Netspar Network for Studies on Pensions, Aging and Retirement

Antoon Pelsser

Jing Cao

Peter den Iseger

Pricing Hybrid Options by an Efficient Monte Carlo Approach

Discussion Paper 2007 - 019

December 20, 2007

Pricing Hybrid options by an efficient Monte Carlo approach

Jing Cao*, Peter den Iseger and Antoon Pelsser

December 20, 2007

*Address corresponding to Jing Cao: University of Amsterdam, Dept. of Quantitative Economics, Roetersstraat 11, 1018WB Amsterdam, The Netherlands, e-mail: J.Cao@uva.nl

Abstract

In this paper, inspired by the Least-squares method introduced by Longstaff and Schwartz (2001), we develop an more efficient Monte Carlo technique to price the Hybrid options. Hybrid option, which is by nature with exotic properties and form, is a combination of two kinds of derivatives, for example a combination of an interest rate derivative with an equity derivative. Therefore, it is a very useful instrument for pension funds to control their funding ratio within required levels. Hybrid payoffs strongly depend both on the evolution of the yield curve as well as the equity underlying. Unfortunately there are no closed-form solutions available for a general (non-zero) correlation between the interest rate and equity process. In the absence of closed-form formulas, we try to price the hybrid options by an modified Least-squares method in Monte Carlo simulation.

1 Introduction

The purpose of this paper is to develop and analyze pricing methodology for hybrid derivatives used by pension funds to hedge the positions in their own funding ratio. These hybrid derivatives are designed specially for pension funds which are willing to sacrifice some upward potentials on their funding ratio to obtain downside protections. Since these kind of financial instruments blend the properties of debt and equity and their payoff structure is just like a collar, moreover, their underlying state variables are often not only one or two, but usually multiple. For pension funds, it is critically important to know the exact value of the derivative at every time point[1] so that they are alert to the status of funding ratio. These factors make it quite difficult for us to pricing these hybrid derivatives accurately and efficiently. Sometimes a closed form formula for a price of such an option is impossible to modeled.

Longstaff and Schwartz in 2001 [1] proposed a simple but efficient least-squares approach to approximate the value of American options by simulation. They make use of least squares to estimate the conditional expected payoff to the option holder from continuation. This approach is readily applicable in path-dependent and multifactor situations where traditional finite difference techniques cannot be used. Inspired by their method, we try to develop an more efficient and numerically stable approach which could

[1]Time point

overcome the limits of Longstaff and Schwartz's method in choosing good "basis functions". We use the properties of Hermite polynomials which are orthogonal with respect to the weight function $e^{-\frac{z^2}{2}}$, and make the conditional expectation values we try to calculate expand on Hermite polynomials. Therefore, we can calculate the estimated coefficients of the conditional expectation functions. Moreover, we also make use of an efficient algorithm of evaluation of Kronecker products in multi-dimension developed by Den Iseger and Oldenkamp [6] in simulation.

We discuss the practical problems that the pension funds are facing, the necessity for pensions to use hybrid options in dealing with their problems and how we could pricing the hybrids in section 2. In section 3, we introduce a practically useful method in pricing options in Monte Carlo simulation–Least squares method. Moreover, we develop a modified Least squares method to improve the accuracy and efficiency. Section 4 presents our numerical results including a simple example: Geometric average option on two assets and a simplified hybrid option with two-factor Hull-White model and one equity Black-Scholes model. Section 5 concludes.

2 Practical problem

To pension funds, funding ratio, which is calculated as the ratio of assets (at market value) to liabilities[2] as reported by each of the funds is a key factor to show the solvability of the schemes, as well as the efficiency of their investment. It is very crucial to keep the funding ratio within certain interval. Because once the funding ratio is smaller than a minimum level[3], the pensions

[2]To make the retirement system more secure by further guaranteeing the solvability of pension schemes, in Netherland, the government requires the pension funds to value their liabilities on a marked-to-market basis from January 2007, making the value of liabilities much more volatile. This will make the matching of assets and liabilities more complicated. However, previously, only the assets of Dutch pension funds were subject to a market valuation. Liabilities were discounted at fixed 4% rate, which made the asset-liability modeling easier.

[3]The minimum level of funding ratio is obviously at least equal to 100% because the liabilities must be fully covered by the assets. For the security reason, some countries have higher "minimum level", e.g, in Netherland, the government requires the pension funds to be supervised by a new framework "Financieel Toetsingskader (FTK)" from January 2007. Under the FTK, pension funds that fall below the minimum coverage ratio of 105% of their indexed liabilities will have to correct their situation within one year.

will have deficits to cover their liabilities (which we call "underfunded") so that it would not be secure to guarantee the solvability any more. On the other hand, if the funding ratio is getting larger than a maximum level (which we call "overfunded"), it will probably require refunds to the sponsor or improvements to the benefits. Therefore, the ideal situation of the funding ratio of the pension funds should stay within the range of $[\alpha_{min}, \alpha_{max}]$, where α denotes the funding ratio level.

Just as we stated above, a stand-alone pension fund is, at time t_0 (the initial time), concerned about the possibility of its funding ratio at some future time t, which we denote as $FR(t)$, falling below a given minimum level α_{min}. It may wish to sacrifice some upward potential (which the pension fund does not like too much) to obtain downside protection by entering into a derivative contract with a third party (usually investment bank). Such a contract is called a collar or a hybrid option. As a result, the pension fund can just keep its funding ratio within the required interval.

Now let us state this problem in an analytical model. Let X denote the state space (state of the economy), such that the liabilities, the funding ratio are all a function of this state space $L_t := L(X_t)$, $FR_t = FR(X_t)$, respectively. Let further $A_t := A(X_t)$ stand for the assets of the pension fund, also (only) dependent on the state of the economy. We consider discrete time (discrete trading) and a horizon T. Let's consider a pension fund wishing to keep its nominal (unprotected) funding ratio $FR_N(t) = A_N(t)/L(t)$, according to the regulation, above a given level α_{min}. In return for achieving such an insurance the pension fund is ready to pay its excess above a level α_{max}[4]. This is in fact an derivative whose payoff is described by a function $D(t)$, defined by

$$D(t) = L(t)\{max[\alpha_{min} - FR_N(t), 0] - max[FR_N(t) - \alpha_{max}, 0]\} \qquad (2.1)$$

[4]The upper boundary of the funding ratio is usually chosen such that the initial value of this derivative contract is zero.

