathbb{E}\left[V_{i+1}\left(X_{i+1}\right)\psi_r\left(X_i\right)|X_i\right]\right]\\ &= \mathbb{E}\left[V_{i+1}\left(X_{i+1}\right)\psi_r\left(X_i\right)\right]\end{aligned}$$

the last passage derives from the tower rule.

Appendix - Derivation of Eq. 3

Finally we can write

$$B_{\psi V} = B_{\psi} \cdot \beta$$

and inverting we obtain the result

$$B_{\psi}^{-1} B_{\psi V} = \beta$$