Stress tests: Objectives, challenges and modelling choices

MATHIAS DREHMANN*

Abstract

Stress tests have become an integral tool for banks' risk management practices as well as for financial stability assessments by central banks. But there has been no debate about the objectives of stress tests, even though an understanding of those is essential when building and evaluating stress testing models. This paper identifies three main objectives: validation, decision making and communication. And it shows that different objectives lead to different and possibly conflicting priorities for model design. In the light of this discussion modelling choices are assessed and two main challenges for stress testing models are discussed: data limitations and the endogeneity of risk.

Keywords: stress tests, objectives, macro feedbacks, endogenous behaviour, liquidity risk, non-linearities

Introduction

The initiation of the Financial Stability Assessment Programmes by the International Monetary Fund (IMF) and the World Bank in 1999 brought stress tests to the forefront of financial stability modelling. Most major central banks in the world have now their own financial stability (FS) stress testing model or are in the process of developing one. Stress tests are also an integral part of banks' risk management practices. This widespread use of stress tests has lead to an emergence of a rough consensus about the model structure of stress tests. For FS stress tests there also seems a broad agreement that the value added from stress tests is derived

* The views expressed in this paper do not necessarily represent those of the Bank for International Settlements (BIS) but only those of the author. The bulk of the paper was written when on secondment to the European Central Bank (ECB) ECB and is based on a presentation for the Stress Testing Workshop at the Riksbank in June 2007, where participants provided helpful comments. I am also grateful for very detailed comments from Claus Puhr, Kasper Roszbach, Matthias Sydow, Iman van Lelyveld and Staffan Viotti which improved the paper considerably. Contact details: Mathias Drehmann, Bank for International Settlements, Centralbahnplatz 2, 4002 Basel, Switzerland, mathias.drehmann@bis.org.

from "an integrated forward looking perspective, a focus on the financial system as a whole and a uniform approach to the assessment of risk exposures across institutions" (p.3, IMF and World Bank, 2003). Therefore, a best practice guide to building stress testing models should be easily established.

This is not the case. As any other model, stress tests can only capture reality in a stylised fashion. Model builders therefore have to make choices on what is essential, what can be represented in a reduced form fashion and what can be ignored. To do this, it is necessary to understand the ultimate objective of the model. It is surprising that there has been little to no debate about this issue. In the first section of the paper I, therefore, provide an overview over different objectives. And I show how they translate into different model requirements which can sometimes be conflicting. In the light of this discussion, I explore how objectives shape modelling choices. One of the key objectives of FS stress tests is to capture the impact of severe (yet plausible) shocks on the whole financial system. So far, this has not been possible as modellers face two important challenges: data limitations and the endogeneity of risk. These issues are discussed in the third section of the paper, where I will also highlight possibilities to start addressing these challenges.

A first distinguishing factor when discussing stress testing objectives is whether they are run for *internal* or *external* purposes. This already implies different model requirements. Whereas *external* models have to focus on the target audience, *internal* models have to be understood and accepted by senior management. As the background and risk management culture differs across institution – especially when comparing commercial and central banks – stress testing models have to reflect these variations. For internal purposes, two broad objectives can be distinguished. The first is *validation*, when stress tests are used to assess for example the robustness of capital models. But, the ultimate internal objective is *decision making*. For private banks, stress testing results may feed into capital decision or business planning and central bank use them as input into the supervisory dialog or for assessing FS vulnerabilities. But for many non-supervisory central banks, the main objective is not internal but external *communication*.

Communication, decision making and *validation* all lead to different model requirements. For the latter two, model accuracy and forecast performance are essential. Whilst these characteristics are important they may not be overriding priorities for communication, which requires primarily that the model is transparent and suitable for storytelling. Unfortunately, transparency, the suitability for storytelling, model accuracy, forecast performance and other priorities cannot always be achieved equally well within the same model. For example, it is well known that simple models such as autoregressive specifications may even outperform the true model with respect to forecast performance (see Clements and Hendry, 1998). But, autoregressive specifications are certainly not granular enough for policy evaluation or communication. Understanding these trade-offs for different model specification is not easy and ultimately depends on the objective. This is not only important for modellers when designing the stress testing framework but also for policy makers. It is not enough to call for more stress testing but it is also important to know what should ultimately be achieved with the stress test results.

Notwithstanding that objectives are different for different stress tests current models share very much a common structure. As pointed out by Summer (2007) this structure is rooted in the quantitative risk management framework. The quantitative risk management framework can be characterised as a model chain starting with a specific shock to systematic risk factors (e.g. stock market returns or macroeconomic variables), followed by a modelled data generating process which captures how systematic risk drivers interact between each other and across time, and finally a model computing how systematic risk drivers impact onto the specific risk measure for exposures (e.g. value at risk). Without undertaking a full survey of the literature (for this see Sorge and Virolainen, 2006 or ECB, 2006) , the main theme throughout my discussion in Section 2 is that different model choices are available, but that data availability and the main objective are the key determinants in choosing an appropriate set-up at each point along the model chain.