Then the real funding ratio FR(t) with a collar will be

$$\begin{aligned} FR(t) &= \frac{A(t)}{L(t)} \\ &= \frac{A_N(t) + D(t)}{L(t)} \\ &= FR_N(t) + \{max[\alpha_{min} - FR_N(t), 0] - max[FR_N(t) - \alpha_{max}, 0]\} \\ &= \begin{cases} \alpha_{min} & \text{if } FR_N(t) < \alpha_{min} \\ \alpha_{max}, & \text{if } FR_N(t) > \alpha_{max} \\ FR_N(t), & \text{if } otherwise. \end{cases} \end{aligned} \tag{2.2}$$

Obviously, the payoff structure of the derivative blends characteristics of debt and equity. Generally, these kind of financial instruments are called "hybrids". Since they are mixed by debt and equity, their market price tend to be influenced by both interest rates as well as the equity price. It turns out that hybrids rely on several risk factors[5], in other words, they are generally high-dimension instruments.

To make sure their safe position regarding to the funding ratio, pensions must have comprehensive information of the hybrid option and meanwhile choose certain efficient investment strategies to hedge their position at every time point. Therefore, the pricing of this derivative becomes an important problem for pension funds.

As we know, many hybrid options have no analytical (closed-form) solutions because they are too complex (high-dimension and the payoff structure is exotic). Consequently, numerical solutions are necessity. In fact, Monte Carlo simulation becomes an important method when dealing with higher dimensions. The reason is that with a grid-based numerical integration or PDE method, the number of grid points grows exponentially with dimensions. Monte Carlo on the other hand is a lot less affected by high dimensions, moreover, it can simulate arbitrarily complicated sample paths of state variables. However, a drawback of Monte Carlo is its high computational cost, especially in a high-dimensional setting. Therefore, we focus on finding an efficient Monte Carlo method to solve the pricing problem of hybrid options in a dynamic economy.

[5]In the pension case, there will be inflation risk, interest rate risk, equity risk, contribution risk of the sponsors and longevity risk of the members and so on.

3 Model and algorithm

There is a fundamental implication of asset pricing theory that under certain circumstances, the price of a derivative security can be represented as an expected value, especially the conditional expected value. Valuing derivatives thus reduces to computing (conditional) expectations. As we mentioned above, in many cases, if we were to write the relevant expectation as an integral, especially when the dimension is large, Monte Carlo method is an attractive approach. Therefore, pricing the hybrid, in principal, becomes computing conditional expectations in Monte Carlo simulation. In our model, we fix a probability space $(\Omega, \mathcal{F}_t, \mathcal{P})$ with a filtration $(\mathcal{F}_t)_{\{X_t\}_{t>0}}$. X_t is the state variable and we assume that $X_{t+1} = \Gamma(X_t, Y_{t+1})$, where Y_{t+1} is a random variable that is independent of X_t. For example, in the standard Black-Scholes model: $dS_t = rS_t dt + \sigma S_t dW_t$, we could have

$$log(S_{t+1}) = log(S_t) + (r - \frac{1}{2}\sigma^2)\Delta t + \sigma\sqrt{\Delta t}Z$$

Here we could understand $X_{t+1} = log(S_{t+1})$, $X_t = log(S_t)$ and $Y_{t+1} = (r - \frac{1}{2}\sigma^2)\Delta t + \sigma\sqrt{\Delta t}Z$. In fact, our assumption of economy is general, which could be extended to not only the Black-Scholes, but other models.

Let $V(t, X_t)$ denote the value of the hybrid option at time t, which could be considered as the expected value conditional on the filtration $\mathcal{F}_t$.

$$V(t, X_t) = \mathbb{E}[V(T, X_T)|\mathcal{F}_t] \tag{3.1}$$

It follows from the Markov tower property that

$$\begin{aligned} V(t, X_t) &= \mathbb{E}[V(T, X_T)|\mathcal{F}_t] \\ &= \mathbb{E}[\mathbb{E}(V(T, X_T)|\mathcal{F}_{t+1})|\mathcal{F}_t] \\ &= \mathbb{E}[V(t+1, X_{t+1})|\mathcal{F}_t] \\ &= \mathbb{E}[V(t+1, X_{t+1})|X_t = x] \end{aligned} \tag{3.2}$$

Clearly[6], the value $V(t, X_t)$ is the expected value of $V(t+1, X_{t+1})$ conditional on the $X_t = x$. Therefore, our problem now is how to estimate this conditional expectation value by Monte Carlo. To deal with this problem, we first introduce a least-squares algorithm proposed by Longstaff and Schwartz (2001), which at first intended to price American-style options.

[6]For simplicity, we assume the discount factor here is one.

