Gaussian Mixture Models for Estimating Conditional Expectations in the Context of Option Valuation

by

Vésteinn Þrymur Ólafsson

Thesis of 30 ECTS credits submitted to the School of Technology, Department of Engineering
at Reykjavík University in partial fulfillment
of the requirements for the degree of
Master of Science (M.Sc.) in Financial Engineering

May 2023

Supervisor:

Ralph Rudd, Supervisor
Assistant Professor, Reykjavík University, Iceland

Jörg Kienitz, Co-advisor
Assistant Professor, University of Wuppertal, Germany

Examiner:

Alex Backwell, Examiner
Senior Lecturer, University of Cape Town, South Africa

Gaussian Mixture Models for Estimating Conditional Expectations in the Context of Option Valuation

Vésteinn Þrymur Ólafsson

May 2023

Abstract

In this dissertation, we review the industry-standard method to approximate the conditional expectations that frequently arise in quantitative finance and compare it to an emerging method that promises both more efficient and more accurate calculations, namely the Gaussian mixture model. The method is data-driven and leads to an analytic expression for a hedging strategy with respect to a chosen proxy. We implement a pricing algorithm for both methods and price American options using both geometric Brownian motion paths and the Heston model. To benchmark our results, we use analytical calculations where available or turn to well-known software packages like Quantlib. We also consider the pricing of "best of" rainbow options using the Gaussian mixture model and the resulting minimal variance deltas and compare them to their respective finite-difference deltas.

Mat skilyrtra væntigilda með Gaussian mixture modeli fyrir verðlagninu valréta

Vésteinn Þrymur Ólafsson

maí 2023

Útdráttur

Í þessari grein könnum við þekkta aðferð Longstaff og Schwartz til þess að meta skilyrt væntigildi sem eru algengt viðfangsefni í fjármálastærðfræði og berum saman við nýkynnta aðferð sem bæði er skilvirkari í notkun og skilar nákvæmari niðurstöður, þ.e. Gaussian mixture model. Aðferðin er drifin áfram af gögnum um undirliggjandi gerninga og leiðir af sér einfalda jöfnu fyrir áhættuvörn með tilliti til undirliggjandi eignar. Við setjum upp verðmatslíkan fyrir báðar aðferðir og verðmetum amerískan valrétt fyrir verðtímaraðir úr bæði geometric Brownian motion og Heston líkani. Þá notum við verðjöfnur sem viðmið þegar þær eru fáanlegar, annars notum við þekkta forritunarpakka á borð við Quantlib til þess að meta frammistöðu aðferðanna. Loks skoðum við verðmat svokallaðra regnbogavalrétta og þá lágmarks dreifni áhættuvörn sem fæst með Gaussian mixture model aðferðinni.

Gaussian Mixture Models for Estimating Conditional Expectations in the Context of Option Valuation

Vésteinn Þrymur Ólafsson

Thesis of 30 ECTS credits submitted to the School of Technology, Department of Engineering
at Reykjavík University in partial fulfillment of
the requirements for the degree of
Master of Science (M.Sc.) in Financial Engineering

May 2023

Student:

....
Vésteinn Þrymur Ólafsson

Supervisor:

....
Ralph Rudd

....
Jörg Kienitz

Examiner:

....
Alex Backwell

....
date

....
Vésteinn Þrymur Ólafsson
Master of Science

Contents

List of Figures

Chapter 1

Introduction

Estimating conditional expectations is a central theme in many problems in quantitative finance [1] [2] [3], for example, in the case of pricing American options. In this article, we investigate an emerging method based on the familiar multivariate Gaussian, namely Gaussian mixtures, and compare it to the current industry standard method. The technique is model-free, that is, it assumes no stochastic dynamics for the underlying but instead relies on observed realizations. Aside from the fitting of the model, the method uses analytical calculations to produce the conditional estimate and additionally provides a smooth proxy hedge. First, we briefly review the Black-Scholes option theory [4] and the associated stochastic dynamics. We then proceed to the American option pricing problem. Having defined a recursive algorithm to price the option, we implement the current industry standard as a benchmark to compare against. Next, we introduce the Gaussian mixture model (GMM) and explain how the popular classification method can be adapted to option pricing. Having compared the two on an underlying driven by geometric Brownian motion (GBM), we add stochasticity to the volatility and introduce the Heston model [5]. Using the characteristic equation as a benchmark, we compare the performance of the two estimation methods for a single continuation value and within a recursive pricing algorithm to price an American option. Lastly, we look at the so-called "best of" rainbow options and compare the resulting deltas implied by the GMM and the finite difference delta for the more conventional least squares Monte-Carlo approach [2].

Chapter 2

The Black-Scholes Model

Introduced in a 1973 paper, the Black-Scholes [4] model (BS) for pricing European options assumes that stock prices follow the log-normal random walk. Consider the stochastic process $S(t)$ on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$ with dynamics

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t) \tag{2.1}$$

where μ and σ are the drift and diffusion coefficients and $W(t)$ is a Wiener process. Black and Scholes further assume constant and known interest rate, the absence of arbitrage opportunities, dividends, and transaction costs, penalty-free short selling along with the ability to borrow any amount to buy or hold the security at the risk-free rate making up what they call "ideal conditions" in the market for the stock and the option. The European option price under these assumptions can now be derived by setting up a self-financing replicating portfolio, $V(S(t), t) = \Delta S(t) + B(t)$, where B is a bank account with dynamics $dB = rBdt$. The value of the option will only depend on the price of the stock, time, and known constants. The change in the portfolio value is therefore given by

$$\begin{aligned} dV =& \Delta dS + rBdt \\ =& (\Delta \mu S + rB)dt + \Delta \sigma S dW. \end{aligned}$$

The value of the option it replicates changes according to the following equation given by Ito's lemma

$$dV = \frac{\partial V}{\partial t}dt + \mu S \frac{\partial V}{\partial S}dt + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}dt + \sigma S \frac{\partial V}{\partial S}dW.$$

Setting $\Delta = \frac{\partial V}{\partial S}$ we can cancel out the Brownian motion terms and substitute $B = V - \Delta S$ to get

$$\begin{aligned} \frac{\partial V}{\partial t}dt + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}dt &= rBdt \\ &= r(V - \frac{\partial V}{\partial S}S)dt. \end{aligned}$$

Rearranging the terms results in the famous Black-Scholes partial differential equation. Notice that the drift term μ has disappeared, and the value of the option is, therefore, independent of μ.

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial S^2} - rV = 0.$$

Assuming the appropriate terminal condition, we can transform the PDE into the heat-transfer equation from physics, which has a known solution. The price of the European call and put options, $c(t, S)$ and $p(t, S)$, under the Black-Scholes model at time t is then given by

$$c(t, S) = S\mathcal{N}(d_1) - Ke^{-r(T-t)}\mathcal{N}(d_2)$$
$$p(t, S) = Ke^{-r(T-t)}\mathcal{N}(-d_2) - S\mathcal{N}(-d_1)$$

where

$$d_1 := \frac{\log(\frac{S}{K}) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}} \quad \text{and} \quad d_2 := d_1 - \sigma\sqrt{T-t},$$

T the time to maturity of the contract and $\mathcal{N}$ is the cumulative standard normal distribution.

2.1 Delta Hedging

The ability to hedge against market movements is a common theme in financial literature, with some models having more easily derived sensitivities than others. The most common is the delta hedge, which is hedging against changes in the price of the underlying and can be achieved by trading said underlying. In the case of the Black-Scholes model, the delta is easily obtained by taking the derivative of the option formula.

$$\delta_{BS,c} = \frac{\partial c(t,S)}{\partial S} = \mathcal{N}(d_1) \quad \& \quad \delta_{BS,p} = \frac{\partial p(t,S)}{\partial S} = -e^{-rT}\mathcal{N}(-d_1)$$

For other options, such as those without an analytical pricing formula, we must rely on numerical methods, as discussed later.

Chapter 3

Pricing American Options

3.1 The Backwards Algorithm

Calculating the value of an American option is considerably more challenging than calculating the value of a European option, even when faced with the idealized assumptions of the Black-Scholes framework. While the American call price may be equal to the European price in the absence of dividends, the same is not true of the put since an optimal exercise strategy might involve an early exercise. We must therefore turn to simulation and a backward calculating algorithm. We follow the method of Longstaff and Schwartz [2] discussed in the next chapter, further details can be found in Glassersmans book Monte-Carlo Methods in Financial Engineering [6].

