Validating Stress-Testing Models
Econometricians need to build models for forecasting or prediction, structural analysis or hypothesis testing, and policy or shock analysis. Each of these applications has an underlying loss or risk function that governs how the model should be built and the properties the preferred model specification should retain. For forecasting, the loss is largely determined by out-of-sample or out-of-time prediction errors, while for hypothesis testing, loss is associated with a test size that closely approximates the stated significance level while maximizing power. In each case, very different models may be found to be optimal, even if the underlying dependent variable is the same in each case. Here we consider stress-testing, which we view as somewhat distinct from other types of econometric analysis.

We define “stress-testing” using the framework applied by the U.S. Federal Reserve in the recent Comprehensive Capital Analysis and Review (CCAR) of large bank holding companies. That is, we consider the case where a specific economic scenario (or scenarios) is provided and stress predictions subsequently computed. We do not assess, like Breuer and Csiszar (2010), whether this is a worthwhile exercise in and of itself. We instead concern ourselves with the question of model validation methodology when the stress-testing protocol is dictated to the modeler in this manner.

In most ways, stress-testing mimics forecasting in the sense that we are trying to use a model to predict the future behavior of the metric of interest. The difference is that, with stress-testing, we are not concerned about baseline behavior directly. We care instead about whether, conditional on the occurrence of the stress event, the model can predict the dependent variable of interest with low error. For the purposes of benchmarking the stress test against baseline conditions, we also care that the baseline projection is accurately determined.

For a dependent variable of interest, $y_t$, assuming squared error loss, the optimal baseline forecasting model will be that which minimizes

$$E(y_{t+i} - \hat{y}_{t+i})^2$$  \hspace{1cm} (1)

where $\hat{y}_{t+i}$ is the predicted value of $y_{t+i}$ at time $t$ and where $i$ is the stated forecast horizon. This risk function can easily be extended to cover multiple forecast horizons if the entire path of the dependent variable is relevant for the problem at hand. In the case of the stress-testing model, this risk function changes subtly, becoming

$$E_\Theta(y_{t+i} - \bar{y}_{t+i})^2$$  \hspace{1cm} (2)

where $\bar{y}_{t+i}$ represents the prediction of the dependent variable at time $t$ under the assumption of a specific future stress event, $\Theta$, and $E_\Theta(\cdot)$ is the expectation conditional on the actual occurrence of the event. We define the optimal baseline forecasting model as that which minimizes (1) and the optimal stress testing model as that which minimizes (2).

Needless to say, different models will likely prove optimal in each case. For example, an optimal baseline forecasting model may, in some situations, turn out to be a univariate AR($p$) specification. It is, however, very difficult to imagine how such a model could be used to construct predictions conditional on a specific stress event, $\Theta$. If the stress event is defined using economic drivers, a logical stress-testing model would involve a regression of the dependent variable of interest on some or all of these drivers. In this simple situation, the optimal baseline prediction model cannot yield stressed predictions and thus cannot even reach the starting line for consideration as a stress-testing model. One possible path forward would then be to use different models to construct the baseline and stressed forecasts. Indeed, ignoring the cost of building and validating models, the optimal approach would be to use different models for baseline and stressed prediction.

In other situations, like the CCAR exercise, the regulator specifies both a baseline and a stressed set of assumptions. In this case, even the baseline model’s risk function changes, becoming

$$E_\Psi(y_{t+i} - \bar{y}_{t+i})^2$$  \hspace{1cm} (3)
where $\Psi$ is the event where the Fed’s baseline scenario actually occurs and where $\tilde{y}_{t+i}$ is the model’s prediction at time $t$ assuming $\Psi$. In this case, the AR($p$) specification based on optimizing risk function (1) is not a feasible approach because the Fed’s baseline prediction may not coincide with the mean of the distribution of possible future economic paths. We would have no way of saying whether the AR($p$) model accurately reflects the Fed’s baseline scenario. The model that minimizes (2), however, could be used to generate predictions consistent with the Fed’s baseline data, though it is doubtful whether it would be optimal given risk function (3). The best approach would still involve producing different specifications to capture baseline and stress conditions.

We now turn to the case where building models is not cost-free, implying that, all else equal, one model is preferable to two. Here, we restrict our attention to models that are capable of producing forecasts under both (2) and (3). We require two objectives to be met, notably that accurate stress and baseline forecasts be produced, but we restrict our attention to cases where one model is used to construct such forecasts. The risk function now becomes

$$\lambda [E_\Theta (y_{t+i} - \tilde{y}_{t+i})^2] + (1 - \lambda) [E_\Psi (y_{t+i} - \tilde{y}_{t+i})^2]$$

(4)

where $0 < \lambda < 1$ is a weight that describes the relative importance of the stressed prediction in the model. We recognize that $\lambda = 0.5$ presents a useful starting point, though we suspect that a higher weight would probably be appropriate given the relative difficulty and importance of determining the effect of stress conditions.

