ARTICLE IN PRESS

International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Stress testing banks

Til Schuermann *

Oliver Wyman, United States
Wharton Financial Institutions Center, United States

ARTICLE INFO

Keywords:
Capital requirements
Leverage
Systemic risk

ABSTRACT

How much capital and liquidity does a bank need to support its risk taking activities? During the recent (and still ongoing) financial crisis, answers to this question using standard approaches, e.g., regulatory capital ratios, were no longer credible, and thus broad-based supervisory stress testing became the new tool. Bank balance sheets are notoriously opaque and susceptible to asset substitution (easy swapping of high risk for low risk assets), so stress tests, tailored to the situation at hand, can provide clarity by openly disclosing details of the results and approaches taken, allowing trust to be regained. With that trust re-established, the cost-benefit of stress testing disclosures may tip away from bank-specific towards more aggregated information. This paper lays out a framework for the stress testing of banks: why it is useful and why it has become such a popular tool for the regulatory community in the course of the recent financial crisis; how stress testing is done (design and execution); and finally, with stress testing results in hand, how one should handle their disclosure, and whether it should be different in crisis vs. "normal" times.

1. Introduction

There are three kinds of capital and liquidity: (1) the capital/liquidity you have; (2) the capital/liquidity you need (to support your business activities); and (3) the capital/liquidity the regulators *think* that you need.[1] Stress testing, regulatory capital/liquidity and bank-internal (so-called "economic capital/liquidity") models all seek to do the same thing: to assess the amount of capital and liquidity which is needed to support the business activities of the financial institution. Capital adequacy addresses the right side of the balance sheet (net worth), and liquidity the left side (share of assets that are "liquid", however defined). If all goes well, both the economic and regulatory capital/liquidity are less than the required regulatory minimum, and their difference (between economic and regulatory) is small, that is, regulatory models do not deviate substantially from the results of internal models.

Prior to their failure or near-failure, financial institutions such as Bear Stearns, Washington Mutual, Fannie Mae, Freddie Mac, Lehman and Wachovia were adequately or even well capitalized, at least according to the regulatory capital rules disclosed in their public filings.[2] This set of institutions spans a broad range of regulatory capital regimes and regulators: the SEC and Basel 2 capital rules (Bear Stearns, Lehman), the OCC and the Federal Reserve and Basel 1 (Wachovia), the OTS (WaMu), and OFHEO (Fannie and Freddie)—the last actually based on a narrow stress scenario. All firms had a broad exposure to residential real estate assets, in the form of either whole loans (mortgages) or securities (MBS), or both, and all had internal risk models which may or may not have deviated materially from the regulatory models (we do not know this, as it is/was firm proprietary information).[3] Yet the answer to

* Correspondence to: 1166 Avenue of the Americas, 29th Floor New York, NY 10036, United States.
E-mail address: til.schuermann@oliverwyman.com.

1 This pithy summary I owe to Peter Nakada.

2 Kuritzkes and Scott (2009) make the case for a more market-oriented assessment of capital adequacy.

3 Lester, Reynolds, Schuermann, and Walsh (2012) report that, out of 16 banks (US and non-US) that publicly disclosed their economic

0169-2070/$ – see front matter © 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

the question of what is the capital you need vs. the capital you have came out wrong in each case. Of course, neither firm-internal (economic) nor regulatory capital and liquidity models can guarantee failure prevention; indeed, that is not their purpose, as every firm accepts some probability of failure, sized by its risk appetite. Nevertheless, the cascading of defaults, and the resulting deep skepticism of the market's stated capital adequacy, forced regulators to turn to a new tool for assessing the capital adequacy of banks in a credible way. That tool turned out to be stress testing.[4]

This paper lays out a framework for the stress testing of banks: why it is useful and why it has become such a popular tool for the regulatory community in the course of the recent financial crisis; how stress testing is done (design and execution); and finally, with stress testing results in hand, how one should handle their disclosure, and whether it should be different in crisis vs. "normal" times. The framework is equally applicable to capital and liquidity adequacy, but for the sake of simplicity, the bulk of the discussion will focus on capital.

A successful macro-prudential stress testing program, particularly in a crisis, has at least two components: first, a credible assessment of the capital strength of the tested institutions, to size the capital "hole" that needs to be filled, and second, a credible way of filling that hole. The US bank stress test in 2009, the Supervisory Capital Assessment Program or SCAP, may serve as a useful example. The US entered 2009 with an enormous uncertainty about the health of its banking system. In the absence of a more concrete and credible understanding of the problems with bank balance sheets, investors were reluctant to commit capital, especially given the looming threat of possible government dilution. With a credible assessment of losses under a sufficiently stressful macroeconomic scenario, the supervisors hoped to draw a line in the sand for the markets: fill this hole, and you won't risk being diluted later because the scenario wasn't tough enough. Moreover, if some institutions could not convince investors to fill the hole, a US government program, namely the Treasury's Capital Assistance Program (CAP), stood ready to supply the required capital. Importantly, the US Treasury was a sufficiently credible debt issuer that the CAP promise was itself credible.[5] All banks with assets greater than $100 bn (YE 2008) were included, accounting for two-thirds of the total assets and about half of the total loans in the US banking system. In the end, ten of the 19 SCAP banks were required to raise a total of $75 bn in capital within six months, and indeed raised $77 bn of Tier 1 common equity in that period.[6] None needed to draw on CAP funds.

The European experience in 2010 and 2011 stands in stark contrast to the 2009 SCAP. Against the background of a looming sovereign debt crisis in the peripheral eurozone countries, the Committee of European Bank Supervisors (CEBS) conducted a stress test of 91 European banks in 2010, covering about two-thirds of the total European bank assets and at least half of that in any given participating country. The stress test included imposing haircuts on the market value of sovereign bonds held in the trading book; however, the bulk of the sovereign exposure was (and is) in the banking book. Of the 91 banks, only seven were required to raise a total of €3.5 bn (<$5 bn at the time) in capital. The level of disclosure provided was rather less than in the SCAP. For instance, loss rates by firm were only made available for two sub-categories: overall retail and overall corporate.[7] By contrast, the SCAP results released loss rates by major asset class such as first-lien mortgages, credit cards, commercial real estate, and so on. Markets reacted benignly nonetheless—until a few months later, when Ireland requested financial assistance from the EU and the IMF. Subsequent stress tests of just the Irish banks, conducted largely by outside independent advisors (BlackRock) revealed a total capital need of €24 bn; all of these banks had previously passed the CEBS stress test. Moreover, to help close the credibility gap, the extent and degree of disclosure was far greater than in any of the stress testing exercises to date.[8] The markets reacted favorably, with both bank and Irish sovereign credit spreads tightening. The stakes for the 2011 European stress test, now conducted by the successor to the CEBS – the European Banking Authority (EBA) – had risen substantially.

At first glance, the results of the 2011 EBA stress test of 90 banks in 21 countries were mild, similar to the previous year's.[9] Eight banks were required to raise a total of only €2.5 bn. However, the degree of disclosure was much more extensive, approaching the high bar set by the Central Bank of Ireland in March 2011, including information on exposure by asset class by geography. Importantly, all bank level results were available to download in spreadsheet form, to enable market analysts to easily impose their own loss rate assumptions. In this way, the "official" results were no longer so final: analysts could (and did) easily apply their own sovereign haircuts on all exposures, and thus test the solvency of any of the 90 institutions themselves.

