A THOUGHT ON FERMI PROBLEMS FOR ACTUARIES

By Runhuan Feng

In physics and engineering education, Fermi problems are named after the physicist Enrico Fermi who was known for his ability to make good approximate calculations with little or no actual data, involving making justified guesses and deducing bounds without using sophisticated tools. As an illustration, we can use the following example:

How many McDonald's restaurants operate in the United States?

There are 10 McDonald's in Champaign-Urbana area, which has a population of about 200,000. ***Assume the number of McDonald's scales with population.*** Since the population of the United States is 300 million, a "back-of-the-envelope" calculation estimates the number of McDonald's at 15,000. The actual number is 14,267 as of 2014.

Using a simplifying assumption is a classic feature of Fermi problems. Although the uniform assumption italicized above is not perfectly correct, the focus of the problem is to produce good and fast approximation, when exact answers are either too time-consuming to determine or too difficult to carry out.

Why Are Fermi Problems of Relevance to Actuaries?

With the increasing complexity of equity-linked products, insurance companies are facing unprecedented exposure to financial risks in addition to traditional insurance risks. Furthermore, the financial risks embedded in equity-linked insurance are often complicated by policyholder behaviors, such as in the investment fund choice/allocation, the use of reset features, and options of withdrawal and surrender benefits, etc.

The current market practice is to develop very complex models and to rely almost exclusively on Monte Carlo simulations. As the computational burden of simulation grows exponentially, many companies simply respond by adding on more and more computing facilities. However, this is an unsustainable solution as the industry practice continues to move toward more detailed modeling and financial reporting involving more stochastic components. In the foreseeable future, the costs of such nested simulations will be prohibitive even for the most resourceful companies. The running time of simulations can still be too long for results to be delivered in a timely manner.

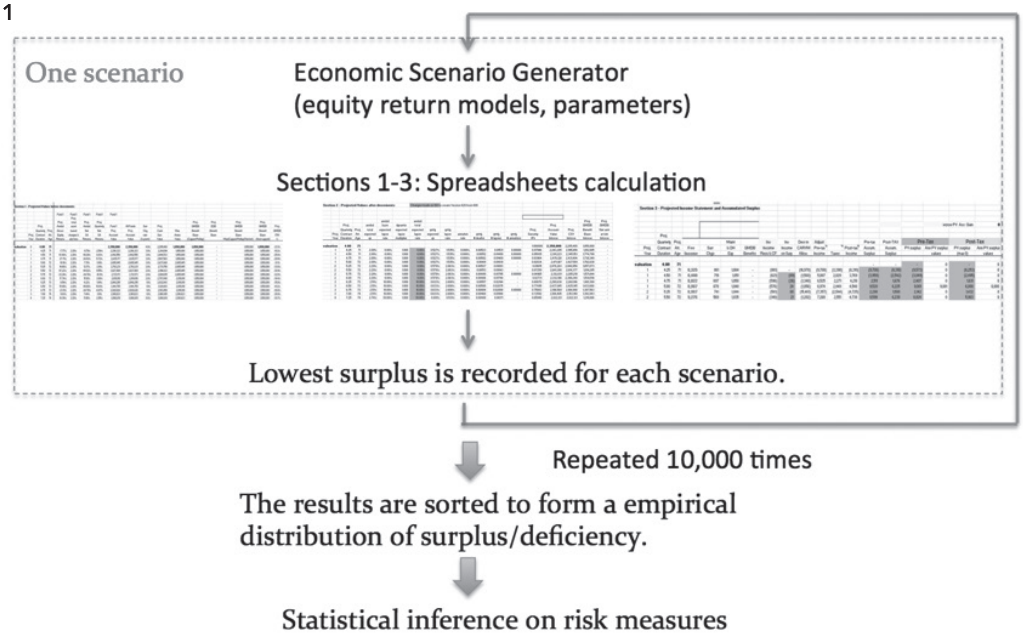
Due to the lack of alternative methods in the current practice, unscientific compromises on the scale of simulations may have to be made in order to save costs and cut running time. However, these "hash" end results may no longer be trusted or may even mislead management to wrong strategic decisions, exposing the industry to substantial model risks and systemic risks in the long term.

