A new hybrid Random Number Generator for more accurate valuation of insurance liabilities

Hervé Andrès
Pierre-Edouard Arrouy
Paul Bonnefoy
Alexandre Boumezoued

Valuing an insurance balance sheet is a complex exercise which requires the use of stochastic economic scenarios. Various tests should be performed to ensure such valuation is produced in a reasonable manner, including the martingale tests on the economic scenarios and the leakage test on the insurance asset-liability management (ALM) model, i.e., initial market value of asset is equal to the sum of the best estimate of liabilities (BEL) and the present value of future profits (PVFP).

In practice, given the run time constraints of typical ALM models, the number of economic scenarios to be considered is limited. Hence, there is a need to develop techniques to ensure that the stochastic valuation of BEL and PVFP converges towards their true values and therefore that the leakage is reasonable and stable between different valuation dates. One potential solution is to enhance the Random Number Generator (RNG) used to generate the stochastic economic scenarios. In this paper, we present a new RNG and demonstrate its efficiency over existing RNGs for the valuation of stochastic BEL. We also discuss the need for universal and interpretable validation strategies for martingale tests for such types of RNGs.

The use of RNGs and their requirements

The increasing use of stochastic economic scenarios for the valuation of insurance liabilities—e.g., Solvency II, International Financial Reporting Standard (IFRS) 17, Long-Duration Targeted Improvements (LDTI), International Capital Standard, Risk-Based Capital regimes in Asia—is putting more pressure on the operational processes of insurance companies, with a particular focus on the ALM model run time when, for example, 1,000 or more stochastic economic scenarios are used for BEL and PVFP valuation. Besides, there is an increasing need to produce reliable and stable valuation estimates over time and for different economic conditions. As one of the key inputs of the Economic Scenario Generator (ESG), the RNG plays a critical role in the quality of the economic scenarios generated and subsequently on the quality of the assessment of the interactions between the assets and liabilities of the insurer, and ultimately on the convergence of the stochastic BEL and PVFP.

At first, any Economic Scenario Generator (ESG)—hence the underlying RNG—shall satisfy the three following conditions:

- Simulations consistently distributed according to the mathematical specification of the model
- Properly correlated risk drivers
- Independent scenarios

However, those conditions are typically not sufficient to guarantee reasonable validation tests performed on the economic scenarios (e.g., martingale tests, repricing tests and correlation tests) if the number of simulations is limited, which is the case in practice. In the worst-case scenario, validation tests may fail even though an appropriate economic scenario generation process is in place. As a consequence, this could lead to a significant leakage as well as issues with Monte Carlo repricing mis-estimation. To this extent, insurance and reinsurance undertakings are required to demonstrate the quality of the RNG they are using. This requirement is clearly mentioned under Solvency II by the following guideline regarding the valuation of the technical provisions:

"Insurance and reinsurance undertakings should ensure that (pseudo) random number generators used in an ESG are properly tested."[1]

Besides, it is also an area of interest for regulators. In particular, in the second half of 2020, the French Prudential Supervision and Resolution Authority (ACPR) carried out a review of the Economic Scenario Generators (ESGs) used by a sample of 15 French insurance companies. This review was based on a series of on-site checks and the key conclusions were set out in a paper[2] summarising the diversity of practices observed and providing insights on some of the best practices, with a particular focus on the assessment and validation of the uncertainty around the stochastic valuation.

The following paragraphs discuss some of the more general market practices mentioned by the ACPR for making and validating stochastic assessments. It is also worth noticing that such practices are also commonly used in other insurance markets across the world.

Practitioners have typically observed discrepancies between the "Monte Carlo" valuation estimate of the BEL and PVFP and their true expected values, with the exact gap changing from time to time depending on economic conditions and/or stress tests performed. This discrepancy is referred to as "leakage." Setting aside the ALM model error as potential source of leakage, this gap is generally due to the convergence error, given that a limited number of risk-neutral economic scenarios is produced in practice, typically between 1,000 and 5,000. To try to make this gap lower and/or more stable, some companies have investigated a few more pragmatic methods, such as:

- **Seed optimisation approach**. Because most of the RNGs depend on a core parameter called a "seed," it is tempting to select the seed such that an overall threshold on leakage is met. However, in practice the seed selection process remains exploratory and even complex when a large number of risk drivers is considered (e.g., interest rate, equity, credit, multiple currencies etc.), and the optimised seed may not always be stable—for example, the leakage may increase under a different set of economic conditions based on the "optimised" seed as some ESG validation tests may improve while some others may deteriorate.
- **Moment matching adjustment.** This refers to adjusting the Monte Carlo estimates of moments (of order 1 or higher) to match their true expected values. Examples include adjusting the Monte Carlo average when assessing the martingale test to its true value (moment matching at order 1), or scaling the variance with an objective to match some volatility targets (moment matching at order 2) or any other distributional characteristics of higher orders (e.g., skewness, kurtosis). As such, this approach modifies the distribution of the outcomes (in particular, scenarios are not independent anymore), with possible adverse impact on repricing tests.

[1] EIOPA. Guideline 59 on the valuation of technical provisions – random and pseudo random number generators.

[2] ACPR (7 December 2020). Economic Scenario Generators: Points of attention and good practices. Retrieved 1 December 2022 from https://acpr.banque-france.fr/generateurs-de-scenarios-economiques-points-dattention-et-bonnes-pratiques.

Additional techniques have also been used by insurance companies in practice and typically produce more stable outcomes or are less subject to expert judgement. One of them relies on **variance reduction techniques**. In particular:

- **Antithetic variables**: The basic principle of antithetic variables is to generate and use stochastic scenarios by pairs (X_1, X_2) that are negatively correlated such that estimates based on $\frac{X_1+X_2}{2}$ have lower variance than those based on independent values of (X_1, X_2). In practice, such approach is relatively straightforward to implement. However, the convergence rate improvement remains rather limited.[3]
- **Control variates**: The underlying idea is to find a variable L^* that is correlated to the liability valuation of interest L and for which the true expected value is known; this can be, for example, economic variables, the asset portfolio or any replicating portfolio.[4] In the end, the expected value of the liabilities is calculated as:

$$\mu_{L^*} + \widehat{\mathbb{E}}[L - L^*],$$

  i.e., by adding the true value of the control variate expectation to the Monte Carlo estimate of the difference between the liabilities and its control variate, the latter having better convergence properties given the lower variance of the estimate at stake due to positive correlation between both.

Note that when antithetic variables are used, scenarios are not independent anymore, so that any estimation error assessment based on the scenarios shall take into account their dependence, especially regarding the underlying variance estimate, as mentioned by the ACPR:

"When variance reduction methods are used (antithetic variables for example), the confidence intervals or the p-values of the statistical tests must be calculated taking into account the lack of independence between the scenarios."[2]

Another technique is based on **quasi-Monte Carlo RNG**. This is covered more specifically in the rest of this paper.

Finally, other techniques such as *reweighting*[5] can help achieving higher convergence by weighting the scenarios distribution in view of better matching of martingale properties and the replication of market prices.

