IMF Working Paper

A New Heuristic Measure of Fragility and Tail Risks: Application to Stress Testing

Nassim N. Taleb, Elie Canetti, Tidiane Kinda,
Elena Loukoianova, and
Christian Schmieder

INTERNATIONAL MONETARY FUND

WP/12/216

IMF Working Paper

Monetary and Capital Markets Department

A New Heuristic Measure of Fragility and Tail Risks: Application to Stress Testing[1]

Prepared by Nassim N. Taleb, Elie Canetti, Tidiane Kinda, Elena Loukoianova, Christian Schmieder

Authorized for distribution by Christopher Towe

August 2012

Abstract

This paper presents a simple heuristic measure of tail risk, which is applied to individual bank stress tests and to public debt. Stress testing can be seen as a first order test of the level of potential negative outcomes in response to tail shocks. However, the results of stress testing can be misleading in the presence of model error and the uncertainty attending parameters and their estimation. The heuristic can be seen as a second order stress test to detect nonlinearities in the tails that can lead to fragility, i.e., provide additional information on the robustness of stress tests. It also shows how the measure can be used to assess the robustness of public debt forecasts, an important issue in many countries. The heuristic measure outlined here can be used in a variety of situations to ascertain an ordinal ranking of fragility to tail risks.

---

[1] This paper benefited from comments by Gianni De Nicolo and Christopher Towe.

Contents   Page

I. INTRODUCTION

Much of the history of the financial crisis can be interpreted broadly as an underestimation of risks, not only of the probability of Black Swans (large impact, unforeseen, random events), but of the financial system's fragility to them. For example, the potential for bank losses and disruption of bank funding markets due to deterioration in the U.S. subprime housing market, or the potential for widespread sovereign and banking distress in the Euro-area triggered by stresses in Greek sovereign finances, were both severely underappreciated.[2]

A great deal of soul-searching since has centered around excessive reliance on financial and economic models that are seen to have led policymakers and financial markets astray, in part by giving an unwarranted level of confidence about the potential size of downside outcomes. Most of this effort has been aimed at, on the one hand, scrapping models altogether, or on the other, seeking to design and develop better models.

This paper stakes out an intermediate ground, showing how existing stress testing models can be used to develop a simple measure of tail risk by taking into account convexity effects. Economic and financial models have limitations and constraints, including misspecification, estimation using assumed and sometimes inaccurate probability distributions, etc. As a result, using such models to estimate the potential impact of shocks may lead to increasingly inaccurate estimates the further in the tails such shocks are. Even when testing is undertaken to try to detect the sensitivity of an outcome to different sized shocks (sensitivity-testing), the focus tends to be on the range of possible levels of outcomes.

In this paper, we apply to stress testing a simple heuristic method proposed by Taleb (2011), —based on methods to detect hidden exposures to volatility in option trading portfolios. This method allows us to evaluate how well tail risks are captured by stress tests. Rather than using point estimates (or ranges) of outcome levels, this method calculates the difference between outcomes to look for potential convexities which, if ignored, might lead stress testers to underestimate (or, likely in fewer cases, overestimate) the impact of tail events. The very simplicity of this heuristic is seen as a virtue. It should focus attention on potential non-linearities in the tails that are rarely given prominence, and yet should be easily understood and taken into account once made explicit.

These non-linear (convexity) effects in the tails can cause, for instance, financial losses or sovereign debt unexpectedly to "blow up" in response to shocks that are only a little larger than anticipated and therefore remain invisible. Even if the shocks are correctly foreseen, model error or parameter uncertainty may also lead to substantial underestimation of risks.

---

[2] Arguably, these were not "Black Swan" events as both sub-prime losses and Greek sovereign distress were in principle foreseeable, but in both cases, the magnitude of the outcomes would at least have been regarded as fairly extreme tail events a year or two prior to their full flowering.

However, with the heuristic, even though one may not be able to rely on estimated levels of potential losses from a given model, looking at how these levels vary in small ranges around the shock being tested should give a sense of whether the difference in loss estimates is growing (or possibly tamping down) as one moves further out the tail of adverse shocks.

Taleb (2011) gives a simple analogy how such a second order test can provide valuable information even when a measuring tool may be flawed. Using an inaccurate tape measure will give a false reading of a child's height (a level measurement). However, if one uses the same tape measure over time, it will give a reliable test of whether the child is growing (a second-order measurement). By the same token, most economic and financial models have limitations, but looking at the differences in estimated outcomes from any given model will be robust under fairly general conditions, thus pointing the stress tester in the right direction.[3]

More specifically, the proposed heuristic takes advantage of a version of Jensen's inequality (applied to higher order terms) to detect convexity in the tails. The idea is to take small, equal-sized perturbations around the results of a tail risk test and see whether the differences in risk measurements indicate convexity (or linearity or concavity). The degree of convexity can then be used as a direct measure of tail risk.[4] In other words, if the estimated models are wrong, the levels of corresponding estimated losses may also be substantially wrong, but the heuristic measure should give a reasonable relative ranking of the "fragility" to such convexities.[5]

In this sense, this heuristic can be seen as a way to elaborate the use of stress testing to develop a more robust measure of relative fragilities, since it should capture the convexity of a loss function in the tails. In sum, the heuristic shows how a firm or government can be exposed to the underestimation of a certain set of tail risks and, what is critical, how vulnerable it can be to model error.

The paper is organized as follows: Section II first provides some concepts and methods to assess fragility in general and then elaborates the proposed heuristic. Section III presents two case studies applying the heuristic to the outcome of bank and fiscal stress tests, respectively. Section IV gives an overview of how the heuristic could be used in stress testing applications. Section V concludes.

---

[3] These conditions include that loss distributions are monomodal, that the bias in the incorrect model (compared to the true model) does not change signs, and higher differences do not carry opposite signs.

[4] The G-20 has called for "the IMF to investigate, develop, and encourage implementation of standard measures that can provide information on tail risks." See IMF/FSB (2009), recommendation #3.

[5] Taleb and Douady (2012) develop a theorem that proves how a nonlinear exposure maps into tail-sensitivity to volatility and model error and produce a transfer function expressing fragility as a direct result of nonlinearity.

II. REVIEW OF CONCEPTS TO ASSESS FRAGILITY

A. The Current State of Stress Testing

The crisis revealed weaknesses in the stress testing exercises performed on financial systems and institutions in several countries, leading the IMF and country financial regulatory authorities recently to pay more attention to stress testing and to overhaul existing methodologies. Specifically, the crisis demonstrated that stress tests with poorly designed scenarios, omitted shocks, based on inappropriate methods or a narrow coverage of institutions, etc., can produce results that provide a false reassurance about the degree of financial stability.[6]

However, there is a more fundamental issue with stress testing, especially seemingly more sophisticated ones: First, many stress tests focus on the point estimates of very few scenarios, and often pay little attention to how the impact would change in case of different scenarios, e.g., a slightly more severe one.[7] Second, if stress tests do not take into account the possibility of model and parameter error, it can be misleading to rely only on the point estimates of even well-designed stress tests. Without considering the potential for these errors, one could miss the convexities/non-linearities that can lead to serious financial fragilities.

