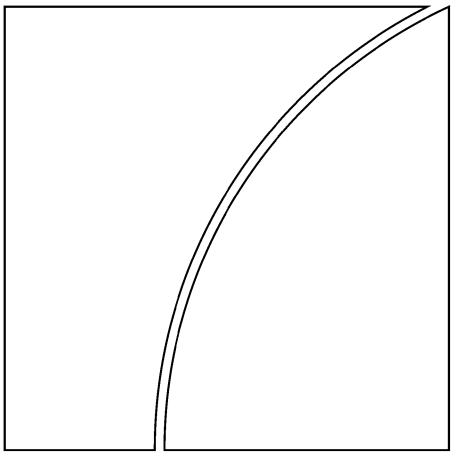
BANK FOR INTERNATIONAL SETTLEMENTS



BIS Working Papers

No 369

Stress-testing macro stress testing: does it live up to expectations?

by Claudio Borio, Mathias Drehmann and Kostas Tsatsaronis

Monetary and Economic Department

January 2012

JEL classification: E30, E44, G10, G20, G28.

Keywords: stress tests, financial instability, macroprudential.

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS.

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)

ISBN 1682-7678 (online)

Stress-testing macro stress testing: does it live up to expectations?

Claudio Borio, Mathias Drehmann and Kostas Tsatsaronis[1]

Abstract

We critically review the state of the art in macro stress testing, assessing its strengths and weaknesses. We argue that, given current technology, macro stress tests are ill-suited as early warning devices, ie as tools for identifying vulnerabilities during seemingly tranquil times and for triggering remedial action. By contrast, as long as properly designed, stress tests can be quite effective as crisis management and resolution tools. We also see additional side benefits, stemming largely from the way such tests can discipline thinking about financial stability. We suggest possible ways to improve their performance.

JEL classification: E30, E44, G10, G20, G28.

Keywords: stress tests, financial instability, macroprudential.

[1] Bank for International Settlements.

Contents

> *"The banking system's reported financial indicators are above minimum regulatory requirements and stress tests suggest that the system is resilient"* (IMF, *Iceland: Financial Stability Assessment – update*, 19 August 2008, p 5)

Introduction[2]

"….and stress tests suggest that the system is resilient". What the IMF said of Iceland in its Financial Stability Assessment released on 19 August 2008 may sound extraordinary to the uninitiated. But it simply echoed the message of stress tests carried out by authorities and banks around the globe ahead of what turned out to be one of the worst financial crises in world history: "The system is sound"; "The institution is strong and resilient". This is the relentless message confronting those of us who were deeply involved in assessing vulnerabilities during the years of the so-called Great Moderation. And, as the quotation highlights, this is the message that persisted even as that Moderation began to show cracks. Had Winston Churchill been still alive, he might well have said: "Never in the history of mankind have so many got it so wrong for so long."

It is, of course, all too easy to criticise stress tests after the fact; but the financial crisis raises a key question: what can and cannot be expected of them, now and in the future? The question is all the more pressing at a time when macro stress testing is becoming a standard weapon in the arsenal of the macroprudential frameworks that the authorities are implementing around the globe (FSB-IMF-BIS (2011)).

It is this question that we begin to explore in our paper. We focus on "macro stress testing", designed to stress the financial system as a whole or sub-sets thereof, rather than on "micro stress testing", designed to stress individual institutions. We argue that, given current technology, macro stress tests are ill-suited as early warning devices, ie as tools for identifying vulnerabilities during tranquil times and for triggering remedial action. By contrast, they can be quite effective as crisis management and resolution tools, since in that context their messages may be more reliable. More generally, macro stress tests can discipline thinking about financial stability risks. In the process, they can yield additional benefits, such as helping to reconcile the widely different perspectives of the various stakeholders (banks, supervisory authorities, central banks and the public at large), foster better communication, cross-check the performance of individual firms' risk models, and identify important data gaps. That said, in order to yield the hoped-for benefits, it is critical to design stress tests properly, tailoring them to the specific purpose.

Whether macro stress tests will *ever* be able to act as effective early warning devices is an open question. Given the analytical challenges, we remain sceptical, although we hope that the efforts underway will prove us wrong. Be that as it may, the tool can only be the beginning, never the end, of a conversation about financial stability risks. It can only be a complement, and never a substitute, for other tools and processes. And what matters most is the mindset of those employing it.

---

[2] This paper was prepared for the conference on "The macroprudential toolkit: Measurement and analysis", sponsored by the Office of Financial Research and the Financial Stability Oversight Council, 1–2 December 2011. We would like to thank Dick Berner for suggesting to us that we write the paper and Anil Kashyap and Kevin Stiroh for excellent discussions and Stephen Cecchetti for comments. The views expressed are our own and do not necessarily reflect those of the Bank for International Settlements.

After defining what macro stress tests are, we organise our discussion around five propositions. We include in text boxes some additional information about the state of the art in stress testing. The conclusion wraps up the discussion.

I. One definition and five propositions

What are the defining characteristics of macro stress tests? What does current practice look like? What can stress tests do and what can they not do? How can they best be designed? We consider these issues sequentially.

What is a macro stress test?

Stress testing originated not in finance but in engineering. In its broadest sense, stress testing is a technique to test the stability of an entity or system under adverse conditions. In finance, it was originally used to test the performance of individual portfolios or the stability of individual institutions ("micro stress tests"). More recently, similar techniques have been employed to test the stability of *groups* of financial institutions that, taken together, can have an impact on the economy as a whole ("macro stress tests").

Any stress test, whether micro or macro, has four elements. The first is the set of *risk exposures* subjected to stress. The second is the *scenario* that defines the *(exogenous) shocks* that stress those exposures. The third is the *model* that maps those shocks onto an outcome (or impact), tracing their propagation through the system. The fourth is a measure of the *outcome*. For example, a typical macro stress test would test the solvency, as measured by the level of capital (outcome), of a group of financial institutions, whose balance sheets and income statements (risk exposures) are subject to a large recession (the scenario defining the shock(s)) by employing a set of reduced-form and/or structural relationships (the model).[3]

While the primary goal of a macro stress test is always to assess the stability of a group of financial institutions, it is worth distinguishing two more specific objectives, depending on the context. One is to identify vulnerabilities in *tranquil times* and provide the basis for addressing them, ie to act as an *early warning device*. The other is to *support crisis management and resolution*. This distinction will be important in what follows.

Proposition 1: Macro stress testing is a toolbox, not a single tool

Despite their common features, stress tests come in many shapes and sizes. They are not a single tool, but a toolbox. We next provide a brief overview (see Box 1 for additional technical detail).[4] We consider, in turn, the set of institutions and exposures assessed, the choice of scenarios, the features of the model, and the measures of the outcome.

[3] By design, therefore, rather than providing a summary measure of the entire distribution of potential outcomes, stress tests focus on the implications of a single (projected) scenario. As such, they are less general but more explicit and transparent in tracing out the system's reactions to specific events.

[4] For a more comprehensive analysis, see the survey by Drehmann (2009).

Box 1

Recent trends in macro stress testing

A decade ago, the IMF started using macro stress tests as part of its Financial Stability Assessment Programs. This method of assessing vulnerabilities also became popular among central banks. Following the outbreak of the current crisis, the main objective shifted from assessing vulnerabilities in tranquil times to supporting crisis management and resolution. This has helped to improve stress testing practices and has allowed modellers to refine their tools. Not least, more data and resources have become available. Yet the underlying techniques have remained broadly the same, as many of the most sophisticated models reviewed in this box had been developed previously.[1]

The earliest stress testing models were very basic, as they relied on equations linking aggregate profits and losses to macro developments (eg Blaschke et al (2001) or Bunn et al (2005)). In a data-poor environment this may still be the only possible approach. But more sophisticated techniques, as for example discussed in Segoviano and Padilla (2006), can help to uncover more robust estimates. Interestingly, instead of relying on complex models, Ong et al (2010) propose to use reverse stress tests as simple tools to uncover vulnerabilities in countries with limited data.

In a seminal contribution, Elsinger et al (2006) develop a model for the Austrian banking sector that integrates market risk, credit risk, interest rate risk and counterparty credit risk in the interbank sector. The model is the first that makes full use of credit register data and can thus achieve a very extensive coverage of on-balance sheet exposures.[2] The model outputs can be represented by loss distributions for the whole financial sector or particular banks or as aggregate value-at-risk (VaR) measures. The model can also be run in stress testing mode. Importantly, given the information about interbank exposures, the model can trace out how a default of one or more banks can spread through the system.[3] More recently, the model has been extended to capture the risk to profits and risks from cross-border exposures as well as to allow for a three-year forecast horizon (Boss et al (2008)).

In a stress testing exercise that integrates credit and interest rate risk in the banking book, Drehmann et al (2010) model assets and liabilities simultaneously. This ensures that banks' balance sheets balance at each point in time during the simulation horizon. Many stress testing models actually ignore this basic accounting identity. Given its granularity, the model provides a suitable framework for exploring the impact on banks' profits and losses of different (assumed) simple rules about the investment behaviour of banks once assets and liabilities mature or profits accumulate.

