FinKont2: Hand-In 2

Gustav Dyhr / msh606

March 2025

I have created a full GitHub of my code used for this Hand-In at https://github.com/guxel/HandIn-2. The main code is in HandIn2.Py, while the rest are supporting objects. Additionally, short code snippets have been inserted where relevant.

The Hand-In references Tomas Björk's book *Arbitrage Theory in Continuous Time* (4th edition), simply refereed to as Björk. The Hand-In also references Rolf Poulsen's *Secrets of Longstaff & Schwartz aged 23 3/4: Part I, the example* (2025) and the original *Valuing American Options by Simulation: A Simple Least Squares Approach* (2001) referred to as Longstaff-Schwartz.

1 Determining Longstaff-Schwartz' σ

Rolf Poulsen's article investigates the simulated asset prices used in section 1 of Longstaff-Schwartz (among other things). The 8 price paths are visible in table 1. We wish to re-create Rolf's table 1, which contains estimates of the volatility used in the simulations and their uncertainty.

$t = 0$	$t = 1$	$t = 2$	$t = 3$
1	1.09	1.08	1.34
1	1.16	1.26	1.54
1	1.22	1.07	1.03
1	0.93	0.97	0.92
1	1.11	1.56	1.54
1	0.76	0.77	0.90
1	0.92	0.84	1.01
1	0.88	1.22	1.34

Table 1: Longstaff-Schwartz example stock price paths.

We can rightfully assume that these paths are created using Geometric Brownian Motion (GBM), and under the bankbook martingale measure. This yields the following process:

$$dS_t = rS_t dt + \sigma S_t dW_t^Q \tag{1.1}$$

We have been told in both papers that $r = 6\%$, so we simply need to determine σ. For this, the article sketches two paths. They both start with converting our asset prices to log returns, which we have done in table 2.

This is a good idea as we know that log asset returns are normally distributed in GBM. To be more precise, we know that they follow the following distribution under Q:

$$\xi_t \sim N(r - \frac{1}{2}\sigma^2, \sigma^2) \tag{1.2}$$

Where we have defined the log returns as:

$$\xi_t = \ln(\frac{S_{t+1}}{S_t}) \tag{1.3}$$

$t = 1$	$t = 2$	$t = 3$
0.0862	-0.0092	0.2157
0.1484	0.0827	0.2007
0.1989	-0.1312	-0.0381
-0.0726	0.0421	-0.0529
0.1044	0.3403	-0.0260
-0.2744	0.0131	0.1560
-0.0834	-0.0910	0.1843
-0.1278	0.3267	0.0938

Table 2: Example log returns.

And we can then determine the distribution using the closed form solution for S_t:

$$ln\left(\frac{S_{t+1}}{S_t}\right) = ln\left(\frac{S_t e^{\left(r-\frac{1}{2}\sigma^2\right)1+\sigma W_1}}{S_t}\right) = ln\left(e^{r-\frac{1}{2}\sigma^2+\sigma Z}\right) = r - \frac{1}{2}\sigma^2 + \sigma Z$$

Here Z is a standard normal random variable. We can now calculate the simple sample volatility of the log returns, dubbed the realised volatility in the article:

$$\sigma_{RV} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(\xi_t - \bar{\xi})^2} = 0.1519 \tag{1.4}$$

We can also estimate the standard error of this estimate using *Ahn & Fessler* (2003) as in the article:

$$SE(\sigma_{RV}) \approx \frac{\sigma_{RV}}{\sqrt{2(N-1)}} = 0.0224 \tag{1.5}$$

This is an approximation, based on the distribution of the variance estimate.

We have hereby completed the first step, now we just need to do the same using the Maximum Likelihood Estimate (MLE) methodology. The idea is that since the mean of the log-returns distribution also depends on σ, then our MLE is no longer the ordinary sample standard deviation. Instead we can get a better estimate by taking the mean into consideration. The idea of MLE is the following:

$$\theta = \arg\max_{\theta} \mathcal{L}(\theta, X) = \arg\max_{\theta} \prod_i f(\theta, x_i) \tag{1.6}$$

Where $f(\theta, X_i)$ is the probability density of the observation X_i ,when the distribution has the parameters θ. In this case $\theta = \sigma$. Inserting our case yields:

$$\begin{aligned} \sigma_{ML} &= \arg\max_{\sigma} \prod_i \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} \\ &= \arg\max_{\sigma} \sum_i \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{x_i - r + 0.5\sigma^2}{\sigma}\right)^2 \\ &= \arg\max_{\sigma} -n\ln\left(\sigma\sqrt{2\pi}\right) - \sum_i \frac{1}{2}\left(\frac{x_i - 0.06}{\sigma} + \frac{1}{2}\sigma\right)^2 \end{aligned} \tag{1.7}$$

We see the impact of the special μ in the last equation, as σ now appears twice. We also log transformed our equation, so we get sums that are easier to work with. This transforms our likelihood function to a log-likelihood function. We must now optimise this sum numerically to determine our σ_{ML}. We find that the optimal solution is $\sigma_{ML} = 0.1484$. We see that the result is similar to σ_{RV}, as expected. It is also slightly closer to 0.15, implying that if we assume that the true value that Longstaff-Schwartz used is 0.15, then it is slightly more accurate. This is expected as it is an unbiased estimate and by definition the most likely estimate. To determine the standard error of the MLE, we take the second derivative of our log-likelihood function:

$$\frac{d^2}{d\sigma^2}\ell(\sigma) = \frac{n}{\sigma^2} + \sum_i \frac{x_i(0.36 - 3x_i) - 0.25\sigma^4 - 0.0108}{\sigma^4} \tag{1.8}$$

As we need the square root of minus the reciprocal value of the second derivative at optimum, then we have:

$$\begin{aligned} SE(\sigma_{ML}) = \sqrt{-\left(\frac{d^2}{d\sigma^2}\ell(\sigma_{ML})\right)^{-1}} &= \sqrt{-\left(\frac{n}{\sigma_{ML}^2} + \sum_i \frac{x_i(0.36 - 3x_i) - 0.25\sigma_{ML}^4 - 0.0108}{\sigma_{ML}^4}\right)^{-1}} \\ &= \sqrt{-\left(\frac{n}{0.1484^2} + \sum_i \frac{x_i(0.36 - 3x_i) - 0.25 \cdot 0.1484^4 - 0.0108}{0.1484^4}\right)^{-1}} \\ &= \sqrt{-(-2203.3705)^{-1}} \\ &= 0.0213 \end{aligned} \tag{1.9}$$

The reason why the Fisher Information is defined as such is that the variance of our estimate is the inverse of the stability of the solution. Our solution must have first derivative (score) 0, as it is a minimum. The second derivative is therefore the sensitivity of the minimum. As the variance of our estimate is the Fisher information, then the standard error is the square root of the Fisher Information. A high stability results in a low estimate standard error. Inversely, then if large changes in σ has a small impact on the optimality of the solution, then the standard error of the estimate must also be large.

