The global financial crisis drew unprecedented attention to the role of stress testing of financial institutions in macroprudential and microprudential surveillance, and its role as an integral element of crisis management to inform policies aimed at restoring confidence in the financial system. Current stress testing practices, however, are not based on a systematic and comprehensive set of principles but have emerged from a trial-and-error approach and practical expediency. The chapter draws on the experience gained from a decade of stress testing in the context of IMF Financial Sector Assessment Programs to propose seven “best practice” principles that are universally applicable, including (1) the intended scope of the stress testing exercise affecting the identification of risks and measurement of vulnerabilities, (2) the macro-financial channels through which shocks are transmitted, (3) the availability of risk-mitigating features, and (4) the effectiveness of communicating the findings. These principles serve as practical guidance on how to tailor stress tests to specific circumstances, including the degree of financial sector development, business models, and the macroeconomic environment in which financial institutions operate.

1. STRESS TESTING: A PRIMER

Over the last 20 years, stress testing has become essential to financial stability analysis. Stress tests first emerged in the late 1990s and have been used since then by financial institutions, regulatory bodies, and international organizations such as the IMF and the World Bank, with the aim of proactively identifying vulnerabilities, and/or determining specific risks for industry sectors, certain business models within these sectors, or systemically relevant institutions.

The global financial crisis placed a spotlight on stress testing of financial institutions, notably banks. The financial crisis had a significant impact on the way stress tests are being carried out, not only by national authorities but also by market participants. The experience highlighted the usefulness of stress tests as a diagnostic tool, but also revealed weaknesses in stress tests undertaken prior to the crisis by the banks themselves, supervisory authorities, and the IMF, all of whom to a greater or lesser extent failed to capture the risks that eventually materialized. In particular, a key lesson from the crisis has been a greater focus on concepts to identify the buildup of financial risks. This has spawned risk-based framework(s) for financial stability analysis, including the examination of macro-financial linkages and the integration of advanced market and risk-based tools for surveillance purposes. At the same time, the crisis underscored the potential of credible and comprehensive stress tests in restor-
ing market confidence in the financial system, as demonstrated by the Supervisory Capital Assessment Program (SCAP) exercise undertaken by the US authorities in 2009. Stress testing, once an arcane subject, has become almost a household name.

As a result of the attention and lessons learned from the crisis, the approaches, underlying assumptions, and uses of stress tests are being scrutinized and actively debated. The large and complex menu of choices in each of these areas has given rise to questions about the interpretation and consistency of findings generated by stress testing exercises. And the continued fragilities in financial systems during the transition to a more robust postcrisis regulatory framework have made the communication of stress test results an increasingly sensitive issue for both supervisors and financial institutions struggling to balance the need for greater transparency with the need to avoid alarming markets and creating self-fulfilling prophecies.

The IMF is well-placed to contribute to this debate, having amassed significant practical experience in applying stress tests in a wide range of countries over a decade. Stress testing of financial systems has been a key component of the Financial Sector Assessment Program (FSAP) launched in 1999 and, subsequently, part of the analytical tools used in the biannual Global Financial Stability Report (GFSR). The IMF staff has also played an instrumental role in developing and disseminating several advanced stress testing models and cooperates closely with technical experts in supervisory agencies and central banks in testing and implementing new techniques, including through the Expert Forum on Advanced Stress Testing Techniques and the Research Task Force (RTF) of the Bank for International Settlements (BIS). And the IMF is providing technical assistance and training to member countries interested in building or expanding stress testing capabilities.

This chapter discusses current practices in stress testing and proposes operational “best practice” principles for their design and implementation. These practical guidelines are derived from the years of the IMF’s own experience in developing and using stress testing tools, including the regular internal review and evaluation of FSAP stress tests. They are not meant to be a “general theory” of stress testing or provide a comprehensive step-by-step stress testing manual. Instead, the key goal is to help set realistic expectations about the effective application of stress tests and explore how their design and implementation can be improved to ensure that they remain useful tools in identifying financial sector vulnerabilities.

The chapter fills a major gap in this debate. To the IMF staff’s knowledge, this is the first time that specific, operational principles have been put forth for system-wide stress tests. The BIS and Committee of European Banking Supervisors (CEBS)—the predecessor of the European Banking Authority (EBA)—have also proposed principles for stress testing, but those were mainly directed to banks performing stress tests as part of their risk-management functions (Appendix 2.1). Greenlaw and others (2012) suggest principles for stress tests focusing on risks that could have system-wide and economy-wide implications, but their principles are more conceptual, aimed at shifting the thinking about the purpose and goals of stress tests away from their microprudential focus on individual institutions toward systemic risk. These considerations are also echoed in the first Annual Report of the Office of Financial Research (OFR) at the US Treasury. Although this chapter concentrates mainly on stress tests conducted for macro-financial surveillance purposes—the key interest for the IMF, central banks, and macroprudential authorities—much of the discussion also applies to stress tests undertaken for other purposes, such as microprudential oversight and institution-specific risk assessment.

In addition, this chapter provides the basis for a more systematic approach to stress testing in FSAPs. The proposed principles establish a yardstick against which individual stress testing exercises can be evaluated, as well as an agenda for improvements to the IMF’s stress testing toolkit. These elements provided important input into the last review of the FSAP (IMF 2013a and 2014a).

The rest of the chapter is organized as follows. Section 2 presents a brief introduction to the basic concepts and tools of stress testing. Section 3 discusses the lessons from the global financial crisis and European sovereign debt crisis for stress tests. Section 4 presents seven “best practice” principles for stress testing and examines how closely actual stress testing practice corresponds to them. The principles are based not only on the IMF’s own extensive experience, but also that of its member countries, on the basis of a survey undertaken for this purpose. Finally, key conclusions and practical implications of these principles for stress testing practitioners are presented in the fifth section.

2 In September 2010, the Executive Board of the IMF made financial stability assessments under the FSAP a regular and mandatory part of the bilateral surveillance under Article IV for jurisdictions with systematically important financial sectors.

3 The survey was conducted in November 2011 and covered (1) the broad use and definition of stress tests; (2) banking sector stress tests (process and organization of stress tests; framework for solvency tests, including risk/scenario selections, macro-financial linkages, determining capital adequacy; framework for liquidity stress tests; communication strategy); and (3) issues regarding the use of stress tests and their application to the nonbank financial sector. A total of 26 central banks and supervisory authorities from 23 different countries responded (in some country cases, the responses were jointly submitted by more than one agency). Among the 23 countries, seven were emerging market and developing economies (EMDEs) and 16 were developed economies, 13 were European, six were Asian, and four were Western Hemisphere countries. A separate background paper (IMF 2012c) provides details of the survey results.
Hiroko Oura and Liliana Schumacher

For an institution to be solvent as a going concern, it would need to maintain a minimum of positive equity capital so that it can absorb potential losses in the event of a shock. Higher amounts of capital than this minimum might be needed to ensure continued access to market funding at a reasonable cost.

A solvency test assesses whether a firm has sufficient capital to remain solvent in a hypothetically challenging environment by estimating profit, impairment losses, and valuation changes. The main risk factors are potential losses from borrowers’ default (credit risk) and securities due to changes in market prices such as interest rates, exchange rates, and equity prices (market risk). A stress test may examine the impact of individual risks (single factor tests) or multiple sources of risks (multiple factor tests). Risk factors could be combined in an ad hoc manner (combined shock test) or generated more coherently using a macroeconomic framework (macroscenario tests).

Solvency tests may cover varying segments of a balance sheet (Figure 2.1). A test for credit risk may cover total loans (including interbank lending) or loans to certain segments (such as corporate, mortgages, or credit card loans). Market risk is assessed for securities in the held-for-trading (HfT) and in the available-for-sale (AfS) accounts, but loans, advances, and debt securities in held-to-maturity (HtM) account may be excluded, because these securities are supposed to be held for longer periods and are not subject to the same market risks.

Figure 2.1  Simplified Bank Balance Sheet

<table>
<thead>
<tr>
<th>Assets</th>
<th>Liabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cash and cash equivalent</strong></td>
<td>Central bank loans</td>
</tr>
<tr>
<td><strong>Money market assets</strong></td>
<td><strong>Money market liabilities</strong></td>
</tr>
<tr>
<td>Interbank lending</td>
<td>Interbank borrowing</td>
</tr>
<tr>
<td>Reverse repurchase agreements and securities borrowing</td>
<td>Repurchase agreements and securities lending</td>
</tr>
<tr>
<td>Certificates of deposit, commercial paper</td>
<td>Certificates of deposit, commercial paper</td>
</tr>
<tr>
<td><strong>Securities</strong></td>
<td>Customer deposits (financial institutions, public sector, firms/SMEs, and household)</td>
</tr>
<tr>
<td>Held-for-trading</td>
<td>Long-term borrowing</td>
</tr>
<tr>
<td>Available-for-sale equities and debt</td>
<td>Held-to-maturity securities</td>
</tr>
<tr>
<td></td>
<td>Debt instruments</td>
</tr>
<tr>
<td>Customer loans (financial institutions, public sector, corporate, and household) and related-party lending</td>
<td>Derivatives</td>
</tr>
<tr>
<td>Derivatives</td>
<td>Other borrowings</td>
</tr>
<tr>
<td><strong>Other assets</strong></td>
<td>Equity</td>
</tr>
<tr>
<td><strong>Off-balance-sheet items</strong></td>
<td></td>
</tr>
<tr>
<td>• Derivatives</td>
<td></td>
</tr>
<tr>
<td>• Contingent claims and liabilities (credit lines, guarantees, [implicit] guarantees to special purpose vehicles)</td>
<td></td>
</tr>
<tr>
<td>• Securitization, resecuritization exposures</td>
<td></td>
</tr>
<tr>
<td>Source: Authors.</td>
<td></td>
</tr>
<tr>
<td>Note: SMEs = small and medium-sized enterprises.</td>
<td></td>
</tr>
</tbody>
</table>

Solvency Tests

An institution is solvent when the value of its assets is larger than its debt, that is, there is a certain amount of positive equity due to a positive net asset value. The values of both assets and liabilities depend on future cash flows, which are uncertain and depend on economic and financial conditions. For an institution to be solvent as a going concern, it would need to maintain a minimum of positive equity capital so that it can absorb potential losses in the event of a shock. Higher amounts of capital than this minimum might be needed to ensure continued access to market funding at a reasonable cost.

A solvency test assesses whether a firm has sufficient capital to remain solvent in a hypothetically challenging environment by estimating profit, impairment losses, and valuation changes. The main risk factors are potential losses from borrowers’ default (credit risk) and securities due to changes in market prices such as interest rates, exchange rates, and equity prices (market risk). A stress test may examine the impact of individual risks (single factor tests) or multiple sources of risks (multiple factor tests). Risk factors could be combined in an ad hoc manner (combined shock test) or generated more coherently using a macroeconomic framework (macroscenario tests).

Solvency tests may cover varying segments of a balance sheet (Figure 2.1). A test for credit risk may cover total loans (including interbank lending) or loans to certain segments (such as corporate, mortgages, or credit card loans). Market risk is assessed for securities in the held-for-trading (HfT) and in the available-for-sale (AfS) accounts, but loans, advances, and debt securities in held-to-maturity (HtM) account may be excluded, because these securities are supposed to be held for longer periods and are not subject to the same market risks.
to be paid in full at maturity as long as the debtor does not default.4 Off-balance-sheet exposures, including contingent claims and securitization exposures, could be affected by both market and credit risks and could potentially have highly nonlinear responses to stress. Bank solvency tests usually do not adjust the value of liabilities for changes in interest rates (market risk) due to their positive duration gap, that is, most bank liabilities are short-term deposits and money market instruments funding lending and investment over longer terms. On the other hand, stress testing for in-money market instruments funding lending and investment that is, most bank liabilities are short-term deposits and interest rates (market risk) due to their positive duration gap, usually do not adjust the value of liabilities for changes in highly nonlinear responses to stress. Bank solvency tests both market and credit risks and could potentially have Basel III) are expected to be introduced or to maintain a warrant, for example when new regulations (for example, but they could be set at different values if circumstances often set at the current minimum regulatory requirement, predetermined threshold or "hurdle rate." Hurdle rates are to "pass" or "fail" the test if the target capital ratio is above a predetermined threshold or "hurdle rate." Hurdle rates are often set at the current minimum regulatory requirement, but they could be set at different values if circumstances warrant, for example when new regulations (for example, Basel III) are expected to be introduced or to maintain a certain level of market funding cost.5 The choice of the hurdle rate is a critical factor in stress testing exercises, especially when the results of the tests are directly linked to capital planning directed by supervisors.

Solvency is measured by various capital ratios, typically following regulatory requirements. The capital impact of the solvency test is reflected in changes of the regulatory capital ratio (capital adequacy) and the extent to which it would drop below the prudential minimum under the assumed stress scenario, resulting in a capital shortfall. Standard choices for banks are the ratio of statutory/core Tier 1/Tier 1 capital to risk-weighted assets (RWA); leverage ratios (capital to assets); losses in percent of capital; or capital shortfalls (the amount of capital needed to maintain a certain capital ratio). Individual institutions or the system at large are said to "pass" or "fail" the test if the target capital ratio is above a predetermined threshold or "hurdle rate." Hurdle rates are often set at the current minimum regulatory requirement, but they could be set at different values if circumstances warrant, for example when new regulations (for example, Basel III) are expected to be introduced or to maintain a certain level of market funding cost.5 The choice of the hurdle rate is a critical factor in stress testing exercises, especially when the results of the tests are directly linked to capital planning directed by supervisors.

Estimating solvency ratios in macroscenario stress tests requires estimating macro-financial models. A macro-financial model estimates the empirical relationship between key risk parameters (net operating profit, non-performing loan ratio, probability of default,6 loss given default, and provisioning) and relevant macroeconomic variables, such as GDP, unemployment, exchange rate, asset prices, and interest rates. It requires the use of econometric models—as well as considerable judgment—and the IMF provides technical assistance in this area. Macrosenario stress tests cover several years (typically one to three in the case of country supervisory authorities or central banks, and often longer in FSAPs), as credit risks materialize gradually in economic downturns. Therefore, preimpairment profits (profits before loan and security losses) also need to be projected, since retained earnings in the test horizon would affect capital. This requires making assumptions about bank behavior (such as dividend payout policies and deleveraging, in case of adverse shocks), which introduces significant degrees of freedom—and complexity—to the exercise (Jobst, Ong, and Schmieder 2013).

**Liquidity Tests**

Liquidity stress tests aim to capture the risk of a bank failing to generate sufficient funding from cash inflows to satisfy short-term payment obligations arising from a sudden realization of liabilities in a stress scenario (Jobst, Ong, and Schmieder 2017). The tests assess the adequacy of the available funding sources over a defined stress horizon. Financial institutions may encounter sudden cash outflows, for instance, because of:

- **Sudden distress with their funding.** Financial intermediaries that engage in liquidity/term structure transformation, particularly banks, have, by the nature of their business, maturity mismatches in their balance sheets. Thus, they need to carefully manage scheduled and unscheduled cash outflows (including the loss of funding sources) against cash inflows that are related to maturing assets, the rollover risk stemming from any existing maturity mismatches, as well as the ability to access unsecured retail/wholesale funding markets. If a large amount of deposits is suddenly withdrawn, or funding markets (such as repurchase agreements and commercial paper) freeze, a bank might no longer be able to meet its current and future cash flow needs even if it is otherwise solvent ("funding liquidity risk"). This would also involve market liquidity risk if the banks cannot sell assets quickly owing to deterioration in its liquidity ("market liquidity risk").

---

4 Values of securities in trading and AfS accounts are mostly assessed at market values (mark-to-market [MtM] valuation). Losses and gains from trading securities are accounted for in the profit-and-loss statement (and therefore affect capital). Unrealized capital gains and losses from AfS securities also affect regulatory capital (that is, Core Tier 1 capital), albeit at varying degrees depending on national regulation and accounting rules. Unrealized losses should in principle be deducted from regulatory capital under the Basel III framework. However, the Basel II framework does not refer specifically to the treatment of unrealized AfS losses, granting national supervisors substantial national discretion in applying prudential filters for gains and losses of AfS securities ("AfS filter"). Under the transitional arrangement of Basel III, a rising fraction of unrealized losses is deducted from capital (in recognition of existing provisions). A standard historical cost approach is applied to HnM securities, which are valued at amortized cost net of any impairment provision (that is, net book value) unless there are persistent and substantial unrealized losses. The precise valuation practices differ across jurisdictions.

5 Pillar 2 of the Basel II framework empowers supervisors to request above minimum requirement capital ratios in line with banks’ risk profiles.

6 Point-in-time (PIT) or through-the-cycle (TTC) default probabilities may be used, but the latter tend to dampen portfolio risk (Rosch and Scheule 2008).

7 Under Basel I or the standardized approach of Basel II, the amount of loss provisions depends on the historical losses (nonperforming loan ratio) and loan classification. Under the internal ratings-based (IRB) or advanced IRB approaches of Basel II (and Basel III), however, banks need to determine provisions based on complete historically credit risk dynamics (that is, probability of default/loss given default or credit rating data). 
• Interlinkages between market and funding liquidity risk. Financial institutions active in taking market positions may face a sudden liquidity need when asset markets become volatile, increasing initial/variation margin requirements. For example, those trading in highly leveraged derivatives markets may face a liquidity shortage if a position becomes out-of-the-money\(^8\) (even if at expiration this result is reversed) or gets downgraded. Even when these positions are taken by legally separate special purpose vehicles (SPVs), some institutions may be forced to support these SPVs for reputational reasons, effectively internalizing the liquidity shortage.