However, any stress testing modeller will meet important challenges. Most of them are well understood (see ECB, 2006), but because of their complexity, limited progress has been made. In Section 3, I discuss the two biggest problems: data and the endogeneity of risk. I argue that data limitations imply that stress testing models are not always econometrically robust, something which should be reflected when communicating the results, e.g. by confidence intervals or by reporting results with different assumptions. That said stress tests can also be used to address data limitations. This is actually frequently done, for example, when modellers do not observe sufficient data to econometrically model the link between risk factors and outcome they may just set values to extreme levels. If the objective is validating the robustness of the capital model, this may be sufficient. If it is communication, it may not as such a stress test does not reveal key parts of the transmission from shock to impact.

It seems to me that when policy makers call for more stress testing they implicitly call for modelling risk endogenously. Endogenous risk is due to endogenous behavioural reactions by agents in the economy including the policy maker. Modelling endogenous behaviour is an important step in addressing issues currently not captured in any satisfactory fashion. For example, current stress tests assume that banks will be passive in light of the stress event and not change their exposures. And central banks are also assumed to act in a mechanical fashion. I discuss, further, how endogenous behaviour can lead to liquidity risk and macro feedbacks. Finally, the endogeneity of risk can imply non-linearities in the model of the data generating process, something often mentioned by policy makers without further specifying what it really means.

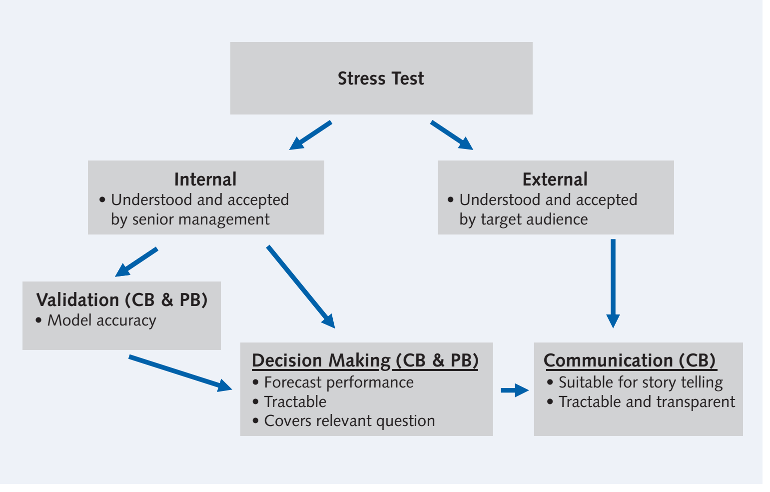
The aim of the Section 3 of this paper is not only to highlight these problems but also to point to possible solutions. But it is apparent that modelling the endogeneity of risk is the key challenge for improving stress testing models. The remainder of the article is structured as follows. Section 1 discusses the objectives of stress testing. Section 2 elaborates on the common structure of current stress testing models and explores how objectives shape modelling choices. Section 3 highlights important challenges and some possibilities to address them. Section 4 discusses the possible next steps and Section 5 concludes.

1 Objectives of stress tests

It is clear that the objective should be the guiding force in shaping the design of any model. In the context of macro models Smith (1998, 2007) provides an excellent discussion about different objectives and how different objectives can imply conflicting priorities when building a model. This also applies to stress testing models. A stylised summary of objectives for private banks (PB) and central banks (CB) is given in Figure 1, which also shows the model requirements different objectives imply.

Whilst many empirical models in the academic literature are built to evaluate economic theory, stress tests are always forecasting exercises. However, the focus of the forecasting exercise is not the mean expected outcome over a specific horizon but the impact of a severe, yet plausible stress event. The results of these forecasting exercises are then used *internally* or *externally*, with different implications for model design. For *external* use the model has to be understood by the target audience. If it is mainly used for *internal* purposes the model structure has to reflect the (risk) management culture of the organisation. Otherwise it is not taken seriously by senior management and thus the model will be ineffective. It is obvious that risk management cultures are different for different institutions, and hence models across institutions should differ as well. As indicated in the introduction, PBs and CBs differ significantly in this respect. This can explain some of the difficulties which arise in discussions between these two parties. PBs approach stress testing from a risk management perspective which is based on finance theory, mathematics and statistics. The majority of senior central bankers on the other hand have a macroeconomic background and therefore require that stress tests take fundamental macroeconomic forces into account. Given different objectives and the state of the macro-finance literature, it is clear that these approaches are not easily integrated and/or comparable, which explains some of the problems for the implementations of stress tests under Basel II. But this point should also be kept in mind, when PBs and CBs engage in discussions about stress testing.

Figure 1. Objectives and model requirements



For internal purposes, stress tests are often used as *validation* tools (van Lelyveld, 2006; CGFS, 2005). For example, historical stress tests can offer insights whether the 99th confidence interval indicated by a capital model may be correct or not. Here model accuracy is at a premium. But the ultimate objective, even for stress tests run for *validation* purposes, is always *decision making*. For example, PBs draw heavily on stress tests when setting capital, trading limits or the risk appetite of the bank (e.g. see BCBS, 2006, CGFS, 2005, or CRMPG II, 2005). For these exercises, similar models are often used as for the day-to-day risk management to ensure comparability of results. Therefore, the forecast performance of the model is essential. However, if the stress test is used for long term business planning it may be more important that the model covers the relevant questions and is tractable so that senior management can actively engage in analysing different scenarios.[1]

