L. van Leeuwen recently started Laura van Leeuwen Consulting which helps companies implementing econometric models. She received her masters in Econometrics and Operations Research (cum laude) and in Quantitative Finance (cum laude) in October 2014 (both VU, Amsterdam). This article is based on her thesis `The Solvency II proxy modelling for life insurance products'. This thesis was supervised by dr. S.A. Borovkova (VU), M. Smith and drs. K. Klein Willink AAG (NN Group). This article is written on her own behalf.

PRACTICAL PITFALLS OF LEAST SQUARES MONTE CARLO

For insurance products with embedded options and guarantees, valuation of liabilities and determination of the Solvency Capital Requirement (SCR) via an analytical approach requires performing nested stochastic simulations and corresponding calculations. Since this is computationally intensive and time consuming, insurers are looking for proxy modelling techniques. Recently a lot of attention has focused on Least Squares Monte Carlo (LSMC). This technique is considered 'a more efficient and useful implementation of curve fitting that can dramatically reduce the runtime of nested stochastic models to the point where they are feasible for use in decision making' (Koursaris). The article written by Van Haastrecht and Plat (2013, De Actuaris) discusses the successful application to an unit linked product with a maturity guarantee.

However, while tested on several insurance products and promoted by multiple researchers and consulting firms, this method is still in the development stage and comes with several practical difficulties that need to be considered before the method can be applied properly. After a short explanation of the LSMC method, I will discuss these difficulties, namely simulation of outer and inner fitting scenarios, risk factor selection and regression function fitting. Finally, I will discuss when LSMC is useful in practice and I conclude that LSMC is a good proxy method if the time-consuming first analysis per product leads to accurate approximations.

Least Squares Monte Carlo

In a complete stochastic-on-stochastic (SoS) valuation we simulate thousands of outer real-world scenarios and thousands of inner risk neutral scenarios per outer scenario, to come up with accurate risk neutral valuations. This is displayed in figure 1a. Because of the dependency of future guarantee cash flows on risk drivers and their dependence on the risk drivers paths, it is common to use multiple inner scenarios per outer scenario in order to ensure estimation accuracy.

In the LSMC approach we still perform thousands of outer real world simulations but we use just a small number, e.g. 3, inner scenarios for each outer one. By carrying out a regression over the resulting inaccurate valuations, we can arrive at reasonably accurate approximations similar to these obtained from a full nested stochastic calculation. This is displayed in figure 1b. The LSMC method is performed by the following steps:

1. Choosing the risk drivers corresponding to the liability value to model.
2. Simulation of the outer and inner fitting scenarios.
3. Estimation of an appropriate regression function by Ordinary Least Squares (OLS).
4. Using real-world scenarios and the fitted function to obtain the liability distribution.

Figure 1: Scenarios and inaccurate valuations
The outer fitting scenarios

The LSMC method requires simulation of thousands of outer fitting scenarios which, in combination with the inner scenarios, are used to compute liability approximations. There are no clear guidelines provided on how the outer scenarios should be simulated.

If real-world scenarios are used for fitting, the relatively small number of scenarios in the tail can lead to a bad tail-value regression fit, while accuracy is most important here. Luckily, the difference in the scenarios used for fitting and estimation allows the use of fitting scenarios considered more appropriate by the modeler. In practice, several modelers choose for an even spread of risk drivers for fitting (Van Haastrecht and Plat, Koursaris) as shown in figure 2b.

Van Haastrecht and Plat showed that for the example of an Unit Linked product with guarantees where for each valuation only 1 inner scenario is used, it takes approximately 20,000 outer fitting scenarios for the average deviation of the liability value to be less than 2 percent. Notice that for some scenarios, the deviation will be higher.

However, an even spread does not necessarily lead to a better (tail) fit. In fact, it can come with a lot of unrealistic scenarios which potentially have a negative effect on the fit of the estimates in the realistic fitting scenarios. In that case, a restricted evenly spread, as displayed in figure 2c, can provide more desirable results.

The inner fitting scenarios

For each of the outer fitting scenarios computed above, the modeler must simulate the corresponding inner scenarios for the time period t=2-maturity. These scenarios must also take the risk factor interdependencies into account. Since insurers might not have these types of scenarios readily available, simulation might be time consuming. For some risk factors it may be economically justified to make simplifying assumptions, i.e. assuming independency w.r.t. other risk factors or their own real-world realizations. The modeler needs to make sure that these assumptions are appropriate and that their impact on the resulting liability estimates are well understood by management.