Financial institutions encounter a liquidity shortage when they cannot generate sufficient cash in response to a shock. Banks that have enough liquid assets counterbalancing capacity can generate sufficient cash either by selling them or by pledging them as collateral for repurchase agreements without making large losses. However, if available assets are mostly nonmarketable loans, or if the market value of these assets declines substantially below the book value (haircut), banks may face liquidity constraints. Several possible hurdle rates may be used in a liquidity stress test, such as the number of days the institution can tolerate a liquidity shock before net cash flows become negative (which would also manifest in a deterioration of stressed liquidity ratios).

Liquidity and solvency stress events are often closely interrelated and hard to disentangle. In the event of funding distress, a liquidity shortage may threaten the bank’s solvency if assets cannot be sold or can be sold only at loss-making prices (“fire sale”). Thus, higher market funding risk in a liquidity stress event is a factor that could translate into solvency stress.

The Typology of Stress Tests

There are four types of stress tests, which are differentiated by their ultimate objectives (for details, see Table 2.1):

• Stress testing as an internal risk-management tool. Financial institutions use stress testing to measure and manage the risks with their investments. One of the early adopters was J.P. Morgan in the mid-1990s, which used value at risk (VaR) to measure market risk. However, early stress testing had limited coverage of risk factors and exposures and little integration with the overall risk management and business and capital planning at firms.

• Microprudential/supervisory stress testing. The Basel II framework requires banks to conduct stress tests for market risk and, in some cases, credit risk as a part of minimum capital regulation (Pillar 1). Additional tests can be required in the context of Pillar 2 that provides supervisors powers to order management actions by banks if deemed necessary. A survey by the Basel Committee for Banking Supervision (BCBS) (BCBS 2012) indicates that supervisory stress tests are increasingly utilized to set capital requirements for specific banks, determine explicit capital buffers, or limit capital distributions by banks in the context of the capital review process under Pillar 2 of the Basel framework. The liquidity risk standards established by the Basel III (and the Solvency II regime for European insurance companies) utilize stress testing as an integral part of the regulatory framework.

• Macroeconomic/surveillance stress testing. Over the past two decades, many country authorities have started using stress test exercises to assess system-wide risks, in addition to institution-specific risks, based on the capital adequacy of the banking sector under adverse macroeconomic conditions. The results are often reported in their financial stability reports. As an integral part of risk-based analytical frameworks supporting macroprudential policy, surveillance stress testing aims at identifying systemic risk ex ante, thereby minimizing the incidence and impact of disruptions in the provision of key financial services that can have adverse consequences for the real economy (and broader implications for economic growth).\(^9\) Such risk to financial stability arises from the collective impact of common shocks on systemically important institutions and/or a material number of firms, possibly amplified by market failures and/or fault lines in the architecture of the financial system, for instance between banking and nonbanking financial sector activities. The IMF has regularly included macroprudential stress testing in FSAPs since the program’s inception in 1999 (Jobst, Ong, and Schmieder 2013). A few country authorities included macroprudential stress testing in FSAPs since the program’s inception in 1999 (Jobst, Ong, and Schmieder 2013). A few country authorities indicated in the staff survey (IMF 2012c) that the FSAP stress tests were the first such exercises undertaken in their countries.

• Crisis management stress testing. Stress tests have also been used, especially after the global financial crisis, to assess whether key financial institutions need to be recapitalized or not, possibly with public support. In particular, the SCAP exercise in the United States and the EU system-wide stress test organized by the CEBS/EBA in 2010 and 2011 attracted attention, because banks were required to recapitalize based on the test results, and the detailed methodology and individual banks’ results were published. In IMF programs with banking sector distress (including

---

\(^8\) The term “out-of-the-money” refers to a call option with a strike price that is higher than the market price of the underlying asset, or a put option with a strike price that is lower than the market price of the underlying asset.

\(^9\) See also Borio and Drehmann 2009.
<table>
<thead>
<tr>
<th>Features</th>
<th>Macroprudential (Surveillance)</th>
<th>Microprudential (Supervisory)</th>
<th>Crisis Management</th>
<th>Internal Risk Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main objective</td>
<td>Unveil the sources of systemic risk and vulnerability in the context of surveillance and regular system-wide monitoring</td>
<td>Assess the health of an individual institution; inform supervision of the institution</td>
<td>Input for bank recapitalization and business restructuring plans</td>
<td>Manage risks from existing portfolio; input for business planning</td>
</tr>
<tr>
<td>Organized by</td>
<td>Central banks, macroprudential authorities, IMF</td>
<td>Supervisors (microprudential authorities)</td>
<td>Macro and microprudential authorities</td>
<td>Financial institutions</td>
</tr>
<tr>
<td>Coverage of institutions</td>
<td>All (or as many as possible) institutions, especially systemically important institutions</td>
<td>Supervised individual institutions (tests for different banks could take place at different times)</td>
<td>Varies, but it should include all distressed and near-distressed institutions.</td>
<td>Own activities</td>
</tr>
<tr>
<td>Frequency</td>
<td>Typically annual or semiannual for country authorities, or in the context of FSAPs</td>
<td>Individual institutions are tested as needed; increasing number of supervisors conduct regular stress tests (with common assumptions)</td>
<td>As needed</td>
<td>High (daily or weekly) for market risks; lower for enterprise-wide exercises</td>
</tr>
<tr>
<td>Nature of shocks</td>
<td>Systemic and common shocks across institutions; shocks tend to be extreme</td>
<td>Often idiosyncratic; common macro assumptions are sometimes made for horizontal or thematic review across institutions</td>
<td>Ongoing systemic stress (baseline) or relatively mild shocks, mainly focusing on solvency risks</td>
<td>Idiosyncratic or systemic (those that matter for the particular institution)</td>
</tr>
<tr>
<td>Capacity to incorporate systemic risks</td>
<td>Through macro and market-level shocks and additional system-wide features (for example, network effects)</td>
<td>Through macro and market-level shocks</td>
<td>Through macro and market-level shocks</td>
<td>Through macro and market-level shocks</td>
</tr>
<tr>
<td>Likelihood of assumed shocks</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Varies</td>
</tr>
<tr>
<td>Assessment criteria (hurdle rates)</td>
<td>Current or prospective regulatory requirements or alternative thresholds (if appropriate)</td>
<td>Current or prospective regulatory requirements or alternative thresholds (if appropriate)</td>
<td>Current or prospective regulatory requirements or alternative thresholds (if appropriate)</td>
<td>Internal risk tolerance indicators and regulatory requirements</td>
</tr>
<tr>
<td>Key output metric</td>
<td>Aggregate indicators for the system and their dispersion</td>
<td>Individual institution indicators</td>
<td>Individual institution indicators</td>
<td>Individual institution indicators</td>
</tr>
<tr>
<td>Follow-up measures after tests</td>
<td>Typically no follow up for individual institutions, but often used as the basis for discussion of potential macroprudential or system-wide measures</td>
<td>Institutions with weak results are often required to explain and take management actions if deemed necessary by supervisors</td>
<td>“Failing” institutions are often required to take major management action, such as recapitalization, possibly with government support</td>
<td>May or may not require management action</td>
</tr>
<tr>
<td>Publication Examples</td>
<td>Often FSAPs, GFSRs, other financial stability reports</td>
<td>Occasionally, for example, U.S. and EU stress tests</td>
<td>SCAP (United States), EU system-wide stress test (CEBS/EBR), exercises in some IMF program countries</td>
<td>No</td>
</tr>
</tbody>
</table>

Source: Authors.
Note: CCAR = Comprehensive Capital Analysis and Review; CEBS = Committee of European Banking Supervisors; EBA = European Banking Authority; ECB = European Central Bank; SSM = Single Supervisory Mechanism; FSAP = Financial Sector Assessment Program; GFSR = Global Financial Stability Report; SCAP = Supervisory Capital Assessment Program; VaR = value at risk.
Ireland, Greece, and Portugal), estimating bank recapitalization needs through stress tests was an important component. As this use of stress tests as a crisis management tool is relatively new, Appendix 2.3 presents the key features, similarities with, and differences from other types of stress tests, using three prominent examples of this type of stress tests (US SCAP, the EU system-wide stress test in 2010 and 2011, and the EBA Capital Assessment Exercise of 2012) as illustrations.10

The risk coverage and methodologies have evolved over time, as the use of stress tests is broadened. Financial institutions are now expected to manage enterprise-wide risks, which cover a broad range of exposures and risk factors in an integrated manner, crossing over the internal segmentations of various business lines. Similarly, macroprudential stress tests have evolved from single-factor tests to macroscenario tests.

Depending on the objective of stress testing, follow-up managerial or supervisory actions may be taken. Macroprudential tests, including in FSAPs, typically do not prescribe bank-specific action, although they could lead to macroprudential policy recommendations. Supervisory stress tests are increasingly used to guide supervisory action, ranging from improving data collection, targeting examinations, and closer monitoring, to requiring bank management actions, such as raising additional capital, reducing certain exposures, capping dividends, and updating individual institutions’ resolution plans. Follow-up is almost certain in crisis management stress tests, which are expressly designed to estimate capital shortfalls.

Stress testing in FSAPs and by the national authorities may be conducted either as top-down or bottom-up exercises or both. Top-down exercises are defined as those conducted by the national authorities or IMF staff (typically as part of FSAPs) using bank-by-bank data and applying a consistent methodology and assumptions. Bottom-up exercises are carried out by individual financial institutions (with or without involvement of supervisory authorities) using their own internal data and models, often under common assumptions. Some supervisory tests include bank-specific risks11 (Table 2.2), including reverse stress tests based on shocks that could render a specific institution insolvent. Liquidity tests are often conducted as bottom-up exercises because they require granular data and depend on banks’ liquidity strategies, and banks are given more flexibility regarding detailed assumptions compared to solvency tests. FSAPs almost always include a top-down test, frequently supplemented by a bottom-up test. Many national authorities use both top-down and bottom-up tests and emphasize the importance of running their own top-down tests to effectively validate bottom-up results.

Communication practices differ across the four types of stress testing. Close communication between banks and supervisors, between supervisory agencies within a country, or between FSAP teams and country authorities is required for effective stress tests. Public communication of stress test

---

10 See also Ong and Pazarbasioğlu 2014.

11 There is some variation in bottom-up practices among country authorities, including bottom-up tests being conducted with institution-specific assumptions or implemented by the supervisory authority using bank-by-bank data.

12 In the United States, the Dodd-Frank Act requires communication of Comprehensive Capital Analysis and Review results by both the US Federal Reserve Board and individual banks (Board of Governors of the US Federal Reserve System 2009 and 2012).
results, on the other hand, is not common, although this has been changing, especially for crisis management stress tests.\textsuperscript{12} Macroprudential/surveillance stress test results are typically reported in financial stability reports—or Financial System Stability Assessments (FSSAs) in the case of FSAPs—but in varying degrees of aggregation, usually without identifying individual institutions. In general, dissemination of stress test results is controversial. Several country authorities have voiced concerns that public dissemination might create unrealistic expectations, lead to misinterpretations in the mass media, and potentially detract from the value of stress tests as a supervisory tool, as banks focus too much on the media impact. For FSAPs, the publication of FSSAs is voluntary, but the majority of countries—including most of the jurisdictions with systemically important financial sectors—publish the documents. The accompanying stress testing Technical Notes, which are more detailed, are published much less frequently.

### Stress Testing Models

Stress tests use a wide range of analytical models. These models are designed to capture key stress factors to solvency or liquidity of individual institutions or financial systems, varying widely in terms of complexity and data requirements. They can be broadly classified into two categories: (1) models predicated on a detailed analysis of balance sheets of individual institutions (sometimes called “fundamental” approaches), and (2) models based on summary default measures for individual portfolios, institutions, or entire systems embedded in market prices, such as stocks, bonds, and derivatives.

Both approaches are complements rather than substitutes. Balance-sheet-based approaches can identify the source of individual vulnerabilities in relation to the accounting identities provided in prudential and/or statutory reporting. Thus, they are more detailed and informative as to the effectiveness of potential remedial actions flowing from the findings of stress tests, and can be applied to emerging market and developing economies (EMDEs), where stock markets tend to be thin or illiquid (and only few financial institutions are publicly listed). But they are backward-looking, data-intensive, hard to update frequently, and not particularly suited for capturing interdependence (portfolio) and contagion effects across institutions. In contrast, market-price-based approaches are more flexible, can easily incorporate portfolio effects and risk factors as perceived by the market, and can be updated as frequently as desired. However, they also involve an identification challenge, which, depending on the model choice, can make it difficult to determine the precise source of vulnerabilities; in addition, market-based risk metrics are sensitive to short-term swings in market perceptions that may have little to do with fundamentals, and cannot be applied to countries or entities with limited or no market price data. Box 2.1 presents an overview of the two families of models, and Table 2.3 presents a detailed comparison of their operational aspects.\textsuperscript{13}

Model choices are somewhat constrained in EMDEs with limited data or weak accounting and regulatory systems, but some options are nonetheless available. At the same time, these countries tend to have simpler financial and economic systems, and key economic risks and vulnerabilities are relatively straightforward to identify. Simple balance-sheet-based tests (such as those illustrated in Čihák 2007) with single or multifactor shocks can be implemented in most countries using basic supervisory data. In cases where supervisory data are patchy or unreliable, or the magnitude of uncertainty is too large to draw strong conclusions from the tests, the priority should be to improve supervision and data quality rather than develop stress testing approaches. Developing a macro-financial linkage model for macroscenario tests can also be a challenge in some EMDEs, given limitations in the quality or availability of long time series data. One option in such cases is to utilize cross-country data.\textsuperscript{14}

### 3. LESSONS FROM THE GLOBAL FINANCIAL CRISIS AND THE EUROPEAN SOVEREIGN DEBT CRISIS

Both the global financial crisis and the European sovereign debt crisis had major impacts on the design, completion, and interpretation of stress tests. First, the global financial crisis revealed some weaknesses of the preexisting stress testing approaches. There is broad agreement that most stress testing exercises before the crisis—whether conducted by industry, national authorities, or the IMF staff in the context of FSAPs—failed to detect key macro-financial vulnerabilities (Haldane 2009a, 2009b). Second, the crises spurred the use of stress tests for crisis management purposes, notably in the United States and Europe, albeit with mixed results (Appendix 2.3). Third, these stress tests engendered a move toward routine system-wide stress testing for surveillance purposes, with stress test results being shared more widely.

So why did precrisis stress tests miss the vulnerabilities that eventually materialized?

\textsuperscript{13} The development of stress testing models has gathered momentum in the wake of the global financial crisis, supporting a plurality of different approaches with their own strengths and weaknesses. As the discussion in some of the following sections illustrates, the IMF staff is very active in this area, in close cooperation with advanced economy central banks and financial stability agencies and the academia. The annual Expert Forum on Advanced Stress Testing Techniques, co-organized by the IMF’s Monetary and Capital Markets Department, is one of the preeminent fora for exploring some of these new approaches among stress testing practitioners.

\textsuperscript{14} For instance, Annex 1.6 in the April 2010 GFSR (IMF 2010b) shows that a nonperforming loan projection model using data from Latin America and emerging market countries in Asia performs reasonably well in predicting loan quality in central and eastern European countries, where time series are relatively short.
Box 2.1. Balance-Sheet-Based and Market-Price-Based Approaches

There are two general approaches to estimating the impact of macro-financial shocks on bank solvency: balance-sheet-based and market-price-based. Balance-sheet-based models cover all material on- and off-balance-sheet exposures and the risks to which these are exposed. In contrast, market-price-based models are based on summary bank default measures embedded in asset prices (such as bank stocks, bonds, and derivatives). These measures are extracted from market prices by solving for the default probability and asset recovery implicit in them, using standard pricing formulas. The IMF staff often uses the contingent claims approach, a market-price-based approach built on option pricing theory on a bank-by-bank basis and/or system-wide to cover joint default risk within a banking sector (Gray, Merton, and Bodie 2008). When used for stress tests, market-price-based methodologies need to project the market-based default measures for the period covered by the test.

Both approaches have advantages and disadvantages. Balance-sheet-based models are more informative but are backward-looking, data intensive, and hard to update given the typical lag of accessible prudential/statutory information. In contrast, market-price-based measures are forward-looking, can be easily updated, and, in many cases, already incorporate portfolio effects (for example, the parametric approach of Credit Risk+ (Avesani and others 2008) as well as the Consistent Information Multivariate Density Optimizing copula function (Segoviano 2006) and systemic contingent claims approach (Jobst and Gray 2013). This alternative valuation approach is fundamentally driven by the intent to quantify possible linkages between financial assets and institutions in support of a more comprehensive assessment of common vulnerabilities than would be allowed by any single approach. However, they also reflect the impact of market swings, which may be unrelated to changing fundamentals.

While data limitations may dictate the preferred modeling approach in a specific set of circumstances, these two methodologies do not convey the same type of information and should be considered complementary rather than substitutes. Balance-sheet-based models are typically used by financial stability/supervisory authorities, who require disaggregated information on the sources of vulnerabilities to adopt risk-mitigating measures. Market-price-based approaches can be used when the focus is placed squarely on understanding the market assessment of banks’ loss absorption capacity under stress (which can also help supervisors prevent bias toward inaction for timely interventions that would not be possible using backward-looking information without any measure of uncertainty).

