Rodrigo Alfaro
ralfaro@bcentral.cl

Mathias Drehmann
mathias.drehmann@bis.org

Macro stress tests and crises: what can we learn?[1]

Few, if any, of the macro stress tests undertaken before the current crisis uncovered significant vulnerabilities. This article examines the reasons for the poor performance by comparing the outcomes of simple stress tests with actual events for a large sample of historical banking crises. The results highlight that the structural assumptions underlying stress testing models do not match output growth around many crises. Furthermore, unless macro conditions are already weak prior to the eruption of the crisis, the vast majority of stress scenarios based on historical data are not severe enough. Last, stress testing models are not robust, as statistical relationships tend to break down during crises. These insights have important implications for the design and conduct of stress tests in the future.

JEL classification: E44, G01, G17.

The current crisis has underlined the importance of complementing the microprudential approach to regulation and supervision with a macroprudential perspective. One important issue is how to measure vulnerabilities and risks on a system-wide level.[2] Macro stress tests are seen as a promising tool. Central banks and the IMF had made extensive use of stress tests prior to the crisis, but generally without identifying significant vulnerabilities. For example, over a third of the countries considered in this article published macro stress testing results as part of an IMF Financial Sector Assessment Program in 2005, 2006 or the first half of 2007. The overwhelming majority concluded that their banking systems were robust even in the face of very severe adverse scenarios.[3] To be sure, not all of these countries subsequently experienced a full-blown banking crisis. But it is remarkable that not more warning flags were raised.

[1] This article was written when Rodrigo Alfaro was a visiting research fellow at the BIS. The authors would like to thank Claudio Borio, Steve Cecchetti, Jean-Michel Mahler, Bob McCauley, Tariq Siddique, Kostas Tsatsaronis and Christian Upper for helpful comments, and Jakub Demski for excellent research assistance. The views expressed are the authors' own and not necessarily those of the BIS or the Central Bank of Chile.

[2] Borio and Drehmann (2009) provide a detailed discussion of how to operationalise frameworks for financial stability in the face of measurement challenges.

[3] In all the studies, the IMF was very careful to highlight potential shortcomings of the stress testing models used.

Why? Several reasons have been suggested: stress scenarios were not severe enough; important risks were missed; and feedback effects within the financial sector as well as between the real economy and the financial sector were ignored. No doubt, all these reasons are valid, and addressing them has sparked an ambitious research agenda that is likely to result in large and complex stress testing models.

In this article, however, we take a step back and ask the why question slightly differently. In particular, we assess – within a very simple framework – three fundamental requirements that any stress test should fulfil to be informative: (i) it should use the correct model to capture the potential unfolding of crises in a realistic, yet stylised, fashion; (ii) scenarios should represent a severe event that ex ante is not beyond the realm of possibility; and (iii) models should be robust, particularly during the crisis periods they aim to simulate.

Disentangling fully which of these requirements, if any, may not have been met prior to crises is not feasible, even ex post. In particular, assessing whether the model used was correct in all respects is an impossibility. But, taking a bird's-eye view, we can provide some indications about the performance of stylised stress testing models around historical banking crises and analyse each of the key requirements in turn. We find that models may not be correct, insofar as the underlying structural assumptions do not match output growth around many crises. We also show that, unless macro conditions are already weak prior to the eruption of the crisis, the vast majority of stress scenarios based on historical data are not severe enough in comparison with actual events. Last, our results raise doubts about the robustness of the models, as many of our simple stress testing models break down during the ensuing crisis. This raises interesting and fundamental questions for future stress testing practices, which are discussed in the concluding section.

Can stress testing models simulate crises in a realistic fashion?

Macro stress testing models can differ significantly in terms of complexity and the risks considered (for an overview, see Drehmann (2009)). However, they all share a similar structure rooted in the quantitative risk management framework. This is the same structure that underpins banks' own risk management and stress testing models (Summer (2007)).