In an uncomfortable parallel to the Irish experience in 2010, the 2011 EBA stress test did nothing to alleviate concerns about the Spanish banking system. Five of the 25 Spanish banks in the EBA stress test did not pass, though once provisions and mandatory bond conversions (to equity) were taken into account, the required additional capital raise was €0. By the spring of 2012, Spain was engaged in or had announced several additional stress tests. First was the IMF's Financial Sector Assessment Program (FSAP), conducted jointly with the Banco de

capital before the crisis, four actually experienced losses exceeding those requirements, all of which were calibrated to at least the 99.9% level (implying an acceptable annual default probability of no more than 10bp).

4 Flannery (2012) argues that stress tests should be evaluated on a fair value (rather than book capital) basis.

5 Note that the act of a sovereign recapitalizing its banks involves that sovereign issuing debt and then investing ("downstreaming") it as equity in the bank(s).

6 http://www.federalreserve.gov/bankinforeg/scap.htm.

7 http://www.eba.europa.eu/EU-wide-stress-testing/2010/2010-EU-wide-stress-test-results.aspx.

8 http://www.centralbank.ie/regulation/industry-sectors/credit-institutions/Documents/The%20Financial%20Measures%20Programme%20Report.pdf.

9 http://www.eba.europa.eu/EU-wide-stress-testing/2011/2011-EU-wide-stress-test-results.aspx.

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

Table 1
Summary of disclosures across stress test exercises.

	Base and stress scenario	Bank level results	Asset/product level loss rates	Exposure detail (asset class, maturity, geography)	Bank vs. supervisory/3rd party estimates
SCAP March 2009	Stress	✓	✓	–	–
CEBS July 2010	Both	✓	Retail, all corporate only	–	–
CCAR March 2011	–	–	–	–	–
Ireland March 2011	Both	✓	✓	Sovereign only	✓
EBA July 2011	Both	✓	Retail, corporate, CRE	High	–
CCAR March 2012	Stress	✓	✓	–	–
Spain (IMF) June 8, 2012	Both	–	–	Asset class (aggregate)	–
Spain (top-down) June 21, 2012	Both	–	✓	Asset class (aggregate)	–
Spain (bottom-up) Sept. 28, 2012	Both	✓	✓	Asset class (aggregate)	–

España. The results of this were released on June 8, 2012,[10] with 11 of the 29 banks requiring a total of €17.7 bn capital using a post-stress hurdle similar to that of the SCAP (4% core Tier 1 capital), or 17 banks requiring a total of €37.1 bn using the higher hurdle of 7% core Tier 1 capital.[11] Second was a short (4-week) top-down exercise conducted by two outside advisers (working in parallel to provide, ostensibly, two further independent assessments), and those results were released on June 21, 2012. No firm-specific results were provided, only an overall capital need. The first estimate, provided by Roland Berger, was €51.8 bn, while Oliver Wyman provided a range of €51–62 bn.[12] A more detailed and intensive bottom-up analysis by Oliver Wyman followed, with results released on September 28, 2012, showing that 7 of 14 the banking groups needed a total of €57.3 bn using the post-stress core Tier 1 threshold of 6%; merger activity had resulted in a significant reduction in independent banking entities.[13]

A summary of the major macro-prudential stress tests to date is provided in Table 3, and a summary of their disclosures is given in Table 1.

[10] http://www.imf.org/external/pubs/ft/scr/2012/cr12137.pdf.

[11] Most European exercises have tested to a post-stress hurdle of 6% core Tier 1; see the discussion in Section 3.

[12] Roland Berger: http://www.bde.es/webbde/GAP/Secciones/SalaPrensa/InformacionInteres/ReestructuracionSectorFinanciero/Ficheros/en/informe_rolandbergere.pdf ; Oliver Wyman: http://www.bde.es/webbde/GAP/Secciones/SalaPrensa/InformacionInteres/ReestructuracionSectorFinanciero/Ficheros/en/informe_oliverwymane.pdf.

[13] http://www.bde.es/f/webbde/SSICOM/20120928/informe_ow280912e.pdf.

Table 2
Features of stress testing, pre- and post-SCAP.

Pre-SCAP	Post-SCAP
• Mostly single shock	• Broad macro scenario and market stress
• Product or business unit level	• Comprehensive, firm-wide
• Static	• Dynamic and path dependent
• Not usually tied to capital adequacy	• Explicit post-stress common equity threshold
• Losses only	• Losses, revenues and costs

The SCAP was the first of the macro-prudential stress tests of this crisis, but the changes at the micro-prudential or bank-specific level were at least equally significant, and they are summarized in Table 2. With the SCAP, stress testing at banks went from mostly single factor shocks (or a handful) to using a broad macro scenario with market-wide stresses; from product or business unit stress testing, focusing mostly on losses, to firm-wide and comprehensive testing, encompassing losses, revenues and costs; and with all of these tied to a post-stress capital ratio to ensure a going concern.

The remainder of the paper proceeds as follows. Section 2 briefly reviews the scant literature, and Section 3 provides a discussion of how to design the stress scenario, including the choice of a post-stress capital hurdle. Section 4 describes modeling approaches for the three components needed to implement stress testing: losses, net revenues (profitability), and balance sheet dynamics. Section 5 reviews the disclosure regimes across the different stress tests to date in more detail, and presents a discussion of disclosure in "normal" times, after which Section 6 provides some concluding remarks.

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

Table 3a
Summary of macroprudential stress tests to date.

	Target capital ratio[a]	# of participating banks	Participation criteria (total coverage)	Balance sheet assumptions	Total required capital raise (for # of banks)	Risk types included: market, credit, liquidity (funding), operational
SCAP March 2009	• 4% T1C • 6% T1	19	• All bank holding companies with at least $100 bn total assets • (∼2/3 of total banking assets)	Constant RWA	$75 bn (19)	M[b], C
CEBS July 2010	• 6%T1	91 (20 countries)	• Largest banks in country until at least 50% of total assets are included • (∼2/3 of total banking assets)	Constant total assets	€3.5 bn (7)	M, C
CCAR March 2011	• 5% T1C	19	• Original SCAP-19	None	—	M, C
Ireland March 2011	• 6% T1C • 10.5% T1C (in base)	4	• Largest banks not in wind-down mode	Allowed for balance sheet shrinkage	€24 bn (4)	M, C, L, O
EBA July 2011	• 5% T1C	90 (21 countries)	• Largest banks in country until at least 50% of total assets are included • (∼2/3 of total banking assets)	Constant total assets	€2.5 bn (8)	M, C, L[c], O
CCAR March 2012	• 5% T1C • 4% T1; 8% Total; 3%–4% leverage	19	• SCAP-19 • An additional 11 BHCs with assets >$50 bn	None	—[d]	M, C, O

[a] T1: Tier 1 capital ratio; T1C: Tier 1 common (or core) capital ratio.
[b] Only banks with at least $100 bn in trading assets were required to conduct the market risk stress test.
[c] Liquidity risk was not assessed directly, though funding stresses were taken into account, especially as they related to sovereign stress impacting the funding costs for financial institutions.
[d] Four of the 19 did not pass, in the sense of not having gained non-objection to their submitted capital plans.

Table 3b
Summary of macroprudential stress tests to date—Spain 2012.

