A Coherent Framework for Stress-Testing

Jeremy Berkowitz
Federal Reserve Board

March 20, 1999
This Draft: July 14, 1999

Address correspondence to:

Jeremy Berkowitz
Trading Risk Analysis
Mail Stop 91
Federal Reserve Board
20th and C Streets, N.W.
Washington, D.C.
jberkowitz@frb.gov

Abstract: In recent months and years both practitioners and regulators have embraced the idea of supplementing VaR estimates with "stress-testing". Risk managers are beginning to place an emphasis and expend resources on developing more and better stress-tests. In the present paper, we hold the standard approach to stress-testing up to a critical light. The current practice is to stress-test *outside* the basic risk model. Such an approach yields two sets of forecasts -- one from the stress-tests and one from the basic model. The stress scenarios, conducted outside the model, are never explicitly assigned probabilities. As such, there is no guidance as to the importance or relevance of the results of stress-tests. Moreover, how to combine the two forecasts into a usable risk metric is not known. Instead, we suggest folding the stress-tests *into* the risk model, thereby requiring all scenarios to be assigned probabilities.

Acknowledgements: I gratefully acknowledge helpful input from Jim O'Brien, Matt Pritsker, Pat Parkinson and Pat White. Any remaining errors and inaccuracies are mine. The opinions expressed do not necessarily represent those of the Federal Reserve Board or its staff.

1. Introduction

In recent months and years both practitioners and regulators have embraced the idea of supplementing VaR estimates with "stress-testing". Risk managers are beginning to place an emphasis and expend resources on developing more and better stress-tests. In this paper, we hold stress-testing and its implications up to a critical light.

The first question that arises in this context is what exactly is the definition of stress-testing? As it stands, the definition is surprisingly vague. To echo Supreme Court Justice Potter Stewart, the professional consensus might be described as, "we can't define it, but we know it when we see it." For example, the 1999 BIS document *Framework for Supervising Information about Derivatives and Trading Activities* says that stress scenarios need to cover "a range of factors that can create extraordinary losses or gains in trading portfolios" and that they should "provide insights into the impact of such event on positions."

The original Market Risk Amendment to the Accord contained two full pages on the importance of stress-testing.[1] Along with the definition given in the 1999 BIS document, the Amendment (p. 46) recommended quantitative and qualitative criteria to "identify plausible stress scenarios to which banks could be exposed." Such criteria should "emphasize that two major goals of stress testing are to evaluate the capacity of the bank's capital to absorb potential large losses and to identify steps the bank can take to reduce its risk and conserve capital." Clearly such exhortations are not specific enough to result in operational stress-testing procedures.

Nor has the private sector gone further in formulating a palpable definition of stress-testing. Anecdotal evidence suggests that risk managers pursue stress-testing as a way of understanding gamma-risk and as a tool for studying portfolio allocation (Kupiec (1999)). Large and complex portfolios containing assets with nonlinear payoffs, such as options, may behave very differently in response to large shocks than would be expected given its valuation in more typical situations. If gamma risk is excessive or allocated in an undesirable way across assets, the portfolio can be modified in response.

Though we lack a formal definition then, it is not really the case that the profession

---

[1]See pages 46-47 of the *Amendment to the Capital Accord to Incorporate Market Risks* (1996), Bank for International Settlements.

disagrees about what constitutes stress-testing. It is understood that stress-testing means choosing scenarios that are costly and rare, and then putting them to a valuation model. The problem of course is that choosing stress-test scenarios is by its very nature subjective. This makes external review of a bank's stress testing program, for example by regulators, extremely difficult.

Regulators may be able to identify "bad" stress-tests. For example, scenarios that feature one or two standard deviation shocks may be ruled insufficiently stressful on theoretical grounds. However, the opposite is not true. Certifying a stress-test as sound would require the two parties to agree on the assignment of probabilities to unusual (or unheard of) events.

We propose a unified framework in which stress-test scenarios are incorporated into the basic risk model used by the firm. The stress-test scenarios are necessarily assigned (possibly subjective) probabilities, which imposes additional discipline on risk managers. In this way, the model produces a single forecasted loss distribution which is internally consistent and amenable to backtesting.