Existing RNGs and their limitations

PSEUDO RANDOM NUMBER GENERATORS

The classic Monte Carlo method usually refers to the simulation of uniformly distributed random numbers on [0,1], called pseudo random numbers, which aims at reproducing random trials, as illustrated in Figure 1. By the central limit theorem, the convergence rate of a standard Monte Carlo method is $\frac{1}{\sqrt{N}}$, where N is the number of simulations.

One of the simplest methods to generate pseudo random numbers is to use a linear congruential generator. Given four integers (a, b, m, u_0), a pseudo random sequence $(x_n)_{n\in\mathbb{N}}$ can be constructed as:

$$x_n = \frac{u_n}{m}$$

with:

$$u_n = au_{n-1} + b \mod m.$$

The integer u_0 is called the seed and should be provided by the user (if not, it is generally chosen using the computer clock), whereas the integers a, b, m are determined at initial stage so as to ensure good statistical properties of the generated sequences. An example of a generator based on linear congruential operations is Witchmann-Hill, which has been used in practice by a number of companies for insurance balance sheet valuation.

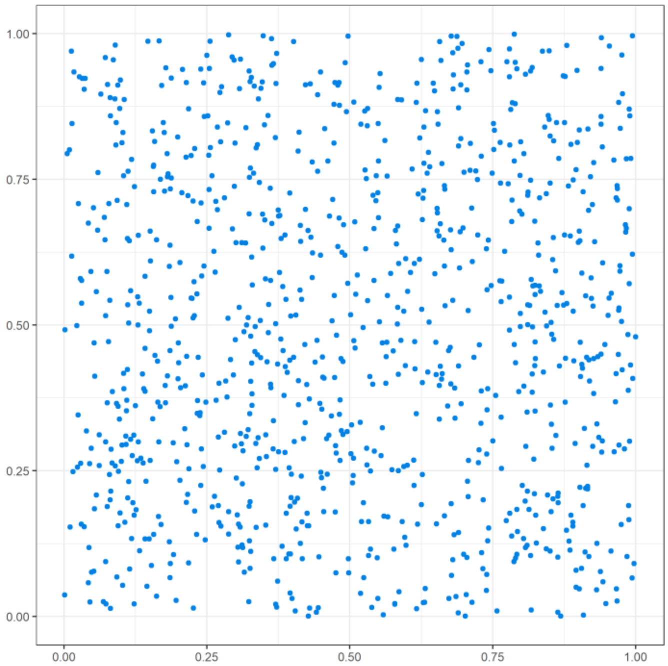
[3] Paskov, S. H. & Traub, J. F. (1995). Faster valuation of financial derivatives. Journal of Portfolio Management 22(1), 113–123.

[4] Hull, J., & White, A. (1988). The use of the control variate technique in option pricing. Journal of Financial and Quantitative analysis, 23(3), 237-251.

[5] Howell, C., Leitschkis, M., & Ward, R. (July 2019). ESG Rebase. Milliman White Paper. Retrieved 2 December 2022 from https://fr.milliman.com/fr-fr/insight/esg-rebase.

A more sophisticated version of a pseudo random generator is Mersenne Twister,[6] which is widely used by ESG practitioners.

FIGURE 1: REPRESENTATION OF THE 1ST (X-AXIS) AND 2ND (Y-AXIS) COORDINATES OF THE FIRST 1,000 PSEUDO RANDOM NUMBERS GENERATED WITH THE MERSENNE TWISTER ALGORITHM



QUASI RANDOM NUMBER GENERATORS

The idea of constructing quasi-random numbers follows the desire to achieve faster convergence than the Monte Carlo rate of $\frac{1}{\sqrt{N}}$. More specifically, consider the problem of approximating the integral:

$$I(f) = \int_{[0,1]^d} f(x)dx.$$

This integral can represent the price of liabilities, of derivatives (option, swaption etc.), a martingale test or the liability value. Quasi-random points $x_1, \dots, x_n$ in $[0,1]^d$ allow us to approximate this integral as:

$$I(f) \approx \frac{1}{N}\sum_{n=1}^{N} f(x_n)$$

The so-called discrepancy of the sequence (x_n), denoted by $D(x_1, \dots, x_N)$, measures how close is the sequence to a uniform distribution. The discrepancy can be used to control the estimation error as follows:

$$\left| I(f) - \frac{1}{N}\sum_{n=1}^{N} f(x_n) \right| \leq C(f)D(x_1, \dots, x_N)$$

where $C(f)$ is a constant, called the Hardy-Krause variation, that depends on the regularity of the payoff function f. It is now acknowledged that sequences having a discrepancy of

$$\frac{(\log N)^d}{N},$$

where d is the dimension at stake are so-called low-discrepancy sequences.[7]

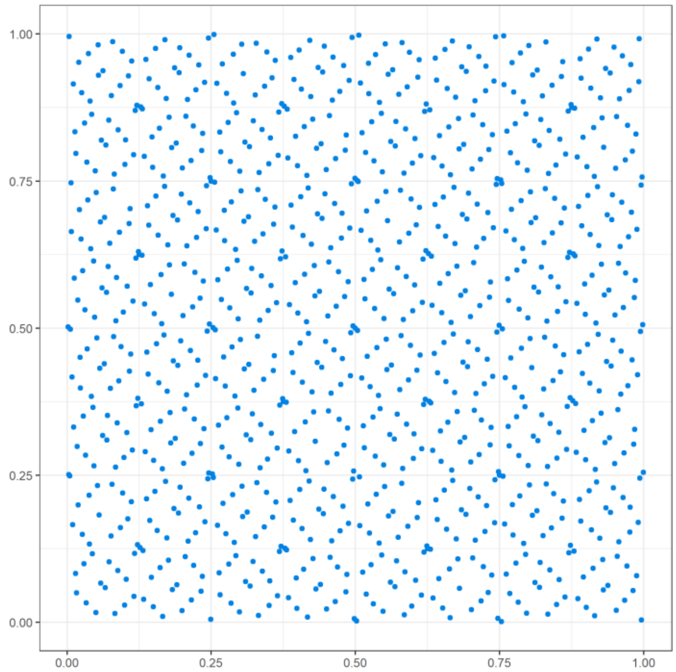
[6] Matsumoto, M., & Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation (TOMACS), 8(1), 3-30.

[7] Niederreiter, H. (1992). Random number generation and quasi-Monte Carlo methods. Society for Industrial and Applied Mathematics.

Such sequence does show a specific deterministic pattern allowing us to achieve a very low estimation error, as illustrated in Figure 2. One of the most famous low-discrepancy sequences is that of Sobol,[8] who provided a theorem that ensures a lower estimation error by using quasi-random number generation with the Sobol sequence rather than using a pseudo random method with a uniform distribution.

It appears that, in many financial applications with reasonable dimensions, quasi-Monte Carlo outperforms ordinary Monte Carlo.[3] However, practitioners are struggling to achieve good convergence properties for higher dimensional problems. As such, a typical example of economic scenarios, as shown in the case study section of this paper below, with 16 risk factors, simulated on a monthly time step and over 60 years of projection, leads to a dimension $d = 60 \times 12 \times 16 = 11{,}520$. With such a high dimension, the convergence rate of the low-discrepancy sequence is relatively low and does not efficiently compete with pseudo random numbers (even with possible additional variance reduction techniques as outlined above).

