www.joim.com

Journal Of Investment Management, Vol. 10, No. 4, (2012), pp. 64–90

THE ROLE OF STRESS TESTING IN CREDIT RISK MANAGEMENT

Roger M. Stein[a,*]

In this paper, we outline some concepts relating to the use of stress testing in credit risk management. We begin by providing a simple taxonomy of stress scenarios and discussing the trade-offs that different approaches require for implementation. Our taxonomy is modeled after one that is common in the credit literature and involves concepts related to reduced-form and structural approaches to credit modeling. Recently, some have expressed the view that the use of distribution-based measures such as VaR and expected shortfall (ES) for credit risk management should be deemphasized in favor of stress testing and scenario analysis. We consider this question in the main portion of this article. We discuss the benefits of stress testing and scenario analysis as well as describing some limitations of using scenario-based approaches as a sole mechanism for assessing portfolio risk. We provide a number of examples to illustrate these limitations. In particular, except in special cases, it is difficult to use stress scenarios alone, ex ante*, for allocating capital across disparate portfolios. However, stress testing and scenario analysis are integral to prudent credit risk management and can complement measures such as VaR and ES, thereby better informing both risk assessment and business strategy development. While neither stress testing nor VaR type measures, in and of themselves, provide complete descriptions of credit portfolio risk, combining both approaches results in more robust risk analysis. This permits risk managers to integrate quantitative measures with managerial intuition and judgment to arrive at more comprehensive assessments of both portfolio risk and overall firm strategy.*

[a]Moody's Corporation, New York, NY 10007.

[*]The views expressed in this article are those of the author and do not represent the views of current or former employers (Moody's Research Labs, Moody's Investors Service, Moody's Risk Management Services, Moody's KMV) or any of their affiliates. Accordingly, all of the foregoing companies and their affiliates expressly disclaim all responsibility for the content and information contained herein.

1 Introduction

A BIS study (Committee on the Global Financial System, 2005) compiled survey results on the stress-testing practices of 64 banks and securities firms across 16 countries. At that time, the authors found that:

"The exercise illustrated the wide range of stress test practices at banks and securities firms. The use of stress tests continues to broaden from the exploration of exceptional, but plausible events—the traditional focus of stress testing—to cover a much wider range of applications. The expanded usage of stress testing derives from its wider acceptance within firms. Aside from its inherent flexibility, it benefits from explicitly linking potential impacts to specific events."

Although banks evidenced increasing growth in the use of stress testing prior to the crisis, the focus at that time appears to have been primarily on applications to market risk. For example, the BIS study found that market risk-related stress-testing exercises represented about 80% of all tests reported. In contrast, stress testing for nontraded portfolios (e.g., loan books, retail, and other credit portfolios) was nascent in most institutions, as were practical methods for integrating the risks of different portfolios across an institution.

Since the crisis, however, there has been considerably more attention given to these issues. This is evidenced by the sudden and pronounced growth in the literature on stress testing (starting in about 2006) and the frequent focus of regulators and risk managers on the design and implementation of stress tests since the onset of the crisis.

Stress tests are being used in a broader variety of contexts than even in the recent past and being applied to a wider set of tasks. These applications range from the more traditional regulatory reporting and risk management to newer uses as part of the due diligence process for acquisition analysis and in strategic planning to set a bank's risk appetite or determine which business segments to grow or stem. Within the regulatory community, stress testing has also emerged as a key tool in monitoring systemic risk.

In this paper, we focus primarily on how stress testing can be used to enhance credit portfolio risk management and the analysis of systemic credit risk. We present perspectives on both the great strengths and the important limitations of stress testing and scenario analysis for these applications. Although our focus is on credit risk, a number of our observations also carry over to stress-testing exercises related to market risk and liquidity risk; we do not, however, consider these in detail here.

The recent resurgence of interest in stress testing has occurred, in part, because there is arguably no more intuitive form of risk analysis than a stress test. The selection of scenarios *for* and the analysis of output resulting *from* stress-testing exercises often precipitate intense and productive discussions between business managers and risk managers within a financial institution.

For example, in setting a firm's risk appetite, stress testing provides means for a bank to go beyond generic statements such as: "*The bank will not take on any risks that put the enterprise at risk.*" to more concrete ones such as: "*The bank should be able to withstand a mild recession while still achieving break-even profitability and maintaining a 4% Tier I capital reserve*," and to provide analysis to support this.

Stress tests also offer alternative perspectives and additional flexibility. It can be difficult to use more quantitative techniques to examine the extreme tails of distributions that fall outside of the dynamics (or the data) of historical experience (Bohn and Stein, 2009). Stress-testing exercises provide a means to associate concrete views on states of the world with model outputs and to evaluate those outputs for reasonableness.

Some market participants have gone as far as to advocate the use of stress testing as a *substitute* for more traditional loss-distribution type measures (e.g., multifactor VaR or expected shortfall [ES]).[1] Our view is more measured. While we find stress testing to be valuable for gaining insight into an institution's portfolios and models, we see it as a *complement to* rather than a *replacement for* distribution-based methods. Said differently, *stress tests may best be used for motivation rather than measurement*.

In this paper, we discuss views on the practical role of stress testing within the credit risk-management function. We differentiate the use of stress testing as a *qualitative approach* to understanding and providing reality checks for a model or portfolio on the one hand, from the use of stress testing in a *quantitative setting* as a capital measure, on the other. While there exist techniques for assigning specific probabilistic interpretations to individual stress scenarios in bespoke analytic settings, in most cases, it appears to be difficult to use stress testing as the sole mechanism for making quantitative statements, *ex ante*, about the probabilities of large future losses on a bank's various holdings; it is also difficult to find a single or small set of general stress scenarios that will be adequate to measure risk consistently across portfolios.[2] This implies that setting capital based solely on stress tests results can be challenging.

This conclusion is based on three observations:

(1) It can be hard, in general, to *order* macroeconomic stress scenarios (from worst to best), even when point probabilities can be assigned to them, except through *ex post* reference to the portfolio-specific losses that are forecast under the scenarios.
(2) The cumulative loss probability under a stress scenario will vary from portfolio to portfolio (or for a single portfolio over time), making designing *generic* stress thresholds hard *ex ante*.
(3) It is difficult to assert that the behavior of market participants during moments of extreme stress will be similar to (or an extrapolation of) the behavior of participants during normal or even "very bad" times. Ultimately, therefore, the *linkage* between an extreme state of the economy and the behavior of assets in that extreme state is a matter of judgment, not empirical fact. While this is not a limitation unique to scenario analysis, it can be particularly pronounced, given the extreme nature of many stress scenarios.

Despite its limitations as a stand-alone capital allocation approach, stress testing is nonetheless a valuable component of a robust credit–risk management program as it can help mitigate model risk and provide insight into macroeconomic cases outside of a model's construct. Said more strongly, *not* performing stress-testing exercises may lead to significant oversights in credit risk management.

In addition to informing the risk management process, however, stress testing can also inform discussions of business *strategy*. As risk-management functions at financial institutions have evolved, considerations of downside risk have become a larger component of strategic discussions regarding an institution's risk appetite. Assessing the impact of stress scenarios on a business provides one effective method for gaining insight into which strategies may lead an institution to operate outside the bounds of its overall risk appetite (as specified by its board-level risk committee). For example, the results of a bank-wide stress-testing exercise may suggest certain business lines for which lending or exposure limits should be reduced or increased, or the results may inform the allocation of risk management staff and resources in the bank. Furthermore, identifying trends in stress-testing results compiled over time can serve as early warning signals for senior managers.

As stress scenarios are easier for business managers to grasp, they can be applied more readily to strategic decisions. Capital sufficiency discussions can then merge with risk-strategy discussions and thus bring the objectives of business managers risk managers closer in line. Stress-testing exercises make risk models more tangible, thereby making it easier for business managers and risk managers to communicate about the implications and severity of extreme losses.[3]

In the remainder of this article, we outline some of the arguments that motivate these observations. We begin in Section 2 by describing a simple taxonomy of stress scenarios and discuss the trade-offs that different stress scenario construction approaches imply. Our taxonomy is modeled after one that is common in the credit risk literature and involves concepts related to reduced-form and structural approaches to credit modeling. Section 3 describes some of the challenges in using stress test results probabilistically across portfolios. Section 4 discusses applications of stress testing in both internal risk management and regulatory settings. Finally, Section 5 concludes by summarizing some of our key observations.

2 Types of stress scenarios

The term "stress scenario" is frequently used without definition. This may be due, in part, to the evolution of stress testing, which started in many institutions primarily as a tool for examining the market risk of portfolios (Committee on the Global Financial System, 2005). In a market-risk context, the definition of a stress test is generally self-evident since the factors in a stress scenario typically involve variables like interest rate levels and term structure, the levels of equity indices, and so forth. The primary concern in market-risk is the level and volatility of the factors themselves, since these directly determine the price of the asset. Therefore, stressing a portfolio is exactly equivalent to stressing underlying market factors. Interest rate swaps derive their market value directly from the underlying rate structure, and equity positions often track closely the value of broad market indices, so shocking interest rates or equity indices forms a straightforward stress test.

However, recently, interest has expanded beyond market risk. As a result, more general macroeconomic factors (such as GDP, oil prices or home prices) or asset–class-specific loss rates (such as default rates for mortgages or small and medium enterprises [SMEs]) have become components of stress tests. In these cases, it is not always easy to find a *direct link*[4] between the factor level and the asset value that depends on it. Instead, some sort of more complex *linking function* is required to translate a change in the factor level to a change in the asset value.

For purposes of this paper, we define two broad types of stress scenarios: those that we term *structural* stress scenarios and those that we term *reduced-form* stress scenarios. We adopt this terminology from the credit-modeling literature in which a structural model refers to one in which there is a *causal*, economically intuitive relationship between the level of a firm's asset value and its probability of default. In contrast, *reduced-form* models treat default events as "surprises," the causes of which are not part of the model, but the behavior of which can be observed.

Though the analogy is imperfect, in our taxonomy, a structural stress scenario is one that posits a particular *state of the economy*, as described by macroeconomic state variables (e.g., the level of unemployment), and relies on some form of model to link this state of the world, in an economically plausible manner, to the resulting state of the assets. A reduced-form scenario, on the other hand, directly posits *the state of the assets* (e.g., the level of default rates), without necessarily providing an economic cause for this state. In the remainder of this section, we discuss the form and properties of these different approaches as well as the constraints that each places on stress-testing exercises.