We can now finally fully recreate the table, which we have done in table 3. We see that the most reasonable value for σ is 0.15, and that it is within the standard errors of both estimates. We also note that these estimates and standard errors are the same as in the updated article.

Method	Estimate	Std' err'
σ_{ML}	0.1484	0.021
σ_{RV}	0.1519	0.022

Table 3: Volatility estimates for the Longstaff-Schwartz' example.

2 Pricing with σ

In this question we will use our volatilities to price European, Bermudan and American put options, as in the article. We have multiple options on how to price the options. One method would be solving the PDE's numerically, this would be especially easy for the European options, but harder for the others. Another would be using a binomial tree, which would be very efficient for the Bermudan options and would also work for the American if we let our time steps go to zero. Another method is to use our new tool, the Longstaff-Schwartz Method (LSM). While it yields a potentially more unstable solution due to the randomness of simulation, it is easy to implement and becomes stable as the amount of simulations increase.

We will first formalise our problem, before choosing. We know that we must be able to replicate the optimal strategy, such that we can pay if our counterpart exercises at the best time possible. We must however also be prepared to keep replicating the option after the optimal exercise time, in case they do not exercise. This is also known as replicating the martingale part of the Snell-Envelope of the discounted payoff. From *Introduction to Stochastic Calculus Applied to Finance* by Lamberton and Lapeyre, we have something similar to a formalisation of the American option pricing problem in section 2.5.1:

$$U_n = S_n^0 \sup_{\nu \in \mathcal{F}_{n,N}} \mathrm{E}^* \left[\frac{Z_\nu}{S_\nu^0} | \mathcal{F}_n \right] \tag{2.1}$$

Which is similar to the discrete time (and substantially nicer) formalisation in *American π: Piece of Cake* by Rolf Poulsen and used formulas that can be found in Björk. What is important here is that, while we do effectively replicate U_n for $n < \nu$, we never actually exercise unless our counterpart does. We define our problem as having to price U_n, so that we are prepared in the worst case where the non-increasing part $-A_n$ is zero. The replication is therefore focused on the martingale part and any loses through A_n will be carried solely by the option holder. We can rewrite the equation as something more familiar and discounted by the bank account with a fixed interest rate:

$$\pi_{AM}(t) = e^{-r(T-t)} \sup_{\tau \in \mathcal{F}_{t,T}} \mathrm{E}_t^Q \left[g(S_\tau) \right] \tag{2.2}$$

Here the we can easily bridge the gap to European and Bermudan options by changing the values that τ can take.

$$\pi_{BM}(t) = e^{-r(T-t)} \sup_{\tau \in B} \mathrm{E}_t^Q \left[g(S_\tau) \right] \tag{2.3}$$

$$\pi_{EU}(t) = e^{-r(T-t)} \mathrm{E}_t^Q \left[g(S_T) \right] \tag{2.4}$$

Where (2.4) is simply the standard risk-neutral valuation formula for European options. B is the set of time-points where excercise is possible for the Bermudan option. In our case $B = \{1, 2, 3\}$. (2.2) and (2.3) are however harder to deal with, so both the binomial model and LSM attempt to discretise the problem:

$$\pi_{AM}(t) = \max \Big(g(S_t), e^{-rdt} \mathrm{E}_t^Q \left[\pi_{AM}(t + dt) \right] \Big) \tag{2.5}$$

An equivalent for this is found as the definition of U_n in *Introduction to Stochastic Calculus Applied to Finance*. Naturally, we again make sure not to liquidate our hedging portfolio before the option is exercised. It is clear that this must be calculated backwards, and that it only holds for $dt \to 0$. For Bermudan options we can simply use (2.5) in LSM for the t where exercising is possible by manipulating dt (assuming they the time points are equidistant). When using models such as the binomial model that approximates continuous processes in discrete time, it would still be required to use a small dt for simulation and then only do the early exercise checks at exercise dates. It is simple to evaluate $g(S_t)$, but $\mathrm{E}_t^Q \left[\pi_{AM}(t + dt) \right]$ is more difficult. Looking at the second-last time point where $t = T - dt$, we have:

$$\begin{aligned} \pi_{AM}(T - dt) &= \max \Big(g(S_t), e^{-rdt} \mathrm{E}_t^Q \left[\pi_{AM}(t + dt) \right] \Big) \\ &= \max \Big(g(S_{T-dt}), e^{-rdt} \mathrm{E}_t^Q \left[\pi_{AM}(T) \right] \Big) \\ &= \max \Big(g(S_{T-dt}), e^{-rdt} \mathrm{E}_{T-dt}^Q \left[g(S_T) \right] \Big) \end{aligned} \tag{2.6}$$

It is clear that if we can evaluate $\mathrm{E}_t^Q\left[f(S_{t+dt})\right]$ then we can evaluate $\mathrm{E}_{T-dt}^Q\left[g(S_T)\right]$ and the following steps $\mathrm{E}_t^Q\left[\pi_{AM}(t+dt)\right]$, for all t. The main difference between Longstaff-Schwartz Model and the binomial model is exactly how they calculate this expectation. The binomial model uses a simple two outcome model, in this case for GBM:

$$\begin{aligned}\mathrm{E}_t^Q\left[f(S_{t+dt})\right] =& q_{up} f(S_{up,t+dt}) + q_{down} f(S_{down,t+dt}) \\ =& \left(\frac{e^{rdt} - e^{-\sigma\sqrt{dt}}}{e^{\sigma\sqrt{dt}} - e^{-\sigma\sqrt{dt}}}\right) f\left(S_t e^{\sigma\sqrt{dt}}\right) + \left(1 - \frac{e^{rdt} - e^{-\sigma\sqrt{dt}}}{e^{\sigma\sqrt{dt}} - e^{-\sigma\sqrt{dt}}}\right) f\left(S_t e^{-\sigma\sqrt{dt}}\right)\end{aligned} \tag{2.7}$$

As defined in section 7.1 in *Finance 1 and beyond* by Rolf Poulsen and David Lando.

The LSM model uses a much simpler method for calculating this, it first simulates N independent stock paths. It then fits linear regressions attempting to estimate the conditional expectation $\mathrm{E}_t^Q\left[f(S_{t+dt})\right]$ using only transformations of S_t. We therefore have:

$$\mathrm{E}_t^Q\left[f(S_{t+dt})\right] = E[f(S_{t+dt}|S_t)] = \hat{F}_M(\omega, t) \tag{2.8}$$

Where $\hat{F}_M(\omega, t)$ is the time t fitted value of $f(S_{t+dt})$ using M basis functions (transformations of S_t in our regression). Where note that here $f(S_{t+dt})$ never directly takes on the value $\pi_{AM}(t+dt)$, instead to avoid biases, it is always a discounted cash-flow from the option being exercised at the optimal time in the future. From Longstaff-Schwartz proposition 2 we know that, with some reasonable constraints, the LSM method will converge on the actual option value for a large enough N and a good choice of basis functions M.