Let $\{t_i\}_{i=1,\dots,N}$ denote a set of simulation steps, each of which is also an exercise date for American options, and $\{S_i\}_{i=1,\dots,d}$ the corresponding set of underlyings. To price our option, we must determine the optimal exercise time along each simulated path $\omega \in \Omega$. The no-arbitrage theory [7] implies that the value of the option is given by the expected discounted cash flows of these optimal exercise times. We begin our pricing algorithm at the second last step, t_{N-1}, where we must choose between exercising immediately and holding the option to revisit the exercise decision at the next exercise time. Working our way back to the initial time step t_0 yields the following recursive relationship

$$O(t_i, \dots, t_N | t_i) = \max(\mathbb{E}[D_{t_i,t_{i+1}} O(t_{i+1}, \dots, t_N | t_{i+1})], h(t_i))$$

where we let $h(t_i)$ represent the value of the exercise payoff at time t_i, $D_{t_i,t_{i+1}}$ the discount factor between times t_i and t_{i+1} and $O(t_{i+1}, \dots, t_N | t_i)$ is the value of an option with strike times $(t_{i+1}, \dots, t_N)$ conditioned on the current time step t_i with $O(t_N) = h(t_N)$. We define the optimal exercise strategy for each path as

$$\tau(\omega) := \min\left(t_i \middle| \mathbb{E}[D_{t_{i+1},t_i} O(t_{i+1}, \dots, t_N | t_{i+1})] < h(t_i)\right)$$

that is the first step where exercising is more valuable than the expected value of holding the option. From the above, it is evident that the value of the option is highly dependent on the conditional expectation of its future payoff. Unfortunately, computing this value is not straightforward since analytic expressions are rarely available. We might proceed with nested Monte-Carlo simulation like discussed in chapter 8.7 in [6], but this means re-running a simulation for each step and each path which quickly becomes very computationally expensive as is demonstrated in the figure 3.1, so we turn to estimation methods.

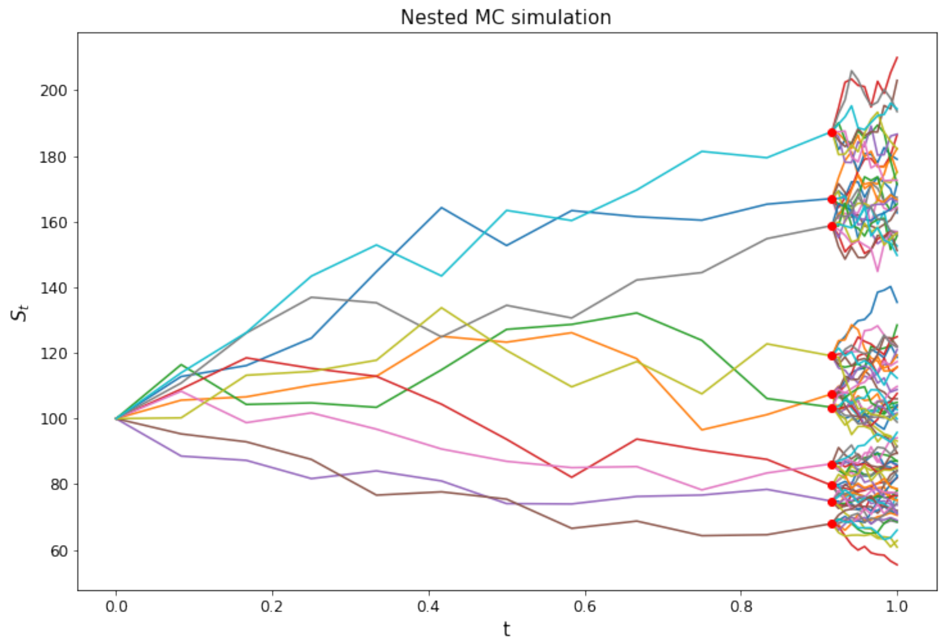

Figure 3.1: Nested Monte-Carlo simulation from the second last step of 10 GBM paths with 10 steps and parameters $S_0 = 100, \sigma = 0.25, r = 0.01, T = 1$.

3.2 The Longstaff-Schwartz Method

The current industry-standard method for valuing American options is the Longstaff-Schwartz [2] Least-Squares Monte-Carlo (LSM) approach, which turns the problem of estimating the conditional expectation into a simple regression problem. We assume that the conditional expectation of the continuation value is some linear combination of a countable set of $\mathcal{F}_{t_i}$ measurable basis functions, that is

$$\mathbb{E}\big[O|\mathcal{F}_{t_i}\big] = F(X, a) := \sum_{j=1}^{\infty} a_j L_j(X)$$

with constant coefficients a_j and some basis functions L_j like the Laguerre, Hermite, Legendre, Chebyshev, Gegnbauer, or Jacobi polynomials. To determine the coefficients $a_j \in A$, we minimize the L_2-distance between the estimate and the observed values, that is

$$\min \sum_{j=1}^{m} \big(O(\omega_j) - F(X(\omega_j), A)\big)^2.$$

We restrict our attention to in-the-money paths since we would only exercise when in the money. This removes unwanted paths from the regression, which otherwise may skew the estimate, and allows us to use fewer basis functions to achieve an accurate approximation. We only use our approximate to evaluate the exercise criterion as seen in equation (3.1) and thus only introduce the approximation error into the decision. This is a vast improvement over other methods, for example, Tsitsiklis and Van Roy [8], which take the approximate as the value of the option and therefore also the error associated with the approximation leading to a lesser estimate, especially for options with many exercise dates.

The following recursive LSM algorithm gives a stopping rule. Still, the value of an American option is based on the most valuable stopping rule of all possible stopping rules, including the one given by the LSM algorithm. We can therefore conclude that the LSM algorithm results in a value less than or equal to the value implied by the optimal stopping rule.

$$O(T_i) = \begin{cases} D_{t_i,t_{i+1}} O(T_{i+1}) & \text{if } h(t_i) < \mathbb{E}[O_{i+1}|\mathcal{F}_{t_i}] \\ h(t_i) & \text{else} \end{cases} \tag{3.1}$$

The LSM method is relatively simple and easy to implement but has drawbacks. As we have alluded to above, the true value of the optimal exercise strategy is the supremum of all possible strategies given by the LSM algorithm, which can be utilized to find the number of basis functions needed. As Longstaff and Schwartz discuss in their original paper, we can increase the number of basis functions J until the value implied by the algorithm stops rising. This, coupled with the findings of Glasserman et al. [9], which showed that the number of paths needed grows exponentially with the degree of the approximating polynomials, can make the method very computationally intensive.

Estimating tails is another shortcoming of the LSM method, as shown in figure 3.2 where we plotted the LSM estimate for the second last timestep of some GBM paths. As discussed previously, we can look at the decision in the second last time step as a choice between exercising and a European option with expiry at the next step. We have, therefore, also plotted the value of the associated Black-Scholes option for reference.

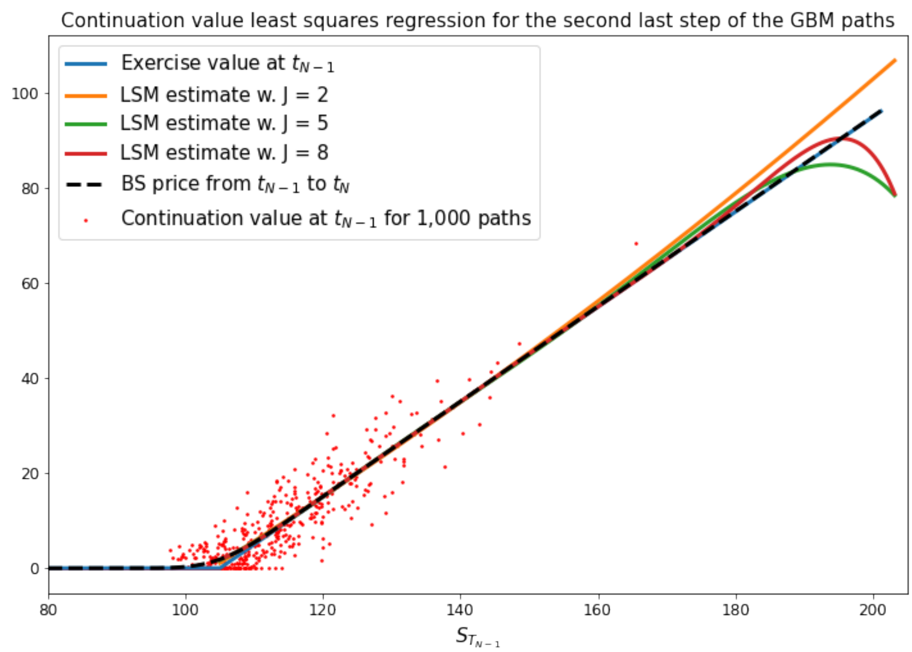

Figure 3.2: LSM regression for the second last step of 1.000.000 GBM paths with 12 steps parameters $S_0 = 100, \sigma = 0.15, r = 0.01, T = 1$ and K = 2,5 and 8 and the value of a corresponding Black-Scholes option. The x-axis is truncated since we only regress on in-the-money paths.

We also plot the errors compared to that same BS curve in figure 3.3. We can see how the LSM method provides a decent estimate when closer to the strike, but deep in the money, the estimate shoots off. As Longstaff and Schwartz point out, the estimate gets better with the increase in the degree of approximating polynomials, but this comes at the cost of the final ascent being steeper, effectively destroying any chance of a good tail approximation. The trick of only considering in the money paths ensures that such problems do not plague the out of the money tail.