We begin with stress tests based on economically motivated scenarios. These involve examining the effects on a portfolio or firm of some particular macroeconomic path, which we term a *structural* scenario. These scenarios can be particularly helpful for communicating risks because a firm's management and risk managers can use the tangible nature of the macroeconomic factors to evaluate the plausibility of the scenario and to imagine how individual factors might evolve, given the states of the others. In order to use such scenarios, the stress-testing exercise must also make use of one or more *linking functions* that serve to relate the value of these macroeconomic factors to asset behavior. For example, stress testing an SME portfolio using, say, changes in GDP, would require a function that translated changes in GDP into changes in the individual or aggregate loan losses.

Virolainen (2004) provides one example of this approach. The author uses a nonlinear aggregate model to determine the relationships between key macroeconomic factors and the aggregate corporate default rates in different sectors in Finland by estimating a model in a SURE framework and then using these estimates to stress the aggregate sector default rates.[5] Otani *et al.* (2009) give another example of such a stress-testing exercise from a regulatory perspective. The authors describe their implementation of an aggregate model for borrower-rating transitions, driven by GDP and Debt. In their model, aggregation occurs at the rating category level.

Although they are intuitive, structural stress scenarios make high demands on stress testers since not only must the movements of the economic factors be internally consistent, but the resulting asset behaviors, given those movements, must also be characterized through the linking function.

Far less demanding are the *reduced-form*[6] scenarios that do not require such links: a stress tester must define only the asset behaviors themselves. The economic mechanism that led to the behaviors is superfluous to the exercise. Prior to the crisis, use of such reduced-form scenarios was common. For example, at the time of the BIS's 2004 study, most banks credit-stress tested their loan books by simply shocking quantities such as PD and LGD levels or rating transition rates (Committee on the Global Financial System, 2005). Rösch and Scheule (2007) offer a more recent example of a reduced-form framework for stress testing retail portfolios at the aggregate level. In their approach, PDs and correlations are stressed in various ways. The authors assume a single-factor credit model for PD that they apply to each (homogeneous) asset class.[7]

Finally, some factors, such as interest rates, may take on either a structural or reduced-form role. In the case of interest rates, exposures such as interest rate swaps or risk-free securities may be evaluated directly through their explicit dependence on the level of interest rates at a specific point on the yield curve. However, for other instruments, interest rates might be a structural factor. For example, mortgages have more complex relationships to interest rate levels since the speed of prepayments, the interest burden, and so forth depend in part on interest rate levels, in part on other macroeconomic factors such as home prices and in part on non-macro factors such as the initial coupon of the mortgage or the loan structure. In this case, more complicated linking functions are required to determine the impact of changes in interest rate levels.

We can understand better the differences in these approaches by examining their respective characteristics. Table 1 provides a brief comparison of the two approaches along some key dimensions.

There is a natural interplay between the dimensions of *Asset Relationships* and *Economic Rationale*, which might be considered the defining characteristics of each approach. In general, the great benefit of the reduced-form approach is the relaxation of the requirement that there be a link between the fundamental macroeconomic factors and the asset behavior. ("Default rates increase by 30%…") The structural approach imposes this relationship. The payoff for this imposition comes in that there is typically a clear economic explanation for *why* the assets behave the way they do ("…because unemployment rates rise to 10%."). The direct causal relationship between a change in a macroeconomic variable and a specific risk parameter is a salient attribute of the structural approach.

The recourse back to fundamental relationships also provides a means for ensuring that the behavior across assets or portfolios is consistent—at least with respect to assumptions about states of the world. In a reduced-form stress scenario, if the default rate on asset class A doubles, should the default rate for another asset class (B) also double, increase by only 1.5 times, or stay flat? The reduced-form approach provides little guidance.

In contrast, under a structural stress scenario, if the default rate on asset A increases by 30% because unemployment rises to 10% on one portfolio, we can examine what happens to the default rate on asset B when unemployment rises to 10%. In this way the default-rate shocks will be more consistent across portfolios. The macro factors serve as an anchor to ensure that the same states of the world obtain in each portfolio—though the behavior is still subject to the individual linking functions.

The issue of how robust a stress scenario is to differences in portfolio structure (last row in Table 1) is one that we deal with in more detail in Section 3.2. It turns out that heterogeneity within and across portfolios implies that a specific macroeconomic stress scenario will be more or less stressful for some portfolios than others, even when the assets are of the same broad type. This is because different macroeconomic behaviors may affect different individual assets differently.

Table 1 A comparison of structural and reduced-form scenario approaches.

Reduced form		Structural
Shock default-rate of portfolio	*Example of a scenario*	Shock levels of home prices or GDP
None	*Asset relationships required*	Link between asset behavior and stress factor
Not required and typically not given	*Economic rationale for portfolio effect*	Explicit changes in fundamental macroeconomic factors
Not usually enforced	*Cross asset class/portfolio coherence*	High due to dependence on common factors
None. Portfolio state given as part of stress scenario	*Impact of within portfolio heterogeneity*	High. Depends on macro factors selected and the factor loadings

Reduced-form stress scenarios simply impose a specific level of asset behavior without reference to the macroeconomic scenario and thereby sidestep this issue.

3 Using stress tests probabilistically

Berkowitz (2000) proposes a formal definition of stress tests and goes on to argue that stress tests be combined with distributional analyses (e.g., VaR) in a coherent framework. The proposed mechanism through which this is achieved is for each scenario (or for the stressed-scenario distribution as a whole) to be assigned a probability (either based on historical data or subjectively) and for the distribution underlying a VaR model be similarly assigned a realization probability. These distributions (of the various scenarios and VaR distributions) are then sampled proportionately to their probabilities. Embedded in the author's recommendation are the assumptions that it is feasible to assign probabilities (and factor distributions) to the stress scenarios and that the objective of the stress testing is to examine a single portfolio, rather than to stress across multiple portfolios.

To some degree these assumptions may hold for single portfolios. However, it does not appear to be the case more generally. One of the emerging requirements for stress testing is that portfolios be comparable both *within* a single institution (e.g., the SME and mortgage portfolios) and *across* multiple institutions (e.g., SME portfolios across large institutions), and in this setting, it is not clear that the implicit assumptions of (Berkowitz, 2000) are practically feasible, when the goal is to keep the cumulative probability of the losses constant across portfolios (e.g., for capital allocation).

As we will discuss in this section, with the exception of some special cases, it is typically not possible *ex ante* to derive, in a manner that permits generalization across portfolios, the probability of losses exceeding the loss obtained under a specific scenario without reference to the full loss distribution of the portfolio.

For example, some mistakenly assert that the cumulative probability of a loss being greater than a loss under a specific stress scenario is equivalent to the probability of the scenario. Thus, if a certain scenario has, say, a 1 in 50 probability of occurring, the assertion would be that 98% of losses on a portfolio would be less severe than the losses under the scenario ($1 - 1/50 = 0.98$). This is generally not true.[8]

In the remainder of this section, we will explore why this might be so.

3.1 The mappings required to move from a scenario to a cumulative loss quantile (economic capital)

Consider the steps needed to move from the definition of a particular stress scenario to determining the probability of a loss exceeding the loss under that scenario, as would be required for capital allocation purposes. (For ease of exposition, we only contemplate the steps of such a mapping for a single portfolio.):

(1) *First, define a method for constructing coherent and realistic scenarios*. By coherent and realistic, we mean that the scenarios reflect reasonable behaviors for interactions of the factors in the scenario. That is, the combination of scenario factors must be plausible. This should be true even if the scenario itself represents an extreme case. For example, it would be unusual to find a state of the world in which unemployment in each major region in a country increased by 1% but the national level unemployment increased by 4%. To be clear, there may be specific reasons for designing such scenarios, but such instances

are rare and their use should be deliberate rather than unintentional or *ad hoc*.

In practice, a number of approaches may be used for generating such stress scenarios. By far the most common is the use of heuristics and judgment. Such approaches are convenient. However, as the number of factors stressed becomes greater (e.g., more than a few) it becomes increasingly difficult to enforce coherence.[9] This is particularly so in instances where forecasts are averaged from multiple sources to form a consensus forecast. In preference to heuristic forecasts, some analysts sample from historical time periods by choosing historically stressful windows or by bootstrapping historical samples. The most sophisticated forecasts are produced by using some form of structural or reduced-form model of the economy, which may then be shocked in various ways to produce more internally consistent forecasts for multiple series simultaneously (Zandi and Blinder, 2010; Stein *et al.*, 2010).

(2) *Assign a probability to the scenario*. Though not strictly required, it is often desired. This may be done by means of expert judgment, the use of more formal Bayesian approaches (Rebonato, 2010), through historical sampling, or through the use of economic simulations. Note that this scenario probability is only the probability that such a scenario will occur, not the probability that other "worse" or "better" scenarios will take place.[10]

(3) *Map the scenario to an asset- or portfolio-specific loss.* This is done through a linking function of some sort. For example, an SME probability of default model might take as input GDP and other factors and produce a point-in-time PD. Alternatively, a mortgage model might make use of the loan-to-value (LTV) of a mortgage in which the "value" component is calculated as a function of the original LTV and change in the home price index for the region under the stress scenario.

(4) *Assign a cumulative probability to the exposure loss.* Strictly speaking, the point probability of the specific loss associated with a specific stress scenario may be the same as that of the scenario as specified in (2). However, for capital allocation purposes we need the *cumulative* probability of the loss, which requires that the losses be ordered. To estimate how much capital to allocate, we also typically need a sufficiently large number of scenarios to be able determine the typical or average loss of an exposure or sub-portfolio when the entire portfolio loss exceeds the capital threshold. Analyses that use only a small number of scenarios may understate the capital usage of a particular exposure or group of exposures. Also note that for reasons we will articulate in more detail below, these probabilities (or even the scenario orderings) cannot usually be transferred from one portfolio to the other or from the same portfolio from one date to a later date after trading has occurred and instruments have become more seasoned.

The difficulty in ordering scenarios *ex ante* makes it challenging to use stress testing as a general technique for capital allocation across portfolios. This is the subject of Section 3.2.