The main focus of stress testing has recently been expanding from solvency and market risk toward liquidity and contagion risks. Compared to solvency stress tests, liquidity stress tests are less developed for several reasons, such as: (i) that liquidity risk was seen as "less of a critical issue" until the recent global crisis; (ii) that liquidity crises were seen as low frequency, but potentially high-impact events; and (iii) that to some extent, all liquidity crises were seen as unique.

Contagion risk stress tests require more elaborated data on individual cross-exposures on the interbank market and across broad financial and economic sectors, in order to build up maps of interconnectedness. This type of data is only now starting to be collected.[8] Thus, even if some network analysis models are developed, data limitations may prevent assessing contagion and interconnectivity risks. However, one of the lessons of this paper is that, by increasing model complexity, the risks of model error increase, which may render stress testing even more vulnerable to the criticism that losses in response to tail events are likely to be severely misestimated.

---

[6] The most notable example is in Iceland, where stress tests were performed just before its liquidity crisis (Ong and Cihak, 2010). See also Borio, Drehmann and Tsatsaronis (2012) for a critical review of the early warning properties of stress tests.

[7] Given the conceivably virtually unlimited number of dimensions that could be covered by stress tests, scenarios will hardly ever be realized precisely as assumed by the stress tests.

[8] See, for example, FSB (2011).

In sum, despite various elaborations of stress testing techniques in the aftermath of the global financial crisis, as well as rethinking about the potential magnitude of shocks, it is still an open question whether tail risks are being captured correctly. Do we estimate risks properly in the face of potential model error, parameter stochasticity and incorrect distributions? How would our estimates respond to a marginal change of the stress scenario?

B. A Simple Heuristic to Detect Fragility

Following Taleb (2011), the heuristic offered below seeks to detect "biases from missed nonlinearities and detection of these using a single 'fast-and-frugal,' model-free, probability free heuristic." Imagine a payoff structure as shown in Figure 1 below. With a linear payoff, the "harm" of an adverse shock is proportional to the size of the shock. But a payoff with concavity (negative convexity) becomes disproportionately larger as the shock (event size) becomes larger. [9] With particularly large "Black Swan" type events, the difference in harm between a linear and negatively convex payoff can escalate exponentially.

Figure 1. Why the Concave is Hurt by Tail Events
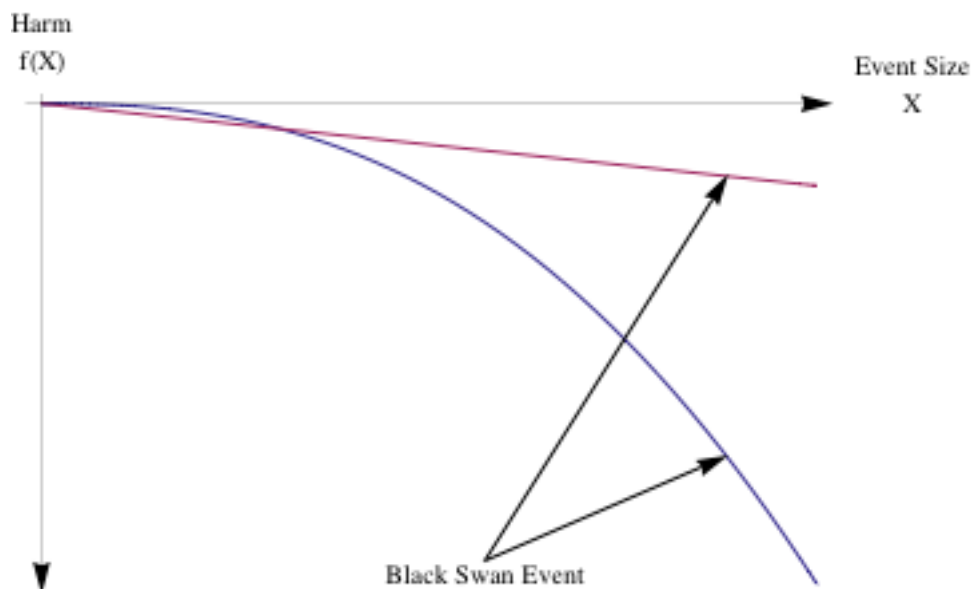
Source: Authors.

Such negative convexity effects are quite frequent in economic and financial situations. For instance, they may result from size. The French bank Société Generale was faced with the

---

[9] Note that if exposure to an event is negative (e.g., a short position), then the concave payoff structure shown in the diagram would actually become convex, and in fact would be exactly the negative of the concave payoff function.

necessity to dump roughly US$70 billion of stock index futures over three days upon the discovery of Jerome Kerviel's rogue trading positions in January 2008.[10] The large size of the fire sale relative to the size of the market forced Société Generale to realize a particularly large loss because the sale itself caused a particularly adverse price reaction. Had the stock futures been dumped, instead, in ten US$7 billion increments over a period, say, of several weeks or months, the effect on prices may have been relatively small. Instead, the price effects of the rapid forced fire sale caused by the size of Société Generale's positions meant that the losses resulting from one US$70 billion sale were far larger than would have been the losses from 10 sales of US$7 billion each.

Negative convexity effects can also be produced by positive (i.e., reinforcing) feedback effects resulting from complexity and interconnectedness of markets. A financial institution can likely face small day to day price variations with relatively little impact on its overall financial position. But if there is a particularly large price variation in a significant position, the financial institution may be forced to sell some of the position (e.g., in order to meet capital requirements or redemption requests). If that sale causes market prices to fall, then further losses can require the financial institution (and other financial institutions) to liquidate even more positions, causing further losses and further liquidations and so on. As a result, the total impact of a large and sudden price decline may be many orders of magnitude larger than the impact of a series of smaller price losses that, over time, amount to the same total price decline. Similar examples could be adduced from fiscal dynamics, where higher debt or rollover requirements lead to higher financing costs that raise debt or rollovers, potentially leading to an out of control debt spiral.

Finally, negative convexity effects may actually be produced by regulation, because of incentives for traders to "hide risks in the tails." For example, if a regulator requires capital to be set aside against the possibility of an adverse market price move, calculated according to a value at risk model at, say, 5 percent probability, a trade that has a large payoff for a tail shock with a 4 percent probability may fall outside the regulator's view, thereby providing traders with an asymmetric incentive to place such a trade.[11]

A simple point estimate from a stress test gives no sense of such potential convexity effects. Such an estimate will be based on a single choice of shock, which can be thought of as applying the model to an average shock in the tails. The heuristic involves averaging the model results over a range of shocks. When convexity effects are present, the average of the

---

[10] According to an investigation into the scandal by Société Generale's own General Inspection department, a €49 billion (US$71 billion) long position on index futures was discovered on January 20 then unwound between January 21 and January 23, leading to gross losses of EUR 6.4 billion. See Mission Green: Summary Report, Société Generale General Inspection Department
http://www.societegenerale.com/sites/default/files/documents/Green_VA.pdf.