Similar to PBs, CBs use stress tests to guide their own policy *decision making* when for example judging the relative importance of different FS vulnerabilities (Haldane et al. 2007). Stress tests can also feed into the supervisory *decision making* process. At a very simple level, FS stress tests can act as a cross-check of PBs' individual stress test results. The IMF and the World Bank (2003) see this as an important benefit of country wide stress tests as they provide a benchmark which is based on a consistent methodology and hence easily comparable. But the use of FS stress tests can go further than that. Given their structure, most stress tests simulate the riskiness of individual PBs before aggregating the results to the system level. The first stages can, therefore, provide some indications about weak PBs. Hence, the stress test can be potentially an additional off-site supervisory tool, which can be used for internal as well as external purposes, e.g. when deciding where to focus supervisory attention or when engaging with PBs in the supervisory dialog. Again different *decision making* processes imply different priorities for model design. For analysing FS it may be most important that the model captures the broad picture of vulnerabilities, whilst for supervisory purposes model accuracy and forecast performance for stress tests run at portfolio level are essential.

The main objective, especially for non-supervisory CBs, is not internal but external *communication* about risks and vulnerabilities to influence risk taking by private agents. Indeed, stress testing results appear in financial stability reports of most CBs (for an overview see Cihak, 2007). Clearly, external *communication* is shaped by internal *decision making*. But model requirements can be different. For FS-stress test to have an impact, the underlying model has to be tractable and transparent so that it can be understood by the target audience such as risk managers in PBs. Most importantly, it has to be suitable for story telling to help illustrating and quantifying the possible unwinding of vulnerabilities, the transmission mechanism from shock to impact and likely ramifications for the financial system. Forecast performance may not be the most important aspect for this objective.

Unfortunately, different priorities can be conflicting. As already discussed in the introduction, simple autoregressive specifications may even outperform the true model with respect to forecast performance – at least for artificially generate data (see Clements and Hendry, 1998). But, autoregressive specifications are certainly not granular enough for policy

[1] Business planning is often based on analysing a set of future scenarios. In essence these are stress testing exercises. However, the horizon is generally much longer (up to 10 years) and tools are often more rules of thumb than full statistical models.

evaluation or *communication* as they do not allow for any story telling or counterfactual policy analysis. A model may also not be able to cover the relevant policy question and be tractable at the same time – something which is important to ensure transparent *communication*. Take an example from the area of macro modelling.[2] Whilst macro models used for monetary policy are now relatively concise and hence tractable, macro models used for fiscal policy decisions often remain large and complex. Policy questions asked of these models can require the modelling of different taxes, tariffs or a wide range of fiscal indicators. Hence, models have to be detailed at the micro level and broad at the same time. Therefore, models are still very much in the spirit of first generation macro models with hundreds of equation, where it can well be the case that a small change in one leads to unreasonable system dynamics. Nor is the forecast performance necessarily good for such large models. However, the model enables the (fiscal) policy maker to undertake thought experiments which highlight transmission channels if certain policy parameters are changed. Whilst the ultimate point estimates may not be accurate such a model based analyses can provide valuable information for the policy debate. It is a reoccurring theme throughout the remainder of the paper that different objectives may lead to different priorities in model design, something which has to be kept in mind when discussing and comparing stress testing models.

2 Modelling choices

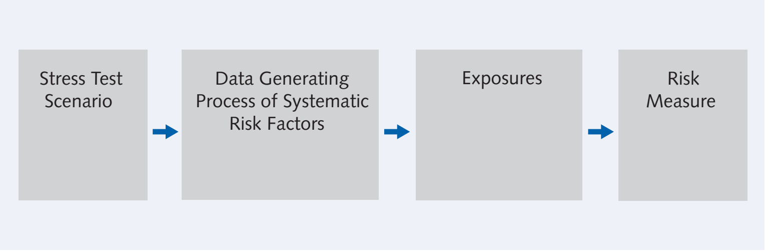
2.1 STYLISED BUILDING BLOCKS OF CURRENT STRESS TESTING MODELS

A consensus view seems to have emerged about the structure of stress tests which is summarised in Figure 2 (see e.g. Bunn et al., 2005, or Sorge and Virolainen, 2006). As Summer (2007) points out, this model structure is essentially rooted in the quantitative risk management framework (see McNeil et al., 2005).

From a modelling perspective the starting point of the quantitative risk management framework are exposures. These could be the trading book of a PB or in the context of FS stress tests the total credit risk exposures of the banking system. It is assumed that the value of these exposures in the future at time T is driven by a set of exogenous systematic risk factors*.* The main part of the stress testing model embodies the data generating process*,* which captures the interdependence of different risk

2 For a detailed discussion see Smith (2007).

Figure 2. Schematic view of building blocks of current stress testing models



factors between each other and across time. Finally, the model captures the impact of systematic risk drivers on a risk measure for the exposures at time *T*. Once the model is in place, different stress tests scenarios can be run. The main objective of the stress test (and data availability) guide modelling choices along the way. The underlying issues will be exemplified in the next section without giving a full survey of the literature. Rather than following the flow of Figure 2, the discussion follows the sequence of questions when setting up a stress testing model. Therefore it starts with a short discussion about exposures, proceeds to risk measures and the data generating process and ends with a short exploration of stress testing scenarios.