It seems very fitting to use our new toy LSM, so we will continue with this method. We use the below code for implementing LSM:

```python
def priceDefault(option: claim,Sim,dt,r,regMethod = "simple",**v):
    """
    Use Longstaf-Schwartz for pricing non-path dependtent options
    """
    n,T = Sim.shape
    CF = np.zeros((n,T-1))

    payOff = option.payoff(S=Sim[:,-1],**v)
    CF[:,-1] = payOff
    DF = np.exp(-r * dt * np.arange(1,T))

    for t in range(T-2,0,-1):
        payOff = option.payoff(S=Sim[:,t],**v)
        ITM = payOff>0

        X = getDesign(Sim[ITM,t],regMethod)
        regCoef = MatrixReg(X,np.dot(CF[ITM,t:],DF[:T-t-1]))

        excercise = np.repeat(False,n)
        excercise[ITM] = payOff[ITM] > np.dot(X,regCoef)

        CF[excercise,:] = 0
        CF[excercise,t-1] = payOff[excercise]


    price = np.mean(np.dot(CF,DF))
    return price
```

The methods works by determining the optimal exercise decisions while using the discrete valuation method in (2.5), and the regression method in (2.8). We initially use the simple regression model described in the Longstaff-Schwartz paper:

$$\mathbf{E}[Y|X] = \beta_0 + \beta_1 X + \beta_2 X^2 \tag{2.9}$$

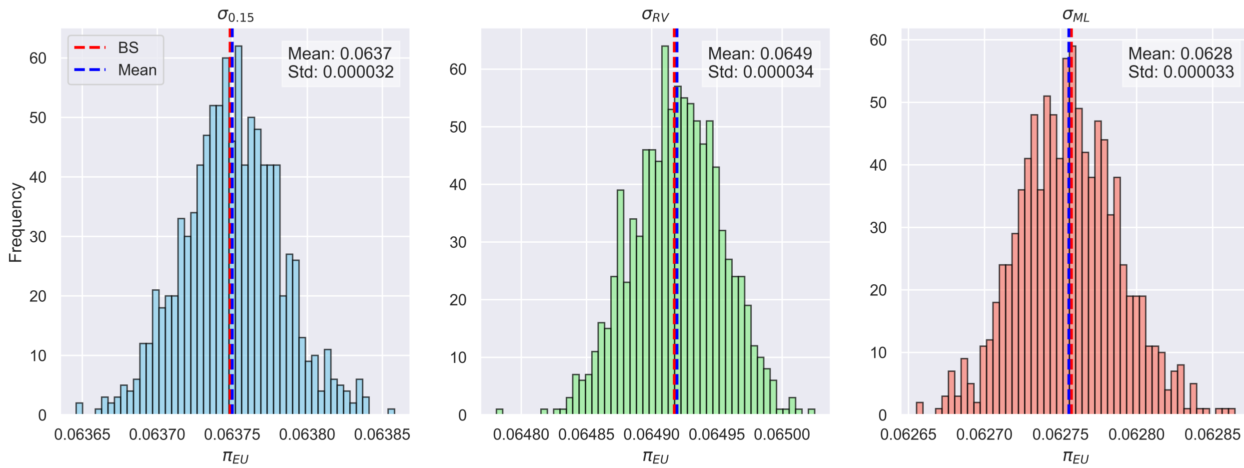

Figure 1: LSM method used on European options with 1,000 prices for each option, each using 10,000,000 simulations

As we wish to model our assets using GBM under Q, then we use the below solution for simulating the asset prices, from Björk's proposition 5.2:

$$S_t = S_0 e^{\left(r - \frac{1}{2}\sigma^2\right)t + \sigma W_t} \tag{2.10}$$

For each option we will calculate multiple different prices, this is an attempt to handle the randomness of the method. We could also instead increase our amount of simulation paths, but due to memory limitations. we have trouble handling more than 10,000,000 at once. Multiple price simulations also gives us an understanding of the errors of ours algorithm., We begin with the European options.

To price European options using LSM we simply have to set our time step dt to $T = 3$. This means that we will effectively be taking the price as the discounted expected payoff under Q, where we evaluate the expectation through taking the mean of simulations. It is therefore following the standard risk-neutral pricing formula, we simply use another method for evaluating the expectation, instead of trying to solve it analytically. We have plotted a histogram of our LSM prices in figure 1 and for comparison we have included the prices according to the Black-Scholes formula.

We see that the prices appear to be normally distributed, with relatively small standard deviations and no apparent tails or skewness. The mean price of each option is approximately equal to the similar Black-Scholes price. The differences are at around $1e-7$. The standard deviation of the prices are increasing in σ, as a larger volatility in our asset price process increases the size of the prices and the volatility of the simulations. We also note that the mean appears to be a slightly better predictor than the median, showing resilience against outliers. We have found the prices:

$$\pi_{EU}(\sigma_{0.15}) = 0.0637$$

$$\pi_{EU}(\sigma_{RV}) = 0.0649$$

$$\pi_{EU}(\sigma_{ML}) = 0.0628$$

They are all about one fourth decimal below the prices in the article. This is an interesting thing that will go again for all of our estimates. Even when using Black-Scholes and comparing with the old article we get:

$$\pi_{EU}^{BS}(\sigma_{0.20}) = 0.0952$$

Exactly 0.0001 below the number in the old article. Investigating the articles attached code convinces us that this is the result of the articles using the binomial model to price the options. This leads to approximation errors as the model attempts to approximate something contentious in discrete time. It is not a significant difference and we are satisfied with our results.

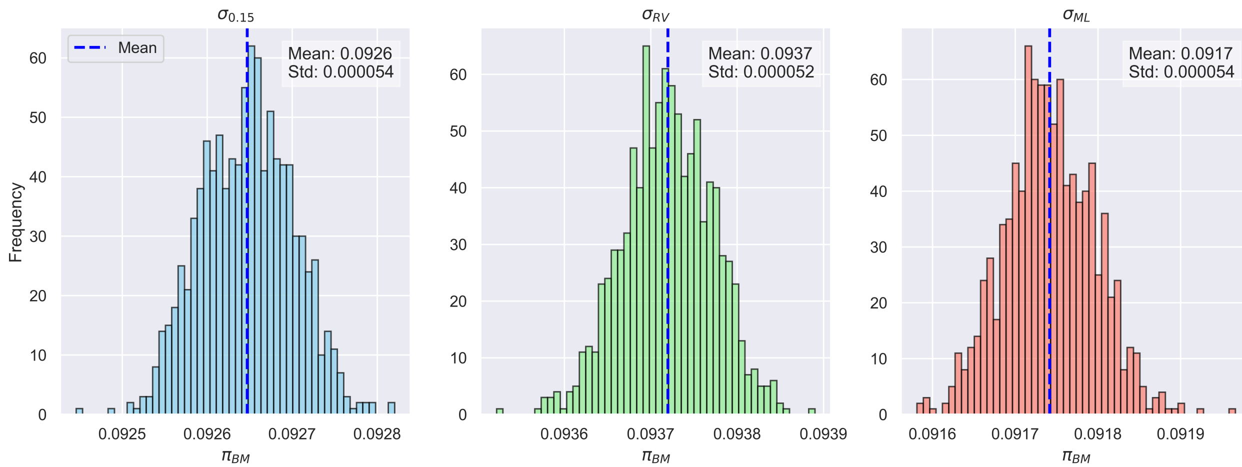

Figure 2: LSM method used on Bermudan options with 1,000 prices for each option, each using 3,333,333 simulations.