The remainder of the paper is organized as follows. Section 2 proposes a formal definition of stress-testing couched within the framework of an internal risk model. Section 3 studies the theoretical implications of our definition given present practices. In section 4, we propose an alternative approach to stress-testing that coheres with the basic internal modeling strategy. Section 5 concludes.

2. A Formal Definition

In this section, we propose a formal definition of stress-testing. Suppose that the firm has a risk model which is used to forecast the distribution of possible returns, y_{t+1}, on some porftolio. Denote the distribution of returns, $g(y_{t+1})$. In this framework, Value-at-Risk modeling reduces to estimating a single percentile, say the 99th, of $g(y_{t+1})$.

Typically such risk models are composed of two parts. First, the model contains a set of risk factors, such as interest rates and exchange rates. Let x_t be the Nx1 vector of factor returns realized at time t. These factors are assumed to follow some distribution (e.g. Normal) or shocks can be resampled from historical observations (historical simulation). Denote the distribution that is utilized, $f(x_t)$. For example, under historical simulation, $f(\cdot)$ assigns probability 1/T to each historical observations, $\{x_1, x_2, ..., x_T\}$.

The second component of the model is a set of pricing rules, $P(\cdot)$, which predict asset values as a function of the underlying factors. For example, we might price options using Black-Scholes. We will call this collection of pricing rules a *valuation model*.

Once the risk model is constructed, the user generates data, $\hat{x}$, from $f(\cdot)$ which is then fed into the valuation model, $P(\cdot)$. If we denote a draw from $f(\cdot)$ as $\hat{x}_f$, then a simulated return can be written $\hat{y}_{t+1} = P(\hat{x}_f)$. This process is repeated a large number of time to build up a forecast of $g(y_{t+1})$. For ease of exposition, I will generally refer to the output of the *model* as the full distribution of $\hat{y}_{t+1}$. This should cause no confusion because a VaR *estimate* could in principle be any percentile of the distribution.

If stress-testing is supposed to put unusual scenarios through the model, then we can understand it as one of the following:

1. Simulating shocks which we suspect are more likely to occur than historical observation suggests.
2. Simulating shocks that have never occurred.
3. Simulating shocks that reflect the possibility that statistical patterns could break down in some circumstances.
4. Simulating shocks that reflect some kind of structural break that could occur in the future.

Note that a scenario falling under the second category, can be understood as a special case, albeit extreme, of the first category. Category three describes structural breaks across states of the world, such as asset correlations increasing during a crisis (e.g., Boyer, Gibson and Loretan (1998) and Kodres and Pritsker (1999)). Category four describes the advent of a systematic structural break over time, such as a switch from fixed exchange rates to floating.

Mathematically, all four categories of shocks entail drawing from some new factor distribution $f_{stress}(\cdot)$ not equal to the distribution typically used, $f(\cdot)$. This leads us immediately to the following definition.

Definition *Consider a risk model generating a forecast distribution,* $g(y_{t+1})$ *as a function of a* *valuation model, $P(\cdot)$, and a factor distribution $f(\cdot)$. Then a stress-test of $P(x_f)$, is a second forecast distribution, $g_{stress}(y_{t+1})$, generated under a modified factor distribution $f_{stress}(\cdot)$.*

According to this definition, stress-tests are conducted through the same valuation model as the basic risk model. A typical simulated stress-test return (the result of a stress scenario) can be written, $\hat{y}_{t+1} = P(x_{fstress})$.

Most familiar stress scenarios are easily couched within this definition. For example, we might be interested in a stock market crash. The modified distribution $f_{stress}(\cdot)$ puts a high probability (perhaps equal to one) on unusual behavior by the equity factor. Since crashes are rare in the historical data, $f_{stress}(\cdot)$ is a distortion of any $f(\cdot)$ that is fit to past returns.[2]

Other types of shocks might change $f(\cdot)$ to allow for unlikely correlations and comovements among factors. To formulate more complex shocks, such as a mid-east crisis, the risk manager might shock oil prices and a subset of bond and exchange rates according to her subjective beliefs as to what such a crisis would look like. The severity of the shocks and which factors to include in the scenario are basically a judgement call. Kupiec (1999) notes that when constructing a particular scenario, it is common practice to "zero out" all but the primary factors of interest. For example, any exchange rates that are not expected to play a key role in a middle east crisis scenario would be left unchanged -- unlike the basic running of the risk model in which all factors move. In that case, $f(\cdot)$ is changed so that the distribution of some factors is degenerate (zeroed out).