A standard macro stress testing model is built in a modular fashion (Figure 1). The stress simulation itself begins with a scenario. But at the heart of the model are a set of exposures that are captured by the analysis. These are often the credit risk exposures of a bank or a banking system in a specific country. More advanced macro stress tests also incorporate market risk or counterparty credit risk in the interbank market. A module then identifies a set of systematic risk factors and models their impact on the analysed exposures, for example with a market or credit risk model. The majority of macro stress tests assume that only domestic macroeconomic variables are systematic risk factors. Therefore, they use as another module some variant of a structural or reduced-form macroeconomic model to capture the impact of the stress

Schematic structure of current stress testing models

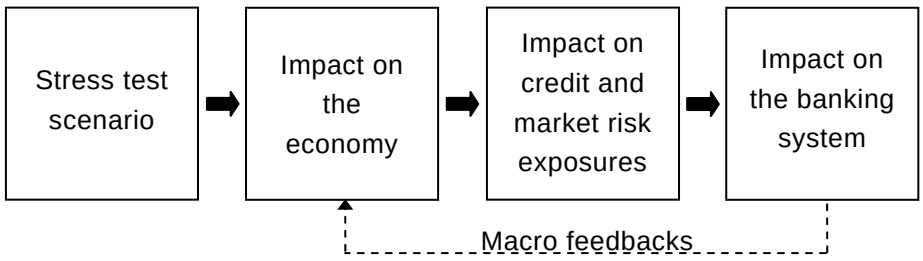


Figure 1

scenario on the economy. Given the state of macro modelling more generally, few stress testing models incorporate feedbacks from the financial sector to the real economy, and those that do tend to be very reduced-form.

The key structural assumption that macroeconomic shocks drive crises ...

In short, current macro stress testing models assume that negative shocks to domestic macroeconomic factors drive stress events in the banking system. If this is a realistic description of how crises unfold, we should observe that domestic macroeconomic conditions weaken significantly ahead of banking crises. We assess whether this is true by looking at 43 banking crises in 30 countries, starting in 1974 and including the latest episodes.[4] And we assume that actual and expected real GDP growth are good summary indicators of broader macroeconomic conditions.

... reflects average behaviour ...

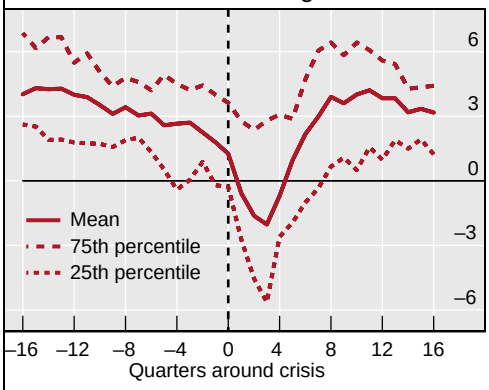
The average evolution of real GDP growth 16 quarters before and after the start of crises (denoted as 0 in the left-hand panel of Graph 1) suggests that the structural assumption may be justified. Average real GDP growth is above 4% four years prior to a crisis. It then starts to decline, with a marked drop one year ahead of the event. Once the crisis materialises, average GDP growth drops to –2% three quarters later. The recovery is V-shaped, and on average the economy returns to it pre-crisis growth path two years after the event. Average expected real GDP growth follows a similar pattern (Graph 1,

[4] Historical banking crises are based on Laeven and Valencia (2008) and Reinhart and Rogoff (2008). From their large sample we exclude all crises where no quarterly GDP data are available at least 10 quarters prior to the crisis. In addition, we do not consider transition economies, given large apparent structural changes in those economies. To avoid overlaps in our analysis of pre- and post-crisis data, we also exclude the 1994 crisis in Brazil, which materialised less than four years after the previous episode. The following crises are included in the sample with the starting quarter in brackets: Argentina (Q4 1989, Q1 1995, Q4 2001), Australia (Q4 1989), Belgium (Q3 2008), Brazil (Q1 1990), Canada (Q4 1983), Denmark (Q4 1987), Finland (Q3 1991), France (Q1 1994, Q3 2008), Germany (Q3 2007), Iceland (Q4 1985, Q4 1993, Q3 2008), Indonesia (Q4 1997), Ireland (Q3 2008), Italy (Q3 1990), Japan (Q4 1997, Q3 2008), Korea (Q3 1997), Malaysia (Q3 1997), Mexico (Q4 1994), the Netherlands (Q3 2008), New Zealand (Q1 1987), Norway (Q4 1991), the Philippines (Q4 1983, Q3 1997), Singapore (Q4 1982), South Africa (Q4 1977), Spain (Q4 1977, Q4 1993, Q3 2008), Sweden (Q3 1991), Switzerland (Q4 2007), Thailand (Q3 1997), Turkey (Q4 2000), the United Kingdom (Q4 1974, Q4 1991, Q2 2007) and the United States (Q4 1988, Q3 2007).

