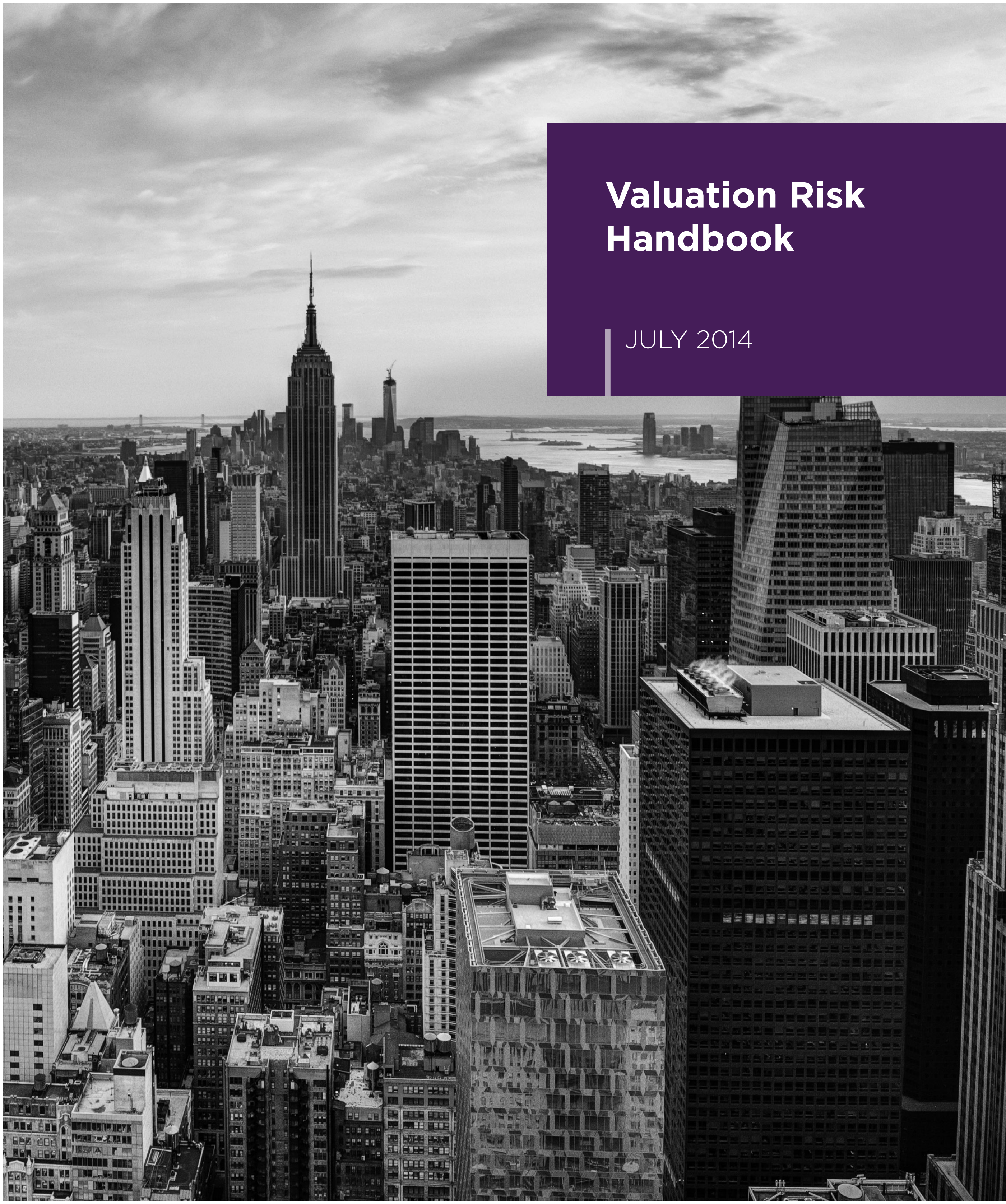
Voltaire ADVISORS



Valuation Risk Handbook

JULY 2014

Model Validation: New Approaches in Testing Mathematical and Financial Correctness of Models

David Eliezer, PhD, Vice President, Head of Model Validation, Numerix LLC (June 2014)

Driven both by regulators and internal pressure to avoid losses due to poor modeling, the validation of derivative pricing has received a burst of renewed interest in recent years. An onslaught of new regulations and revived attention to this field have lit the spark for exploration into best practices for model validation - including enlightened new approaches in testing mathematical and financial correctness of models and using automated model tests to improve derivative pricing models.

