Article from

CompAct

April 2019

Issue 59

Need for Speed: How to optimize models for maximal run efficiency

By Vincent Xuan, Housseine Essaheb and Benjamin Stirewalt

Producing a crystal-clear balance roll-forward on time, obtaining a fresh model output the first time or providing model results for your business partners with a tight turnaround all depend on one thing: how fast the model engine runs.

Why is model run time so important? First, financial reporting is usually not negotiable. Your finance colleagues need to report financial results by a specified date following the quarterly or year-end closing, and they need actuarial analytics to explain the results in a timely manner. Additionally, actuaries need to produce various analytics for senior management's internal management purposes. Capital, surplus and profitability analyses are in regular demand from risk management and treasury partners, and it is critical to be able to provide complex data quickly. Another consideration is that model development requires shorter turnaround time for faster iterations. Finally, long model run time usually means more grid core hours or more cloud computing usage, which results in more modeling and technology staff support. All these factors drive up the computation bill and directly impact the company's bottom line.

Here are some run time challenges and ways to minimize their impact:

- **The size and complexity of the business** is a major contributing factor. The larger the in-force block is, the more policies the model needs to process; therefore, longer run time is required. One potential solution is population compression. For companies with multiple lines of business, modelers need to balance the pros and cons of merging into one model. Within each line of business, companies facing various generations of products can use mapping techniques to avoid modeling each product exactly as described in the product spec, which may significantly drive up the complexity and run time. Modelers can also seek reasonable model simplifications for complex product features.
- **Multiple uses of the same model** can also cause an increase in model run time. In most cases, one model is used for multiple purposes, such as financial reporting, internal financial forecasting, capital management and risk hedging. When modeling multiple reporting bases—including statutory, U.S. GAAP, IFRS and tax—modelers should strive to centralize the common calculation segment as much as appropriate. One possibility is to combine statutory with tax calculations after the recent tax reform. Another run time multiplier is the number of the economic scenarios and assumption sensitivity runs. Model users should be encouraged to trim down the number of scenarios and assumption shocks, especially for ones with muted impact.
- **Model structure inefficiencies** should be regularly examined. The same logic should be programmed and run once instead of multiple times. For example, the liability cash flow generation segment could be calculated once and then shared across different bases for further calculations instead of calculating multiple times for the same outputs. Modelers should consider periodical peer reviews and seek advice from the vendor system on code efficiency. Sometimes an overall run time diagnosis can reveal some unknown run power consumptions.
- **Infrastructure automation and process control** should be considered along with the calculation engine optimization. When redundant manual interventions are involved, it is hard to increase the end-to-end process speed. Try to find ways to eliminate manual feed or handover and instead automate the process. For example, instead of setting up the models manually for different runs, using a batching tool or robotic technology to automate the model runs is highly preferable. Another infrastructure consideration is to optimize parallel run capabilities to improve grid or cloud efficiency. Managing the process control will also help minimize risk instead of creating excessive approval stops.

POPULATION COMPRESSION

There are several approaches for population compression, including randomized selection, clustering and model point creation. These techniques may be used for analytics and model development and testing, even if not for financial reporting.

1. **Random selection:** In this approach, a random subset of the full seriatim in force is selected, and then the calculated results are scaled up. The selection can be randomized in several ways. The simplest way is to sort all the records, for example, by contract number and then select every Kth record. Alternatively, each record can be assigned a random or pseudo-random number between zero and one, and

all records whose numbers are less than K percent can be selected. In either case, results are uniformly scaled up for each record by the compression factor K.

Because the algorithm is rather straightforward, this method is generally easy to implement, in both ad hoc and production settings. The compression ratio is easily controlled by setting the value of the factor K.

This approach is best suited for aggregate results across larger in-force blocks that can take advantage of the law of large numbers. The results may not converge as well for medium-sized blocks or more granular reporting metrics.

2. **Clustering:** In this approach, the contracts are grouped together into clusters based on similar characteristics, such as product type, issue age, gender and moneyness/net amount at risk. A sample is then formed by choosing the best representatives from each cluster. Finally, the results are scaled for each representative based on a measure of the "size" of its parent cluster, which is usually the total account value or benefit face amount.

   Given the complexity of the algorithm, this method can be more complicated to implement in practice. The criteria that define the clusters must be determined, and this generally requires testing several iterations until all criteria are fully specified. As the in-force changes over time, the criteria would also have to be monitored periodically. It can also be harder to achieve a specific compression ratio, as the size of the subset is a function of how strictly one sets the clustering criteria. Several iterations should be tested in order to achieve a desired target ratio. Finally, additional infrastructure components would generally be required to employ this approach in a production setting.