We now use the method for the Bermudan options. The Bermudan options allow for exercising at the end of every time period ($t = 1, 2, 3$), so we incorporate this discrete optionality by setting $dt = 1$. Therefore, we get a LSM model similar to the Longstaff-Schwartz exampl. We compute if it's optimal to exercise at 2 time-points, given the conditional expectation of the future option payoffs. We have plotted the results in figure 2.

The prices again appear to be following pretty normal distributions, now with higher standard deviations. This is likely both caused by the fewer simulations and the added complexity. Extracting the means, we find the prices:

$$\pi_{BM}(\sigma_{0.15}) = 0.0926$$

$$\pi_{BM}(\sigma_{RV}) = 0.0937$$

$$\pi_{BM}(\sigma_{ML}) = 0.0917$$

We see that these prices are higher than for the European options. This is expected as we now have more optionality. The options can be exercised early for greater value in most paths. We also note that they again are exactly one 4'th decimal below the numbers in the article.

We now look at the American option. If we assume that the time-unit is in years, then the option has a 3-year maturity. This is a fair assumption, as $r = 6\%$ looks like a yearly interest rate. In that case, a reasonable choice for dt for an American option might be $1/252$, which would equal an evaluation on if the option should be exercised on every (trading) day. The only real reasonable dt for American options is $dt \to 0$, but as we have limited patience, and as the article uses the same time gap, then we will use $dt = 1/252$. As we now have significantly more options to evaluate in our LSM model, we will also have to bring down the amount of simulations in each batch for the same reason. We plot the results in figure 3.

The prices are again normally distributed, but this time with significantly higher deviations. This as a significantly lower amount of simulations used. The added complexity also does not help. We find the prices:

$$\pi_{AM}(\sigma_{0.15}) = 0.1088$$

$$\pi_{AM}(\sigma_{RV}) = 0.1096$$

$$\pi_{AM}(\sigma_{ML}) = 0.1081$$

Interestingly, we see that the difference between our fitted values and those in the new article has grown from 0.0001 to about 0.0004. This is a systematic bias and is immune against new simulations and sampling. There are multiple potential reasons that could be the cause for this.

First, we have to remember that we are approximating an American option with a Bermudan option with many exercise opportunities. As the amount of exercise opportunities increase, the optionality of the option increases and

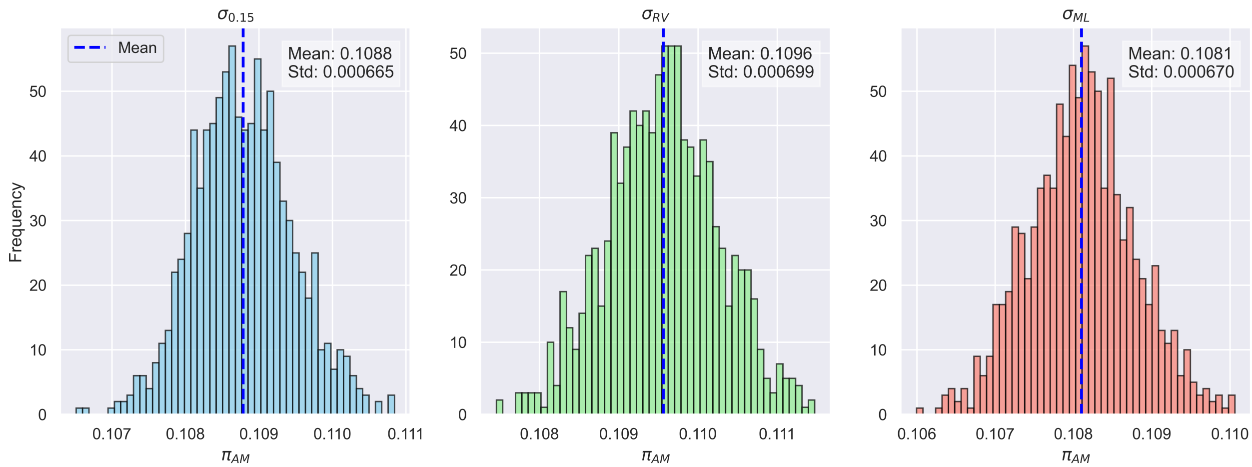

Figure 3: LSM method used on American options with daily exercise evaluations and 1,000 prices for each option, each using 10,000 simulations.

thereby the price. This will then converge on the price of the American option. Knowing this, a cause could be a difference in the amount of exercised opportunities used to approximate the continuous exercising. This is however not the case, as we used $252 \cdot 3$ time points like the article.

Another potential reason is bias in the LSM implementation. As we will see in the next question, calculated the expected future value of the option is not easy and bias can sneak in from multiple sources. We know from Longstaff-Schwartz proposition 1 that with a large enough N, LSM will yield a price smaller than or equal to the correct price. We also know from proposition 2 that for a fitting set of basis functions, the error will converge to 0. This is something that happens as we need a strong regression to be able to properly approximate the optimal strategy, all other guesses will be sup-optimal (unless we start fitting to the noise, as we will examine closely in question 3). While the second degree polynomial model is very stable for a low amount of simulations, we have more now so we can attempt using a stronger regression method with more basis functions. To choose we plot some alternatives in figure 4.

We see that the impact of the choice of regression appears to grow with the time till maturity. We see that using a model with an intercept and three weighted Laguerre polynomials strikes a nice balance between optimal fitting, intuitive development and no bends in the ends caused by over-fitting. This is also the regression model used for pricing american put options in section 3 of Longstaff-Schwarz. One can see in the first and second plot that the bends in the ends of the higher degree fits can be very problematic. This happens to the fifth degree polynomial as it over-fits to the noise and thereby gains extreme and unintuitive developments for very low X. This can have a high influence on if some far ITM are exercised very early. It should however also be noted that the size of difference between the exercise now value and the fitted future value is unimportant. Only the direction matters (positive vs. negative difference). We still choose to continue with the simpler model to avoid any potential over-fitting:

$$\mathbf{E}[Y|X] = \beta_0 + \beta_1 e^{-X/2} + \beta_2 e^{-X/2}(1-X) + \beta_3 e^{-X/2}(1-2X+X^2/2) \tag{2.11}$$

To avoid over-fitting, we also further increase our amount of simulated paths to 100,000. Meanwhile we have half the paths as antithetic to reduce Monte-Carlo sampling variance. This time we only calculate 100 prices for each σ to limit computing time. We see the results in figure 5.