To date, the most comprehensive approach is RAMSI, the risk assessment model by the Bank of England (Aikman et al (2009)). Using Drehmann et al (2010) as one building block, the approach aims to model all the key channels highlighted Graph 1, including all relevant feedback mechanisms. So far it captures counterparty credit risk in the interbank market and allows for feedback channels arising from market and funding liquidity risk. Given a lack of data for estimating equations econometrically, liquidity risk is modelled by a range of indicators that change in stressed conditions in line with rules of thumb, calibrated to past crises (Kapadia et al (2011)).

Macroeconomic feedbacks are the focus of the work by Jacobson et al (2005). They propose a reduced-form approach for Sweden consisting of an aggregate vector autoregressive model (VAR) that includes the average default frequency of companies as a measure of financial stability, a model linking macro and balance sheet specific factors with defaults of companies, and a module tracing the evolution of balance sheets in response to macro factors. By integrating these three building blocks, they show that there are significant feedback effects from financial stability back to the real economy. De Graeve et al (2008) use the same methodology but proxy financial stability more directly, as they model the default probability of banks in Germany. They find that bank capitalisation has significant implications for the transmission mechanism of shocks to banks' balance sheets and back.[4] The new generation of dynamic stochastic general equilibrium (DSGE) models that include a financial sector may also at some point be useful for stress testing purposes, as a means of capturing macroeconomic feedbacks (eg Meh and Moran (2008) or Christiano et al (2010)). At this stage, though, the models are not yet rich and robust enough for policy exercises.

An alternative to more balance-sheet based models is to rely on contingent claims analysis. In a series of papers, Gray and various co-authors develop this method as a tool for macroprudential analysis (eg Gray et al (2006) or more recently Gray and Jobs (2010)). This allows them to derive

the mark to market value of interlinked sectoral balance sheets – including that of the government sector – in normal and stressed conditions. Given its origins in the Merton model (Merton (1974)), this approach in principle captures some non-linearities, specifically those around default boundaries. Merton-type models for particular sub-sectors have been used more broadly by, for instance, Pesaran et al (2006) or Düllmann and Erdelmeier (2009). The main innovation of the paper by Pesaran et al (2006), though, is to propose a global VAR framework for modelling national and international macroeconomic risk factors jointly, which has made it quite attractive for many other stress tests (eg Castren et al (2008)).

The biggest improvements in the area of stress testing in recent years have undoubtedly been in the treatment of liquidity risk. The work by the Bank of England discussed above is one such example. A similar approach is followed by Barnhill and Schumacher (2011), who calibrate the link between solvency and liquidity risk based on developments during the recent crisis. Van den End (2008) follows a different strategy. Rather than trying to build an overarching model capturing all risks simultaneously, he concentrates more specifically on liquidity risk, which allows for a richer analysis. Looking at the Dutch banking sector, he finds that once stress emerges in one bank it can quickly spread through the system.

_____

[1] For a detailed survey of the stress testing literature see Drehmann (2009). [2] Data from credit registers are now used by several countries for stress testing purposes (see Foglia (2008) for an overview). [3] Interestingly, Elsinger et al (2006) find that second-round effects associated with counterparty risk in the interbank market are of second order importance in their model. Joint defaults of banks are mostly driven by common exposures, ie exposures to systematic risk factors. [4] In particular, they find that the impact of a monetary policy shock can be six times larger when the banking system is weakly capitalised.

In principle, one would like to subject the whole financial system to a macro stress test. In practice, tests have considered parts of the overall system. Not surprisingly, the banking sector is the most common object of analysis, given its undisputed importance for financial stability. But stress tests have sometimes also covered other institutions, such as insurance companies and pension funds. Tests have tended to assess the strength of institutions in *individual* jurisdictions, although typically including their consolidated balance sheets worldwide. The only coordinated *multi-country* tests have been the recent exercises in the European Union.

Historically, macro stress tests have focused on credit risks in the banking, as opposed to the trading, book. Given the size of these exposures, this generally represents the core of the analysis. But the tests have also covered market risk in the trading book, risks to future income and counterparty credit risk in the interbank market. The most sophisticated variety also seeks to capture liquidity risk. While some risks are routinely considered together (eg credit risk in the banking book and future income risk), others are often considered individually. This is regularly the case for market risk or liquidity risk, as it has so far proved very hard to integrate them consistently with credit risk in the banking book.

Graph 1 provides a schematic overview of the structure of a typical macro stress test for banks. Clearly, the structure is simpler if the exercise addresses only one type of risk.

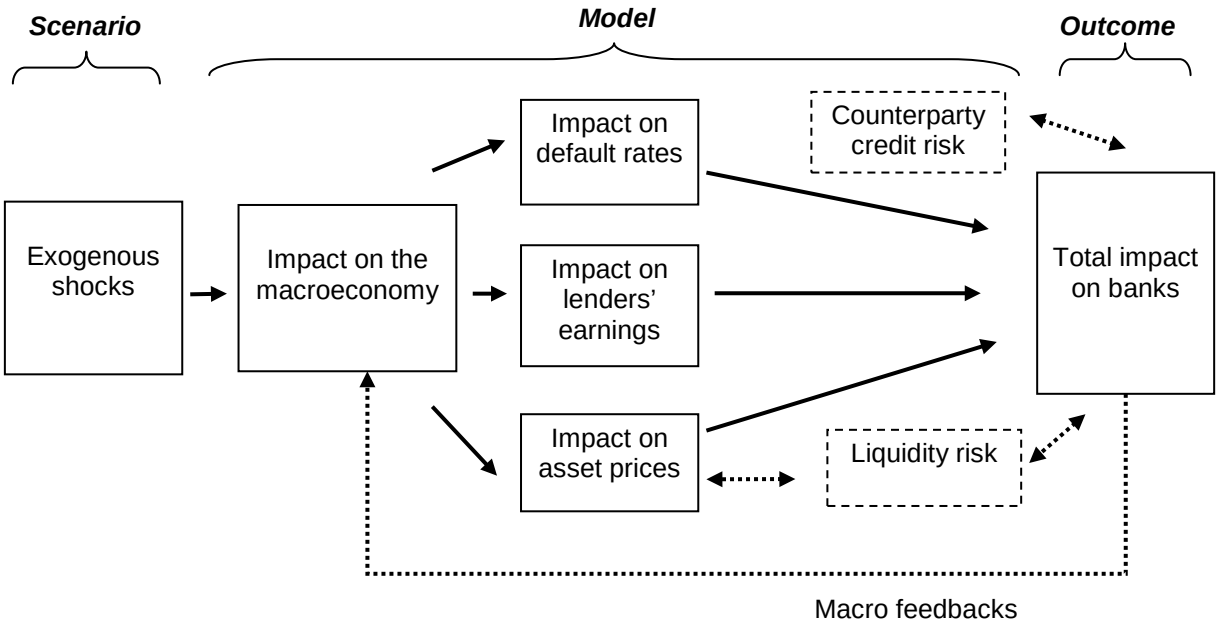
Any stress test starts with the set of exogenous shocks that capture the scenario.[5] As defined so far, best practice calls for "severe yet plausible" scenarios: severe enough to be meaningful yet plausible enough to be taken seriously (eg Quagliarello (2009)). Beyond this, the objective of the exercise largely determines the choice.

[5] In the literature, "scenarios" can describe two different things: (i) the set of exogenous shocks or (ii) the set of exogenous shocks together with their impact on the macroeconomy, as captured by the model. Analytically, the former is a cleaner approach, while for communication purposes the latter may be more convenient.

When the objective is to support crisis management or resolution, the key risks are often apparent. For instance, if the crisis has originated in exposures to property markets, it is natural to stress them further.

Graph 1

The structure of macro stress tests: Schematic overview[1]



[1] Schematic overview of the structure of the current macro stress tests that seek to evaluate the strength of banks. Bold lines represent the components captured by the majority of stress tests; dotted lines indicate the feedback effects that only the more sophisticated versions are able to capture, and even then only partially.

When the objective is to uncover vulnerabilities in tranquil times, scenario design becomes more difficult. There are two types of approaches. The first type is to rely directly on history. One may replicate specific historical episodes. Undoubtedly the 2008 crisis will become a future standard, just as the stock market crash of 1987 and the financial turbulence in the autumn of 1998 already are. Alternatively, one may draw shocks from the tail of the historical distribution of specific risk factors. The second type of approach is to use judgement to avoid the risk of relying excessively on the past. In this case, one may run hypothetical scenarios or else try to identify the shocks that would cause most damage to the system ("reverse stress tests"). That said, in the end plausibility is often judged according to historical experience.

Given their focus, macro stress test scenarios generally consider weak macroeconomic conditions. Typical scenarios that were run ahead of the crisis included severe drops in property prices, sharp adjustments to exchange rates or severe and sustained recessions (eg IMF (2005) and CGFS (2005)).

The "model" that maps scenarios into outcomes is, in fact, a process that involves a variety of steps and tools. The process may be top-down or bottom-up, or a combination of the two. In the bottom-up case, a central authority provides individual institutions with a common scenario, the institutions use their own models to estimate the impact of the shocks on their performance, and the central authority then aggregates the results. In the top-down case, the central authority does not involve individual banks directly but relies on its own internal model(s) to produce the results, possibly on the basis of detailed position data. In practice, many IMF and national stress tests have combined both processes, as was the case for the

Supervisory Capital Assessment Program (SCAP) in the United States (Board of Governors of the Federal Reserve System (2009)) and the tests carried out by the European Banking Authority in 2011 (EBA (2011)).