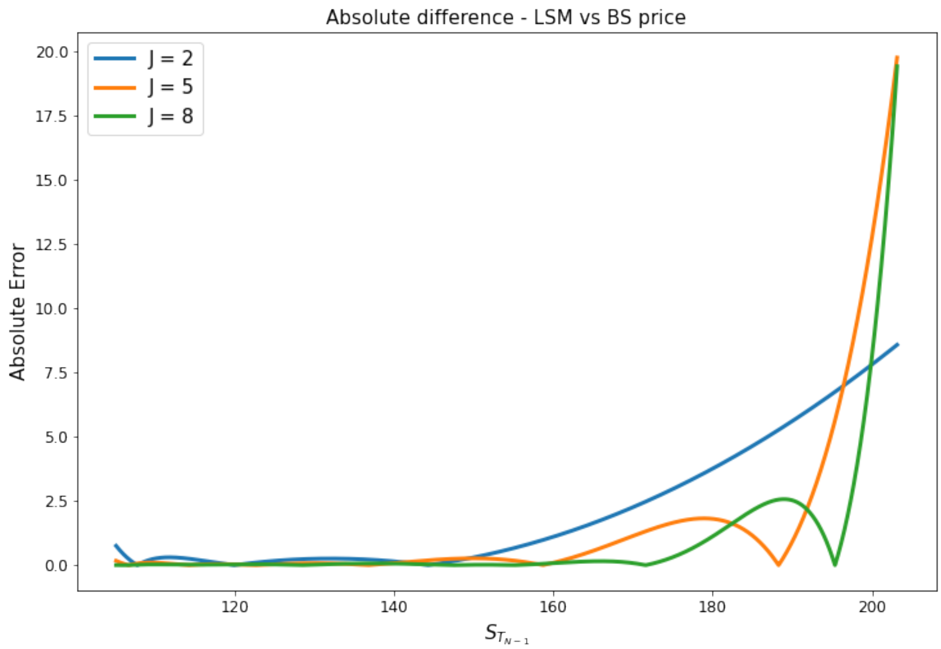

Figure 3.3: The absolute difference between the LSM estimates and the Black-Scholes option.

Finally, we point out the inherent computational effort it takes to compute the LSM option sensitivities when relying on finite difference methods, as is often the case with Monte-Carlo based methods. We introduce the notation $LSM(S)$ for the Least Squares Monte-Carlo/Longstaff-Schwartz price of an option on an underlying with spot price S and define the forward and backward difference estimators as

$$\Delta_F = \frac{LSM(S + \Delta S) - LSM(S)}{\Delta S}$$
$$\Delta_B = \frac{LSM(S - \Delta S) - LSM(S)}{\Delta S}$$

along with the central difference estimator

$$\Delta_C = \frac{LSM(S + \Delta S) - LSM(S - \Delta S)}{2\Delta S}.$$

Calculating these deltas requires re-running the pricing algorithm, further increasing the computational inefficiency of an already intensive algorithm.

Chapter 4

Gaussian Mixture Models

As a possible solution to some of the shortcomings discussed above, we consider Gaussian mixture models as discussed (GMM) by Kienitz in [3]. We begin with a short overview of Gaussian Mixture models, a method typically seen in classification problems, before considering its application to option valuation. Let $d \in \mathbb{N}$, $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$ be a positive-definite symmetric matrix. Then, the d-dimensional multivariate Gaussian distribution has the following joint probability density:

$$p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\Big\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big\}$$

where μ is the associated mean vector and Σ the covariance matrix. Let $\mathbf{z}$ be a J dimensional binary random variable for which the values of z_j satisfy $z_j \in \{0,1\}$ and $\sum_j z_j = 1$. The marginal distribution of $\mathbf{z}$ is given by

$$p(z_j = 1) = \pi_j$$

and the parameters π_j we call the mixing coefficients, which, to qualify as valid probabilities, must satisfy $0 \leq \pi_j \leq 1$ and $\sum_j \pi_j = 1$. The conditional distribution of x, given some value for $\mathbf{z}$, can be written in the following form

$$p(x|z_j = 1) = \mathcal{N}(x|\mu_j, \Sigma_j).$$

The marginal distribution of x is obtained by forming the joint distribution of x and $\mathbf{z}$, $p(z_j = 1)p(x|z_j = 1)$, and summing over the different states of the variable $\mathbf{z}$.

$$p(x) = \sum_{\mathbf{z}} p(z_j = 1)p(x|z_j = 1) = \sum_{j=1}^{J} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j).$$

The Gaussian mixture distribution for the random variable x can now be viewed as a weighted sum of J multivariate Gaussians.

4.1 Application to American Options

We are now ready to investigate the application of GMMs to approximate the conditional expectation needed for proper option valuation. We will begin by looking at American options and compare the GMM method to the current industry practice, the Longstaff-Schwartz, before venturing off to more exotic options.

Let $\mathcal{X} \in \mathbb{R}^{N\times d}$ be a matrix of realizations of stochastic risk factors, the prices $S(t) = (S_1(t), \ldots, S_d(t))$ of the d stocks in the case of our example. Let $\mathcal{Y} = (y_1, \ldots, y_N), y_n \in \mathbb{R}$ be a function of the risk factors, for example, the discounted payoff of an option, and $\mathcal{Z} = (z_1, \ldots, z_n), z \in \mathbb{R}$ some realizations of an associated random variable we intend to use as a control variate. We will start by using the underlying itself as the control variate since we have it already available, but this could easily be swapped out for some other proxy of choice, as might want to do, for example, for rainbow options. Now we form the training set X.

$$X = \begin{bmatrix} | & | & | \\ \mathcal{Z} & \mathcal{Y} & \mathcal{X} \\ | & | & | \end{bmatrix} \overset{\text{for example}}{=} \begin{bmatrix} | & | & | \\ S_{T_N} & O_{T_N} & S_{T_{N-1}} \\ | & | & | \end{bmatrix}$$

To determine the model parameters π_j, μ_j, and Σ_j, we turn to the maximum likelihood framework, which generally, when applied to GMMs, is known to cause significant singularity-related problems. We, therefore, consider the log-likelihood function given by

$$\ln p(\mathbf{X}|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{j=1}^{J} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j) \right\}.$$

To find the maximum likelihood solutions we take the derivative with respect to each of the unknown parameters π_j, μ_j, and Σ_j, using equations (58), (61) and (86) from from The Matrix Cookbook by Petersen and Pedersen [10], and set to zero. In the case of π_j we can use Lagrange multipliers to ensure that the mixing coefficients sum to one. The resulting equations are then used with the expectation-maximization algorithm [11]. Now that we have fitted the Gaussian mixture to the data and obtained the parameters of the mixture for the J components

$$\mu = \begin{bmatrix} \mu_{Z,1} & \mu_{Y,1} & \mu_{X,1} \\ \vdots & \vdots & \vdots \\ \mu_{Z,J} & \mu_{Y,J} & \mu_{X,J} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \begin{bmatrix} \Sigma_{ZZ,1} & \Sigma_{ZY,1} & \Sigma_{ZX,1} \\ \Sigma_{YZ,1} & \Sigma_{YY,1} & \Sigma_{YX,1} \\ \Sigma_{XZ,1} & \Sigma_{XY,1} & \Sigma_{XX,1} \end{bmatrix} & \ldots & \begin{bmatrix} \Sigma_{ZZ,J} & \Sigma_{ZY,J} & \Sigma_{ZX,J} \\ \Sigma_{YZ,J} & \Sigma_{YY,J} & \Sigma_{YX,J} \\ \Sigma_{XZ,J} & \Sigma_{XY,J} & \Sigma_{XX,J} \end{bmatrix} \end{bmatrix}$$

$$\pi = \begin{bmatrix} \pi_1, \ldots, \pi_J \end{bmatrix}$$

all that remains are analytical calculations. We construct $\mu_{Y|X,j}$, $\Sigma_{Y|X,j}$ and $\tilde{\pi}_j$ for the conditional distributions by plugging the mixture parameters into the following equations. For derivations, see the appendix.

$$\mu_{Y|X,j} = \mu_{Y,j} + \Sigma_{YX,j}\Sigma_{XX,j}^{-1}(X - \mu_{X,j})$$

$$\Sigma_{Y|X,j} = \Sigma_{YY,j} - \Sigma_{YX,j}\Sigma_{XX,j}^{-1}\Sigma_{XY,j}$$

$$\tilde{\pi}_j \sim \frac{\pi_j \mathcal{N}(\mu_{X,j}, \Sigma_{XX,j})}{\sum \pi_l \mathcal{N}(\mu_{X,l}, \Sigma_{XX,l})}$$

Using these results, we can now form the conditional density of the GMM.

$$p(Y|X) = \sum_{j=1}^{J} \tilde{\pi}_j \mathcal{N}(Y|X, \mu_{Y|X,j}, \Sigma_{Y|X,j})$$