Because of their detailed nature, balance-sheet-based models face severe limitations in capturing all risks in an integrated manner. For this reason, in practice, some forms of risk have received overwhelming attention (such as default risk of private counterparties in the banking book) at the expense of others (such as sovereign default risk; downgrading risk of sovereign and private counterparties; counterparty risks of derivatives; and liquidity risks, including funding costs). Methodologies aimed at the valuation of all positions held by banks and the assessment of all risks have been introduced (see, for instance, Barnhill, Papapanagiotou, and Schumacher 2002; Barnhill and Schumacher 2011) but are data-intensive and require refinement. In addition, joint default risk within a system of firms varies over time and depends on the individual firm’s likelihood to cause and/or propagate shocks arising from the adverse change in one or more risk factors, which requires a closer examination of their historical volatility. Factors may be weakly correlated under normal economic circumstances but highly correlated in times of distress (as correlations become very high in downturn conditions).

A common challenge for both types of models is finding a way to stress test individual institutions to system-wide risks. Capturing the intrinsic dependencies across different types of risk is a major challenge for balance-sheet-based models. For example, counterparty risk is inherently dependent on the evolution of market risk; and the global crisis has shown that systemic liquidity risk cannot be assessed without considering the solvency profile of institutions. If risk factors are not fully correlated, it is reasonable to account for their dependence structure and combinations of stress testing parameters in which the individual impact of each risk is lower than the appropriate percentile for that risk in isolation (that is, the magnitude of change of a single risk factor relative to historical experience). This requires assessing measures of default dependence to produce estimates of systemic loss distribution. In the context of the balance-sheet-based models, this extension can be undertaken by superimposing a network of claims to keep track of default effects of one institution onto the others. Market-price-based models, on the other hand, typically treat the banking system as a portfolio of banks and derive a distribution of systemic losses using portfolio analysis techniques like those used for individual bank portfolios.

- **The institutional perimeter of stress tests was too narrow.** Stress tests did not, as a rule, cover what came to be known as “shadow banking” (for example, money market funds and insurance companies writing credit insurance), which played a key role in originating or transmitting shocks.
- **Key shock transmission and propagation channels were not covered.** Interconnectedness among key financial institutions through cross-exposures propagated and amplified shocks (as in the case of Lehman Brothers). Second-round feedback effects between the financial sector, the real economy, and sovereign risk were not incorporated in macro-financial stress testing.
- **Several important risk factors were not included.** Shocks hit multiple markets and countries at the same time due to common stress or contagion from one to another, which increased systemic risks. Several specific risk factors were also missed, including counterparty, basis, and contingent risks (BCBS 2009, 2012).\(^{15}\)
- **Balance sheet valuations did not fully reflect economic value.** Stress tests based on regulatory and accounting norms underestimated the resilience of the financial

\(^{15}\) Counterparty risk is a type of credit risk (for example, risk of default of a counterparty to a derivatives transaction). Basis risk for hedging is the risk that hedging becomes imperfect because there is either (1) difference between the asset, whose price is to be hedged, and the asset underlying the derivative; or (2) a mismatch between the expiration date of the hedge and the actual selling date of the asset. Contingent risks could arise either from legally binding credit and liquidity lines or from reputational concerns related to, for example, off-balance-sheet vehicles.
### TABLE 2.3
Comparing Balance-Sheet-Based and Market-Price-Based Approaches

<table>
<thead>
<tr>
<th><strong>Balance-Sheet-Based Approach</strong></th>
<th><strong>Market-Price-Based Approach</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary input data</strong></td>
<td>• Accounting data (balance sheet, income statement, overview of interbank exposures)</td>
</tr>
<tr>
<td><strong>Secondary input data</strong></td>
<td>• Probability of distress (and loss-given-default) or nonperforming loan ratios/loan classification of borrowers (for credit risk)</td>
</tr>
<tr>
<td></td>
<td>• Market data (equity prices, exchange rates, interest rates, price volatility) to calibrate shocks</td>
</tr>
<tr>
<td><strong>Type of test</strong></td>
<td>• Solvency, liquidity, and network analyses</td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td>• Varies depending on the reporting cycle (quarterly, semiannual, annual)</td>
</tr>
<tr>
<td><strong>Application</strong></td>
<td>• Most banks or financial systems (including emerging markets and developing countries) as long as financial reporting or supervisory data exist and are available</td>
</tr>
<tr>
<td><strong>Link(s) to macroscenarios</strong></td>
<td>• Possible, by estimating additional macro-financial model(s), linking macroscenario variables and risk factors (for example, probability of default of borrowers, nonperforming loan ratios)</td>
</tr>
<tr>
<td><strong>Estimation of systemic effects</strong></td>
<td>• By considering common macroshocks across banks (for example, GDP, inflation, and unemployment); and incorporating network effects (interbank exposures)</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>• Various capital ratios</td>
</tr>
<tr>
<td></td>
<td>• Liquidity ratios</td>
</tr>
<tr>
<td></td>
<td>• Capital shortfalls</td>
</tr>
<tr>
<td><strong>Strength</strong></td>
<td>• Pinpoints the type of risk exposing the vulnerability (for example, credit losses from housing loans, market valuation losses from exposures to sovereigns, losses from currency mismatches)</td>
</tr>
<tr>
<td></td>
<td>• Possible to adjust for supervisory weakness (for example, under-provisioning, and forbearance)</td>
</tr>
<tr>
<td><strong>Weakness</strong></td>
<td>• Data intensive (especially for network analysis)</td>
</tr>
<tr>
<td></td>
<td>• Quality of the analysis depends on the granularity and availability of the data.</td>
</tr>
<tr>
<td><strong>Selected Examples</strong></td>
<td>• Stress tester 101 (Čihák 2007)</td>
</tr>
<tr>
<td></td>
<td>• Next generation stress testing (Schmieder, Puhr, and Hasan 2011)</td>
</tr>
<tr>
<td></td>
<td>• Network analysis (Espinosa-Vega and Solé 2010)</td>
</tr>
<tr>
<td></td>
<td>• CreditRisk+ (Credit Suisse Financial Products)</td>
</tr>
<tr>
<td><strong>Market-price data</strong></td>
<td>• Financial market data (equity prices, bond yields, CDS spreads or equity option-based probability of distress)</td>
</tr>
<tr>
<td><strong>Secondary input data</strong></td>
<td>• Balance sheet data (combination of equity prices and accounting data to obtain key input variable, such as the Expected Default Frequency by Moody’s CreditEdge)</td>
</tr>
<tr>
<td><strong>Type of test</strong></td>
<td>• Largely focused on solvency and its interdependence among key financial institutions; incipient research on testing for liquidity indicators (Jobst 2014) or liquidity stress</td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td>• Daily or lower frequency</td>
</tr>
<tr>
<td><strong>Application</strong></td>
<td>• Limited to market data-rich countries and publicly listed institutions that is, cannot cover savings/cooperative banks, privately-held, or government-owned companies*</td>
</tr>
<tr>
<td><strong>Link(s) to macroscenarios</strong></td>
<td>• Stand-alone analysis for subsidiaries may be difficult</td>
</tr>
<tr>
<td><strong>Estimation of systemic effects</strong></td>
<td>• Possible, by estimating additional macro-financial model(s), linking macroscenario variables and risk factors (for example, probability of default of banks, volatility or leverage of banks)</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>• By considering common macroshocks across banks and incorporating interdependence (portfolio) effects among banks, which may be estimated using asset prices</td>
</tr>
<tr>
<td><strong>Strength</strong></td>
<td>• Expected losses</td>
</tr>
<tr>
<td></td>
<td>• Unexpected losses</td>
</tr>
<tr>
<td></td>
<td>• Probabilities of spillover among banks</td>
</tr>
<tr>
<td></td>
<td>• Less data intensive than the accounting-based approach</td>
</tr>
<tr>
<td></td>
<td>• Focuses on systemic risks/losses and tail events</td>
</tr>
<tr>
<td></td>
<td>• Incorporates risk factors priced by the market</td>
</tr>
<tr>
<td><strong>Weakness</strong></td>
<td>• The causes of different risks are difficult to disentangle (“black box”)</td>
</tr>
<tr>
<td></td>
<td>• Estimated vulnerability measures may be very volatile during periods when markets are under significant stress, and links with balance sheet fundamentals may be obscured</td>
</tr>
<tr>
<td></td>
<td>• Systemic CCA (Jobst and Gray 2013)</td>
</tr>
<tr>
<td></td>
<td>• CoVaR (Adrian and Brunnermeier 2008)</td>
</tr>
<tr>
<td></td>
<td>• Distress dependence (Segoviano and Goodhart 2009)</td>
</tr>
<tr>
<td></td>
<td>• SES and MES (Acharya and others 2010)</td>
</tr>
</tbody>
</table>

Source: Authors

Note: CCA = contingent claims analysis; CoVaR = conditional value at risk; MES = marginal expected shortfall; SES = systemic expected shortfall.

* The IMF staff has developed an adjustment approach for nonlisted firms for the systemic CCA-based stress test in the context of the Financial Sector Assessment Program for Hong Kong, SAR (IMF 2014b).

---

System. Market pressures on sovereign and financial sector exposures, for instance, were not fully reflected in the balance sheet of banks, which did not value every security at market value. And hurdle rates in solvency stress tests were based on regulatory minima, which proved insufficient when markets demanded larger capital buffers due to higher uncertainty about bank valuations, resulting in some banks “passing” stress tests only to face severe distress shortly thereafter.

- The shocks were not severe enough and did not examine genuine “tail events.” In some cases, the stress scenarios turned out to be too benign compared to the actual shocks. In other cases (for example, the unraveling of complex structured products), the empirical basis was too limited for calibrating the sensitivity of these instruments to changes in macro-financial conditions affecting their credit performance. Thus, time-varying correlation and extreme market risks,
which were the main contributors to tail risks, were not well reflected in stress tests (Rosch and Scheule 2008).

- Many shocks to identified risk factors were considered “unthinkable,” and, thus, were not included in the stress test. For instance, severe liquidity risks, involving the closure of key funding markets and its impact on bank solvency, were not included in stress testing exercises. Similarly, sovereign default risk in advanced economies was not covered. Both scenarios were considered too extreme to be plausible.

The design and implementation of stress tests should ideally be based on “best practice” principles that incorporate the lessons from past crises, include severe but plausible scenarios, and be sufficiently operational. Such principles should provide practitioners with operational guidance on how to tailor stress tests to a wide variety of country and sector circumstances while maintaining minimum standards that:

- ensure comparability across different exercises (which is particularly important for FSAPs (Jobst, Ong, and Schmieder 2013, 2017);
- protect against their most obvious pitfalls, especially considering the lessons from the global financial crisis; and
- interpret their results appropriately.

Efforts are being made to improve stress testing frameworks in light of this experience, but this is still an unfinished agenda. While it is important not to “fight the war of the past,” financial sector crises offer important insights into market dynamics (including the reaction function of key market participants) under stress. Current practices have incorporated some lessons learned from past crises, notably in terms of the types of scenarios and the severity of shocks, partly because shocks that materialized during the global financial crisis provide a good benchmark for tail risks. However, challenges remain with incorporating all relevant risk transmission channels, as well as feedback effects between the financial sector and the real economy. Appendix 2.4 presents some of the emerging methodological approaches at the frontier of stress testing techniques that attempt to tackle these challenges.

4. “BEST PRACTICE” PRINCIPLES AND ACTUAL STRESS TESTING PRACTICES

Stress testing practices have thus remained largely unsystematic. In many instances, established practices in various institutions, including the IMF, often reflect constraints in human, technical, and data capabilities on the institutions undertaking stress testing. They have emerged mostly out of trial-and-error efforts, or, in some cases, even as a matter of habit or convenience. Existing approaches do not always reflect a systematic effort to build a structured approach to stress testing.

This chapter proposes seven “best practice” principles for stress testing to address this issue with a view to promoting greater consistency and completeness. These principles are mainly focused on stress tests for macroprudential surveillance but they also contain elements that are generally applicable to all types of stress tests. As such, they are related to, but do not overlap with, the Principles for Sound Stress Testing Practices (BCBS 2018), which focus on banks’ own stress testing practices (Appendix 2.1). The remaining sections discuss the implications of these principles and evaluate to what extent actual practices correspond to them, based on the survey results (IMF 2012c). The seven principles are:

- **Coverage.** Define appropriately the institutional perimeter for the tests.
- **Risk transmission.** Identify all relevant channels of risk propagation.
- **Scope.** Include all material risks and buffers based on the total balance sheet.
- **Interpretation.** Make use of the investors’ viewpoint in the design of stress tests.
- **Calibration.** Focus on tail risks.
- **Communication.** When communicating stress test results, speak smarter, not just louder.
- **Limitations.** Beware of the “black swan.”

**Principle 1 (Coverage): Define Appropriately the Institutional Perimeter for the Tests**

This principle targets the selection of the institutions to be included in the tests. For system-wide tests, this involves a choice of which institutions to include and which to leave out. This choice requires an assessment of which banks are systemically important (that is, capable of triggering or amplifying systemic risk). Size, substitutability, complexity, and interconnectedness are the criteria that are used to assess the systemic importance of internationally active banks (see IMF/BIS/FSB 2009; BCBS 2011). These criteria are often mirrored in the assessment of systemic importance in the domestic context. A bank’s distress or failure is more likely to cause damage to other banks, markets, or the economy if its activities comprise a large share of financial intermediation. The larger the network of a bank’s contractual obligations in which it operates, the higher the likelihood that its failure will materially impact the default probability at other institutions. The systemic impact of a bank’s distress or failure is expected to be negatively related to its degree of substitutability as a market participant or service provider (in the

©International Monetary Fund. Not for Redistribution
case where specific institutions provide critical services, such as a market infrastructure). Thus, the systemic relevance is higher for more complex institutions (all else being equal), as the cost and time needed to resolve them are greater.

While size, degree of substitutability, and complexity are observable features, an assessment of interconnectedness requires the use of sophisticated network approaches. Network approaches can be useful in identifying systemic institutions that should be covered by stress tests (Box 2.2). Financial institutions hold claims against each other, which forms a network that can be thought of as a matrix of bilateral claims. Network models provide rich dimensions of interconnectedness and can identify systemically important institutions or groups of financial institutions that are at the center of the network, going beyond simple metrics of cross-exposures. They can also be used directly to measure the likelihood of multiple defaults by systemically important institutions, a key feature of systemic risk (stress testing individual institutions’ solvency and simply aggregating the outcome would tend to underestimate systemic risk—see Principle 5).

Separately, stress testing requires identifying the appropriate perimeter of financial activities that fall within the scope of the exercise. Some of the largest financial institutions have a wide range of activities, often straddling national boundaries and covering different industries (banking, insurance, pension fund, investment fund, various financial SPVs, and even nonfinancial activities). The ultimate ownership structure and economic links may not be always clear. And although some of the SPVs may be legally independent, the crisis illustrated that they could still imply contingent liabilities for the “parent” companies if the latter choose to provide support beyond legal obligations for reputational reasons. Analyses of global banks in their home countries often tend to focus on their activities in that country, but cross-border exposures may be relevant in assessing inward and outward spillover effects. However, in the context of host countries, limited prudential information on group-wide activities available to host country supervisors might put the most important risk—the health of the parent company—outside of the scope of stress tests. The stress tester needs to use judgment as to whether these activities should be consolidated or segregated.

Finally, defining the proper perimeter also calls for the inclusion of nonbank institutions that may trigger or propagate systemic risk, such as financial market infrastructures (FMIs) and insurance companies. The concern about FMIs, such as payment systems, central securities depositories, securities settlement systems, and central counterparties, is not their balance sheet per se, but rather their safe and reliable functioning. Their systemic importance is given by the key role played by the services they provide. Stress tests of insurance companies are also becoming more common. Although traditional insurance activities are not very likely to trigger systemic risk, insurance companies may engage in nontraditional, noninsurance (NTNI) activities, such as nonhedging-related derivatives trading and securities lending/repo, which can create systemic risk through their links with banks. In this regard, a network approach integrating all financial institutions would be less likely to leave out sources of potential systemic risk.

Applying this principle requires a good understanding of the system’s main features before undertaking the stress tests. This includes knowledge about the relevant market participants as well as their operations, business models, types of transactions, areas of risk concentration, and the likely channels of risk transmission. A formal mapping of this understanding into a network of claims and potential claims would facilitate the job, but other, more heuristic (and qualitative) approaches may also be used to get an understanding of the system, including market intelligence and conversations with market participants.

How well do actual stress testing practices correspond to this principle?

- Bank stress tests tend to be comprehensive and cover either all institutions in the system or focus on systemically important institutions. FSAP stress tests typically cover at least 70 percent of the locally incorporated commercial banks (including subsidiaries of foreign banks and state-owned banks) by assets (Jobst, Ong, and Schmieder 2013). Country authorities’ exercises focus on private domestically owned commercial banks, followed by foreign subsidiaries and state-owned banks. Based on the IMF staff survey results (IMF 2012c), the coverage ranges from 60 to 100 percent of the system by assets (with the median being 85 percent and 16 banks). The number of banks in the sample varies from below five to over 1,000.
- However, the methodology used to define the perimeter of the most relevant institutions tends to vary. The size of the balance sheet and interconnectedness are key factors, followed by the size of local retail activities and legislative definitions. Formal network approaches are increasingly used to assess interconnectedness, although the more sophisticated models are still used in a minority of cases, even in advanced economies. Various indicators of interconnectedness, in addition to size, played critical roles for determining the 25 jurisdictions with systemically

---

17 For this reason, the Committee on Payment and Settlement Systems (CPSS) and the Technical Committee of the International Organization of Securities Commissions (IOSCO) requires an assessment of FMIs using stress testing.