As this analysis suggests, macro stress tests generally rely on more than one technical tool or "model" in the narrow sense. Typically, only some of the building blocks shown in Graph 1 are integrated into a single such tool, but the flow from shocks to impact is generally similar. As a rule, the initial component is a macro model that provides estimates of how the exogenous shocks affect the economy. Because standard macro models do not include the variables relevant to the assessment of risks on banks' balance sheets, the outputs of the macro model are fed into auxiliary models that do incorporate them. Examples include models for the default rates of borrowers (for credit risk), for a broader range of asset classes (for market risk) and for the future earnings of banks (for income risk). These then determine what are sometimes called the "fundamental" losses in the stress scenario. The analysis often stops here. The more sophisticated stress tests also seek to assess the size of various potential feedback effects (dotted lines in Graph 1). By now, tools for the treatment of counterparty credit risk in the interbank market are reasonably well developed, albeit still rather mechanical. By contrast, the modelling of liquidity and macro feedbacks is at a much more preliminary stage.

The last component of a stress test is a measure of the outcome, which captures the final impact of the shocks on banks' balance sheets and income statements. The most common metrics are portfolio losses or capital and, less frequently, liquidity adequacy. Other metrics include the number of defaults or the size of capital injections needed to recapitalise the system.[6] If stress tests are used as tools for crisis resolution, the outcomes are typically set in terms of the amount of capital required to restore adequate strength.

A key question for any stress test is the horizon over which the impact of the shocks on banks' balance sheets is assessed, ie the forecast horizon. In a very influential contribution, Elsinger et al (2006) choose one quarter. They do so because their model – as most others – does not allow for behavioural reactions: in particular, banks are assumed not to restructure their portfolios in the stressed environment. Over such a short horizon, the assumption is more easily justified. Usefully, the current standard is a longer two-to-three year horizon, as it is otherwise nearly impossible to produce severe losses, given the lag structures embedded in most models. Some models allow for the possibility that banks adjust their balance sheets in response to the shocks, although so far only through mechanical rules of thumb.

The foregoing discussion suggests a number of general observations about the properties of the models.

First, as practised today, macro stress tests are still largely partial equilibrium exercises. As pointed out by Summer (2007), the model structure is rooted in the quantitative risk management framework that underpins the risk management models used by banks for business and regulatory purposes (McNeil et al (2005)). In such a setup, it is assumed that the evolution of the value of a given set of exposures is driven by a set of exogenous systematic risk factors.

Such a framework does not allow for feedbacks, even though these lie at the heart of financial instability. Disruptive spirals between market and funding liquidity risk played a crucial role in spreading distress after the Lehman failure (eg Brunnermeier (2009), Gorton (2009)), just as in previous episodes (eg Borio (2003)). And policymakers are equally concerned about credit crunch effects, through which the banks' tightening of credit terms in response to losses can weaken the economy. There is a consensus that stress tests should

[6] For an overview of different measures, see Cihak (2007).

capture such feedback effects. But doing so in practice has proved very difficult. Not least, these feedbacks depend on market participants' behavioural responses that are exceedingly hard to model.

Second, models are likely to be mis-specified econometrically. For one, just as with old-style macro models, hundreds of separate equations are sometimes estimated to try to cover all relevant aspects. The risk of modelling errors in such a setup is very high. More importantly, most models are estimated as if the true world behaved in a log-linear fashion, when in fact it does not. Non-linearities are at the heart of stress episodes (eg Drehmann et al (2007), Juselius and Kim (2011)). If the interest lies in studying the impact of small shocks around the equilibrium, linear approximations can be justified.[7] But they cannot capture effectively the dislocations caused by financial distress. Unsurprisingly, linear models tend to show signs of structural breaks at those times (Alfaro and Drehmann (2009)). And even if non-linear estimation methods are employed, the degree of statistical confidence in the results is exceedingly limited: the relevant episodes are very rare and the data available generally poor.[8] Hence, models tend to perform worst precisely in the conditions stress testing is designed to capture.

These properties of stress tests have significant implications for what we can and cannot expect stress tests to do.

Proposition 2: Beware of macro stress tests as early warning devices

To our knowledge, no macro stress test carried out ahead of the crisis identified the build-up of vulnerabilities. The message was overwhelmingly: "The system is sound". Rather than being part of the solution, stress tests turned out to be part of the problem. They lulled policymakers and market participants into a false sense of security. There is a serious risk that, unless their limitations are fully understood, they will continue to do so in the future.

Two sets of limitations stand out. The first relates to the technical aspects of the approach, ie the "model" used to simulate financial distress. The second relates to the broader context in which the stress tests are run.

Consider the models first. The previous analysis clearly indicates that the current generation of models is a long way from providing a realistic picture of the dynamics of financial distress. The models can hardly capture the non-linearities involved. The approach is still largely a partial equilibrium one: to the extent feedback effects are included at all, they are rather weak. All this shifts the burden of producing any damage from the properties of the models to the size of the shocks, which end up being "unreasonably" large. Market participants complained loudly that the crisis was generating twenty-plus standard deviation moves in asset prices and outcomes. But this was not, to put it mildly, an accurate reflection of the rarity of the event: as some observers have already pointed out, it would be tantamount to saying that those events are unique in the life span of the universe (Haldane (2009)). Rather, it reflected serious shortcomings in the models, for both micro and macro stress tests. No matter how hard one shook the box, little would drop out.

More generally, the models are the antithesis of what financial instability is all about (Borio and Drehmann (2011)). The very essence of financial instability is that *normal-size shocks cause the system to break down*. An unstable financial system is a fragile financial system; it

---

[7] A linear model can always be interpreted as a first-order Taylor-series approximation to the true, possibly non-linear data generating process.

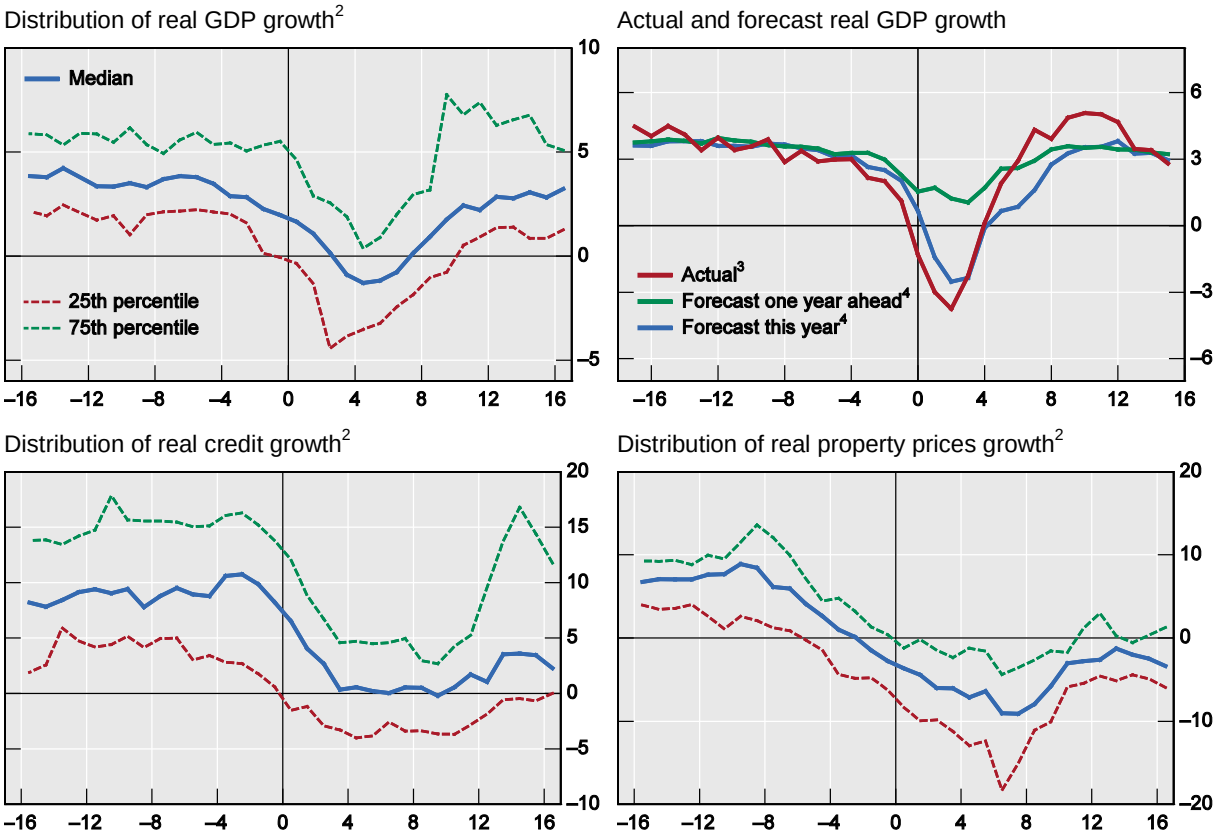
[8] Interestingly, stress tests generally report no error bands around their forecasts but only provide point estimates of the impact of the shock on banks' balance sheets. Quite apart from how rare crises are, very large standard errors will be present whenever the model involves a large number of estimated relationships.

is not one that would break down only if hit by severe macroeconomic shocks. And yet this is typically what stress tests need to assume.