One potential solution to this efficiency problem is to follow the spirit of Fermi's estimate, which is to find good approximation with reasonable simplifying assumptions. For complicated problems such as the modeling of guaranteed benefits, it is unrealistic to expect a simple one-size-fits-all rule of thumb. Nevertheless, there are many computational techniques developed in the academic literature that can be used to construct a modern-day "back-of-the-envelope" calculation, which requires only modest computational efforts. We take the guaranteed minimum death benefit (GMDB) as an example to show that some deterministic techniques can be used to reach fast and efficient results.

A Simplified Example of Stochastic Reserving

While the actuarial practice on reserve calculation can be highly complex and vary company by company, we shall take a minimalistic approach here to summarize the basic principles of reserving under Actuarial Guideline (AG) 43. Policyholders make purchase payments to buy variable annuity products, which offer a selection of fund allocations at the discretion of policyholders. Then the policyholders' account values are linked to the particular equity-index/fund allocation in which they invest. The GMDB is a type of ben-

Figure 1



efit that offers a minimum account value guarantee at the time of policyholders' death. The insurer's gross liability is in essence a put option, which pays the amount by which the guarantee base exceeds the then-current account value at the time of death.

However, the rider is much more complicated than a conventional put option, as the option is funded not by an upfront fee, but rather by a stream of mortality and expenses (M&E) charges that are fixed percentages of account values deducted on a periodic basis. The financial risk is not only on the liability side from investment volatility interacting with mortality risk, but also present on the income side. For example, in adverse economic scenarios where the equity-index/fund values are persistently low over time, the insurer's liability is high, as the account value is expected to be lower than the guarantee base. Meanwhile, the problem is exacerbated by the fact that these lower account values also generate lower-than-expected fee income.

Valuation actuaries typically quantify and assess the above-mentioned losses by running spreadsheet calculations through simulated scenarios of fund performances, or using other software that performs much the same procedure. The calculations can be summarized as illustrated in figure 1.

1. A set of economic scenarios is generated to reflect a company or regulator's expectation of the variability of economic outcomes over the lifetime of the policies.
2. Under each scenario, policyholders' account values are projected according to certain accounting conventions and model assumptions. Spreadsheets are used to compute the outcomes of profitability measures such as the present value (PV) of accumulated surplus/deficits.
3. Combining all scenarios, the profitability measures are then ranked to form an empirical distribution.
4. The reserves/risk capitals are then determined by an estimation of a chosen risk metric, such as value at risk or conditional tail expectation (CTE), applied to the distribution of profitability measures.

CONTINUED ON **PAGE 24**

While the simulation procedure is easy to implement and works generally for all product designs, one should bear in mind that simulation-based techniques are statistical procedures for which estimation errors are unavoidable.

It is a well-known fact that the sampling error of Monte Carlo simulation for averages in general goes down by $1/\sqrt{n}$ as the sample size n increases. In other words, the sample size has to increase a hundredfold in order for the estimate to improve one significant digit. What makes it more challenging is that practitioners are typically interested in sensitivity measures, such as sensitivity of profitability to interest rate shocks, Greeks for hedging, etc. A small sampling error of the profitability measures can lead to a huge relative error of sensitivity measures.

Analytical Alternative Solutions

There is typically a trade-off between efficiency and generality of computational methods. In contrast with simulations, which can be used for all models despite their inefficiency at times, analytical approximations are only efficient under specific model assumptions, as less computation is achieved through careful use of analytical properties of the underlying model.

An example would be the celebrated Black-Scholes option pricing formula. It is developed under the normality assumption of log equity returns and flat term structure of interest rates, which were all proved wrong by empirical evidence. Yet it is widely accepted for pricing and hedging with some fine-tuning of the volatility parameter in the formula. Some practitioners have described this as "the wrong number in the wrong formula to get the right price."