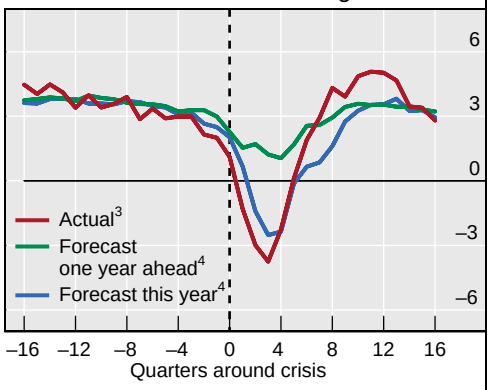
Real GDP growth around crises[1]

In per cent

Distribution of real GDP growth[2]



Actual and forecast real GDP growth



[1] The mean and the percentiles are based on the historical dispersion of real GDP growth taken at the specific quarter across all crises in the sample. The vertical line indicates the starting point of crises, normalised to 0. [2] Distribution is based on all crises in the sample. [3] Average real GDP growth across crises, where forecasts are available. [4] Average consensus forecasts for real GDP growth across crises, where forecasts are available.

Sources: IMF; OECD; © Consensus Economics; national data; BIS calculations. Graph 1

right-hand panel).[5] However, it appears that, on average prior to crises, consensus forecasts overestimate GDP growth. But once crises materialise, the average forecasts for the current year tend to underestimate the initial drop in output as well as the speed of the recovery. This aligns well with the current experience.

Two important caveats are worth highlighting. First, the timing of crises is not always unambiguous. We rely on Laeven and Valencia (2008) and Reinhart and Rogoff (2008), who define the beginning of crises by the emergence of large-scale policy assistance or the default of important players in the financial system. However, whether this method pinpoints the exact starting date is unclear. Boyd et al (2009), for example, show that stress often materialised beforehand. For example, Laeven and Valencia's (2008) approach dates the beginning of the current crises in Belgium, Iceland and Ireland after the bankruptcy of Lehman Brothers. An alternative starting date could be the first emergence of strains in global interbank markets in August 2007. If anything, such a dating would strengthen some of the messages in this paper.[6]

… which conceals many exceptions

Second, looking at averages conceals the fact that domestic macroeconomic conditions remained rather robust around the beginning of a large fraction of banking crises. Some of the cross-crisis differences are apparent from the 25th and 75th percentiles of the distribution of real GDP growth (Graph 1, left-hand panel). The experience during the recent crisis is

[5] The sample of crises for which we observe consensus forecasts is smaller than our full sample. The consensus indicator is not as good a leading indicator as we would wish. The forecast for the current (next) year always refers to the current (next) calendar year independently of the quarter.

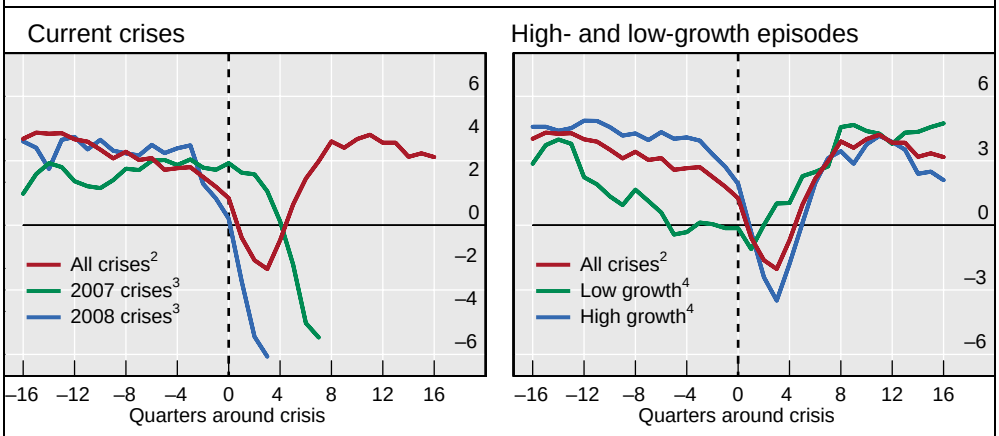
[6] For example, both the shape and the level of average GDP growth for the 2007 and 2008 crisis countries would have been almost identical in Graph 2 (left-hand panel).