	Target capital ratio[a]	# of participating banks[b]	Participation criteria (total coverage)	Balance sheet assumptions	Total required capital raise (for # of banks)	Risk types included: market, credit, liquidity (funding), operational
IMF June 8, 2012	• 7% T1C	29	• Large and medium banks and cajas, together making up ∼90% of total bank assets	Deleveraging	• €37.1 (17) under 7% T1C	C, L
Top-down June 21, 2012	• 9% T1C [base] • 6% T1C [stress]	14 entities	• Large and medium banks and cajas, together making up ∼90% of total bank assets	Deleveraging	• €16–25 [base] • €51–62 [stress]	C, L
Bottom-up Sept. 28, 2012	• 9% T1C [base] • 6% T1C [stress]	14 entities	• Large and medium banks and cajas, together making up ∼90% of total bank assets	Deleveraging	• €24.1 (5) [base] • €57.3 (7) [stress]	C, L

[a] T1: Tier 1 capital ratio; T1C: Tier 1 common (or core) capital ratio.
[b] The 14 entities are the result of mergers.

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

2. Stress testing in the literature

Stress testing has been part of the risk manager's toolkit for a long time. It is perhaps the most basic of risk-based questions to want to know the resilience of an exposure to deteriorating conditions, be it a single position or loan or a whole portfolio. Typically, the stresses take the form of sensitivities (spreads double, prices drop, volatilities rise) or scenarios (black Monday 1987, autumn of 1998, post-Lehman bankruptcy, severe recession, stagflation). These types of stresses lend themselves naturally to understanding financial risks, particularly in a data rich environment such as that found in a trading operation. Nonfinancial risks, such as operational, reputational and other business risks, are much harder to quantify and parameterize yet rely heavily on scenario analysis (earthquakes and other natural disasters, computer hacking, legal risks, and so on). While the original Basel I Accord of 1988 did not make any formal mention of stress testing, it merited its own section in the Market Risk Amendment of 1995, and thus became embedded in the regulatory codex. Indeed, evidence of stress testing capabilities is a requirement for regulatory approval of internal models.

Risk management as a technical discipline came into its own with the publication of the RiskMetrics technical document in 1994, and stress testing (of both kinds, sensitivities and scenarios) is mentioned throughout. The first edition of Jorion's standard-setting VaR book (Jorion, 1996) had a subsection devoted to the topic (which was elevated to a chapter in subsequent editions), and there must surely be earlier examples. Stress testing as a risk management discipline was found largely in the relatively data rich environment of the trading room, with the closely related treasury function of conducting interest rate scenarios and shocks.[14] The Committee on Global Financial Systems (CGFS) of the BIS conducted a survey on stress testing in 2001, and it reinforces this view.[15] In their summary of the CGFS report, Fender et al. (2001) point out that most of the scenarios involve shocks to market rates, prices or volatilities. Typical examples are equity market crashes such as October 1987, rates shocks such as 1994, credit spread widening such as during the fall of 1998, and so on. Such stress scenarios have the virtue of being unambiguously articulated and defined, and are thus transparent and easy to implement and communicate, at least on assets that have themselves natural market prices or analogs, as is mostly the case in the trading book. More typical banking assets, such as corporate loans (especially to privately held firms) and consumer loans (e.g. auto loans), are less naturally amenable to this approach.

Formal stress testing of the banking book, which is dominated by credit risk, is more recent, partly because quantitative credit risk modeling is itself a newer discipline.[16] Perhaps stimulated by the success of RiskMetrics, the late 1990s saw a spurt of activity in the development of credit portfolio models, with the two most prominent examples being CreditMetrics (Gupton, Finger, & Bhatia, 1997) and CreditRisk+ (Wilde, 1997).[17] However, stress testing did not feature in these papers. At the same time, as Koyluoglu and Hickman (1998) show quite clearly, all of these credit portfolio models share a common framework of mapping outcomes in the real economy, often represented by an abstract state vector, to the credit loss distribution, and thus, they should lend themselves naturally to stress testing. With that in mind, Bangia, Diebold, Kronimus, Schagen, and Schuermann (2002), broadly following the CreditMetrics framework, show how to use credit migration matrices to conduct macroeconomic stress tests on credit portfolios. Foglia (2008) provides a survey of the literature (at least through to late 2008) of stress testing credit risk, both for individual banks or portfolios and for banking systems. More recently, Rebonato (2010), with his suggestively titled book *Coherent stress testing* (we return to the problem of coherence below), argues for a Bayesian approach to financial stress testing, i.e., one which is able to formally include expert knowledge in the stress testing design, with an emphasis on exploring causal relationships using Bayesian networks.

With few exceptions, regulatory requirements for stress testing were thin prior to the crisis, though considerable expectations of stress testing capabilities were voiced in supervisory guidance in the US. Examples include the Joint Policy Statement on Interest Rate Risk (SR 96-13), guidance on counterparty credit risk (SR 99-03[18]), and country risk management (SR 02-05). However, banks had a significant degree of discretion with regard to the specific design and implementation of their stress tests. Brian Peters, then head of risk in bank supervision at the New York Fed, observed at an industry conference in March 2007 that no firm had a fully-developed program of integrated stress testing that captured all major financial risks on a firm-wide basis.[19] Market risk stress tests were most advanced, while corporate or enterprise-wide stress testing, whereby all businesses were subjected to a common set of stress scenarios, was in a developmental phase at best.

3. Stress testing design

Perhaps the most fundamental choice in stress testing design is the risk appetite of the authorities: how severe and how long the stress scenario should be, and what the post-stress hurdle is. To take a sailing analogy: how severe and how long is the storm, and how solid does the boat still need to be once the storm has passed? In stark contrast to standard capital regimes, the target calibration is not strict

[14] See Berkowitz (2000) and Kupiec (1998) for more extensive discussions of VaR-based stress testing.

[15] See CGFS (2001) and the summary of its principal findings by Fender, Gibson, and Mosser (2001).

[16] Of course, the credit rating agencies, having been in the business of rating corporate bonds for nearly a century, probably employ stress testing in their bond rating methodology, but old documentation to this effect is hard to come by.

[17] For an excellent overview and comparison of these and related models, see Koyluoglu and Hickman (1998).

[18] The most recent guidance on counterparty credit risk, SR 11-10, has greatly expanded on stress testing expectations. All SR letters can be found at http://www.federalreserve.gov/bankinforeg/srletters/srletters.htm.

[19] Presentation delivered at Marcus Evans conference "Implementing stress tests into the risk management process", Washington DC, March 1–2, 2007.

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

solvency (i.e., just enough capital to have a positive net worth), but rather some notion of adequate capitalization *post-stress*. For instance, the 2009 SCAP in the US presented a two-year scenario with a post-stress hurdle of 4% Tier 1 common capital. The 2012 bottom-up Spanish stress test used a three-year scenario with a post-stress hurdle of 6% core Tier 1 capital, suggesting a lower risk appetite by the Spanish authorities than by the American.

While length and post-stress hurdles are easy to compare across macro-prudential stress tests, scenario severity is not. Authorities are reluctant to make statements like "a 1 in 100 scenario" which would allow such comparisons, in part because such a statement is very difficult to make credibly. In its stress testing program, the Federal Reserve makes available time series of relevant variables to allow users to assess the severity of a given scenario, at least for those variables.[20] Of course, a multivariate assessment is much more difficult.

Once the risk appetite has been established, one of the principal challenges faced by both the supervisors and the firms when designing stress scenarios is coherence. The scenarios are inherently multi-factor: we are seeking to develop a rich description of adverse states of the world in the form of several risk factors, be they financial or real, taking on extreme yet coherent (or possible) values. It is not sufficient to specify only high unemployment or only a significant widening of credit spreads or only a sudden drop in equity prices; when one risk factor moves significantly, the others move too. The real difficulty is in specifying a coherent joint outcome of all of the relevant risk factors. For instance, not all exchange rates can depreciate at once; some have to appreciate. A high inflation scenario needs to account for likely monetary policy responses, such as an increase in the policy interest rate. Every market shock scenario resulting in a flight from risky assets – "flight to quality" – must have a (usually small) set of assets that can be considered safe havens. These are typically government bonds from the safest sovereigns (e.g., the US, Japan, Germany, Switzerland). Of course, as sovereign government budgets are increasingly strained, questioning the ultra-low risk assumption of such treasury instruments would certainly be a worthwhile stress scenario, but it would need to define an alternative "risk-free" asset class to which capital can flee.