3.2 Ordering individual stress scenarios

It is typically not possible to order macroeconomic scenarios themselves *ex ante*. For example:

- *Is a 2% rise in national unemployment worse than a 10% drop in national home prices?*
- *Is a 10% drop in national home prices worse than a $20/barrel rise in the price of oil?*

In both cases, the resolution of which scenario is "worse" depends on a number of factors, including an understanding of which types of assets are being stressed. Consider an institution that has an active business lending to energy producers and airlines. For this portfolio, a $20/barrel rise in oil prices might be a very bad thing as its borrowers will face economic constraints in their businesses. However, if the lending institution also holds a portfolio of RMBS tranches that are sensitive to home prices, the ranking of stress cases becomes more complicated. Inevitably, the question of which scenario is "worse" raises the corresponding retort "For whom?"

This has only a little to do with whether economists can assign probabilities to different scenarios. Even if the probabilities are given exogenously and we can rely on them, determining the ranking of two scenarios, in terms of their *generic* severity, is typically possible for only the starkest cases (e.g., a 1% increase in GDP vs. a 1% decline in GDP).[11] Knowing that a particular scenario has a 1 in 50 or a 1 in 25 probability of occurring does not usually imply a ranking of its severity *for a particular portfolio*.

Some authors explicitly recognize this. Breuer *et al.* (2008), for example, suggest a search approach for identifying portfolio-specific "bad" stress scenarios in order to "be sure not to miss out any harmful but plausible scenarios, which is a serious danger when considering only standard stress scenarios." The authors assume factor returns are elliptically distributed and then use the Mahalanobis distance (between the scenario and the "mean" economic path) as a measure of plausibility. Flood and Korenko (2010) propose an alternative methodology, also under the assumption of elliptically distributed factors, but based on an efficient grid search. While these papers presents specific approaches to finding stress scenarios, the difficulty in identifying "universally stressful" scenarios has also led to more general proposals by regulators for "reverse stress testing," in which a firm is required to search for scenarios representing states of the world that would result in high losses on the firm's individual portfolio (FSA, 2008). We discuss this topic briefly in Section 4.2.

To delve a bit deeper into the scenario ordering problem, in the next subsection we decompose macroeconomic stress scenarios along two dimensions: (1) the number of factors included in the stress scenario (i.e., a single factor or more than one factor); and (2) the number of (time) periods over which the factors are projected. Both of these attributes affect the ability to order scenarios.

3.2.1 Cross-portfolio coherence of scenarios of more than one factor

In the case of a single factor, all asset behaviors are determined by a single macroeconomic series. In this case, assets will behave differently depending on their factor loadings. However, if we add another factor (or two, three,...), the presence of multiple factor loadings for each asset makes determining the severity ordering for scenarios more involved.

For example, imagine two scenarios

A. *US national home prices drop by 5%. Each state experiences a 5% decline.*
B. *US national home prices drop by 4.5%. Each state experiences a 4% decline but New York experiences a 25% decline.*

Determining which scenario is "worse" is not straightforward. For many portfolios, Scenario B, which involves a national home price decline of only 4.5% is less stressful than Scenario A, which involves a 5% national home price decline. However, for a New York-based banking institution that is heavily exposed to New York real estate, Scenario A, with a 5% decline in national home prices may actually be preferred to the 4.5% national decline under Scenario B, since the decline for New York properties under Scenario B is 25% versus only 5% under Scenario A.

3.2.2 Cross-portfolio coherence of scenarios of more than one period

The time dimension raises similar, but a bit more subtle issues. Imagine three 10-year home price stress scenarios[12]:

A. *Slowdown in growth, but growth remains positive*: Home prices rise ½% each year over 10 years.
B. *Prices drop*: Home prices decline by 5% over 5 years (ending in year 5 at pre-decrease levels minus 5%). After year 5, prices rise at 4.5% per year.
C. *Prices drop severely*: Home prices decline by 25% over the first 3 years and then rise to pre-decrease levels minus 5% over the subsequent 2 years. After year 5, prices rise at 4.5% per year.

These three scenarios are shown in each of the panels of Figure 1.

Also shown are the coupon reset dates for three homogenous portfolios of mortgages:

(1) A portfolio of 3/27 loans (loans that pay a low fixed coupon payment for the first 3 years and then convert to a floating coupon, with a typically higher interest payment);
(2) A portfolio of 5/25 loans (loans that pay a low fixed coupon payment for the first 5 years and then convert to a floating coupon, with a typically higher interest payment); and
(3) A portfolio of 7/1 loans (loans that pay a low fixed coupon payment for the first 7 years and then convert to a floating coupon, with a typically higher interest payment) that resets each year.

For simplicity, in this example, we focus on one dimension of loan performance: payment resets (and the borrower's ability to refinance to avoid resetting to a higher monthly coupon payment). We also assume that most borrowers tend repay

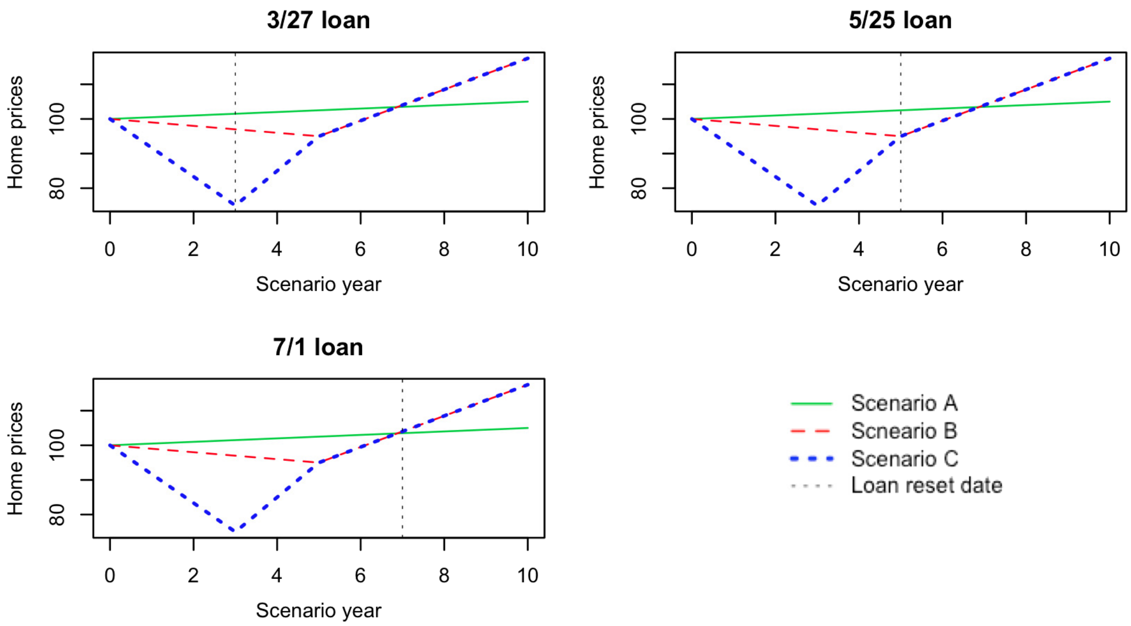

Figure 1 The impact of multiperiod stress scenarios on different loans.

their loans when coupon rates reset at the end of the fixed-rate period.[13] However, it can be difficult to refinance a mortgage when the value of the property is less than the face value of the loan. This implies that when home prices have declined substantially, the risk is highest that the borrower will not be able to refinance and will thus be forced to accept higher interest payments, which, in turn, place a higher financial burden on the borrower.

When we consider the three scenarios generically, it is clear that Scenario C (*prices drop severely*) is worse than Scenario B (*prices decline*) since the peak-to-trough home price decline is more severe in Scenario C than in B, and the two are identical after year 5. Scenario A (*slowdown in growth, but growth remains positive*) would be viewed by many as the least stressful. From the figure, though, it is obvious that the timing and levels of home prices affect the reset risk of each type of mortgage differently.

From a reset risk perspective, Figure 1 shows that given the different structure of the exposures in each portfolio, there is *no clear ordering* of worst or best scenarios. The scenario orderings for each portfolio are summarized in Table 2, and discussed in more detail in the Appendix.

From the table, it is clear that regardless of the probability associated with a specific *economic* outcome, the impact on reset risk will be high or low depending on the structure of the loans in the portfolio being analyzed. It should also be clear that for any scenario chosen to stress portfolios, there is another scenario that is as bad or worse for others.

In this example, we only focused on the borrower's ability to avoid increases in monthly interest payments by prepaying. Clearly, losses on real loans and loan portfolios are governed by a host of other behaviors that interact in a variegated fashion.[14] All of these can affect the ordering of losses under different scenarios for a specific portfolio. The high dimensional nature of the asset behavior makes the ordering problem more acute.

3.3 A schematic for the ease of ranking scenarios

We can generalize our discussion of scenario ranking, albeit in a stylized fashion. In Figure 2, we construct a 2×2 matrix that outlines the stress scenario dimensions we have been using. The x-axis defines the number of factors used in constructing the stress scenario and the y-axis describes the number of periods.

Starting in the upper left, the quadrant represents those assets for which, in general, we can rank single-factor single period scenarios.[15] These cases apply to assets that do not exhibit path dependence of the sort in our mortgage example. Moving down to the lower left, if we extend the single period framing to a multiple period one, but remain in a single-factor world, we can now add path dependent assets, provided they are homogeneous in their characteristics (both within and across portfolios).

In the upper right, we move to a multifactor world, but revert back to a single-period scenario (this

Table 2 Summary of implied reset risk on different mortgage portfolios under different stress scenarios.

	Portfolio 1 (3/27)	Portfolio 2 (5/25)	Portfolio 3 (7/23)
Highest reset risk in	C	B or C	A
Lowest reset risk in	A	A	B or C

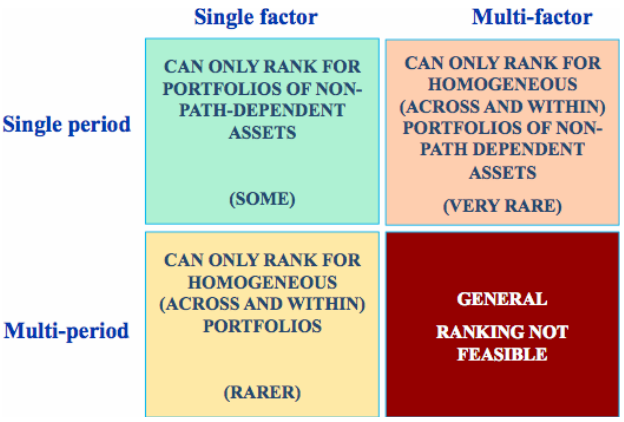

Figure 2 Schematic describing scenarios characteristics and ease of ranking.

would be analogous to "shocking" a number of key macro variables). In this set-up, there are relatively few assets for which we can rank scenarios. Those assets are characterized as being homogeneous and nonpath dependent. Finally, the lower right represents the richest (and most realistic) set of scenarios that allow for multiple factors and a multiperiod setting. This setup permits the most detailed representation of asset behavior, but at the cost of forgoing any practical ability to rank scenarios in a manner that is consistent across portfolios.