[11] Traders' compensation systems generally provide extremely asymmetrical incentives since traders will receive large bonuses for highly risky trades that pay off, but an equivalent amount cannot be clawed back from them should the trade result in large losses, since their salaries will be bounded at zero. Such asymmetric incentives can lead traders to take on more risk than they would if there were a symmetrical incentive scheme.

model results will not be equal to the model results of the average shock. The heuristic is a scalar that measures the extent of that deviation, and is calculated as H, where:

$$H = \frac{f(\alpha - \Delta) + f(\alpha + \Delta)}{2} - f(\alpha)$$

f(x) is the profit or loss for a certain level α in the state variable concerned, or a general vector if we are concerned with higher dimensional cases. Δ is a change in α, a certain multiple of the mean deviation of the variable. The severity of the convexity expressed by H should be interpreted in relation to the total capital (for a bank stress test, or GDP for a sovereign debt stress debt), and can be scaled by it, allowing for comparability of results, and hence an ordinal ranking of fragilities, among similar types of institutions. When *H=0* (or a small share of the total capital) the outcome is robust, in the sense that the payoff function is linear and the potential gain from a smaller (by the amount Δ) x is equal to the potential loss from an equivalently sized larger x. When *H <0*, and significantly so with respect to capital, the outcome is fragile, in the sense that the additional losses with a small unfavorable shock (i.e., compared to a given tail outcome) will be much larger than the additional gains with a small favorable shock.[12] Thus, volatility is bad in such a situation; i.e., we can say that an institution for which H is negative is "fragile" to higher volatility.[13]

C. How Can the Simple Heuristic Enhance Stress Tests?

The heuristic provides a technique to assess how non-linear tail risks are and thereby to assess the sensitivity of the outcome of stress tests vis-à-vis different risk drivers, in a way that is more robust to model error. As macro stress tests are usually based on a very limited number of stress scenarios, the heuristic enhances the scope of macro stress tests with limited additional effort and thereby fills an important gap in the stress testing toolkit.

As shown in Figure 1, the outcome of solvency stress tests, measured in terms of changes in capitalization, is usually highly skewed to the left, i.e., there is a limited chance that capitalization will go up significantly, while there is considerable risk that capitalization will drop sharply in response to a stress (fat tail). Most macro stress tests explore a very limited "area" of the distribution function, often limited to a baseline scenario and a few stress

---

[12] Note that if the stress test involves something where a larger result represents the adverse case, such as in the change in the net debt/GDP ratio examined in Section III.C, fragility will be represented by H > 0.

[13] Taleb (2012, forthcoming) posits the opposite situation. If we start with profits, and H>0, then greater volatility leads to a more profitable outcome than lower volatility. Such a situation is termed "antifragility." This is not the same as robustness, since with robustness, higher volatility provides neither significant harm nor benefit.

scenarios, whereby one computes a few point estimates, but the sensitivity of the outcome to changes in key risk drivers remains hidden.[14]

As the distribution tends to be particularly non-linear in the tails, which is when banks (and systems) come close to the brink of failure, it is essential to understand these non-linearities. Rather than running a series of additional scenarios, the heuristic allows testing the sensitivity of the outcome in an efficient and robust manner.[15] As such, the heuristic could be used as a standardized method to measure tail risks.

The outcome of the heuristic in the box on the left hand side of Figure 2 is illustrated further in Figure 3, which displays the possible non-linearities that would be tested by the simple heuristic.[16]

III. THE HEURISTIC APPLIED TO THE OUTCOME OF STRESS TESTS
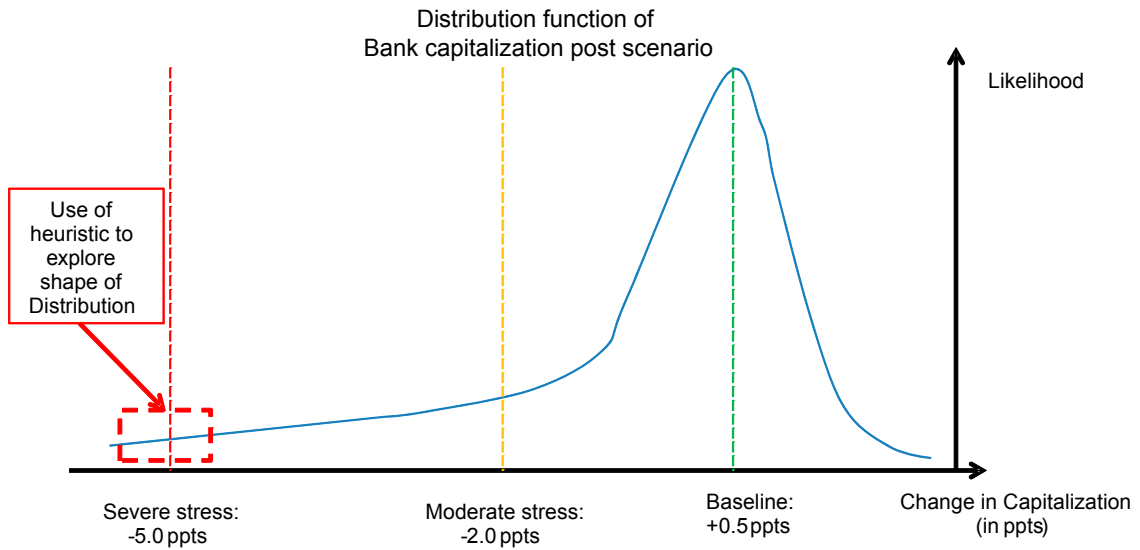
A. Purpose for the Use of the Heuristic

As outlined above, macroeconomic stress tests are usually limited to the computation of a small number of scenarios (e.g., the point estimates in Figure 2 represented by the vertical lines), which are then used to draw policy conclusions. The drawback of this procedure is that the sensitivities of the outcome to small changes in inputs, although possibly something the team actually running the test may have a feel for, remains hidden for the wider audience, notably including the decision-makers that use the results. Ultimately, the outcome of the scenario is thereby taken at face value. In fact, stress test results are often presented as having

---

[14] For example, the most prominent recent bank stress tests, the Supervisory Capital Assessment Program (SCAP) and the subsequent Comprehensive Capital Analysis and Review in the United States, and the two sets of published stress tests conducted by the European Banking Authority (EBA) used only a baseline scenario and one adverse stress scenario.