2b. Implications

Several points emerge from the above description of stress-testing. First, stress-testing has no effect on the pricing model. If scenario analysis has no direct effect on $P(\cdot)$, does that mean pricing rules cannot be shocked? They can but such experiments at present come under "sensitivity analysis" (see, for example, the BIS document, *Credit Risk Modeling: Current*

---

[2]Clearly the practice of over-sampling (exponential weighting) or discarding "old" observations can be viewed as modifying an empirically based distribution.

Practices and Applications). Scenario analyses often involve changes in factor correlations but apparently in practice never include shocking the pricing models. To avoid confusion and to maintain a link with actual practice, we adopt the same distinction in the present paper.

A second point is that, in theory, VaR's calculated through historical simulation are only logically consistent with a particular subset of stress-tests. Under historical simulation, the factor shocks are drawn from historical realizations -- there is no assumed distribution. If there has ever been a market crash, then a crash-like scenario will be fed into the model. There is no parameter uncertainty if the model is nonparametric! This logic breaks down if we are interested in scenarios that have never occurred -- a reasonable possibility. The problem unfortunately remains that there is no way to evaluate their relevance or backtest if they have never taken place.

The final and most important implication is that stress-testing is easily embedded *within the VaR framework*. The distinction between the basic risk (perhaps VaR) model and the stress-tested model is an artificial one. To see this, note that the basic model's output (a simulated distribution) is $g(\hat{y}_t)$ tabulated from the simulated returns,

$$\hat{y}_t = P(\hat{x}(f)) \qquad (1)$$

This is the distribution of portfolio returns, given by a pricing model and factors, $\hat{x}_t$, simulated under distribution to $f(\cdot)$. The stress-test forecast distribution is simply $g_{stress}(\hat{y}_{s,t})$ tabulated from:

$$\hat{y}_{s,t} = P(\hat{x}(f_{stress})) \qquad (2)$$

where the factors now have distribution f_{stress}.

At present, the difference between model (1) and (2) is simply that $f_{stress}(\cdot)$ is a deformation of the basic $f(\cdot)$, which features typical moves in equity markets, interest rates and the like. It seems likely that models (1) and (2) could somehow be combined into a single model. Such a consolidation would not be a mere mathematical nicety -- it is critical for coherent risk management. In the next section, we detail the problems associated with having two separate forecasted loss distributions.

3. Making Use of Stress-Tests

There is no theoretical reason to distinguish between stress-tests and VaR but perhaps there is an operational motive. Perhaps there is some numerical or operational reason to keep unusual scenarios outside the basic model. If this is the case, the regulator or risk manager will have *two* predicted returns distributions. What should she do with these two separate results?

Let's begin by defining the bank's and/or regulator's objective function as follows:

Definition *The bank's objective criterion is $U(g(y_{t+1}))$, a convex function, $U(\cdot)$, of the expected (firm-wide) portfolio return distribution.*

This definition would seem to encompass any reasonable objective. It simply says that the firm's goals depend on its expected financial returns in some way.

After stress-testing, we have two such predicted distributions, $\{g(\hat{y}_t), g_{stress}(\hat{y}_t)\}$, so the bank's problem becomes:

$$\max_{g,\ g_{stress}} U(\cdot)$$

The firm would like to maximize its objective function but now has two separate forecasts, $g(\hat{y}_t)$, $g_{stress}(\hat{y}_t)$. Clearly, the solution is some kind of forecast combination -- write this optimal combination as $h(g(\hat{y}_t), g_{stress}(\hat{y}_t))$. The function $h(\cdot)$ combines the basic model forecast with the results of the stress-test into a single "best" forecast. What does this function $h(\cdot)$ look like? The next two propositions show that there is no simple answer to this question.

Proposition 1 *Assuming the scenarios built into stress-testing could occur with some non-zero probability, the best forecast of $\hat{y}_{t+1}$ cannot in general be described by a linear combination of $g(\hat{y}_{t+1})$ and $g_{stress}(\hat{y}_{t+1})$.*

Proof *See appendix.*

The import of this proposition is that, even under the best circumstances, we do not know what to do with the results of the stress-test. There may be some nonlinear function of the basic VaR and the stress-test that could be used -- but we do not even know the form of such a rule. The form depends on the objective function h(·) in a nontrivial way. Stress-testing *separately* from the basic risk model is not a coherent approach to risk management.