3.1 Least-squares method

Longstaff and Schwartz (2001) [1] introduced a least-squares approach to value American options by simulation, which actually focuses on calculating the conditional expectation value of the option. They use regressions across simulated sample paths to evaluate the conditional expectation of the continuation value of the option and compare this expectation to the immediate exercise value at all future dates along each simulated path. Since there is no closed-form expression for the conditional expectation function, Longstaff and Schwartz suggest to select a set of basis functions such that their weighted combinations is "close" to the true function. Still, we generate random vectors X_t^k for $k = 1, \ldots, m$ and $t = 1, \ldots, T$, which are independently and identically distributed samples. $(p_i(X))$ denotes basis function and A_i represents the weights on the basis functions. They try to minimize the difference between the true value function and the weighted combinations:

$$min_{A_i} \| V(t, X_t) - \sum_{i=1}^{\infty} A_i p_i(X_t) \| \tag{3.3}$$

so that the weighted combinations can be used to approximate the true value function. With a large number of simulation paths, the Law of Large Numbers guarantees the weighted combinations to converge to the actual price of the option. Let $C(t, X_t) = \sum_{i=1}^{\infty} A_i p_i(X_t)$, simply speaking, there exists a relation:

$$\begin{aligned} V(t, X_t) &= \mathbb{E}[V(t+1, X_{t+1} | X_t = x)] \\ &\approx \sum_{i=1}^{\infty} A_i p_i(X_t) \\ &= C(t, X_t) \end{aligned} \tag{3.4}$$

Thus, in least-squares method, a linear combination of basis functions is fitted to the data via least squares regression in order to approximate the conditional expectation over the entire state space. Specifically, approximating the conditional expected function by a linear combination of known functions of the current state and using least squares regression to estimate the best coefficients in this approximation, which, then, provides a direct estimate of the conditional expectation function. Therefore, at time t, it is assumed that $V(t, X_t)$ can be expressed as a linear combination of orthonormal basis

functions ($p_i(X)$) such as Laguerre, Hermite, Legendre, Chebyshev or Jacobi polynomials. In practical implementation, $C(t, X_t)$ is approximated by

$$\hat{C}_I(t, X_t) = \sum_{i=1}^{I} A_i p_i(X_t), A_i \in \mathbb{R} \tag{3.5}$$

where I is the number of basis functions we used in estimation. Once this subset of basis functions has been specified, $\hat{C}_I(t, X_t)$ is estimated by projecting or regressing the discounted values of $V(t+1, X_{t+1})$, which is given by: $V(T, X_T) = D(T, X_T)$[7], onto the basis functions for the paths where the option is in-the-money[8] at time t.

Least-squares method has been proved to be a efficient and accurate approach to price derivatives in Monte Carlo. However, its accuracy depends on the choice of a "good" set of basis functions used in the regression. That's the main drawback of this method. If we did not choose good basis functions (regression variables), we would throw away useful information and find a biased price. Moreno & Navas (2003) [5] analyzed the robustness of the LSM approach relative to the type and number of basis functions. Their overall conclusion is that least-squares method is very robust when pricing the American put option. In this case, they found, for a given degree, using different polynomials as basis functions produces very similar results; but for complex options, the choice of basis functions is not clear, since combining polynomials with other functions to represent the information set at each exercise date is necessary. For these options, their conclusion is that the robustness seems not to be guaranteed and the type and numbers of basis functions can slightly affect option prices [5].

And the choice of basis functions [1] has implications for the statistical significance of individual basis functions in the regression. In particular, some choices of basis functions are highly correlated with each other, resulting in estimation difficulties for individual regression coefficients akin to the multicolinearity problem in econometrics. If the choice of basis functions leads to a cross-moment matrix that is nearly singular, then numerical inaccuracies in some least-squares algorithms may affect the functional form of estimated conditional expectation function.

[7]The value of the option at final time is equal to the payoff.

[8]The use of only in-the-money paths in the regressions is to increase efficiency and decrease computation time.

Our objective is to price hybrid options with very complicated features and usually multidimensional, it is usually difficult to find a "perfect" set of basis functions to estimate in the regression.

Moreover, speaking of the numerical stability of the least-squares method, it cannot be guaranteed in dealing with high-dimensional problems. When using pure matrix notation the least-squares problem can be written as

$$Y = bX + \varepsilon$$

Therefore, the least-squares estimator for b is

$$\hat{b} = (X^T X)^{-1} X^T Y$$

(where X^T is the transpose of X). and the sum of squares of residuals is

$$Y^T (I_n - X(X^T X)^{-1} X^T) Y$$

When it comes to high-dimension, which means X-vector becomes large, it is quite possible that the matrix $X^T X$ is not invertible. That would be the reason the results from least-squares estimation become instable. (Maybe write it in appendix)

Requirement: The quadratic matrix $X^T X$ is symmetric, which means it is a square matrix with corresponding elements above and below the diagonal. Consequently it equals its own transpose. In addition we will assume the design matrix X is such that the $X^T X$ matrix is positive definite and thus nonsingular.

Thus we have to find a way to overcome the limitations of least-squares method.

3.2 Modified least-squares method

As we discussed, Least-squares method estimates the coefficients of the conditional expectation function by regressing the subsequent realized cash flows from continuation on a set of basis functions of the values of the relevant state variables. The fitted value of this regression is an efficient unbiased estimate of the conditional expectation function. The basis functions are usually certain orthogonal polynomials. To overcome the limitations "how to choose good basis functions and how many of them do we need", we can make use of some special properties of Hermite polynomials to develop a more efficient way to estimate the coefficients of the conditional expectation function.

Hermite polynomials $H_j(Z) : \mathbb{R} \to \mathbb{R}$ are j_{th} polynomials for $j = 0, 1, 2, 3 \cdots$. These polynomials are orthogonal with respect to the weight function $e^{-\frac{Z^2}{2}}$, i.e., we have

$$\int_{-\infty}^{\infty} H_j(Z) H_k(Z) \frac{e^{-\frac{Z^2}{2}}}{\sqrt{2\pi}} dZ = \delta_{jk} = \begin{cases} 1, & k = j; \\ 0, & k \neq j. \end{cases} \tag{3.6}$$

Where δ_{jk} is the Kronecker delta. Obviously, we can say that Hermite polynomials are thus orthogonal with respect to the standard normal probability density function. They form an orthogonal basis of the Hilbert space of functions satisfying

$$\int_{-\infty}^{\infty} |f(x)|^2 e^{-\frac{x^2}{2}} dx < \infty$$

on which the inner product (g(x) is an arbitrary function in Hilbert sapce) is given

$$< f \cdot g >= \int_{-\infty}^{\infty} f(x) \overline{g(x)} e^{-\frac{x^2}{2}} dx$$