The distributions look less normal, due to their small sample sizes, but the standard deviation of the prices are significantly smaller than before. This indicates that our new optimal exercise strategy is more stable and optimal. The larger amount of simulation paths also have a direct effect on bringing down the volatility of the price estimates. The small sample size does however limit the precision of our mean. We note the mean prices:

$$\pi_{AM}^{Better}(\sigma_{0.15}) = 0.1091$$

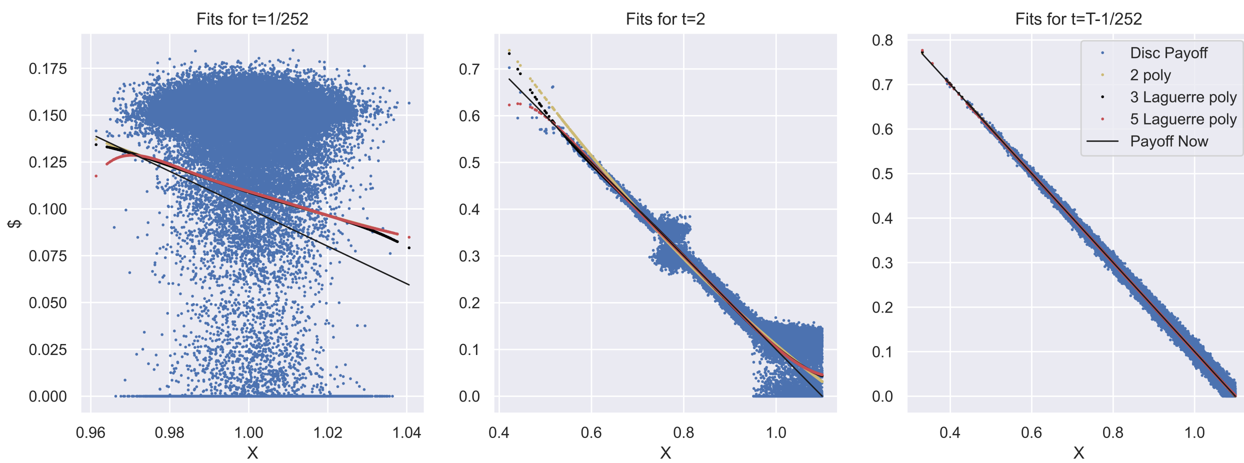

Figure 4: Different regressions for the expectation of the discounted option payoff, conditioned on the current asset price. Done at respectively time $t = T - 1/252$, $t = 2$ and $t = T - 1/252$.

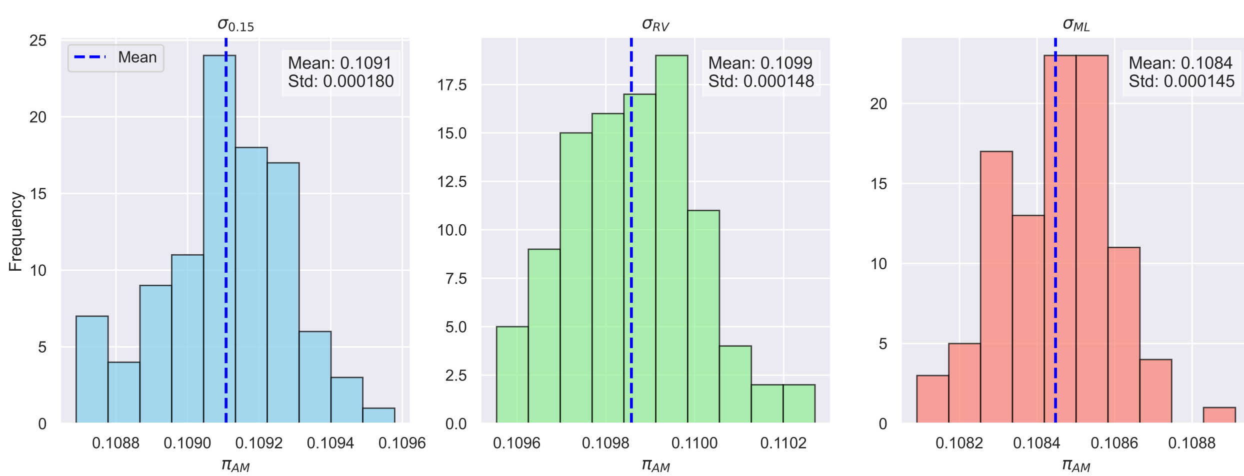

Figure 5: LSM method used on American options with daily exercise evaluations and 100 prices for each option, each using 100,000 simulations. Conditional expectations fit using intercept and third degree weighted Laguerre polynomials.

$$\pi_{AM}^{Better}(\sigma_{RV}) = 0.1099$$
$$\pi_{AM}^{Better}(\sigma_{ML}) = 0.1084$$

All are exactly 0.0001 below the prices in the article, like our other options. As including more basis functions in our fit would lead to risk of over-fitting, and as the errors are equal to our other errors, we accept these prices as true and consider the algorithm as converged. We also note that the prices are above $K - S_0 = 0.1$, so there is no wish to exercise instantly. If the opposite was the case, then we would have to correct our prices to $\max(\pi_{AM}, K - S_0)$ as LSM cannot handle instant exercise.

We have calculated all our prices and collected the results in table 4.

Method	σ	American	Bermudan	European
ML	0.1484	0.1084	0.0917	0.0628
Guess	0.1500	0.1091	0.0926	0.0637
RV	0.1519	0.1099	0.0937	0.0649

Table 4: Option prices using the volatility estimates from the Longstaff-Schwartz' example.

We see that as expected, the prices are increasing in both volatility and optionality. The American option with high volatility is the most expensive, and the European option with low volatility is the cheapest. It is also visible that the change from European to Bermudan has a greater impact than the change from Bermudan to American. This implies that the marginal utility of optionality decreases strongly as the amount of exercise opportunities increase. We also observe that impact of the changes in volatility on the price decreases as the optionality increases. An intuitive explanation would be that some Bermudan and many American options are exercised early, and thereby they do not experience the full impact of the volatility change under their shorter lifespan.

We also note that the Longstaff-Schwartz example price 0.1144 is a fair bit away from the actual price $\pi_{BM}(\sigma_{0.15}) = 0.0926$, which can be expected from such a small simulation. In question 4, we will find out just how likely this price is.

3 Expanding the regression

We must investigate what happens in the Longstaff-Schwartz example when we expand the regression model to also involve polynomials of higher degrees. We must also examine if using a build-in regression function changes anything.