On top of that, the modeler must choose an optimal number of inner scenarios. Notice that some extreme liability values may only occur in case of both an extreme outer and an extreme inner scenario.

Therefore, using only a single deterministic inner scenario per outer one, may result in systematically overlooking some type of risk. As a result, some products will require the use of e.g. at least 10 inner scenarios per outer scenario in order to fully cover the risk profile of the portfolio.

Risk factor selection and high dimensionality

Before we can simulate scenarios, we must have a set of risk drivers that corresponds to the portfolio risk exposure. Although most insurers by now know the main risk drivers of their product, we here discuss how LSMC handles risk factor selection and how high dimensionality may cause difficulties.

For each possible risk-factor combination, the modeler must evaluate both the risk factor significance and the appropriateness of the corresponding scenarios for the regression function fitting (as discussed above). An indication of risk factor significance in the univariate or bivariate case can be obtained by plotting the liability as a function of the risk factor(s). However, for higher dimensions (caused by larger sets of risk factors), plotting is no longer possible and analysis gets more complicated.

Besides the function fitting, the modeler must perform scenario simulation for all possible risk-factor combinations. If we use an evenly spread of risk factors, an additional risk factor will not impact the outer fitting scenarios of the original risk factors. However, given the interdependencies of the risk factors and the dependencies of inner scenarios on the outer scenarios, it might be necessary to adjust the inner scenarios of multiple risk factors. Notice that the complexity of this task grows exponentially in the number of risk factors considered.

The regression function fitting and high dimensionality

After simulation of the outer and inner fitting scenarios in terms of the risk drivers, the next step is to describe the liability value as a function of the risk factors. More precisely, we must formulate the regression function which describes the liability value as a function of the outer risk factor realizations in a way that allows for OLS estimation.

There are no clear guidelines provided for obtaining an appropriate type of regression curve. For some examples (Koursaris) it seems that the modeler has used an exponential regression function, where in
other examples a simple linear regression function shows a sufficient accurate fit.

In the univariate or bivariate case with simple relations, the relation between the risk factors and the liability value is easily found. However, recall that the whole necessity of SoS approximation is caused by the complexity of the dependency between the products and their various risk factors. Especially for more exotic insurance products, it is somewhat naive to assume a simple linear relation between all relevant risk factors and the liability estimates.

Univariate non-linearities might be analytically solved and estimated by finding appropriate transformations. A basic approach is simply adding more powers of a risk factor as explanatory variables. But in addition to non-linearity, the effects of the various risk factors may not be additive which further complicates the fitting. It may occur that the functional form considered to be most appropriate, does not allow for OLS estimation. Rejection of the OLS-restriction and allowing more time-consuming estimation processes will make the LSMC less attractive. Especially for larger sets of risk-factors and exotic insurance products this method can fail to result in a function that well describes the relation between the risk factors and the liability values. Choosing the tail scenarios and choosing an appropriate regression function form that can be estimated by OLS becomes a complex exercise. Notice that this complexity increases exponentially in the number of dimensions.

When to use LSMC
We have seen that LSMC leaves several choices to the modeler. The combination of risk factor selection, fitting scenario simulation, and regression function fitting, can be a complex exercise that doesn't necessarily leads to good results. In order to guarantee accuracy of the LSMC results, the modeler must first validate the LSMC proxy function. Only if the model succeeds in producing real-world approximations which are sufficiently similar to those resulting from a full MC-study, can we use the LSMC approximations for valuation. Thus, as the other proxy methods, the LSMC method must always be used in combination with a full SoS approach. For LSMC to be useful, the proxy function must be used several times before it requires re-fitting.

At first instance we suggest the use of LSMC for large portfolios whose value is of key importance to the company, and where frequent valuations obtained in real-world time are useful for management purposes. For such a portfolio the first analysis is most likely to be worth the effort. Especially for lower dimensions we can expect the estimates to be accurate, also in the tails.

Easy to use, better approximations, fast computation?
Now, is LSMC a bad estimation technique? No, as Van Haastrecht and Plat illustrated, LSMC can provide accurate approximations in real-time, and therefore can be a major contribution to the set of proxy techniques used by insurers today. But application of LSMC leaves the modeler with a time-consuming first analysis per product. Only when the modeler succeeds in producing accurate real-world approximations similar to those resulting from a full MC-study, the model is ready for use.

References
Koursaris A. (2011): The advantages of Least Squares Monte Carlo, Barrie & Hibbert insights
Van Haastrecht A., Plat R. (2013): Toepassing Least Squares Monte Carlo op unit linked met eindgarantie, De Actuaris