In the context of mathematical finance and option valuation, we are typically most interested in the conditional expectation. Let $\mathcal{X}^* = (x_1^*, \ldots, x_M^*)$ be a test set, this could either be an entirely new set or $\mathcal{X}$ itself for an in-sample forecast like we use for our examples The output of our model $y^* = (y_1^*, \ldots, y_M^*), y_i^* \in \mathbb{R}$ is now the conditional expectation, $y_j^* \approx \mathbb{E}[Y|X = x_j^*]$ and is shown in figure 4.1 along with the absolute error compared to the Black-Scholes option in figure 4.2. We can see how the GMM provides a reasonable estimation when out of the money and until generously in the money. However, as we get deeper into the tails, the quality of the estimate compared to the analytically calculated BS option deteriorates. We introduce a control variate (CV) to combat this and stabilize the model. Our revised estimate is

$$Y^* := Y|X + \beta_X(Z|X - \mu_{Z|X})$$

where β can be seen as a parameter controlling the effect of the CV. Note that since $\mu_Z(x) := \mathbb{E}[Z|X = x]$, adding the control variate will not affect the expected value of the estimate since the expected value of the bracket added is 0. We have already fitted the distribution to Z and, therefore, modeled the covariance structure and can calculate the conditional mean $\mu_{Z|X}(x)$ analytically just like we calculated $\mu_{Y|X}$. Then, to minimize the variance of our estimate, we derive the β corresponding to that minimized variance. The derivation can be seen in the appendix.

$$\beta_{X=x} = -\frac{Cov[Y, Z|X = x]}{Var[Z|X = x]}$$

When Z is the underlying, as in our example, the β is nothing but the time discrete conditional minimal variance delta. It can be viewed as a trading strategy that minimizes the variance for the self-financing portfolio $\Pi = S(t) + \beta Y^*$ and, thus, a proxy hedge.

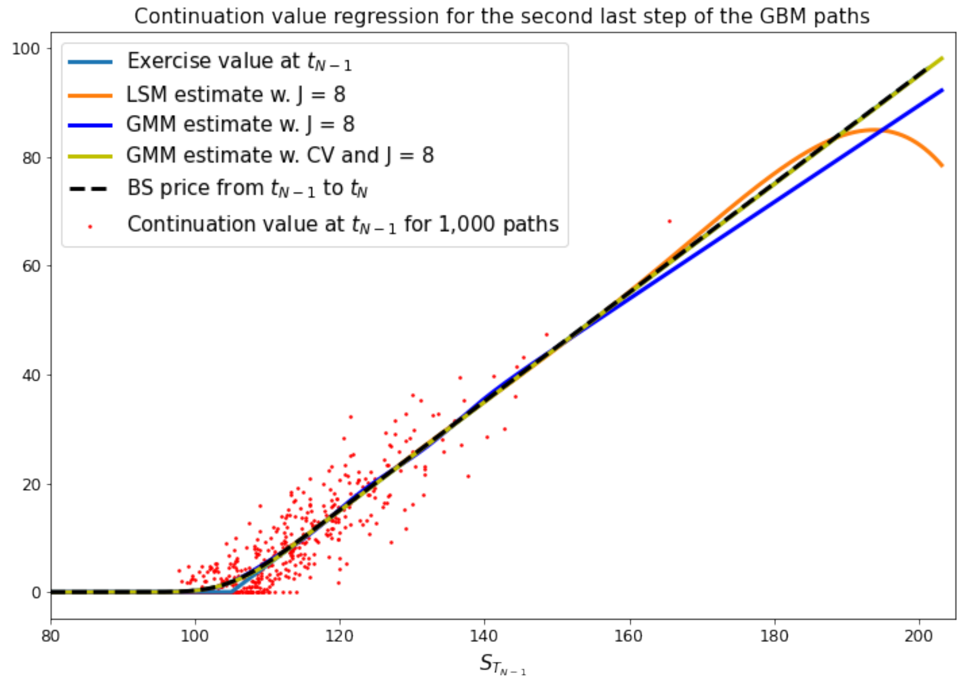

Figure 4.1: Updated figure 3.2 with the GMM estimate, with and without the control variate. Notice how the CV smooths out the estimate.

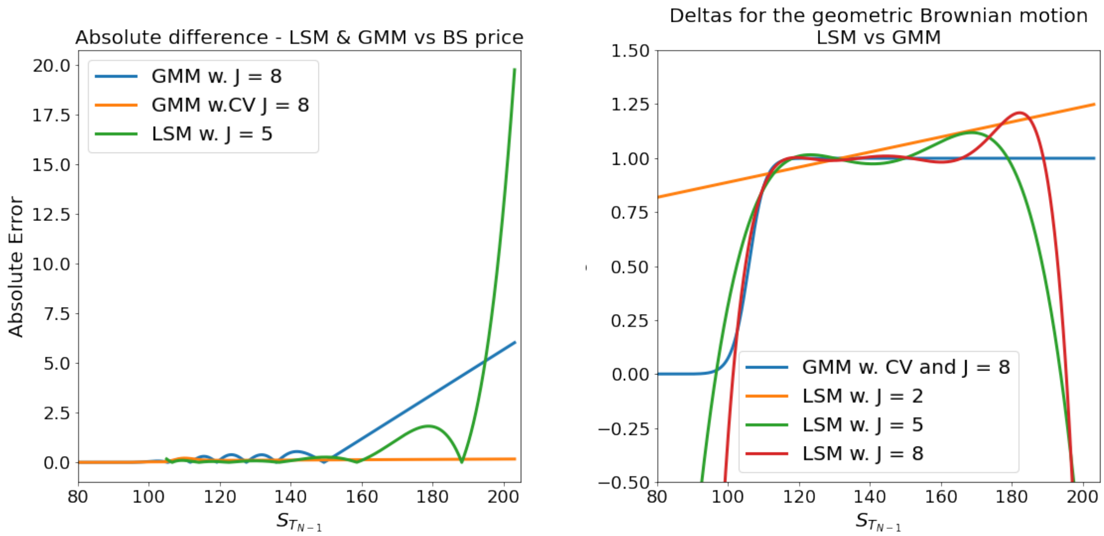

Figure 4.2: The absolute difference of the estimates and the Black-Scholes call option. The effect of the CV is clearly visible, especially deep in the money. On the right, we plot the minimal variance delta from the GMM method against the LSM deltas given by differentiating the polynomial approximations.

Now that we have a working model to approximate the conditional expectation, we can price options based on that model. We proceed almost like the LSM algorithm but replace the least squares approximation with the new Gaussian mixture based approach. To benchmark our method, we use the Quantlib library, and to further stress the GMM and the LSM methods, the benchmark Quantlib simulation runs 1.000.000 times while the GMM and LSM only use 10.000 paths. The American call and put results are plotted in a moneyness plot in figures 4.3 and 4.4 along with the absolute difference compared to the benchmark Quantlib simulation. Looking at the graphs, we can clearly see the difference in efficiency, the relatively low number of paths seems to affect the LSM algorithm more, but overall, both methods appear to converge to the benchmark price. Further simulations confirm that increasing the number of paths improves this convergence.

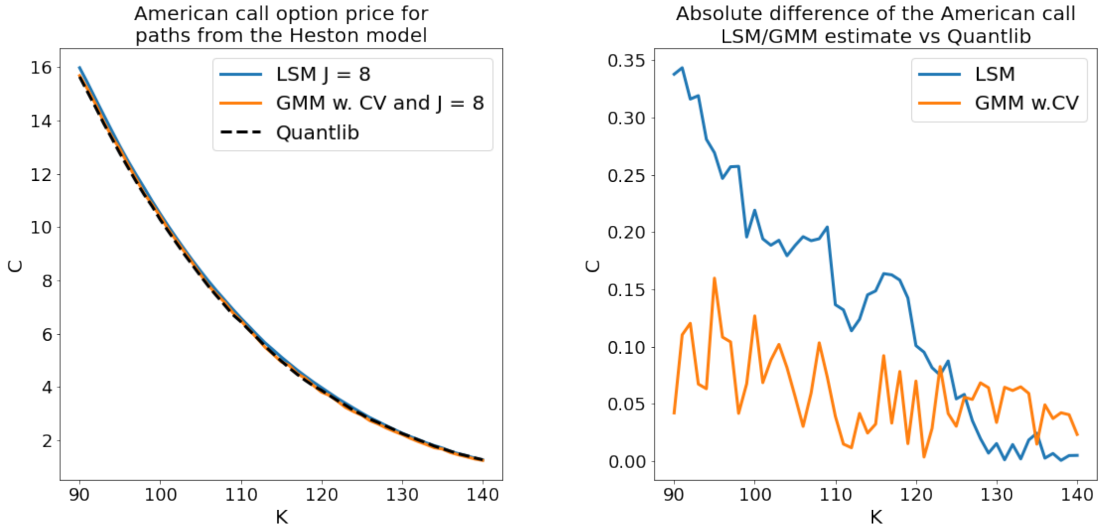

Figure 4.3: American call price as a function of the strike K and the absolute difference of the estimates and the benchmark call option calculated with the Quantlib library.