Not surprisingly, empirical evidence is inconsistent with the implicit assumption of macro stress tests that crises occur as a result of unusually large negative shocks (Graph 2). As shown by Alfaro and Drehmann (2009), financial crises generally do not begin *after* output has collapsed, but *before* it contracts significantly. This is shown in Graph 2, which traces the average evolution of real GDP, actual and forecasted, around 43 banking crises in 30 countries (top panels). Moreover, on average, real property prices have not fallen substantially at that point (lower right-hand panel) and, partly as a consequence, credit growth is still well in positive territory (lower left-hand panel).

Graph 2

The evolution of GDP, credit and property price growth around crises[1]



[1] The horizontal axis depicts plus\minus 16 quarters around a crisis, which is indicated by the vertical line. [2] Distributions are based on a large set of crises from 1960 to the present. For details see Drehmann et al (2011a). [3] Average real GDP growth for the crises for which forecasts are available. [4] Average consensus forecasts for real GDP growth for the crises for which forecasts are available, see Alfaro and Drehmann (2009).

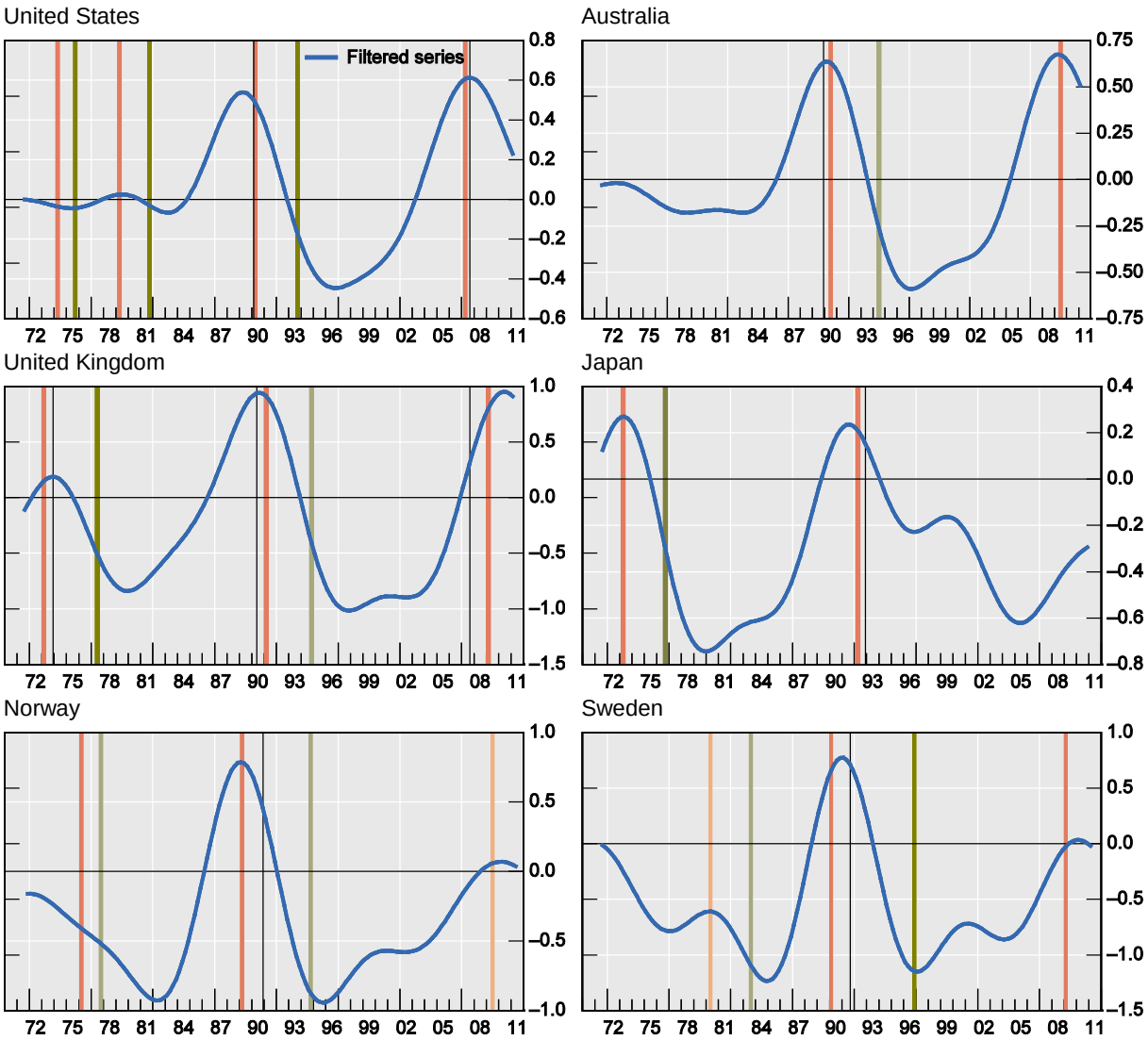
Sources: IMF; OECD; Consensus Economics; national data; BIS calculations.

Confirming this picture, recent work suggests that crises tend to begin at the *peak* of the medium-term financial cycle, not during the depth of the *bust* (Drehmann et al (2011b)). Graph 3 illustrates this for six countries. The graph shows that the systemic banking crises

(black vertical lines) coincided with the peak of the medium-term financial cycle, captured by the joint behaviour of credit and property prices.[9] This is true regardless of whether those peaks are estimated through turning-point methods (brown vertical lines) or frequency-based statistical filters (blue lines).

Graph 3

The financial cycle



Note: The graph is based on Drehmann et al (2011b). That paper empirically characterises the financial cycle using information from the evolution of property prices and credit in a given economy. It draws on two methods – turning-point and frequency-based filter analysis – to identify distinct medium-term financial cycles that are considerably longer than traditional business cycles. Pink and green bars (light pink and light green, if they are only weakly identified) indicate peaks and troughs of the cycle using the turning-point method. The frequency-based cycle (blue line) shows the results for the frequency-based filters. Black vertical lines indicate the starting point for banking crises, which in some cases (United Kingdom 1976 and United States 2007) are hardly visible as they coincide with a peak in the cycle.

Source: Drehmann, Borio and Tsatsaronis (2011b).

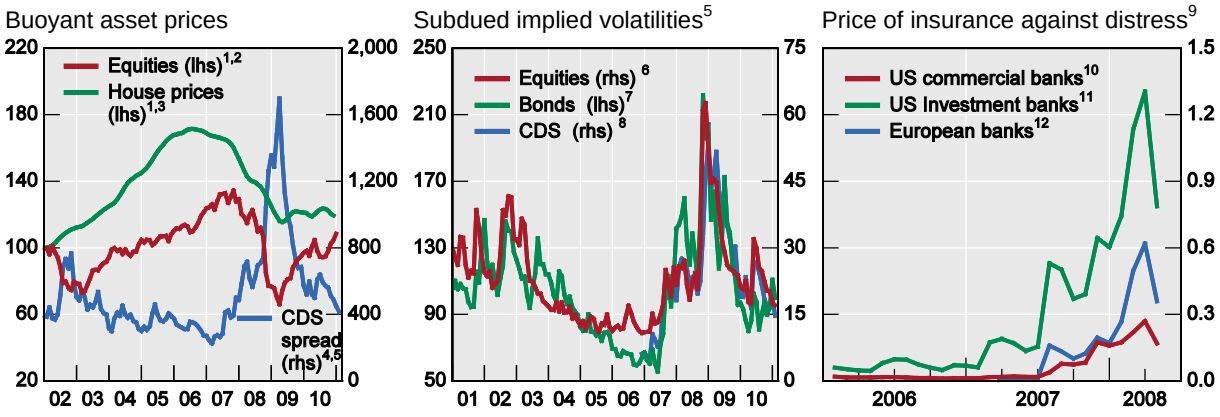
[9] Drehmann et al (2011b) seek to obtain a parsimonious empirical measure of the financial cycle. They analyse a broad range of indicators in seven countries from 1960 to the present. Using correlations at different frequencies and past crises as reference points, they conclude that financial cycles are medium-term phenomena and that they are best characterised by the joint behaviour of credit and property prices.

Next, consider the context. The key concept here is what one might call the "paradox of financial instability" (Borio and Drehmann (2011)): the system looks strongest precisely when it is most vulnerable. Credit growth and asset prices are unusually strong, leverage measured at market prices artificially low, profits and asset quality especially healthy, risk premia and volatilities unusually low precisely when risk is highest. What looks like low risk is, in fact, a sign of aggressive risk-taking. Graph 4 illustrates this point based on the behaviour of market prices during the run-up to the crisis in the United States (left-hand and centre panels). This perverse behaviour infects more formal measures of systemic risks that use market prices, including correlations, such as the implied price of insurance against a systemic event (right-hand side panel). Clearly, these measures were unusually subdued ahead of the crisis and showed signs of trouble only once overt financial market stress emerged in mid-2007. Indeed, ahead of the crisis the most common question was: "where has risk gone?"; no one could find it, regardless of where one looked (Knight (2007)).