18 Although traditional insurance activities have not contributed to systemic risk during the financial crisis, the IAIS (2013) identified some vulnerabilities from NTNI activities in its initial assessment methodology for the identification of globally active, systemically important insurance firms. The weighted indicator-based approach for globally active, systemically important insurance firms is similar in concept to that used to identify global systemically important banks, but also introduces additional indicators that are germane to insurance activities.
Box 2.2. Network Models in Financial Stability Analysis and Stress Testing

Network models help measure interconnectedness, that is, linkages among financial institutions, systems, or entire countries, through claims held against each other or other channels. The importance of interconnectedness as a channel of contagion has been studied for some time. For example, Allen and Gale (2000), among others, explore how linkages among banks through direct exposures could be a source of contagion, which was underscored by the failure of Lehman Brothers in 2008.

Stress testing for financial market infrastructures also makes extensive use of network analysis. Nevertheless, it is important to keep in mind that the relationship between interconnectedness and financial stability is not simple and monotonic: interconnectedness may enhance or reduce financial stability, depending on the degree of cross-institution or cross-border integration and the precise pattern of cross exposures or other linkages (Čihák, Muñoz, and Scuozarella 2011).

Simple network models measuring interconnectedness through cross exposures have been used to add contagion channels to stress testing. The IMF’s introductory stress testing kit by Čihák (2007) includes a simple feature analyzing how a failure of a bank may affect other banks directly if it defaults on its borrowings. Espinosa-Vega and Solé (2010) and Tressel (2010) incorporate an additional channel: the failure of a bank may affect other banks indirectly because it stops the failed bank from lending to the other banks, thus eliminating a source of liquidity in the system. Such contagion analysis has been part of the joint Early Warning Exercise of the IMF and Financial Stability Board (with a focus on cross-border contagion using cross-border bank exposure data from the Bank of International Settlements) and stress testing in Financial Sector Assessment Programs. Network effects can be incorporated into stress tests in an ad hoc manner (selecting randomly the bank[s] that fail, or “trigger” banks) or incorporated into the main (balance-sheet-based) stress testing exercise, where the institution(s) failing the solvency or liquidity tests under a particular scenario become the “trigger banks.”

Advanced network models can analyze further dimensions of networks. These models typically examine four measures of interconnectedness, depending on the structure of cross-exposure links, to identify key nodes (institutions, financial systems, or entire countries) in the network: (1) “in-degree,” which is the number of links (or vertices) that point to a particular node; (2) “closeness,” which is the inverse of the average distance from one node to others; (3) “betweenness,” which focuses on the shortest path between nodes; and (4) “prestige,” which assigns increasing scores to nodes that are connected to other high-scoring nodes. When applied to financial data, these models can, among other things, describe the importance of financial centers in transmitting shocks around the world or identify banks that may be small but can play a critical role in connecting financial centers.

Cluster analysis separates the network into subgroups ("clusters") of nodes that have closer connections to each other than to those outside of the cluster. It can help identify subgroups of nodes with close connections and "gatekeeper" institutions or systems that bridge across different clusters (see Figure 2.2.1). This technique was used by the IMF to identify the 25 jurisdictions with the most systemically important financial sectors that are at the center of the global financial network (see Figure 2.2.2) (IMF 2010a). A similar technique was applied by the Reserve Bank of India to identify core institutions, including banks and nonbanks (Figure 2.2.3).

These network models can help both model contagion and provide input to stress tests. For instance, the Reserve Bank of India’s network model (Figure 2.2.3), based on the tiered network with a highly clustered central core, showed that a failure of a bank with large exposures to the insurance and mutual funds segments of the financial system could cause distress to 10 other institutions, including three insurance companies. Moreover, by identifying systemically important institutions at the core of the financial system, this approach can help set the right perimeter for stress testing.
important financial sectors that are required to undergo mandatory stability assessments under the FSAP (IMF 2010a).

- **Practices are much more uneven for the inclusion of nonbank institutions in stress tests.** FSAPs always test the banking sector, and occasionally the insurance sector; other sectors are rarely tested. In cases in which insurance is covered, effort is made to align the economic shocks to those used for the banking stress tests (Jobst, Sugimoto, and Broszeit 2014). But since each segment of the financial system may react differently to a certain macroeconomic scenario, the main scenario needs to be complemented with sector-specific scenarios. Moreover, insurers can also be vulnerable to structural and/or nonmacro risks (for example, catastrophic risks in the case of nonlife insurers, or long-term changes in mortality, the case of life insurers), which are captured by single-factor shocks but tend to be rarely covered in FSAPs.

- **Among country authorities, about 40 percent of the respondents indicate they only test the banking sector and another 45 percent also test the insurance sector.** Tests of pension funds and FMIs are undertaken on a much more ad hoc basis, if at all. The US SCAP and Comprehensive Capital Analysis and Review exercises include a life insurance company (MetLife) and the former auto loan arm of General Motors (Ally), in addition to investment and commercial banks, though these companies currently operate with bank holding company licenses. The Dodd-Frank Act allows the US Federal Reserve Board to use its discretion to include banks and nonbanks in the stress tests.

**Principle 2 (Risk Transmission): Identify All Relevant Channels of Risk Propagation**

In addition to network effects among financial intermediaries, there are other channels of shock propagation that relate financial intermediaries to each other and to other agents in the economy. Key examples of these propagation channels, illustrated in Figure 2.2, include:

- **The feedback between liquidity and solvency risks** (Wong and Hui 2009). This includes the (highly nonlinear)

---

19 For instance, banks and insurers often react to interest rate shocks differently, as banks typically have positive duration gaps, losing from rising rates, but insurers have negative duration gaps, gaining from higher rates.

20 The Dodd-Frank Act allows the US Federal Reserve Board to use its discretion to include banks and nonbanks in the stress tests.
Box 2.3. Integrating Liquidity and Solvency Risks and Bank Reactions in Stress Tests

Banks have numerous ways to react to credit and funding shocks. High-quality capital and profits are usually the first line of defense, and retained earnings can help buffer banks’ capital levels. Banks have an inherent capacity to generate liquid assets by using high-quality eligible securities as collateral for market or central bank funding if interbank markets freeze. As seen post-Lehman, fire sales of securities are also an option, but at considerable cost in an environment of sharply declining asset prices. Deleveraging, especially targeted at assets with higher risk weights, is a way to raise capital adequacy ratios by reducing risk-weighted assets. In practice, banks have been using a combination of these, as well as other hybrid measures, ranging from debt-to-equity conversions to issuance of convertible bonds to optimizing risk-weighted assets, to react to shocks.

Incorporating banks’ reactions to shocks is a critical input into the design of informative stress tests, especially over longer time horizons. This, however, requires modeling solvency and liquidity shocks in a coherent manner because first, when banks react to financial stress, the source of the shock (solvency or liquidity) is not always clear; and second, the measures banks take in reaction to these shocks have both capital and liquidity aspects that are not easy to disentangle.

A relatively simple (but somewhat ad hoc) way to integrate solvency and liquidity shocks is to conduct two-round stress tests, with a bottom-up first round and a top-down second round. If, for example, most banks report the sale of particular asset classes in response to stress, the source of the shock (solvency or liquidity) is not always clear; and, when banks react to financial stress, the measures banks take in reaction to these shocks have both capital and liquidity aspects that are not easy to disentangle.

Several analytical approaches have attempted to integrate solvency and liquidity more systematically:

- Schmieder and others (2012) simulate the increase in funding costs resulting from a change in solvency, indicated by a change in a bank’s (implied) rating.
- The Dutch Central Bank developed a stress testing model that tries to endogenize market and funding liquidity risk by including feedback mechanisms that capture both behavioral and reputational effects. Several central banks and bank supervisors have been successfully using this framework.
- The Hong Kong Monetary Authority sought explicitly to capture the link between default risk and deposit outflows. Their framework allows simulating the impact of mark-to-market losses on banks’ solvency positions, leading to deposit outflows, asset fire sales by banks, and a consequent sharp increase in contingent liquidity risk.
- Barnhill and Schumacher (2011) developed a more general empirical model, incorporating the previous two approaches that attempt to be more comprehensive in terms of the source of the solvency shocks and compute the longer-term impact of funding shocks.
- Another attempt to integrate funding liquidity risks and solvency risk is the Risk Assessment Model for Systemic Institutions developed by the Bank of England. The framework simulates banks’ liquidity positions conditional on their capitalization under stress, and other relevant dimensions such as a decrease in confidence among market participants under stress.

The feedback from financial stress to the real economy (Krznar and Matheson 2017). Bank reactions to financial stress (for example, deleveraging and capital flight) can have adverse effects on the real activity, which in turn further degrades bank asset values (second-round effects).

The feedback between financial sector stress and sovereign risk (Gray and Jobst 2010). Traditionally, the identification of risk from the relation between financial activities and the sovereign focused on the magnitude of contingent liabilities from implicit (or explicit) government guarantees to the banking sector (and the financial system more generally). However, the European sovereign debt crisis between 2010 and 2011 demonstrated that this relation is much more complex. This is discussed in more detail in Appendix 2.4.

Policy feedback. Policy reactions (or lack thereof) can have a significant impact on risk transmission and on the duration of a crisis. Aït-Sahalia and others (2012) show the importance of both the right choice and timing of policy interventions to be effective during the time of a crisis.

The design of stress tests requires a careful examination of transmission channels and a sound understanding of the range of possible reactions of financial institutions and capital market behavior to different shocks. While progress is being made on all of these fronts, there are still gaps in our understanding of the interaction between the real economy with the financial sector—the macro-financial framework—and of the role of the FMI and business practices in amplifying and transmitting negative shocks.

The operational implementation of this principle remains a major challenge, especially when it comes to the feedback effects between the real and financial sectors. As noted previously, reliable stress tests require the identification and calibration of propagation channels (with historical information or expert judgment) and their incorporation in stress test design and implementation in the face of incomplete information, including dealing with the tail risk arising from “unknown unknowns.” Progress is being made rapidly on models addressing some of the transmission channels highlighted in Box 2.3. Although recent research has

©International Monetary Fund. Not for Redistribution
enhanced our understanding of amplification mechanisms, such as those arising from funding shocks (for example, the adverse liquidity spiral as described in Brunnermeier and Pedersen [2009]), their incorporation in actual stress tests is difficult without a fully specified macro-financial model. Thus, some of the more salient transmission channels and effects are sometimes integrated in an ad hoc manner, by adding a second-round top-down stress test to bottom-up, bank-by-bank stress tests to assess potential bank reactions to the first-round shock. For example, a possible identification strategy could be informed by discussions with financial institutions about their responses to certain stresses (for example, portfolio allocation, deleveraging, and/or access to central bank funding). Such responses could help improve the design of stress test exercise, for example, designing the second-round shock by financial stability authorities. Against this background, it is no surprise that stress testing practice generally falls short in this area:

- Most country authorities that responded to the survey (IMF 2012c) and almost all FSAPs incorporate liquidity shocks in their stress tests, typically assuming the withdrawal of deposits and, in many cases, reductions in interbank exposures and access to secured funding (due to “haircuts” on liquid assets). In addition, in a few cases, countries also account for liquidity needs from off-balance positions and withdrawals of other types of wholesale funding. About half of the respondents consider domestic currency and foreign currency liquidity positions separately. However, most of these liquidity stress tests are implemented independent of solvency tests, excluding the possibility of experiencing a more severe run on liabilities when banks are likely to make substantial losses (illustrated in the case of Bear Stearns and the experience of some banks in the European periphery). Most large European banks compute their maximum risk tolerance (ECB 2008) utilizing a stochastic approach, which aims at estimating their “liquidity-at-risk” (maximum liquidity gap within a certain time horizon and for a given confidence level) or their “liquidity value-at-risk” (maximum cost of liquidity under certain assumptions). These approaches comprise single factor shocks and do not incorporate traditional credit risk stress or links to macroeconomic scenarios.
- Feedback effects from the financial sector to the macro-economy or the complex bank-sovereign linkages are rarely incorporated. In some FSAPs, fiscal-financial linkages have been explicitly modeled by incorporating findings from fiscal debt sustainability analysis on the valuation of bank exposures and funding costs (IMF 2012a), or by relying on market-price-based approaches (Appendix 2.4). But other types of feedback effects, such as the mutually reinforcing effect of banking sector deleveraging and declining credit growth, are not captured. This area would require a new generation of macroeconomic models that include a fully specified financial sector. Such research has been spurred by the crisis, in both policy institutions and academia, but it will take some time before sufficiently operational models are developed.22

- Network analysis or market-based systemic risk measures are increasingly used for the analysis of spillover effects. The FSAP for Luxembourg (IMF 2011c) examined network effects including bank-by-bank cross-border exposures, including exposures to parents and subsidiaries in the same financial group. Among country authorities, stress testing models such as the Bank of England’s Risk Assessment Model for Systemic Institutions model add contagion effects using network models. Market-based approaches (see Box 2.1) reflect interconnectedness across institutions in a reduced-form manner in a process of estimating systemic losses accounting for interdependence among institutions, for instance, in the case of the FSAP for the United States (IMF 2010a).

**Principle 3 (Scope): Material Risks and Buffers**

Capturing all quantifiable risks is key to obtaining reliable stress test results. Until the global financial crisis, stress tests typically focused on credit risk from customer loans and market risk from marketable securities. The crisis revealed that this coverage was incomplete, and other sources of risk, such as sovereign, funding, systemic liquidity, and counterparty risks, should also be included in the stress tests to widen the coverage of potential sources of shock. For internationally active financial institutions, incorporating cross-border exposures, credit and market risk of off-balance-sheet positions, and funding (including parent-subsidiary funding and liquidity transfers) is important for both home and host country authorities to assess the full risk profiles of the financial institutions.

Nevertheless, there are limits to the scope of risk factors to be included in stress tests. Some risks (for example, own sovereign) are so large and hard to hedge that financial institutions and—especially—entire systems are likely to be very vulnerable to them, and any risk reduction or mitigation (for example, through additional capital) is bound to be so costly as to be infeasible. Similar issues arise for system-wide li-

---

22 Two examples of the former include the Global Financial Stability Report (GFSR) (IMF 2011b), which presents a dynamic stochastic general equilibrium model with a banking sector and an econometric model incorporating the interactions between the financial sector and macroeconomy developed by the staff at the Bank of Japan (Ishikawa and others 2012). The IMF’s Research Department has also expanded its global dynamic stochastic general equilibrium model—Global Integrated Monetary and Fiscal model—to explicitly include the financial sector.
Stress test results depend not only on shock-related capital losses but also on preimpairment income (profit before losses from lending and investments), which contributes to capital. This factor is particularly important to consider in stress tests with longer risk horizons. In many cases, the size of preimpairment income can be substantial compared to the shock-related losses. But projecting preimpairment income is complex, and there is no widely agreed methodology to do so. Instead, stress testers employ a range of techniques.

The most straightforward way to project bank profit in line with macroeconomic conditions is to estimate the elasticity of preimpairment income (relative to capital) to economic growth. Hardy and Schmieder (2013) estimate this elasticity using BankFocus data for more than 16,000 banks. They find a decline of GDP growth by one percentage point reduces the ratio between preimpairment income and capital by between 1 and 1.5 percentage points. The average "preimpairment-income-to-capital ratio" in both advanced and developing economies is about 12 to 15 percent, and a moderate stress event (a 4 (6) percentage point drop in annual GDP growth in advanced [developing] economies) would reduce it by about half. However, under severe stress conditions, the elasticity increases to a factor of 4 in advanced economies, implying a strongly nonlinear relation between macroeconomic conditions and bank profits.

Depending on data availability, the main components of preimpairment income—net interest income, fee and commission income, and noninterest expenses (including salaries)—may be estimated separately. Net interest income could be projected in line with the interest rate assumption included in the macroeconomic scenario and assumptions for interest rate pass-through to bank borrowers. Hardy and Schmieder (2013) find that net interest income is more closely linked to macroeconomic conditions than other sources of income (such as fees and commissions), except for trading income under highly adverse conditions. There are also several studies (Ennis, Fessenden, and Walter 2016; IMF 2013b, 2017) that document a large influence of the level and slope of yield curve on banks’ net interest margin. Fees and commission income could be projected by examining their sources: if the majority is related to the sales and trading of securities, it is reasonable to project them in line with asset prices; if fees are mostly related to loan origination and credit cards, they should be linked to credit growth or employment.

The estimation of net interest income under stress could also incorporate interactions between solvency and liquidity stresses. When depositors and wholesale fund providers suspect that a bank may incur substantial losses, they would tend to reduce their exposure and/or raise the cost of funding due to higher counterparty risk. Higher funding costs, if not sufficiently passed onto borrowers, could reduce net interest margins. This is particularly true for wholesale funding costs, which are much more elastic than deposits (since at least part of the depositors’ funds are protected by deposit insurance) and harder to pass on to borrowers in weak economic conditions. Schmieder and others (2012) and Jobst (2014) have therefore constructed a method linking a bank’s funding cost to its solvency condition by using structural models for credit spreads (similar to Moody’s KMV). While funding costs remain relatively stable for well-capitalized banks, they rise sharply once banks approach the minimum regulatory capital requirement. Depending on how much of the higher funding costs banks can pass on, their income (that is, net interest income) can shrink substantially.