This returns us to a common mathematical modeling trade-off: the most stylized representations of the world enjoy the nicest analytic regularities, while the most realistic representations are messy and inconvenient to deal with. As our mortgage example demonstrated, it is unfortunate that many assets about which we might be concerned do exhibit path dependence. Furthermore, the behaviors of many assets are driven by more than a single factor[16] and thus different portfolio construction strategies create different state-contingent payoffs. Duffie (2010) articulates this succinctly:

"Essentially any stress measurement system is subject to a financial risk management analogue of the Heisenberg Uncertainty Principle, by which increasing the precision of one's measurement of one aspect of a system merely increases uncertainty regarding other dimensions of the system. …[I]f a regulator measures the exposure of a bank to a 25% change in the value of an asset, the bank could buy and sell options on the asset so as to lower this particular exposure, while raising its exposure to a 30% change in the value of the asset."

In light of observations such as these, it is reasonable to consider *how much* of a difference the ranking of scenarios might make. To give some sense of this, we present the results of an informal experiment, structured as follows: we took a single (multifactor/multiperiod) scenario and, using a set of linking functions, estimated the portfolio losses under the scenario for a set of mortgage portfolios.[17] We then simulated a full loss distribution for each portfolio, using the same linking functions as in the scenario analysis.[18] Each of the three mortgage portfolios contained an the same number of mortgages (just under 20,000) which were drawn randomly from a large database of over 20 million US mortgage loans).

Because we conducted a full simulation of portfolio losses, in addition to examining losses under the stress scenario, we had access to the full loss distribution for each portfolio. As a result, we were able to assign a loss percentile to the stress-scenario loss for each portfolio. For example, if under the stress scenario, a Pool X experienced a loss of 15% of par, we would look up 15% in the loss distribution for Pool X to determine what percent of losses were greater than 15%.

To simplify the presentation, we focus on evaluating the amount of capital required so that losses will be no greater than L with probability $1 - \alpha$. This is commonly referred to as the "$1 - \alpha$ value at risk level" or the "$1 - \alpha$ VaR."

In Figure 3 we show both numerically and graphically the VaR levels that would be implied by the stress scenario. It is clear from the figure that, even under identical linking functions (the

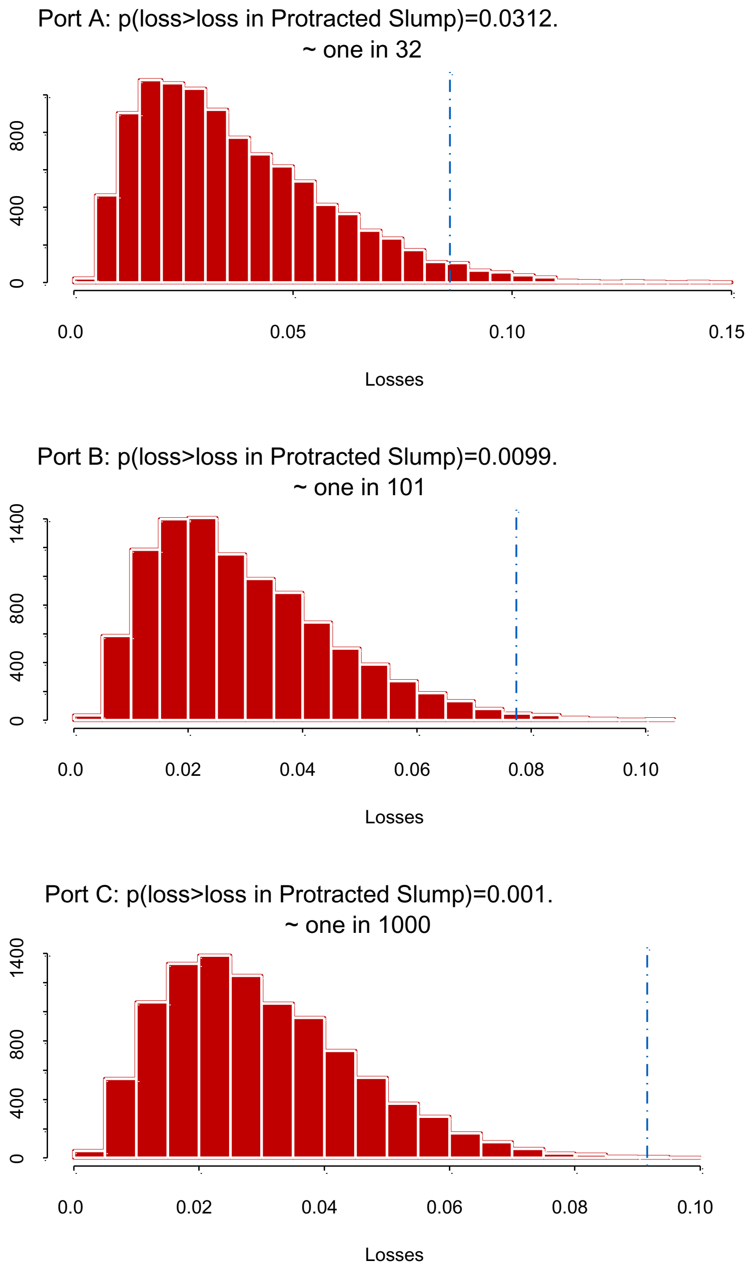

Figure 3 VaR level implied by the same scenario on different portfolios.

same linking models were used in all cases), the probability of exhausting the capital required to "pass" a stress test using this one scenario can vary greatly. For one portfolio (i.e., Portfolio C), this capital would represent a 99.9% VaR, while for another (e.g., Portfolio A), it would offer a far weaker buffer, equating to only a 97% VaR. Said differently, for the identical stress scenario, the probability of exhausting "stress capital" in Portfolio C would be 1 in 1,000 while the probability of exhausting "stress capital" in Portfolio A would be only 1 in about 30.[19] In other experiments we can observe examples of even higher variability (e.g., see Chinchalkar and Stein [2010], Appendix for examples of the same stress scenario resulting in percentiles ranging from 99.99% down to 75% or 80%).

The reason for this disparity is that the scenario is both multifactor and multiperiod in nature. Thus loans in different geographic regions and loans of different types, ages, etc. will all experience the scenario differently.

3.4 Behavior of assets in extreme economic environments

While it is clear that stress tests should explore states of the economy beyond those contained in the historical record, in some cases, acknowledging that the *macroeconomic environment* can be different and potentially worse than those observed in the historical record is not sufficient. In many settings, *borrower and market behaviors* also change materially during a crisis in ways that do not permit simple extrapolation of the relationships observed during normal or even "pretty bad" times.

On its face, there is nothing new here. Introductory statistics texts admonish students to carefully distinguish between relatively simpler interpolation problems and more tenuous extrapolation ones. However, the point is more subtle. There are at least two ways in which the world may be different during times of stress than our linking functions suggest.

First, the underlying relationships between a macroeconomic factor and, for example, the default rate of an asset may be nonlinear in a way that makes it hard to understand the true relationship from historical data. Figure 4 shows an example of this type of "different" behavior during a crisis. In this example, it is not until a key macroeconomic factor begins to move above its historical range (during a crisis) that the fuller shape of the relationship becomes clear.

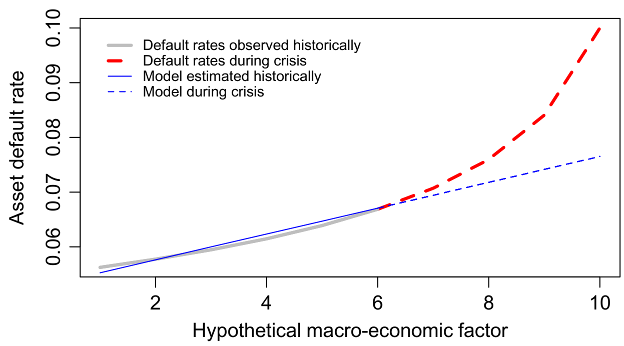

Figure 4 Hypothetical example of a factor whose relationship to default is not clear until a crisis pushes it to new levels.

In a sense, this type of model misspecification is of the traditional "model risk" type and frequent recalibration and validation, along with careful managerial judgment can, at least in part, help to mitigate these limitations. However, a second way in which the realized outcome of a macroeconomic stress may deviate from a model's estimate is that the manner in which individuals and institutions *react* to the state of the economy and markets may change. This type of structural break arises not only from an increase in market stress, but also from a fundamental change in behaviors.

For example, the recent crisis saw the advent of "jingle mail"—a colloquial term used to describe strategic default on the part of mortgage holders.[20] Historically, borrowers tended to be reticent to default on a mortgage for fear of the social stigma associated with foreclosure and for fear of the negative impact a default would have on their credit records. However, in the past several years the reported incidence of strategic default has increased significantly as entire neighborhoods found the values of properties falling well below the values of the mortgages on them[21]; and as investors in real estate, with no ties to the community or attachment to their properties, walked away from what had become bad investments.

The combination of this type of dynamic with model misspecification can diminish the effectiveness of macroeconomic stress scenarios since the links between macroeconomic environments and asset behavior become less reliable in extreme settings. Based on an analysis of pre-crisis stress-testing models, Alfaro and Drehmann (2009) explicitly describe the need to accommodate the type of change in market dynamics that make crisis environments different than environments that are just "very bad." They note:

"[Our] results highlight that the structural assumptions underlying stress-testing models do not match output growth around many crises. Furthermore, unless macro conditions are already weak prior to the eruption of the crisis, the vast majority of stress scenarios based on historical data are not severe enough. Last, stress-testing models are not robust, as statistical relationships tend to break down during crises."