While the problem of coherence is generic to scenario design, it is especially acute when considering stress scenarios for market risk, i.e., for portfolios of traded securities and derivatives. These portfolios are typically marked to market as a matter of course, and risk managed in the context of a value-at-risk (VaR) system. In practice, this means that the hundreds of thousands (or more) of positions in the trading book are mapped to tens of thousands of risk factors, which are tracked on a (usually) daily basis and form the "data" used to estimate risk parameters like volatilities and correlations. Finding coherent outcomes in such a high dimensional space, short of resorting to historical realizations, is daunting indeed.

Compounding the problem is the challenge of finding a scenario in which the real and financial factors are jointly coherent. The 2009 SCAP had a rather simple scenario specification. The state space had only three dimensions – GDP growth, unemployment, and the house price index (HPI) – and the market risk scenario was based in historical experience: an instantaneous risk factor impact reflecting changes from June 30 to December 31, 2008. This period represented a massive flight to quality, with the markets experiencing the failure of at least one global financial institution (Lehman), and risk premia at the time arguably placed a significant probability on the kind of adverse real economic outcome painted by the tri-variate SCAP scenario. This solution achieved a loose coherence of the real and financial stresses. However, the price that one pays for choosing a historical scenario is the usual one: it does not test for anything new. Figs. 3 and 4 compare some of these risk factors (real GDP, unemployment, equity and home prices indices) across the four US stress tests to date, both to each other and to actual realizations since 2008 Q4.

For the 2011 EBA test, the supervisors specified over 70 risk factors for the trading book, eight macro-factors for each of 21 countries (macro-factors such as GDP growth, inflation, unemployment, real estate price indices, both residential and commercial, short and long term government rates, and stock prices), plus sovereign haircuts across seven maturity buckets. The macroeconomic stress scenario was generated by economists at the ECB with reference to the EU Commission baseline economic forecast.

All supervisory stress tests to date have imposed the same scenario on all banks. Naturally, any scenario may be more severe for some banks and much less so for others, depending on the business mix and geographic footprint. This one-size-fits-all approach is analogous to the problem of regulatory vs. internal economic capital models: the former is the same for all banks by design, while the latter, being limited to a given bank, takes the particular business mix of that bank into account directly. This problem of same vs. specific stress scenarios becomes especially acute when we move from crisis times, when there may be less debate about what a relevant adverse scenario might look like, to "normal" times. The US CCAR program, which has been in operation since 2011, recognized this problem and asks banks to submit results using their own scenarios (baseline and stress) in addition to results under the common supervisory stress scenario. This was an important step forward from the 2009 SCAP: by asking banks to develop their own stress scenario(s), thus revealing the particular sensitivities and vulnerabilities of their specific portfolio and business mix, supervisors could learn what the banks themselves thought to be the high risk scenarios. This is useful not just for micro-prudential supervision – learning about the risk of a given bank – but also for macro-prudential supervision, by allowing for the possibility of learning about common risks across banks which may hitherto have been undiscovered or under-emphasized. With this dual approach, supervisors could compare results across banks from the common scenario directly, without sacrificing risk-discovery.

[20] See http://www.federalreserve.gov/bankinforeg/bcreg20121115a3.xlsx.

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

4. Executing the stress scenario: losses and revenues

With the macro-scenario in hand, how does one arrive at the corresponding micro-outcomes: losses and revenues under adverse market and macroeconomic conditions? To date, there has been very little discussion in the public domain on how to solve this problem, except perhaps for stress testing the trading book. Indeed, one of the more important contributions of the supervisory stress tests in the US and Europe has been the accompanying methodology documents that have been disclosed by the supervisors, which are, understandably, more heavily focused on the banking book.[21]

4.1. Modeling losses

For a firm which is active in many markets (product and geography), the first task is to map from the few macro-factors to the many intermediate risk factors that drive losses for particular products by geography. The EBA was forced to confront the problem of geographic heterogeneity directly, since it spans 21 sovereign nations with rather different economies. US supervisors, in stress testing an economic region only slightly smaller than that of the EBA, left the task of accounting for the not-inconsiderable geographic heterogeneity to individual firms. Regional differences are critical in modeling losses for real estate lending (residential and commercial), but are hardly limited to those products. Since the US experiences regional business cycles – the national business cycle obscures a considerable degree of variation across states – nearly all lending has some geographic component. For example, credit card losses are especially sensitive to unemployment, and in July 2011, with the national rate at 9.1%, the state-level unemployment rate ranged from 3.3% in North Dakota to 12.9% in Nevada. Similar dynamics are at work in wholesale lending, particularly for SME (small and medium enterprise) lending, whose performance has a strong geographic component.

The problem of mapping from macro to more intermediate risk factors is not limited to geography. An interesting example is auto lending and leasing, where the collateral assets are used cars. While auto sales invariable decline during a recession, and the decline in 2008–2009 was unprecedented in the post-war period, used car sales typically suffer less. Yes, households buy fewer cars in a recession, but if they do need to purchase a car, it is relatively more likely to be a used car. Thus, even if the default rate on auto loans increases significantly during a recession, the corresponding loss given default (LGD) or loss severity need not. A useful indicator of the health of the used car market, and thus the collateral of an auto lending portfolio, is the Manheim index. Over the course of the most recent recession (Dec. 2007–June 2009), the index rose 4%, while total new auto and light truck sales declined by 37%.

The problem of loose coupling of the loss severity to the business cycle is not limited to auto loans. Acharya, Bharath, and Srinivasan (2007) show that for corporate credit, an important determinant of LGD is whether the industry of the defaulted firm is in distress at the time of default. The authors make a compelling asset specificity argument: if the airline industry is in distress, and a bank is stuck with the collateral on defaulted aircraft loans or leases, it will be hard to sell those aircraft except at very depressed prices. The healthcare sector may be relatively robust at the time, as indeed it was in the recent recession, but it is difficult to transform an airplane into a hospital.

The EBA disclosure on methodology is especially rich. In the March 2011 document, for example, detailed guidance is provided on stressed probabilities of default (PDs) and stressed LGDs. Note that such guidance presumes that a bank has implemented an internal credit rating system for its commercial loan portfolio. For a Basel II bank this may not be unreasonable, since internal ratings, mapped to a common external scale such as those used by the rating agencies, are a cornerstone of the Accord. With a credit rating (internal or external) in hand, computing stressed default rates for the portfolio becomes a straightforward exercise, either by assigning higher PDs to a given rating, or by imposing a downward migration on the current portfolio.[22] Since the EBA stress test was based on risk weighted assets (RWA) computed using Basel II risk weights, which are ratings sensitive, banks were forced to make use of stress migration matrices to compute not only increased defaults (the last column of the matrix), but also the entire future ratings distribution, to arrive at the correct RWA value. The US stress tests were conducted under Basel I risk weights, which are not obligor ratings sensitive. The fuss about RWA calculations is important, since the denominator of capital ratios, used to determine whether or not a bank needs to raise capital, is RWA. Clearly, this complicates any comparison of US and European stress test results.