Graph 4

Footprints of the paradox of financial instability

The US example



[1] End 2001 = 100. [2] S&P 500. [3] S&P Case-Shiller index, 20 cities. [4] Five-year on-the-run CDX.NA.HY 100 spread. [5] In basis points. [6] VIX index (implied volatility on S&P 500). [7] MOVE index (implied volatility on treasury options). [8] Implied volatility on the five-year-on-the-run CDX.NA.HY 100 spread. [9] In per cent, based on CDS spreads. Risk-neutral expectation of credit losses that equal or exceed 15% of the corresponding segments' combined liabilities in 2006 (per unit of exposure to these liabilities); risk-neutral expectations comprise expectations of actual losses and attitudes towards risk. Taken from Tarashev and Zhu (2008). [10] Ten banks headquartered in the United States. [11] Eight banks headquartered in the United States. [12] Sixteen universal banks headquartered in Europe.

Sources: Bankscope, Bloomberg, Datastream; JP Morgan, Markit; Tarashev and Zhu (2008), author's calculations.

Moreover, the temptation to argue that "things are different this time", that risks have disappeared, is especially strong when, as is typically the case, these booms go hand-in-hand with rapid financial innovation (eg Reinhart and Rogoff (2009)). Financial innovation holds out the promise of a much better management of the risks and, at the same time, stacks the deck against disproving this proposition. By construction, no historical data exist for new products and extrapolating from the performance of similar ones can severely underestimate risks (Box 2).

All this means that macro stress testing faces an uphill struggle. Technically, the size of the shock has to be very large to get any action *in the model*, regardless of initial conditions in the system. And, compounding the problem, those *initial conditions*, both in terms of balance sheets and earnings capacity, appear unusually strong prior to the crisis. No wonder the macro stress tests carried out ahead of the crisis did not identify serious vulnerabilities. Behaviourally, even if the stress tests were successful in pointing to potential vulnerabilities, it would be hard to have participants take them seriously. The tests would be run precisely

when hubris was at its highest and prudence at its lowest.[10] For instance, who could have taken seriously a test that assumed that the spread between interbank rates and overnight index swap would rise to more than 300 basis points, as it did during the crisis, when the spread had generally fluctuated between 10 and 15 basis points and, in "efficient markets", should be expected to approach zero?

The bottom line is simple. The fact that (macro) stress tests lulled policymakers and market participants into a false sense of security in the run-up to the recent crisis was not happenstance. It was an accident waiting to happen. We consider below what improvements could be made and whether they might be sufficient to overcome the limitations inherent in the approach. But one thing is certain: as devices to identify vulnerabilities in tranquil times, stress tests have a huge challenge ahead. The deck is stacked against them.

Box 2

Financial liberalisation and innovation – a key problem for stress tests[1]

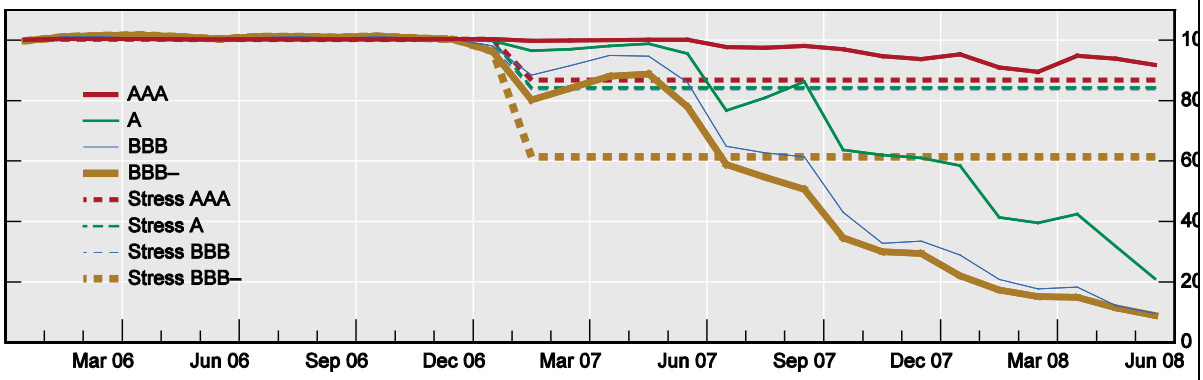
All stress tests – like all models – rely on historical data to estimate empirical relationships. Given typical econometric techniques, these models reflect average past relationships among the data series, rather than how the series interact under stress. Their reliance on past data also means that these models are not well suited to capturing innovations or changes in market structure. And yet, innovations – be they financial, such as structured credit products, or "real", such as the invention of railways – are often at the centre of the build-up of financial imbalances and the following distress.[2] Similarly, it is not uncommon for financial liberalisation episodes to trigger a boom that may prove unsustainable while at the same time changing the characteristics of the economy.

As always, assumptions are necessary to stress test new products. It is common practice to approximate the characteristics of new products by those of others for which historical information is available. This process involves potential pitfalls, which can result in a severe underestimation of risk.

Graph B 2.1

Stress testing new products[1]

A simple test that proxies ABS with corporate bonds[2]



[1] Solid lines: actual market prices for ABS index from JP Morgan for January 2006. One vintage for different ratings. Dotted lines: impact of the hypothetical stress test for different ratings. Impact for BBB ratings worse than for A, but hard to distinguish in the graph. [2] ABS tranches are assumed to behave like bonds of the same rating category. Stress test scenario starts in February 2007. An unspecified shock is assumed to lead to defaults in each rating category equal to the highest default rates ever observed for corporate bonds in that category. In addition, non-defaulted exposures experience a drop in prices which is three times the worst annual return on corporate bond indices for the various ratings over the period 1990 to beginning of 2007.

Sources: JPMorgan Chase; BIS calculations.

[10] Many observers point to weak scenario design as an important factor explaining the poor performance of stress tests before the crisis (eg Ong and Cihak (2010)). This is true but scenario design will always be difficult especially in good times, given the context (see below for a further discussion).

To illustrate this point, we carry out a micro stress test for a portfolio of asset-backed securities (ABS) exposures, following a procedure that was not uncommon prior to the crisis. The typical assumption was to proxy the default characteristics of ABS by those of corporate bonds of the same rating category. Based on this assumption, we implement a severe stress test scenario starting in February 2007.[3] An unspecified shock is assumed to lead to defaults in each rating category equal to the highest default rates ever observed for corporate bonds in that category. Essentially, these are default rates from the Great Depression. In addition, non-defaulted exposures experience a drop in prices which is three times the worst annual return on corporate bond indices for the various ratings over the period 1990 to the beginning of 2007.

Only for AAA ratings is the outcome of this stress test worse than actual developments, while the impact for all other categories is much more benign. Admittedly, more appropriate pricing models would have fared better. But to replicate actual price developments, given the typical assumptions used at the time, extreme scenarios would have been needed – something which would have easily been dismissed as implausible.

This is the typical conclusion reached at the time. The UBS report to its shareholders highlights this point (UBS (2008)). Given the evolution of historical data for super senior CDO tranches, the report notes that stress tests carried out ahead of the crisis concluded that the range of losses was so small that the bank could have protected itself with only partial hedges or even none at all. As it turned out, actual losses were so high that UBS needed state funds to survive the crisis.

A more general point is apparent from eyeballing the graph. By definition, only limited data are available for new products and none of those would be taken from a crisis. Understanding the "true" statistical properties is therefore difficult, if not impossible, from an ex-ante perspective. Arguably, measurement models built on these statistical relationships will break down in precisely the scenarios that they aim to capture beforehand – a problem that applies to many financial time series more generally (Danielsson (2008)).

---

[1] This box draws on Borio and Drehmann (2011). [2] Thakor (2011) shows that banks can have strong incentives to provide loans for which there is not sufficient data to assess risks fully. While this fosters innovation, it increases the risk of crises substantially. [3] Historical prices are based on the ABX index from JP Morgan for the January 2006 vintage for different ratings. The treatment of correlations is crucial for the pricing and evolution of structured credit products (eg Fender et al (2008)). This stress test implements a very simplistic correlation structure: it assumes that defaults occur independently but that price changes are fully correlated.

Proposition 3: Macro stress tests can greatly help in crisis management/resolution…

For much the same reasons, macro stress tests can be more effective as tools in crisis management and resolution. Here, the deck is stacked in their favour, or at least not so obviously against them. The crisis has already erupted. Sources of vulnerability have become apparent,[11] as have strong non-linear behavioural responses. Initial conditions are weak: the macroeconomy is suffering, balance-sheet quality has overtly deteriorated and financial institutions are incurring losses. Hubris has given way to prudence. The balance of power has shifted from business areas to risk controllers, and from the financial industry to the official sector.

In such an environment, the technical shortcomings of the tests are less of an issue. It is easier to identify relevant scenarios. It is easier to take them seriously. And the system does not need to be shaken so hard to reveal weaknesses.

---

[11] Alfaro and Drehmann (2009) show that the vast majority of stress scenarios based on historical data fall short of actual events unless macroeconomic conditions are already weak.

This is true regardless of the specific objective of the test, which varies somewhat with the stage of the crisis. One possibility, as highlighted by Greenlaw et al (2011), is to identify how much capital should be injected into the overall system to prevent a credit crunch, as, for instance, was done in the United States in 2009. This is most appropriate in the early stages of financial distress. Another is to weed out weak institutions from strong ones, with a view to resolving those that have no future – which is closer in spirit to what was done in Japan in 2004. This objective is more compelling once institutions are closer to bankruptcy.