Preimpairment income represents an important buffer to liquidity shocks. This has led many to question the usefulness of stress testing these types of risk. Other risks have simply fallen out of fashion: for example, in an environment of low and stable inflation and interest rates due to more effective monetary policy regimes, refinancing and reinvestment risks arising from sudden asset-liability mismatches due to interest rate shocks, are often neglected. Regardless of the validity of some of these arguments, the incorporation of all risks in a stress test remains an issue of paramount importance to obtain a complete picture and guide the search for risk-mitigating solutions. Not all potential risks need to be addressed with additional capital, while comprehensive and candid stress tests could help gauge the consequences of inaction or delay (for example, in addressing sovereign debt sustainability).

Preimpairment income is an important buffer to the impact of risk factors contributing to losses and should also be part of stress tests. Most stress tests in FSAPs cover risk horizons of two years or longer. Over such a time period, profitability can significantly influence the stress test results. For instance, in many of the prudential stress tests before the crisis projected preimpairment profits exceeded estimated losses; and in the first EU system-wide stress tests, preimpairment profits absorbed the impairment losses due to credit and market risks. However, overly optimistic expectations about net interest margins, investment income, and potential fees and commissions in a stress scenario can mask the impact of losses. Therefore, it is important to properly understand the business conditions under stress. Separately, test results could differ depending on how the impact of stress scenarios on RWAs is incorporated; as with income, changes in RWAs can affect the stress test results, in some cases substantially.

Modeling the transmission channels between macroeconomic stress and nonimpaired components of income is a challenging task. It is particularly difficult for top-down exercises due to (1) the lack of sufficiently granular information (for example, which assets and liabilities are based on fixed or floating interest rates, maturities, and banks’ hedging practices); (2) the complexity of the sources of bank income; and (3) the likely changes in banks’ behavior under stress as they protect their balance sheets (for example, many banks attempt to shield profitability through higher fees and commissions when entering a downturn). Box 2.4 describes approaches used in FSAP stress tests to tackle these challenges. In bottom-up tests, banks have many degrees of freedom, partly because there is no widely accepted single model for parameterization for nonimpaired profits. Careful examination of individual banks’ models and comparison of assumptions and results across institutions with similar business models should thus be an integral part of any stress testing exercise.
The implementation of this principle requires a thorough understanding of the key transmission channels between risks and business conditions affecting bank portfolios in order to support a credible capital assessment under stress. This would enable the stress tester to capture all relevant risks and incorporate buffers appropriately in the design of the exercise. It also requires taking as comprehensive an approach to risks as possible, even if some of them may be “too big to mitigate.” Finally, this principle also provides an argument in favor of bottom-up over top-down tests to the extent that stress tests conducted by the banks themselves (with adequate supervisory scrutiny) may be more comprehensive and informative, since banks know better their portfolios and risks.

The experience of recent crises has spurred improvements in the way stress tests incorporate all material risks and buffers, but important gaps remain:

- **The global financial crisis and the European sovereign debt crisis, as discussed in Section 3, illustrate that many of the relevant risk factors were not sufficiently addressed in stress testing exercises.** Since then, the coverage of stress tests has expanded to include additional risk factors and transmission channels, often with greater specificity. Before the crises, stress tests focused on shocks to credit risk and asset prices (mainly equity prices and real estate) as well as liquidity and funding risks (IMF 2012c). After the crisis, liquidity and funding risks have gained greater prominence, and new risks emerged as important elements of stress test scenarios, including sovereign risk, regulatory risks, and spillover/contagion risks due to interconnectedness.

- **Sovereign risks have increasingly relevant in both FSAP and national authorities’ tests. System-wide stress tests in both Europe and the United States (EBA 2012, 2016, 2014; Board of Governors of the US Federal Reserve System 2012) as well as FSAP stress tests in advanced economies have covered sovereign risk since 2011.** In most cases, these are modeled as expected losses from market risks by applying valuation haircuts on sovereign debt securities. However, the methodologies differ across exercises, especially regarding the coverage of exposures (for example, stressing HtM exposures to own sovereigns, which is discussed more under Principle 4) and the size of the shock (which could be too small if shocks are calibrated based on historical data, as discussed under Principles 5 and 6).

- **Stress tests are expected to complement regular prudential oversight.** Since stress tests should primarily reflect the economic impact of certain shocks to the capital assessment, reported pretest capital adequacy ratios might need to be adjusted for potential under-provisioning, weaknesses in loan classification, collateral overvaluation, forbearance, and concentration risk from large exposures. However, these adjustments are judgmental and often involve large estimation errors, and should therefore be cross-checked with the expert judgment of the supervisory authority.

- **For system-wide exercises involving banks with different business models, it is challenging to ensure methodological consistency in projecting preimpairment income and capital intensity of exposures (that is, unexpected losses) without dismissing each institution’s idiosyncratic characteristics.** In bottom-up exercises, the supervisory authority or the FSAP team typically attempt to impose some harmonization by enforcing uniform assumptions on preimpairment income (for example, interest pass-through). However, the impact of a shock on RWAs can vary across countries and institutions, partly reflecting differences in regulatory treatment.

---


Market perceptions of solvency and asset values matter for the design of the stress test. Prior to the global financial crisis, the

---

23 Ong and Čihák (2010) discuss how the precrise stress tests on Iceland generated deceptively benign results by ignoring liquidity risk (deposit withdrawal). They illustrate that liquidity stress tests using detailed precrise disclosure information on funding positions could have indicated a vulnerability that materialized later.

24 Sovereign risk is ultimately a specific kind of credit risk. However, “sovereign risks” for advanced economies typically means sovereign market or spread risks, which are (unrealized) MtM valuation losses upon changes in market prices of sovereign bonds rather than outright default risks (Hannoun 2011).
This idea was implemented by the EU Capital Exercise (EBA 2012), which requires banks to value all HtM or AfS sovereign debt exposures to European Economic Area member countries to be valued at fair market value.26

26 This idea was implemented by the EU Capital Exercise (EBA 2012), which requires banks to value all HtM or AfS sovereign debt exposures to European Economic Area member countries to be valued at fair market value.

economic capital (and provide a more conservative basis for the capital assessment under stress).

- **Stressing market risk appetite.** Market risk appetite can be explicitly stressed when designing stress test scenarios. For example, extreme adverse scenarios could incorporate an increase in the market price of risk similar to that experienced during a severe global crisis—say, of the magnitude of the 2008–09 crisis. This historical scenario would have an important impact on the bank’s risk-adjusted balance sheet, leading to higher credit spreads and funding costs, higher potential losses to bank creditors, and lower equity values. If these effects are not accounted for, the distribution of bank losses under an adverse scenario may be underestimated and lead to an overly optimistic assessment of bank capital during stress.27

- **Imposing hurdle rates based on targeted funding costs.** Hurdle rates based on regulatory ratios reflect what regulators consider an adequate solvency ratio, but the markets’ assessment of a bank’s solvency may be different. And markets may demand—and banks would have an incentive to target—capital ratios that enable them to attain a certain risk rating and keep their funding costs under a certain ceiling. Using hurdle rates that reflect market views (in addition to the regulatory minima) in stress tests recognizes this simple but stark reality. Otherwise, if stress materializes, banks that “passed” the stress test may fail the market-based hurdle rate of capital adequacy. Box 2.5 presents two approaches developed by the IMF staff to calculate market-based hurdle rates. For supervisory authorities or central banks, which typically have better access to bank-specific and market information, the identification of the risk-return trade-off as perceived by the market should be simpler. Imposing hurdle rates that may be higher than regulatory minima is particularly important for macroprudential stress tests. Financial institutions must be sufficiently capitalized not only to ensure their own viability in the event of a system-wide shock, but also to prevent from becoming sources of risk propagation. This may well require higher capital than is necessary when a financial institution is considered on a stand-alone basis (OFR 2012).

This principle also has implications for the public disclosure of stress test results. Publishing stress test results can...
stress tests, including, but not necessarily limited to, capital injections.

Actual stress testing practices generally fall short of this principle:

- **Market-based hurdle rates, such as a market-based capital adequacy ratio, are generally not used by country authorities, and only rarely used in FSAPs (IMF 2010a; IMF 2011a).** However, following the crisis, funding costs are increasingly being modeled and incorporated explicitly in stress tests. While this is an important improvement, it is not enough. As shown in Box 2.5, the relationship between funding costs and bank

**Box 2.5. Market-Based Hurdle Rates**

As a complement to regulatory hurdle rates, market-based hurdle rates should be used in stress tests. This market-based hurdle rate can be calculated directly from the trade-off between banks’ capital ratios and their funding costs. For example, Schmieder and others (2012) derive a nonlinear relationship between solvency (measured as the capital adequacy ratios under the internal ratings-based approach consistent with banks’ implied default probabilities) and funding costs for a sample of German banks based on the following steps:

- Funding costs (that is, bank interest expenses measured as the excess spread above the government bond rate) were estimated for an average German bank for 12 quarters during 2007–2009.
- German banks’ weighted average MKMV’s expected default frequency were plotted against the funding costs for the same 12 quarters.
- The expected default frequencies are translated into capitalization ratios using the internal ratings-based formula to establish a link between funding costs and capital adequacy ratios as shown in Figure 2.5.1 (the mapping includes an additional capital cushion of 2.5 percentage points, in line with empirical evidence).

The results suggest that a regulatory capital ratio of about 15 percent would be a good benchmark for stable funding costs, which would be equivalent to a common Tier 1 capital ratio of about 11 percent. Below this level, the funding costs are very elastic to a decline in capitalization.

Alternatively, the contingent claims analysis, using option pricing theory, can be used to recover the banks’ probability of default implied in bank stock prices, the market value of bank assets (that is, the value of assets adjusted by the banks’ default risk), and the market value of risky debt. The latter can also be thought of as the risk premium required by the bondholders to compensate for the expected losses on their claims. This information can be used to construct the market-based capital adequacy ratio (that is, market capitalization to market value of implied assets), which can be mapped onto the corresponding spread required by banks’ creditors. Figure 2.5.2 presents these results for a sample of French banks. Again, the relationship is nonlinear and provides a methodology to determine a market-based hurdle rate.

![Figure 2.5.1. Basel II (EDF-Implied) Capital Ratio vs. Funding Costs for a Sample of German Banks, 2007–2009](image1)

**Sources:** Schmieder and others 2012; and IMF staff calculations.

**Note:** EDF = expected default frequency according to Moody’s KMV.

**Figure 2.5.1. Basel II (EDF-Implied) Capital Ratio vs. Funding Costs for a Sample of German Banks, 2007–2009**

![Figure 2.5.2. CCA CAR vs. Five-Year FV CDS Spread, for a Sample of French Banks, 2007–2009](image2)

**Sources:** Bloomberg L.P.; Moody’s KMV; and authors.

**Note:** CAR = capital adequacy ratio; CCA = contingent claims analysis; FV = fair value. CAR and spreads are (daily) median value for Crédit Agricole, Société Générale, BNP Paribas, and Natixis.

**Figure 2.5.2. CCA CAR vs. Five-Year FV CDS Spread, for a Sample of French Banks, 2007–2009**

help remove asymmetric information during periods of uncertainty and restore market confidence. Even in the case of stress tests undertaken for surveillance purposes during noncrisis periods, the public communication of their results could create awareness of risks, promote more realistic risk pricing, and enhance market discipline—which, in turn, could reduce the probability of future crises. However, for the publication of results to yield these benefits, stress tests need to be candid assessments of risk, explicit about the coverage and limitations, and the announcement of their results needs to be accompanied by measures that will convincingly redress any vulnerabilities unveiled by the

©International Monetary Fund. Not for Redistribution
solvency is nonlinear, which suggests that there are some market-related solvency thresholds that can and should be used as hurdle rates.

- **The assessment of sovereign risk is another area where market considerations are generally not incorporated in stress tests.** Regulatory and accounting standards could substantially underestimate the valuation risks of sovereign securities when a significant share of these securities is held in HtM accounts. Most FSAPs now include unrealized losses from securities in HtM accounts by applying MtM valuations in response to yield changes, recognizing that it is the economic rather than the accounting valuation that matters for sustainability. But some national authorities or financial institutions have not embraced this approach. However, the Bank of Japan regularly tests interest rate shocks on all securities, including those in HtM accounts, as part of its regular macroprudential surveillance.

- **Adjusting the initial capital for already realized shocks is often an integral part of supervisory and crisis management exercises.** For instance, the EU capital exercise (EBA 2012) required banks to mark to market all their sovereign bond holdings incorporating the large yield changes materialized in 2011. The stress test exercises conducted by Blackrock for Ireland and Greece for estimating recapitalization need included reassessment of existing loans by going through re-underwriting process. However, surveillance stress tests often do not include these adjustments.

### Principle 5 (Calibration): Focus on Tail Risks

The rule of thumb for stress tests has traditionally been to apply “extreme but plausible” shocks, but there is no systematic way to determine these. Typically, the size of shocks defined by a unit of measure or a qualitative attribute of severity, such as “worst-in-a-decade” events, “1 percent probability” tail events, or a “x-times standard deviation” shock, which is calibrated to a historical scenario of one or more (macro) variables. One obvious problem with this approach is that historical experience varies across countries and changes over time. A worst-in-a-decade scenario today would look very different than the same scenario in 2007. Another constraint, which became particularly relevant for stress tests conducted before the crisis, is the lack of sufficient historical price information on new financial products (for example, asset-backed securities) to calibrate adverse shocks. And obviously, an approach based on historical data would not work when considering an event that has never happened. BCBS (2004) has provided some quantitative guidance for determining tail risks for single-factor shock stress tests. However, there is no comparable guidance on macroeconomic scenarios.

This concern is especially pressing when stress tests take place in crisis or near-crisis situations. In these cases, the financial institution or system is already experiencing significant distress. Thus, some supervisory authorities may be reluctant to apply excessively negative tail scenarios to an already stressed baseline projection. Moreover, publishing the outcome of stress tests that incorporate extreme scenarios in these circumstances could trigger self-fulfilling crises. Political economy and legal constraints could also come into play when choosing scenarios, especially when the results could serve as basis for determining the restructuring or resolution of a failing banks (or the determination of public sector support). On the other hand, as indicated by the contrasting experience of SCAP and EBA tests (Appendix 2.3), compromising on the severity of scenarios could undermine the credibility of the exercise and prolong the crisis. These trade-offs are not easy to tackle. In principle, an effective stress test in a near-crisis situation should not compromise on the severity of scenarios, but instead mitigate the possible adverse market impact of the results by having credible support measures in place. If this is not feasible, it might be preferable not to conduct stress tests at all, but instead release critical exposure data and provide qualitative analysis to enhance transparency.

Another shortcoming of traditional approaches, regardless of the size of assumed shocks, is that they ignore the interdependence among shocks and among affected institutions or systems. For stress tests to provide a reliable assessment of the resilience of an institution or an entire financial system, they must consider not just the potential size of individual risk factors but also their dependence under all plausible scenarios. In addition, risk factors may be weakly correlated under normal economic circumstances but highly correlated in times of distress. Similarly, the joint default risk within a system varies over time and depends on the individual firms’ likelihood to cause and/or propagate shocks arising from individual risk factors. Given that large shocks are transmitted across entities differently than small shocks, measuring nonlinear dependencies in stress tests can deliver important insights about the joint tail risks.

But while it is relatively straightforward to generate stress test results based on the effect of a single risk on a static measure of capital, combining multiple risks and the extent to which all of these risks are correlated, the extent to which shocks to one subset of risk factors induce negative feedback on another subset of risk factors, and the extent to which shocks to one subset of risk factors induce positive feedback on another subset of risk factors can become extremely complex.

30 BCBS (2004) suggests for G10 currencies either a ± 200-basis-point parallel rate shock or the first and 99th percentiles of observed interest rate changes using a one-year risk horizon period based on a minimum of five years of observations for calibration.
31 In these instances, there is a considerable shift of the average away from the median (“excess skewness”) and a narrower peak (“excess kurtosis”) of the probability distribution. If distributions become highly skewed, large tails may even cause the mean to become undefined, which is an important complication when using stress tests.
which firms’ default risks may be correlated under different scenarios is very complex (Box 2.6):

- In the context of balance-sheet-based models, network models can help capture spillover effects and the probability of many institutions defaulting (Box 2.2). However, most conventional balance-sheet-based stress test models do not formally account for default dependencies across institutions. In this case, complexity arises from the amount of information necessary to identify the right network.
- In market-price-based models, it is possible to model joint default probabilities. In this case, the key challenge is modeling the dependence structure. One approach used frequently is estimating conditional VaR, which measures the VaR of the financial system conditional on the distress of one or more financial institutions in the system. The value of an institution and distress correlations among institutions are estimated using equity or other market price data. Another related approach is the marginal expected shortfall (Acharya and others 2010) that specifies historically expected losses of a financial institution conditional on the financial system having breached some systemic risk threshold. Adjusting marginal expected shortfall by the degree of firm-specific leverage and capitalization yields the systemic expected shortfall, which yields the average, linear, bivariate dependence between banks when the entire banking sector is undercapitalized. More complex approaches require a significant departure from conventional statistical methods (Box 2.6).

Stress testing practices in this area are evolving rapidly in line with the development of new analytical tools.