We use the script from question 2 and the simulations from question 1. We start by using the standard model defined in equation (2.9). This yields the price:

$$\pi_{2D}(\text{Example}) = 0.1144$$

Exactly as in Longstaff-Schwartz. We can now attempt to increase the amount of polynomials to three, yielding the conditional expectation formula:

$$\mathbf{E}[Y|X] = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 \tag{3.1}$$

This results in the price:

$$\pi_{3D}(\text{Example}) = 0.1154$$

And continuing we get:

$$\pi_{4D}(\text{Example}) = 0.1243$$

$$\pi_{5D}(\text{Example}) = 0.1116$$

We see that the prices increase substantially, and then they decrease again. This signals instability and over-fitting. But before we can conclude on what is going on, we should examine the regression coefficients. We have done so in table 5. In the last column we have included the estimation error, comparing the estimate to the real price of a Bermudan option with $\sigma = 0.15$ ($\pi_{BM}(\sigma_{0.15}) = 0.0926$).

Degree	Regression Coefficients						MSE	Pred Error
0	0.10						4.80e-03	0.022
1	0.47	-0.39					2.46e-03	0.023
2	-1.07	2.98	-1.81				2.18e-03	0.022
3	49.12	-162.26	178.01	-64.70			7.95e-05	0.023
4	213.78	1005.47	-1756.42	1351.69	-386.76		8.61e-15	0.032
5	-103.41	410.93	-482.24	-7.2	334.57	-152.5	3.84e-03	0.019

Table 5: Coefficients and errors of different regression models in the Longstaff-Schwarz example. Regression coefficients and MSE are for $t = T - 1$.

We see that the last model has significantly higher MSE and the earlier ones. This should be impossible, as in the worst case it should set $\beta_5 = 0$ and have the same MSE as the fourth degree polynomial model. This happens due to the high collinearity between the variables, making the inverse transformation problematic numerically. We try fitting the same regressions using scikit-learn's *LinearRegression* (Python alternative to *lm*). We show the results in table 6. The coefficients deviate slightly for the fourth degree polynomial regression, and strongly for the fifth degree polynomial regression. As the scikit-learn function uses Singular Value Decomposition (SVD), which is more stable, it handles the problematic design matrices better. We will not describe SVD in detail, but it is a method for decomposing a matrix into orthogonal matrices and a diagonal matrix of singular values (not to be confused with singular matrices, which is something very different. There is however a connection.). One can then efficiently compute the pseudo-inverse of the decomposition, while handling singularities in the original matrix. The pseudo-inverse can then be used for further computing in the regression.

Comparing the in-sample MSE (objective) of these two models, we see that the SVD one strongly outperforms with 6.862e-27 against 0.0038. We will therefore use the second table for interpreting, it is however important that one is aware of what method they implement and the weaknesses/strengths of the selected method has as while they attempt to solve the same problem, they are definitely not equivalent.

Degree	Regression Coefficients						MSE	Pred Error
0	0.10						4.80e-03	0.022
1	0.47	-0.39					2.46e-03	0.023
2	-1.07	2.98	-1.81				2.18e-03	0.022
3	49.12	-162.26	178.01	-64.70			7.95e-05	0.023
4	213.78	1005.48	-1756.43	1351.70	-386.76		6.79e-26	0.032
5	-88.18	329.7	-308.73	-191.97	432.61	-173.23	6.86e-27	0.032

Table 6: Coefficients and errors of different regression models in the Longstaff-Schwarz example. Regression coefficients and MSE are for $t = T - 1$. Calculated using SVD.

These results are easier to interpret, the errors appear to have a minimum at the intercept only model or the second degree polynomials and after that they increase. We also see that the regression coefficients explode in size simultaneously with this. We note two reasons for this. One is the high collinearity between the different polynomials, the other is the low amount of observations. This leads to high variance inflation and general over-fitting. This would be a much greater problem for out of sample prediction, but it is still an issue for our in-sample-fitting.

There is a lot of noise between our current asset prices X, and the future payoffs after the assets have moved randomly. By including more regressors, we start modelling this noise. This is very evident in the error column, where the models with third, fourth and fith degree polynomials have large errors. These errors are all from estimated prices larger than the actual price. This happens as the extra information that the model fits about the random development in the future is equivalent to the option holder having information on what is likely to happen in the future. Therefore, the regression deviates from simply modelling the expected future value given to current asset price, and becomes closer to the expected future value given the future asset price. This is obviously not ideal, as it allows for too optimal exercise strategies and therefore too high of a valuation of the option. This is only the case for regressions with few observations. If we had more, then all the paths would unfold independently and one would need even more variables to model the randomness in the different paths.

Interestingly the error for the fourth degree polynomial regression and the fifth degree are equal. In figure 6 we have plotted the different fitted values in the time 2 regression, in an attempt to understand why.

We see a lot of interesting things in the regression plot. The main line of interest is the straight black *Payoff Now* line indicating the value of exercising the option at time 2. The LSM algorithm will exercise for all fitted values below this line, and refrain from exercising for all fitted values above. Naturally, the optimal exercise strategy would be using the *Disc Payoff* as our indicator, but that would require knowing exactly what happens in the future. Therefore, the models that exactly stays on the same side of the black line as the *Disc Payoff* follows the optimal strategy (and therefore make use of future information, and have an upward bias in their prices). We see a lot of unnatural things in the models with higher than second degree polynomials, such as periods where the expected future payoff increases significantly, while the price of the underlying also increases. This is highly unreasonable for put options and a clear result of modelling the randomness. We can also see here why the fourth and fifth degree polynomial model gives the same price, as they both manage to stay on the "correct" side of the payoff line in every simulation. They therefore correspond to models where one knows exactly what happens in the future, with very little uncertainty. One must also note that the time 1 regression also plays a part in the pricing, which we have not graphed here. The result should however be very similar.

Looking at Longstaff-Schwartz' comments on the choice of regression model in section 2.2, they are relatively open and simply recommend different possible series of transformations of X ("basis functions"). They do however mention that the fitted values only converge on the conditional expectation if the amount of simulations goes towards infinite. 8 is clearly not enough. In addition, they name a method for checking the stability of our fit, which we will implement in question 5.

Because the models with higher than second degree polynomials fit to the noise, gain bias and act unreasonably, we deem them as ill-fitting. As the relationship between the future payoff and the current asset price is non-linear (due to the optionality making OTM payoffs 0 and not negative), our best model for the example **is** the two polynomial model used by Longstaff-Schwartz in their example.