However, this implies that, in general, the only approach, *ex ante*, to systematically accommodating such structural breaks in building stress-testing models is to apply managerial judgment and modelers' subjective views.

In light of these types of unobservable dynamics, it can be informative to conduct stress-testing exercises using combinations of both reduced-form and structural approaches. For example, a macroeconomic stress scenario can be run and, at the same time, the default probabilities within the linking functions can be increased as well (for certain assets or the entire portfolio) to determine how this might impact the results of the stress test.

Typically, this will require more than simply, for example, doubling the losses on a portfolio, particularly if path-dependent assets are included. (For example, one could also imagine decreasing prepayment rates, increasing LGD, adjusting recovery times, assuming that financial guarantees are not honored, etc.) Though such an approach is less satisfying theoretically, it does provide a means to contemplate *ex ante* unknown structural changes, albeit in an abstract sense.

3.5 "Plausibility" and probability

In considering probabilistic interpretations of scenario analysis, it is also natural to think about the notion of the plausibility of a stress scenario. The topic of plausibility has been explored in other literature streams in varying contexts.[22] In the domain of stress testing, it is relevant in that many definitions of a stress test require that the scenarios chosen be "exceptional, but plausible." Although these terms are used often, their definitions are not always explicit.[23]

In considering stress testing, some authors (e.g., Breuer *et al.,* 2008) take the term to be a statistical one and measure plausibility in terms of deviations from an "average" case. However, such representations may not capture the potential for very different outcomes than have been observed historically (or than those that might be drawn from a candidate distribution of some sort).

An alternative view is that plausibility is more subjective but that reasonable individuals should be able to agree on whether a specific scenario is plausible or not. In this sense, plausibility initially involves not the analysis of the precise probability of a scenario occurring, but rather simply determining that the scenario is *not practically impossible*. Once it has been established that a scenario *could* take place, the problem reduces again to that of assigning probabilities. The admissibility of a stress scenario as "plausible" may be further refined by also stipulating that an exceptional scenario have a sufficiently high probability to be considered relevant. (To also be "extreme," the probability must similarly be sufficiently low that it not be something typically encountered in historical data or daily observation.)

Thus, plausibility may be thought of as a probability assignment problem with the added requirement that a (subjective) probability threshold for plausibility also be defined. There are a number of mechanisms one might contemplate for both generating and evaluating potential "exceptional, but plausible" scenarios. However, practically, it may be difficult to elicit useful probabilities for very rare events, even from experts. This can be particularly so of events with negative consequences (Tversky and Kahneman, 1986). As such, the assessment of plausibility remains challenging and largely subjective for very extreme scenarios.

4 Discussion of internal risk management and regulatory applications

Despite some drawbacks as a stand-alone capital allocation measure, stress testing does have a valuable role to play in credit–risk management.

In addition to providing a snapshot of the exposures that a firm or markets face, given a specific scenario, stress testing can provide decision makers with a monitoring tool that allows them to measure credit risks over time, observing trends and changes in the risk profile of those entities relevant to their analysis. In this regard, stress testing seems well suited to providing both risk managers and senior management with a broad directional view on the holdings and portfolio of their institution; it can also provide regulators with tools to understand the evolution of risk in a regulated entity or across a market over time.

Importantly, some forms of stress testing may also provide a means to mitigate model risk by enabling intuitive interpretations of states of the world that may cause a portfolio or organization to experience high losses. This intuitiveness makes stress testing useful in evaluating a model's behavior in general, and the appropriateness of a model's linking functions in particular. Their intuitiveness also permits more transparent communications about models and risks and thereby fosters considerations of credit risk as part of a firm's broader business strategy.

In this section, we discuss the role of stress testing in the risk management of individual institutions' credit portfolios and strategies as well as its role in aiding regulators in monitoring the stability of individual institutions and of the financial system overall.

4.1 The role of stress testing for internal credit–risk management and strategic planning at financial institutions

Stress testing provides a unique means to understanding both risk models and the portfolios that institutions analyze with them. In fact, even if risk models (e.g., VaR tools) were perfect (i.e., had no error associated with their estimates), stress tests would still provide a measure of intuition that is generally not feasible otherwise.

The exercise of selecting factors, creating scenarios and evaluating the impact of those scenarios on a portfolio induces a connection to both the models and the risks in the portfolio that is typically far richer than with quantitative portfolio analytics alone. The scenarios provide intuitive descriptions of states of the world that might occur (but may never have been seen in the historical record) and the losses associated with those states (under a linking function). This provides insights into both the model's behavior and the drivers of credit risk for the portfolio.

It is natural to consider using the two approaches in combination. For day-to-day risk management, VaR (or ES,...) provides useful mechanisms for sizing capital and for identifying which positions contribute the most to the tail-risk of the portfolio. This also leads naturally to the use of such measures in implementing transfer-pricing mechanisms within a financial institution. Credit transfer-pricing can be useful as a common language across the organization for aligning the incentives of those using the institution's capital (i.e., those who create risk exposures through lending and trading operations) and those managing the risk of the institution. This type of transfer-pricing also naturally produces warnings about dangerous concentrations that may be developing in the portfolio and, at the same time, it provides disincentives to originate additional exposures that contribute to those concentrations.[24]

However, even the best designed portfolio tools cannot always contemplate states of the world that are very, very different, both in magnitude and character, from those in the historical record (or that are outside of the theoretical constructs of a model). Stress testing provides a bridge to permit just this analysis. It is one of the most direct and intuitive ways for managers to impart a judgmental overlay on routine risk management.

Stress testing also provides a reality check that can help mitigate some types of model risk. A user can evaluate losses under a stress scenario and compare these to the simulated distribution of losses (e.g., as in the results shown in Figure 3). If the losses appear much higher than under most states of the simulation, the user can use this information to delve into the source of the differences. It may also happen that losses under what the user considers to be a very stressful scenario do not appear high compared to those under, say, extreme outcomes in a portfolio simulation. In this case, again, the user can take advantage of this information to better understand the drivers of portfolio risk. Differences between a user's expectation and the model's output for a stress case may be due to limitations of the model. They may also arise because of disagreements about the relative likelihood of the stress case.

To this end, recently, Rebonato (2010) has proposed a promising approach for developing coherent subjective probabilities for scenarios to be used on single portfolios. The approach relies on the use of Bayesian networks, a form of probabilistic directed graph, popularized in the artificial intelligence community in the 1990s, as a means for collapsing and calculating conditional probabilities. A key feature of this approach is the ability to reduce the dimensionality of conditional distributions through the careful application of Bayes rule. The author recommends eliciting probability distributions from experts for collections of factors that are relevant to a particular portfolio and then using these elicited scenario distributions to form loss distributions for each asset class.

For institutions particularly concerned about specific stress scenarios, the results of a stress test can also be used as a capital hurdle. That is, an institution may elect to set capital levels based on the *greater of* the stress-test results and the results of, say, VaR analysis. In this sense, having adequate capital to pass the stress test becomes a *necessary, but not sufficient condition*.

Using both portfolio simulation and scenario-based approaches permits users to combine managerial and analytic judgment with the portfolio simulation results in a way that is often more satisfying and informative that either one alone.

This may be done by benchmarking losses under a stress scenario to quantiles of a loss distribution generated under a portfolio model, as in Figure 3. But stress testing can also provide a much richer and more flexible setting in which to ask and answer questions. By trading off quantitative specificity of the outcomes against the breadth of the scenarios examined, risk managers can experiment with the impact of quite different assumptions than those they might typically consider.

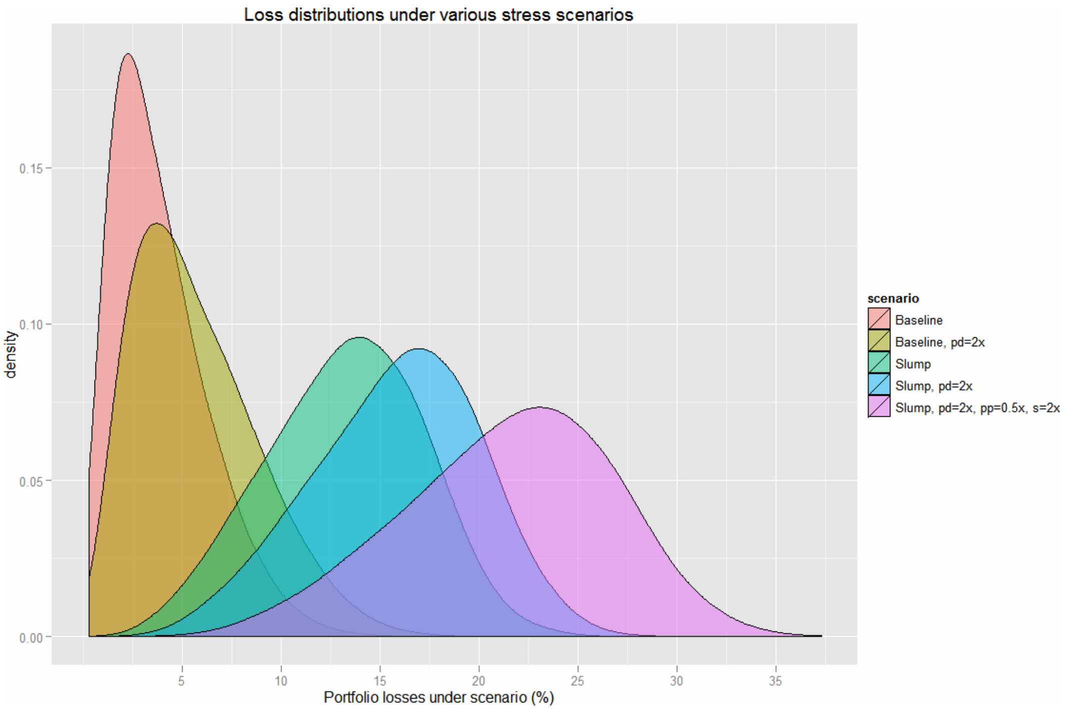

Figure 5 Example combining reduced-form stress testing, structural stress testing and portfolio simulation to assess the changes in the shapes of loss distributions assuming different "mean" economic paths in the simulation.