even more illustrative. In the left-hand panel of Graph 2, we show the average across all crises as well as the average real GDP growth for the countries where the banking crises began in 2007 (2007 crises) and the countries which experienced systemic banking strains only after the collapse of Lehman Brothers (2008 crises). Up to one year ahead of the crisis, average real GDP growth for the latter group is broadly in line with the evolution of average real GDP growth for all crisis episodes.[7] Real GDP growth then drops significantly and falls deep into negative territory after the start of the crisis. Given limited data, we can obviously not show the recovery. The profile of average real GDP growth of the four 2007 crisis countries is significantly different. Until half a year after the crisis erupted, real GDP growth fluctuates around 2.5%. Only then does it begin to drop sharply, and it remains negative even two years after the event.

Overall, only just over one half of the crises in our sample were preceded by adverse macroeconomic conditions.[8] We define this broadly as a sharp drop in real GDP growth immediately prior to the crisis, or high macroeconomic volatility or low average growth in the previous three years.[9] For example, it seems that the structural assumptions about the pattern of GDP growth in stress testing models are justified for crises that are preceded by several years

Real GDP growth around crises for different subsamples[1]

In per cent



[1] The mean is based on the historical dispersion of real GDP growth taken at the specific quarter across all crises in the specified subsample. The vertical line indicates the starting point of crises, normalised to 0. [2] Average real GDP growth across all crises. [3] Average real GDP growth across crises which started in 2007 and 2008, respectively. [4] Average real GDP growth across crises, preceded by low (high) growth, defined as annual GDP growth being less than (more than) 2% on average in the three years prior to the crisis.

Sources: IMF; OECD; © Consensus Economics; national data. Graph 2

[7] This holds even if the average is calculated without the 2007 and 2008 crises.

[8] If crises were dated earlier, this share would be even lower.

[9] We always take a three-year horizon prior to the beginning of crises. Crises with high volatility are episodes where the standard deviation in output growth in this period is larger than the total average. A significant drop in GDP growth is defined as GDP growth dropping by more than 3 percentage points year on year. Low average growth is defined as average annual GDP growth in this period of less than 2%.

of low growth (Graph 2, right-hand panel). By contrast, for high-growth episodes, GDP growth is nearly 3% only two quarters prior to the beginning of the crisis.

Many crises are not preceded by weak macroeconomic conditions

Our finding that a large fraction of banking crises is not preceded by weak domestic macroeconomic conditions shows that current stress testing models are not able to replicate the dynamics of many past crises. This could be a result of stress tests considering the wrong risk factors and missing those which were the actual drivers of crises. For example, for many countries the current crisis was driven not by domestic exposures but by large shocks to foreign assets. Another explanation is that the underlying model structure is wrong and crises are not simply a result of large negative shocks to exogenous risk factors.

This dichotomy is already reflected in the literature on banking crises: one strand argues that banking crises are driven by shocks to fundamentals (eg Gorton (1988)). In other models, crises can emerge even when conditions are good. The classic panic-based bank run models (eg Diamond and Dybvig (1983)) are possibly the most obvious example. The same holds true for another intellectual tradition, which sees financial distress as the result of the build-up in risk-taking over time, owing to self-reinforcing feedback mechanisms within the financial system and between this and the real economy (Minsky (1982) and Kindleberger (1996)). In these models, the actual trigger for the crisis may be exceedingly small and unobservable (eg a change in mood). This contrasts strongly with current stress testing models, which require large shocks to generate a large impact. By construction, the endogenous build-up of vulnerabilities is also not possible with current stress testing models.

Different theoretical models have different implications for how stress testing models should be enhanced. We will discuss this below. For now, we remain within the assumed structure of stress testing models and assess whether it is possible to construct severe, yet plausible, scenarios.

Can we construct severe yet plausible scenarios?

Even though the preceding discussion showed that for a large fraction of crises the structural assumptions underlying stress testing models may not be met, output drops substantially in nearly all of our observed crises once stress emerges. A pragmatic approach to stress testing could use this as a starting point to construct scenarios, independently of whether falls in output truly reflect or cause crises.