FIGURE 2: REPRESENTATION OF THE 1ST (X-AXIS) AND 2ND (Y-AXIS) COORDINATES OF THE FIRST 1,000 QUASI-RANDOM NUMBERS GENERATED WITH SOBOL SEQUENCES



To summarise, pseudo RNGs are relatively flexible and universal, in the sense that they can be used when dealing with a high number of dimensions (driven by the number of currencies, risk factors, time steps) because they achieve a similar convergence rate, although this convergence rate is far from optimal. On the other hand, quasi-random numbers have appealing convergence properties, but the convergence rate deteriorates as the dimension increases.

As such, the objective of the next section is to set out the key steps for the construction of a hybrid RNG that takes advantage of key features of both the pseudo RNGs and the quasi-RNGs.

[8] Sobol, I. Y. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 7(4), 784-802.

A hybrid RNG with improved characteristics

THE SO-CALLED HYBRID RNG RELIES ON THREE CORE FEATURES:

1. Using a *quasi-RNG* as a basis in a dimension that is reasonable in order to preserve an efficient convergence rate; in the present case, this is achieved by using the quasi-random numbers for the time steps at a lower frequency (e.g., annual) than the final simulated scenarios (e.g., monthly). The core sequence will be selected among a collection of existing sequences (e.g., Sobol, Faure, Torus) based on the quality of outcomes from the ESG, including martingale and repricing tests.
2. Relying on a *Brownian bridge* technique to recompose paths of Brownian motions at the refined timestep (e.g., monthly) to support the approach mentioned in the first point above. The Brownian bridge will rely on pseudo random numbers in order to preserve the low dimensionality of the use of the quasi-random numbers that will be restricted to a less granular projection time step (e.g., annual).
3. A so-called *randomisation* approach is considered. It consists in perturbing the (deterministic) quasi-Monte Carlo sequence by a random transformation (such as a shift), in order to derive better convergence properties.

The three core ingredients are further detailed in the following.

LOW-DISCREPANCY SEQUENCE

The so-called Van der Corput sequence is a low-discrepancy sequence in dimension 1, which is therefore not of practical use, but is the building block of many other low-discrepancy sequences. Indeed, a sophisticated generalisation of the Van der Corput sequence in dimension d is the Faure sequence, which will be a first RNG of interest in this paper. The Faure sequence is theoretically well adapted to high-dimension problems. Indeed, its discrepancy satisfies:

$$D(x_1, \dots, x_N) \sim F_d \frac{(\log N)^d}{N}.$$

where F_d converges to 0 as d converges to infinity

Like the Faure sequence, the well-known Sobol sequence results from transformations of the Van der Corput sequence. Differences between the Sobol and the Faure sequences include that the calculations underlying the Sobol numbers generation are more computationally efficient, and that there are structural limitations about the current availability of Sobol sequences in terms of dimension within existing packages. The discrepancy of the Sobol sequence in dimension d satisfies:

$$D(x_1, \dots, x_N) \sim \frac{2^{H(d)}(\log N)^d}{N},$$

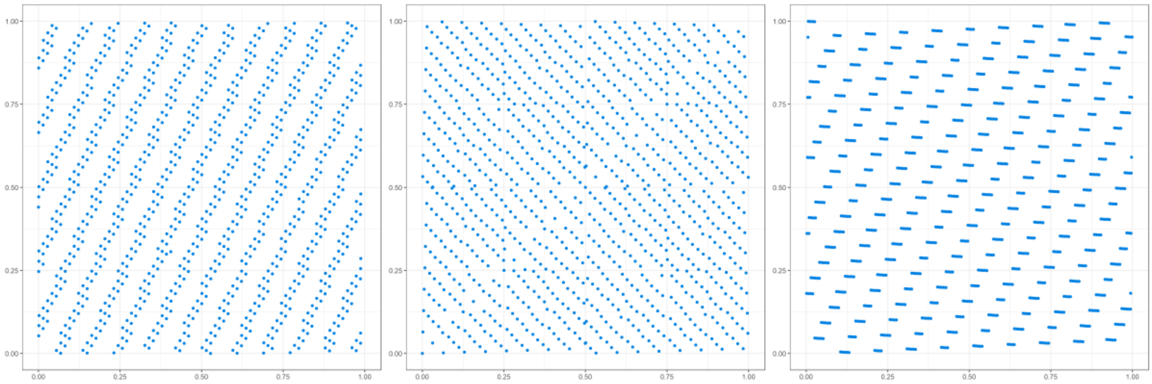
where $H(d)$ is a function of d growing faster than d but slower than $d \log d$

Although this theoretical bound is definitely worse than for the Faure sequence in high-dimension problems, the Sobol sequence seems to be the most competitive low-discrepancy sequence when using practical dimensions, as shown in the literature and as will also be the case in this paper.

Unlike the previous sequences, the Torus sequence, which will be the third approach of interest, is not built from the Van der Corput sequence and is derived by a dedicated method based on predefined prime numbers.

An illustration of the three quasi-Monte Carlo sequences is provided in Figure 3.

FIGURE 3: REPRESENTATION OF THE 9,999TH (X-AXIS) AND 10,000TH (Y-AXIS) COORDINATES OF THE FIRST 1,000 POINTS OF THE FAURE (LEFT), SOBOL (MIDDLE) AND TORUS (RIGHT) SEQUENCES IN DIMENSION 10,000



BROWNIAN BRIDGE

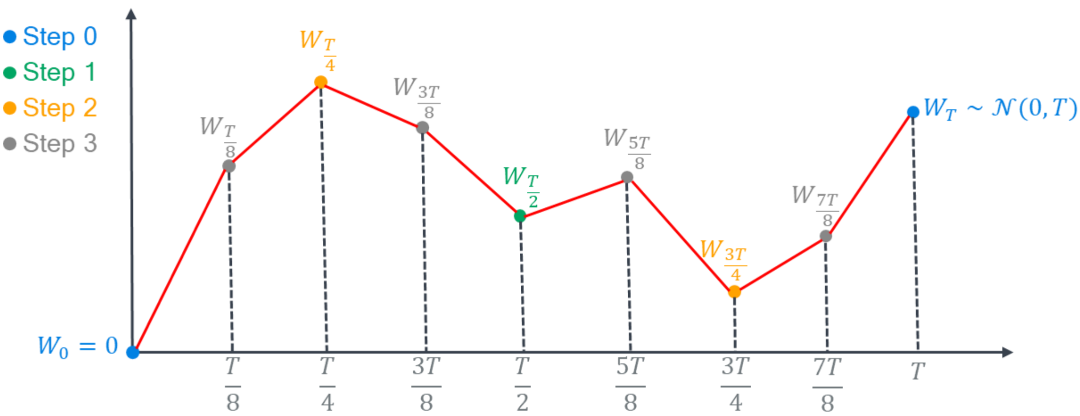
The Brownian bridge is a well-known approach to simulate paths of a Brownian motion. Instead of simulating paths in a forward manner, as in the Euler scheme for example, the Brownian bridge allows us to simulate a path conditionally on past and future outcomes being given. It relies on the following result, which states that, given outcomes $W_{s_1}, \dots, W_{s_m}$ at some less granular time steps $s_1, \dots, s_m$, the distribution of W at some point s within an interval $[s_i, s_{i+1}]$ is:

$$W_s = \frac{(s_{i+1} - s)x_i + (s - s_i)x_{i+1}}{s_{i+1} - s_i} + \sqrt{\frac{(s_{i+1} - s)(s - s_i)}{s_{i+1} - s_i}} G,$$

where $G \sim \mathcal{N}(0,1)$ is independent from $W_{s_1}, \dots, W_{s_m}$

Such a formula can then be applied recursively to build the Brownian path within an overall interval, as illustrated in three main steps in Figure 4.