For example, Figure 5 shows the impact of using different baseline macroeconomic paths for VaR analysis. The figure shows the estimated loss distributions for a portfolio of about 20,000 prime US mortgages under different stress scenarios.[25]

The median path for the simulator used to estimate the leftmost two loss distributions is a baseline economic forecast. However, the remaining loss distributions were simulated using a macroeconomic simulator that was calibrated so that the median path was a generally more severe one. In this way, structural stress scenarios are used to inform the VaR analysis. Some of the simulations are also further stressed using reduced-form shocks to different processes (e.g., default or severity) within the simulations.

While it is unclear how to interpret precisely the many distributions that one might generate in such a setting (we have only shown a few), the insight that this analysis provides can still be valuable.[26] For example, from the figure, it is clear that in addition to the expected movement of the mean losses to the right as scenarios get more stressful for this portfolio, we also observe that the *shapes* of the distributions change as well. As more stressful scenarios are introduced, the associated loss distributions become broader, suggesting more uncertainty, while also exhibiting less skewness (and actually reversing skewness at one point).

Finally, as an institution's senior management develops its business strategy, stress testing can serve as a mechanism to bring to bear insights about the impact of various strategic options on the risk profile of the firm. Conversely, it can serve to highlight risk management challenges for which strategic solutions are sought. Often, financial institutions relegate considerations of risk management to a compliance-based review of a final strategy rather than involving risk managers at the outset as part of the process of developing the strategy itself. Discussion of stress scenarios and stress tests can be useful in strategy development as these discussions may motivate managers to alter business plans in order to build more sustainable franchises, so too can discussions of trends in stress test results over time.

Even in cases in which an organization may have less confidence in the robustness of a specific linking model (particularly when applied to extreme economic outcomes), stress test results can be used to inform a financial institution's management about on trends in their credit risk exposure. For example, by forming a time series of stress test results, a management team may gain at least directional insight as to whether their franchises may be becoming more or less risky over time.[27]

In addition to exploring the impact on the state of a portfolio (or the whole firm) of a macroeconomic stress scenario, stress testing at the strategic level may involve exploring *the reaction* of a firm to a stress event as well. For example, consider the most extreme version of a reduced form stress scenario. This type of test is perhaps better termed a "thought experiment" or "war game." The scenario might take the form of an open-ended question such as[28]:

> "The firm has just lost 10% of its capital, and most of our competitors have also lost between 5% and 20% of their capital. Hedge funds are suffering withdrawals and are pulling positions back from their prime brokers. How do you respond?"

By forcing managers to think through this type of scenario, stress testing aids senior managers in understanding the implications of key strategic decisions. It can also highlight weaknesses in business strategies that make tacit assumptions about how markets function or about the flexibility with which the firm can operate in them. The observations from such stress tests can eventually form the basis for fail-safe plans that better prepare an organization for future economic shocks.

4.2 *The role of stress testing in regulatory monitoring and systemic credit risk analysis*

It is useful to consider what our observations on stress testing imply from a regulatory and systemic risk measurement perspective. Here we focus in particular on the *macroprudential* (systemic) perspective, since much of the preceding discussion on internal risk management and strategy also applies to regulatory uses of stress testing for the *microprudential* (individual firm) perspective.

4.2.1 *Microprudential applications*

Before leaving the discussion of microprudential stress testing, it is useful to reiterate the difficulty of defining one or a few scenarios for use across many portfolios and institutions. The discussion in Section 3.2 suggests that stress scenarios may need to be tailored to individual institutions if the goal is to determine the robustness of the institutions themselves to financial shocks. Said differently, each financial institution will likely have a different set of scenarios that constitute the worst outcomes for their specific portfolios. Thus, it is not likely that one set of scenarios will be equally relevant (in a credit risk sense) to all financial institutions even if the institutions are of similar size and operate in similar markets.

For regulators, the challenge may be to develop scenarios that are rigorous enough to provide confidence in an institution's resilience, given the risks to which that institution is exposed, but that at the same time do not favor or disadvantage any one institution. Reverse stress testing may provide a partial solution, as might the assignment of subjective probabilities, by regulators, to a larger set of stress scenarios as a first approximation to ensuring that the (different) scenarios used at different institutions have approximately the same (subjective) probabilities.

The use of subjective probabilities for a broad set of factor outcomes has the potential to result in a large set of factors and a correspondingly large number of probability assignments. However, by using, for example, the approach described in (Rebonato, 2010), the dimensions of the probability space may be reduced considerably. Nonetheless, this approach explicitly requires that stress testers enumerate "all relevant factors" for a portfolio, which, in many realistic settings may be is a difficult task (e.g., see Footnote 10). Also, by construction, applying this approach *across* portfolios would require a large enough number of factors to sufficiently blanket all portfolios.

Reverse stress testing offers an alternative, model-driven approach. However, there may be trade-offs here as well. On the one hand, sampling of macroeconomic states in search of the particularly severe ones can provide macroeconomic stress scenarios that are customized for an individual portfolio; examining these bespoke scenarios can provide new insights into regions of fragility in the portfolio. On the other hand, performing such a search using more complex macro economic models can produce results that become decoupled from intuition, reducing the search to a less informative statistical exercise. In this sense, the stress test may actually *introduce* model risk rather than help mitigate it.

4.2.2 Macroprudential applications

From the *macroprudential* perspective, the objectives are different than those that focus on individual banks. While it is clearly regrettable when any institution fails, macroprudential stress tests are most concerned with the failure of one or more of the *key links* in the financial system since failures of such institutions may cascade through the system, spreading financial distress. To this end, the goals of stress tests for systemic risk may be less ambitious from a precision perspective than in the microprudential case, even as the implementation of the systemic stress tests becomes more complex.

Given the large number of institutions potentially involved in such a systemic stress test, the scenarios, by necessity, must be standardized across institutions at the expense of analytic detail.[29] For example, Duffie (2010) has proposed an approach[30] that both holds great promise and appears to be gaining wide acceptance due to its practical feasibility. Under this approach, a regulator requires the most significant N financial institutions to report their exposure to their largest K counterparties under each of M stress scenarios, where N, K and M are not too large (e.g., $\boldsymbol{O}(10)$).Under this method, to ensure that asset-specific scenarios could be run, the institutions would choose the K counterparties stress-scenario by stress-scenario. Once the results of each scenario had been computed by each institution, the regulator would then aggregate these results, scenario-wise, to permit the regulator to get a snapshot of the state of the financial system "one tick after" the scenario takes place. The method is a general one that could be applied to stress testing many forms of risk.

An important feature of (Duffie, 2010) is the recursive nature of the method. Conceptually, if a regulator observes that a nonreporting counterparty appears to represent a large exposure for one or more of the reporting entities, the nonreporting entity would then be asked to similarly report stress test results for its own exposures (thus effectively becoming a significant entity itself, thereby resulting in $N + 1$ reporting entities).

Almost certainly, systemic monitoring will require some form of network analysis as well.

Given the number of entities reporting and the myriad of counterparty relationships that naturally emerge as a result, a network representation is a natural one. This also affords regulators yet another means to combine both reduced-form and structural approaches to stress testing.

For example, having reviewed the results of tests under an initial stress scenario, a regulator may identify a specific hedge fund as a systemically important counterparty in the banking system. A reasonable next stress test might be to ask firms to report their hypothetical losses should that fund default *for any reason* under the same (or different) macroeconomic conditions. Such exposures can be determined readily in a network setting. Figure 6 shows an example of one such hypothetical stress test from the network perspective. In this hypothetical example, the (nonreporting) hedge fund "Hedge Fund A" is of interest systemically, given the large volume of exposures for which it is a counterparty.[31]

Finally, it is worth noting that the high-dimensional nature of financial markets and the interactions of the instruments, individuals and institutions that make them up imply that it is unlikely that the source of some future financial crisis will be exactly the same as, or even very similar to, the scenarios that were examined during a stress-testing exercise.[32] Some detractors highlight this as a reason to forgo stress testing entirely. There is an alternative view, however, that argues that *even if* a future crisis is caused by

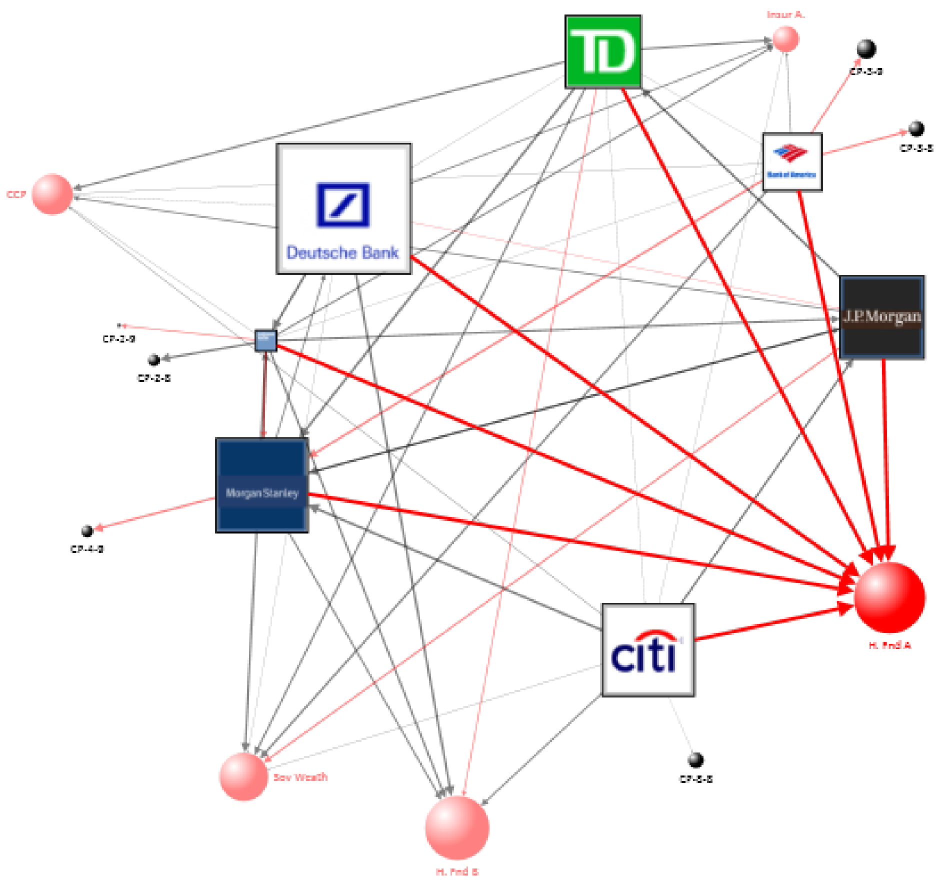

Figure 6 Example of network representation of stress test.

such a very different event, stress testing is still valuable in that, for the scenarios for which stress testing *was* done, the tests provided a measure protection against just those sorts of events, which is still better than not examining any stress scenarios at all.[33] Furthermore, the very act of preparing for and performing stress tests, provides the participants with training on how to think about dire scenarios.