[15] Of course, ideally, losses would be derived in a closed-form expression that would allow the stress tester to trace out the complete arc of losses as a function of the state variables, but it is exceedingly unlikely that such a closed-form expression could be tractably derived, hence, the need for the simplifying heuristic.
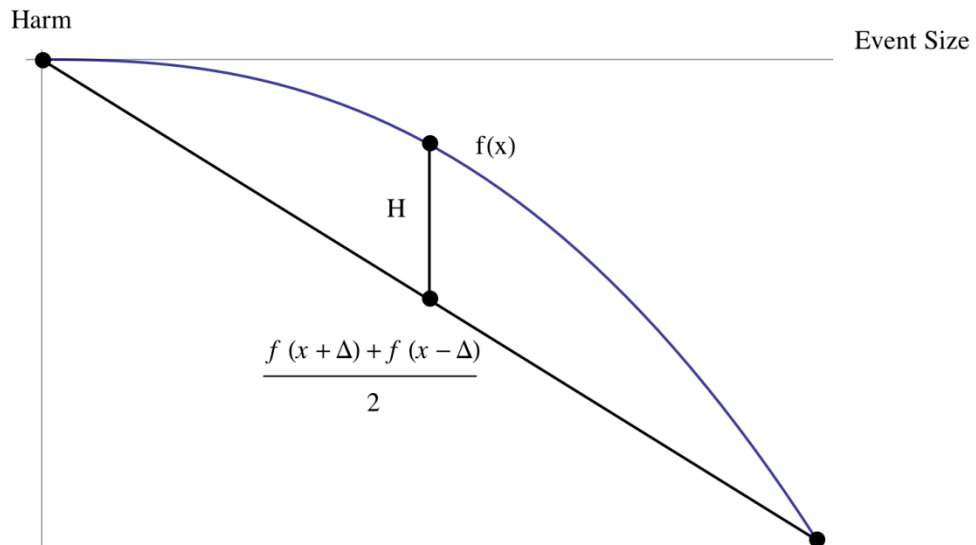
[16] The mathematical logic that (1) the heuristic reveals tail "fragility" and (2) that it also reveals model error is demonstrated in Taleb and Douady (2012) as follows. The definition of fragility below a certain level K is the sensitivity of the tail integral—the partial tail expectation—between minus infinity and K to changes in parameters, particularly the lower mean deviation. By the "fragility transfer theorem", such sensitivity for a variable Y is caused by the second derivative of the function φ, such that $Y=\varphi(x)$ hence a direct result of the convexity of such function. By the "fragility exacerbation theorem", the increase in fragility is mapped as a direct effect of such convexity. So it becomes a matter of detecting the convexity, hence the heuristic. Furthermore, parameter imprecisions in a model are considered as fragilizing if they are capable of causing an increase in the left tail. Taleb and Douady (2012) also shows that the convexity bias, that is the mis-estimation of the effect of Jensen's inequality can be obtained by setting K at Infinity, to take the effect of the convexity on the total expectation.

Figure 2. Illustration of the Use of the Heuristic

Distribution function of
Bank capitalization post scenario

Likelihood

Use of
heuristic to
explore
shape of
Distribution

Severe stress:
-5.0 ppts

Moderate stress:
-2.0 ppts

Baseline:
+0.5 ppts

Change in Capitalization
(in ppts)

Source: Authors.

Figure 3: Fragile and Antifragile Outcomes of Stress Tests

Harm

Event Size

f(x)

H

$$\frac{f\,(x+\Delta)+f\,(x-\Delta)}{2}$$

a binary "pass/fail" outcome. However, as financial risks are usually highly non-linear, drawing policy-conclusions based on a very few point estimates can produce misleading conclusions. Rather than presenting a series of additional outcomes to policy-makers, the simple heuristic summarizes the fragility of the result in a single number. Moreover, its very simplicity forces the observer directly to confront the possibility (likelihood?) of error in the level estimates, and the asymmetrical costs of such inaccuracies.

B.   Case Study I: The Simple Heuristic Applied to Bank Stress Tests

We use the outcome of a stress test for a sample of 12 large U.S. banks to illustrate the functioning of the heuristic. The tests were performed based on the framework developed by Schmieder, Puhr and Hasan (2011).The stress test was on projected end-2011 data, using end-2010 balance sheet data, complemented with bank data from the first half of 2011. Stresses were applied for the period from 2012-2016. The outcome of stress was measured in terms of Tier 1 capitalization.

The test projected one single macro scenario, namely a near-zero GDP path with a cumulative deviation from the WEO baseline by 10 percentage points during 2012–16. In historical terms for advanced countries (based on the period from 1980–2010) this scenario could be expected to occur with a likelihood of about 4 percent.[17] The scenario was simulated including a feedback loop between stress in the banking system and macroeconomic growth, i.e., banks' capital needs and the pertinent deleveraging was simulated to have a negative impact on output, based on Vitek and Bayoumi (2011).[18] Together with additional stress elements, especially with respect to losses of trading income and increases of funding costs under stress, the scenario constitutes a highly adverse tail risk scenario.

The impact of macroeconomic stress on key financial risk drivers, namely credit losses, credit growth and pre-impairment income, have been estimated using so-called "satellite models" using panel regressions. The test assumes no recapitalization other than through retained earnings and allows for deleveraging in case of stress. Further details on the test are provided in Appendix I.

The heuristic was applied to bank capitalization, first, separately for each of the specific risk drivers (credit growth, credit losses, trading income) and then all three together with the impact of GDP growth (Table 1).

---

[17] More specifically, scenarios with a cumulative deviation from average growth rates by 10 percentage points (independent from when the deviation occurs) have been used to compute the likelihood for the occurrence of the zero-growth scenario.

[18] One of the virtues of the heuristic is that it may reveal non-linearities in the stress testing model, e.g. arising from feedback loops, even when such non-linearities were not explicitly or intentionally built into the stress testing model. This could reveal either a true non-linearity, or a need to refine the model.

The outcome indicates that the tail stress test produces non-linear results in the majority of cases (Table 1). For most banks, the outcome is fragile with respect to all of the risk drivers, namely macroeconomic stress (GDP growth change,[19] scenarios 1–4), credit growth (5–8), credit losses (9–12), and income (13–16), as well as to trading income as part of total income (21–24).[20] The analysis also includes a combined scenario that stresses credit growth, credit losses and income by one additional mean deviation (17–20). Fragility is especially high for the banks with the worst outcomes, i.e., those experiencing the most substantial drop of Tier 1 capitalization under stress (row 0), where all results are fragile (see also Table 2). On the other hand, although Bank 3 and 4 would not be found to be the most vulnerable in terms of the impact of stress on capitalization (capitalization in 2016 vs. pre-stress capitalization, row 0) and the conclusion might therefore be that they are resilient, the heuristic reveals fragilities for those banks.

However, there are also some antifragile outcomes, reflecting risk mitigating effects such as a drop of RWAs (through credit losses), deleveraging that helps to mitigate stress, as well as other factors such as banks' relative levels of losses and income (i.e., risk and return) and other effects. These cases are instances of deleveraging (due to a reduction in assets under stress as a result of credit losses), which has a positive and non-linear marginal impact on the risk profile of a bank under stress.[21]

In addition to information whether the tail is non-linear as such, the ratio can also provide information on how non-linear it is, and thereby allows for ordinal comparisons. The outcome allows computing the additional impact of a further drop of GDP growth by 1 (or 2) mean absolute deviations (MDev) on various key drivers of bank capitalization, for example.