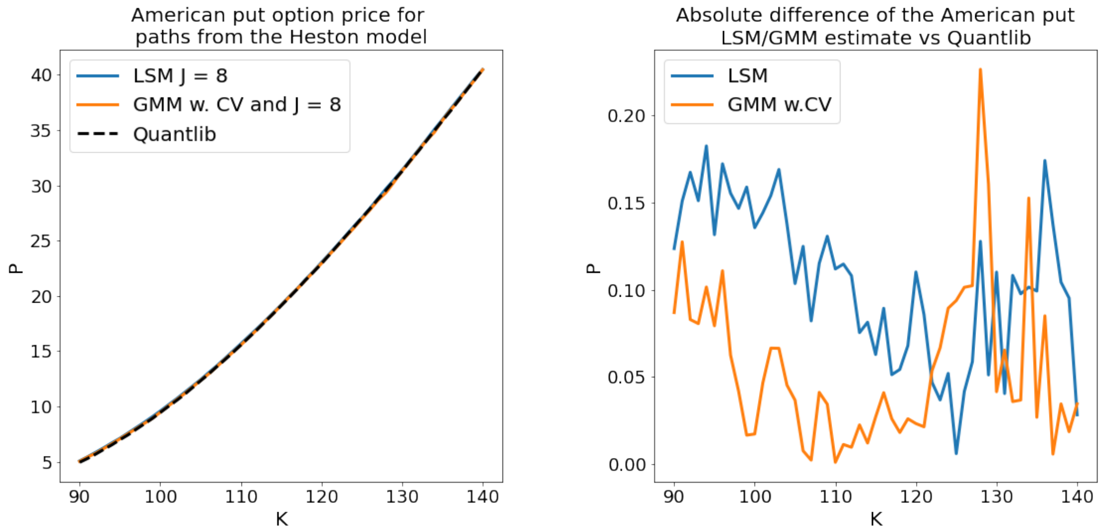

Figure 4.4: American put price as a function of the strike K and the absolute difference of the estimates and the benchmark put option calculated with the Quantlib library.

4.2 GMM on a Heston Model

Until now, we have only considered valuations utilizing the rather simplistic geometric Brownian motion. We will now consider a more challenging model, namely the Heston model [5]. Under the Heston model, we assume that the spot price of an asset S at time t follows the familiar stochastic process

$$dS(t) = \mu S(t)dt + \sqrt{v(t)}S(t)dW_1(t)$$

but now the volatility is changed from a constant to an Ornstein-Uhlenbeck process as used by Stein and Stein [12]. The resulting variance process can be written as the square-root process used by Cox-Ingersoll-Ross [13].

$$dv(t) = \kappa\big(\theta - v(t)\big) + \sigma\sqrt{v(t)}dW_2(t)$$

where θ is the long run variance, κ the rate of mean reversion, σ is the volatility of the variance and the Wiener process $W_1(t)$ has correlation ρ with $W_2(t)$. By the no-arbitrage arguments employed by Black and Scholes, the value of an asset $U(S, v, t)$ must satisfy the partial differential equation

$$\begin{aligned}\frac{1}{2}vS^2\frac{\partial^2 U}{\partial S^2} + \rho\sigma vS\frac{\partial^2 U}{\partial S\partial v} + \frac{1}{2}\sigma^2\frac{\partial^2 U}{\partial v^2} + rS\frac{\partial U}{\partial S}\\ +\{\kappa[\theta - v(t)] - \lambda(S, v, t)\}\frac{\partial U}{\partial v} - rU + \frac{\partial U}{\partial t} = 0\end{aligned} \tag{4.1}$$

where the unspecified term λ represents the market price of volatility risk. Seeking a solution in a form corresponding to the Black-Scholes model, Heston starts with a guess solution for a European call option with strike K and maturity T

$$C(S, v, t) = SP_1 - KP(t, T)P_2$$

where P_1 is the option's delta and P_2 is the conditional probability that the option expires in the money, both of which must satisfy the partial differential equation (4.1). Instead of solving the PDE directly, Heston builds a solution using the characteristic functions of the probabilities defined via the inverse Fourier transformation

$$P_j(x, v, t; \log(K)) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{e^{-i\phi\log(K)}f_j(x, v, t; \phi)}{i\phi}\right]d\phi$$

where he lets $\Re$ denote the real part of the complex function. The characteristic function solution is

$$f_j(x, v, t; \phi) = e^{C(T-t;\phi)+D(T-t;\phi)v+i\phi x}$$

where

$$
\begin{aligned}
C(\tau;\phi) &= r\phi i\tau + \frac{a}{\sigma^2}\left\{(b_j - \rho\sigma\theta i + d)\tau - 2\log\left[\frac{1-ge^{d\tau}}{1-g}\right]\right\} \\
D(\tau;\phi) &= \frac{b_j - \rho\sigma\phi i + d}{\sigma^2}\left[\frac{1-e^{d\tau}}{1-ge^{d\tau}}\right] \\
g &= \frac{b_j - \rho\sigma\phi i + d}{b_j - \rho\sigma\phi i - d} \\
d &= \sqrt{(\rho\sigma\phi i - b_j)^2\sigma^2(2u_j\phi i - \phi^2)}
\end{aligned}
$$

with $u_1 = 0.5 = -u_2$, $a = \kappa\theta$ and $b_1 = \kappa + \lambda - \rho\sigma = b_2 - \rho\sigma$.

The Heston characteristic function does, however, have two possible values for the complex root d as pointed out by Albrecher et al. [14], and the second root is equal to $-D$. This seemingly innocent-looking difference, often called Heston's little trap, has minor but important implications in practice. The original formulation crosses the negative real axis when increasing ϕ, which can lead to mispricing, especially at longer maturities. We will, therefore, use the second formulation.

We now test the GMM method on these more challenging paths from the Heston model, starting as before by estimating the continuation value and then the full American option price. First, we see how the LSM method fares and then introduce our GMM model. Then, since the method is entirely data-driven, we calculate all the necessary quantities exactly as before, swapping out the GBM paths for the new Heston paths.

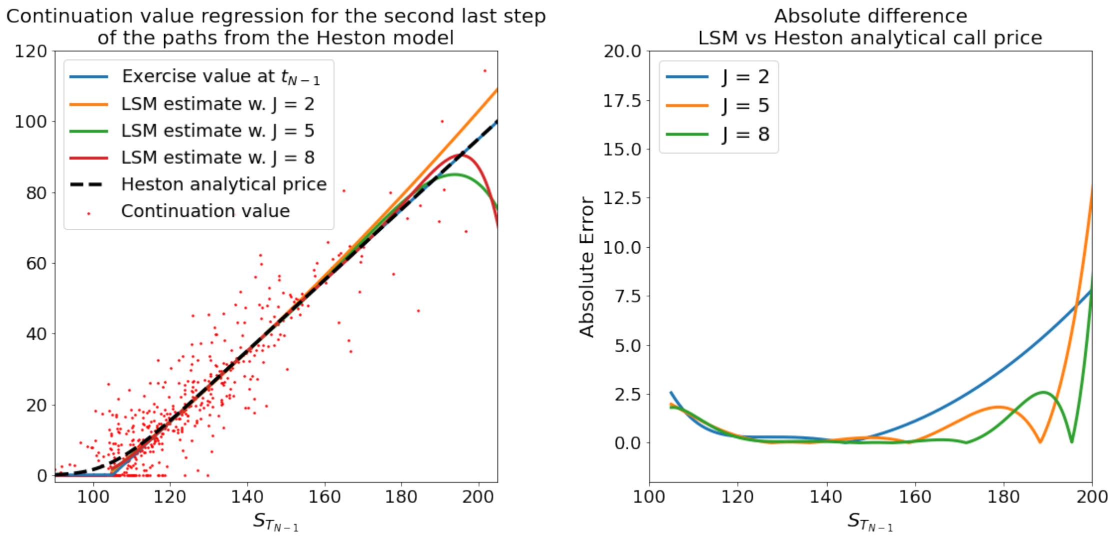

Figure 4.5: LSM regression for the second last time step of 1.000.000 12 step paths generated with the Heston model compared to the value of a corresponding analytically calculated option. The model parameters are $S_0 = 100, v_0 = 0.9, \theta = 0.1024, \rho = -0.5, \kappa = 0.2\sigma = 0.3, r = 0.01$ and $T = 1$. On the right is the absolute difference between the estimates and the analytical Heston call option.

Using the analytical call price for the Heston model as a reference point, we can see that the LSM estimate is less effective at the money than in the GBM case and experiences the same tail estimation problems. However, the same is not true for the GMM method, which provides an excellent approximation when paired with the control variate compared to the value calculated analytically via the characteristic equation.