Proposition 2 *If the scenarios built into the stress-test cannot ever occur, then the optimal rule is to not stress test.*

Proof *See appendix.*

Proposition 2 does nothing more than formalize the obvious notion that if stress-testing is comprised of unrealistic scenarios then stress-testing is unnecessary.

Beyond the math, the point here is that we cannot recommend stress-testing because it does not fit into a *coherent* risk management framework. Under current practice, the crisis events are outside the basic model and hence need not have probabilities assigned. To make the stress scenarios useful, they must be assigned probabilities. Incorporating stress scenarios within the basic risk model, would then insure that they are associated with probabilities.

Why are probabilities so important? Imagine that some stress scenarios are put to the valuation model. It is impossible to *act on* the results without probabilities. A doomsday scenario might bankrupt the company, but be so unlikely as to be irrelevant. Senior management cannot make responsible decisions based on shocks that may or may not be too unlikely to care about.[3]

4. A Unified Framework

Up to this point we have emphasized that a basic problem with stress-testing is that it results in two competing forecast distributions. The forecasted distributions are generated under different assumed factor distributions. A natural question is whether we could define a meta-distribution of factors, say $f_{combined}(\cdot)$, that assigned a positive probability to all scenarios, both crashes and more typical moves.

If so, this new "model" would embed both equations (1) and (2) within it. In this section, we argue that such a combination is possible. Begin by defining a new factor distribution,

---

[3]In fact, there is anecdotal evidence that, at least at some firms, the results of stress-testing are generally ignored for precisely this reason (pers. comm.).

$f_{combined}(x)$ such that

$$x \sim f(\cdot), \quad \text{with probability } (1-\alpha) \qquad (3)$$
$$x \sim f_{stress}(\cdot), \quad \text{with probability } \alpha$$

where α is the (possibly subjective) probability assigned to a particular stress scenario. Both $f(\cdot)$ and $f_{stress}(\cdot)$ are n-variate distributions describing the joint behavior of the n factors. For simplicity, we begin by considering only a single stress-test. The approach is generalized to allow for several stress tests below.

It is straightforward to simulate factor realizations under (3) using standard Monte Carlo techniques. For example, we might define a random variable η~Uniform(0,1) so that x is drawn from $f(\cdot)$ if $\eta<1-\alpha$ and from $f_{stress}(\cdot)$ when $\eta>1-\alpha$.

The remaining steps proceed exactly as within the basic model framework. That is, we evaluate the simulated factor shock with the valuation model, $P(\hat{x})$, to produce simulated portfolio returns. By repeating this many times, we can tabulate a forecast distribution $g_{combined}(y_{t+1})$. This distribution incorporates the subjective importance of unusual or unseen scenarios while leaving the remainder of the probability mass on historically typical moves.

It is instructive to verify that the combined distribution can indeed handle all four categories of stress-tests outlined in section 2. Scenarios in the first two categories, those that are composed of unusually large movements in a subset of factors, are trivially embedded within (3). We merely define $f_{stress}(\cdot)$ in such a way as to generate the desired shocks. Similarly, a stress-test featuring a structural break across states of the world clearly fits into this framework. For example, suppose we are interested in a scenario in which correlations between foreign and domestic equities increases. If s_t and s_t* are the domestic and foreign equities, then we specify $f_{stress}(\cdot)$ such that $E_{fstress}(s_t\, s_t^*) = k$, the desired stress-test covariance. Of course, $k > E_f(s_t\, s_t^*)$, the more typical covariance. To deal with structural breaks over time, we must consider a distribution $f(\cdot)$ that is inherently dynamic -- one in which today's factor distribution depends on past events. Suppose, for example, that factor $s_t \sim N(\rho\, s_{t-1}, \sigma^2)$ under $f(\cdot)$. In other words, $s_t = \rho\, s_{t-1} + \varepsilon_t$,an AR(1) conditional on the factor's past realization. In this case, a structural break over time could take various forms. We might imagine a systematic increase in the factor's mean value so that, under $f_{stress}(\cdot)$, $s_t = \mu + \rho\, s_{t-1} + \varepsilon_t$, where μ is some positive constant. Alternatively, we might be interested in a scenario in which the factor's persistence increases: $s_t = (\mu+\rho)\, s_{t-1} + \varepsilon_t$, under $f_{stress}(\cdot)$.