Table 2 sheds some more light on the fragility of banks in relative terms. To do so, banks are ordered according to (i) the impact of the adverse scenario in terms of capitalization (the outcome of the actual standard stress tests, shown in row 0 in Table 1), and the heuristic for (ii) scenarios 2/4 and (iii) scenarios 18/20. The latter two cases are assumed to be a proxy for the overall fragility of the banks.

---

[19] GDP affects credit losses, income and credit growth through the satellite models, which makes scenarios 1–4 complementary to the scenarios 17–20.

[20] By way of a numerical example, the heuristic for Bank 1 under the 1 standard deviation GDP shock (-0.035) is computed as follows. Under the stress test, the Tier 1 ratio is 1.639 percent and under additional GDP shocks of +/- one standard deviation is .2.835 and . 373, respectively. Thus, the heuristic is calculated as: (0.373 +2.835)/2-1.639 = -0.035, the same as the computation using the changes of capitalization as shown in Table 1: (-1.27–1.2)/2.

[21] This impact explains the positive H for banks 5, 8, 9, and 11 in the case of the GDP shock.

Table 1. The Heuristic Applied to the Outcome of Macroeconomic Stress Tests for the Largest U.S. Banks

Table shows change in T1 capitalization conditional on the scenarios	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6	Bank 7	Bank 8	Bank 9	Bank 10	Bank 11	Bank 12
Scenario												
Impact of adverse scenario	-9.60	-13.65	-8.48	-2.82	-8.90	-15.07	-1.61	-0.56	2.27	-18.62	-13.31	-7.63
GDP growth 1 MDev worse (1.5 percentage points, cumulative)	-1.27	-2.71	-1.86	-0.80	-0.11	-0.24	-0.71	-0.29	-0.12	-2.01	-1.63	-1.67
GDP growth 2 MDev worse (3.0 percentage points, cumulative)	-2.62	-5.79	-4.01	-1.94	-0.23	-0.48	-1.48	-0.63	-0.24	-4.25	-3.22	-3.49
GDP growth 1 MDev better (1.5 percentage points, cumulative)	1.20	2.52	1.28	0.76	0.11	0.24	0.58	0.31	0.12	1.85	1.68	1.56
GDP growth 2 MDev better (3.0 percentage points, cumulative)	2.34	4.88	2.28	1.48	0.23	0.43	1.16	0.63	0.24	3.67	3.42	3.06
H +/- 1 MDev	-0.035	-0.096	-0.294	-0.019	0.000	-0.001	-0.066	0.007	0.000	-0.080	0.025	-0.053
H +/- 2 MDev	-0.141	-0.451	-0.863	-0.230	0.000	-0.025	-0.162	0.000	0.001	-0.293	0.097	-0.216
Credit growth 1 MDev more (3.4 percentage points)	-0.99	-1.71	-1.40	-0.62	-0.03	-0.09	-0.46	-0.28	-0.02	-1.61	-1.51	-1.34
Credit growth 2 MDevs more (6.8 percentage points)	-2.07	-3.63	-2.96	-1.57	-0.06	-0.19	-1.16	-0.57	-0.05	-3.39	-3.13	-2.82
Credit growth 1 Mdev lower(3.4 percentage points)	0.90	1.53	0.86	0.47	0.03	0.09	0.38	0.26	0.02	1.46	1.41	1.22
Credit growth 2 MDevs lower (6.8 percentage points)	1.73	2.96	1.43	0.91	0.05	0.17	0.73	0.50	0.04	2.71	2.73	2.33
H +/- 1 MDev	-0.043	-0.090	-0.269	-0.076	-0.001	-0.003	-0.039	-0.009	-0.001	-0.076	-0.049	-0.061
H +/- 2 MDev	-0.171	-0.334	-0.764	-0.332	-0.003	-0.011	-0.214	-0.035	-0.003	-0.341	-0.198	-0.244
Credit Losses 1 MDev higher (0.11 percentage points)	-1.21	-2.45	-1.68	-0.68	-0.04	-0.13	-0.60	-0.24	-0.02	-2.07	-1.62	-1.55
Credit Losses 2 MDev higher (0.22 percentage points)	-2.50	-5.16	-3.59	-1.65	-0.09	-0.26	-1.21	-0.47	-0.04	-4.40	-3.21	-3.25
Credit Losses 1 MDev lower (0.11 percentage points)	1.13	2.26	1.16	0.64	0.04	0.13	0.48	0.24	0.02	1.88	1.63	1.45
Credit Losses 2 MDev lower (0.22 percentage points)	2.21	4.44	2.04	1.28	0.09	0.26	0.96	0.49	0.04	3.70	3.28	2.83
H +/- 1 MDev	-0.036	-0.097	-0.262	-0.019	0.000	0.000	-0.057	0.002	0.000	-0.094	0.009	-0.052
H +/- 2 MDev	-0.148	-0.360	-0.773	-0.184	0.001	0.000	-0.127	0.010	0.001	-0.353	0.037	-0.210
Pre-impairment income/TA 1 MDev less (0.17 percentage points)	-1.72	-4.51	-3.44	-2.06	-0.72	-1.17	-1.76	-1.00	-0.96	-1.33	-1.97	-2.66
Pre-impairment income/TA 2 MDevs less (0.34 percentage points)	-3.68	-10.72	-7.80	-4.95	-1.47	-2.46	-4.55	-2.25	-1.93	-2.76	-3.97	-5.63
Pre-impairment income/TA 1 MDev more (0.17 percentage points)	1.55	3.93	2.00	1.56	0.69	0.88	1.31	0.89	0.96	1.24	1.97	2.33
Pre-impairment income/TA 2 MDevs more (0.34 percentage points)	2.99	6.39	3.67	2.75	1.36	1.53	2.35	1.79	1.93	2.35	3.94	4.29
H +/- 1 MDev	-0.085	-0.291	-0.718	-0.250	-0.014	-0.144	-0.224	-0.051	0.000	-0.043	0.000	-0.162
H +/- 2 MDev	-0.346	-2.163	-2.065	-1.101	-0.058	-0.466	-1.096	-0.230	0.000	-0.202	-0.016	-0.670
All three drivers one MDev in adverse direction	-4.34	-10.24	-7.58	-4.45	-0.80	-1.35	-3.86	-1.60	-1.01	-5.63	-5.29	-6.07
All three drivers two MDev in adverse direction	-9.47	-33.65	-19.50	-10.92	-1.64	-3.09	-9.95	-4.73	-2.02	-13.11	-11.31	-14.13
All three drivers one MDev in benign direction	3.33	6.17	3.38	2.43	0.76	1.03	2.02	1.36	1.00	4.87	4.87	4.23
All three drivers two MDev in benign direction	5.17	9.00	6.23	4.19	1.48	1.72	3.67	2.65	2.00	7.91	7.97	6.61
H +/- 1 MDev	-0.505	-2.035	-2.098	-1.012	-0.020	-0.162	-0.924	-0.123	-0.003	-0.649	-0.212	-0.922
H +/- 2 MDev	-2.147	-12.326	-6.638	-3.362	-0.082	-0.685	-3.139	-1.040	-0.012	-2.602	-1.670	-3.756
Trading income/TA 1 MDev less (0.1 percentage points)	-0.48	-1.10	-0.57	-0.26	-0.50	-1.43	-0.25	-0.11	-0.19	-0.49	-0.47	-0.36
Trading income/TA 2 MDev less (0.2 percentage points)	-0.99	-2.12	-1.19	-0.52	-1.03	-3.40	-0.51	-0.20	-0.38	-1.01	-0.95	-0.74
Trading income/TA 1 MDev higher (0.1 percentage points)	0.45	0.97	0.50	0.26	0.48	1.13	0.19	0.11	0.19	0.47	0.47	0.35
Trading income/TA 2 MDev higher (0.2 percentage points)	0.89	1.84	0.98	0.51	0.94	1.96	0.37	0.21	0.38	0.92	0.95	0.69
H +/- 1 MDev	-0.013	-0.063	-0.035	-0.001	-0.012	-0.147	-0.033	0.000	0.000	-0.011	0.000	-0.007
H +/- 2 MDev	-0.053	-0.140	-0.107	-0.005	-0.047	-0.720	-0.070	0.003	0.000	-0.045	0.000	-0.027