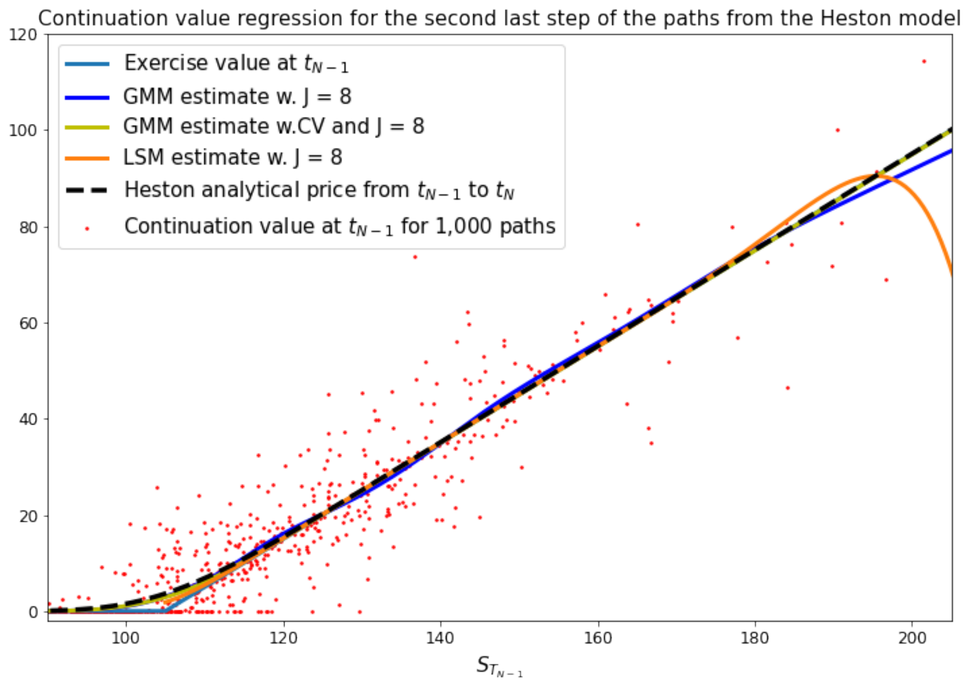

Figure 4.6: The GMM and LSM regression for the second last time step of the Heston from above with 2,5 and 8 basis functions compared to the value of a corresponding analytically calculated option.

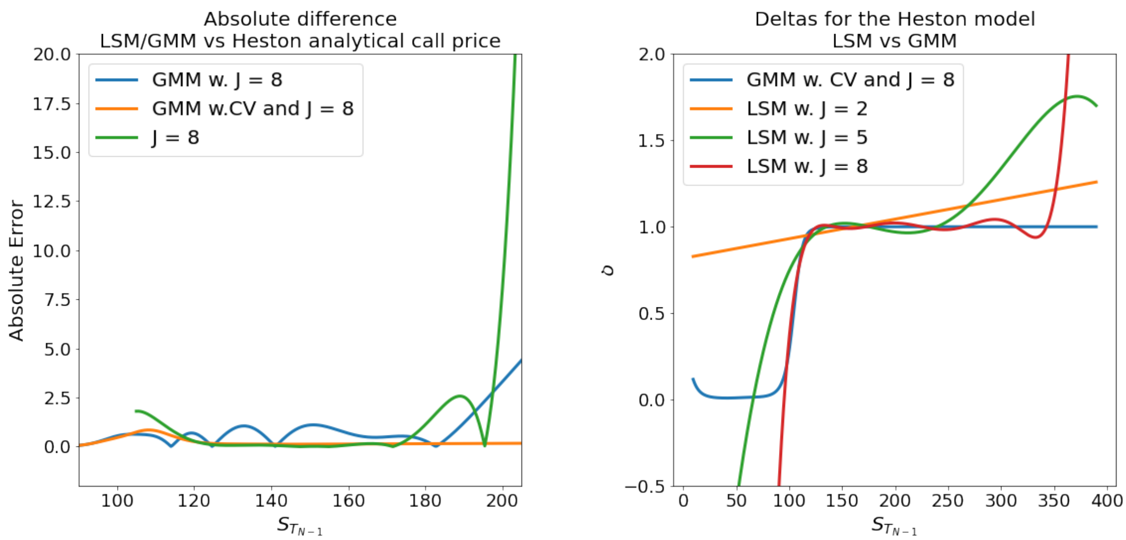

Figure 4.7: The absolute difference between the estimates and the Heston call option. On the right, we plot the minimal variance delta from the GMM method against the deltas resulting from the LSM method.

As with the GBM paths, we will use Heston paths to price an American option. To benchmark our American call, we will use the Heston characteristic equation. Although it only prices a European call option, in the absence of dividends, the European price should equal the American. Like with the GBM paths, we simulate 10.000 paths and feed them into our pricing algorithms.

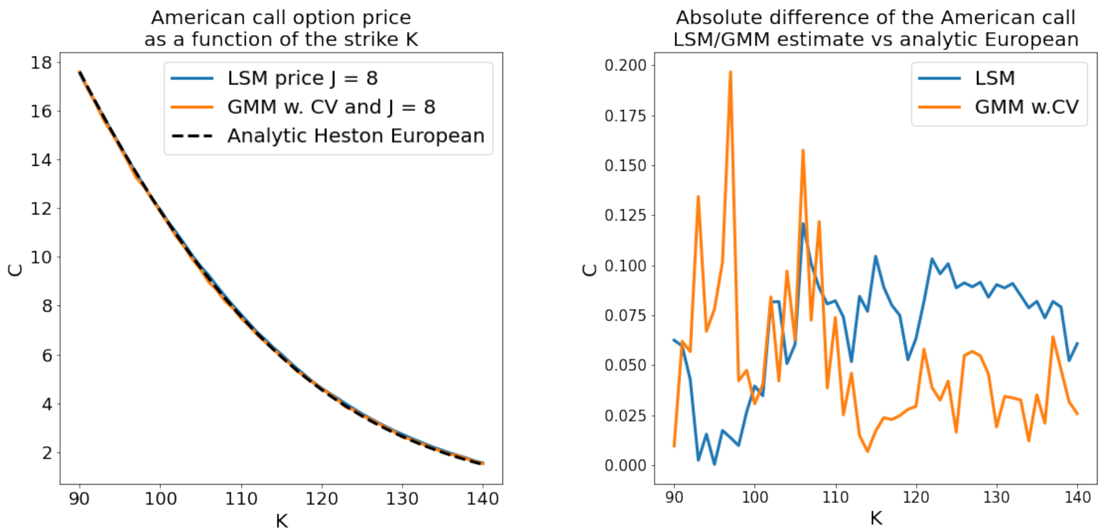

Figure 4.8: American call price as a function of the strike K and the absolute difference of the estimates and the benchmark call option calculated with the Quantlib library.

4.3 Application to Rainbow Option

Rainbow options are options whose performance is derived from multiple underlyings, typically through "best of" or "worst of" payoff configurations. Specific colors are sometimes assigned to each of the underlyings, so when put together, they form a rainbow, hence the name rainbow options. For demonstration purposes, we will consider a best-of option on 5 assets driven by a 5-dimensional Heston model and running 1.000.000 simulations. To correlate the assets, we consider the following asset-asset and asset-variance correlation matrices.

$$C_{aa} = \begin{bmatrix} 1.0 & 0.2 & 0.0 & 0.5 & 0.7 \\ 0.2 & 1.0 & 0.4 & 0.0 & 0.1 \\ 0.0 & 0.4 & 1.0 & 0.3 & 0.2 \\ 0.5 & 0.0 & 0.3 & 1.0 & 0.25 \\ 0.7 & 0.1 & 0.2 & 0.25 & 1.0 \end{bmatrix} \quad \& \quad C_{av} = \begin{bmatrix} -0.7 & 0 & 0 & 0 & 0 \\ 0 & -0.8 & 0 & 0 & 0 \\ 0 & 0 & -0.9 & 0 & 0 \\ 0 & 0 & 0 & -0.8 & 0 \\ 0 & 0 & 0 & 0 & -0.3 \end{bmatrix}$$

The 5 asset Heston model has 10 stochastic paths, 5 asset paths and 5 variance paths and can therefore be seen as a 10 dimensional model. However, our correlation matrices are only 5 dimensional, and we are missing the correlation structure between asset i and variance j when $i \neq j$ and the correlation between the different variances. To create the full correlation matrix, we use an algorithm called "complete the correlation" by Günther and Kahl [15]. The algorithm uses the known correlation between the Wiener processes $W_{S,i}$ and $W_{V,i}$, $dW_{S,i}dW_{V,i} = \eta_i dt$ for asset i, which in shorthand is written as $W_{S,i}W_{V,i} = \eta_i$, and the correlation between assets $dW_{S,i}dW_{S,j} = \rho_{i,j}$ to fill in the missing pieces marked ? in this incomplete correlation matrix

$$C = \begin{bmatrix} \rho_{1,1} & \cdots & \rho_{1,n} & \eta_1 & & ? \\ \vdots & \ddots & \vdots & & \ddots & \\ \rho_{n,1} & \cdots & \rho_{n,n} & ? & & \eta_n \\ \eta_1 & & ? & 1 & & ? \\ & \ddots & & & \ddots & \\ ? & & \eta_n & ? & & 1 \end{bmatrix}.$$

To bridge the gap in correlations, Kahl and Günther link the volatilities through their respective assets and their correlations to define the correlation between S_i and V_j as

$$c_{s_i,v_j} = W_{S,i} \cdot W_{V,j} = (W_{S,i} \cdot W_{S,j})(W_{S,j} \cdot W_{V,j}) = \rho_{i,j}\eta_j.$$

The cross correlation between two volatilities V_i and V_j can be defined similarly.