   However, for medium in-force blocks or more granular metrics, the results should converge.

3. **Model points:** This approach begins like the clustering approach, except that once the policy clusters are formed, rather than selecting representatives, all the contracts in each cluster are combined into a single model point and treated as an actual contract. This can be accomplished by adding the seriatim values together within each cluster, and results would not need to be scaled up.

   This method shares the complications of the clustering method previously mentioned. Additionally, it may be harder to trace the integrity of the values comprising the model points.

   As with clustering, this approach should converge even for medium-sized in-force blocks.

ONE VERSUS MULTIPLE MODELS

To consolidate or not to consolidate? This is the question every modeler should ask. Actuarial models are an integral component in the actuarial profession, as they are heavily relied upon for all actuarial work. Actuaries use models for pricing new products, satisfying regulatory requirements for financial reporting and supporting management decisions. The decision whether to use a consolidated model—such as implementing new products, implementing new regulatory requirements or adding new projections capabilities—will depend on its use and actuarial judgment.

> Producing a crystal-clear balance roll-forward on time … or providing model results with a tight turnaround … depends on how fast the model engine runs.

The benefits of having a consolidated model include:

- End-to-end IT infrastructure: A consolidated model will leverage an already fully integrated IT infrastructure, allowing for a shared set of existing input and output facilities. This also allows for centralized aggregation, which uses a common data warehouse to analyze across products and different purposes (e.g., valuation and forecast).

- Implementation efficiency: With a consolidated model, regulatory changes, new reinsurance treaty arrangements or assumption updates need be performed only once rather than duplicating effort into separate models.

- Risk management: Existing modeling controls can be leveraged for additional model implementations rather than creating separate controls specific to different models. For example, a shared business requirement can be used in a consolidated model.

- Cost optimization: Whether the company uses in-house models or vendor software, building upon an existing model reduces costs associated with training, infrastructure support, existing system retrofits and potential fees associated with new vendor models.

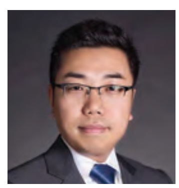
Two examples most actuaries may be familiar with are new product implementation and developing forecasting capabilities.

When a new product line is launched, does a separate standalone model need to be created, or should it be consolidated into a main model? How would one go about consolidating? Is it as simple as mapping to a prior product with some tweaks, or should it be built from the bottom up? These are all questions to consider.

For example, if a company decides to roll out a new enhanced death benefit (DB) rider onto a base variable annuity product, there is an opportunity to leverage the original DB rider. Suppose the original DB rider returns the initial deposit to the beneficiary when the insured dies. This type of rider protects in situations where the market depreciates prior to the annuitant's death, resulting in current account value to be lower than the promised DB. The enhanced DB rider resets the death benefit above and beyond the initial deposit, which resets periodically at the highest account value over a certain contractual duration. Given the similarities in the calculation to the original death benefit, the modeler needs to modify only the existing code to accept inputs related to the enhanced death benefit, such as the highest account value, rather than the initial deposit and fees associated with the enhanced DB.

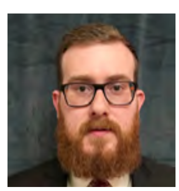
When building a projection model for forecasting capabilities, consider the benefits of consolidating by building off the valuation model versus creating and maintaining a separate model. One key benefit of a consolidated model is the ability to share the same methodology so that the forecast model is always in sync with the valuation financial reporting model. This also enables more sophistication than stand-alone forecast models, which tend to be less complex and use simplification techniques that can potentially cause mismatches between actual and forecast results. The consolidated model will ease the attributions by eliminating model differences.

Renowned computer scientist Dr. Donald Knuth once said, "Premature optimization is the root of all evil." Model consolidation and population compression should be part of all optimization discussions, but before blind pursuit, the pros and cons should be laid out clearly for model users to ensure understanding of limitations and to evaluate the options. ■



Vincent Xuan, FSA, CFA, MAAA, is a vice president and actuary with Prudential Financial. He can be reached at *xu.xuan@prudential.com*.

Housseine Essaheb, FSA, CERA, MAAA, is a director and actuary with Prudential Financial. He can be reached at *housseine.essaheb@prudential.com*.



Benjamin Stirewalt is a director and actuary with Prudential Financial. He can be reached at *benjamin.stirewalt@prudential.com*.