FIGURE 4: BROWNIAN BRIDGE



Thus, when a Brownian motion trajectory is simulated using the Brownian bridge, the overall shape of the trajectory is determined by the first standard normal variables that have been used in the first step. This property is particularly relevant in the context of low-discrepancy sequences in high dimension. Indeed, some coordinates of a low-discrepancy sequence may exhibit better uniformity properties than others; for instance, Figure 3 illustrates a deterioration of uniformity with high dimensions. The Brownian bridge allows us to leverage the uniformity property by using the first coordinates to generate the coarse structure of the Brownian paths. Hence, this approach is particularly useful for low-discrepancy sequences that can't be generated beyond a certain dimension (such as the Sobol sequence) or whose high-dimensional behaviour deteriorates.

RANDOMISATION

In the context of quasi-Monte Carlo generators, randomisation consists in randomising low-discrepancy sequences using pseudo random numbers. Starting from a point set $P_n = \{x_1, \dots, x_n\}$ consisting of n terms of a low-discrepancy sequence, randomisation produces a point set $\tilde{P}_n = \{\tilde{x}_1, \dots, \tilde{x}_n\}$ satisfying:

- Each point $\tilde{x}_i$ is uniformly distributed over $[0,1]^d$
- The set $\tilde{P}_n$ still has the low-discrepancy property

It can be shown that, in some settings, randomisation improves the accuracy of the quasi-Monte Carlo approximation. As an example, one result[9] states that the approximation error using some specific randomisation is $O(1/n^{1.5-\epsilon})$ while the approximation error is $O(1/n^{1-\epsilon})$ without randomisation. This result has only been proven for smooth integrands, but nevertheless it indicates that randomisation can be a solution to reduce the approximation error.

Several randomisation techniques exist but they can't be applied to all kinds of low-discrepancy sequences. The case study presented in the next section introduces a randomisation, which consists in distorting quasi-random points based on additional pseudo random perturbation. This randomisation works for the Faure and Sobol sequences.

ESG case study

SETTING

The proposed case study relies on economic scenarios generated for three currencies and a number of asset classes as detailed in the table in Figure 5; for each model, the number of risk factors is also presented in order to better appreciate the high number of dimensions.

FIGURE 5: CURRENCY, ASSET CLASS, AND RISK FACTORS

CURRENCY	ASSET CLASS	NUMBER OF RISK FACTORS
EUR	Nominal interest rate	2
	Real interest rate	2
	Equity	1
	Real estate	1
USD	Nominal interest rate	2
	Real interest rate	2
	FX rate	1
GBP	Nominal interest rate	2
	Real interest rate	2
	FX rate	1

Under this setting, 1,000 stochastic risk-neutral scenarios are simulated over a projection horizon of 60 years using a monthly time step.

We consider the following methods in this numerical experiment:

- One ordinary Monte Carlo method relying on the Mersenne Twister RNG and making use of antithetic variables. This method is referred to as "MT-AV."
- Three hybrid methods as specified in the table in Figure 6.

[9] Owen, A. B. (1997). Scrambled net variance for integrals of smooth functions. Annals of statistics, 25(4), 1541-1562.

FIGURE 6: HYBRID METHODS

METHOD	LOW-DISCREPANCY SEQUENCE	RANDOMISATION
FaureG-DS-hybrid60	Faure Generalised	Digital shift
Sobol-DS-hybrid60	Sobol	Digital shift
Torus-hybrid60	Torus	No randomisation

RESULTS ON THE ECONOMIC SCENARIOS

The comparison between the various methods is produced using the root mean squared relative error (RMSRE), which is computed for all martingale and repricing tests for all asset classes. This RMSRE is defined as follows:

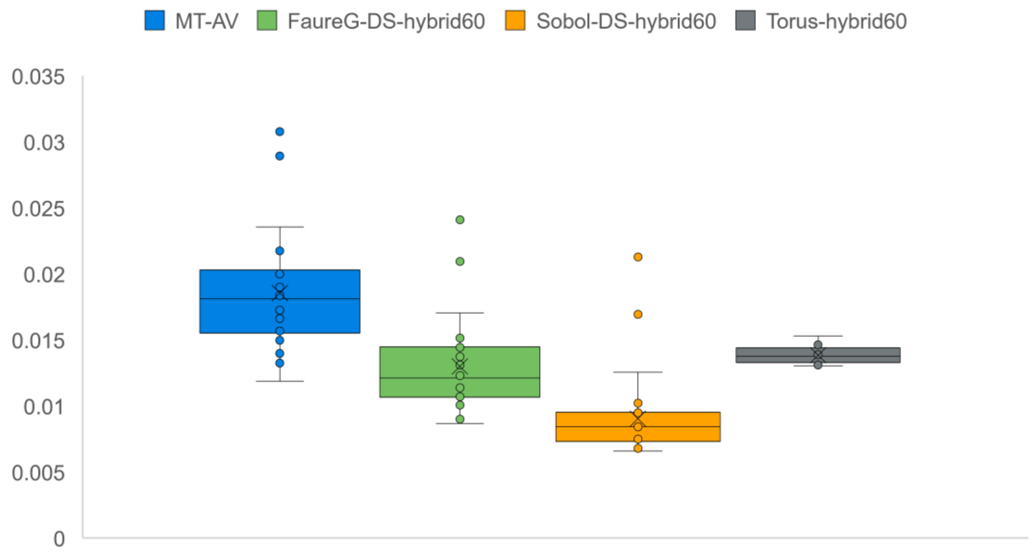
$$RMSRE = \sqrt{\frac{1}{|\mathcal{T}|}\sum_{i\in\mathcal{T}}\left(\frac{\hat{E}_i - E_i}{E_i}\right)^2},$$

where $\mathcal{T}$ is the set of all martingale and repricing tests, $\hat{E}_i$ is the estimated value over the simulations for test indexed by i and E_i is the corresponding true value

For example, if i is the index of a martingale test of the discount factor of maturity T, then E_i is given by the zero-coupon bond price $P(0,T)$ at time 0 and maturity T and $\hat{E}_i$ is given by the empirical mean $\frac{1}{N}\sum_{j=1}^{N} D^{(j)}(0,T)$, where $D^{(j)}(0,T)$ is the discount factor value of maturity T in the j-th simulation. Overall, the lower the RMSRE, the better the method.[10]

Each method is illustrated for 30 different seeds, resulting in 30 values of the RMSRE for each. Note that pure quasi-Monte Carlo methods do not depend on a seed but here we consider quasi-Monte Carlo methods in combination with hybrid Brownian bridge and digital shift, which involve pseudo random numbers that do depend on a seed. The results are presented in Figure 7.