$$c_{v_i,v_j} = W_{V,i} \cdot W_{S,j} = (W_{V,i} \cdot W_{S,i})(W_{S,i} \cdot W_{S,j})(W_{S,j} \cdot W_{V,j}) = \eta_i \cdot \rho_{i,j}\eta_j$$

Our completed correlation matrix now becomes

$$C = \begin{bmatrix} 1.0 & 0.2 & 0.0 & 0.5 & 0.7 & -0.7 & -0.16 & 0.0 & -0.4 & -0.21 \\ 0.2 & 1.0 & 0.4 & 0.0 & 0.1 & -0.14 & -0.8 & -0.36 & 0.0 & -0.03 \\ 0.0 & 0.4 & 1.0 & 0.3 & 0.2 & 0.0 & -0.32 & -0.9 & -0.24 & -0.06 \\ 0.5 & 0.0 & 0.3 & 1.0 & 0.25 & -0.35 & 0.0 & -0.27 & -0.8 & -0.075 \\ 0.7 & 0.1 & 0.2 & 0.25 & 1.0 & -0.49 & -0.08 & -0.18 & -0.2 & -0.3 \\ -0.7 & -0.14 & 0.0 & -0.35 & -0.49 & 1.0 & 0.12 & 0.0 & 0.28 & 0.147 \\ -0.16 & -0.8 & -0.32 & 0.0 & -0.08 & 0.112 & 1.0 & 0.288 & 0.0 & 0.024 \\ 0.0 & -0.36 & -0.9 & -0.27 & -0.18 & 0.0 & 0.288 & 1.0 & 0.216 & 0.054 \\ -0.4 & 0.0 & -0.24 & -0.8 & -0.2 & 0.28 & 0.0 & 0.216 & 1.0 & 0.06 \\ -0.21 & -0.03 & -0.06 & -0.075 & -0.3 & 0.147 & 0.024 & 0.054 & 0.06 & 1.0 \end{bmatrix}.$$

To simulate the Heston model, we need a few more parameters.

$$\begin{aligned} S_i(0) &= 100 \quad \forall i, \quad r = 0.03, \quad T = 1, \quad K = 0.9 \\ v(0) &= \begin{bmatrix} 0.03 & 0.02 & 0.023 & 0.04 & 0.03 \end{bmatrix} \\ \kappa &= \begin{bmatrix} 0.1 & 0.13 & 0.2 & 0.21 & 0.15 \end{bmatrix} \\ \sigma &= \begin{bmatrix} 0.2 & 0.23 & 0.3 & 0.14 & 0.3 \end{bmatrix} \\ \theta &= \begin{bmatrix} 0.032 & 0.026 & 0.023 & 0.02 & 0.024 \end{bmatrix} \end{aligned}$$

The payoff function for our "best of" call option is

$$h(t) = \max\left[\max\left[\frac{S_1(t)}{S_1(t_0)}, \cdots, \frac{S_5(t)}{S_5(t_0)}\right] - K, 0\right].$$

Like before, our training set consists of three components, the stochastic risk factors X, some function of the risk factors Y, and the control variate Z. We will consider each underlying as the control variate by fitting the GMM model to a 100.000 paths subsample of the original 1.000.000 paths. Our components are

$$\begin{aligned} X &= \begin{bmatrix} S_1(t), \dots, S_5(t) \end{bmatrix} \\ Y &= h(T) \\ Z &= S_i(T) \end{aligned}$$

yielding the following training set

$$X = \begin{bmatrix} | & | & | & | & | & | & | \\ S_i(T) & O(t) & S_1(t) & S_2(t) & S_3(t) & S_4(t) & S_5(t) \\ | & | & | & | & | & | & | \end{bmatrix}.$$

After fitting the mixture model, the analytical calculations are carried out in almost the same way as before, adjusting for the increased dimensionality of the risk factors. The resulting minimal variance deltas are plotted in figure 4.9 along with their respective finite difference deltas. As expected, the GMM deltas appear smooth, unlike the finite-difference deltas, but the shapes of the curves vary more. This is likely due to the relatively low number of simulations as finite-difference methods are, like many Monte-Carlo-based methods, known to be simulation hungry.

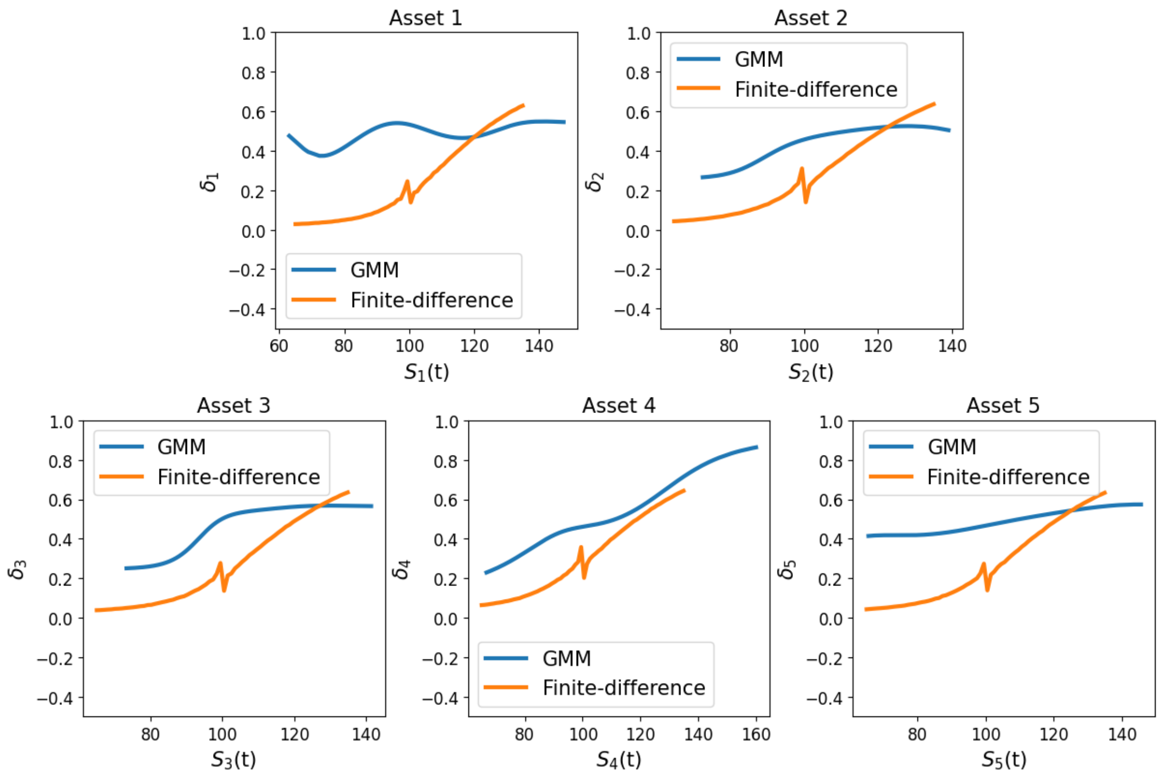

Figure 4.9: The approximated deltas, finite difference estimated for the LSM method and the minimal variance from the GMM method.

Chapter 5

Conclusions

Gaussian mixture models are a new and promising way to approximate conditional expectations for quantitative finance. The method is entirely data-driven and provides semi-analytical formulas for the conditional expectations. Not only does it appear to be more efficient in calculating the conditional expectation, requiring fewer paths than the current industry standard method, but it also provides the minimal variance delta via analytical calculations and, therefore, drastically reduces the computational effort needed to estimate this fundamental quantity of risk management.

In this article, we applied the method to evaluate conditional expectations using both geometric Brownian motion and paths from the Heston model as well as valuating American and rainbow options. Paired with the appropriate control variate, the method provides estimates on par with or even more accurate than the current industry standard and is a viable alternative to the least squares estimation of the Longstaff-Schwartz algorithm.

Bibliography

[1] C. R. Harvey, "The specification of conditional expectations", *Journal of Empirical Finance*, vol. 8, no. 5, pp. 573–637, 2001, On the Predictability of Asset Returns, ISSN: 0927-5398. DOI: https://doi.org/10.1016/S0927-5398(01)00036-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927539801000366.

[2] F. A. Longstaff and E. S. Schwartz, "Valuing American Options by Simulation: A Simple Least-Squares Approach", en, *Review of Financial Studies*, vol. 14, no. 1, pp. 113–147, Jan. 2001, ISSN: 0893-9454, 1465-7368. DOI: 10.1093/rfs/14.1.113. [Online]. Available: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/14.1.113 (visited on 03/28/2023).

[3] J. Kienitz, "Semi-analytic conditional expectation", *RISK*, vol. 7, 2022.

[4] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities", *The Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973. [Online]. Available: http://www.jstor.org/stable/1831029.

[5] S. L. Heston, "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options", en, *Review of Financial Studies*, vol. 6, no. 2, pp. 327–343, Apr. 1993, ISSN: 0893-9454, 1465-7368. DOI: 10.1093/rfs/6.2.327. [Online]. Available: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/6.2.327 (visited on 03/29/2023).

