Validating Simulation Models using Hadoop

Jim Rogers, COO, TFG Financial Systems.

15th July 2015

Abstract

A standard technique used in both market and credit risk management for determining the risk of a portfolio of complex financial transactions is to value those transactions on a large number of possible market scenarios to see how changes in market prices could affect the value of the portfolio. A simulation model is used to produce a large number of possible future scenarios that represent how the markets could evolve. Validating these scenarios, given the large amounts of data involved, has typically been a cost and computational challenge. However, technologies such as Hadoop allow this to be accomplished at a much lower cost.

Introduction

A standard technique for measuring the risk in a financial portfolio involves the generation of scenarios that represent the possible future levels of the numerous market rates and prices that impact a portfolio’s value, valuing the portfolio on each possible scenario, then comparing the possible value to the current value. For statistical accuracy a large number of scenarios need to be considered and the scenarios need to accurately represent what the future might look like. To validate the simulation model actual market rates are captured for a period of time such as a year and compared to what was simulated. The majority of actual observations should fall within the simulated range, based on the desired confidence interval. If too many are outside the simulated values the model is underestimating risk, and if too few are outside the simulated values the model is overstating risk. How do you perform this validation before using the model, and without waiting for a long period of time? In the past firms have done “spot checks” on limited samples, waited the period of time, then compared actual rates to predicted. If the results were not within tolerance the model was tweaked and the process repeated. The problem with this approach is you never knew if your simulation model is working until you use it and by then it was too late. An alternative approach would be to run the simulation model repeatedly over a prior period as it would have run, capture the results, and compare them to actually observed rates. In this way models could be changed and re-validated with confidence quickly. The problem with this approach is that the amount of data that needs to be produced,
stored, and analyzed can easily become very large making the solution a technical challenge and cost prohibitive.

**Hadoop**

The problem at its simplest is producing, storing, and analyzing a large dataset at low cost. The data is test data and can be reproduced, so data integrity requirements can be relaxed a little. The problem is not real-time or transactional, but more of a batch nature, and is not typically done regularly, but when models are changed. This type of problem is ideally suited to Hadoop. The low cost disk based storage of Hadoop can store large amounts of data cheaply and the grid of compute engines make analyzing this data easy.

Using Hadoop, either in the cloud or on an internal grid, enables firms to validate models quickly, at lower costs, than was previously possible. The result should be an improvement in the thoroughness of testing financial simulation models both by financial firms and vendors. Cost is no longer be an issue.