2.1 EXPOSURES

It is evident that the objective is guiding the choice of exposures which should be considered. For PBs, this is generally not a difficult question as the exposures modelled in stress tests are the same as the ones PBs base their capital models on, i.e. the banking and/or the trading book. For FS stress tests the ambition is generally to model the "whole financial system". But model and data limitations make this often impossible. The starting point for many FS stress tests is therefore to assess the most important risks for the financial system. In some sense this is a circular question: Without a full FS measurement model it is impossible to know which risk type is most important for FS. Given PBs hold capital buffers it is not necessarily the case that the most important risks for commercial banks (or other financial intermediaries) are the most important ones from a system perspective.

The more practical approach for most FS stress tests is to start with the banking system because of its pivotal role in the transformation of savings into investments and, hence, its position in transmitting FS shocks back to the real economy. Within the banking system, stress testing models usually confine themselves to domestic credit risk (see ECB, 2006).[3] On the one hand, this is due to the fact this risk type is considered to be the most important one by size of exposures. On the other hand, the focus on domestic rather than international credit exposures is driven by data availability and a desire to keep focused, even though international credit risk has been modelled as well (see Peseran et al., 2006).

Given different data availabilities some stress tests also model market risk exposures of the trading book (Boss et al., 2006), the insurance and pension sector (van den End, 2007) or counter party credit risk in the interbank market (e.g. see Elsinger et al, 2006). The surprising conclusion of the latter study is that counterparty credit risk is of second order importance. Results for pure contagion models seem to support this view (see Upper, 2007). For model builders these results suggest that for a first order approximation it is not necessary to model counterparty credit risk. And it may be useful to concentrate modelling efforts on other areas. That said, counter party credit risk models are relative straight forward to incorporate. Once the matrix of interbank exposures is known (or derived by maximum entropy from data on banks' balance sheets as commonly done), a clearing mechanism ala Eisenberger and Noe (2001) is simple to implement.

An important question is whether to consider liabilities. So far, most stress tests focus on assets, even though including liabilities is essential when aiming to model liquidity risk, which I will discuss in Section 3.2.2. But the liability side of the balance sheet is also important when looking at net-interest income, which remains the most important source of profits for commercial banks even though it has been declining in recent years. De Bandt and Oung (2004), Bunn et al. (2005) and Boss et al. (2008) capture net-interest income in a reduced form fashion, whilst Drehmann et al. (2007) derive a structural model. Their model reveals interesting dynamics about the system and confirms that the maturity missmatch between assets and liabilities can be an important source of vulnerability for PBs.

2.2 THE RISK MEASURES

As pointed out the objective determines the exposures modelled, which in turn determine the choice of risk measure to a large extent. But even for the same exposures, different risk measures used as summary variables of the stress test can be useful for different objectives.

[3] Even when only domestic credit exposures are modelled, it is not necessarily the case that these are only driven by domestic risk factors. Pesaran et al. (2006) show that for large internationally active firms national and international macroeconomic factors are important systematic risk factors.

A basic issue is whether to use a risk measure based on a market-to-market perspective or an accounting perspective. Whereas the market-to-market perspective provides a long term view of banks' health based on economic fundamentals, an accounting perspective assesses whether there could be future regulatory or liquidity constrains (e.g. when there are significant losses in the short run but sufficient profits in the long run so that the bank is fundamentally sound but capital adequacy is threatened over a 1 year horizon).[4] The choice of perspective should be aligned with the accounting standards in the country. This is obvious for PBs. But it should also be the case for FS stress tests to enhance *communication* and to ensure comparability of results of private and public players.

The ultimate choice of the risk measure is linked to the objective and whether results are used internally or externally. The focus for PBs is the optimal risk return trade-off, and risk measures used are for example capital adequacy or future profitability. So stress test results are expressed in similar terms. Given a different perspective, risk measures for FS stress tests could and maybe should be different. CBs use a host of different measures in practice. The choice is often guided by data limitations, for example when only loan loss provisions rather than credit risk write-offs or firm specific data are available. More sophisticated stress tests can look not only at losses but at net profits, the number of PB defaults or possible lender of last resort injections to recapitalise the banking system (for an overview of different measures see Cihak, 2007).

FS stress tests often normalise losses by capital to assess whether the banking system is robust or not. This is problematic for two reasons. First, banks generally make positive profits which act as the first buffer against any losses. Hence, the risk of the stress scenario is possibly overestimated, if profits are not stress tested as well (which is rarely the case). Second, banks set capital against all risks including market, credit, operational, business and reputational risk. All these risks impact on profits and losses, but are normally not stress tested, even though they may also crystallise in severe scenarios. Hence, the buffer indicated by capital maybe too large. Presentations of the results should highlight these uncertainties.

Another key problem for FS stress test is that representing the financial system with aggregate variables may be misleading. Take for example average capital adequacy for a banking system. Two different stress tests may result in average capital adequacy ratios which are well above minimum requirements, even though in one case all banks are solvent whilst in the other a major player defaults. From a FS perspective these scenarios are clearly different. Stating all individual results, on the other hand, may

4 For a detailed discussion on this issue see Drehmann et al. (2007).

not be too useful either. First, this may be resisted because of confidentiality agreements. But second, even presenting interquantile ranges or anonymous minima and maxima may distort the message. For overall FS, it is not just the capital adequacy but the size of the affected institution which is key. A failure of a very small player can be generally absorbed by the financial system whilst a large player can create financial instability leading to severe real losses. A possible solution could be to not derive the simple average but the (size) weighted average, which so far is – surprisingly – not done.