[6] P. Glasserman, *Monte Carlo methods in financial engineering* (Applications of mathematics 53), en. New York: Springer, 2004, ISBN: 978-0-387-00451-8.

[7] S. A. Ross, "The arbitrage theory of capital asset pricing", *Journal of Economic Theory*, vol. 13, no. 3, pp. 341–360, 1976, ISSN: 0022-0531. DOI: https://doi.org/10.1016/0022-0531(76)90046-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022053176900466.

[8] J. Tsitsiklis and B. Van Roy, "Regression methods for pricing complex American-style options", en, *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 694–703, Jul. 2001, ISSN: 10459227. DOI: 10.1109/72.935083. [Online]. Available: http://ieeexplore.ieee.org/document/935083/ (visited on 04/21/2023).

[9] P. Glasserman and B. Yu, "Number of paths versus number of basis functions in American option pricing", en, *The Annals of Applied Probability*, vol. 15, no. 2, May 2005, arXiv:math/0503556, ISSN: 1050-5164. DOI: 10.1214/105051604000000846. [Online]. Available: http://arxiv.org/abs/math/0503556 (visited on 04/21/2023).

[10] K. B. Petersen and M. S. Pedersen, "[http://matrixcookbook.com]", en,

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the *EM* Algorithm", en, *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, Sep. 1977, ISSN: 00359246. DOI: 10.1111/j.2517-6161.1977.tb01600.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1977.tb01600.x (visited on 04/22/2023).

[12] E. M. Stein and J. C. Stein, "Stock Price Distributions with Stochastic Volatility: An Analytic Approach", en, *Review of Financial Studies*, vol. 4, no. 4, pp. 727–752, Oct. 1991, ISSN: 0893-9454, 1465-7368. DOI: 10.1093/rfs/4.4.727. [Online]. Available: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/4.4.727 (visited on 05/18/2023).

[13] J. C. Cox, J. E. Ingersoll, and S. A. Ross, "A Theory of the Term Structure of Interest Rates", en, 2023.

[14] H. Albrecher, P. Mayer, W. Schoutens, and J. Tistaert, "The little heston trap", *Wilmott*, pp. 83–92, Mar. 2007.

[15] C. Kahl and M. Günther, "Complete the Correlation Matrix", en, in *From Nano to Space*, M. H. Breitner, G. Denk, and P. Rentrop, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 229–244, ISBN: 978-3-540-74237-1. DOI: 10.1007/978-3-540-74238-8_17. [Online]. Available: http://link.springer.com/10.1007/978-3-540-74238-8_17 (visited on 04/26/2023).

[16] C. M. Bishop, *Pattern recognition and machine learning* (Information science and statistics), en. New York: Springer, 2006, ISBN: 978-0-387-31073-2.

Appendix A

Derivation of the multivariate Gaussian components

This derivation is largely based on chapter 9.2 in Bishop's book, Pattern Recognition and Machine Learning [16]. Let $Z = (Z_1, \dots, Z_m, Z_{m+1}, \dots Z_{m+n=d})^\intercal$ be a d dimensional vector such that $Z \sim \mathcal{N}_d(Z|\mu_Z, \Sigma_Z)$ with the partition $Z = (Y, X)$ where we take Y to form the first m components of Z and X the remaining $n = d - m$ components. We also define the associated mean vector and covariance matrix partitions

$$\mu_Z = \begin{bmatrix} \mu_Y \\ \mu_X \end{bmatrix} \quad \& \quad \Sigma = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix}$$

as well as the inverse of the covariance matrix known as the precision matrix

$$\begin{bmatrix} \Lambda_{YY} & \Lambda_{YX} \\ \Lambda_{XY} & \Lambda_{XX} \end{bmatrix} = \Lambda := \Sigma^{-1}$$

with

$$\Lambda_{YY} = (\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY})^{-1} \tag{A.1}$$

$$\Lambda_{YX} = -(\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY})^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}. \tag{A.2}$$

We note that the apparent symmetry $\Sigma^\intercal = \Sigma$ implies that Σ_{YY} and Σ_{XX} are symmetric while $\Sigma_{XY} = \Sigma_{YX}$. The same is true for the precision matrix. Recall the quantity in the exponent of the multivariate Gaussian distribution, $\Delta^2 = \frac{1}{2}(Z - \mu_Z)^T\Sigma_Z^{-1}(Z - \mu_Z)$. Inserting for our partition, we get

$$\begin{aligned} \Delta^2 = &-\frac{1}{2}(Y - \mu_Y)^T\Lambda_{YY}(Y - \mu_Y) - \frac{1}{2}(Y - \mu_Y)^T\Lambda_{YX}(X - \mu_X) \\ &- \frac{1}{2}(X - \mu_X)^T\Lambda_{XY}(Y - \mu_Y) - \frac{1}{2}(X - \mu_X)^T\Lambda_{XX}(X - \mu_X). \end{aligned} \tag{A.3}$$

Fixing X and looking at this expression as a function of Y, it is clear that this is a quadratic form like in the typical Gaussian exponent, and so the conditional distribution $p(Y|X)$ is also Gaussian. Since Gaussian distributions are completely described by their mean and covariance, we will construct them from equation (A.3). In general, the exponent of a Gaussian distribution can be written as follows

$$\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = -\frac{1}{2}x^\top\Sigma^{-1}x + x^\top\Sigma^{-1}\mu + C$$

where all terms independent of X are kept constant and represented by C. When comparing both sides of equation (A), we can equate the second-order terms in X and the linear terms. Applying the same procedure to equation (A.3) we get, for the second order terms in Y

$$-\frac{1}{2}x^T\Sigma_{Y|X}^{-1}x = -\frac{1}{2}Y^T\Lambda_{YY}Y$$

from which we conclude that the covariance matrix of the conditional distribution $p(Y|X)$ is given by $\Sigma_{Y|X} = \Lambda_{YY}^{-1}$. Equating the linear terms in Y and using $\Lambda_{XY}^\top = \Lambda_{YX}$ we get

$$\begin{aligned}
x^\top\Sigma_{Y|X}^{-1}\mu_{Y|X} = &\frac{1}{2}\Big\{Y^\top\Lambda_{YY}\mu_Y + \mu_Y^\top\Lambda_{YY}Y - Y^\top\Lambda_{YX}X \\
&+ Y^\top\Lambda_{YX}\mu_X - X^\top\Lambda_{XY}Y + \mu_X^\top\Lambda_{XY}Y\Big\} \\
=&Y^\top\{\Lambda_{YY}\mu_Y - \Lambda_{YX}(X-\mu_X)\}.
\end{aligned}$$

Rearranging and using equations (A.1) and (A.2) to substitute Λ_{YY}^{-1} and Λ_{YX} we get

$$\begin{aligned}
\mu_{Y|X} &= \Sigma_{Y|X}\{\Lambda_{YY}\mu_Y - \Lambda_{YX}(X-\mu_X)\} \\
&= \Lambda_{YY}^{-1}\{\Lambda_{YY}\mu_Y - \Lambda_{YX}(X-\mu_X)\} \\
&= \mu_Y - \Lambda_{YY}^{-1}\Lambda_{YX}(X-\mu_X) \\
&= \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X-\mu_X).
\end{aligned}$$

From the definition of conditional density and the fact that $p(X) = \sum_l \pi_l\mathcal{N}(X|\mu_{X,l}, \Sigma_{XX,l})$ we have

$$\begin{aligned}
p(Y|X) &= \frac{p(Y,X)}{p(X)} \\
&= \sum_{j=1}^{J}\frac{\pi_j\mathcal{N}(Y,X|\mu_j,\Sigma_j)}{p(X)} \\
&= \sum_{j=1}^{J}\frac{\pi_j\mathcal{N}(X|\mu_{X,j},\Sigma_{XX,j})}{\sum_l \pi_l\mathcal{N}(X|\mu_{X,l},\Sigma_{XX,l})}\mathcal{N}(Y|X,\mu_{Y|X,j},\Sigma_{Y|X,j}) \\
&= \sum_{j=1}^{J}\tilde{\pi}_j\mathcal{N}(Y|X,\mu_{Y|X,j},\Sigma_{Y|X,j}).
\end{aligned}$$

where we have defined

$$\tilde{\pi}_j \sim \frac{\pi_j\mathcal{N}(X|\mu_{X,j},\Sigma_{XX,j})}{\sum_l \pi_l\mathcal{N}(X|\mu_{X,l},\Sigma_{XX,l})}.$$

Appendix B

Derivation of the minimal variance delta

Our model output is an expression of the form

$$Y^* = Y|X + \beta_{X=x}\left(Z|X - \mu_{Z|X}\right).$$

Taking the variance of this estimate results in

$$Var(Y^*) = Var(Y|X) + 2\beta_{X=x}Cov(Y|X, Z|X) + \beta^2 Var(Y|X).$$

Differentiating with respect to β and setting equal to 0 gives

$$0 = 2Cov(Y|X, Z|X) + 2\beta Var(Y|X).$$

Now, we solve for β and obtain an expression for the minimum variance β

$$\beta_{X=x} = -\frac{Cov[Y, Z|X = x]}{Var[Z|X = x]}.$$