It should be said that the combined factor distribution (3) does not come without costs. The risk manager must decide upon a crisis probability, α. We emphasize, however, that such probabilities should be assigned anyway. The unified approach imposes discipline in the form of requiring hard numerical probabilities on the stress-scenarios. In current practice, such probabilities may never be formally declared. This leaves stress-testing in a statistical purgatory. We have some loss numbers but who is to say whether we should be concerned about them?

In reality, risk-managers are likely to be interested in a set of scenarios, not just one. Fortunately, generalizing (3) to multiple stress-tests is straightforward. Suppose there are m stress-tests of interest, each assigned a (subjective) probability, $\{\alpha_1, \ldots \alpha_m\}$. Then we would like to construct a combined factor distribution such that:

$$\begin{aligned} &x \sim f(\cdot), && \text{with probability } (1-\Sigma\alpha_i) \\ &x \sim f_{stress,1}(\cdot), && \text{with probability } \alpha_1 \\ &&& \ldots \\ &x \sim f_{stress,m}(\cdot), && \text{with probability } \alpha_m. \end{aligned} \tag{4}$$

Again, we could generate this distribution by defining a "dummy" random variable η~Uniform(0,1) and draw x from $f(\cdot)$ if $\eta<1-\Sigma\alpha_i$. The range between $\Sigma\alpha_i$ and one is partitioned into m sections, corresponding to the m stress-tests. Specifically, we draw x from $f_{stress,1}(\cdot)$ if $\eta\in(1-\Sigma\alpha_i, 1-\Sigma\alpha_i+\alpha_1)$. Draw x from $f_{stress,2}(\cdot)$ if $\eta\in(1-\Sigma\alpha_i, 1-\Sigma\alpha_i+\alpha_1+\alpha_2)$ and so forth.

4b. Conditioning on Recent Events

Inspection of the combined factor distribution (3), immediately reveals that $f_{combined}(x)$ assigns a probability to stress events that is constant over time. If there are times when stress events are more or less likely, this is an assumption we would like to relax. One way to allow the crash probability α to vary over time is to build a statistical model that allows for dependence upon recent factor movements. For example, we might specify:

$$\alpha_t = \bar{\alpha} + \rho\ (x_{t-1} - \hat{x}) + \varepsilon_t \tag{4}$$

where $\hat{x}$ is the sample mean of x_t and ε_t is a residual. To insure that the crash probability is between zero and one, equation (4) should be modified by logit or probit transformation. For example, a logistic distribution would give

$$\mathrm{pr(crash)} = \frac{e^{\bar{\alpha}+\rho\,(x_{t-1}-\hat{x})}}{1+e^{\bar{\alpha}+\rho\,(x_{t-1}-\hat{x})}}. \tag{5}$$

Of course, the user must specify the average crash probability, $\bar{\alpha}$, and persistence, ρ.

Alternatively, we might want α to vary with subjective beliefs or extra-sample information rather than past factor shocks. There is nothing to prevent a risk manager from allowing α_t to change in any way deemed appropriate and then defining

$$f_{\text{combined}}(x) = \begin{cases} x \sim f(\cdot), & \text{with probability } (1-\alpha_t) \\ x \sim f_{\text{stress}}(\cdot), & \text{with probability } \alpha_t. \end{cases} \tag{6}$$

Constant crash probabilities are of course embedded within (6) as a special case. There might, however, be added danger in allowing α_t to vary arbitrarily over time. If α is constant or evolves according to (4), external model review by regulators or senior management is relatively straightforward. The two parties simply need to agree on realistic scenarios and crash parameters. With subjectively time-varying α_t, model, no two models are ever the same.

These concerns can be addressed to some extent if the model is backtested. Incorporating the stress-tests into the basic model, results in a single forecasted loss distribution. As such, the standard backtesting tools are immediately applicable. Indeed, this is a powerful argument in favor of a unified stress-testing approach. At present, stress-tests are separate from the basic model and are therefore not subject to backtesting!