5 Conclusions

The recent increased interest in stress testing among academics, regulators and practitioners has led to much new discussion of the topic. While there has been a sizeable increase in the volume of published research *describing* stress-testing approaches, there has been relatively little in the way of corresponding work on *theory*. This paper does little to change that trend. However, the intent has been to provide some basis for risk managers and regulators to think about the appropriate use of credit stress-testing exercises in the context of their broader activities.

Stress testing provides users with a rich palette to explore the impacts of changes in the state of the world on the financial performance of portfolios, institutions and the broader financial system. However, often, stress scenarios are drawn from the realm beyond historical data and the models estimated on it. This inherently requires that judgment be applied in the construction of scenarios, the evaluation of the probabilities of the scenarios, the linking of scenarios to losses and the analysis of the results of the stress tests themselves.

This is not necessarily a bad thing.

The exercise of constructing stress tests requires that the parties to the test engage in active discussion and analysis of all aspects of the stress scenarios, the models that translate them into portfolio or institutional losses and the interpretation of the results. As a qualitative component of a risk-management program, stress testing and scenario analysis provide an important complement to quantitative risk-management approaches. The accessible, intuitive nature a stress scenario and the resulting stress test result also provide a bridge between discussions of credit risk and discussions of strategy that senior management can use to evaluate the impact of different business options on their firm's risk appetite.

It is also reasonable to expect that the increased application of stress testing—both by institutions as part of a risk-management program and by regulators as a means to understanding the fragility of a single institution or the broader financial system—will lead to improvements in information technology, data quality and data infrastructure. These improvements will have benefits that extend beyond transparency and risk management.

Acknowledgments

I am grateful to Jeff Bohn, Ashish Das, Darrell Duffie, Mark Flood, Joe Langsam and an anonymous referee for detailed comments on earlier drafts of this paper. I had useful conversations on a number of the topics in this paper with Richard Cantor, Shirish Chinchalkar, Francis Gross, Juan Licari, Yaacov Mutnikas, and Mark Zandi. All errors are, of course, my own.

Appendix

Details of scenario orderings from example in Section 3.2.1

Recall that three 10-year home price stress scenarios in this example, are defined as follows:

A. *Slowdown in growth, but growth remains positive*: Home prices rise ½% each year over 10 years.

B. *Prices drop*: Home prices decline by 5% over 5 years (ending in year 5 at pre-decrease levels minus 5%). After year 5, prices rise at 4.5% per year.
C. *Prices drop severely*: Home prices decline by 25% over the first 3 years and then rise to pre-decrease levels minus 5% over the subsequent 2 years. After year 5, prices rise at 4.5% per year.

We also assume three homogenous portfolios of mortgages are defined as follows:

(1) A portfolio of 3/27 loans (loans that pay a low fixed coupon payment for the first 3 years and then convert to a floating coupon, with a typically higher interest payment).
(2) A portfolio of 5/25 loans (loans that pay a low fixed coupon payment for the first 5 years and then convert to a floating coupon, with a typically higher interest payment);
(3) A portfolio of 7/1 loans (loans that pay a low fixed coupon payment for the first 7 years and then convert to a floating coupon, with a typically higher interest payment) that resets each year.

Consider now how each scenario affects Portfolio 1 (3/27 loans) and Portfolio 2 (5/25 loans). For both portfolios, Scenario A is the least disruptive from a refinancing perspective. As the coupon reset approaches, loans in both portfolios have realized positive equity growth and (assuming refinancing makes sense from an interest rate environment standpoint) positive equity will allow them to refinance in order to avoid increased coupon payments.

The other scenarios are less clear:

- For Portfolio 1 (3/27 loans), scenario C (*prices drop severely*) is far more challenging than Scenario B (*prices drop*) and refinance risk will be higher under C than B. This is true because at the very time the interest payments on the mortgage are due to reset to a higher rate in year 3, the borrower has experienced substantial declines in the value of home equity. Home prices have dropped 25% since origination and for many borrowers, their mortgages will be "underwater." Thus, even though they would like to refinance, they may not be able to, due to the negative equity. This is shown in the upper left panel of Figure 1.

  In contrast, under Scenario B (*prices drop*) these same borrowers will have experienced a much smaller decrease in equity, making refinancing still sensible in many cases. For these borrowers, the sharp increase in coupon payments will be avoided.
- For Portfolio 2 (5/25 loans), both Scenario B and Scenario C affect losses similarly. This is because by the time the loans reset in year 5, home prices are the same level under both scenarios and they then move identically in both cases. This is shown in the upper right panel of Figure 1.

Now consider how the three scenarios affect Portfolio 3 (7/1 loans, shown in the lower left panel of Figure 1). In this case, either one of the "bad" scenarios (Scenario B or C) is slightly *preferable* to Scenario A (*slowdown in growth, but growth remains positive*). To see why consider that when the loans in the portfolio are due for rate resets in year 7, the home prices under Scenarios B and C will have experienced continued growth at 4.5% per year, which, starting from a 5% decline in year 5, puts the home prices at about 3.75% over the initial value. In contrast, under Scenario A, the 10 year growth has been a bit slower with 7 years of 0.5% growth resulting prices levels of about 3.5%. Thus, under Scenario A, the loans in Portfolio 3 will have experienced a bit less home price appreciation than under B or C.

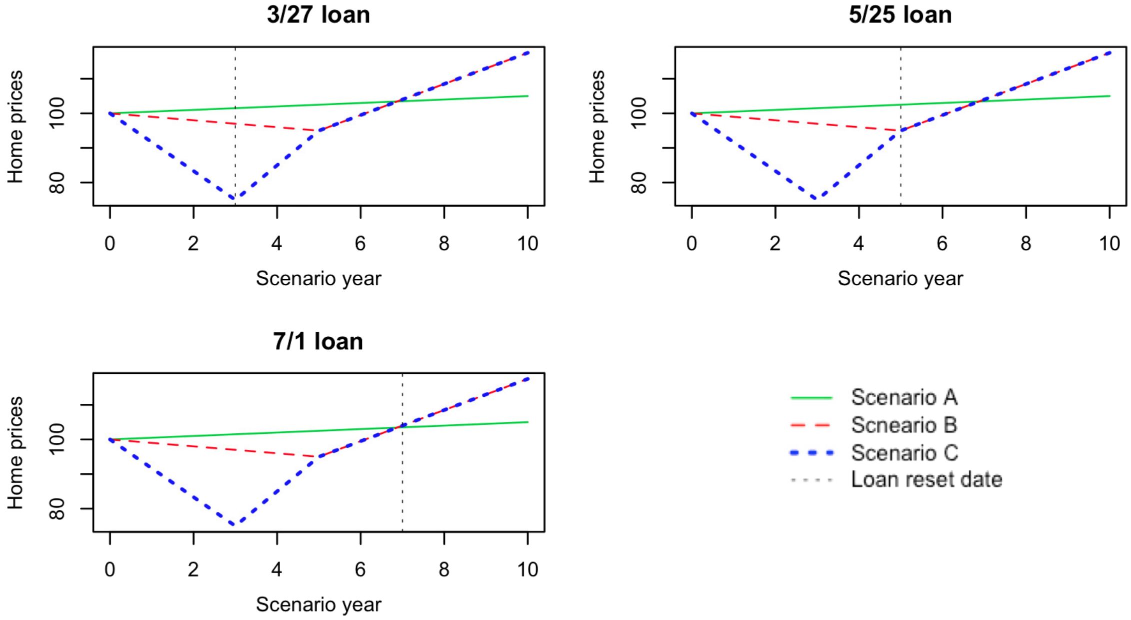

Figure 1 The impact of multiperiod stress scenarios on different loans.

Figure 1, reproduced below, shows these relationships.

Notes

[1] Note that when we refer to VaR, we are not referring to single-factor copula type models or percentiles of historical data, but rather to more sophisticated models that capture more fully the structure of individual assets and heterogeneous portfolios.

[2] Interestingly, Alfaro and Drehmann (2009) report on a study of stress tests done prior to the Crisis of 2007–9 and conclude that most stress tests were not adequate, since the majority did not raise any red flags with respect to banking system fragility.

[3] While we do not discuss it here, the alignment of incentives must naturally be addressed in any robust risk management program. Transparency and intuition, while useful in providing managers with insight into model behavior and usage, cannot prevent individuals from acting against the long-run interests of an institution. However, by increasing intuition for the risks being addressed in a particular scenario, stress-testing exercises make some of the assumptions underlying risk assessments more concrete and thus make the analyses themselves more accessible to both technical and non-technical managerial oversight.

[4] Berkowitz (2000) uses the term *pricing function* to describe a similar concept, though primarily in a market risk setting. We prefer the term *linking function* both for its statistical connotations and for its generality to a broad variety of asset behaviors (e.g., default likelihood, cashflow generation, prepayment likelihood, etc.) as well as to other more general cases such as the impact of an event on the revenue of a business segment.

[5] The author implicitly assumes that the composition of firms (with respect to factor sensitivities) remains constant and also that behavior in the stressed economic environment is a direct extrapolation from the unstressed behavior (the author alludes to this in Footnote 23 of the article).

[6] Note that Lopez (2005) terms these tests, *scenario* and *sensitivity* tests, respectively. Other authors use the term *sensitivity* to refer to any perturbing of key risk model parameters for purposes of better understanding the model.

[7] Note also that in some practical settings, regardless of how the stressed state of the assets is determined (through structural or reduced-form scenarios) it may be necessary to further translate the stressed state of the assets an ultimate portfolio loss through the use of some portfolio tool. This translation might be the case for an institution using a software tool that takes as input the asset state (e.g., PDs or CDR curves), and then translates these into losses through the application of functions

to calculate cash flows or lost interest under such scenarios.