Implementation in the trading book is more straightforward, and has been discussed extensively in the public domain; see inter alia Allen, Boudoukh, and Saunders (2004), Jorion (2007), or Rebonato (2010). In a nutshell, existing positions are simply repriced using the stress scenario risk factors, subject to the proviso that the risk factor mapping problem, discussed in Section 3, has been solved. The corresponding problem of stressing the counterparty credit risk that comes with the activities of derivatives has received less attention.[23] Counterparty credit risk arises when, in a derivative transaction which is revalued to the stress scenario, the bank finds itself in the money (i.e., enjoys a

[21] For SCAP, see http://www.federalreserve.gov/bankinforeg/bcreg20090424a1.pdf. For EBA, see http://www.eba.europa.eu/EU-wide-stress-testing/2011/The-EBA-publishes-details-of-its-stress-test-scena.aspx. For the 2011 and 2012 CCAR, see http://www.federalreserve.gov/newsevents/press/bcreg/bcreg20110318a1.pdf and http://www.federalreserve.gov/newsevents/press/bcreg/bcreg20120313a1.pdf respectively.

[22] Of the 90 participating banks, 59 were so-called IRB (internal ratings based) banks, meaning that their internal models were validated to the supervisor's satisfaction for at least one regulatory portfolio (e.g., corporate, commercial real estate, etc.). Non-IRB banks were given very non-specific guidance (EBA, 2011, Section 5.5.1.1).

[23] For an excellent treatment, see Canabarro (2010) and Hopper (2010).

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

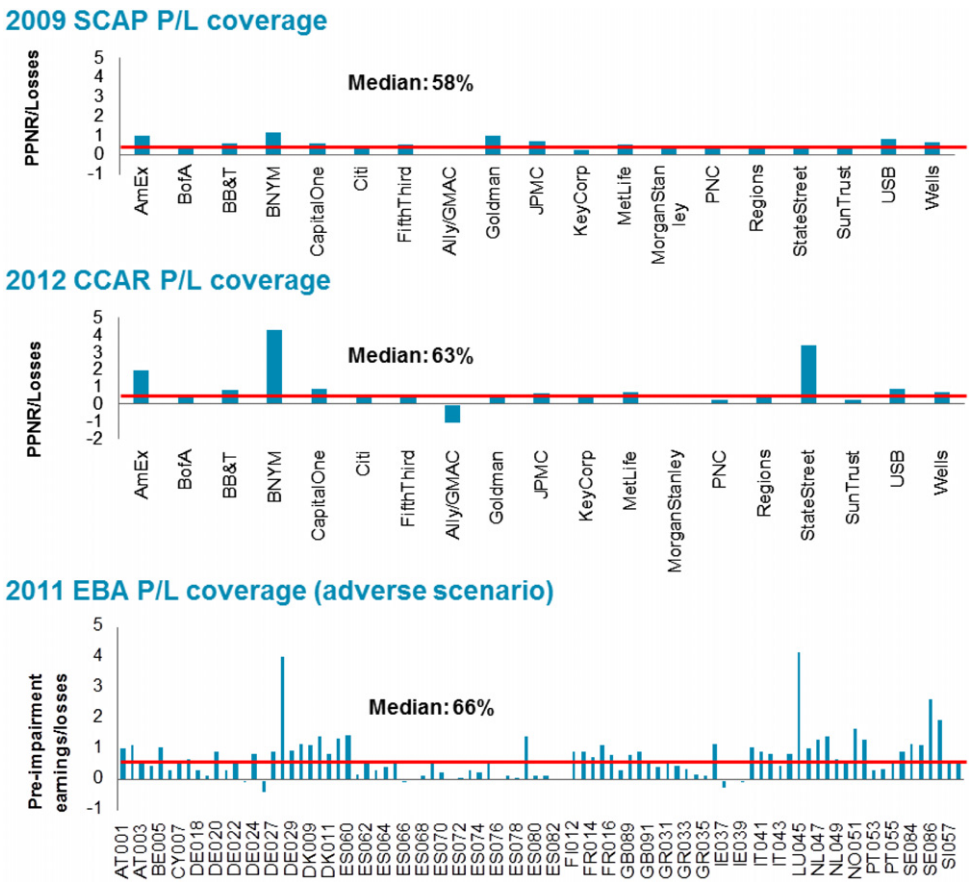

Fig. 1. Projected coverages of losses with profits in the 2009 SCAP and 2011 EBA stress tests.

derivative receivable), but cannot be sure that the counterparty to the transaction will be solvent in order to make good on the payment. Thus, the value is discounted, where the discount is a function of the expected default likelihood of the counterparty *under the stress scenario*, which is presumably higher than today. This adjustment is called a credit value adjustment (CVA), and banks with significant derivative activities manage CVA as a matter of course. As Canabarro (2010) and Hopper (2010) point out, the modeling challenge of stress testing counterparty credit risk is considerable. Not only does the PD of the counterparty change in a stressful environment, the exposure does likewise. Thus, any CVA stress test involves two distinct simulation exercises. If the collateral posted by the counterparty is anything other than cash or a cash equivalent, a revaluation of that collateral under the same stress scenario needs to be added to the process.[24]

4.2. Modeling revenues

Implementing stress scenarios on the revenue side of the equation remains largely a black box, and seems far less well developed than stress testing for losses. Neither the 2009 SCAP nor the otherwise richly documented 2011 EBA disclosures devoted much space or revealed much detail about the methods and approaches for computing revenues under stressful conditions. Banks' total income can be divided roughly into interest and non-interest income. The interest income is clearly a function of the yield curve and credit spreads posited under the stress scenario, but the net impact of rising or falling rates on bank profitability remains ambiguous, perhaps in part because of interest rate hedging strategies (English, 2002; Purnanandam, 2007). The impact of stress scenarios on the non-interest income, which includes service charges, fiduciary, fees, and other income (e.g., from trading), is far harder to assess, and there has been precious little discussion of its determinants in the literature. This is concerning, since Stiroh (2004) shows that not only has the share of non-interest income in US banks been rising steadily, from 25% in 1985 to 43% in 2001, but it is associated with a greater volatility and lower risk-adjusted returns. If we compare the 2009 SCAP, the 2011 EBA and the 2012 CCAR stress tests, the median bank in the US was able to cover about 58% of its total projected losses with profits (including reserve releases, if any) in 2009 and 63% in 2012,[25] compared

[24] There is the added complication that major derivatives dealers actively manage CVA risk using a range of strategies and instruments that themselves vary in price and availability depending on market conditions.

[25] PPNR calculations in the 2012 CCAR were net of operational risk related losses and OREO expenses, as well as mortgage repurchase and

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

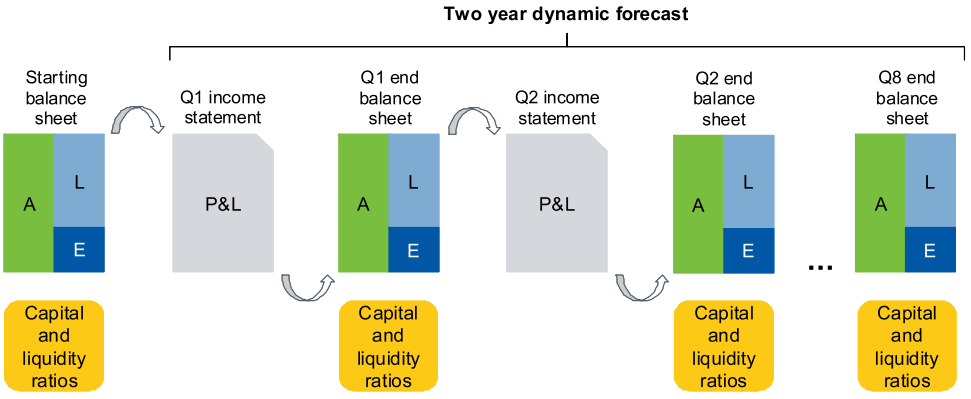

Fig. 2. Stress testing balance sheet and income statement dynamics.

with 66% in the European case. As Fig. 1 shows, there is a considerable degree of variability across banks, especially in the EBA test, where in some cases profits are projected to outpace losses 4:1, even under the stress scenario!