5. Concluding Remarks

The recent rise in prominence of stress-testing coupled with its vague definition could potentially lead risk managers to view stress-testing as a kind of silver bullet against disaster. Such a view might instill a false sense of security because, at present, stress-testing is conducted without a probabilistic structure. Management should be skeptical of scenarios that are chosen subjectively, not given probabilities and not backtested. Regulators, who cannot know the firm's true objective function nor how the results of stress-tests affect firm behavior, should be all the more wary.

Certainly, a primary concern of regulators and managers is to understand how well the predicted distribution of $\hat{y}_t$ compares to the actual realized distribution of portfolio returns. However, there seems to be some confusion regarding *stress* testing as opposed to *backtesting*. Stress-testing is often touted as providing a “check” on the basic VaR or accounting for some of the model uncertainty in the forecast. Such a view is mistaken -- stress-testing is not model validation! A second loss forecast does not in any way provide information on model performance. However, if we incorporate stress-testing into the basic model, we can proceed with backtesting and examine whether a particular model matches historical data. Recent methodological advances in backtesting (e.g., Christoffersen (1998), Berkowitz (1999)) make this possibility especially attractive.

In addition, unification of stress-testing with the standard risk model imposes much needed discipline on the risk manager. It requires that the scenarios that make up stress-testing be assigned probabilities. Although one would assume that such probabilities are at least implicitly attached to the results of stress-tests, it need not be the case under current practice.

References

Berkowitz, J (1999), "Evaluating the Forecasts of Risk Models," Federal Reserve Board, Finance and Economics Discussion Series, #1999-11.

Boyer, B. H. Gibson, M. S. and M. Loretan (1997), "Pitfalls in Tests for Changes in Correlations," Federal Reserve Board, International Finance Discussion Papers #597.

Christoffersen, P. F. (1998), "Evaluating Interval Forecasts," International Economic Review, 39, 841-862.

Kodres, L. E. and M. Pritsker (1999), "A Rational Expectations Model of Financial Contagion," Federal Reserve Board, Finance and Economics Discussion Series #1998-48.

Kupiec, P. H. (1999), "Stress-Testing in a Value at Risk Framework," *Journal of Derivatives*, 6, p. 7-24.

Appendix

Proof of Proposition 1

Let the optimal combination of $g(\hat{y}_t)$ and $g_{stress}(\hat{y}_t)$ be denoted $h(g, g_{stress})$, where $h(\cdot)$ is an unknown function. By definition we have that $\dim[(f,f_{stress})] < \dim[\hat{y}_t]$, so that the number of assets in the portfolio exceeds the number of factors.

Now consider the set of differentiable rules H. We can restrict attention to this set of rules, $h(\cdot) \in H$, because H contains all linear and some nonlinear rules. In general, the distributional cross-derivatives $\partial^2 h(\hat{y})/\partial f^i \partial f^j_{stress} \neq 0$, where f^i is a VaR factor f^j_{stress} is a stress scenario. This implies that $h(g, g_{stress,} \theta)$ cannot be written as a linear combination of $g(\hat{y}_t)$ and $g_{stress}(\hat{y}_t)$. To see this, consider any coefficients a,b. Then,

$$\partial \{ a\, g(\hat{y}_t) + b\, g_{stress}(\hat{y}_t) \} /\partial f^i = a\, \partial g/\partial f^i$$

and so

$$\partial^2 \{ a\, g(\hat{y}_t) + b\, g_{stress}(\hat{y}_t) \} /\partial f^i \partial f^j_{stress} = 0.$$

Proof of Proposition 2

Let Ω_i be the event space on which risk factor i is defined and let $\mathfrak{I}_i$ denote the sigma field generated by Ω_i. Then the Cartesian product $\Omega = \{\mathfrak{I}_1 \times \mathfrak{I}_2 \times \ldots \mathfrak{I}_n\}$ contains the space of all possible outcomes for all the risk factors 1 through n.

An optimal forecast rule assigns to each possible portfolio return, $\hat{y}(\omega)$, $\omega \in \Omega$, a probability which is a unique function of the probability measure defined over Ω. That is, the optimal forecast is a mapping from Ω to R^1.

If the factor shocks built into the stress test are not possible, then $f_{stress} \not\subset \Omega$. The results of stress tests are irrelevant. If stress-testing is at all costly, then the optimal rule is to not stress test.