How is one formula based on seemingly unrealistic assumptions preferred more than simulations under more complex and realistic models on interest rates, equity prices, volatilities, etc.? This goes back to the idea of Fermi's problem, which is to provide good estimates with a minimal amount of computational effort. While trying not to get into mathematical details, we can point out some quantitative tools developed in the literature that have not yet gained much attention in the practitioners' world.

Differential Equation Methods

The basic step of the above-mentioned spreadsheet calculation is to determine incremental changes in surplus for each valuation period.

Changes in surplus = Fee income + Interest on surplus – Benefit payments – Expenses

From a mathematical point of view, the spreadsheet calculations are essentially numerical algorithms based on difference equations. Each row in a spreadsheet corresponds to a recursive formula, aka a first-order difference equation. Such equations typically do have explicit solutions that can be represented in terms of the sample paths of account values.

Here is an example of the PV of accumulated deficiency for a GMDB rider, which was derived in Feng et al. (2015) based on a practitioner's spreadsheet calculation. The policy lasts for T years and fees are collected n times each year as a fixed percentage m_d of account value $F_t (0 \le t \le T)$. The initial guarantee base is G with a rollup rate δ and the interest rate is r per annum. The symbols p and q are standard actuarial notation for survival and mortality rates. In this formula for the PV of accumulated deficiency, the first term is the PV of all benefit payments (put options) over all periods up to maturity and the second term is the PV of all fee income. For simplicity, we ignore non-asset-value-based expenses.

$$L = \underbrace{\sum_{j=1}^{nT} e^{-rj/n} \, {}_{(j-1)/n}p_x \, {}_{1/n}q_{x+(j-1)/n} \, (Ge^{\delta j/n} - F_{j/n})_+}_{\text{Benefit Outgo (death benefit of put options)}} - \underbrace{\left(\frac{1}{n}\right) \sum_{j=0}^{nT-1} e^{-rj/n} \, {}_{j/n}p_x \, m_d F_{j/n}}_{\text{Fee Income (fixed percentages of acct values)}}$$

Bear in mind that the basic principle of AG43 reserving is to determine the 70%-CTE of the PV of accumulated deficiency.

$$\mathrm{CTE}_p(L) := \mathbb{E}[L|L > \mathrm{VaR}_p] \qquad \mathrm{VaR}_p(L) := \inf\{y : \mathbb{P}[L \le y] \ge p\}$$

There are several advantages of using the above explicit solution over simply running spreadsheet calculations.

(1) The cash flow structure becomes more apparent and it is easier to identify scenarios under which outgoes exceed incomes.

(2) The above solution is an additive functional of the underlying fund values $F_t (0 \le t \le T)$. The tail distribution of an additive functional can be determined or approximated by analytical methods. See below for an example of comonotonic approximations.

(3) The computational complexity of summation is typically less than that of (spreadsheet) recursion.

If we shrink the valuation period to zero, then the difference equation (corresponding to spreadsheet calculation) goes to its limit—a differential equation. In other words, it can be treated as if the fees are taken continuously and benefits are payable immediately upon death. Even though the assumptions are not realistic, the continuous-time approximation can be very close to the discrete-time true value, just as we often use continuously paying annuities to approximate monthly paying annuities for pricing annuities-certain. There are many well-established numerical methods for solving ordinary/partial differential equations, such as finite-difference, finite-elements, etc. An example of such approximations can be found in Feng (2014).

Furthermore, there are many cases of guaranteed benefits, for which analytical formulas for risk-neutral valuation and risk measures are available under certain model assumptions. For example, we recently worked out closed-form pricing formulas for guaranteed minimum withdrawal benefit (GMWB) and guaranteed lifetime withdrawal benefit (GLWB). They can be used to remove inner components of nested simulation for financial reporting on these complex riders

Comonotonic Approximations

Observe that the accumulated deficiency L is essentially determined by a weighted sum of fund values at various time points. The adverse scenarios of the accumulated deficiency are results of tail events of the weighted sum. Instead of going through all the trouble with simulating or analyzing the complex dependency among various fund values, one can study the extreme events by looking at a single random variable that characterizes the dependency.