By the property of Hilbert space, a function of standard normal variable has the following relation:

$$W_t(Z_t) = \sum_{i=1}^{I} < W_t(Z_t) \cdot H_i(Z_t) > H_i(Z_t) \tag{3.7}$$

Clearly, if we want to know W_t, what we have to compute is the inner product $< W_t \cdot H_i >$, I is the number of Hermite polynomials. And the inner product could be written as

$$\begin{aligned} < W_t \cdot H_i > &= \int_{-\infty}^{\infty} W_t(Z_t) H_i(Z_t) e^{-\frac{Z_t^2}{2}} dZ_t \\ &= \mathbb{E}[W_t(Z_t) H_i(Z_t)] \end{aligned} \tag{3.8}$$

Moreover, in simulation, the expected value

$$\mathbb{E}[W_t(Z_t) H_i(Z_t)] \approx \frac{1}{M} W_t^{(k)}(Z_t) H_i(Z_t^{(k)})$$

where k denotes the k_{th} simulated path, $k = 1, 2, \cdots, M$.

Now let us return to the problem: estimating the conditional expectation functions. For equation (3.4), we choose the Hermite polynomials as basis functions in order to make use of the property (3.7), then

$$V(t, X_t) = \mathbb{E}[V(t+1, X_{t+1}|X_t = x)] \approx \sum_{i=1}^{I} A_i H_i(X_t) \tag{3.9}$$

But since our value function $V(t, X_t)$, in general, is not a function of standard normal variable, we cannot directly expand it on Hermite polynomials. Therefore, we have to connect our value function with a certain function of standard normal variable. Therefore, we choose a function ϕ so that

$$X_t = \phi_t(Z_t), (i.e.) Z_t = \phi_t^{-1}(X_t) \tag{3.10}$$

where Z is a standard normal variable. Let

$$W_t(Z_t) = V(t, X_t) = V(t, \phi_t(Z_t)) \tag{3.11}$$

and expand W_t

$$\begin{aligned} W_t &= \mathbb{E}[V(t+1, X_{t+1})|X_t = \phi_t(Z_t)] \\ &= \mathbb{E}[W_{t+1}(\phi_{t+1}^{-1}(X_{t+1}))|X_t = \phi_t(Z_t)] \end{aligned} \tag{3.12}$$

where $W_{t+1}(\phi_{t+1}^{-1}(X_{t+1})) = V(t+1, X_{t+1})$ when $t+1 = T$, by assumption, is already known. Combining the equation (3.7) and (3.9) we could write

$$\begin{aligned} V(t, X_t) &= W_t(Z_t) \\ &= \sum_{i=1}^{I} < W_t(Z_t) \cdot H_i(Z_t) > H_i(Z_t) \\ &= \sum_{i=1}^{I} A_i H_i(Z_t) \end{aligned} \tag{3.13}$$

Now we have clues to estimate W_t, however, to choose the function ϕ so that the value function could be written as a function of standard normal variable: $W_t(Z_t)$, we need to use some tricks in sampling in Monte Carlo. Assume we don't know exactly the distribution of state variable X_t, which is the general case in practice. Moreover, we have to always pay attention

to keeping the sample of variables on the domain of Hermite Approximation so that we can guarantee that the Hermite polynomials' properties can be applied appropriately. By choosing certain function ϕ to map standard normal variables Z_t to state variable X_t, and we first derive the cumulative density function (cdf)[9] from the known scenarios of X_t. Then according to the principle of inverse transform sampling[10], the cdf of X_t are uniform distributed which can be easily transformed into normal variables. Naturally, we can get the correlation covariances of the normal variables. This is the first phase that we get comprehensive information of the distribution of state variables X_t. Second phase, we first generate independent standard normal variables $\hat{Z}_t$, then transform them by multiplying the covariance matrix we got at first phase. Then by conducting the procedures of the first phase backwards, we could get the scenarios of $\hat{X}_t$ which has direct relation with independent standard normal variables and can be expanded on the Hermite polynomials. Simply speaking, our simulation procedure is

$$X_t \to F_{X_t}(x) = Pr\{X_t \leq x\} \to U_t = F_{X_t} \to Z_t = Trans(U_t) \to \Sigma = covariance(Z_t)$$

Then we need to connect the random standard normal variables with the state variables $\hat{X}_t$

$$\hat{Z}_t \to \varphi = \Sigma \hat{Z}_t \to \hat{U}_t = invTrans(\varphi) \to \hat{F}_{X_t}(x) = \hat{U}_t \to \hat{X}_t$$

The procedure above guarantees that the simulated scenarios of $\hat{X}_t$ has the same correlations with the original data and most importantly, it builds up the connection between the independent standard normal variables and the state variables $\hat{X}_t$. Therefore we can make expansions on the domain of Hermite Approximation.

Last but not the least, we could generalize Hermite polynomials to multi-dimensional cases. Define[11] $\hat{H}_J : \mathbb{R}^d \to \mathbb{R}$,

$$\hat{H}_J(Z_1, \cdots, Z_d) = \prod_{s=1}^{d} H_{j_s}(Z_s) \tag{3.14}$$

e.g.two-dimensional case $\hat{H}_{J=j_1,j_2}(Z_1, Z_2) = H_{j_1}(Z_1)H_{j_2}(Z_2)$. Equation (3.13)

[9] By kernel fit, see Appendix.

[10] See the Appendix.