8 Mathematically, we might write that the probability of an event e being worst than a particular event E_i, p_i^c, is not equivalent to the probability of a loss l being worse than the loss. L_i under event E_i, $p_i^l : p(e \prec E_i) = p_i^c \neq p_i^l = p(l > L_i)$.

9 For example, there are 50 states in the United States. Using the NAR index of median home prices, the change in State-level median home prices between 2005Q1 and 2010Q4 varied substantially from state to state, from a maximum decline of −57% (NV) to a 4% net increase (OK). Interestingly, the paths of the changes were also quite different, with some States hitting their lowest values in 2006Q1 and others still in decline as of 2010Q4. Clearly, assuming the national level home price change of −26% with a low in 2009Q2 would severely understate declines in some regions while overstating it in others. Depending on the exposures of a particular portfolio, this would make the losses more or less extreme.

10 If the number of factors in a scenario is small and the number of factor-states is similarly constrained, it is feasible, under the assumption that the factors fully span the space of all portfolio behaviors, to map out a fuller distribution of economic factors in a Bayesian context as in Rebonato (2010). However, even in this setting, the distribution will not necessarily order the scenarios with respect to the severity losses for a given portfolio.

11 Even here, it is easy to construct cases in which the severity of two simple scenarios is reversed for two portfolios. A trivial way to accomplish this is to create two portfolios on identical assets and to reverse the direction of the position (long or short) in the second portfolio for every asset in the first portfolio.

12 This example is based on one given in (Chinchalkar and Stein, 2010).

13 This is consistent with data on loan prepayments, which typically exhibit a major spike around the time of the rate reset. However, empirically, after month 12, there is often also a much more modest increase in prepayments as borrowers with lower FICO scores take advantage of their newly established track record of mortgage payments to re-negotiate.

14 For example, the default probability of a 5/25 mortgage with a 70% LTV is clearly affected by a 25% decline in home prices differently than a mortgage with an LTV of 85%. After the decline, the borrower on the 70% LTV loan still retains approximately 5% equity in the home, while the borrower on the 85% LTV home is underwater with significant negative borrower's equity of approximately −10%. This difference materially affects the borrower's propensity to default, as do numerous other loan and borrower-attributes.

15 Here again, we assume strictly long-only or short-only portfolios.

16 Extending the mortgage example a bit more demonstrates this point. Mortgage defaults can be driven not only by home-price changes, but also by the levels and changes of unemployment and interest rates.

17 The scenario, provided by Moody's Economy.com, was a severe downturn stress case. It contained national-level interest rates and home price and unemployment series forecasts for several hundred local regions.

18 The models use macro-economic simulation to generate loss distributions and are described more fully in (Stein *et al.*, 2010).

19 We also estimated 95% bootstrap confidence bounds on the estimates as (0.0278, 0.0346), (0.0079, 0.0118) and (0.0003, 0.0016), respectively, suggesting that the observed differences are reasonably robust.

20 The "jingle" refers to the notion that defaulting borrowers would simply drop the keys to a property into an envelope and mail them to the lender before walking away from their mortgage.

21 A recent study (Guiso *et al.*, 2009) provides an analysis of survey respondents' reported willingness to default on a mortgage, should it be economically sensible to do so. The authors find that "the most important variables in predicting the likelihood of a strategic default are moral and social considerations. Social considerations are directly affected by the frequency of foreclosures (in the same zip code) and the probability that somebody knows somebody else who strategically defaulted."

22 The notion of plausibility was particularly actively discussed in the Artificial Intelligence community. For early examples, see Nilsson (1986) or Collins and Michalski (1989) for logical frameworks for calculating plausibility; or Shafer (1976) for an approach to calculating plausibility as an upper bound on probability in evidential reasoning.

23 I am indebted to Mark Flood for suggesting that this topic be included in this paper and for useful discussions on the topic.

24 See Bohn and Stein (2011) for a discussion of transfer pricing as a link between managing the risk of a credit portfolio and the risk of a financial services franchise.

[25] The macro-economic scenarios used in this example were provided by Moody's Economy.com and the various loss distributions were generated using Mortgage Portfolio Analyzer.

[26] Berkowitz (2000) provides an approach to combining stressed distributions with a non-stressed distribution subject to an assumed scenario probability. This can be generalized to multiple stress scenarios, provided probabilities were known for each of the stress scenarios. However, this implicitly assumes that the universe of all stress scenarios is enumerated, which will typically be challenging. (Not explicitly including a particular stress scenario implies assigning a zero probability to it.)

[27] Because the usefulness of such indicators depends heavily on the quality and characteristics of the linking functions, such analyses may best be considered warning flags, rather that being viewed the sole means for performing such analysis.

[28] I am grateful to Joe Langsam for this example.

[29] This may be more a statement about the current state of data infrastructure and availability in financial institutions than an immutable tenant. With appropriate data infrastructure, standards and ontology, in principle it would be possible to run much more detailed scenarios. There have been proposals that move in the direction of more common and standardized data infrastructure (Flood, 2009; Gross, 2010), though implementations of these may take some time.

[30] The approach was actually originally presented informally in 2007.

[31] For examples of applying network analysis using only publicly available data, see (Billio, Getmansky, Lo and Pelizzon, 2010).

[32] Indeed, in hindsight, it is interesting to note that over 85% of the banks reporting in a pre-crisis survey did not indicate that they performed residential real-estate-specific stress tests (Committee on the Global Financial System, 2005).

[33] Darrell Duffie first brought this observation to my attention.

References

Alfaro, R. and Drehmann, M. (2009, December). "Macro stress tests and crises: what can we learn?" *BIS Quarterly Review*.

Basurto, M. A. and Padilla, P. (2006). *Portfolio Credit Risk and Macroeconomic Shocks: Applications to Stress Testing Under Data-Restricted Environments.* International Monetary Fund, Monetary and Capital Markets Department. International Monetary Fund.

Berkowitz, J. (2000). "A Coherent Framework for Stress-testing." *Journal of Risk* **2**, 1–11.

Billo, M., Getmansky, M., Lo, A. W. and Pelizzon, L. (2010). "Measuring Systemic Risk in the Finance and Insurance Sectors", MIT Sloan School Working Paper 4774-10.

Bohn, J. R. and Stein, R. M. (2009). *Active Credit Portfolio Management in Practice.* Hobokken, NJ: Wiley.

Bohn, J. R. and Stein, R. M. (2011). *Some Observations on Improving a Bank's Share Value with Credit Portfolio Management, Credit-Transfer Pricing and Stress Testing.* New York, NY. URL: available at www.moodysresearchlabs.com

Breuer, T., Jandačka, M., Rheinberger, K., and Summer, M. (2008). "Macro Stress and Worst Case Analysis of Loan Portfolios." Working Paper, Research Centre PPE.

Chinchalkar, S. and Stein, R. M. (2010). *A Comparisson of Loan-Level and Pool-Level Mortgage Analysis.* New York, NY: Moody's Research Labs.

Collins, A. and Michalski, R. (1989). "The Logic of Plausible Reasoning: A Core Theory." *Cognitive Science* **13**, 1–49.

Committee on the Global Financial System. (2005). *Stress Testing at Major Financial Institutions: Survey Results and Practice.* Bank for International Settlements, Committee on the Global Financial System. Bank for International Settlements.

Drehmann, M. (2008). Stress tests: objectives, challenges and modelling choices. *Riksbank Economic Review* (June).

Duffie, D. (2010). *Systemic Risk Exposures: A 10-by-10-by-10 Approach.* Graduate School of Business, Stanford University.

Flood, M. D. and Korenko, G. G. (2010). *Systemic Scenario Selection: A Methodology for Selecting a Representative Grid of Shock Scenarios from a Multivariaiate Elliptical Distribution.* Washington, DC: Federal Housing Finance Agency.

Flood, M. (2009). "Embracing Change: Financial Informatics and Risk Analytics." *Quantitative Finanace* **9**(3), 243–256.

Financial Services Authority. (2008). "Stress and Scenario Testing." Consultation Paper, Financial Services Authority, London.

Gross, F. (2010 йил 15-16-December). International Reference Data Utility: a necessary infrastructure for measuring systemic risk. *Presentation at the Measuring Systemic Risk Conference—The Milton Friedman Institute*. Chicago. URL: http://mfi.uchicago.edu/events/20101215_systemicrisk/ppts/12.16.2010_FGross.ppt

Guiso, L., Sapienza, P., and Zingales, L. (2009). *Moral and Social Constraints to Strategic Default on Mortgages.* European University Institute.

Howard, S. (2008). *Stress Testing with Incomplete Data: A Practical Guide.* IFC Bulletin, IFC.

Lopez, J. A. (2005). "Stress Tests: Useful Complements to Financial Risk Models." *FRBSF Economic Letter* (No. 2005-14).

Nilsson, N. J. (1986). "Probabilistic logic." *Artificial Intelligence* **28**(1), 71–87.

Otani, A., Shiratsuka, S., Tsurui, R., and Yamadaa, T. (2009). *Macro Stress-Testing on the Loan Portfolio of Japanese Banks.* Bank of Japan, Financial Systems and Bank Examination Department.

Rebonato, R. (2010). *Coherent Stress Testing: A Bayesian Approach to the Analysis of Financial Stress.* West Sussex, UK: John Wiley & Sons, Ltd.

Rösch, D. and Scheule, H. (2007). "Stress-Testing Credit Risk Parameters—An Application to Retail Loan Portfolios." Working Paper.

Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton: Princeton University Press.

Stein, R. M., Das, A., Ding, Y., and Chinchalkar, S. (2010). *Moody's Mortgage Metrics Prime: A Quasi-Structural Model of Prime Mortgage Portfolio Losses.* New York, NY: Moody's Research Labs.

Tversky, A. and Kahneman, D. (1986). "Rational Choice and the Framing of Decisions." *The Journal of Business* **54**(1), S251–S278.

Virolainen, K. (2004). *Macro Stress Testing with a Macroeconomic Credit Risk Model for Finland.* Bank of Finland, Research Department.

Zandi, M. and Blinder, A. (2010). *How the Great Recession was Brought to an End.* West Chester, PA: Moody's Analytics.

Keywords: Stress testing, scenario analysis, portfolio analysis, systemic risk, model risk, VaR, macroeconomic analysis, credit risk