Deriving scenarios from historical data …

The standard rule for scenario selection says that stress scenarios should be severe yet plausible (eg Quagliariello (2009)). But what does this mean in practice? Often scenario construction is guided by history in that either scenarios simply replicate historical stress events or shocks to risk factors are expressed in terms of high multiples of standard deviations of the historical distribution. For example, the shocks used for the UK IMF Financial Sector Assessment Program broadly corresponded to events three standard deviations away from the mean of a particular variable. The statistical

distributions, in turn, were based on the error variance of the Bank of England's macro model (Hoggarth and Whitley (2003)). It is also common practice to consider hypothetical scenarios designed to address current concerns without being constrained by the past. However, even then history provides a reference framework for judging the plausibility of the event: it is hard to argue that something is plausible if it is beyond the realm of anything that has been experienced.

We implement a historical approach to scenario selection in a highly stylised fashion to assess whether severe yet plausible scenarios can be constructed. We continue to consider real GDP growth as the main risk factor in our hypothetical stress testing models. And we assume that it depends only on its own past behaviour and random shocks.[10] To replicate the information available to policymakers before crises, we estimate a different model for each crisis in each country, using only data up to the crisis itself. For example, we observe two crises in the United States: the savings and loan crisis in the late 1980s and the current episode. Hence, one model is estimated with US data up to 1988, whilst the second model includes all information, including the past crisis, up to 2007.

As stress scenario, we use the worst negative forecast error of our crisis-specific models, regardless of whether this coincided with a banking crisis or not. We shock our models with these scenarios four quarters before the beginning of the crisis and compare the maximum drop in GDP growth during the stress test with the maximum drop during the actual episode.[11] This provides a rough benchmark to assess whether, based on information available before crises, a severe yet plausible scenario can be constructed. If so, we should find that the stress test we simulate is at least as severe as actual developments.

... may under-estimate vulnerabilities

We find strong evidence that a historical perspective does not always provide the right framework for scenario construction. In nearly 70% of all cases, the hypothetical stress scenarios fall short of the severity of actual events (Table 1). Interestingly, for none of the 11 countries that have experienced a banking crisis after 2007 do our stress tests anticipate the severe drop in GDP growth, even though several of these economies had previously experienced crises.

However, stress tests seem to be a useful tool to gauge the potential impact of further adverse shocks if macro conditions are already weak. In 64% of all low-growth episodes, stress scenarios are severe enough. This contrasts starkly with high-growth episodes, where in over 80% of all crises a stress test could not have generated the actual sharp decline in GDP growth.

---

[10] Our stress testing models are simple autoregressive processes. Based on econometric selection criteria, we choose either an AR(1) or an AR(2) model as the best specification for each crisis. It is interesting to note that simple autoregressive models often outperform more complex ones in terms of forecast performance (eg Clements and Hendry (1998)).

[11] The maximum drop in GDP growth during the crises is calculated as the difference between GDP growth four quarters prior to the crises and the minimum GDP growth two years after the crises.

Comparison of the impact of stress tests with actual events[1]			
	Number of crises	Stress test less severe than actual events[2]	Stress test more severe than actual events[2]
All crises	43	67%	33%
Previous crises[3]	32	56%	44%
Current crises[3]	11	100%	0%
High growth[4]	29	83%	17%
Low growth[4]	14	36%	64%

[1] Comparison of the maximum drop in GDP growth during the stress test with the maximum drop during the actual episode. The stress scenario is the worst negative forecast error of our crisis-specific models. [2] Percentage of crises in each category. [3] Current crises are all crises which started in 2007 or 2008. Previous crises are those occurring before. [4] Crises which were preceded by low (high) growth, defined as annual GDP growth of less than (more than) 2% on average in the three years prior to the crisis.

Sources: IMF; OECD; national data. Table 1

Some of our results may be due to the inability of our simple model to capture macroeconomic feedbacks. Graphs 1 and 2 indicate that negative feedback spirals from the financial sector to the real economy seem to emerge during crises, as average real GDP growth drops sharply after crises began. On the other hand, we are also unable to capture the impact of policy actions implemented to contain the negative effects of banking crises. In both regards, our model is as limited as the vast majority of macro stress testing models currently in use.