Basel 3, Dodd-Frank, Solvency II, and of course, the Supervisory Guidance on Model Risk Management (OCC Bulletin 2011-12) all continue to force a major shift in model risk governance and current model validation best practices. This has led us to explore new model validation approaches, including testing with mathematical identities.

NEW TESTING APPROACHES, COMPUTE POWER AND SATISFYING REGULATORS

We all know that the validation of derivative pricing models can be a slow, labor intensive and expensive exercise. In addition, it often provides a limited amount of certainty on the correctness of the pricing models, because methods for validating pricing models are often generic methods that can be applied to any kind of software. However, since pricing models are mathematical models, they satisfy mathematical identities which can provide strong tests that leave very little possibility for error. Furthermore, these tests provide failure conditions that require no human judgment, that can be automated, and that can therefore run over tens of thousands of test scenarios.

Significant advances in computing power such as the cloud and grid computing now make automated testing more accessible to market practitioners—making the testing inexpensive, rapid, and easy. (Testing of this kind is completely and straightforwardly parallelizable, and small-scale experiments have shown speedups as large as a factor of 12). Automated, parallelized testing results in a faster time-to-market for validating models, and a stronger case can be made to regulators that models are thoroughly tested and correct. In this article, we will examine model validation as it is typically practiced today and then explore new approaches, including the benefits of testing with mathematical identities.

MODEL VALIDATION AS IT IS PRACTICED TODAY

How can we begin to seriously mitigate the risks due to model failure? A model's dynamics are implemented by a fairly large set of complicated code. How can we be certain that such a mass of computation is free of bugs? Model validation, when implemented correctly, is one form of protection.

Parallel Implementation

The current standard for validation of stochastic dynamics is the method of parallel reimplementation—in which the model is validated by having another quant independently implement the model from the model documentation alone. This alone doubles the cost of the model development. Then, the two models are reconciled in a point-by-point fashion, and through this fairly laborious process, the bugs are hopefully worked out. Model development is expensive to begin with, and validating that model by parallel reimplementation makes it far more expensive.

Parallel reimplementation relies on the improbability of two quants not making the same errors. This works well most of the time, but there are occasionally misconceptions that are common to many quants, and errors of this kind can get through this test. One example of this is the reflecting boundary condition that is necessary in CIR-type processes: It was very commonly omitted in implementations, until a flurry of papers in the literature cleared up this point.

Convergence and Stress Tests

In addition to this, sanity checks, such as convergence tests and stress tests can also be performed. Stress tests involve taking market inputs, payoff parameters, or model parameters to extreme values, and looking for unreasonable behavior, such as singularities, or violations of arbitrage conditions. This is designed to detect problems of implementation that don't reveal themselves under ordinary conditions, but only when conditions are extreme (as in a crisis, when debugging suspicious behavior in pricing functions falls to the bottom of the priority list.) Stress tests like this require human judgment, and cannot be automated, because most of the failure conditions are not easily definable.

Convergence testing involves simply evaluating the pricing function at various values of its approximation parameters (e.g. number of paths, number of timesteps) in the hope of observing either visual or analytical evidence suggesting that the function is settling down to an asymptote as the approximation parameter approaches its limit. A proof of convergence, in the sense understood by mathematicians, is not feasible in a numerical setting—but, empirical evidence suggesting convergence is adequate for the purposes of finance.

TESTING WITH MATHEMATICAL IDENTITIES

Derivative pricing is a mathematical process, and it satisfies mathematical identities. These identities are what guarantee the model is what it is. By checking them, we can actually come up with tests that make use of the certainty that theorems provide. They aren't full mathematical proofs, but they come as close as we can get with numerical experiments. Identities are far easier to check, and can be checked far more thoroughly. The conditions, being mathematical, require no judgment, and can be automated.