The ultimate variable of interest from a welfare perspective is the real economy and hence GDP. Measuring the impact of shocks on banks' capital adequacy is in essence only a proxy variable for this. Stress tests have not successfully tackled this problem because of formidable technical challenges. As will become apparent in Section 3.2.3, current models which capture the feedbacks from the financial sector back to the real economy are in a highly reduced form and therefore not too useful for *communication,* nor is it clear how robust they are. Hence, they are of limited use to achieve objectives for either CBs or PBs.

2.2.1 The horizon of interest

Related to the ultimate risk measure used is the question over which horizon the stress test should be run. Some guiding principles are given by the regulatory framework which specifies a ten day horizon for market risk and a one year horizon for credit risk. Early FS stress tests also used a one year horizon but it has been acknowledge that the emergence of severe credit risk losses takes time to trickle through the system. By now CB practitioners therefore often use a three year horizon. However, Drehmann et al. (2007) show for one stress test scenario that, while credit risk losses take three years to fully impact on banks' balance sheets, the maximum loss occurs after less than two years, if interest rate risk and net-interest income are modelled appropriately.

Ultimately, the horizon chosen requires a trade-off between the time a vulnerability needs to crystallise and the realism of the modelled behavioural responses by market participants and policy makers in times of stress. The shorter the horizon, the more realistic it is to assume that participants do not or cannot react for example by reducing exposures – hence the ten day focus for market risk. However, this may not be sufficient for severe stress to emerge. In essence this question is linked to the problem of endogeneity of risk, which I will discuss in Section 3.2. But the horizon is also important to assess the key systematic risk drivers which should be captured by the model. It is for example obvious that for a ten day horizon macro factors will not play an important role as they fluctuate at a much lower frequency. Overall, there is no golden rule for the optimal horizon to consider in a stress test. And again, this question has to be decided by the ultimate objective of the stress testing exercise.

2.3 SYSTEMATIC RISK FACTORS AND THE MODEL OF THE DATA GENERATING PROCESS

The choice of exposures, the risk measure and the time horizon will determine the set of risk factors which should be sensibly considered. By now there is a large literature on this issue, especially in the context of FS stress tests (see ECB, 2006, or Sorge and Virolainen, 2007). Therefore, I limit my discussion at this point to highlight that the way risk factors are modelled should be driven by the objective of the stress tests and the availability of data. For explanatory purposes I do this for credit risk only. CreditRisk+ is a model heavily used by financial institutions and CBs. Model developers admit that the single risk factor specification for CreditRisk+ is driven by the lack of data on default correlations (see Credit Suisse, 1997). Further, they argue that "*no assumptions are made about the cause of default. This approach is similar to that taken in market risk management, where no assumptions are made about the causes of market price movements. This not only reduces the potential for model error but also leads to…an analytical tractable model*" (Credit Suisse, p.7, 1997).

The objective of CreditRisk+ is risk management and capital adequacy. Hence, the forecast performance and computational speed implied by an analytical tractable model are at a premium. Given the statement above, it is apparent that the ability to tell stories how a change in a risk factor may ripple through the economy and finally impact on defaults of exposures is not considered to be important for the model.

An unspecified risk factor is not well suited for story telling and conflicts with objectives for stress testing models most central bankers implicitly have in mind.[5] Further, it is extremely difficult to link the data generating process of the single risk factor with observable variables. Hence, stress test scenarios, based on historical data or a hypothetical scenario, are hard if not impossible to implement in a robust and realistic fashion. This is one of the key problems PBs face when trying to implement stress test for Pillar II purposes of Basel II.

[5] Even though central bankers generally do not approach it this way, it should be pointed out that stress tests are possible in a model with unobservable risk factors as well. These stress tests would assess changes in the underlying data generating process. A stress test in the case of CreditRisk+ could for example be an increase in the volatility of the unobserved factor.

Since the main policy instruments of CBs in the FS arena is primarily *communication* and because of a more macroeconomic background, stress tests used by policy makers generally specify macro economic factors as systematic risk factors. Models considering aggregate variables such as write-offs or non-performing loans as their risk measure only assume macroeconomic variables (e.g. GDP, interest rates, exchange rates) as systematic risk factors (see e.g. Bunn et al., 2005 or Basu et al., 2006). Models which look at firm specific probabilities of default to construct loss distributions for credit portfolios sometimes assume that macroeconomic factors are the only important systematic risk drivers (Pesaran et al., 2006, Jacobsen et al. 2005 or Castren et al., 2007). Other papers taking firm specific defaults as inputs into their risk measures follow Wilson (1997a,b) and consider firm specific and macroeconomic variables as key explanatory variables (see e.g. Boss, 2002). While macro factors are informative about the level, firm specific factors help to better predict the cross-sectional distribution of default risk (see Carling et al., 2007). Therefore, it would seem best to consider both. However, questions then have to be asked how firm specific factors change during stress. Some, such as size, are likely to be unaffected, whilst others such as profitability or leverage certainly vary.