Source: Authors.

Overall, the rank order tends to be consistent across the three measures, but with several exceptions. Bank 9, for example, is the least vulnerable according to both measures. However, while a "traditional" stress test would single out Bank 6 as vulnerable, the heuristic classifies this bank as a less fragile one. The opposite is true for banks 3, 4, 7, and 12. Intuitively, what this means is that for these banks, their capital would be asymmetrically impacted by negative shocks (compared to similar sized positive shocks), and thus are fragile to a more volatile environment.

Table 2. Overall Fragility of Banks

	Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6	Bank 7	Bank 8	Bank 9	Bank 10	Bank 11	Bank 12
Rankorder (Impact of shock based on Stress Test)	5	3	7	9	6	2	10	11	12	1	4	8
Rankorder based on Heuristic (Scenario 2,4)	7	2	1	4	10	8	6	9	11	3	12	5
Rankorder based on Heuristic (Scenario 18,20)	7	1	2	4	11	10	5	9	12	6	8	3

Source: Authors.
Note: The lower the rank, the higher the impact of stress.

C. Case Study II: The Simple Heuristic Applied to Public Debt

The heuristic can be also used to predict the effect of debt and the underestimation of the risks of higher than planned deficits.

The recent financial crises have been highlighting the uncertainty around the future path of growth, particularly in advanced economies. Debt and deficit outturns have sometimes been considerably worse than expected, underlining the importance of assessing how a worse-than-expected growth scenario over the medium term might impact public finances. This section assesses the sensitivity of public debt to adverse growth shocks for a number of advanced economies.[22] More specifically, it analyzes how errors in the estimate of growth shocks could lead to significantly higher error in the estimate of countries' debt dynamics. Indeed, if non-linearities exist, underestimating growth shocks could lead to disproportionately higher underestimation of countries' corresponding debt levels.

Various growth shock scenarios are considered in order to analyze the non-linearity. A central scenario assumes that growth is 2 percentage points less than in the baseline (September 2011 WEO) per year between 2012 and 2016. This implies a near zero real growth scenario for most of the countries in the sample. In addition to this central scenario, four tail risk scenarios are considered to assess the non-linearity and thus the fragility of the results. These tail risk scenarios assume that growth is one or two mean deviations above or below the central scenario. The mean deviations are estimated over the period 1981–2010 for each country.

---

[22] See Appendix II for details on the methodology.

The results indicate that the impact of tail growth shocks on debt is non-linear in all cases, implying that all outcomes are fragile (Table 3). Based on a diverse sample of countries, Table 3 illustrates changes (in percentage points of GDP) in net debt as a result of tail growth shocks. The sample includes countries with both low and high initial debt levels, countries with low trend growth, countries under market pressure, as well as countries with large automatic stabilizers. The results illustrate that large negative growth shocks have a disproportionately higher impact on net debt compared to smaller or positive growth shocks, indicating a non-linearity (Figure 4).

Due to the non-linearity, there is a disproportionately higher cost to underestimating the growth shock than overestimating it.[23] As illustrated by Table 3, the results illustrate that non-linearities can be important. Thus, when the stress tester is uncertain about the appropriate size of tail shock to choose, e.g., because growth is particularly volatile, symmetrical stress tests around the chosen central shock could help shed light on the impact of higher volatility on debt dynamics.

Table 3. Change in Net Debt Under Various Scenarios
(Percentage points of GDP)

	Country A	Country B	Country C	Country D	Country E	Country F	Country G
Central shock (2% points less)	26.2	13.5	20.9	29.6	28.1	34.5	23.7
	Fragility/Antifragility Tests: Change Relative to the Central Shock.						
GDP growth 1 M Dev. less	14.0	13.1	16.1	18.5	28.9	39.6	37.5
GDP growth 2 M Dev. less	28.6	27.1	33.2	37.9	60.2	83.6	80.1
GDP growth 1 M Dev. more	-13.5	-12.2	-15.2	-17.6	-26.7	-36.0	-33.3
GDP growth 2 M Dev. more	-26.4	-23.7	-29.5	-34.3	-51.5	-68.7	-63.1
H: +/- 1 M Dev.	0.3	0.4	0.5	0.4	1.1	1.8	2.1
H: +/- 2 M Dev.	1.1	1.7	1.9	1.8	4.4	7.4	8.5

Source: Authors.
Note that since the change in the net debt/GDP ratio is presented in positive numbers, a positive heuristic implies an increase in risk.

---

[23] Under the hypothesis that the impact of a growth shock on debt is linear, the impact should be similar when a growth shock is augmented and reduced by a similar constant.