[11] J is a d vector

which is a one-dimensional case can be represented as

$$\sum_{i=1}^{I} A_i H_i(Z) = \begin{bmatrix} H_1 & H_2 & \cdots & H_I \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \cdots \\ A_I \end{bmatrix} \tag{3.15}$$

We can also rewrite the two-dimensional case as matrix structure. To express the notation easily to understand, we say, for example, the number of the basis function for each dimension is 3.

$$\sum_{i=1}^{3}\sum_{j=1}^{3} A_{ij} H_i(Z_1) H_j(Z_2) = \begin{bmatrix} H_1(Z_1)H_1(Z2) & H_2(Z_1)H_1(Z2) & H_3(Z_1)H_1(Z2) \\ H_1(Z_1)H_2(Z2) & H_2(Z_1)H_2(Z2) & H_3(Z_1)H_2(Z2) \\ H_1(Z_1)H_3(Z2) & H_2(Z_1)H_3(Z2) & H_3(Z_1)H_3(Z2) \end{bmatrix} \begin{bmatrix} A_{11} \\ A_{21} \\ A_{31} \\ A_{12} \\ A_{22} \\ A_{32} \\ A_{13} \\ A_{23} \\ A_{33} \end{bmatrix} \tag{3.16}$$

It is no longer the usual matrix multiplication but can be recognized as the Kronecker product[12] and rewritten as

$$\sum_{i=1}^{3}\sum_{j=1}^{3} A_{ij} H_i(Z_1) H_j(Z_2) = (H_I \otimes H_J) vec(A) \tag{3.17}$$

Where $vec(A)$ denotes the vectorization of the matrix A formed by stacking the columns of A into a single column vector. Obviously, for high-dimensional cases, what we need to calculate is

$$(H_I \otimes H_J \otimes \cdots \otimes H_D) vec(A)$$

However, direct evaluation of these Kronecker products, by sheer size, would cause memory overflow. In practical implementation for multi-dimensional problems, it is very necessary to use some high-dimension transformation

[12]If I denotes the number of basis functions we chosen and d is the dimension of the state variables, clearly, we have to summate I^d terms. It is a quite a burden for calculation in practical implementation. We have to resort to efficient transformation techniques.

techniques to reduce the computation time and increase the convergence rate. Den Iseger and Oldenkamp [6] developed an efficient algorithm with Kronecker product, especially for multi-dimension[13]. We will use their algorithm to solve our problem.

3.3 Algorithm

we can sum up the algorithm as following:

The first phase (forward part):

1. From the known input variables (e.g. d-dimension) of $\{X_1^k, X_2^k, \cdots, X_d^k\}$, estimate the cumulative density functions $F_{X_1^k}(x), F_{X_2^k}(x), \cdots, F_{X_d^k}(x)$;
2. Compute the uniform distributed variables by $U_1^k = F_{X_1^k}(x), \cdots, U_d^k = F_{X_d^k}(x)$;
3. Transform the uniform distributed variables into normal distributed variables by $Z_1^k = N^{(-1)}(U_1^k), \cdots, Z_d^k = N^{(-1)}(U_1^d)$;
4. Compute the covariance matrix Σ of $\{Z_1^k, Z_2^k, \cdots, Z_d^k\}$

The second phase (backward part):

1. Simulating the sample paths $\{\hat{Z}_1^k, \cdots, \hat{Z}_d^k\}$, which are independently and identically standard normal distributed, $k = 1, \cdots, M$, M is the number of simulation paths;
2. Transforming the $\{\hat{Z}_1^k, \cdots, \hat{Z}_d^k\}$ into $\{\varphi_1^k, \cdots, \varphi_d^k\}$ by $\varphi = \Sigma\hat{Z}$;
3. Making the backward transformation: $\hat{U}_1^k = N(\varphi_1^k), \cdots, \hat{U}_d^k = N(\varphi_d^k)$ and $\hat{X}_1^k = F^{(-1)}(\hat{U}_1^k), \cdots, \hat{X}_d^k = F^{(-1)}(\hat{U}_d^k)$, therefore we get the simulated paths of the variables $\{\hat{X}_1^k, \hat{X}_2^k, \cdots, \hat{X}_d^k\}$ which can be expanded on the Hermite space;
4. Simulating (for each dimension) a fixed number N of $\hat{X}_{t+1}^{(k,j)}$ from $\hat{X}_t^k$ by the transformation function $\hat{X}_{t+1}^{(k,j)} = \Gamma(\hat{X}_t^{(k)}, Y_{t+1}^{(j)})$, $j = 1, \cdots, N$, therefore we can avoid exponential growing of the estimators[14];

[13] See appendix.

[14] By fixing the second time simulation numbers, we can make use of the sobol sequence (Quasi-Monte Carlo). For practitioner, it will save a lot computation time because of fast

5. Inverse transforming $\hat{Z}_{t+1}^{(k,j)}$ from $\hat{X}_{t+1}^{(k,j)}$ by the same techniques above;

6. $W_{t+1}^{(k,j)}(\hat{Z}_{t+1}^{(k,j)}) = V_{t+1}(\phi_{t+1}(\hat{Z}_{t+1}^{(k,j)})) = V_{t+1}(\Gamma(\hat{X}_t^k, Y_{t+1}^j))$;

7. Calculating $W_t^k = \mathbb{E}[W_{t+1}^{k,j}|\hat{Z}_t] = \frac{1}{N}\sum_{j=1}^{N} W_{t+1}^{(k,j)}(\hat{Z}_{t+1}^{(k,j)}) = \frac{1}{N}\sum_{j=1}^{N} V_{t+1}(\Gamma(\hat{X}_t^k, Y_{t+1}^j))$;

8. $A_i = \frac{1}{M}\sum_{k=1}^{M} W_t^k \hat{H}_i(\hat{Z}_t^k)$, $i = 1, \cdots, I$, I is the number of basis functions we chosen;

9. computing the value function $V(X_t) = \Sigma_{i=1}^{I} A_i \hat{H}_i(\hat{Z}_t^k)$

Therefore, we computed the inner product $< W_t \cdot H_i >$, and then estimate the value of W_t, which is the value of the option at time t: $V(t, X_t)$ by equation (3.13). Then repeat this procedure step by step, at time 0, we could get the price of the option.