The choice of systematic risk factors is closely connected to the model of the data generating process. Credit risk models specify the data generating process differently. Models range from calibrated distributions of the unobserved factor (see e.g. CreditRisk+, 1997), autoregressive processes for each underlying macro variable (see e.g. Wilson 1997 a/b, or Duffie et al., 2007), reduced form vector autoregressive macro models (see e.g. Pesaran et al., 2006) or structural macro models (see e.g. Haldane et al. 2007). All these models have different benefits in terms of forecast performance, computational simplicity and speed, tractability and ability to tell stories which means that modellers have to find the right-offs, which suit their main objective.

If *communication* is the main objective for a FS stress test, unobservable factors may not be the first modelling choice as they are unsuited for storytelling. In contrast, using general equilibrium structural macroeconomic models to forecast the impact of shocks on credit risk may be very good in highlighting the key macroeconomic transmission channels. However, macro models are often computationally very cumbersome. As they are designed as tools to support monetary policy decisions they are also often too complex for stress testing purposes. For example, it may not be straight forward to undertake simple stress tests such as a 40% drop in house prices as the model requires that only deep parameters (e.g. preferences or productivity) are shocked. This may speak to reduced form vector autoregressive macro models, which could provide an optimal trade-off between complexity, ability to tell stories and forecast performance.

It should be pointed out that from a stress testing perspective it is certainly hard to use autoregressive processes as model for the data generating process. By design, such an approach does not capture the interdependence of systematic risk factors. Hence, it constraints the set of sensible stress test scenarios which could be undertaken with the model. Therefore, when setting out the model of the data generating process the desired stress test scenarios should also be kept in mind. Obviously this depends on the ultimate objective as well.

2.4 THE STRESS TEST SCENARIO

Stress test scenarios can fully determine the changes of all systematic risk factors up to the end of the stress test horizon. This is the case when undertaking a historical stress test and all risk factors are changed in line with previously observed changes in the stress period. Classical examples for such stress test scenarios in the context of market risk are the LTCM crisis or 9/11 (see CGFS, 2005).[6] But scenarios often contain only factor changes for one or two risk factors during a specific time period and the model of the data generating process then determines the change of all other risk factors conditional on the stress scenario. This is for example a scenario which would be impossible to run if the data generating process is model by autoregressive processes for each systematic risk factor. Often, these stress tests are referred to as hypothetical stress tests and common examples in are an increase in oil prices or a drop house prices. Furthermore, some stress tests not only specify the initial stress scenario but also make assumptions on the data generating process. For example, market stress tests often assume a correlation break down in stressed conditions. Other stress tests change model parameters to judge the impact of new products (e.g. see Bunn et al., 2004). Berkowitz (2000) is one of the few who looks at stress tests from a more conceptual perspective. He argues that all scenarios can be categories as one of the four types:

1) simulating shocks which are suspected to occur more frequently than historical observations suggest;
2) simulating shocks which have never occurred;
3) simulating shocks that reflect the possibility that statistical patterns could break down in some circumstances;

[6] Historical stress tests generally require some manipulation as well. For example, modellers have to decide whether to use relative or absolute changes of the respective systematic risk factors considered in the stress. Further, they have to come up with ways to model new financial products.

4) simulating shocks that reflect structural breaks which could occur in the future.

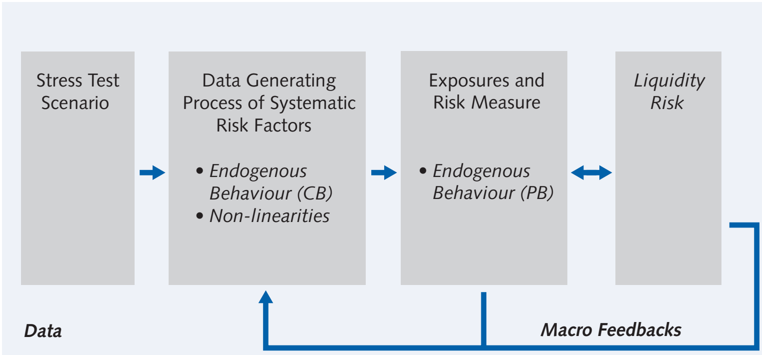
Within the model framework, these stress tests essentially imply that the modeller believes that the true data generating process is different from the one embedded in the model. This is certainly a key aspect of stress testing. However, Berkowitz' view on stress testing is too narrow. His implicit objective is assessing the robustness of capital. However for *validation, decision making* or *communication*, running single scenarios using the data generating process embedded in the model can provide fruitful insights. It has long been known that even experts have problems of correctly assessing the impact of a low probabilities event on one variable (e.g. see Kahnemann and Tversky, 1972) let alone on a portfolio of correlated exposures. It is, therefore, hard to intuitively understand what a severe adverse shock to unidentified risk factor would mean not only for losses (which are given by the model) but also for wider economic conditions. But understanding these are essential for *decision making*. It is much easier to conjecture the impact of a shock similar to the LTCM crisis in 1998. Given that such a scenario can be intuitively grasped it is easier linked to setting the risk appetite of a PB or just deriving possible reactions as contingency plans if such a scenario were to unfold.

This discussion highlights what is obvious: the objective needs to inform the choice of stress testing scenarios. This should be the main reference point in deciding whether to run a historical stress test or a hypothetical stress test, to draw the scenario randomly from the data generating process or specify it and change the data generating process at the same time.