This approach has the advantage that it is not labor-intensive, and does not rely on the unlikelihood of two quants making the same errors. It uses computer time instead of human time, and so is less expensive. What this means is that we can check this condition not for just a few, typical values, but thoroughly, over a large testing domain, by choosing thousands of examples from within the domain, on which to run the test. The clear-cut nature of the failure conditions means that a human need not examine each such example to decide with his judgment whether the test passes.

Partial Differential Equation (PDE) Test

The simplest check one can make is to use the uniqueness theorem for equations of heat type, which states that the function that satisfies the boundary conditions given by the payoff and the PDE of the model is unique: it is the model's price for that payoff. This means that one can simply check the model's payoff (i.e. this can be done by setting the volatility to zero, and adjusting the model's drift parameters until every desired payoff scenario is confirmed to yield the quantity specified in the payoff.) The only requirement then, is to check that the function satisfies the PDE of the model it is using.

How can this be done? It is tricky, because a pricing function is an approximation, and is not expected to satisfy the PDE exactly. We cannot evaluate the exact PDE operator on the solution anyway, in a numerical setting. The key is to have not just a single function, but a family of functions, indexed by a single approximation parameter, along which we may take limits. It is straightforward to show rigorously that the limit of the solutions does satisfy the exact, limiting PDE, if an approximate PDE operator acting on an approximate function vanishes in the simultaneous limit of both approximation parameters reaching their limits—along with a few other convergence conditions.

Of course, when probing a pricing function numerically, mathematical proofs of convergence are not really possible. Instead, one can amass evidence for the convergence conditions that the testing theorem above requires, in the same way that one does for ordinary convergence tests. Note that testing throughout the domain is a necessity for the PDE test to conclude that the pricing function is valid, otherwise the pricing function, even at tested points, might be very different from the true solution to the PDE.

Calibration Round Trip Test

The PDE test does not address the selection of model parameters, by the process of calibration. Calibration is a difficult problem, famously so, and is often ill-defined. In fact, most models in use today encapsulate dynamics represented by just a few-parameter-family of volatility surfaces with which to fit the full universe of possible option prices that we see in the market. When the market's volatility surface is very similar to the model's, then the solution to the calibration problem is likely to be meaningful, if not unique. But if the market volatility surface is very different from any of the model's volatility surfaces, a least squares solution defining the model's parameters won't be very meaningful, because the calibrated volatility surface won't look anything like the market volatility surface. The result will depend sensitively on details of the definition of the problem, such as the choice of objective function; and, in particular it is very difficult to know whether the solution that the calibrator has found is near to the optimal solution. Furthermore, it is irrelevant, as such a solution is unlikely to lead to conspicuously better hedging.

Therefore, the calibration problem is really only financially meaningful when the market volatility surfaces are "close" in some sense to the model's volatility surfaces. In this situation, the sensitivity to the problem setup is greatly reduced, and the problem becomes very well defined—precisely when the market data comes from the model, although not with a unique solution. In this case, there is an exact solution, and the optimal solution is one in which the error in all option prices is zero, although there may be more than one solution. We may decide very simply whether the calibration problem has been solved - the calibration error should be within the calibrator tolerance.

Thus, one may construct a strong test for the calibrator by picking some model parameters for our model, and some calibration instruments (i.e. options), with their strikes and maturities, and we price the options in the model with its given parameters. This will provide us with a volatility surface that is exactly compatible with the model we are testing. Calibrating another instance of the same model to these options should result in zero calibration error, and if it fails, the calibrator did not find the optimal solution.

As in the PDE test, the availability of a well-defined failure condition means that one may run this test thousands of times, with different input parameters. This allows the tester to discover regions where the calibrator performs well, and where it performs poorly.

Even with the narrowed scope of this test, calibration is still a very difficult problem to solve. It generally requires some analytic work to find smart initial guesses, as purely numerical solvers rarely can find the true global minimum in such a complex problem as this one. Finding a calibrator that can always pass even this narrow test may not always be possible. In this case, calibrators should be compared by the percentage of cases that pass, and by how much the calibrator failed.