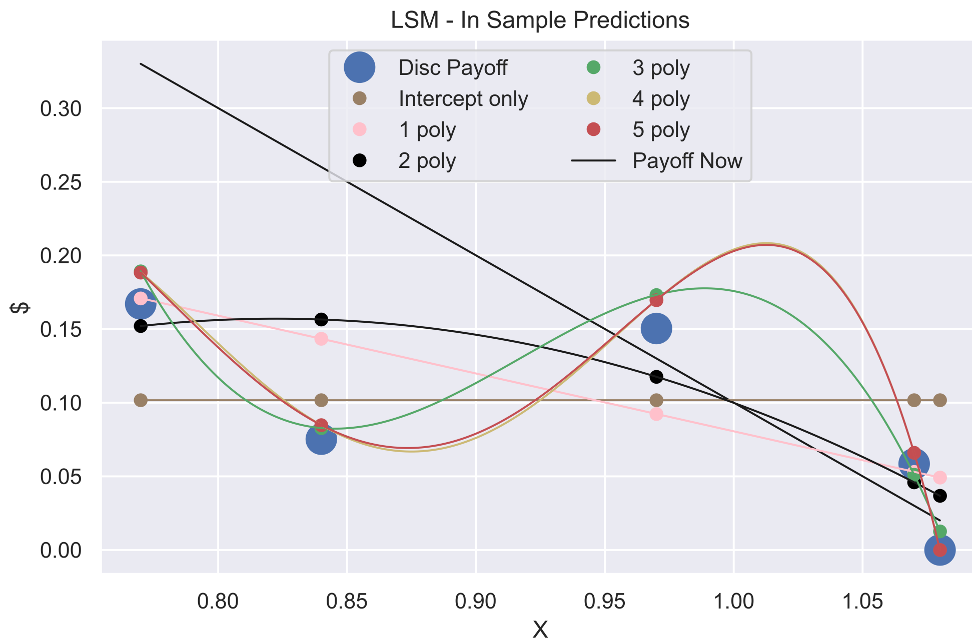

Figure 6: The time T discounted payoff modelled for time $T-1$ ITM options using the time $T-1$ asset price.

4 Recreating the Example

We now have the repeat the Longstaff-Schwartz example, but with our own simulated data. We again only use 8 simulation paths each time and only 3 time points. We use our LSM method from question 2. While the question asks for $\sigma = 0.2$, we assume that it means $\sigma = 0.15$ as in the example. We here note that repeating this experiment yields slightly different results. We could increase the amount of simulated prices for a higher accuracy, but as the question asks for 10,000, then we will use that and **assume that our result matches the real distribution**, as it would for a higher amount of simulations.

We plot the results of the experiment in figure 7. We do not use antihetic simulation, to get the full randomness of ordinary Monte-Carlo simulation. We see that it looks normally distributed, but a closer inspection reveals a right skew. The skew is formed by the boundary at 0, as the payoffs, and therefore option values, cannot go negative. They also have an upper limit as the put payoff never can exceed the strike $K = 1.1$, but all simulations are far from that limit. The excess kurtosis is close to 0, indicating tails no heavier than that of the normal distribution. We therefore believe that the simulations follow a right skewed normal distribution. We have fit such a distribution in black and we note that it appears to fit the simulated distribution extremely well.

We note that the sample expected value of the distribution is:

$$\mathrm{E}[\mathrm{dist}] = 0.1069 \tag{4.1}$$

Which is a step above the actual price 0.0926 (Assuming that our Bermudan price from question 2 is correct). This is caused by both the randomness of the distribution (resampling yields another mean) and the in-effectiveness of the conditional expectation. Not being able to fit the optimal exercise strategy likely adds a downward bias. There is however a much greater upward bias, from over-fitting in cases where there are very few ITM paths. In these cases the option will be exercised too optimally, using future information as we saw happen in question 3. While this isn't a big problem for a second degree polynomial regression in other cases, there will be plenty of simulations with few paths in the money, including many with 2 or less.

The sample standard deviation of the distribution is:

$$\sqrt{\mathrm{Var}[\mathrm{dist}]} = 0.0363 \tag{4.2}$$

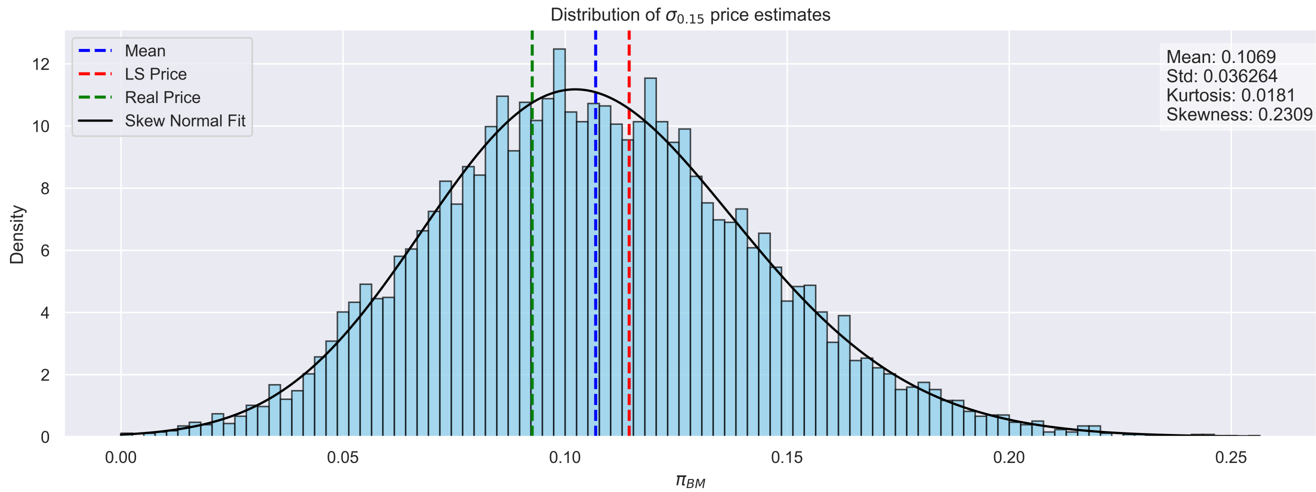

Figure 7: Distribution of the Longstaff-Schwartz example.

Which is very high compared to the mean and to the standard deviations that we got in question 2. We also see that the real price is within 1 standard deviation of the mean. So is the price that Longstaff-Schwartz got in their example, 0.1144. We can actually determine the probability of getting a price that big or bigger using our distribution:

$$\mathrm{P}[\pi_{BM}^{\text{sample}} \geq 0.1444|\text{dist}] = 40.75\% \tag{4.3}$$

We can therefore conclude that the result that Longstaff-Schwartz got is relatively close to the median and a likely draw from the distribution. The 8 simulation LSM also gives a result surprisingly close to the actual value when taking a mean of multiple samples, but with an upward bias from over-fitting. The individual samples are not very trust-worthy, due to the high standard deviation of the distribution.