Figure 4. Illustration of Debt Dynamics Under Various Scenarios
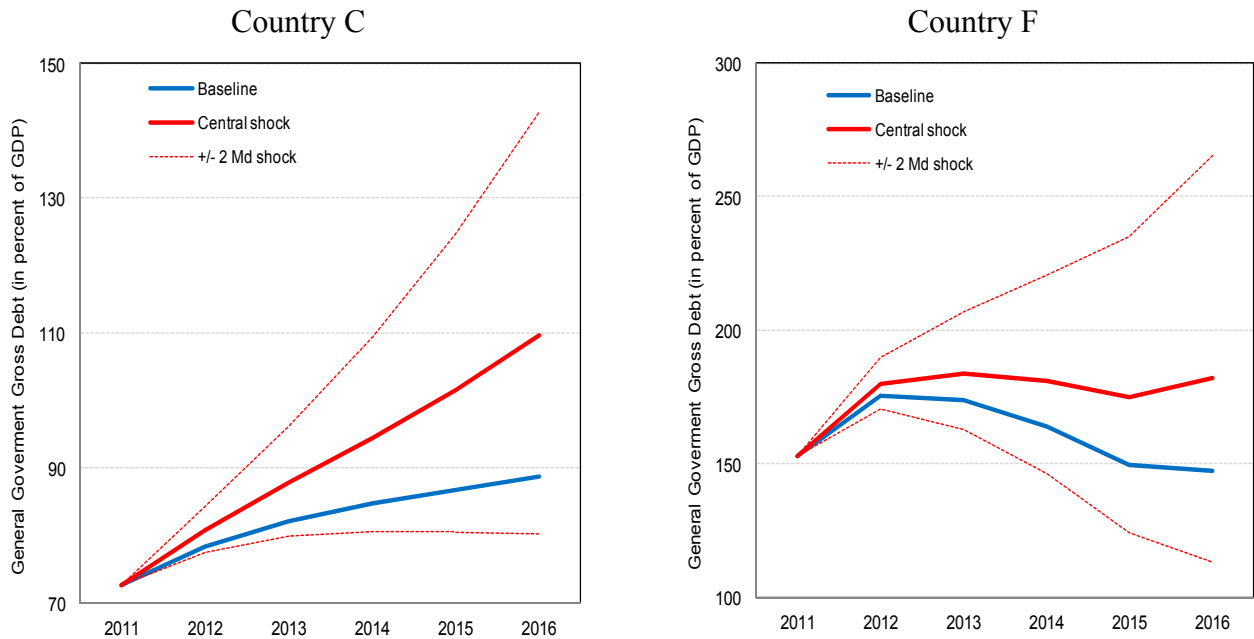
Source: Authors.

## IV. HOW TO APPLY THE SIMPLE HEURISTIC IN IMF STRESS TESTS

The heuristic can be used as a standard element in IMF bank and public debt stress test analysis, for example, as displayed in Figure 5. The heuristic can explore potential non-linearities in some range around the typically-sized stress test (often a stress of two standard deviations in the state variable). The size of the delta in the stress test, which determines the size of that range, can be chosen with a view to exploring where the stress tester may suspect that non-linearities could arise.[24] The basic procedure would be as follows:

- First, run stress tests and obtain results (bank capital ratios, liquidity ratios, NPL ratios, Net interest income, ROA, ROE, public debt, etc.)

- Second, take the stress test scenario and construct two additional scenarios: $X_t + \Delta X$, $X_t - \Delta X$ where ΔX is an estimated mean deviation of the variable $X(t)$ over the pre-determined time period.

- Third, compute the outcome for these two additional scenarios.

---

[24] For example, if the non-linearity arises from traders hiding risks in the tails (as posited in Section II.B.) then a relatively small delta plus and minus around the 5 percent probability level could reveal such a non-linearity.

- Fourth, compute the heuristic accordingly.

- Fifth, draw conclusions and reiterate:

  - If the heuristic indicates fragility to higher volatility (positive when a higher outcome is adverse, negative when a lower outcome is adverse), then the stress tester would conclude that an even greater adverse stress could make the outcome substantially worse. The test can also allow for a rank ordering of fragility to specific risk drivers. This would serve as an additional measure of riskiness beyond the typical "level" stress test, and furthermore, one that is more robust to some types of model error.

  - A finding of fragility to higher volatility could suggest reiterating the procedure in order to explore further how outcomes can vary in different parts of the tail. That is, additional scenarios with smaller or larger adverse shocks could be used to explore other areas of the risk distribution function. In that case, one could re-iterate from Step 1 (a comprehensive reiteration with a different set of stresses) or Step 2 (a limited reiteration choosing different Δ's).[25]

Figure 5. The Simple Heuristic as an Integral Part of Stress Test Frameworks
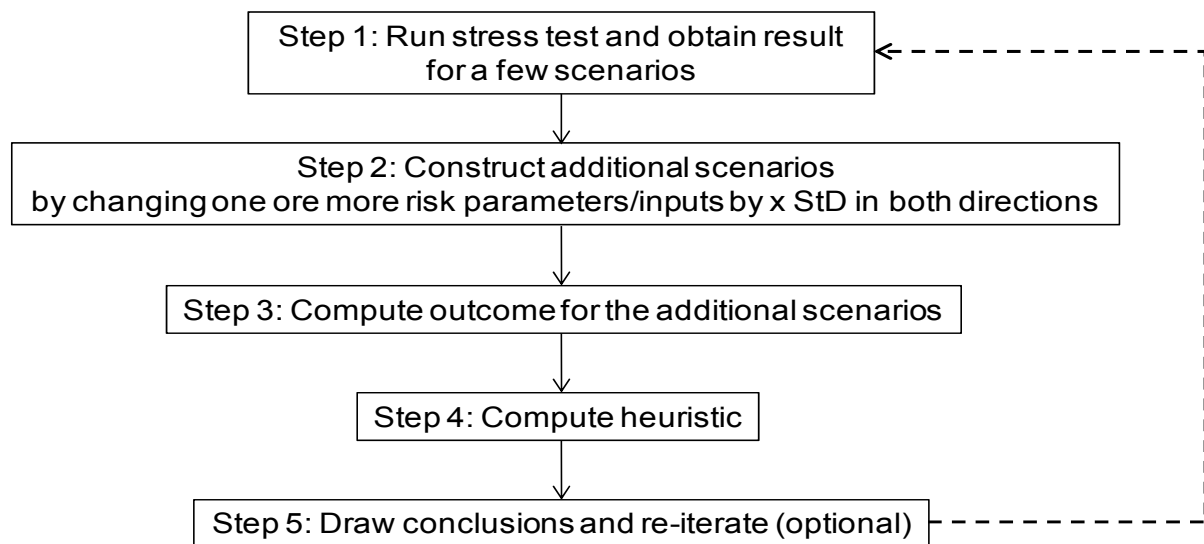
Source: Authors.

---

[25] There will be no obvious procedure to specify when the iterations should stop, since the functional form of the relationship between outcomes and stressors may be almost limitlessly complex. Rather, the iteration of the procedure will be used to expand the stress tester's knowledge about behavior in the tails, but the stress tester should never have a pretence to complete knowledge of the functional form governing how the outcome will react to different stressors.

V. CONCLUSION

This paper has presented a conceptually and operationally simple heuristic, developed in Taleb (2011), to expand information that can be extracted from the results of stress tests. When there is uncertainty about the stress testing model and/or the potential size of tail shock to be tested, as will generally be the case, the level results, generally presented in a simple pass/fail form, will be imprecise in some measure (as stress testers themselves will be aware). By examining the convexity of losses in the tails of the distribution, the stress tester can gain an extra order of information on the fragility of what is being tested. These convexities can be important, since they can cause financial losses, sovereign debt, or some other financial variable to "blow up" in response to a shock that may be only modestly larger than analyzed by the stress tester.