FIGURE 7: RMSRE FOR THE DIFFERENT RNG METHODS



[10] The metric used here is for illustration purposes. In practice, this could be adjusted to each company's risk profile, by considering weights in line with the asset and liability sensibility, in order to give more or less importance to the martingale and/or repricing deviations, depending on their impact on the liabilities. It is also worth mentioning that correlation tests could be included in the metrics.

We observe that the three hybrid methods outperform the Monte Carlo method (MT-AV) and that the Sobol-DS-hybrid60 is the best among the three tested. In particular, we see some differences due to the selection of hybrid seeds. This is however compensated by the fact that the worse seeds for the hybrid Sobol RNG provide RMSRE results close to most seeds for the pseudo RNG with antithetic variables. It is also worth highlighting that the hybrid Torus RNG is not randomised, hence showing lower variance in RMSRE but with higher variance in average compared to Faure and Sobol.

In order to illustrate the average improvements of quasi-Monte Carlo over ordinary Monte Carlo in terms of martingale and repricing tests, we plot some of them for both approaches and for an "average seed," i.e., for a seed yielding a RMSRE close to the average in the previous box plots. In Figure 8, the discount factor martingale tests for the GBP currency are shown, while in Figure 9 the martingale test for the foreign exchange (FX) rate from USD to EUR is presented. As it is shown, the use of an "average seed" leads to martingale tests that are closer to their true expected values when using the hybrid RNG as opposed to the pseudo RNG with antithetic variables. It is also worth noting that the estimate of the empirical variance is also more precise when using the low-discrepancy RNG, leading to more robust confidence interval estimates.

FIGURE 8: DISCOUNT FACTOR MARTINGALE TESTS FOR MT-AV (LEFT) AND SOBOL-DS-HYBRID60 (RIGHT)

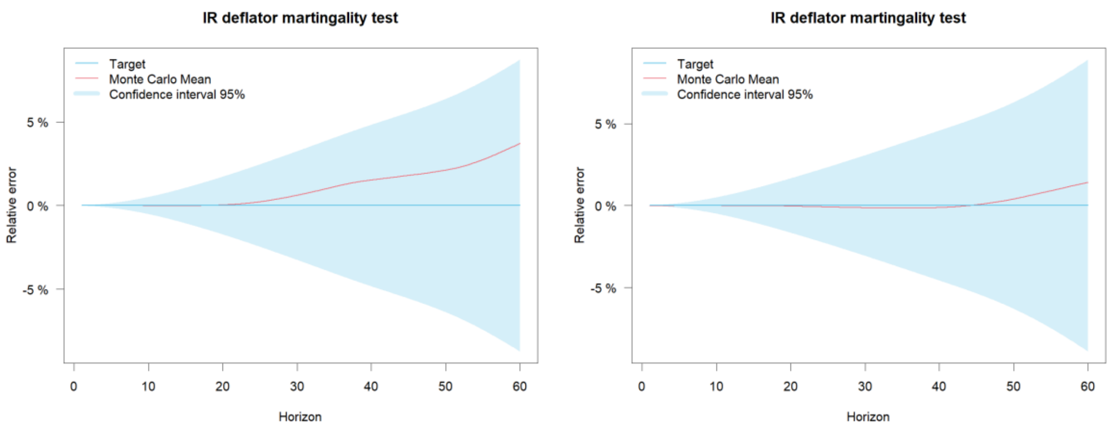
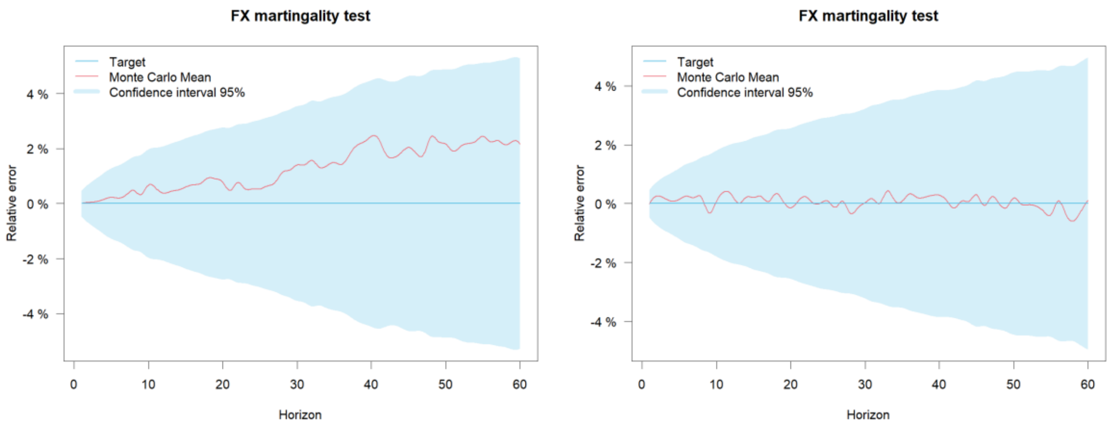
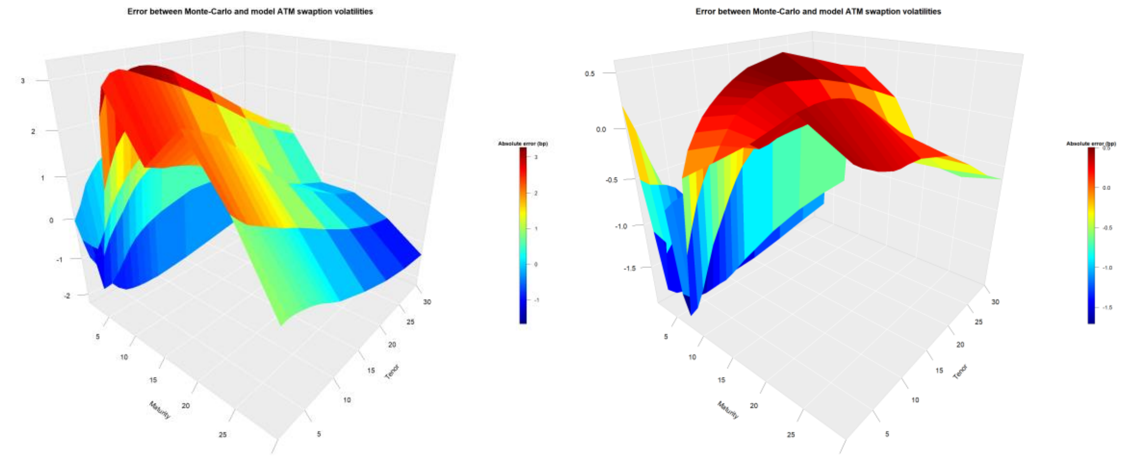


FIGURE 9: FX MARTINGALE TESTS FOR MT-AV (LEFT) AND SOBOL-DS-HYBRID60 (RIGHT)



We also show in Figure 10 the difference between Monte Carlo swaption volatilities (at-the-money, for USD currency as an illustration) and the volatilities repriced by the model (through approximate formulas, as used in the calibration process). We can see in particular that the discrepancies between Monte Carlo and model swaption volatilities are lower with the hybrid RNG compared to the pseudo RNG with antithetic variables. Note that it is not expected to get an error of zero as the model prices are derived by approximations (including the so-called freezing) in order to recover tractable numerical methods for pricing. Nevertheless, this tends to show that the improved RNG not only materialises with better martingale tests, but also better repricing tests in the sense that the Monte Carlo prices are closer to their true value, so that additional sampling noise around Monte Carlo repricing is further reduced. As such, this provides a tool to better assess pure approximations in pricing formulas, and then better governance around validation of repricing tests.