CONTINUED ON **PAGE 26**

Here is a simplified description of the mathematics behind this. For any continuous random vector $(X_1, X_2, \cdots, X_n)$, the sum of the random vector dominates the sum of the vector conditioned on another random variable Λ in the sense of the so-called "convex order."

$$S^l := \sum_{i=1}^{n} \mathbb{E}(X_i|\Lambda) \leqslant_{cx} S := \sum_{i=1}^{n} X_i$$

The left-hand-side single random variable is known as a comonotonic bound. Ignoring the technical definition of convex order, the most important consequence of such a relation is that the CTEs of the two sums are also ordered regardless of the choice of Λ.

$$\mathrm{CTE}_p(S^l) \leq \mathrm{CTE}_p(S)$$

Note, however, that the computation of risk measures for the single random variable S^l is much easier than that of the sum of highly dependent random variables, *S*. The goal of comonotonic approximation is to find the best choice (optimization) of Λ so that the CTE of S^l is as close as possible to *S*.

Perhaps the best way to visualize a comonotonic approximation for the reserving exercise is to think of the fund values at various time points as competing horses in a chariot. Even though the horses have different speeds individually, their average performance can be characterized by the speed of the chariot. Keeping track of the speed of the chariot is much easier than averaging the individual measurements of the horses at all times. In a similar way, comonotonic bounds usually provide much more efficient solutions with only small compromises on accuracy, in the same spirit as in Fermi problems.

A recent work in Feng et al. (2015) gives several examples where comonotonic approximations can be used to estimate various risk measures of the PV of accumulated deficiencies. Here we reproduced the table for the comparison of efficiency among comonotonic approximations (labeled "optimizations") and Monte Carlo simulations (standard deviations in brackets). If we only need accuracy up to three decimal places, then the approximations (labeled "75% reduced") can be about 30 times faster than simulations based on a sample size of 1 million scenarios. To achieve accuracy up to four decimal places, the approximation (labeled "50% reduced") runs roughly 750 times faster than simulations based on the sample size of 100 million.

Method	$\mathrm{VaR}_{0.9}$	$\mathrm{CTE}_{0.9}$	Time (secs)
Nonlinear optimization	0.03035	0.06126	69.97
Nonlinear optimization (50% reduced)	0.03031	0.06123	30.50
Nonlinear optimization (75% reduced)	0.03018	0.06111	7.83
Monte Carlo (1 million)	0.03059 (0.00013)	0.06137 (0.00010)	226.16
Monte Carlo (100 millions)	0.03035 (0.00002)	0.06128 (0.00002)	22602.80

Note

Computational Risk Management Lab is a research lab, housed in the Department of Mathematics at the University of Illinois, whose mission is to develop and implement efficient computational solutions for risk management problems. We invite practitioners to make use of our state-of-art research resources and to provide us with challenging research questions. Interested practitioners should contact rfeng@illinois.edu for further information.

References

Feng, R., and H.W. Volkmer. 2012. Analytical Calculation of Risk Measures for Variable Annuity Guaranteed Benefits. Insurance: *Mathematics and Economics* 51 (3): 636-648.

Feng, R. 2014. A Comparative Study of Risk Measures for Guaranteed Minimum Maturity Benefits by a PDE Method. *North American Actuarial Journal* 18(4): 445-461.

Feng, R., X. Jing and J. Dhaene. 2015. Comonotonic Approximations of Risk Measures for Variable Annuity Guaranteed Benefits with Dynamic Policyholder Behavior. Preprint. Available on author's website.

Runhuan Feng, FSA, CERA, Ph.D., is an assistant professor of actuarial science at the University of Illinois at Urbana-Champaign. He can be reached at *rfeng@illinois.edu*.