More recent FSAPs are making a systematic effort to include severe shocks that are comparable across countries.32 One rule of thumb applied in many FSAPs is to assume a two-standard-deviation shock on GDP growth rate for two years based on a long (20–30 year) history, unless a different magnitude of shock (in either direction) is justified. Among country authorities, most survey respondents adopt tail events with small probability (ranging from 1–5 percent), while some use qualitative criteria, such as in line with, or worse than, historical worst (IMF 2012c). Since the global economy experienced extremely sharp economic deterioration in 2008–09, shocks calibrated using a sample that includes the crisis time observations should generate fairly severe shocks. Another frequently used approach is to test for a shock of similar magnitude as the global financial crisis. At the same time, all historical data-based approaches always carry the risk of complacency. This needs to be managed by more qualitative approaches (Principle 6).

Incorporating risk interdependence in stress tests continues to remain a difficult technical challenge. Based on the survey results, several central banks and supervisory authorities attempt to capture risk interdependence heuristically by applying common macro and financial shocks and assessing interbank contagion through network models (IMF 2012c). The IMF staff has also been increasingly using network models on a stand-alone basis or as a part of a stress testing model. Beyond these heuristic approaches, some FSAPs have

---

**Box 2.6. Risk Interdependence in System-Wide Stress Testing**

For stress testing to inform a realistic assessment of capital adequacy, the objective is to consider both the variability of risk factors and their mutual dependence under all plausible scenarios (in addition to their likelihood and severity). While it is generally straightforward to generate stress test results based on the effect of a single risk on a static measure of capital, combining multiple risk factors under different scenarios—and on a system-wide basis—complicates a reliable capital assessment.

Estimating the dependence of risk factors to account for the low probability of joint (negative) extreme outcomes (with no or little historical precedent) is not straightforward and requires a significant departure from conventional statistics. The traditional correlation coefficient detects only linear dependence between two variables (or risk factors) whose marginal distributions are assumed to be normal. This statistical inference assumes an empirical relation (or the lack thereof) based on relatively more central (and more frequent) observations, and implies that the bivariate distribution of these variables is elliptical, which is hardly encountered in reality. Alternative measures of dependence between risk factors can capture the nonlinear dynamics of changes in variables far removed from the median. For example, an expedient nonparametric method of investigating the bivariate empirical relation between two random vectors is to ascertain the incidence of shared cases of cross-classified extremes via a refined quantile-based chi-square statistic of independence. Similarly, copula functions and other nonparametric methods provide the possibility of combining two or more distributions of variables based on a more flexible specification of their dependence structure at different levels of statistical significance. These approaches generate measures of so-called “joint asymptotic tail dependence,” which define the expectation of common extreme outcomes.

The measure of nonlinear dependence of risk factors can be combined with their individual severity distribution (or the risk profile of each bank) to derive a tail-sensitive estimate of joint default risk and the system-wide capitalization of multiple institutions under stress. For instance, closed-form methods under extreme value theory, such as the generalized Pareto distribution, are frequently used to help define the limiting behavior of different risks affecting the operating performance of a bank in extreme situations.1

---

1 Two prominent examples are the generalized extreme value distribution and the generalized Pareto distribution. While generalized extreme value helps model the asymptotic tail behavior of the order statistics of normalized maximums (or minimums) drawn from a sample of dependent random variables, the generalized Pareto distribution represents an exceedance function that measures the residual risk of extremes beyond a given threshold (that is, a designated maximum [or minimum]) as the conditional distribution of mean excess.
One example of this approach in IMF stress testing is the two-step estimation of joint expected losses within the Systemic Contingent Claims Analysis (Jobst and Gray 2013) framework. This alternative approach is fundamentally driven by modeling extreme dependence to quantify possible linkages between expected losses in support of a more comprehensive assessment of common vulnerabilities than would be allowed by balance-sheet-based approaches. After estimating the nonparametric dependence function of individual expected losses, it is combined with their marginal distributions, which are assumed to conform to a generalized extreme value (GEV) distribution, which identifies the asymptotic tail behavior of normalized extremes. The dependence function is estimated iteratively on a unit simplex that optimizes the coincidence of multiple series of cross-classified random variables. Finally, the conditional value at risk of joint expected losses is determined as the point estimate of the probability weighted residual density beyond a prespecified statistical confidence level.

The estimates of joint default risk under extreme scenarios can be used to cross-validate results from traditional, accounting-based stress testing approaches. In several Financial Sector Assessment Programs, most notably, in the cases of Germany, Sweden, the United Kingdom, and the United States (IMF 2010a, 2011d, 2011e, 2011f), institution-level stress tests of the banking sector were combined with systemic contingent claims analysis to derive capital assessments from a system-wide perspective—after controlling for the market-implied dependence structure of risk factors and their sensitivity to extreme events.

In general, incorporating time-varying extreme dependence of risk factors—ideally together with a market-derived measure of capital adequacy—offers a more realistic capital assessment in stress tests. This approach identifies the amount of capital shortfall to current capital levels based on market expectations on solvency that far exceed the minimum regulatory requirements. For instance, the assumption of higher uncertainty about the realization of expected losses (that is, greater historical volatility) reflects the notion that, especially during distress periods, firms would need to have higher capital buffers in place to absorb the realization of losses above existing provisioning levels so that current capital levels remain unaffected over a specific risk horizon.

Figure 2.6.1. Key Conceptual Differences in Loss Measurements—Implied Capital Requirement under “Distributional Approaches”

Sources: Jobst and Gray 2013; and authors.
Note: VaR = value at risk.

---

2 In contrast to the traditional (pairwise) correlation-based approach, this method links the univariate marginal distributions of expected losses in a way that formally captures both linear and nonlinear dependence in joint asymptotic tail behavior over time.
tried to include some of these nonlinear dependences formally in the stress testing framework. In the FSAPs in Germany, Sweden, the United Kingdom, and the United States (IMF 2010a, 2011d, 2011e, 2011f), institution-level stress tests were combined with an attempt to derive capital assessments from a system-wide perspective, which considered losses from a portfolio of key financial institutions after controlling for the dependence structure of risk factors and the stochastic nature of input parameters. Technical limitations, however, mean that it will take some time before these approaches are “mainstreamed” in FSAPs.

Principle 6 (Communication): When Communicating Stress Test Results, Speak Smarter, Not Just Louder

The experience of the global crisis underscored the importance of effective public communication of stress test results. Central banks and supervisory authorities in several countries published stress test results in their financial stability reports even before the crisis, at varying degrees of detail. FSSAs—most of which are published with the consent of national authorities—always include stress test assumptions and aggregate results (often supplemented by a more detailed description of data coverage and methodologies in Technical Notes). In the wake of the financial crisis, national authorities enhanced the disclosure of stress test results, especially in the United States and Europe, where policymakers saw it as a way to shore up market confidence. In the United States, the publication of the stress test results was established in national law.33 This spurred greater public interest in—and scrutiny of—stress tests, which, in turn, increased pressure for greater disclosure.

Public disclosure of stress tests can yield significant benefits but is no panacea. The US SCAP successfully restored market confidence in the banking sector, allowing investors to differentiate between banks based on their resilience, which arguably facilitated the raising of additional capital from private sources. In contrast, the 2011 EU system-wide stress test did not fully achieve this objective. This was not due to differences in the scope of public disclosure between the two exercises—in fact, the latter was praised for its transparency. Rather, it reflected differences in the design of the stress tests and the context within which their results were published. While the setup of the SCAP was credible (combined with a fiscal backstop), clearly communicated, and in place ahead of time, the European stress test was considered mild and not fully capturing the risk profile of weaker national banking systems. And perhaps more importantly, the follow-up actions and policy backstops for failing banks were considered ambiguous (Appendix 2.3).34

Public disclosure may also create difficult trade-offs in some cases. Public disclosure of stress test methodologies, underlying exposures, assumptions, and results can help (1) raise public awareness of risks; (2) promote more realistic risk pricing and strengthen market discipline, thereby reducing the probability of future sudden reversals of investors’ sentiment; and (3) inform more effective financial stability policies. Even when the results are weak, public communication can have a positive impact if accompanied by credible contingency plans and support measures for financial institutions that fail the tests, reflecting the authorities’ recognition of the problems and commitment to financial stability. At the same time, greater disclosure carries risks. It may (1) entice financial institutions to make portfolio choices to “game” the tests; (2) increase moral hazard and encourage complacency if investors rely excessively on published stress test results—which are always subject to a margin of error (Principle 7)—at the expense of other bank soundness indicators; and/or (3) undermine confidence if the necessary support measures are not in place (for political economy or other reasons). Also, as stress tests become more common, disclosure of different results can cause confusion—a problem encountered in FSAPs for EU countries when they took place concomitantly with EU system-wide stress tests.

Balancing the benefits and costs of disclosure depends partly on the circumstances and the nature of the tests. In a crisis situation, when market confidence is at a premium, a strong case can be made in favor of greater public disclosure of system-wide stress test results. In normal times, a more conservative approach might be warranted. Even then, however, regular publication would familiarize market participants to stress tests, making them a more effective tool at times of crisis. This is particularly true for system-wide stress tests (rather than microprudential stress tests as part of capital adequacy reviews), which are more likely to apply consistent assumptions across financial institutions, making the results comparable and guiding the attention on systemic risk factors that are relevant for maintaining or restoring market confidence.

Realizing the benefits of greater disclosure depends on a number of crucial preconditions. The stress tests should be credible. For this, they need to cover the relevant risks and transmission channels, assume serious shocks, set appropriate hurdle rates, and produce a candid assessment. And crucially, they should be accompanied by a convincing framework for follow-up action, including government support, if needed. If these preconditions are not met, disclosure would not be informative and might do more harm than good.

Lastly, the disclosure of stress tests should be set in the perspective of the broader communication strategy of financial stability policies. Just as stress tests should be one of several assessment tools, so should their disclosure be part of a coherent overall communication strategy. Publishing stress test results is likely to be much more effective if done in the context of regular outreach aimed at informing markets and

33 The Dodd-Frank Act requires the Board of Governors of the US Federal Reserve System to disclose summary results of stress tests for large banks.

34 See Ong and Pazarbasioglu 2014.
the public about financial stability issues, and complemented by disclosure of a broad set of indicators.

The survey of actual practices confirms the trend toward greater disclosure (IMF 2012c):

- Most surveyed countries communicate the results of macroprudential stress tests (almost 85 percent in the case of solvency tests and 50 percent in the case of liquidity tests). Usually, this takes place in annual (or semiannual) financial stability reports. In most cases, results are communicated using system aggregates, often with some distribution metrics, without disclosing the identity of individual institutions—with the notable exceptions of the EBA and US Federal Reserve tests. FSAP stress test results are reported in FSSAs, which typically do not disclose the identity of individual institutions. Most FSSAs are published. Technical Notes on stress tests, which report much more detailed results, are increasingly published, especially in advanced economies.
- Raising public awareness of financial stability issues, achieving greater transparency, and providing information to market participants are mentioned as the main objectives of public communication. Although public communication is seen positively by all respondents who publish, many expressed concerns about the risk of exaggerated expectations placed on stress tests, inconsistent interpretation of stress tests by mass media, and excessive focus by banks on published stress test results that could undermine their effectiveness as a supervisory tool.

Principle 7 (Limitations): Beware of the “Black Swan”

There is always a risk that the “unthinkable” will materialize, regardless of how extensive the coverage of risk factors, how refined the analytical models, how severe the shocks incorporated in the stress tests, and how careful the communications strategy. Stress tests provide a measure of the resilience of a financial institution or a system to given shocks but cannot predict the future. However, future shocks might arise from new sources and unexpected events, which have historically shown little volatility or have not materialized for such a long time that they have been forgotten (for example, advanced country sovereign defaults). What practical ways are there to incorporate these factors into stress test design?

One approach is to design hypothetical scenarios based on expert judgment and new information, where available, rather than simply be guided by history. The Bank of England (Haldane, Hall, and Pezzini 2007) proposed using current vulnerabilities as a guide for the choice of hypothetical shocks. This means, for example, that systems concentrated in real estate deserve a stress test on the impact of a large decline in real estate prices regardless of the probability of such a shock. Alternatively, the US Federal Reserve Board typically uses two scenarios for stress tests. One is unique to each institution and chosen by it; the other is common. In this way, the institutions are assessed under scenarios that they themselves consider particularly damaging. Reverse stress tests by individual institutions (Appendix 2.1) and surveys of such exercises across institutions could help extend the frontier of tail risks.

Another approach is the application of distribution theory to the scenarios themselves, as opposed to the current practice of choosing just one adverse scenario. This reflects the recognition that the future is stochastic and can be represented by a number of event combinations, each of which with a probability of realization. A scenario distribution approach was used by the IMF staff in the first FSAP for South Africa. Based on the statistical properties of the historical distribution of price changes, and using Monte Carlo simulations, each scenario was represented by a combination of changes in prices, including credit spreads, which were used to revalue bank assets. The final outcome represented a distribution of bank capital ratios for each bank, in which each point of the distribution was associated with a particular scenario.

Ultimately, the principle of being aware of the “black swan” is more about the context of stress test scenarios than about the mechanics of their design and implementation. It serves as a reminder that stress tests should not be undertaken in isolation and their results should not be taken too literally. No matter how much a stress tester tries, stress tests always have margins of error. Their results will almost always turn out to be optimistic or pessimistic after the fact. In addition, there will always be model risk, imperfect data access, or underestimation of the severity of the shock. One should therefore set stress test results in a broader context.

Stress testing is just one of the many tools to assess key risks and vulnerabilities in financial institutions or entire systems. They should be treated as complements to other tools that can provide information about potential threats to financial stability, such as qualitative and quantitative bank risk analysis, early warning indicators, models of debt sustainability, and informed dialogue with supervisors and market participants, among others. Final conclusions about the resilience of the institution or system should draw on all these sources and not just on the results of stress tests.

---

35 The term “black swan” was first used by Taleb (2004) to indicate highly improbable events that have a major impact.

36 There are also policy trade-offs in designing shocks. On one hand, it is important to prevent a Type II error of allowing weak banks to pass the test. Dexia is a good example of a bank that passed the July 2011 EBA stress test but failed shortly afterwards. On the other hand, it is not useful to design shocks that are too severe, causing an excessive failure rate of banks that would be deemed sound under most adverse scenarios.

37 The methodology for this approach is discussed in Barnhill and others 2002, which is similar to Borio, Drehmann, and Tsatsaronis 2014, which suggests that stress tests themselves need to be stress tested by assessing the sensitivity of results to changes in assumptions.
Stress test design, models, and implementation should be “back-tested” to the extent possible and regularly reassessed. Back-testing can take the form of a comparison of stress test outcomes under baseline scenarios with actual outcomes. For adverse scenarios, one could have stress test results reviewed by a panel of external experts to assess their rationale and consistency of results across banks. Checking the robustness of the results for variations in key parameters (in other words, stress testing the stress test), assessing the impact of new tools and new approaches and, last but not least, remaining vigilant for the emergence of new risks are crucial to ensuring more reliable tests.

5. CONCLUSIONS AND OPERATIONAL IMPLICATIONS

This chapter contributes to the debate on stress test design and implementation by proposing a set of operational best practice principles. The wide variety of stress testing approaches and underlying assumptions have raised important questions about the interpretation of the test results and their comparability. Setting best practice principles and ensuring adherence to these could improve the integrity of stress testing exercises, promote greater transparency and comparability of stress tests across countries and over time, and ultimately contribute to more meaningful and effective financial stability assessments.

A key goal of this chapter was to set realistic expectations about what stress tests can and cannot accomplish. Stress tests are forward-looking tools to assess financial institutions’ solvency and liquidity and the resilience of the entire financial system under possible adverse scenarios, but they do not predict the likelihood of these scenarios materializing. As such, regardless of refinements and improvements, they will always remain hypothetical statements. One should therefore always be cautious about using stress test results in isolation: a well-rounded risk assessment should use stress tests in conjunction with other tools to broaden the understanding of vulnerabilities.

The discussion highlighted several important decisions in the design and implementation of stress tests. These involve the choice of (1) risk scenarios—in terms of both the coverage of all relevant risk factors and their severity, (2) types of tests so that they cover all relevant transmission channels and include realistic assumptions about buffers, and (3) appropriate hurdle rates. These decisions are key for the effectiveness of stress tests and the reliability of their results. When pressures from the industry (especially in the case of bottom-up tests) or political economy or other considerations unduly influence these decisions, stress tests can do more harm than good. Deriving benign conclusions from stress tests that assume modest shocks, include optimistic projections of future income, and use trivial hurdle rates can lead to complacency. Disseminating publicly such results could undermine the credibility of the exercise.

However, the success of stress tests cannot be reduced to the choice of a few parameters but should be seen in the broader context outlined by the proposed principles. Stress tests are complex exercises with many “moving parts.” Their effectiveness does not depend on just a few parameters or on the degree of public disclosure of their results but also on the context within which they are conducted. This context includes (1) a clear understanding of the stress tests’ scope and objectives; (2) knowledge of the key individual financial institutions in the system, their business models, and main channels of risk transmission; (3) appropriate decisions on the tests’ perimeter and coverage; (4) consideration of complementary assessment tools; (5) a communications strategy tailored to the circumstances and purpose of the tests; and (6) a credible commitment to take the measures that may be required to address vulnerabilities uncovered by the tests.