3 Challenges

So far I tried to show that modelling choices have to be driven by objectives. However, any stress testing modeller will meet important challenges. Most of them are well understood (see ECB, 2006), but because of their complexity, limited progress has been made. Figure 3 extends the schematic representation of stress tests from Figure 2 to indicate where these challenges lie along the modelling chain. Data problems are an overarching issue, which I discuss first. Then I look at the endogeneity of risk which arises because of endogenous behaviour by PBs and the CB, liquidity risk and macro feedbacks. In the end I will briefly touch on nonlinearities which can be a result of endogenous behaviour.

Figure 3. Challenges for stress testing models



Note: Words in italics indicate challenges for stress testing models.

3.1 DATA

Data are rare, especially for severe stress episodes. This is a long standing problem for FS stress tests. As has been seen already, data availability is often an important consideration in what exposures to model and which risk measure to adopt. It should be stressed that full data availability would not solve all problems. For example, very few stress tests endogenise cyclical variations in loss given default (LGD) because of data problems. It is well known that LGDs are cyclical (see Altman et al., 2005) even though recent research has highlighted that the underlying relations are complex and the ultimate amount recovered is highly contract specific (e.g. see Davydenko and Franks, 2008). An all encompassing model should take these issues into account. However, the desire to incorporate every contractual detail would inevitably lead to an enormous model with millions of underlying data points and hundred of equations. Such a model would likely have limited forecast performance and would be rather intractable and therefore not suitable for any objective.

One frequently cited reason for the lack of data is a result of rapid innovation in financial markets, with a multitude of players and products emerging continuously. Historically stress often emerges after periods of market liberalisation or around financial innovations, with the recent turmoil providing an excellent example. Innovations are nearly impossible to capture by models as they can change the endogenous relationships within the system and hence parameters used to model the data generating process.

Rather than arguing that a lack of data and changing financial systems invalidate stress testing models, I argue the opposite: stress tests are well suited to address these problems as long as users are aware of the assumptions made. Going back to LGDs. As long as cyclical LGDs do not change the dynamic properties of the data generating process[7] but only the level of losses in case a debtor defaults, it may for example be sufficient for risk management purposes to "stress test" a portfolio credit risk model by assuming that LGDs are as low as the lowest historical observation or even lower. In practice this is often done as such a stress test can provide some comfort about the level of capital. This may already be sufficient, depending on the objective. However, when communicating the results it should be kept in mind that numbers generated by such a stress testing model are not accurate point estimates but only provide an assessment about possible downside risks.

The same can be done for new financial instruments where no data are available, especially not during stressed conditions. For example, Bunn et al. (2004) assess the impact of buy-to-let mortgages on FS. In the early years of this decade, buy-to-let mortgages were a new financial product in the UK, mentioned frequently as a possible FS risk. As no historic default series were observed, the authors just assumed that buy-to-let mortgages were up four times more sensitive to changes in house prices than normal mortgages. By running several stress test scenarios, this exercise showed that even with extreme assumptions buy-to-let mortgages hardly posed a FS threat to the UK.

Whilst stress test can address some of the problems posed by a lack of data and a changing FS system, the underlying problem remains: standard parametric econometric techniques require sufficient data which is often not given. This can lead to large errors in the econometric specification of the data generating process. Therefore, a lack of data will limit the forecast performance of any stress testing model – especially as the focus of stress tests are the tails of the distribution.

As a first step, model outputs should therefore not only present point estimates. Even though this is generally not done, error bands could either be based on standard statistical techniques (see Hanson and Schuerman, 2006, or Drehmann et al., 2006) or on judgement (see Haldane et al., 2007). The lack of model robustness also implies that many assumptions have to be made. These can significantly alter the results. As a second step, this means that models should be run under different assumptions as Cihak (2007) rightly stresses. And it is also important that this has to be made transparent to the broader audience of the model results.

However, modellers may want to consider different econometric methods. It is for example surprising that Bayesian methods are not more frequently used for risk management models. Such an approach seems

7 Apriori feedbacks cannot be excluded. In the case of mortgages, for example, higher LGDs are linked to lower house prices and lower house prices in turn may induce higher mortgage defaults in subsequent periods.

especially useful in an environment where people's priors deviate from observable data as is for example the case for low default portfolios. Alternatively, Segoviano and Padilla (2007) successfully deal with limited data by using non parametric models based on entropy. Their model provides robust estimates even for data series with less than 20 data points.

3.2 THE ENDOGENEITY OF RISK

Danielsson and Shin (2002) discuss endogenous and exogenous risk by reviewing the problems of the Millennium Bridge in London. The Millennium Bridge is a foot bridge over the Thames. After the opening weekend, it had to be immediately closed down because of safety concerns. When designing the bridge engineers had assumed a limited number of pedestrians which would walk with different rhythms so that the resulting vibrations would cancel each other out. In essence they assumed risk to be exogenous. However, an unexpected large number crowd wanted to see the bridge for the opening weekend. Given the flexible design of the bridge small oscillations emerged. In turn these lead people to sway in step, increasing the amplitude of the oscillation and amplifying the original shock in an ever increasing feedback loop. This story highlights how small exogenous shocks can have a disproportional impact because of an endogenous behavioural response.