It is also possible that our results are driven by the pre-crisis data sample, even though we use all the data that are readily available. We could take a longer-term perspective. For example, Haldane (2009) shows for the United Kingdom that the current crisis is not out of the ordinary in comparison with a historical perspective going back as far as 1693 for equity prices or 1857 for GDP growth. However, swings in output and the stock market index are very large if judged against the 10 years preceding current events.

The drawback of taking such a long-term perspective is that it ignores structural change. How could we assume that the economy has not evolved since the time when the United States was still a colony? Did the IT revolution not transform the interrelations within the banking system more recently? Clearly, the world is constantly changing. If models are not fully structural and parameters are not invariant to change, the estimated statistical relationships should be expected to change over time as well. From a stress testing perspective, it is especially important to ask whether such changes are likely to occur in an abrupt manner during crisis periods. If so, models will not be robust and the third requirement will be not fulfilled. We will explore this in the next section.

Are models robust during crises?

Model robustness is a crucial, but implicit assumption in any forecasting or simulation exercise. For stress testing models, it is generally assumed that the statistical relationships estimated prior to a crisis also describe the economy

adequately during a crisis. We use our simple model to test this implicit assumption for the crises in our sample.[12]

Statistical relationships tend to change during crises …

The results are discouraging: for 28 of our 43 crises (65%), the statistical relationships break down around the crisis date (Table 2). Models are particularly fragile after the beginning of the crisis.[13] And it seems that this is the case regardless of whether we look at high- and low-growth episodes or at current and previous crises. Even then, the current crisis stands out, as *all* our models experience a structural break after it started.

In principle, this result may be driven by the fact that we use the wrong model. But more realistic models are unlikely to fare better. We only look at a model with one variable. Cutting-edge stress testing models may have hundreds of equations, often estimated on an equation by equation basis. For the whole to be robust, we have to be confident that all equations are free of major structural breaks.

Both theory and the experience of past crises also make it very likely that reduced-form statistical models break down during crises. In most theoretical models, crises are associated with an abrupt change in the behaviour of economic agents. Independently of whether crises are assumed to be driven by fundamentals or not, these models imply that observables change suddenly and dramatically: depositors withdraw all their money, the interbank market freezes, banks ration credit, etc. And large public interventions are often the policy response prescribed by these models.

Structural breaks around crises

	Overall	Before[1]	During[1]	After[1]	Only before[1]	Only after[1]
All crises	28	18%	75%	82%	0%	25%
Previous crises[2]	18	22%	83%	72%	0%	17%
Current crises[2]	10	10%	60%	100%	0%	40%
High growth[3]	20	20%	80%	90%	0%	20%
Low growth[3]	8	13%	63%	63%	0%	38%

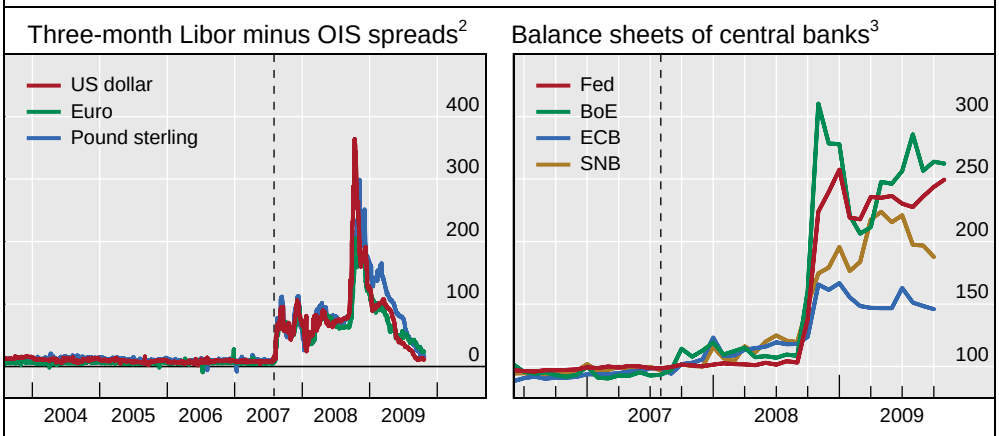
[1] We estimate each model inserting dummies one year before and up to two years after crises started. We define a structural break if dummies in four consecutive quarters are jointly significant. (Only) Before / During / (Only) After indicates whether dummies are significant (only) before / during / (only) after the beginning of the crisis. Percentage of structural breaks in each category. [2] Current crises are all crises which started in 2007 or 2008. Previous crises are those occurring before. [3] Crises which were preceded by low (high) growth, defined as annual GDP growth of less than (more than) 2% on average in the three years prior to the crisis.