That said, the distinction between these two objectives is not clear-cut. If the stress tests are tough enough, as they should be, in most cases one would expect a mixture of outcomes. And to embark on the exercises with a specific objective in mind, eg raising capital regardless of the underlying conditions of the banks, would risk prejudging the final result. Financial crises tend to be preceded by unusually strong credit and asset price booms. These booms leave in their wake bloated balance sheets and an overhang of debt. Cleaning up balance sheets is a precondition for balance-sheet repair.[12] Raising or preserving capital, by itself, cannot do this. Indeed, unless accompanied by determined attempts to enforce losses, it may even exacerbate the excess capacity that typically prevails in the financial system in such circumstances and that undermines its functioning (Borio et al (2010)). The ultimate objective should be to ensure that the financial system is healthy, so that it does not *artificially* constrain or *misallocate* the supply of credit. This, in turn, would help re-establish the basis for the system's long-run sustainable profitability.

Proposition 4: …and their additional benefits should not be underestimated….

Whether employed primarily as tools to uncover vulnerabilities in tranquil times or to support crisis management and resolution, macro stress tests can yield benefits that go beyond the promotion of those objectives narrowly defined. The benefits derive from the way that stress tests can help to discipline and structure thinking about financial stability among the many parties involved (the "stakeholders"). They can help to inform and reconcile different perspectives. They provide an indispensable common language and reference point.

Stakeholders come from very different backgrounds; they all have a partial and compartmentalised view of the world. Macroeconomists have grown accustomed to working with models that do not feature financial institutions and in which most financial variables beyond interest rates play, at most, a peripheral role (eg Woodford (2003)). Finance specialists have a tendency to consider the macro economy – if they consider it at all – as setting background conditions, and rarely as a factor influenced by financial conditions and the decisions of financial firms. Risk managers are accustomed to think in terms of risk factors, with only a vague mapping onto explicit macroeconomic and financial variables (McNeil et al (2005)). Loan officers have tended to focus on individual loans and customers at the risk of losing sight of the macroeconomic conditions that underlie their common performance (eg scoring models in retail lending, such as FICO scores; Frankel (2006)). And prudential supervisors have historically tended to focus very much on the riskiness and health of institutions on a standalone basis, ie from a "microprudential" standpoint (eg Borio (2011), Brunnermeier et al (2009)).

Macro stress testing provides a natural platform to reconcile these widely different perspectives. In fact, it *forces* their reconciliation. Macroeconomists are forced to incorporate financial institutions in their thinking. Finance specialists and risk managers are forced to map macroeconomic variables onto the anonymous risk factors that drive the performance of their risk models. Loan officers are forced to take greater account of macroeconomic

[12] On the Japanese experience, see eg Caballero et al (2008).

conditions. Prudential supervisors are forced to take a more systemic or system-wide perspective. And the final results are discussed at the highest levels of the institutions involved.

There is little doubt that the macro stress-testing exercises carried out so far have helped to edge the various stakeholders towards a better, albeit still limited, understanding of the nature of financial stability. In the process, they have also helped improve the availability and use of valuable historical data, which would otherwise have been disregarded, thrown away or never collected in the first place (eg property prices, interbank exposures, credit registry data). It is all too easy to forget how poor the starting point was and to underestimate the progress made so far.

If these benefits are cumulative and accrue only over time, as stakeholders become increasingly familiar with the common language, others are more tightly linked to individual exercises. Comparing bottom-up with top-down outcomes can improve the dialogue about the risks faced in a specific situation. More importantly, it can help to improve banks' own stress-testing practices,[13] validate their models and assess their risk management systems, exposing obvious outliers in cross-sectional comparisons or a generalised tendency to underestimate risks. And aggregation can also reveal obvious inconsistencies in individual results, such as the violation of adding-up constraints. Typical examples include firms that, in response to the shock, report inconsistent increases in market shares, aggregate improvements in earnings beyond reasonable historical experience, or reactions that are hard to reconcile with the assumed changes in market prices.

Proposition 5: …but if you do them, do them right!

Whatever the inherent limitations of macro stress tests, their benefits depend on the way they are structured and carried out. We next explore a number of preconditions for their usefulness and then suggest areas in which payoffs from improvements appear highest.

Elements of good practice[14]

We see three key elements of good practice: having the will to really stress the system; ensuring buy-in by all the stakeholders; and entertaining a clear follow-up plan in line with the specific objective of the exercise. We consider each of these elements in turn.

The current technical limitations of macro stress tests put a premium on a *strong will to stress the system*. This calls for conservative estimates of the mapping of shocks onto losses and for severe scenarios as a means to overcome those limitations.

Conservative estimates can be achieved in at least two complementary ways. One is judgmental adjustments to statistical estimates that, almost inevitably, rely too much on data drawn from tranquil times. Another is through the cross-checking of outputs from different models, including by combining top-down and bottom-up approaches. For example, Hirtle et al (2009) note that the adoption of multiple and independent approaches to come up with loss and revenue estimates added credibility to the SCAP stress tests.

Especially severe scenarios would include seemingly unrealistic shocks to asset prices and macro variables and, ideally, the protracted evaporation of funding and market liquidity. Such

[13] There is anecdotal evidence that some banks enhanced their stress testing capabilities following the SCAP and have started to use these tools for their own risk management.

[14] For a set of good practice principles for micro stress tests, see BCBS (2009).

scenarios should not be overly constrained by historical experience.[15] To be sure, the choice of such severe scenarios partly runs against the general advice that they should be "severe yet plausible". No doubt, all else being equal, plausibility facilitates buy-in. But, in our view, the current shortcomings of the models leave no choice. The risk and cost of fostering a false sense of security are simply too high. Fully recognising the shortcomings of current technology is a precondition for success and a better basis for buy-in.

Moreover, the plausibility of the scenario is arguably irrelevant in two specific cases. One is when the macro stress test is used as a cross-check for banks' own internal models. What matters is whether under extreme conditions the model produces plausibly extreme losses. The other is when the test is designed to find the breaking point of the system, a variant of reverse-stress testing.[16] While common in engineering, to our knowledge such tests have not as yet been employed for financial stability purposes. In this case, *if approached with a healthy degree of* scepticism, they can also shed some light on the shortcomings of the macro stress testing model itself. They can act as reality checks, whenever the breaking point appears unrealistically far away.

The *buy-in of all stakeholders* is critical. Buy-in is a precondition for the commitment of time and resources on the part of the various stakeholders and for follow-up. It is especially precious given the substantial role of judgement in the exercise. Governance matters a lot here. Clarity in the objectives and in the assignments of responsibilities and attention to incentives can help. This applies both to the official authorities and to the private sector participants. When multiple authorities are involved, their perspective can differ and misunderstandings about the specific purpose of the exercise can arise. This is true both within national jurisdictions, where tensions between micro- and macro-prudential perspectives can emerge, and across them, where incentive problems are naturally more prominent. For the private sector, concerns about the follow-up inevitably loom large, as the tests can result in the need to strengthen capital and liquidity buffers or even in more intrusive forms of intervention. Market participants tend to see the main value added as coming from the greater information they receive as a direct result of their involvement. This can be an analysis of system-wide risks, enhanced data on aggregate positions of financial firms, or feedback on the performance of their risk models. That said, the ultimate benefit, which is harder to perceive, comes from the follow-up that should make the system more stable.

A clear *follow-up* plan, tailored to the specific objective of the exercise, is essential. For example, if the primary objective is to uncover vulnerabilities in tranquil times, the authorities should always entertain the option of taking targeted action to strengthen the system's defences, possibly through firm-level intervention. If the primary objective is to support crisis management and resolution, system-wide public-sector liquidity and capital backstops are critical. Without them, no exercise can be credible. Moreover, the suspicion would be irresistible that the test was not ambitious enough precisely in order to justify subsequent inaction. And the risk of undermining market confidence would be all too real. In addition, specific design features should be carefully calibrated. In particular, in order to limit the risk of an unintended credit crunch, capital targets should be set as absolute amounts rather than as ratios to assets or risk-weighted assets (Greenlaw et al (2011)).

---

[15] The experience of UBS is instructive in this regard, even though it relates to a micro stress test. The losses incurred by UBS during the crisis were so severe that they prompted the intervention of the Swiss authorities. As the report to shareholders acknowledges (UBS (2008)), stress tests – and risk management more broadly – failed as they relied exclusively on historical data, which excluded severe stress in the US housing market.

[16] As suggested to us by Anil Kashyap, a possible starting point for reverse stress tests could be banks' own analysis of "death threats".

In any follow-up, communication issues figure prominently. It is not possible to do justice to them in the space available. How much to communicate, in what form, and to whom, are perennial, exceedingly tough questions. The answers will again partly depend on the nature of the exercise and the context, including the broader communication strategy for financial stability policy. We would argue, however, that on balance the bias has generally been on the side of communicating too little rather than too much. Indeed, the positive reaction to the disclosure of greater information about individual firms in the latest stress tests in the United States and Europe is encouraging in this respect.[17] Provided the exercise is done well, public communication can be an important tool in restraining hubris during booms and instilling confidence during busts.

Two sometimes underappreciated risks deserve specific attention in this context. One, more pertinent during booms, is what we would call "risk-spotting fatigue". Since, as noted, the build-up of financial imbalances takes many years, frequent exercises may be counterproductive, because too little changes from one to the next. This can undermine both the support for tests and their credibility. Even correct messages pointing to the build-up of vulnerabilities could be called into question, including by those in charge of the tests. The other risk, more pertinent during busts, is embarking on the exercise with the objective of showing to the markets that the situation is not as bad as they think. The authorities should always approach the tests with an open mind and be seen to do so.