Overall, the modified least-squares method is inspired by Longstaff and Schwartz's method and overcome its limitation in choosing good basis functions which is key point to get better estimation results of the valuation, especially for complicated, high-dimensional derivative securities. Moreover, the modified least-squares method is numerically more stable and

4 Numerical Examples

4.1 Geometric average option on two assets

Geometric average option is a type of derivative whose payoff is based on the difference between a the average value of the underlying assets during the life of the option and a fixed strike, i.e.,the payoff upon exercise of this option (five underlying assets) is $V(t, X_t) = ((X_t^1 \cdots X_t^2)^{(\frac{1}{2})}, K)^+$. We use this option because the pricing problem can be reduced to a single-asset American option, which can be priced accurately using a one-dimensional binomial tree. Therefore, we can compare our results from modified least-squares method with the "true" price.

convergence rate. The traditional Monte Carlo (MC) using (pseudo) random numbers, has a convergence rate of only $O(N^{-\frac{1}{2}})$ (think with Central Limit Theorem arguments or see Niederreiter, 1992, Theorem 1.1, p.4). Note that this rate is independent of the dimension d, it depends only of the number of simulations N; Quasi-Monte Carlo (QMC) rate of convergence can be much faster with errors approaching size of $O(N^{-1})$ in optimal cases (see e.g. Morokoff, 2000, pp.765-766, and Birge, 1995, p.2).

4.2 simple Hybrid: two factor Hull-White Model and Black-Scholes Equity model

In the two-factor Hull-White model[15] the instantaneous short rate r is evolved by

$$dr(t) = [\theta(t) + \mu(t) - \bar{a}r(t)] + \sigma_1 dZ_1(t), r(0) = r_0$$
$$d\mu(t) = -\bar{b}\mu(t)dt + \sigma_2 dZ_2(t), \mu(0) = 0$$

where μ can be seen as the (stochastic) mean reversion level satisfying the differential equation and with (Z_1, Z_2) a two-dimensional (correlated) Brownian motion. We use two-factor Hull-White model because it is a rich and flexible model, and is capable of fitting the main interest rate market.

For simplicity, we chose Black-Scholes as equity model.

5 Conclusion

We present a method for pricing hybrid options with multiple underlying variables by simulation. This approach which can be called a modified Least-squares method is intuitive, accurate, easy to apply, and computationally efficient. We illustrate this technique using two examples, including the valuation of a geometric average option on two assets as well as a realistic example: a simple Hybrid with two factor Hull-White Model and Black-Scholes Equity model.

It is meaningful to know the exact price of the hybrid options, especially, for pension funds. On the one hand, they can be aware of how much is exactly their true funding ratio and whether it is within a required interval. On the other hand, once it is possible to pricing the hybrid options, pension funds can try to replicate the options by choosing simple financial instruments so that they don't have to spend quite an amount of money to buy the hybrid derivatives from investment banks.

[15] We have chosen the Hull-White two-factor model, because it has a number of desirable properties among other short rate models: first the model is analytically tractable, which means there are closed-form formulas for bond prices, as well as for the main interest rate derivatives like caps or swaptions. Because of this the model can be calibrated fast and accurately, which are both must-haves for the financial industry. Secondly the model is rich enough to fit the revelent interest rate derivatives prices of the main interest rate markets. Thirdly the model has significant mass around zero, just as the risk-neutral density implied by the market volatility skew/smile indicates.

6 Appendix

6.1 Proof of equation (3.3)

6.2 The inverse function

Let g be the inverse function of f i.e. $g = f^{-1}$. Let
$\{\Phi_k : k \in \mathbf{N}_0\}$ be the Legendre polynomials. The Legendre polynomials $\{\Phi_k : k \in \mathbf{N}_0\}$ are a complete orthogonal polynomial set in $\mathcal{L}^2([0,1])$. These polynomials are given by

$$\Phi_k(t) = \frac{\sqrt{2k+1}}{k!}\partial^k(t^k(1-t)^k) \tag{6.1}$$

Since the shifted Legendre polynomials $\{\Phi_k(.-j) : k \in \mathbf{N}_0, j \in \mathbf{Z}\}$ are a complete orthogonal set in $\mathcal{L}^2(-\infty,\infty)$, any function $f \in \mathcal{L}^2(-\infty,\infty)$ can be expanded into

$$f = \sum_{k=0}^{\infty}\sum_{j=-\infty}^{\infty} < f, \Phi_k(.-j) > \phi_k(.-j) \tag{6.2}$$

Moreover, since the support of these polynomials is $[0,1]$, we obtain that

$$f(t) = \sum_{k=0}^{\infty} < f, \phi_k(.-j) > \phi_k(t-j) \quad t \in [j, j+1] \tag{6.3}$$

If we let $\mathcal{D}$ be the derivative operator we have using the Legendre expansion

$$\mathcal{D}g(t) = \sum_{k=0}^{n} < \mathcal{D}g, \Phi_k > \Phi_k(t), \quad t \in [0,1] \tag{6.4}$$

with

$$\begin{aligned} < \mathcal{D}g, \Phi_k > &= \int_0^1 \mathcal{D}g(t)\Phi_k(t)\,dt \\ &= \int_0^1 \Phi_k(t)\,dg(t) \end{aligned} \tag{6.5}$$

Then we change of variable and we define $y = g(t)$ so we obtain that $t = g^{-1}(y) = f(y)$ and

$$
\begin{aligned}
< \mathcal{D}g, \Phi_k > &= \int_{g(0)}^{g(1)} \Phi_k(f(y))\, dy \\
&\approx \sum_{k=1}^{n} w_k \Phi_k(f(y_k))
\end{aligned}
\tag{6.6}
$$

Here we had approximated the integral by a finite summation using the Legendre Gaussian quadrature. The two numbers w_k and y_k are given. There is a problem which is that as we don't know the function g, we can't compute the value $g(0)$ and $g(1)$. But we can solve this problem as follows.