Adherence to these principles is uneven in practice. Table 2.4 summarizes key operational implications that flow from each principle. For Principles 1–3, the recommendations focus on the preparatory work that needs to be conducted by the stress tester to conduct tests with adequate coverage of institutions, risks, and channels of risk transmission. For Principles 4–7, the recommendations focus on stress test design and communication. A survey of central banks and supervisory authorities in 23 countries and stress tests in FSAPs showed that despite major improvements since the crisis, practices still fall short of these principles (IMF 2012c; Jobst, Ong, and Schmieder 2013). Shortcomings are particularly notable in three areas, which reflects both gaps in the analytical toolkit and weaknesses in implementation: (1) identifying the channels of risk propagation, (2) using the investors’ viewpoint, and (3) focusing on tail risks. The proper interpretation of stress test requires a sufficient understanding of the implications of these gaps, which the current generation of stress testing exercises aims to close with priority.
TABLE 2.4
Practical Implications of “Best Practice” Principles for Stress Testers

<table>
<thead>
<tr>
<th>Principle</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td><strong>Define the appropriate institutional perimeter.</strong></td>
</tr>
<tr>
<td>-</td>
<td>For system-wide stress tests, if universal coverage is not an option, identify systemically important institutions to be covered in the tests, including relevant nonbanks and financial market infrastructures. Criteria for assessing systemic importance should include an assessment of interconnectedness.</td>
</tr>
<tr>
<td>-</td>
<td>Gain a basic understanding of the structure of financial conglomerates, if present, and cover in the tests any of their banking or nonbanking activities that may have a significant impact on financial stability during times of stress.</td>
</tr>
<tr>
<td>2.</td>
<td><strong>Identify all relevant channels of risk transmission.</strong></td>
</tr>
<tr>
<td>-</td>
<td>Identify and understand the main channels of risk propagation, including but not limited to: (1) the relation between solvency and liquidity conditions, (2) risk transfers between banks and the sovereign, (3) potential bank reactions to stress, and (4) feedback effects between financial sector stress and the macroeconomy (for example, through deleveraging and lower GDP growth).</td>
</tr>
<tr>
<td>-</td>
<td>Incorporate as many of these channels in the design of the stress tests as possible. If modeling them explicitly is not feasible, use heuristic means, such as a second-round top-down test that maps the impact of first-round tests on individual institutions onto key macro-financial variables (asset prices, credit growth, aggregate liquidity) or expert judgment, to capture all potential channels of risk transmission.</td>
</tr>
<tr>
<td>3.</td>
<td><strong>Include all material risks and buffers.</strong></td>
</tr>
<tr>
<td>-</td>
<td>Understand the tested institution’s— or, for system-wide tests, the systemically important institutions’—business model, including markets where they operate and their sectoral or cross-border exposures.</td>
</tr>
<tr>
<td>-</td>
<td>Be as comprehensive as possible in including potential sources of risk in the tests. Do not omit risks because they may be “too big to mitigate.”</td>
</tr>
<tr>
<td>-</td>
<td>Assess whether any banks may be exposed to risks from cross-border exposures, non-banking activities (such as employee pension funds and investment funds) or other institutions that may be legally separate but with significant economic links (such as parent companies, subsidiaries, and off-balance-sheet vehicles), and include as many of these as possible in the tests.</td>
</tr>
<tr>
<td>-</td>
<td>If stress tests include non-banking institutions, consider relevant noneconomic risks that may have a major impact on them and the rest of the system (such as natural disasters for insurance companies).</td>
</tr>
<tr>
<td>-</td>
<td>In stress tests with longer time horizons, model explicitly likely buffers, such as preimpairment income. Preimpairment income should be scenario-dependent and conservatively projected.</td>
</tr>
<tr>
<td>4.</td>
<td><strong>Make use of the investors’ viewpoint in the design of stress tests.</strong></td>
</tr>
<tr>
<td>-</td>
<td>Adopt market-consistent valuation of bank assets and liabilities and point-in-time parameters to measure expected and unexpected losses.</td>
</tr>
<tr>
<td>-</td>
<td>Adjust institutions’ initial capital for supervisory practices or other factors that result in systematic over-reporting of capital, such as underprovisioning, collateral overvaluation, and forbearance. Use economic capital estimates as a complement to statutory capital in the stress tests.</td>
</tr>
<tr>
<td>-</td>
<td>Stress market risk appetite by including explicitly shocks involving increases in the market price of risk.</td>
</tr>
<tr>
<td>-</td>
<td>Use hurdle rates reflecting market views (for example, target funding costs) in addition to regulatory minima.</td>
</tr>
<tr>
<td>5.</td>
<td><strong>Focus on tail risks.</strong></td>
</tr>
<tr>
<td>-</td>
<td>Use a variety of approaches to determine “extreme but plausible” shocks to be used in stress tests, including calibration on historical data, cross-country experience, and “worst-ever” events.</td>
</tr>
<tr>
<td>-</td>
<td>Whenever possible, adopt methodologies that capture tail events and risk dependencies. Ideally, use both balance-sheet-based models and market-price-based models, as the latter can incorporate more easily measures of risk dependencies using market prices. If using balance-sheet-based models alone, supplement stress tests of individual financial institutions with tools that measure spillovers and joint default probabilities, such as network models.</td>
</tr>
<tr>
<td>6.</td>
<td><strong>When communicating stress test results, speak smarter, not just louder.</strong></td>
</tr>
<tr>
<td>-</td>
<td>Tailor the communication of stress test assumptions, methodologies, and results to circumstances and the goals of the tests. Disclosure can yield substantial benefits for macroprudential stress tests, especially in crisis situations, when market confidence is at a premium.</td>
</tr>
<tr>
<td>-</td>
<td>Before disclosing stress test results, ensure that the tests cover relevant risks and transmission channels, assume serious shocks, set appropriate hurdle rates, produce a candid assessment, and are accompanied by a credible framework of follow-up action, including government support, if needed.</td>
</tr>
<tr>
<td>-</td>
<td>To the extent possible, set the communication of stress test results in the context of a broader communication strategy for financial stability objectives and policies, and complement them by other financial soundness indicators.</td>
</tr>
<tr>
<td>7.</td>
<td><strong>Beware of the “black swan.”</strong></td>
</tr>
<tr>
<td>-</td>
<td>Supplement statistical approaches to the identification of potential tail events with expert judgment and new information. Consider analyzing the sensitivity of results to variations in the assumed shock, instead of choosing just one adverse scenario.</td>
</tr>
<tr>
<td>-</td>
<td>Supplement stress tests with other assessment techniques, such as qualitative and quantitative bank risk analysis, early warning indicators, models of debt sustainability, and informed dialogue with supervisors and market participants, among others. Final conclusions on the resilience of an institution or a system should be based on all these sources, not just the results of stress tests.</td>
</tr>
<tr>
<td>-</td>
<td>Evaluate and regularly reassess the institutional coverage, risks, and channels of risk transmission included in the tests as well as the models and other approaches. Constantly assess the suitability of new tools. Remain vigilant for the emergence of new risks.</td>
</tr>
</tbody>
</table>

Source: Authors.
Appendix 2.1.
Existing Supervisory Guidelines for Stress Testing by Banks

The Basel III framework (BCBS 2017a) requires banks to maintain rigorous stress testing programs. For instance, EU banks that apply the internal ratings-based approach conduct regular stress tests to determine the regulatory capital for market risk under Pillar 1 of the Capital Requirements Directive, which applies the Basel framework to the European context. Internal ratings-based banks are further asked to run credit risk stress tests in order to examine the robustness of their internal approach. However, banks are not only exposed to the three risks covered under Pillar 1; they are also exposed, for example, to securitization, concentration, liquidity, business, and residual credit risks. The Supervisory Review and Evaluation Process, which is the European version of the capital review under Pillar II, requires banks to take a forward-looking and more comprehensive view on risk, which in turn should determine both capital and strategic planning. Supervisors review and evaluate the banks’ internal stress testing framework according to the principle of proportionality and make regular and comprehensive assessments of the processes, strategies, and systems that banks integrate into their Internal Capital Adequacy Assessment Process. Supervisors consider banks’ results to assess capital adequacy and, if necessary, require additional capital and liquidity buffers.

The Basel Committee on Banking Supervision and country authorities have issued guidance on the implementation of stress tests by banks. During the global financial crisis, the Basel Committee developed *Principles for Sound Stress Testing Practices and Supervision* (BCBS 2009) after reviewing supervisory authorities’ implementation of stress tests. These principles were mirrored in guidelines developed by supervisory authorities (CEBS 2010). They state that stress scenarios should (1) reflect bank-specific risks, (2) take into account system-wide interactions, and (3) be flexible enough to adapt to changes in portfolio composition, the emergence of new risks, and specific risks related to businesses, entities, and products.38 These principles also anchor stress testing in the corporate governance and the risk-management culture of a bank. Senior management is required to (1) identify and communicate the level of risk appetite according to the bank’s business model, (2) identify relevant and plausible stress scenarios that are tested on a firm-wide level, and (3) ensure that results feed into the bank’s decision-making process, including strategic business planning and capital and liquidity planning.

However, the BCBS (2012) found that these guidelines were not followed systematically. Risk managers were not able to communicate the purpose, results, and implications of their assessments to the risk owners within the banks. In many cases, stress testing remained a very technical exercise, which was completed in a rather isolated and mechanical manner without being sufficiently integrated with business lines. Consequently, the results were often interpreted as unrealistic, not credible, or just too technical. In many cases, these tests were performed separately for different business units, which prevented a comprehensive, firm-wide perspective across different units and risks. Consequently, risks were largely underestimated. Moreover, precision stress scenarios were most often based on data that did not cover heavy downturn scenarios and the simultaneous realization of several risks. This, in turn, caused a systematic underestimation of the scope and relevance of various shocks during the global financial crisis. The Basel Committee further argued that banks failed to integrate guidelines on reputational risk and risks arising from off-balance-sheet vehicles, as well as the risks arising from highly leveraged counterparties and deficiencies in risk-mitigating techniques. In 2017, the Basel Committee revised the 2009 stress testing principles, which are now stated at a higher level to enhance greater acceptance and applicability while preserving necessary flexibility to accommodate new developments (BCBS 2018).39

---

38 Also, scenarios should feature a wide range of alternatives, from optimistic forecasts (“baseline scenario”) to tail events challenging the bank’s business model (that is, “reverse stress tests”).

39 The Basel Committee published these principles following public consultation (BCBS 2018).
Appendix 2.2. Stress Testing of Nonbank Financial Institutions: Insurance Companies and Financial Market Infrastructures

INSURANCE COMPANIES

Stress testing of the insurance sector is gaining growing acceptance as a risk-management tool, especially as developments in macroprudential surveillance have warranted greater focus on identifying systemic risk affecting the insurance sector. Insurers increasingly use it to assess and manage their risks. Insurance supervisors use it to assess the risks facing specific insurers and to identify possible vulnerabilities of the sector as a whole. Even though traditional insurance activities have not contributed to systemic risk during the financial crisis, there are some vulnerabilities from nontraditional, noninsurance activities (IAIS 2013), which could change the transmission channels of risks affecting the solvency and liquidity conditions of insurance companies under stress. If an insurer is significantly involved in activities that are closely linked to the broader financial sector—such as providing protection against credit exposures—then its failure could have systemic implications. Similarly, the failure of a large insurance company for which there is no quick substitution in the market, would also make a systemic case (Geneva Association 2010).

However, insurance stress testing has become an important element of stress testing within Financial Sector Assessment Programs (FSAPs) following the global financial crisis. Prior to that, it had typically played a secondary role relative to the analysis of the banking sector risks. Before the crisis, only 13 FSAPs (out of a total of almost 170) contained insurance stress tests (of which 11 were completed in advanced economies). This might be explained not only by the fact that insurers are considered less systemically relevant in many jurisdictions but also by the unique conceptual challenges that emerge from the different balance sheet structure of insurance companies and lack of a global solvency and valuation standards, which has resulted in a greater reliance on national supervisory frameworks for more resource-intensive stress testing in bottom-up approaches (Jobst, Sugimoto, and Broszeit 2013).

In FSAP insurance stress tests, efforts have been made to align the economic shocks to those used for the banking stress tests. However, insurers are also vulnerable to other types of risk that exceed the traditional parameters of bank stress testing and are difficult to capture in general macroeconomic scenarios. For example, natural events, such as floods, earthquakes, and windstorms, can be important to non-life insurers; and life insurers might be severely affected by a pandemic or by long-term improvements in mortality. The effects of stresses on the insurance sector are particularly difficult to test on a top-down basis because of the contract-level linkage between assets and liabilities for many life insurance products and the effects of insurer-specific reinsurance programs on the financial condition of non-life insurers. In addition, many insurance supervisors—even in developed markets—do not have the detailed data and models needed to perform such tests. It might be argued that top-down insurance stress testing at a meaningful level of granularity is impossible in most jurisdictions, at least until supervisory data and modeling capabilities have evolved, and that efforts should instead be made to improve the quality of bottom-up stress testing.

STRESS TESTING OF FINANCIAL MARKET INFRASTRUCTURES (FMI)40

Stress testing has also been applied to assessing the resilience of financial market infrastructures. In contrast to what is done for banks, it is not their balance sheet that is tested, but their proper functioning in case a risk materializes. This risk may be of an operational, credit, or liquidity nature. What matters is the immediate reaction of the FMI, the way it will finish the day and be able to operate the next few days following the shock. Both central banks, as FMI operators and overseers, and FMI's

40 FMIs refer to payment systems, central securities depositories, securities settlement systems, and central counterparties.

©International Monetary Fund. Not for Redistribution
themselves conduct stress tests. In FSAPs, the IMF staff helps specify the stress testing requirements embedded in the Committee on Payment and Settlement Systems (CPSS)/International Organization of Securities Commissions (IOSCO) standards and determines whether FMIs satisfy those requirements.

**Stress Testing by Central Banks**

Central banks play a crucial role in safeguarding the safety and soundness of payment systems. They also seek to ensure payment systems operate in a practical and efficient manner for users and for the economy as a whole. In most cases, central banks operate the main national large-value payment system, which is the backbone of the entire financial market infrastructure. In addition, central banks are often in charge of overseeing core payment systems and other FMIs, such as central counterparties (CCPs) and securities settlement systems, such as central securities depositories (CSDs).

As operators and supervisors of the main payment system, some central banks use simulation tools to analyze the underlying payment flows and participant behavior in different scenarios—in addition to a standard-based qualitative approach. Operational disruptions of the FMI itself (or a major participant) are often tested as well as the financial default of major participants. Incidents are simulated in order to identify recovery times, critical participants, contingency measures, and stop-sending limitations under different parameters, such as the concentration level, availability of liquidity, back-up procedures, reactions of nondefaulted banks, and structure of the money market. They can also help quantify the impact on liquidity by simulating the suspension of a participant’s outgoing payments, which results in liquidity accumulations for the failing participant and liquidity shortages for other participants and thereby disrupts settlements. Financial defaults of major participants are also tested to check whether the payment system will be able to handle them properly and to assess the impact on remaining participants. Findings from simulation exercises could give rise to operational, organizational, or financial changes, such as the implementation of new and/or more robust risk-mitigation facilities.

**Stress Testing by FMIs**

In the wake of the financial crisis, many FMIs have developed their own stress testing approaches to manage liquidity and credit risks (BIS/IOSCO 2012).

**Payment and Settlement Systems**

Aside from central banks, several large private settlement systems have developed stress testing processes. The Clearing House Interbank Payment System, the US private sector large-value system, has instituted a program to ensure that participants understand the consequences of a failure of one or more banks to honor their closing positions and to encourage participants to develop liquidity contingency plans. The Continuous Link Settlements Bank, which operates payment-versus-payment settlement services in 17 currencies, covering more than 60 percent of foreign exchange transactions worldwide, conducts a range of stress testing and back-testing scenarios to review the adequacy of its risk-management procedures. The simulations include, among other things, the adequacy of haircuts and the failure of the settlement member with the single largest funding obligation in a single currency. Since the global financial crisis, the Continuous Link Settlements Bank has been working on more extreme scenarios, such as the failure of all settlement members of a given currency to complete their funding and the simultaneous failure of all liquidity providers in the same currency. Some securities settlement systems, in particular international CSDs, which also offer securities loans and credit lines to participants, face potentially large credit and liquidity risks. Most of them conduct regular stress tests of their financial resources.

**Central Clearing**

Among FMIs, CCPs exhibit the highest concentration of liquidity and credit risks. Their core service is to become principal to every transaction that they clear, which implies that market participants no longer have credit exposures to their trading operations.
counterparties, but only to the CCP. Therefore, CCPs concentrate credit risk and would face large liquidity needs if a participant defaults because they need to fulfill the settlement obligations of the defaulting participant, potential losses when the cleared position or related collateral are liquidated, and the cash flows relating to possible hedge transactions.

Stress testing is key in managing the credit and liquidity risks of CCPs. Stress tests consider extreme but plausible market conditions and are typically framed in terms of the number of participant defaults a CCP can withstand. Current CPSS/IOSCO standards (BIS/IOSCO 2012) prescribe that CCPs should be able to withstand the default of the participant with the largest exposure, but some CCPs have chosen to be more stringent and test for the default of several major participants. CCPs’ models that calculate margin requirements, default fund contributions, collateral requirements, and other risk control mechanisms are expected to be subjected to rigorous and frequent stress tests that reflect their product mix and other risk-management choices. Key elements of stress tests are the assumed market conditions and default scenarios and the test frequency. A CCP should assume extreme market conditions (that is, price changes significantly larger than the prevailing levels of volatility), and evaluate the potential losses in individual participants’ positions. Other stress tests may consider the distribution of positions between the defaulting participants and their customers in evaluating potential losses. These should consider the resources of the potential defaulters that are available to a CCP (margins, clearing fund contributions, or other assets), as well as the CCP’s own resources, to provide perspective on the potential size of the losses and liquidity gaps of the CCP.