Sources: IMF; OECD; national data; BIS calculations. Table 2

[12] Graphs 1 and 2 indicate that there is no permanent statistical break during crises as GDP growth recovers after several quarters. Therefore standard statistical break tests do not apply. Hence, we estimate each model using all available data including crisis periods and insert dummies one year before and up to two years after crises started. We define a structural break if dummies in four consecutive quarters are jointly significant (at the 10% level). This is a relatively strong test as it requires that the statistical relationships break down over a one-year horizon.

[13] Similar results have been found in the literature. Cecchetti et al (2009b) find that 50% of crises in their sample experience a structural break in the level or the trend in real GDP within one or two years of the beginning of the crisis.

Central bank balance sheets and spreads for the current crisis[1]



[1] The vertical line indicates August 2007 as the starting point of the current crisis. [2] In basis points. [3] Index, 2007 average = 100.

Sources: Bloomberg; Datastream; BIS calculations. Graph 3

… in ways that are impossible to predict

Looking at past crises, we also find that observables change drastically after a crisis, and often in ways which could not have been anticipated prior to the events. For example, before August 2007, the spread between interbank rates and overnight index swap rates had fluctuated between 10 and 15 basis points (Graph 3, left-hand panel). Ex post, it seems obvious that these spreads can widen dramatically. But ex ante, any hypothetical stress test that would have implied spreads climbing beyond 300 basis points would certainly not have passed the plausibility test. Policy also reacted in unforeseeable ways. Central banks around the globe undertook unprecedented policy operations. It would have been very hard to anticipate the degree of quantitative easing and, by implication, the ballooning of central banks' balance sheets (Graph 3, right-hand panel). Even more difficult would have been to foresee the effects of these policy interventions, as they are still not fully understood.

Stress tests and crises – what do we learn?

Assumptions underlying stress testing models are often not met

For this article, we undertook stress tests prior to past crises and compared results with actual outcomes. In particular, we examined the performance of three fundamental requirements which should be fulfilled for stress tests to provide useful information. First, the correct model should be used. But it is questionable whether the current modelling framework aligns well with observables around historical banking crises. In nearly 50% of the analysed crises, the evolution of GDP growth does not seem to be in line with the structural assumptions of current stress testing models. Second, the stress scenario should be severe yet plausible. But unless macro conditions are already weak prior to the eruption of the crisis, we show that the vast majority of stress scenarios based on historical data are not severe enough in comparison with actual events. Last, models should be robust. Our results also question whether this can be generally assumed as 64% of our simple models break down during the following crisis.

What do these findings imply for macro stress testing? For stress tests to be useful, the underlying structure has to be improved to better capture crisis dynamics. An important avenue of future research is to incorporate more risk factors, such as international interlinkages or non-macro factors. But it is also crucial that future stress testing models should be able to replicate endogenous cycles, which are often the underlying driver of crises. This is a major challenge as macro models more generally are currently far from being able to do so (eg Cecchetti et al (2009a)).

The impossibility of fully capturing all drivers of crises …

But there is a fundamental problem. Like any other model, stress testing models can only capture reality in a stylised fashion.[14] Model builders therefore have to make choices concerning what is essential, what can be represented in a reduced-form fashion and what can be ignored.[15] This is not an easy task: as Caballero and Kurlat (2009) point out, ex post we may well be able to understand how models failed, but ex ante this is different. For the current crises, for example, we would have had to fully identify the dangers of structured investment vehicles and structured products. However, the prevailing view at the time was that these innovations were, on balance, highly beneficial, as they would shift risk to those better able and willing to bear it.

… calls for great caution

For the foreseeable future, the challenges in modelling crises appropriately seem enormous. And as we have argued, it is doubtful that the statistical models will be free of structural breaks once crises emerge. As a consequence, it is likely that stress tests will continue to underestimate the risks to the economy, as they did prior to the current crisis. There is, therefore, a real danger that stress testing results will continue to lull users into a false sense of security (Borio and Drehmann (2009)). We suggest three practical steps to reduce this risk.[16]

First, model outputs should not be taken at face value and all results should be interpreted with great caution. It is important that this is understood by all users of the output, be they policymakers, commercial banks or the media. One way to highlight this problem would be to publish stress testing results with confidence intervals, as is often done for macro forecasts.