The heuristic will also be important in the way stress test results are presented. It provides an intuitively understood measure of the bias stemming from the likely imprecision necessarily involved in stress testing. Alternatively, it can be seen as a measure of fragility to volatility in the stresses (as represented by the state variables), with fragility defined as the property that stresses bring disproportionately higher harm as the stress increases.

The heuristic is calculated by conducting additional stress tests around the central stress scenario, by varying the relevant state variable (or vector), plus and minus, by a multiple of its mean deviation. The heuristic itself is a simple scalar that can be easily understood.

This paper constructs heuristics for two different stress testing applications, bank capitalization and public sovereign debt. In both cases, there are à priori reasons to be concerned about dynamics that could give rise to nonlinearities of the type that can cause fragility. Indeed, in both sets of stress tests, we find cases in which such non-linearities appear. This finding should lead the stress tester towards a more cautious interpretation of the robustness of the results of a single point stress test, or consider running more scenarios. The heuristic, as a kind of second-order stress test, may well give a different ordinal ranking than would be found by examining only the "level" of stress found in the point test. That is, the heuristic presented in this paper could lead to different (or at least additional) conclusions to the pass/fail results that are typically presented as the main outcome of stress tests. Thus, the heuristic explicitly highlights the potential for harm (or conceivably benefit) from high volatility of stresses in the tails.

Such results may have important policy implications. For example, if a country finds the structure of its public finances make it particularly fragile to growth shocks, it may conclude it has less room for countercyclical deficits. By the same token, a stress test could help a bank or a banking supervisor find otherwise hidden fragilities coming from, say, large illiquid positions subject to fire sales in some conditions, derivatives exposures with nonlinear payoffs, or feedback loops between losses and funding costs.

More broadly, the heuristic could be of quite general application whenever a system is being subjected to a stress test. The authors believe, in particular, that calculation and presentation of the heuristic could become a useful standard statistic in stress testing applications, as well

as a diagnostic tool that could be used to suggest when the stress tester may need to dig deeper to explore the robustness of the results of the original stress test.

income remains positive, otherwise retained income (i.e., loss) is fully retained. A tax rate of 25 percent was applied to all banks, in line with empirical evidence.

Funding costs

We assumed a funding cost increase conditional on the capitalization level (during stress), that is, funding costs increase faster as capital falls to low levels, reflecting empirical evidence. Hence, the funding costs add a dynamic element to the stress test. We assumed that banks are able to pass on (only) 50 percent of funding cost increases to their customers, given competition pressures. Further information is given in Schmieder, et. al. (2012).

APPENDIX II: DETAILS ON PUBLIC DEBT STRESS TEST

The following standard debt dynamic equations illustrate the main parameters influencing debt dynamics:

$$\begin{cases} d_t = d_{t-1}(1+r_t) - pb_t \\ r_t = \dfrac{i-g}{1+g} \\ pb_t = pb_t^{base} + (\varepsilon_R - \varepsilon_E)\Delta og_t \end{cases}$$

d_t represents the ratio of public debt to GDP, pb_t is the primary balance, and r_t is the growth adjusted interest rate. The growth adjusted interest rate is a function of the nominal interest rate (i) and the nominal GDP growth rate (g). The nominal interest rate is derived from the ratio of interest payments during the current year to the end-period stock of debt during the previous year. The primary balance (pb_t) depends on the primary balance under the baseline (pb_t^{base}), revenue and expenditure semi-elasticity to changes in the output gap ($\varepsilon_R, \varepsilon_E$),[31] and the change in the output gap between the baseline and different scenarios (Δog_t). All scenarios assume that growth shocks do not affect potential GDP and governments do not take any discretionary corrective measures to smooth their impacts.[32] As a consequence, growth shocks affect debt ratios through the size of automatic stabilizers and changes in the GDP base.

Three macroeconomic variables will therefore affect each country's debt dynamics: trend growth, the size of the initial (pre-shock) stock of public debt, and the size of the automatic stabilizers. Trend growth would particularly matter for countries with projected low growth rates during the period 2012–2016. For these countries, a negative growth shock would lead to a significantly higher build-up of public debt than in high growth countries. The initial stock of public debt would be particularly important for highly indebted countries, notably those that experienced a surge in their debt ratios as a result of the crisis. The size of automatic stabilizers matters more in countries with particularly high welfare spending, as the relationship between tax revenues and economic activity tends not to vary greatly across countries.

---

[31] Revenue and expenditure elasticity to the output gap are from Girouard and André (2005). When not available, an elasticity of one is assumed for revenue and zero for expenditures.

[32] This is a partial equilibrium simulation that also assumes no change in the nominal interest rate as a result of the growth shock.

REFERENCES

Borio, Claudio, Drehmann, Mathias, and Kostas Tsatsaronis, 2012, "Stress-testing macro stress testing: does it live up to expectations?," BIS Working Paper no.369, January.

Financial Stability Board, 2011, "Understanding Financial Linkages: A Common Data Template for Global Systemically Important Banks: Consultation Paper," October 6, 2011.

Financial Stability Board Secretariat and IMF, 2009, "The Financial Crisis and Information Gaps: Report to the G-20 Finance Ministers and Central Bank Governors," October 29, 2009.

Girouard, Nathalie and Christophe André, 2005, "Measuring Cyclically-Adjusted Budget Balances for OECD Countries," OECD Economics Department Working Paper No. 434 (Paris: Organization for Economic Cooperation and Development).

Ong, Li Lian and M. Cihak, 2010, "On Runes and Sagas: Perspective on Liquidity Stress Testing Using an Iceland Example," IMF Working Paper 10/156 (Washington: International Monetary Fund).

Schmieder, Christian, Puhr, Claus and Maher Hasan, 2011, "Next Generation Balance Sheet Stress Testing," IMF Working Paper 11/83 (Washington: International Monetary Fund).

_____, Hesse, Heiko, Neudorfer, Benjamin, Puhr, Claus and Stefan W. Schmitz, 2012, "Next Generation System-Wide Liquidity Stress Testing," IMF Working Paper 03/12 (Washington: International Monetary Fund).

Taleb, Nassim N., 2011, "A Map and Simple Heuristic to Detect Fragility, Antifragility, and Model Error", NYU-Poly working paper, SSRN.

_____, 2012, forthcoming), Antifragile: Things That Gain from Disorder, Fall 2012, Random House (US) & Penguin (UK).

_____, and Raphael Douady, 2012, "Mathematical Definition and Mapping of (Anti)Fragility", NYU-Poly Working Paper.

Vitek, Francis, and Tamim Bayoumi, 2011, "Spillovers from the Euro Area sovereign debt crisis: A macroeconometric model based analysis," CEPR Discussion Paper, 8497.