4.3. Modeling the balance sheet

Recall that capital adequacy is defined in terms of a capital ratio, roughly capital over assets. Of course, both the numerator and denominator are nuanced. All supervisory stress tests have insisted, to varying degrees, that the relevant form of capital be common equity. The 2010 CEBS test allowed for some forms of hybrid capital which are typical of state participations, but the requirements were tightened a year later. As was discussed in Section 4.1, the denominator is typically risk-weighted assets (RWA), where the risk weights are determined by the prevailing regulatory capital regime, namely Basel I (in the US cases of the SCAP and CCAR) and Basel II (in the European stress tests). The many subtleties of what this implies are beyond the scope of this paper; suffice it to say that a bank may be forced to raise capital under one regime but not the other, and there is no way to know which regime will result in a more favorable treatment without knowing about the portfolio in considerable detail.

Regardless of the risk weight regime, determining the post-stress capital adequacy requires modeling of both the income statement and the balance sheet, both flows and stocks, over the course of the stress test horizon, which is typically two years.[26] This is illustrated in Fig. 2 below. The point of departure is the current balance sheet, at which point the bank meets the required capital (and, if included, liquidity) ratios. The starting balance sheet generates the first quarter's income and loss, which in turn determines the quarter-end balance sheet. The modeler is then faced with the problem of considering the nature and amount of new assets originated and/or sold during the quarter, and any other capital depleting or conserving actions such as acquisitions or spin-offs, dividend changes or share (re-)purchase or issuance programs, including employee stock and stock option programs. The problem of balance sheet modeling exists under a static (be it in raw form, as in the 2011 EBA, or in risk weighted form, as in the 2009 SCAP) or dynamic balance sheet assumption. The bank should not drop below the required capital (and liquidity) ratios in any quarter. Moreover, at the end of the stress horizon, the bank needs to estimate the amount of reserves needed to cover expected losses on loans and leases for the following year. In this way, the stress tests are really three years (or $T + 1$ years for a T-year stress test).

5. Stress testing disclosure

Stress testing is here to stay, whether because it is just good risk management practice, or because it is enshrined in legislation (through the Dodd–Frank Act). In the debate on disclosure regimes, it is not clear that more is always better. We divide the discussion into crisis and non-crisis or normal times, with the simple point that normal times may not require or even desire the same degree of transparency as is clearly needed in times of crisis.

We have seen very large differences in disclosure across the different supervisory stress tests, as summarized in Table 1. The SCAP in 2009 opened Pandora's box by disclosing projected stress losses for each of the 19 participating banks, for eight different categories or asset classes, as well as resources other than capital for absorbing losses (mostly pre-provision net revenue and reserve releases, if any). Until then, regulatory disclosures (e.g., Y-9C reports for US bank holding companies) reported only realized losses (the past), not projected losses (a possible future). This allowed the market to check the severity of the stress test easily, not just in terms of the scenario, but also, and much more importantly, in terms of the resulting outcomes at the bank level. Given the crisis of confidence which was prevalent in the market at the time, this amount of transparency was crucial. Two years later, the CCAR displayed a radically different disclosure regime: only the macro-scenario was published, with no bank-level results. The only indications of bank-level outcomes were the subsequent dividend and other capital actions announced by some banks: banks which were allowed to raise their dividends

put-back costs, meaning that these items were not reported separately (though they totaled $115 bn for all 19 banks) (Board of Governors, 2012).

[26] The horizon is 9 quarters for the CCAR, as it is based on Q3, not Q4, balance sheets.

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

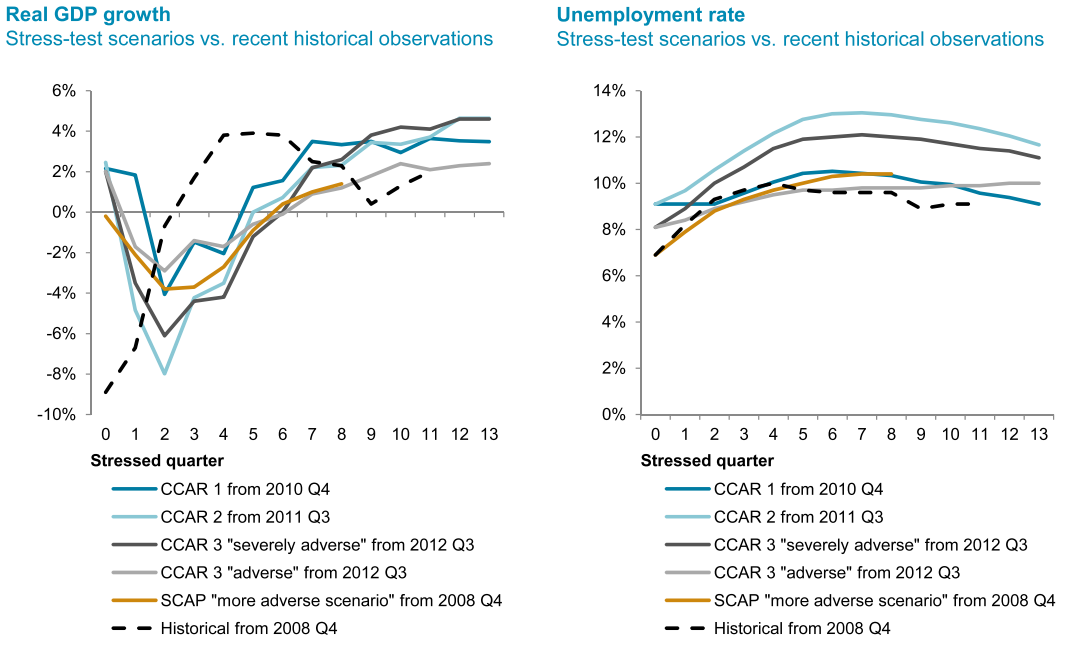

Fig. 3. US real GDP and unemployment scenarios compared.
Source: Fed, The Supervisory Capital Assessment Program: Design and Implementation, 24 April 2009; Fed, Comprehensive Capital Analysis and Review: Objectives and Overview, 18 March, 2011; Fed, "Comprehensive Capital Review" document and "Capital Plan review" 22 November 2011; Fed, "Supervisory Scenarios" 15 November 2012; Datastream.

