PART II: REGRESSION FUNCTIONS IN LEAST-SQUARES MONTE CARLO SIMULATIONS

In "Part I: An Introduction to Least-Squares Monte Carlo Simulation" we learnt the definition of a nested simulation and that the computational workload can be reduced if a Least-Squares Monte Carlo (LSMC) technique is used. Furthermore, it was stated that a central aspect of LSMC is the regression used to compensate for the highly reduced number of inner scenarios. In this part, we will introduce an efficient way of estimating and calibrating regression functions in a LSMC environment.

A SIMPLE REGRESSION FUNCTION

Typically, the LSMC approach aims at finding a regressed function that can be used to estimate a value of a portfolio or instrument based on the underlying risk factors. The first step towards finding a regressed function is to generate another function, or a set of functions, to perform the regression over. This function would then consist of an arbitrary set of terms that the underlying portfolio or instrument could depend on, i.e. risk factors such as stock return, volatility, interest rates etc. To demonstrate how a regression function used to determine a portfolio value could look, we can apply the following expression

$$V_{\text{Portfolio}} = \beta_0 + \beta_1 R + \beta_2 \sigma + \beta_3 r + \epsilon, \tag{1}$$

where β_i are the regressed coefficients for the risk parameters, R is the stock return, σ is the volatility, r is the rate and ϵ is an error term.

SELECTING AN APPROPRIATE REGRESSION FUNCTION

The equation stated above is a simplified version of a regression function, since the functions used in LSMC also may consist of several risk factors of higher order, as well as cross terms between the risk factors. There are different approaches when deciding on the polynomial and here we will use a general polynomial that can be written as

$$y = f(\mathbf{x}, \beta, g\{\cdot\}), \tag{2}$$

where $\mathbf{x} \in R^{N\times 1}$ is a vector containing the risk factors and $\beta \in R^{(pN+M+1)\times 1}$ is a vector containing the coefficients for each risk factor. Moreover, $g\{\cdot\}$ is

a set of functions, each describing a cross term, N is the number of risk factors, p is the order of the polynomial and M is the number of cross terms. Thus, we have the following expression for our polynomial

$$y = f(\mathbf{x}, \beta, g(\cdot)) = \beta_0 + \sum_{i=1}^{N}\sum_{j=1}^{p} \beta_{(i-1)p+j} x_i^j + \sum_{i=1}^{M} \beta_{pN+i} g_i(\mathbf{x}). \qquad (3)$$

While analysing the expression we realise that if we add variables to our regression and allow high polynomial orders, the combination of possible cross terms will rapidly increase and result in an extensive function. This raises the question, is it necessary to include all terms in the regression? If not, do we need to limit the amount of risk factors included in our regression, lower our polynomial order, or simply exclude some of the terms from the regression without sacrificing accuracy? By including only the terms we consider significant[1], we can manage to reduce the size of the function without losing accuracy. By doing so, we will:

- ◇ lower function complexity
- ◇ avoid overfitting of the model, i.e. that we model noise instead of real data.

As a starting point, we will restrict ourselves to polynomials of the form described in Equation (3), where the cross terms considered are limited to include two risk factors. By *simple polynomial*, we refer to a regression function that includes the maximum number of possible terms given the polynomial order and said restriction. We then attempt to reduce the number of terms included, thereby introducing what we call *optimised polynomials*.

OPTIMISATION OF THE REGRESSION FUNCTION

When selecting which terms to include in our function, we use an optimisation algorithm based on linear regression with additional constraints. Essentially, the algorithm first sets some of the risk factors' coefficients to zero. It then repeatedly solves the optimisation while adding the risk factor that increases the model's explanatory power the most in each iteration.