Finally, the communication strategy and the interaction between supervisors and banks should be designed so as not to undermine the effectiveness of banks' own stress testing strategies. For one, there is a risk that the scenarios are perceived as *the* key vulnerabilities, crowding out more bank-specific micro stress tests. In addition, tests under supervisory guidance could also turn into box-ticking exercises rather than being used as creative risk management tools. Policymakers should make it clear that macro stress tests cover only a sub-set of relevant scenarios and should avoid linking regulatory requirements to the results of voluntary micro stress tests run by banks.

A way forward

Looking forward, which areas deserve special attention and which ones may be less likely to provide high payoffs? We consider, in turn, the use of complementary information from reduced-form real-time leading indicators of financial distress; the universe of institutions included in the tests; the relative treatment of common exposures and bilateral interlinkages; and the balance between bottom-up and top-down approaches.

The use of *complementary information* from leading indicators of financial distress can help constrain the limitations of macro stress tests as early warning devices in seemingly tranquil times. The recent literature suggests that it is possible to develop reduced-form real-time indicators that provide a fairly reliable signal of systemic financial distress a few years ahead, even out of sample (eg Alessi and Detken (2009), Borio and Drehmann (2009)). One such variant relies on the *joint* deviation of the ratio of credit-to-GDP and asset prices, notably property prices, from historical trends. These indicators seek to exploit the paradox of financial instability to their advantage: they interpret unusually exuberant behaviour in financial quantities and prices as signs of fragility rather than strength. They seek to distinguish sustainable from unsustainable booms. And, to do so, they focus on the most systematic and general signs of the build-up of risks across policy regimes and historical

[17] In fact, providing sufficient information for market participants to carry out their own stress tests may be a solution in cases where the authorities wish to avoid sending the wrong signal. The markets positively received the decision by the European authorities to provide more detailed information about individual bank sovereign exposures rather than just pick a specific scenario.

periods – they focus, that is, on what is common to the various episodes, rather on what differs across them.

This information could support macro stress tests in various ways. Generally speaking, when these indicators flashed yellow or red, policymakers could increase the severity of the tests. They could, for instance, increase the size of the shocks and severity of the scenarios. They could tighten the scrutiny of the models and outcomes. And since, by construction, the reduced-form indicators can at best provide a rather general sense of the build-up of risks, they could follow up with more targeted assessments of pressure points, partly on the basis of the macro stress tests themselves.

There is scope to improve the selection of the *universe of institutions* subject to macro stress tests. One way of doing this, as suggested by Greenwald et al (2011), is to extend it beyond banks to cover a larger portion of the financial system. That said, probably an even higher priority is to extend stress tests beyond national borders. An exclusively national focus, assessing one national system at a time, sits uneasily with an increasingly global financial system. The recent financial crisis has reminded us that financial distress does not stop at national borders. To be sure, confidentiality issues loom large: the experience of the stress tests in the European Union highlights the difficulties that exist even in comparatively closely integrated regions. But, over time, those difficulties could be overcome. One could then adjust the set of institutions included in the exercise to suit the specific scenario under consideration. Another, complementary, possibility would be to run macro stress tests on the most important global financial institutions, such as the so-called Global Systemically Important Financial Institutions (G-SIFIs), (BCBS (2011)).

The relative treatment of *common (similar) exposures and bilateral interlinkages* bears close watching. To our mind, too much attention is being paid to bilateral interlinkages and network analysis. True, this information can be very helpful in understanding the geography of the financial system. It is also necessary to estimate meaningful balance-sheet measures of sectoral or aggregate leverage: the capital available to absorb losses in any given sector is overstated unless interlinkages within the sector are taken into account (eg the well known "double leverage" phenomenon). As such, it may cast some light on the tail of the distributions (eg Drehmann and Tarashev (2011)). And it can be helpful in crisis management, as long as it is very detailed and available in real time. But it is very unlikely to yield *substantial* benefits in the context of macro stress tests. Common exposures of institutions, on both their asset and liability sides, together with undiscriminating responses by investors and counterparties, are the main drivers of the dynamics of financial distress. A financial crisis acts like a tsunami that sweeps away all before it, not like a force that knocks down one domino after another along a specific path. Considerable empirical evidence points in this direction.[18]

Achieving the right *balance between top-down and bottom-up approaches* is not easy. Both have merits and should probably be used simultaneously. As already noted, they can act as a useful cross-check for each other, foster communication and help reconcile perspectives. That said, we remain sceptical of approaches that seek to aggregate individual reaction functions with a view to measuring systemic risk or eliciting information about endogenous responses through iterative procedures (eg Brunnermeier et al (2010), Duffie (2011)). This is so regardless of whether the reaction functions are estimated from the data or reflect survey responses (CGFS (2005)). Estimation is exceedingly hard given the challenges involved (limited number of relevant data points, instability across episodes etc); and the responses to surveys should be taken with more than a pinch of salt, given the incentives to misreport and

[18] See Elsinger et al (2006) for empirical evidence on this point; see Upper (2007) for a critical survey of contagion analysis based on networks.

the shortcomings of the firms' models.[19] Moreover, even if both types of information could be taken at face value, it is hard to imagine that one could develop a reliable iterative mapping between responses and outcomes. The cost-benefit balance does not appear to be particularly attractive.[20]

Conclusion

Macro stress tests are set to become a core element of the macroprudential frameworks being put in place across the globe. As offspring of the (micro) stress tests carried out by individual financial institutions, their ascendancy has gone unchallenged. And yet, stress tests failed spectacularly when they were needed most: none of them helped to detect the vulnerabilities in the financial system ahead of the recent financial crisis.

In this paper we have argued that it is important to understand what stress tests can and cannot do. We should not set expectations unrealistically high. Ironically, macro stress tests are best suited to crisis management and resolution; currently, they are not reliable, in our view, for identifying vulnerabilities in seemingly tranquil times – the purpose for which they were originally designed. They can help, and have helped, discipline and improve the dialogue about financial stability vulnerabilities; but, unless properly interpreted, they risk taking that dialogue astray. They can help, and have helped, spot shortcomings in our models of systemic risk and financial crises; but they have so far largely done so because of what they have *failed* to produce (crisis warnings), rather than for what they have produced (comforting outcomes).

We have discussed ways to improve the performance of macro stress tests. From a technical perspective, it is well recognised by now that generating more realistic non-linearities and feedback effects is a priority. We remain sceptical, however, of attempts that see the secret of success in modelling network effects or the iterative bottom-up aggregation of individual responses.

From a broader perspective, process and governance are critical. We have suggested that the severity of the scenarios could be increased based on the signals from reduced-form leading indicators of financial distress, such as those based on unusually strong cumulative increases in credit and asset prices. Those signals could also be used as a trigger for more specific drill-down risk assessments, in which stress tests could play a part. We have also suggested that a more global focus, rather than a jurisdiction-by-jurisdiction approach, would be helpful. And we have argued that a focus on common exposures is more promising than one on interlinkages. Ultimately, however, improvements in the performance of stress tests depend on a change in mindset. No stress test can succeed unless there is a strong will to stress the system hard and to distrust rosy results.

And here lies the problem. The importance of the right mindset has been appreciated ever since the inception of stress tests (eg CRMPG (1999)). But this proved to be no check on the generalised hubris that prevailed before the recent crisis among market participants and policymakers alike. Will it be any different next time?

---

[19] Bottom-up stress tests can only provide useful insights if banks' internal models can capture the relevant risks. This cannot be taken for granted, as highlighted by the report to UBS shareholders (2008). UBS only partly hedged its super senior CDO tranches, which turned out to be the major source of its losses: historical data indicated that the partial hedges were sufficient to fully protect the bank from any losses. From the outset, internal models netted these exposures to zero. Hence, even if the actual crisis had been run as a scenario, such a stress test would not have uncovered any vulnerabilities.

[20] See CGFS (2000) for an early analysis of the aggregation of stress tests.

References

Aikman, D, P Alessandri, B Eklund, P Gai, S Kapadia, E Martin, N Mora, G Stern and M Willison (2009): "Funding liquidity risk in a quantitative model of systemic stability", *Bank of England Working Paper*, no 372.

Alessi, L and C Detken (2009): "Real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity", *ECB Working Paper,* no 1039.

Alfaro, R and M Drehmann (2009): "Macro stress tests and crises: What can we learn?", *BIS Quarterly Review*, December, pp 29–41.

Basel Committee on Banking Supervision (BCBS) (2009): *Principles for sound stress testing practices and supervision*.

_____ (2011): *Global systemically important banks: Assessment methodology and the additional loss absorbency requirement*, November.

Barnhill, T and L Schumacher (2011): "Modeling correlated systemic liquidity and solvency risk in a financial environment with incomplete information", *IMF Working Paper*, WP/11/263.

Blaschke, W, M Jones, G Majnoni and S Peria (2001): "Stress testing of financial systems: An overview of issues, methodologies, and FSAP experiences", *IMF Working Papers*, 01/88.

Board of Governors of the Federal Reserve System (2009): *The supervisory capital assessment program: Design and implementation*, April.