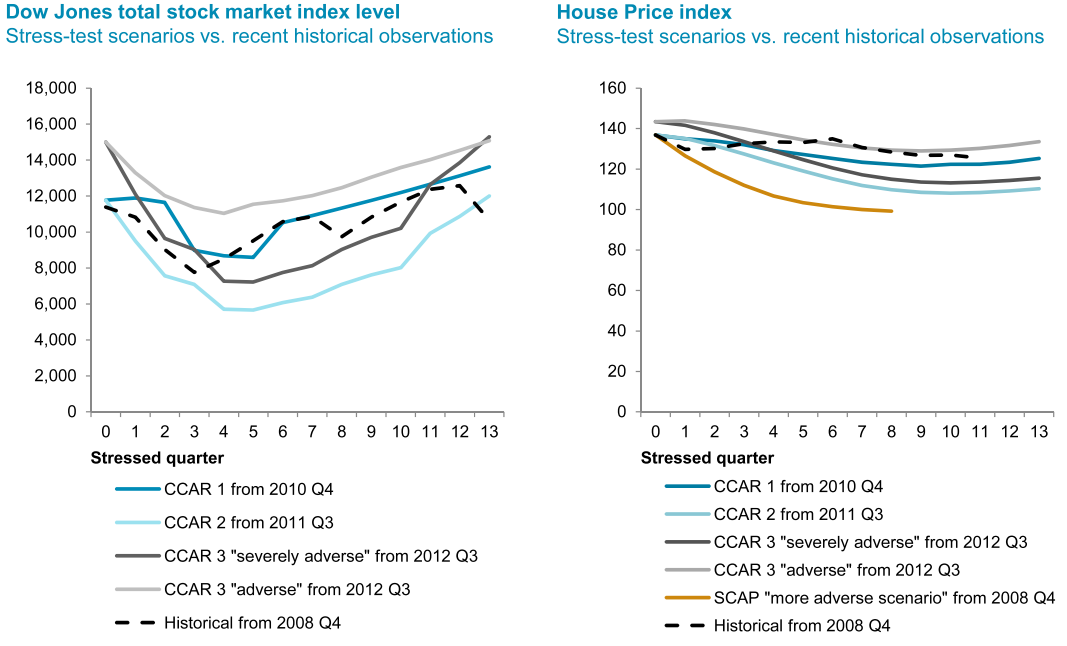

Fig. 4. US equity and house price indices compared.
Source: Fed, The Supervisory Capital Assessment Program: Design and Implementation, 24 April 2009; Fed, Comprehensive Capital Analysis and Review: Objectives and Overview, 18 March, 2011; Fed, "Comprehensive Capital Review" document and "Capital Plan review" 22 November 2011; Fed, "2013 Supervisory Scenarios" 15 November 2012; Datastream.

were interpreted as having "passed" the stress test. The market digested this meager information event without a hiccup.

Dodd–Frank, however, requires the Fed to disclose the results of regular stress testing, and the 2012 CCAR, with the accompanying rules (final and proposed[27]), gave a glimpse of what regular disclosure might look like. The

[27] http://www.gpo.gov/fdsys/pkg/FR-2011-12-01/pdf/2011-30665.pdf.

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

2012 CCAR disclosed nearly the same level of detail as the 2009 SCAP, namely bank-level loss rates and dollar losses by major regulatory asset classes (following the categories of the FR Y-9C bank holding company reports): first and second lien mortgages, commercial and industrial (C&I) lending, CRE, credit cards, other consumer, and other loans. In addition, the Fed reported the dollar PPNR, gains/losses on the AFS/HTM securities portfolio, and trading and counterparty losses for those firms who were required to conduct the trading book stress.[28] Again, as with the 2009 SCAP, the numbers reported were supervisory estimates, not the banks' own estimates of losses (and PPNR) under the stress scenario.

By contrast, the 2011 Irish and 2011 Europe-wide EBA stress tests, both of which were disclosed after the CCAR, were considerable in their detail, including comparisons of bank and third-party estimates of losses in the Irish case (revealing the bias that any bank is likely to have when estimating its own potential losses), and data in electronic, downloadable form in the EBA case. Ireland in particular was suffering from an acute credibility problem, having emerged from the CEBS stress test with flying colors in July 2010, only to require massive external aid four months later.

This difference in experiences between Europe and the US provides some hints on how to design a disclosure regime during "normal" times. The discussion of the benefits *and* costs of stress test disclosures by Goldstein and Sapra (2012) is helpful. They argue persuasively that in a world with frictions and strategic environments, the benefits (better market discipline) may not outweigh the costs: banks may make poor portfolio choices which are designed to maximize the chance of passing the test (window dressing), thereby giving up longer term value; while traders may place too much weight on the public information of stress test disclosure and lose their incentive to produce private information about the banks; and finally, with the information content of market prices having been damaged, market discipline is harmed, and supervisors will find market prices less useful for policy decisions (micro- as well as macro-prudential).

Clearly, some disclosure is still preferable to no disclosure, and Goldstein and Sapra propose the disclosure of aggregated but not necessarily bank-specific results, with sufficient information about category outcomes (loss rates by major asset class, for instance). Aggregation has the advantage of being less wrong, since the idiosyncratic errors in estimating bank conditions under hypothesized stress scenarios are averaged out. In this way, supervisors can still provide the useful macro-prudential information which only they can provide – loss rates by asset class, total capital decline in the system (or significant fraction of the banking system) – without drowning out signals about individual banks from the market participants themselves. Such a disclosure gives the market an anchor point for system-wide possibilities, without diluting the incentive to dig hard into a particular firm's financials.

During times of crisis, with the enormous uncertainty about the health of the banking system, the benefit of detailed bank-specific stress test disclosure is significant, given the ability of supervisors to assess the health of individual firms correctly, and the resulting inability of the market distinguish between a good bank and a bad. Indeed, Goldstein and Sapra argue that stress test disclosures, when more disaggregated, ought to be accompanied by detailed descriptions of the exposures of the banks. This is precisely what was done in the Irish bank stress test of 2011, an acute case of loss of confidence (and a subsequent regaining of confidence), as well as in the 2011 EBA stress test. Because the credibility of European supervisors was rather low by that point, only with a very detailed disclosure, bank by bank, of their exposures by asset class, by country and by maturity bucket, could the market do its own math and arrive at its own conclusions.

Between March 2009 and March 2011, the 19 SCAP banks had raised about $300 bn in capital and the S&P500 had increased by 65%; by the end, the economy was no longer in recession, and, arguably, the supervisory agencies had regained credibility. The non-event of the non-disclosure of the 2011 CCAR suggests that the market seems content to live in a state of "symmetric ignorance", to borrow a term from Dang, Gorton, and Holmstrom (2010). Of course, this might change were the economy to receive another adverse shock, but until it does, it is not clear that an EBA-like disclosure regime is necessarily either desirable or stability-enhancing. In contrast, Europe is not yet out of the woods (at the time of writing); yet even the EBA was not limitless with its disclosure of the 2011 stress test results. It is worth noting that funding liquidity was also stressed for banks, but without disclosing the results. Because liquidity positions are highly dynamic, and thus subject to rapid change, snapshot disclosure, especially with a delay (the as-of date for the 2011 EBA stress test was YE 2010), is unlikely to be informative at the time of disclosure.[29]

Recall the discussion in the introduction: regulatory capital models (risk weighting), internal economic capital models and stress testing all have the same goal, namely to determine the amount of capital needed to support the business (risk taking) of the bank. Both regulatory and economic capital models (and especially the former) evolve very slowly, and thus have difficulty in adapting to financial innovations and rapidly changing macro conditions. Indeed, some of the innovation is motivated by those slowly evolving, one-size-fits-all regulatory capital rules. Moreover, bank balance sheets are notoriously opaque and subject to easy-to-hide asset substitution (higher risk for lower risk assets); see Morgan (2002). Stress tests, especially macro-prudential supervisory stress tests, are adapted to the then-current environment and bank portfolios by construction. Between balance sheet opacity, asset substitution and regulatory arbitrage, it is easy to see the value of a "pop quiz" in the form of bespoke stress testing (Acharya, Mehran, Schuermann, & Thakor, 2011).

[28] In 2012, these were the six institutions with the largest trading portfolios.