FIGURE 10: USD SWAPTION REPRICING TESTS FOR MT-AV (LEFT) AND SOBOL-DS-HYBRID60 (RIGHT)



Note that Figure 10 represents the discrepancies between the Monte Carlo swaption volatilities (calculated on the sample of risk-neutral simulations for interest rates) and the model volatilities (computed through the closed-form approximations, as used in the calibration process). The discrepancies are represented in the scale in terms of basis points (bps), over two axes, namely maturities and tenors of the swaptions.

Impact study on a cash flow model

To support the above analysis, we have developed a case study assessing the impact of the RNG on the key quantitative metrics of an insurance company. To this extent we simulate 3,000 stochastic paths of the financial risk factors at 31 December 2021 with the Milliman ESG, the Cloud Hosted Economic Scenario Simulator (Milliman CHESS™),[11] using both the Mersenne Twister with antithetic variables (MT-AV) RNG and the Sobol-DS-hybrid60 RNG introduced in this case study. These scenarios are then read as input of a cash flow model of a typical representative French life insurance company. In the following, we compare three key indicators, namely:

- The best estimate of liabilities (BEL). The BEL values presented in this section have been normalised by the total asset market value.
- The value of in-force (VIF), which represents the present value of future profits of the insurance company. As for the BEL, the VIF has been normalised by the total asset market value.
- The ALM leakage defined as the initial market value minus the sum of the BEL and the VIF (before reallocation of the model leakage). In the following the model leakage is given as a percentage of the total market value of assets.

COMPARISON OF THE RNGS FOR SEVERAL SEEDS

Firstly, we compare the key indicators obtained with the MT-AV and the Sobol-DS-hybrid60 RNGs. Because the methods rely on a limited number of scenarios (here 3,000), the value of the key indicators can depend on the seed that has been used to initialise the RNG. In order to measure the sensitivity of the key indicators to the seed, the results are produced using 12 different seeds. As mentioned earlier, although the Sobol-DS-hybrid60 method uses quasi-random numbers, the seed has an influence because this hybrid RNG encompasses for Brownian bridging and randomisation features, which both rely on pseudo random numbers. For reference, the graphs in Figures 11 to 13 also provide the average BEL, VIF and leakage computed over the two RNGs and all the 12 seeds.

[11] See https://www.milliman.com/en/products/milliman-chess.

FIGURE 11: COMPARISON OF THE BEL OBTAINED WITH THE MT-AV AND THE SOBOL-DS-HYBRID60 RNG (EACH POINT REPRESENTS A SEED)

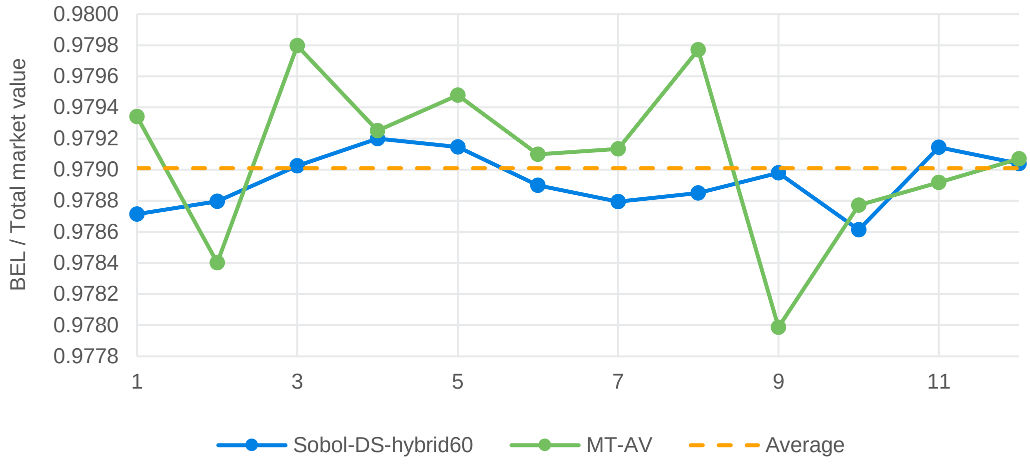


FIGURE 12: COMPARISON OF THE VIF OBTAINED WITH THE MT-AV AND THE SOBOL-DS-HYBRID60 RNG

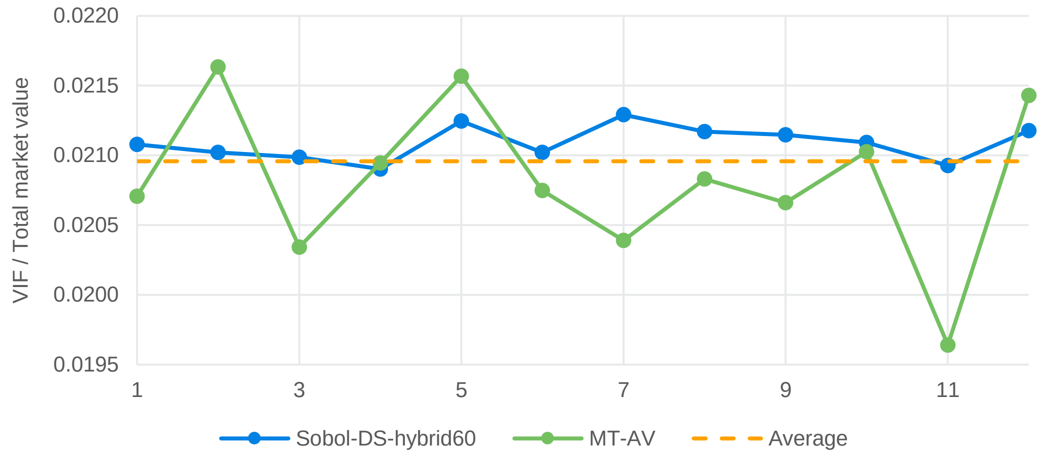
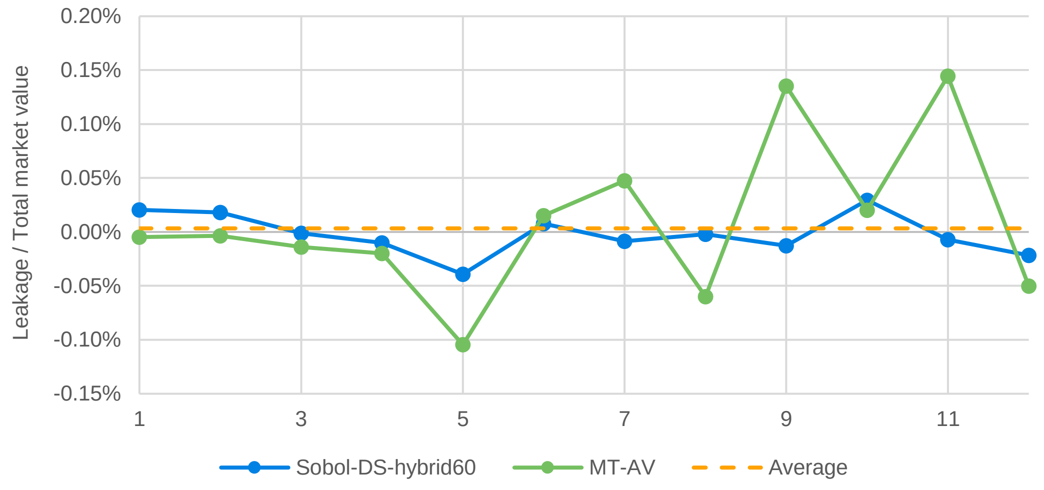