Since the financial crisis, the attention of regulators and supervisors has been drawn to CCPs’ stress testing, in particular when clearing over-the-counter (OTC) derivatives. Central clearing of OTC derivatives presents more challenges than clearing listed or cash-market products because of their complex risk characteristics. Following the G20 commitment to strengthen regulation of the OTC derivatives markets, improved rules took effect at the end of 2012. In some jurisdictions, this has included guidance on stress testing. For example, the European Commission adopted technical standards that specify the types of tests to be undertaken for different classes of financial instruments and portfolios, the involvement of clearing members or other parties in the tests, the frequency of tests, and the time horizon. In addition, CCPs’ supervision and oversight have generally been strengthened since the crisis, for example by systematically analyzing CCPs’ internal risk-management models, including stress testing parameters.

**Stress Testing by the IMF**

The IMF often examines FMI stress testing arrangements in the context of FSAPs as well as Stability Modules (including reviews of standards and codes and reviews under the Offshore Financial Center Assessment Program). So far, the IMF has not conducted stress tests on FMIs. However, the IMF specifies the stress testing requirements embedded in the CPSS/IOSCO standards (BIS/IOSCO 2012) and checks whether those requirements are met by the respective FMIs.
Appendix 2.3.  
Prominent Crisis-Management Stress Tests

Since the global financial crisis, stress tests have increasingly been used as instruments for crisis management. Their results feed into political decision-making processes, and in many cases determine the allocation of public support measures.

The most prominent examples for crisis management stress testing exercises were those conducted in the United States and Europe: (1) the US Federal Reserve’s Supervisory Capital Assessment Program (SCAP), which was completed in May 2009 (Board of Governors of the US Federal Reserve System 2009); (2) the EU system-wide stress tests performed by the Committee of European Banking Supervisors in 2010 (CEBS 2010) and the European Banking Authority (EBA) in 2011 (EBA 2011a, 2011b); and (3) the EBA’s 2011–2012 Capital Exercise (EBA 2012). In contrast, the US Federal Reserve’s Comprehensive Capital Analysis and Review (Board of Governors of the US Federal Reserve System 2009), which evolved from the SCAP in 2011, the EU system-wide stress tests since 2014, which evolved into a biannual exercise (EBA 2014, 2016, 2017), and routine tests required under the Basel frameworks constitute typical supervisory stress tests.

KEY FEATURES OF CRISIS-MANAGEMENT STRESS TESTS

Crisis-management stress testing exercises incorporate characteristics of both microprudential and macroprudential (or surveillance) stress tests. Like in supervisory stress tests, the individual institutions’ resilience to shocks is assessed on a bank-by-bank basis. The tests incorporate common macroeconomic and market-level shocks, and evaluate the banks’ performance relative to either existing (or future) regulatory requirements or alternative thresholds and definitions for capital ratios specifically designed for the particular exercise.46 Potential follow-up actions for banks that do not pass the test include private recapitalization and/or the acceptance of governmental support, mandatory restructuring, and changes in business lines/models.47

Crisis-management exercises are conducted for specific (crisis management) purposes and do not take place on a regular basis. The shock scenarios usually involve higher probability shocks that are more likely to materialize than the extreme-but-plausible tail risks typically tested in surveillance and supervisory tests. The exercise is comprehensive, covering both the banking and trading books, and usually consider all off-balance sheet positions. Portfolios are broken down by geographic regions, industry sectors, and asset classes, which in turn implies that potential losses can differ across banks not only because of their different portfolio composition, but also reflecting differences in asset quality.

Crisis-management stress tests are designed as traditional balance-sheet-based solvency tests. The SCAP, EU system-wide stress testing exercises, and the EBA Capital Exercise focused on solvency risk in individual banks; liquidity risk was either not tested (SCAP) or assessed in a separate exercise (EBA 2011b).48 Spillover effects and default dependencies across institutions were, to some extent, indirectly considered through the design and structure of adverse scenarios, which typically take into account several transmission channels. These exercises applied a form of static or constant balance sheet assumption to allow banks to shrink balance sheets or adapt business models during the forecasting period. Such approaches eliminate the scope for strategies to boost capital ratios by reducing risk-weighted assets. At the same time, this comes at the price of disregarding behavioral response functions.

46 For instance, in the 2011 EU system-wide stress test, the EBA focused on banks’ core capital and considered a specific Core Tier 1 capital definition that was based on the Capital Requirements Directive II definition of Tier 1 capital net of deductions for participation in financial institutions, excluding hybrid instruments but including existing preference shares and existing governmental support measures. This definition should not be confused with the Common Equity Tier 1 under Basel II (and implemented in the EU under the Capital Requirements Directive IV). In the 2009 SCAP, the US Federal Reserve applied a modified definition of Tier 1 capital, which excluded preferred stock, minority interest in subsidiaries, and less qualifying trust preferred securities.

47 At the same time, crisis-management stress tests do not limit their focus to assessing the health of individual banks but are equally concerned about the stability of the system as a whole. In order to evaluate the resilience of the banking system and the potential need for macroprudential or system-wide measures, aggregate indicators are taken into consideration.

48 The 2011 EU system-wide stress test, however, included a shock to banks’ funding costs (that is, funding liquidity) within the solvency stress testing framework. A traditional liquidity stress test assessing banks’ liquidity profiles was performed separately. The results of this assessment were not published.
Crisis-management stress tests combine elements of the bottom-up and top-down approaches.\(^{49}\) Based on a detailed methodology designed by the supervisor, banks examine the impact of one or more common macroeconomic scenarios on their portfolios. The results are checked for completeness, consistency, and plausibility by the supervisor. But the tests also include centralized components.\(^{50}\) The idea is to maximize the benefit of bottom-up approaches and, at the same time, to cross-check or validate the banks’ results through top-down elements.\(^{58}\)

### COMPARING THE CRISIS-MANAGEMENT STRESS TESTING EXERCISES

The three most prominent crisis-management exercises had different purposes. The goal of the SCAP was to “[e]nsure adequate system capital to promote lending and restore investor confidence” (Board of Governors of the US Federal Reserve System 2009). The exercise was designed to “estimate losses, revenues, and reserve needs” for the large bank holding companies, and evaluated the size of governmental capital injections contingent on the banks’ performance in the stress test (Board of Governors of the US Federal Reserve System 2009). The EU system-wide stress testing exercises aimed at examining the resilience of the banking system, identifying vulnerabilities, and informing policymakers about the current capacity of banks to absorb shocks and the banking system’s dependence on public support measures (EBA 2011a, 2011b). EBA’s Capital Exercise was specifically designed to “create an exceptional and temporary capital buffer to address current market concerns over sovereign risk and other residual credit risk related to the current difficult market environment” (EBA 2012) but also informed a more objective capital assessment of euro area banks as national authorities were preparing for the integrated supervision of systemically important institutions (via the Single Supervisory Mechanism).

The design and severity of macroeconomic stress scenarios depended on the objectives and the timing of each exercise. The adverse scenarios in both the SCAP and the EU system-wide stress tests involved a simultaneous realization of several risk factors. EBA’s Capital Exercise, in contrast, was based on a EU system-wide baseline scenario provided by the European Commission, with a focus on sovereign risks. Banks were required to conservatively assess the value of direct exposures to European Economic Area sovereigns held in the banking book (that is, loans and nontraded assets) at market prices as of September 2011. The adverse scenario of the 2011 (2010) EU system-wide stress test translated into a cumulative GDP shock of \(-4.1\) (\(-3.1\)) percentage points, compared with a \(-2.8\) percentage point shock in the SCAP in 2009. Since the SCAP took place at the peak of the crisis, the assumed macroeconomic shock implied a deeper contraction compared with the European tests (that is, a cumulative two-year contraction in GDP of \(-2.7\) percent under the SCAP, compared to a cumulative contraction of \(-0.2\) [\(-0.4\)] percent in the 2011 [2010] EBA exercise).

Another key difference between these exercises was the organizational complexity of the EU system-wide tests. The EU exercises involved 21 countries, 24 national supervisors and authorities, around a dozen European institutions, and 91 participating banks (Appendix Figure 2.3.1). They had to deal with 19 different languages and seven currencies. None of these complexities applied to the SCAP.

The experience with these exercises was mixed. While both the SCAP and the EU system-wide stress testing shared (almost) identical goals, their impact was rather different:\(^{52}\)

- **The SCAP demonstrated convincingly that stress testing can be a powerful instrument for crisis management.** The setup of the SCAP was credible, and backstop measures (including, crucially, government support) were clearly communicated and in place ahead of time. The US Federal Reserve argued that the test was “an important turning point in the financial crisis” and that confidence improved as banks raised capital, mainly in the private markets. The results allowed markets to differentiate between banks based on their ability to withstand the shocks; after the publication of the results, the correlation among major banks’ stock prices fell by more than 10 percent. Capital injections for banks that did not pass the tests were mandatory. Banks unable to raise capital in the private markets within a preset time period were asked to accept governmental capital support. Consequently, both goals of the SCAP were achieved—confidence in the US banking sector was restored and capital buffers were replenished to withstand further shocks to the system.

\(^{49}\) As the supervisory methodology constrains banks in assessing the impact of the scenarios, the EBA has coined the term “constrained bottom-up” tests.

\(^{50}\) The 2009 SCAP basically followed a top-down approach and, at the same time, incorporated several decentralized components of bottom-up frameworks, like the integration of banks’ own projections of losses, operating profits, and loan loss provisions under the given scenario. According to the detailed methodology designed by the supervisor, the CEBS/eba tests asked banks to examine the impact of two scenarios on their portfolios. The results were checked and challenged internally by CEBS (2010) and within a multilateral review process, which was also flanked by top-down calculations by the European Systemic Risk Board in the 2011 exercise.

\(^{51}\) This is most crucial for preimpairment income, which serves as a first, and substantial, cushion against losses. Since banks have a better understanding of their business conditions, they are in a better position when discussing their bottom-up forecasts with the supervisor. While the US Federal Reserve applied top-down stress tests to challenge the banks’ submissions, the EBA chose to cap net interest income at 2010 levels in its 2011 EU system-wide test.

\(^{52}\) While the SCAP exercise morphed into the annual Comprehensive Capital Analysis and Review (with the Dodd-Frank Annual Stress Testing as stress testing element) (Board of Governors of the US Federal Reserve System 2009, 2012), European authorities also have established system-wide stress testing as a regular macroprudential exercise [at two-year intervals] (EBA 2011a, 2011b, 2014, 2016, 2017).
The 2011 EU system-wide stress test was in many aspects a successful exercise but did not fully achieve its goals. The exercise remedied several shortcomings of previous stress tests in 2010, and was praised for the risk coverage, tighter definition of capital components, quality assurance process, and detailed disclosure of exposures and results. The inclusion of stress on funding costs, sovereign exposures, and securitization positions expanded the test’s risk coverage substantially. While the tougher capital definition was generally welcomed, it was also criticized for being inconsistent with the Basel III definition of Common Equity Tier 1. The comprehensive publication of the methodology and results made the test an important exercise in disclosure, allowing analysts to replicate the findings and draw their own conclusions. However, the whole exercise also contained several shortcomings. First, the scope and severity of the shock to sovereign exposures was deemed insufficient. While sovereign risk was, in principle, covered, the stress applied on sovereign exposures did not reflect the dislocations in European sovereign debt markets. Second, the adverse scenario was designed as a common scenario for all EU member states, but this meant that it did not adequately reflect the specific risk profiles of some national banking sectors, which was underscored by the failure of the Belgian bank Dexia in 2011 (which passed the stress test but become insolvent only a few weeks after the conclusion of the exercise). And third, backstops were seen by the markets as ambiguous, along with uncertainty over actions for banks ending up slightly above the hurdle rate, which considerably undermined attempts to restore investor confidence.

53 “The banking component can no longer be separated from sovereign and institutional developments. This is why Friday’s publication of stress tests results, while useful, is unlikely to be the game-changer it could have been two years ago (Véron 2011).”
This page intentionally left blank
This appendix presents some obvious conceptual challenges for stress testing and various analytical approaches that might be used to address them, including work being done at the IMF.

INTEGRATING SOVEREIGN RISK AND BANKING-SOVEREIGN FEEDBACKS IN STRESS TESTS

Conceptual and technical difficulties related to how to account for the impact of sovereign-bank links became evident in the EU system-wide stress testing. However, it is difficult to design stress testing models that account for banks’ impact on sovereign risk and the feedback from sovereigns to banks through several transmission channels (Appendix Figure 2.4.1). The mark-to-market fall in the value of sovereign bonds held by banks reduces bank asset values, and distressed banks increase explicit or implicit contingent liabilities of the public sector. Higher contingent liabilities might lead to rising sovereign spreads, which, in turn, can raise bank funding costs. If the sovereign is distressed enough, the value of official support (guarantees) to the bank is eroded. This can have knock-on effects on foreign banks and other sovereigns, as shown in Appendix Figure 2.4.1.

To include sovereign risk in stress testing, the key interlinked risk exposures between the government and financial sector should be analyzed in a comprehensive framework. A stylized framework starts with the economic, that is, risk-adjusted, balance sheets of the financial sector, which can be treated as a portfolio and linked to the government’s balance sheet. In this specification, distressed financial institutions can lead to large government contingent liabilities, which in turn reduce government assets and lead to higher risk of default on sovereign debt. Dynamic macro-financial linkage models used for bank stress tests can also be linked to sovereign risk models, together with the feedback of banking risk to sovereigns via contingent liabilities and sovereign spreads affecting bank funding costs.

Sources: IMF 2010b; Jobst and Gray 2013; and authors.

Appendix Figure 2.4.1. Spillovers from Sovereigns to Banks and Banks to Sovereigns
ENHANCING ANALYSIS OF MACROECONOMIC AND BANKING SECTOR FEEDBACK AND CONTAGION BETWEEN FINANCIAL INSTITUTIONS

A typical stress testing exercise uses macroeconomic scenarios to assess the impact on the banking/financial sector risk and capital adequacy without feedbacks to the macroeconomy. However, in many cases, distress in the banking/financial sector precedes the contraction of credit, leading to lower GDP growth. This feedback effect tends to be absent in stress tests. Models that contain such feedback channels would be useful, for example, a shock to the financial sector might be used to estimate the reduction in GDP growth and other factors, which would in turn have negative impacts on corporate and household borrowers, which would in turn increase credit risk on banks’ balance sheets. Dynamic factor models can help enhance the modeling of the feedback between the banks and the macroeconomy.

The global financial crisis demonstrated the potential for strong contagion among financial institutions. Traditional stress tests that use macroeconomic factor models to link to banking risk have a certain built-in correlation between banks, which comes from their common correlations to the macroeconomic factors. However, this does not capture correlations, dependencies, and feedback between the institutions. Contagion effects can be modeled with networks, joint default probabilities or expected losses, and a variety of other models that have been developed. Enhanced stress testing can include some of the features of these systemic risk models to improve the analysis of interdependence and joint risk.

TECHNIQUES TO IMPROVE MODELING OF NONLINEARITIES IN CHANGES IN BANK ASSETS, CAPITAL, CREDIT RISK, AND FUNDING COST

The risk-adjusted balance sheet can be helpful in illustrating the trade-offs between the nonlinear impact of changes in bank’s assets and (risky) debt on credit/funding spreads and capital. The fundamental conceptual framework of the risk-adjusted balance sheet comes from the contingent claims approach (Merton 1973) and risk-neutral valuation (Cox and Ross 1976). A bank’s liabilities (equity and [risky] debt) are claims on underlying bank assets that are uncertain over this time horizon, and the degree of uncertainty (that is, volatility) affects the risk premiums and values of equity and debt liabilities. There are different ways to construct the risk-adjusted balance sheet. One can use the estimated loan portfolio loss distribution and other components of the bank’s balance sheet. Using risk-neutral valuation, the probability distribution of the bank’s risky loans and distribution of assets (over a specific time horizon) can be estimated. This asset distribution is then combined with the promised payments on debt and deposits to construct a risk-adjusted (contingent-claims-analysis-type) balance sheet. This technique does not rely on market prices. A second method estimates the market-implied asset and asset volatility of a bank from the observed market value and volatility of the bank’s equity and the book value of its debt and deposits. Comparing the results of the two methods can provide insights into the dynamics and differences between the “fundamental” loan portfolio loss approach and the market-implied view (which will vary between calm and stress periods).

This modeling approach can also be used to estimate the increase in funding cost dependent on the risk-free interest rate, the bank’s credit spread, market risk appetite, and the impact of the government’s (implicit and explicit) guarantees. In addition, the risk-adjusted balance sheet of a bank already incorporates the dynamic changes between assets and market capital. Changes in assets lead to changes in market equity and changes in expected losses to bank creditors or guarantors. The magnitude of these losses depends on the level of distress of the bank. Thus, the change in bank market capitalization due to the decline in implied asset values is analogous to an aggregate risk-weighted asset adjustment factor in the Basel III framework, and can be interpreted as a “quasi-risk-weighted assets” dynamic adjustment factor.

54 See Kranar and Matheson 2017.
55 More work is needed on capturing the nonlinearities between changes in bank assets and capital. In balance-sheet-based stress testing models, fixed correlations between exposures and static risk weights can lead to underestimation of the capital shortfall under stress. This can be improved by using time-varying correlation between exposures when estimating portfolio loss distributions.
REFERENCES


©International Monetary Fund. Not for Redistribution