[29] Reuters, Sept. 2, 2011, "EBA won't seek disclosure of bank liquidity". Available at http://www.reuters.com/article/2011/09/02/idUSL5E7K23PI20110902.

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003

6. Conclusion

The problem of sizing the amount of capital needed to support a bank's risk taking is not new, but the use of broad-based supervisory stress tests for an entire banking system is. The first use of such tests was in the US in 2009, and its success there has made it the supervisory and risk management hammer for dealing with all nails. A critical component of the exercise is the disclosure of the results. The reason why stress testing became an imperative was precisely because existing approaches that were publicly disclosed, such as regulatory capital ratios, were no longer informative, and were heavily (if not entirely) discounted by the market. In order to regain their credibility, supervisory authorities needed to disclose enough to allow the market to "check the math".

However, broad-based supervisory stress testing has not been universally successful, as the 2010–2011 European experience has shown. Nor is it clear how useful such broad supervisory stress testing with concomitant disclosure would be as a matter of routine. Its value in the crisis was undoubtedly due to its "pop quiz" nature. It was sprung on the banks at short notice, and thus was very difficult for them to manipulate through careful pre-positioning; and it was tailored to the situation at hand, genuinely revealing new information to all participants and the public. As a result, trust was regained. Once that trust has been re-established, the cost-benefit of stress testing disclosures may tip away from bank-specific towards more aggregated information. This still provides the market with unique information (after all, supervisors have access to proprietary bank data) without taking away market participants' incentives to produce private information and trade on it—with all the downstream benefits of information-rich prices and market discipline.

Acknowledgments

I would like to thank John Fell, Mark Flannery, Itay Goldstein, Bengt Holmstrom, Bill Janeway, Umit Kaya, Ugur Koyluoglu, Andy Kuritzkes, John Lester, Clinton Lively, Hashem Pesaran, Brian Peters, Barry Schachter, Hal Scott, and members of the Committee for Capital Markets Regulation for helpful discussions and suggestions. I am also thankful to Cary Lin for helpful research assistance, as well as the participants in the workshop on "Predicting Rare Events" sponsored by the IF/Federal Reserve Bank of San Francisco. All errors remain mine, of course.

References

Acharya, V. V., Bharath, S. T., & Srinivasan, A. (2007). Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of Financial Economics*, *85*, 787–821.

Acharya, V., Mehran, H., Schuermann, T., & Thakor, A. (2011). *Robust capital regulation.* Federal Reserve Bank of New York Staff report no. 490.

Allen, L., Boudoukh, J., & Saunders, A. (2004). *Understanding market, credit and operational risk: the value at risk approach*. Blackwell: New York, NY.

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C., & Schuermann, T. (2002). Ratings migration and the business cycle, with applications to credit portfolio stress testing. *Journal of Banking and Finance*, *26*(2–3), 235–264.

Berkowitz, J. (2000). A coherent framework for stress testing. *Journal of Risk*, *2*, 1–11.

Board of Governors of the Federal Reserve System (2012). Comprehensive capital analysis and review 2012: methodology and results for stress scenario projections. 13 March, 2012. Available at: http://www.federalreserve.gov/newsevents/press/bcreg/bcreg20120313a1.pdf.

Canabarro, E. (2010). Pricing and hedging counterparty risk: lessons re-learned? In E. Canabarro (Ed.), *Counterparty credit risk (Chapter 6)*. London, UK: Risk Books.

Committee on the Global Financial System (2001). A survey of stress tests and current practice at major financial institutions. Available at: http://www.bis.org/publ/cgfs18.htm.

Dang, T.V., Gorton, G., & Holmstrom, B. (2010). *Financial crises and the optimality of debt for liquidity provision.* Working paper. Available at: http://mfi.uchicago.edu/publications/papers/ignorance-crisis-and-the-optimality-of-debt-for-liquidity-provision.pdf.

English, W. B. (2002). Interest rate risk and bank net interest margins. *BIS Quarterly Review, December*, 67–82.

European Banking Authority (2011). 2011 EU-wide stress test: methodological note. 18 March 2011. Available at: http://www.eba.europa.eu/EU-wide-stress-testing/2011/The-EBA-publishes-details-of-its-stress-test-scena.aspx.

Fender, I., Gibson, M. S., & Mosser, P. C. (2001). An international survey of stress tests. *Current Issues in Economics and Finance*, *7*(10), Federal Reserve Bank of New York.

Flannery, M. J. (2012). *Measuring equity capital for stress-testing large financial institutions*. Working paper.

Foglia, A. (2008). *Stress testing credit risk: a survey of authorities' approaches*. Banca d'Italia occasional paper, No. 37.

Goldstein, I., & Sapra, H. (2012). *Should banks' stress test results be disclosed? An analysis of the costs and benefits*. Working paper. Available at: http://finance.wharton.upenn.edu/~itayg/Files/disclosure.pdf.

Gupton, G. M., Finger, C., & Bhatia, M. (1997). *CreditMetricsTM—technical document.* This version: April 2. J.P. Morgan. Available at: http://www.defaultrisk.com/_pdf6j4/creditmetrics_techdoc.pdf.

Hopper, G. (2010). Stress testing and scenario analysis: some second generation approaches. In E. Canabarro (Ed.), *Counterparty credit risk (Chapter 11)*. London, UK: Risk Books.

Jorion, P. (1996). *Value at risk: the new benchmark for managing financial risk* (1st ed.). New York, NY: McGraw Hill.

Jorion, P. (2007). *Value at risk: the new benchmark for managing financial risk* (3rd ed.). New York, NY: McGraw Hill.

Koyluoglu, H. U., & Hickman, A. (1998). Credit risk: reconcilable differences. *Risk*, *11*(10), 56–62.

Kupiec, P. H. (1998). Stress testing in a value at risk framework. *Journal of Derivatives*, *6*(1), 7–24.

Kuritzkes, A., & Scott, H. (2009). Markets are the best judge of bank capital. *Financial Times*, September 23.

Lester, J., Reynolds, P., Schuermann, T., & Walsh, D. (2012). *Strategic capital: defining an effective real world view of capital.* Oliver Wyman financial services report. Available at: http://www.oliverwyman.com/strategic-capital-defining-an-effective-real-world-view-of-capital.htm.

Morgan, D. P. (2002). Rating banks: risk and uncertainty in an opaque industry. *American Economic Review*, *92*(4), 874–888.

Purnanandam, A. (2007). Interest rate risk management at commercial banks: an empirical investigation. *Journal of Monetary Economics*, *54*, 1769–1808.

Rebonato, R. (2010). *Coherent stress testing: a Bayesian approach to the analysis of financial stress*. New York: John Wiley & Sons.

Stiroh, K. (2004). Diversification in banking: is noninterest income the answer? *Journal of Money, Credit and Banking*, *36*(5), 853–882.

Wilde, T. 1997. CreditRisk+ —a credit risk management framework. Available at: http://www.csfb.com/institutional/research/assets/creditrisk.pdf.

Wyman, O. (2012a). Bank of Spain stress testing exercise. Available at: http://www.bde.es/webbde/GAP/Secciones/SalaPrensa/InformacionInteres/ReestructuracionSectorFinanciero/Ficheros/en/informe_oliverwymane.pdf.

Wyman, O. (2012b). Asset quality review and bottom-up stress test exercise. Available at: http://www.bde.es/f/webbde/SSICOM/20120928/informe_ow280912e.pdf.

Please cite this article in press as: Schuermann, T. Stress testing banks. International Journal of Forecasting (2013), http://dx.doi.org/10.1016/j.ijforecast.2013.10.003