Validation

In practical backtesting applications, identifying the best model under risk function (1) is straightforward. Given sufficient history, we simply retain a holdout sample, estimate various competing models under the restricted sample, and choose the one that predicts best through the holdout period. Validation under risk functions (2) and (3) is much more problematic. This is especially the case for risk function (2), where stress circumstances occur. In both of these cases, a specific set of projections needs to be assumed for the economic data and all projections need to be assessed, as closely as possible, conditional on these economic projections actually unfolding. In the case where we strictly apply risk function (2), for example, the performance of the model during historical episodes of economic health and prosperity are entirely irrelevant. We are interested only in past performance under macroeconomic stress.

Fortunately, we are in a position, for many banking portfolios, of having a recent, clearly defined period of stress over which we can validate stress-testing models. Consider U.S. mortgages. If our risk function is (2), and the model closely predicts outcomes for the period between the collapse of Lehman Brothers in September 2008 and the end of 2009, using a starting point around June 2008, we can be fairly confident that our model can correctly represent stress events. In different industries we may be able to clearly identify other periods of undue stress; in yet other industries, no such past stress events may be identifiable. If stress events cannot be identified, developing a model that reflects baseline type conditions accurately may be the best we can hope for. In applying risk function (3), this is closer to the simple case of (1) though the modeler needs to be cognizant of responsiveness to economic data in building the model.

Under the composite risk function (4), the modeler needs to be able to reflect behavior under stressed circumstances and then be able to relate performance under stress to performance under standard baseline conditions. We perceive two distinct ways of approaching this in practice. One approach involves using two distinct holdout samples and fitting a model that minimizes prediction errors through both periods. In the case of a mortgage portfolio with a long historical record, we imagine aggregating the post-Lehman prediction errors with the performance of the model during 2004 and early 2005. Such a period occurred in the aftermath of the 2001 recession and jobless recovery and before the housing market boomed—conditions that are akin to those assumed under the CCAR baseline.

The second approach uses a single holdout sample but takes advantage of regional fluctuations. As is well known, mortgage performance in California and Florida in late 2008 and early 2009 was much worse than similar portfolios of Texan loans. If the model is able to accurately capture the behavior of all three regions, we may tentatively conclude that it is capable of producing sound predictions under stress and baseline conditions.

Conclusion

As McCloskey (1985) pointed out in a seminal paper, it is fundamentally important that the loss function not be mislaid when conducting statistical exercises. Too often, we see models used for portfolio forecasting or stress-testing being validated in ways that are unrelated to their use. In building models for stress-testing, validation against standard forecasting criteria is not appropriate. Instead, validation criteria must be established that make it likely that the model will perform correctly should stressed circumstances actually occur. We recognize that building models is a costly exercise, but we suggest that multiple approaches be employed when comparing stressed outcomes with baseline projections.
References

or http://dx.doi.org/10.2139/ssrn.1328022

About the Authors

Anthony Hughes

Tony Hughes is senior director of Credit Analytics at Moody’s Analytics, where he manages the company’s credit analysis consulting projects for global lending institutions. An expert applied econometrician, Dr. Hughes also oversees the Moody’s CreditCycle and manages CreditForecast.com. His varied research interests have lately focused on problems associated with loss forecasting and stress-testing credit portfolios.

Now based in the U.S., Dr. Hughes previously headed the Moody’s Analytics Sydney office, where he was editor of the Asia-Pacific edition of the Dismal Scientist web site and was the company’s lead economist in the region. He retains a keen interest in emerging markets and in Asia-Pacific economies.

A former academic, Dr. Hughes held positions at the University of Adelaide, the University of New South Wales, and Vanderbilt University and has published a number of articles in leading statistics and economics journals. He received his PhD in econometrics from Monash University in Melbourne, Australia.
About Moody's Analytics
Economic & Consumer Credit Analytics

Moody's Analytics helps capital markets and credit risk management professionals worldwide respond to an evolving marketplace with confidence. Through its team of economists, Moody's Analytics is a leading independent provider of data, analysis, modeling and forecasts on national and regional economies, financial markets, and credit risk.

Moody's Analytics tracks and analyzes trends in consumer credit and spending, output and income, mortgage activity, population, central bank behavior, and prices. Our customized models, concise and timely reports, and one of the largest assembled financial, economic and demographic databases support firms and policymakers in strategic planning, product and sales forecasting, credit risk and sensitivity management, and investment research. Our customers include multinational corporations, governments at all levels, central banks and financial regulators, retailers, mutual funds, financial institutions, utilities, residential and commercial real estate firms, insurance companies, and professional investors.

Our web and print periodicals and special publications cover every U.S. state and metropolitan area; countries throughout Europe, Asia and the Americas; and the world’s major cities, plus the U.S. housing market and other industries. From our offices in the U.S., the United Kingdom, and Australia, we provide up-to-the-minute reporting and analysis on the world’s major economies.

Moody's Analytics added Economy.com to its portfolio in 2005. Its economics and consumer credit analytics arm is based in West Chester PA, a suburb of Philadelphia, with offices in London and Sydney. More information is available at www.economy.com.