FIGURE 13: COMPARISON OF THE LEAKAGE OBTAINED WITH THE MT-AV AND THE SOBOL-DS-HYBRID60 RNG



We observe that the hybrid Sobol RNG is more stable when assessing the key indicators; in particular this method induces BEL and VIF values nearer to the overall average as well as a model leakage closer to the target value of 0. The overall average leakage value can be interpreted as the limit when increasing the number of simulations; it is very close to 0, as expected.

We set out in the table in Figure 14 a few additional descriptive statistics on the VIF. Note that similar findings can be drawn from the analysis of the BEL. The minimum (or maximum) variation is computed as the minimum (or maximum) observed VIF divided by the average VIF computed over the 12 seeds.

FIGURE 14: VIF STATISTICS

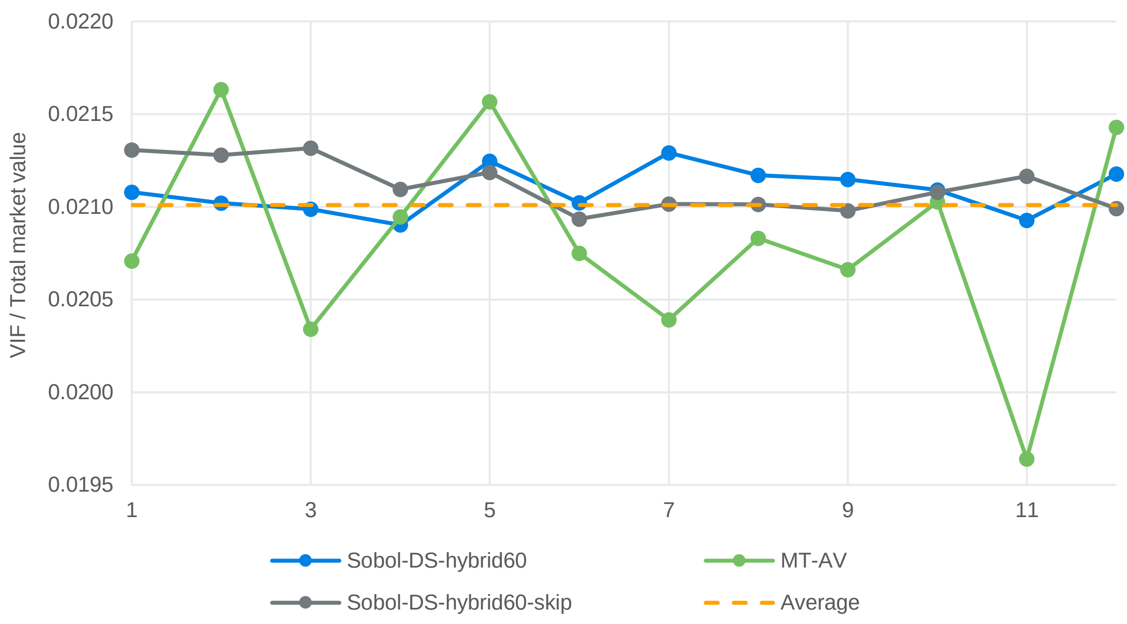
	MEAN	STANDARD ERROR	MIN. VARIATION COMPARED TO MEAN	MAX. VARIATION COMPARED TO MEAN
MT-AV	0.02083	0.056%	-5.7%	3.9%
SOBOL-DS-HYBRID60	0.02109	0.012%	-0.9%	1.0%

These results confirm that the MT-AV RNG has the highest variability, with a range of uncertainty reaching up to 9.5% of the total VIF. In comparison, the Sobol-DS-hybrid is significantly more stable, as the maximal variation of the VIF is about four to five times smaller.

IMPACT OF THE SOBOL SEQUENCE STARTING POINT

Sobol sequences attribute a polynomial function to each dimension of the problem to be sampled. Each simulation k is then given an integer i_k that is passed to the polynomial functions, outputting a “random number” for each dimension. Generally, this integer is determined as the simulation index, formally $i_1 = 1, \dots, i_n = n$ (with n the number of paths to be generated). The results previously discussed for the Sobol-DS-hybrid60 RNG rely on this choice. Nevertheless, one could skip the first Sobol numbers to address the impact of the Sobol sequence starting point, e.g., taking $i_1 = n, \dots, i_n = 2n$. The graph in Figure 15 additionally plots the VIF computed by skipping the first n points over the 12 seeds; this approach is labelled the Sobol-DS-hybrid60-skip.

FIGURE 15: IMPACT OF THE SOBOL SEQUENCE STARTING POINT ON THE VIF



The table in Figure 16 shows a few additional descriptive statistics on the VIF.

FIGURE 16: VIF STATISTICS

	MEAN	STANDARD ERROR	MIN. VARIATION COMPARED TO MEAN	MAX. VARIATION COMPARED TO MEAN
MT-AV	0.02083	0.056%	-5.7%	3.9%
SOBOL-DS-HYBRID60	0.02109	0.012%	-0.9%	1.0%
SOBOL-DS-HYBRID60-SKIP	0.02111	0.014%	-0.8%	1.0%

This analysis shows that modifying the starting point of the Sobol sequence has no significant impact on the Sobol-DS-hybrid60 results, because the variability remains limited, and the average value of the VIF is not materially different (about 0.1% variation). These observations are also valid for the BEL and the leakage. There are several discussions in the scientific literature about dropping the initial Sobol points. However, there is no theoretical evidence regarding the benefits of such an approach. On this topic, a recent paper[12] even demonstrates that skipping the first point can have drawbacks for some applications.