The problem can be stated as the following optimisation problem

$$\begin{aligned} \min_{\beta} \quad & \frac{1}{2}\|\mathbf{A}\beta - \mathbf{b}\|^2 = \frac{1}{2}(\mathbf{A}\beta - \mathbf{b})^T(\mathbf{A}\beta - \mathbf{b}) \\ \text{s.t.} \quad & \beta_i = 0, i \in D, \end{aligned} \qquad (4)$$

[1] Significant terms are those that contribute with explanatory power to the model.

where $\mathbf{A}$ is a matrix of the risk factors, $\mathbf{b}$ is a vector of the estimated values (e.g. portfolio values etc.) and D is a set representing the excluded risk factors. By way of explanation, we wish to minimise the error between the regressed risk factors and the portfolio values. To simplify our calculations, we can choose to only consider terms that do not satisfy the condition $\beta_i = 0, i \in D$. In practice, this reduces our problem to

$$\min_{\bar{\beta}} \frac{1}{2}\|\bar{\mathbf{A}}\bar{\beta} - \mathbf{b}\|^2, \text{ where } \bar{\mathbf{A}} \text{ and } \bar{\beta} \text{ are reduced matrices.} \tag{5}$$

To decide which term to add to the regression function, we calculate to what extent each excluded term improves the objective function. Primarily, we repeatedly solve Equation (4),[2] adding one term (term i) in each iteration while holding all $\bar{\beta}$-values constant. If N_1 is the amount of terms in the reduced optimisation problem, the problem can now be written as

$$\begin{gathered} \min_{\beta_i} \frac{1}{2}\|\tilde{\mathbf{A}}\tilde{\beta} - \mathbf{b}\|^2, \\ \text{where } \tilde{\mathbf{A}} = (\bar{\mathbf{A}} \;\; \mathbf{A}_i) \in \; R^{n \times N_1 + 1}, \; \tilde{\beta} = \begin{pmatrix} \bar{\beta} \\ \beta_i \end{pmatrix} \in R^{N_1 \times 1 + 1}, \\ \mathbf{A}_i \in R^{n \times 1} \text{ and } \beta_i \in R^{1 \times 1}. \end{gathered} \tag{6}$$

When solving Equation (6) we are interested in finding an optimal value for each $\beta_i, \; i \in D$, i.e.

$$\frac{\partial}{\partial \beta_i}\left(\frac{1}{2}\|\tilde{\mathbf{A}}\tilde{\beta} - \mathbf{b}\|^2\right) = 0. \tag{7}$$

Secondly, when all optimal β_i have been identified, we calculate each term's improvement of the objective function to decide which one to add. The decision is based on the evaluation of the terms in the objective function value ($\frac{1}{2}\|\tilde{\mathbf{A}}\tilde{\beta} - \mathbf{b}\|^2$) that contain β_i. Finally, the term that optimises the objective function value is added and the problem is then resolved, adding a new term in each iteration.

Since the algorithm is iterative, some additional constraints have been included to enhance effectiveness. The constraints are based on the following measurements:

- ◇ Change in adjusted R^2 - assures that the extra term improves the goodness of fit.
- ◇ Statistical T-test - assures that the regression coefficient for the included term is (with statistical significance) non-zero.
- ◇ Maximum amount of terms - assures that we do not exceed a predefined upper limit of terms to include in the function.

[2] The equation can be solved with QR-decomposition of the reduced matrices.

Due to constraints, we expect our resulting optimised model to have a satisfying amount of terms with a good fit but with reduced risk of overfitting to the input data.

SUMMARY

The main purpose of this article has been to provide insight into how an optimised regression can be performed in a LSMC simulation. We have explained general concepts regarding regression and described ways to choose an appropriate regression function in LSMC. In addition, we shared an algorithm used to optimise the LSMC regression in order to further improve its performance. The following article, "Part III: Effects of Least-Squares Monte Carlo Simulation", will show the results that can be achieved when using LSMC with optimised regressors.

CONTACT US

Do you want to know more about Kidbrooke Advisory?
Please contact us for more information.

e. info@kidbrooke-advisory.com

a. Kidbrooke Advisory Aktiebolag
Birger Jarlsgatan 20
SE-114 34 Stockholm
Sweden

FREDRIK DAVÉUS
Managing Partner
+46(0)761 170 009
fredrik.daveus@kidbrooke.com

EDVARD SJÖGREN
Partner
+46(0)733 182 583
edvard.sjogren@kidbrooke.com

DISCLAIMER