Suppose we want to know the function g over the interval $[x_0, y_0]$. First we search the two numbers a and b such that $f(a) < x_0 < y_0 < f(b)$. Then we define the function $\mathcal{D}g_\Delta(t) = \mathcal{D}g(f(a) + \Delta t)$ with $\Delta = f(b) - f(a)$ and we obtain by formula (6.4)

$$
\mathcal{D}g_\Delta(t) = \sum_{k=0}^{n} < \mathcal{D}g_\Delta, \Phi_k > \Phi_k(t), \quad t \in [f(a), f(b)] \tag{6.7}
$$

with

$$
\begin{aligned}
< \mathcal{D}g_\Delta, \Phi_k > &= \int_0^1 \mathcal{D}g_\Delta(t)\Phi_k(t)\, dt \\
&= \int_0^1 \mathcal{D}g(f(a) + \Delta t)\Phi_k(t)\, dt
\end{aligned}
$$

Changing of variable we obtain

$$
\begin{aligned}
< \mathcal{D}g_\Delta, \Phi_k > &= \int_{f(a)}^{f(b)} \mathcal{D}g(t)\Phi_k\left(\frac{t - f(a)}{\Delta}\right) dt \\
&= \int_{f(a)}^{f(b)} \Phi_k\left(\frac{t - f(a)}{\Delta}\right) dg(t)
\end{aligned}
\tag{6.8}
$$

Then if we let $y = g(t)$ we have $t = g^{-1}(y) = f(y)$ and as above in the equality (6.6) we obtain

$$< \mathcal{D}g_\Delta, \Phi_k >= \int_{g(f(a))}^{g(f(b))} \Phi_k\left(\frac{f(y)-f(a)}{\Delta}\right) dy \tag{6.9}$$

However, $g(f(a)) = a$ and $g(f(b)) = b$ by definition of g then we obtain

$$\begin{aligned} < \mathcal{D}g_\Delta, \Phi_k > &= \int_a^b \Phi_k\left(\frac{f(y)-f(a)}{\Delta}\right) dy \\ &= (b-a)\int_0^1 \Phi_k\left(\frac{f(a+(b-a)x)-f(a)}{\Delta}\right) dx \qquad (6.10) \\ &\approx \sum_{k=1}^{n} w_k \Phi_k\left(\frac{f(a+(b-a)x_k)-f(a)}{\Delta}\right) \end{aligned}$$

where w_k and x_k are given numbers.

Hence we can compute the derivative of the function g over the interval $[f(a), f(b)]$ with the Legendre expansion given by the formula (6.7). But our aim is to compute the value of the function g then we have to integrate the function $\mathcal{D}g$. We define the operator $\mathcal{I}$ as follows

$$\mathcal{I}f(t) = \int_0^t f(t)\, dt \tag{6.11}$$

Before integrating the function $\mathcal{D}g$ in order to obtain the function g, we can remark that the matrix $\mathbf{I}$ defined by

$$\mathbf{I}_{jk} = < \mathcal{I}\Phi_k, \Phi_j > \tag{6.12}$$

is 3-diagonal.

First, we can easily see that

$$\mathcal{I}\Phi_k \in \Pi_{k+1} \tag{6.13}$$

where Π_{k+1} is the set of polynomial functions which have degree lower or equal to $k+1$.

Second, we remark that the operator $\mathcal{I}$ has the following property

$$< \mathcal{I}\Phi_k, \Phi_j >= - < \Phi_k, \mathcal{I}\Phi_j > \quad \forall k, j \quad j \neq 0, k \neq 0 \tag{6.14}$$

This equality can be proved as follows

$$\begin{aligned}
< \mathcal{I}\Phi_k, \Phi_j > &= \int_0^1 \int_0^x \Phi_k(t)\, dt\, \Phi_j(x)\, dx \\
&= \int_0^1 \int_t^1 \Phi_j(x)\, dx\, \Phi_k(t)\, dt \\
&= \int_0^1 \int_0^1 \Phi_k(t)\, dt\, \Phi_j(x)\, dx - \int_0^1 \int_0^t \Phi_j(x)\, dx\, \Phi_k(t)\, dt \\
&= \left(\int_0^1 \Phi_k(t)\, dt \right) \left(\int_0^1 \Phi_j(x)\, dx \right) - < \Phi_k, \mathcal{I}\Phi_j > \\
&= < \Phi_j, 1 >< \Phi_k, 1 > - < \Phi_k, \mathcal{I}\Phi_j >
\end{aligned}$$

As $\{\Phi_k : k \in \mathbf{N}_0\}$ is a complete orthogonal set we have

$$< \Phi_j, 1 >= 0, \quad j > 0 \tag{6.15}$$

Then it follows that

$$< \Phi_j, 1 >< \Phi_k, 1 >= 0, \quad j > 0, \quad k > 0 \tag{6.16}$$

Finally

$$< \mathcal{I}\Phi_k, \Phi_j >= - < \Phi_k, \mathcal{I}\Phi_j >, \quad j > 0, \quad k > 0 \tag{6.17}$$

We can see that

$$< \mathcal{I}\Phi_k, \Phi_j >= 0 \quad j > k + 1 \tag{6.18}$$

because as $\mathcal{I}\Phi_k \in \Pi_{k+1}$ we can write

$$\mathcal{I}\Phi_k = \sum_{n=0}^{k+1} < \mathcal{I}\Phi_k, \Phi_n > \Phi_n \tag{6.19}$$

Then we have

$$< \mathcal{I}\Phi_k, \Phi_j >= \sum_{n=0}^{k+1} < \mathcal{I}\Phi_k, \Phi_n >< \Phi_n, \Phi_j > \tag{6.20}$$

As $0 \leq n \leq k+1$ we have

$$< \mathcal{I}\Phi_k, \Phi_j >= 0 \quad j > k+1 \tag{6.21}$$

In the other hand we can also see that

$$< \mathcal{I}\Phi_k, \Phi_j >= 0 \quad j < k-1 \tag{6.22}$$

because

$$\begin{aligned} < \mathcal{I}\Phi_k, \Phi_j > &= - < \Phi_k, \mathcal{I}\Phi_j > \\ &= - < \mathcal{I}\Phi_j, \Phi_k > \\ &= 0, \quad k > j+1 \text{ i.e. } j < k-1 \end{aligned}$$

From (6.18) and (6.22) we can conclude that the matrix $\mathbf{I}$ is 3-diagonal.