Stress tests as catalysts for further analysis

Second, macro stress tests should not be seen as the final output but as the starting point for an effective discussion about potential financial stability threats. The modelling challenges imply that meaningful stress testing exercises will have to involve discussions and judgments. As Bunn et al (2005)

---

[14] It is important to understand the ultimate objective of the model in order to make modelling choices (Drehmann (2008)). A model will never serve all objectives equally well, as model requirements can sometimes conflict. For example, the model with the highest forecast ability may not necessarily be the one which is most tractable and suited for story telling.

[15] The choices are often guided by history or banks' own risk management models and the risks these highlight. Relying on banks' own risk management models to identify risks raises an interesting conundrum. If banks' own stress testing models are useful, the results should feed into banks' capital and liquidity decisions (taken either voluntarily or through regulatory pressure) and thereby reduce the related risks.

[16] An important side benefit of stress testing is that repeated stress tests help to organise available data in a coherent and user-friendly fashion. The ready availability of this information can be highly valuable during crises or in addressing other policy questions.

have pointed out, one important value added of stress testing models is in providing a coherent framework within which to consider the implications of differing judgments, for example on how new financial products may change the dynamics of crises or how a vulnerability may crystallise in different ways. Taking account of judgments and a range of views across the organisation is also one of the key recommendations of the Basel Committee (BCBS (2009)) in its principles for sound stress testing practices.

Third, scenario design is critical. Regardless of how elaborate models or stress testing processes become, the outcome will always depend on judgments and the stress scenario. As we have shown in this article, a statistical approach to scenario selection will certainly fall short for many future crises. However, there is no easy answer to the question of how to do it best otherwise. An interesting starting point could be new products which grow rapidly, or business areas where banks make large profits. Historically, these areas could have identified some of the vulnerabilities in the run-up to crises. Scenario design will certainly require creative thinking and the courage to ask unusual questions because, as history has shown, once crises emerge we should expect the unexpected.

References

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

Borio, C and M Drehmann (2009): "Towards an operational framework for financial stability: 'fuzzy' measurement and its consequences", *BIS Working Papers*, no 284.

Boyd, J, G de Nicoló and E Loukoianova (2009): "Banking crises and crisis dating: theory and evidence", *IMF Working Paper*, WP/090/141.

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 and P Kurlat (2009): "The 'surprising' origin and nature of financial crises: a macroeconomic proposal", paper presented at the Jackson Hole Symposium on Financial Stability and Macroeconomic Policy, August.

Cecchetti, S, M Kohler and C Upper (2009a): "Financial crises and economic activity", paper presented at the Jackson Hole Symposium on Financial Stability and Macroeconomic Policy, August.

Cecchetti, S, P Disyatat and M Kohler (2009b): "Integrating financial stability: new models for a new challenge", paper prepared for the joint BIS-ECB workshop on *Monetary policy and financial stability*, Basel.

Clements, M and D Hendry (1998): *Forecasting economic time series*, Cambridge University Press.

Diamond, D and P Dybvig (1983): "Bank runs, deposit insurance and liquidity", *Journal of Political Economy*, no 91, pp 401–19.

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

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

Gorton, G (1988): "Banking panics and business cycles", *Oxford Economic Papers*, no 40, pp 751–81.

Haldane, A (2009): "Why banks failed the stress test", speech given at the Marcus-Evans Conference on Stress Testing, 9–10 February.

Hoggarth, G and J Whitley (2003): "Assessing the strength of UK banks through macroeconomic stress tests", *Bank of England Financial Stability Review*, June.

Kindleberger, C (1996): *Maniacs, panics and crashes*, Cambridge University Press.

Laeven, L and F Valencia (2008): "Systemic banking crises: a new database", *IMF Working Paper*, WP/08/224.

Minsky, H (1982): "Can it happen again?", in *Essays on Instability and Finance*, Armonk: M E Sharpe.

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

Reinhart, C and K Rogoff (2008): "Banking crises: an equal opportunity menace", *NBER Working Papers*, no 14587.

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,* Frankfurt.