Validation of RNG outcomes in the ESG process

As presented in this paper, hybrid RNGs can be designed as a core quasi-Monte Carlo method, augmented by Brownian bridging techniques (to control the dimension) as well as randomisation (to improve convergence properties). As such, they offer competitive RNGs to improve the quality of liability valuation estimates when the number of scenarios is fixed and limited, and to this extent outperform pseudo RNGs with antithetic variables in particular, even after—in some cases—using posteriori adjustments as moment matching. This type of new RNG for insurance applications is expected to lead the way towards an improved market practice regarding the choice of RNG, along with a better stability of the leakage over different economic conditions and stresses.

As different and even new techniques for generating random numbers become available within ESGs, a heterogeneity of practices to validate economic scenarios is observed, especially regarding martingale tests. This results in overall accuracy being different from one application to another, hence leading to a natural question about the way the leakage can be assessed at economic scenario levels in a consistent and interpretable manner.

CONFIDENCE INTERVAL VS. ESTIMATION ERROR

We can distinguish two specific concepts:

- **Model-based confidence interval** is the confidence range provided by the model governed by random sources (Brownian motions in general). By nature, the model-based confidence interval does not depend on the random number generation. The aim of the confidence interval is to validate the statistical reasonableness of the empirical mean deviations in the light of the original model. Moreover, as it relies on a closed-form confidence interval (where the variance can be estimated on the original sample), there is no computational burden related to the calculation of model-based confidence intervals.
- **Estimation error** is the uncertainty on the estimator of the empirical mean (martingale tests, repricing tests, correlation tests). Hence, estimation error typically depends on the RNG. For low-discrepancy sequences, estimation error is significantly reduced, leading to improved leakage, or to a better convergence of Monte Carlo repricing and correlation estimates. Measuring estimation error is often tackled via bootstrap techniques, which are computationally intensive. As an example, for a reasonable 95% estimation error estimate with 10,000 bootstrap simulations, and in the context where 5,000 risk-neutral paths are used for valuation, the computational cost corresponds to 5×10^7, which is equivalent to a nested stochastics application such as within economic capital models.

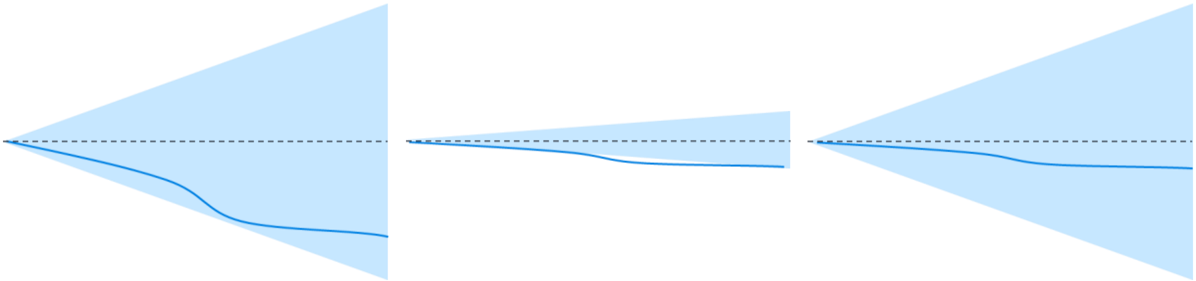
[12] Owen, A. B. (2022). On dropping the first Sobol point. In International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (pp. 71-86). Springer, Cham.

VALIDATION STRATEGY

Improvement of the RNG is by construction paired with reduction of the size of the confidence interval based on the estimation error. Thus, if insurance companies wish to measure acceptance/rejection thresholds based on estimation error, using such an approach will impose more adverse thresholds, although the Monte Carlo estimators are more accurate. For this reason, in the context of a model validation, the use of the model-based confidence interval, built from the classical variance estimator, appears appropriate.

Indeed, as illustrated in Figure 17, using estimation error-based thresholds could lead to invalidating a martingale test while the relative deviation from the theoretical value is small compared to the ordinary Monte Carlo relative deviation.

FIGURE 17: ILLUSTRATION OF VALIDATION REGARDING MARTINGALE TESTS



In Figure 17, three simplified martingale tests are represented where the blue line corresponds to the relative deviation at several time steps, the blue cone represents the associated estimation error and the dashed line corresponds to the target value. In the left graph, the martingale test passes with large relative deviations because the blue line remains in the limits of the confidence interval. In the central graph, the martingale test would not pass when the interval measuring estimation error is too narrow, while the relative deviations are much smaller than on the left graph. Hence any rejection of the scenarios set for the middle graph could be questionable, as absolute differences between the martingale outcome and the lower bound of the interval is small. In the right graph, the martingale test passes with small relative deviations, within the model-based confidence interval.

As a consequence, from the validation perspective, a key advantage of using the confidence interval from the original model is to be able to compare outcomes (e.g., from ESG martingale tests) within a universal framework. As such, it is possible to interpret validation thresholds in terms of absolute value from an original confidence interval, which has a similar meaning for all applications, therefore allowing for comparability over different use cases and companies. Also, this introduces no specific disincentive to consider a hybrid RNG with high convergence accuracy, as again the martingale thresholds are defined a priori, possibly in the light of their own impacts on the liabilities, instead of varying by type of RNG.

That being said, it appears that deriving an estimation error threshold, which could be of interest to set a priori expectations on deviations, is far from easy to implement for hybrid RNGs. Measuring accurate estimation errors for hybrid RNGs has for a long time been an intractable problem and it is only recently that some attempts were provided to measure estimation error from hybrid RNGs. As such, some studies[13] showed that even the Gaussian assumption for asymptotic deviation of the randomised quasi-RNG is not valid, advocating for adding an additional margin to the estimation error measurement. This shows the complexity in deriving such uncertainty estimates. It is expected that this will continue to be a core area of research in computational mathematics in the coming years, from which the insurance space could take advantage to further understand and stabilise valuation metrics.

[13] Tuffin, Bruno (9 May 2008). Randomization of Quasi-Monte Carlo Methods for Error Estimation: Survey and Normal Approximation. Monte Carlo Methods and Applications vol. 10, no. 3-4, 2004, pp. 617-628. Retrieved 2 December 2022 from https://doi.org/10.1515/mcma.2004.10.3-4.617.

Milliman is among the world's largest providers of actuarial, risk management, and technology solutions. Our consulting and advanced analytics capabilities encompass healthcare, property & casualty insurance, life insurance and financial services, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

milliman.com

CONTACT
chess@milliman.com

Hervé Andrès
herve.andres@milliman.com

Pierre-Edouard Arrouy
pierre-edouard.arrouy@milliman.com

Paul Bonnefoy
paul.bonnefoy@milliman.com

Alexandre Boumezoued
alexandre.boumezoued@milliman.com

© 2022 Milliman, Inc. All Rights Reserved. The materials in this document represent the opinion of the authors and are not representative of the views of Milliman, Inc. Milliman does not certify the information, nor does it guarantee the accuracy and completeness of such information. Use of such information is voluntary and should not be relied upon unless an independent review of its accuracy and completeness has been performed. Materials may not be reproduced without the express consent of Milliman.