Finally, we have

$$\mathcal{D}g = \sum_{k=0}^{\infty} < \mathcal{D}g, \Phi_k > \Phi_k \tag{6.23}$$

Then by the integral linearity we have

$$\mathcal{I}\mathcal{D}g = \sum_{k=0}^{\infty} < \mathcal{D}g, \Phi_k > \mathcal{I}\Phi_k \tag{6.24}$$

and we can conclude that

$$
\begin{aligned}
g \;=\; \mathcal{I}\mathcal{D}g \;&=\; \sum_{k=0}^{\infty} < \mathcal{D}g, \Phi_k > \mathcal{I}\Phi_k \\
&=\; \sum_{k=0}^{\infty} < \mathcal{D}g, \Phi_k > \sum_{j=0}^{\infty} < \mathcal{I}\Phi_k, \Phi_j > \Phi_j \qquad (6.25) \\
&=\; \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} < \mathcal{D}g, \Phi_k >< \mathcal{I}\Phi_k, \Phi_j > \Phi_j
\end{aligned}
$$

As the matrix $\mathbf{I}$ is 3-diagonal, we can easily compute the function g.

6.3 Inverse transform sampling

Inverse transform sampling also known as the probability integral transform, is a method of sampling a number at random from any probability distribution given its cumulative distribution function (cdf). The probability integral transform states that if X is a continuous random variable with a strictly increasing cumulative distribution function F_X, and if $Y = F_X(x)$, then Y has a uniform distribution on $(0, 1)$.

Let F be a continuous cumulative distribution function, and let F^{-1} be its inverse function: [9]

$$F^{-1}(u) = inf\{x|F(x) = u, 0 < u < 1\}$$

Claim: If U is a uniform random variable on $(0, 1)$ then $F^{-1}(U)$ follows the distribution F.

Proof:

$$
\begin{aligned}
Pr(F^{-1}(u) \leq x) &= Pr(inf\{x|F(x) = u\} \leq x)^{16} \\
&= Pr(u \leq F(x))^{17} \\
&= F(x)^{18}
\end{aligned}
$$

6.4 Efficient implementation with Matrix-Kronecker product

In mathematics, the Kronecker product, denoted by $\otimes$, is an operation on two matrices of arbitrary size resulting in a block matrix. In practice, direct

evaluation of the Kronecker products would cause memory overflow. Den Iseger and Oldenkamp [6] developed an efficient algorithm to avoid direct computation by reshaping and rewriting.

$$(A_{m_A\times n_A}\otimes B_{m_B\times n_B})\,x_{n_An_B\times 1}=\begin{bmatrix} a_{11}B & \cdots & a_{1n_A}B \\ \vdots & & \vdots \\ a_{m_a1}B & \cdots & a_{m_An_A}B \end{bmatrix}\begin{bmatrix} x^1_{n_B\times 1} \\ \vdots \\ x^{n_A}_{n_B\times 1} \end{bmatrix}.$$

With simple matrix notation this comes down to

$$(A_{m_A\times n_A}\otimes B_{m_B\times n_B})\,x_{n_An_B\times 1}=\sum_{j=1}^{n_A}a_{kj}\left(B_{m_B\times n_B}x^j_{n_B\times 1}\right).$$

The following algorithm computes this equation in Matlab®.
Algorithm:
Input: $A_{m_A\times n_A}$, $B_{m_B\times n_B}$, $x_{n_An_B\times 1}$
Output: $Z=(A\otimes B)x$
$x=\text{reshape}^{19}(x,n_B,\text{numel}(x)/n_B)$;
$y=B^*x$;
$y^T=\text{reshape}(\text{transpose}(y),\text{numel}(y)/n_A)$;
$Z=A^*y^T$.

References

[1] Francis A. Longstaff and Eduardo S. Schwartz, 2001. *Valuing american options by simulation: A simple least-squares approach*, The Review of Financial Studies Spring 2001 Vol.14.

[2] Paul Glasserman, 2000, *Monte Carlo methods in Financial Engineering*, Springer.

[3] Mark Broadie and Paul Glasserman, 1997, 2004. *A stochastic mesh method for pricing high-dimensional american options*, Journal of Computational Finance Volume 7 / Number 4, Summer 2004.

[4] Emmanuelle Clement, Damien Lamberton, and Philip Protter, 2002. *An analysis of a least squares regression method for American option pricing*, Finance and Stochastics, Vol.6 October 2002.

[19] Reshapes vector x into no of rows$(x)/n_B=n_A$ pieces of size $n_B\times 1$ vectors

[5] Manuel Moreno and Javier F. Navas, 2003. *On the Robustness of Least-Squares Monte Carlo (LSM) for Pricing American Derivatives*, The Review of Derivatives Research, 6, 2003.

[6] Peter den Iseger and Emőke Oldenkamp *High dimensional transformation algorithms*, working paper, 2007.

[7] Roland van Gaalen, 2004. *Pension funds: Funding index, mismatch risk premium and volaitility*, Society of Actuaries.

[8] Roland van Gaalen, 2003. *Pension funds: A capital adequacy test with a soft mismatch cushion.*

[9] Luc Devroye, *Non-Uniform Random Variate Generation*, New York: Springer-Verlag, 1986.

[10] Roger A. Horn, Charles R. Johnson, 1991, *Topics in Matrix Analysis*, Cambridge University Press.

[11] J.Hull and A.White, *One factor interest rate models and the valuation of interest rate derivative securities*, Journal of Financial and Quantitative Analysis, Vol 28, No.2 June 1993.

[12] J.Hull and A.White, *Numerical procedures for implementing term structure models 2*, Journal of Derivatives 2, 1994.

[13] Theo P. Kocken, 2006. *Curious Contracts: Pension fund redesign for the future.*

[14] Janwillem Engel, Mark Willemse, 2007. *Quantifying the full impact of FTK*, Global Pensions, July 2007.