Borio, C (2003): "Market distress and vanishing liquidity: Anatomy and policy options", in A Persaud (ed), *Liquidity black holes: understanding, quantifying and managing financial liquidity risk*, Risk Publications, December. Also available (revised and updated) as *BIS Working Papers*, no 158, 2004.

_____ (2011): "Implementing a macroprudential framework: Blending boldness and realism", *Capitalism and Society*, vol 6(1), January, pp 1–23. An earlier version is also available at http://www.bis.org/repofficepubl/hkimr201007.12c.htm.

Borio, C and M Drehmann (2009): "Assessing the risk of banking crises – revisited", *BIS Quarterly Review*, March, pp 29–46.

_____ (2011): "Towards an operational framework for financial stability: 'Fuzzy' measurement and its consequences", in *Financial stability, monetary policy and central banking*, Central Bank of Chile. Also available as *BIS Working Papers*, no 284, June 2009.

Borio, C, B Vale and G von Peter (2010): "Resolving the financial crisis: Are we heeding the lessons from the Nordics?", *Moneda y Crédito,* 230, pp 7–47. Also available as *BIS Working Papers*, no 311, July 2010.

Boss, M, G Fenz, G Krenn, J Pann, C Puhr, T Scheiber, S W Schmitz, M Schneider and E Ubl (2008): "Stress tests for the Austrian FSAP update 2007: Methodology, scenarios and results", *OENB Financial Stability Report,* 15.

Brunnermeier, M (2009): "Deciphering the liquidity and credit crunch 2007–2008", *Journal of Economic Perspectives*, vol 23(1), pp 77–100.

Brunnermeier, M, A Crockett, C Goodhart, M Hellwig, A Persaud and H Shin (2009): "The fundamental principles of financial regulation", *Geneva Reports on the World Economy*, no 11.

Brunnermeier, M, G Gorton and A Krishnamurthy (2010): "Risk topography", mimeograph.

Bunn, P, A Cunningham and M Drehmann (2005): "Stress testing as a tool for assessing systemic risk", *Bank of England Financial Stability Review*, June.

Caballero, R, T Hoshi and A Kashyap (2008): "Zombie lending and depressed restructuring in Japan", *American Economic Review*, vol 98, pp 1943–77.

Castren, O, S Dees and F Zaher (2008): "Global macro-financial shocks and corporate sector expected default frequencies in the euro area", *ECB Working Paper*, no 875.

Cihak, M (2007): "Introduction to applied stress testing", *IMF Working Paper*, 07/59.

Committee on the Global Financial System (CGFS) (2000): *Stress testing by large financial institutions: Current practice and aggregation issues*, no 14, April.

_____ (2005): *Stress testing at major financial institutions: Survey results and practice*, no 24.

Counterparty Risk Management Policy Group (CRMPG) (1999): *Improving counterparty risk management practices*, June.

Christiano, L, R Motto and M Rostagno (2010): "Financial factors in economic fluctuations", *ECB Working Paper*, no 1192.

Danielsson, J. (2008): "Blame the models", *Journal of Financial Stability,* no 4, pp 321–8.

De Graeve, F, T Kick and M Koetter (2007): "Monetary policy and financial (in)stability: An integrated micro-macro approach", mimeo.

Drehmann, M (2009): "Macroeconomic stress testing banks: A survey of methodologies" in M Quagliariello (ed), *Stress testing the banking system: Methodologies and applications,* Cambridge University Press.

Drehmann, M, C Borio and K Tsatsaronis (2011a): "Anchoring countercyclical capital buffers: The role of credit aggregates", *International Journal of Central Banking,* 7(4).

_____ (2011b): "Characterising the financial cycle: don't lose sight of the medium term!", paper presented at the Reserve Bank of Chicago-ECB 14th Annual International Banking Conference, *The role of central banks in financial stability: How has it changed?*, Chicago, 10–11 November.

Drehmann, M, A Patton and S Sorensen (2007): "Non-linearities and stress testing", in *Risk measurement and systemic risk, Proceedings of the fourth joint central bank research conference, ECB*.

Drehmann, M, S Sorensen and M Stringa (2010): "The integrated impact of credit and interest rate risk on banks: A dynamic framework and stress testing application", *Journal of Banking & Finance,* 34, pp 735–51.

Drehmann, M and N Tarashev (2011): "Systemic importance: Some simple indicators", *BIS Quarterly Review*, March, pp 25–37.

Duellmann, K and M Erdelmeier (2009): "Crash testing German banks", *International Journal of Central Banking,* 5, pp 139–75.

Duffie, D (2011): "Systemic risk exposures: A 10-by-10-by-10 approach", Stanford University, mimeo.

European Banking Authority (EBA) (2011): *Overview of the EBA 2011 banking EU-wide stress test.*

Elsinger, H, A Lehar and M Summer (2006): "Risk assessment for banking systems", *Management Science*, vol 52(9), September, pp 1301–41.

Fender, I, N Tarashev and H Zhu (2008): "Credit fundamentals, ratings and value-at-risk: CDOs versus corporate exposures", *BIS Quarterly Review*, March, pp 87–101.

Foglia, A (2008): "Stress testing credit risk: a survey of authorities' approaches", *Bank of Italy Occasional Papers*, no 37.

Frankel, A (2006): “Prime or not so prime? An exploration of US housing finance in the new century”, *BIS Quarterly Review*, March, pp 67–78.

FSB-IMF-BIS (2011): “Macroprudential policy tools and frameworks”, *Report to the G20*, March.

Gorton, G (2009): “Information, liquidity, and the (ongoing) panic of 2007”, *American Economic Review*, vol 99(2), pp 567–72.

Gray, D and A Jobst (2010): “New directions in financial sector and sovereign risk management”, *Journal of Investment Management*, vol 8, no 1.

Gray, D, R Merton and Z Bodie (2006): “A new framework for analyzing and managing macrofinancial risks of an economy”, *NBER Working Paper*, 12637.

Greenlaw, D, A Kashyap, K Schoenholtz and H Shin (2011): “Stressed out: macroprudential principles for stress testing”, paper presented at the US Monetary Policy Forum, New York, 25 February.

Haldane, A (2009): “Why banks failed the stress test”, speech at the Marcus-Evans conference on stress testing, 9–10 February.

Hirtle, B, T Schuermann and K Stiroh (2009): “Macroprudential supervision of financial institutions: lessons from the SCAP”, *Federal Reserve Bank of New York Staff Reports*, no 409.

IMF (2005): *Financial sector assessment: A handbook*.

Jacobsen, T, J Linde and K Roszbach (2005): “Exploring interactions between real activity and the financial stance”, *Journal of Financial Stability*, vol 1 (3), pp 308–41.

Juselius, M and M Kim (2011): “Sustainable financial obligations and crisis cycles”, *Helsinki Economic Centre of Research Discussion Papers*, 313.

Kapadia, S, M Drehmann, J Elliot and G Stern (2011): “Liquidity risk, cash flow constraints and systemic feedbacks”, in J Haubrich and A Lo (eds), *NBER publication on Quantifying Systemic Risk*, (forthcoming).

Knight, M (2007): “Now you see it, now you don’t: Risk in the small and in the large”, speech delivered at the Eighth Annual Risk Management Convention of the Global Association of Risk Professionals, 27–28 February. http://www.bis.org/speeches/sp070302.htm.

McNeil, A, R Frey and P Embrechts (2005): *Quantitative risk management: Concepts, techniques and tools,* Princeton Series in Finance, Princeton University Press.

Meh, C, and K. Moran (2008): “The role of bank capital in the propagation of shocks”, *Bank of Canada Working Papers*, 2008-36.

Merton, R (1974): “On the pricing of corporate debt: the risk structure of interest rates”, *Journal of Finance*, vol 29, pp 449–70

Ong, L, and M Cihak (2010): “Of runes and sagas: Perspectives on liquidity stress testing using an Iceland example”, *IMF Working Paper*, WP/10/156.

Ong, L, R Maino and N Duma (2010): “Into the great unknown: Stress testing in weak data”, *IMF Working Paper*, WP/10/282.

Pesaran, M, T Schuerman, B Treutler and S Weiner (2006): “Macroeconomic dynamics and credit risk: a global perspective”, *Journal of Money Credit and Banking*, vol 38(5), pp 1211–62.

Quagliarello, M (2009): *Stress testing the banking system: Methodologies and applications,* Cambridge University Press.

Reinhart, C and K Rogoff (2009): *This time is different: Eight centuries of financial folly*, Princeton University Press.

Segoviano, M and P Padilla (2007): “Portfolio credit risk and macroeconomic shocks: Applications to stress testing under data-restricted environments”, *IMF Working Papers*, no 06/283.

Summer, M (2007): “Modelling instability of banking systems and the problem of macro stress testing”, Paper presented at the ECB conference on “Simulating financial instability”, 12–13 July, Frankfurt.

Tarashev, N and H Zhu (2008): “Market perceptions of systemic risk in the banking industry”, *BIS Quarterly Review*, March, pp 6–8.

Thakor, A (2011): “Incentives to innovate and financial crises”, *Journal of Financial Economics,* 103, pp 130–48.

UBS (2008): *Shareholder Report on UBS’s Write-Downs*.

Van den End, W (2008): “Liquidity stress-tester: A macro model for stress-testing banks’ liquidity risk”, *DNB Working Paper*, no 175.

Woodford, M (2003): *Interest and prices: Foundations of a theory of monetary policy*, Princeton University Press.