Financial Correctness and Hedge Performance

Is the model suitable to the current market? The previous tests were designed to test the mathematical conditions defining the pricing equation for the model. But there is a deeper mathematical condition that can be tested against the market data, forming a test that evaluates how similar are the model's dynamics—including both marginal and conditional distributions—to those realized in the market data.

Recall that the defining condition of the Black-Scholes Merton pricing framework in any framework is the minimization or elimination (when markets are complete) of the variance in the portfolio consisting of the derivative together with its hedge. In practice, even models with complete hedges will not actually experience zero variance in real life because the model is not perfect. The variance that is experienced, when compared with that of the portfolio without the hedge, is an excellent measure of the quality of the model, and of its model risk.

The hedge ratios can be thought of as regression coefficients, and so the model's quality can be measured exactly as we do that of a regression, by the fraction of variance that the model's hedge removes from the instrument's performance, (i.e. by R^2.)

How can this be done? One takes a path of historical data, and just as is done in real life, calibrates and prices the model every day, and computes the hedge ratios. With this data, one may simulate actual trading and hedging of the instrument, keeping track of the PnL attribution, and hedge bleed, etc., just as traders do in real life. The change in the instrument's price, set against the hedge's change in value, is the hedging error.

This error is not supposed to be zero. In fact, it should be the cost of hedging the upfront price. But, it is supposed to be non-random, even though derived from random quantities (the remarkable fact of the Black-Scholes-Merton arbitrage-free pricing.) How can this be checked?

The difficulty is that we have only one true past history, and measurements of hedging error randomness (i.e. by variance or some other measure) requires many historical paths. This problem is addressed in other context (i.e. historical VaR) by constructing pseudo-historical paths. This is done by a variety of methods, but one common method is to compute all the daily increments of asset prices, or yields or implied volatilities, and store them as a probability distribution of asset price/yield/volatility movements. Then, paths can be constructed by starting with an initial set of market data, and choosing a succession of increments to add on to the initial market data. Within this framework, there are many different ways to compute the increments, but it is straightforward to find methods that preserve all basic qualitative behavior—ensuring that overall volatility, skews, kurtosis, yields and asset prices all go up and down when they are supposed to.

In this way, one may construct the paths necessary to compute the variance of the hedged portfolio. Additionally, we may regress the hedge error back on the hedge instruments. The hedge is supposed to be optimal, so the regression coefficients should come out close to zero. Their nonzero value serves as another measure of the imperfection of the model, and in fact guides as to how to improve the model. If the regression coefficients are significantly positive, then the hedge is an under-hedge—if negative, it is an over-hedge. The level of significance can be estimated by the increase in R^2, due to replacing the hedge ratios with 'optimal' ones according to this regression. If the improved hedge ratios lead to a significantly greater R^2, then it is clear that it is worth changing the hedge to improve the model. This may even be done with new, proposed hedge instruments.

The variety of incrementization methods means that this test should properly be done with several such methods, to ensure that a consistent message emerges. If consistency is found, however, the test may be used as an effective measure of the fidelity of the model to the true market probabilities.

MODEL VALIDATION WRAP-UP

Great improvements to our process and to our industry are possible through a smart approach to model validation. Increasing regulations and significant losses due to model failures have beefed-up interest in the field of model validation, paving the way for new approaches. Mathematical identities of models can be used to provide a strong and automatable testbed for pricing model implementations. Backtesting of pricing models can similarly be strengthened by testing the mathematical properties the model is designed to have, the optimal reduction of variance. This kind of backtesting provides tremendous opportunities for deeper analysis to find the causes of model breakdown, and the path to model improvement.

AUTHOR BIOGRAPHY

David Eliezer, PhD, Vice President, Head of Model Validation, Numerix LLC

Mr. Eliezer has been a quant on Wall Street for 18 years, at Goldman-Sachs, Morgan Stanley, General Re Financial Products, and Bloomberg, among others. He has published work on option pricing, and on modeling liquidity in finance. He runs the internal testbed for Numerix models, and he leads the Model Validation group at Numerix.

1 References: Board of Governors of the Federal Reserve System, Office of the Comptroller of the Currency, Supervisory Guidance on Model Risk Management (OCC Bulletin 2011-12) dated April 4, 2011