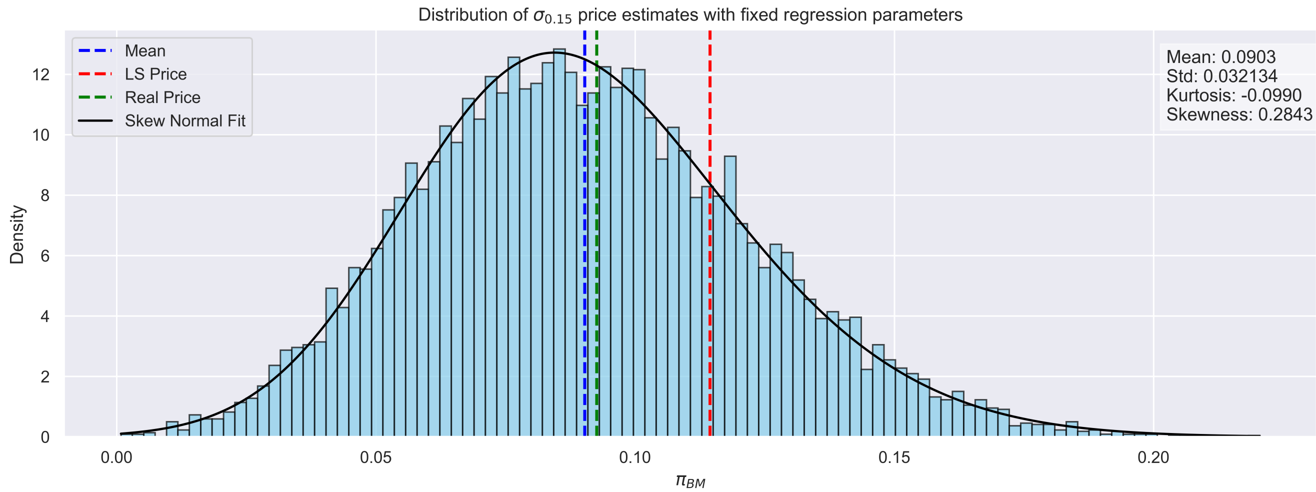

Figure 8: Distribution of the Longstaff-Schwartz example with fixed regression coefficients equal to those of the example.

5 Fixed regression

We now attempt to repeat the distribution experiment, but we fix the regression coefficients to those from the Longstaff-Schwarz example:

$$\mathrm{E}_1[Y|X] = 2.037512 - 3.335443X + 1.356457X^2 \tag{5.1}$$

$$\mathrm{E}_2[Y|X] = -1.069988 + 2.983411X - 1.813576X^2 \tag{5.2}$$

Doing this is (as mentioned in Longstaff-Schwarz chapter 3) a good way to test the stability of a solution. By taking the optimal stopping model (the conditional expectations) from one set of paths and applying them to another set of paths, we can investigate if the prices change significantly. If our solution is stable, then the out-of-sample prices should be approximately the same as the in-sample one. We compute the experiment and display the results in figure 8. We have again calculated 10,000 simulates, and while the result varies highly, we will pretend that our simulation reflects the actual distribution perfectly.

We see a shifted and even more skewed distribution. The expected price is now 0.0903, very different to both the price before and the price in the example. This indicates that the example unsurprisingly has not converged on the true solution, the example's price was heavily affected by bias from over-fitting. The new price is however extremely close to the correct option price.

We also note that the method has slightly higher stability, expressed in a smaller variance and kurtosis. The change in standard variation is also partly caused by the generally smaller prices, leading to a smaller scale. The kurtosis is driven more by the fewer extremely high prices caused by over-fitting. The skewness still increased, as the distribution is closer to the lower bound 0.

The model is no longer able to over-fit to the noise and thereby predict exactly what will happen in the next period. We therefore simply have the downward bias left, from a suboptimal exercise strategy. This is possible to bring down be fitting in a simulation with more paths and possibly also by including more basis functions in the regression. If it had been the optimal exercise strategy, then the Longstaff-Schwarz example price should also have been close to the median of our distribution, as the solution should have been stable. The miss-pricing is however small in size at around \$0.002 or 2.5%.

We can therefore conclude that result in question 4 was dominated by over-fitting upward bias and that by using fixed regression parameters we can purge this bias. We however instead gain a downward bias, as our fixed regressions are less optimal than the optimal exercise strategy.

If we out of interest extract the regression coefficients from one of the simulations that we made in question 2, we see a better result. The simulations each used 3,333,333 paths and selecting one at random and applying it yields figure 9. We see that this new distribution has a mean exactly matching the same option price as the original.

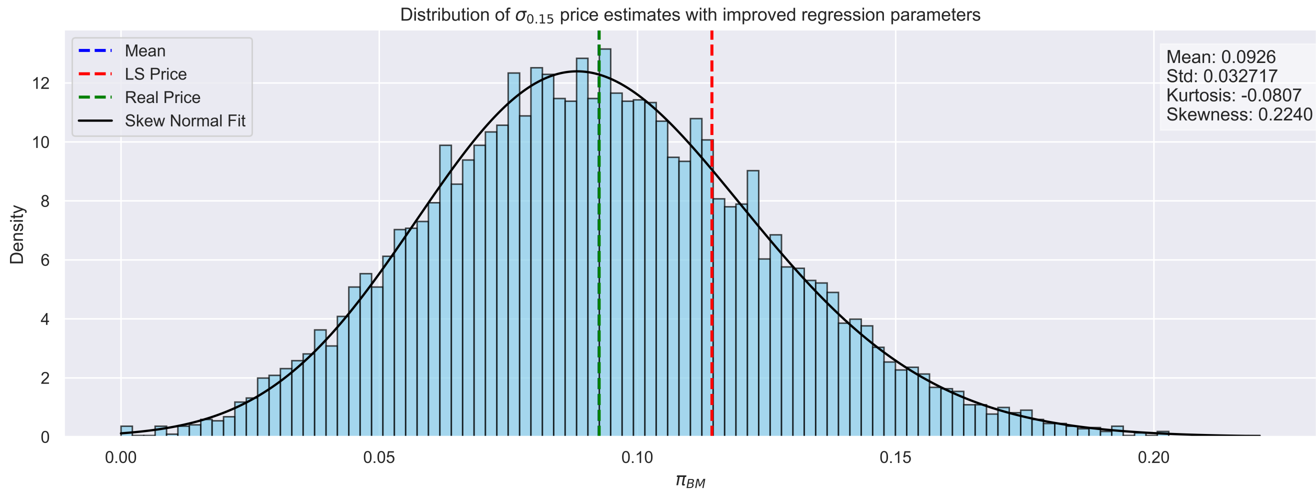

Figure 9: Distribution of the Longstaff-Schwartz example with fixed regression coefficients from a large simulation.

We therefore now have a model with no bias, and we can conclude that our LSM model in question 2 converged on the real price with the optimal exercise strategy and no over-fitting. It is therefore possible to estimate (or rather approximate) the optimal exercise strategy using a second degree polynomial for a three step Bermudan option, as long as the amount of simulation paths is large enough. This can then be used for further simulations with less paths, which will yield an unbiased estimate of the option price (with some random noise with mean 0 and size depending on the amount of simulations).

We have hereby pulled apart the Longstaff-Schwarz example (question 1-3), recreated it (question 4) and fixed it (question 5).