In financial systems, the endogeneity of risk can emerge for several reasons. Below I discuss endogenous behaviour by PBs, the CB, macro feedbacks and liquidity risk in more detail. Endogenous behaviour will also give rise to non-linearities which are discussed in the last sub-section. Developing models which are able to capture the endogeneity of risk is possibly the most important challenge risk management models and stress tests need to meet. Once risk is endogenous, it is well known that standard risk management models break down (see e.g. Danielsson, 2002).

3.2.1 Endogenous behaviour

Endogenous risk is essentially due to endogenous behavioural reactions by agents in the economy including the policy maker. This is also true for macro feedbacks and liquidity risk. Given their importance it is however worth to discuss them separately. As shown in Figure 2, stress tests are a chain from an exogenous shock via the data generating process to the impact on banks' balance sheets. Exogenous behavioural responses are important at several stages (see Figure 3).

First, in standard stress tests exposures only change because some default occurs and/or their market value changes. Implicitly, PBs are sitting on their initial portfolio allocation during the stress event without trying to hedge losses or realign their portfolio. Over a one to three year horizon, this is clearly unrealistic. Most modellers are aware of this problem. But the model structure hides this issue. Once the maturity structure of PBs balance sheets is taken into account, endogenous behaviour has to be addressed. Drehmann et al. (2007) model assets and liabilities as well as their respective maturity structure to capture interest rate risk in the banking book. A large part of the book has a maturity below one year. Hence, even for a one year stress test, the question has to be addressed what the PB (and depositors) is doing once assets (and liabilities) mature or can be repriced.

An ideal model would consider full portfolio optimisation. This has been done by the operations research literature discussing stochastic programming models for dynamic asset and liability management.[8] But even the latest papers can only undertake this modelling exercise for tradable assets funded with a simple cash account (see Jobst et al., 2006). Drehmann et al. (2007) therefore use a simple rule of thumb. They assume that PBs as well as depositors are passive, i.e. that they continue to invest in the same assets with the same risk characteristics as before.[9] Clearly, rules of thumb are not ideal. However, it is a first step in modelling endogenous behaviour. De Bandt and Oung (2004) follow a different strategy. Rather than building a structural model, they establish a relationship between the demand and supply for credit and the state of the economy. Hence, balance sheet adjustments by PBs are accounted for in a reduced form fashion.

Second, when talking about endogenous behaviour, it is crucial to consider the policy makers as well. If the model of the systematic risk drivers is reduced form, then the (average past) CB policy response is already embedded in the data generating process. To clarify this, assume that the events of the summer 2007 are run as a historical stress test scenario using the observed changes in market prices. The latter are, however, a result of the stress event as well as CBs liquidity interventions. Hence, by re-running this scenario, a similar CB reaction is implicitly assumed. Market participants may reasonably expect this in the future if a similar scenario would unfold again. However, if this stress test is run by a central banker aiming to explore the robustness of the system with and without policy interventions, this is obviously a problem.

---

8 For an overview over this literature see Zenios and Ziemba (2007).

9 Drehmann et al. also have to make an assumption about the re-investment of positive profits. The modelling framework is generally flexible enough to also look at rules of thumb where balance sheets are increasing or portfolio realignment occurs. Alessandri and Drehmann (2007) base a full economic capital model on the same set-up.

Reduced form macro models representing the data generating process share a similar problem as past interest rate decisions are embedded in the data generation process. If a structural macro models is used to capture the dependence of macro risk factors, most stress tests do not model the behaviour explicitly but rely on an estimated Taylor rule, i.e. they assume that the CB mechanically sets interest rates to minimise deviations in inflation and output. It is well known that this assumption imposes problems in generating severe stress scenarios. For example, a severe shock to the housing market would lead to a reduction of the interest rate, which dampens the impact of the initial shock on banks balance sheets as interest rates are an important driver for corporate and household defaults. Without additional shocks to inflation, it is therefore very hard to generate consistent scenarios where interest rates rise whilst house prices fall, even though this was the case for example in the early 1990s recession in the UK, which was a stress event for PBs.

No easy answers can be given how best to start modelling endogenous behaviour. However, it is already important to be aware of the problem. A first step could be to explore rules of thumb for both policy makers and PBs. In practice this would force PBs to start thinking not only about their behavioural responses but also what other market participants would do. The latter step is important for understanding systemic risk in general and liquidity risk in particular.

3.2.2. Liquidity risk

Liquidity risk crystallizes because of an endogenous behavioural response by agents. In the classic paper on liquidity risk by Diamond and Dybvig (1983), agents run because other agents run.[10] The modern literature on bank runs (Goldstein and Pautzner, 2005) establishes that behaviour by depositors is not driven by sun-spots but by negative information. It is not only funding liquidity risk but also market liquidity risk which can affect the overall stability of the financial system. Markets may be illiquid because of informational frictions as is the case in the classical markets for lemons (Akerlof, 1970). And market liquidity more broadly may dry up because of behavioural responses by agents when they for example withdraw their money from weak performing funds (see Vayanos, 2004) or if there is a negative feedback loop between funding and market liquidity

[10] In lack of a common definition, I mean by market liquidity the ability to sell assets at fair value with immediacy and by funding liquidity the ability to satisfy the demand for money (see Drehmann and Nikolaou, 2008).

risk (see for example Gromb and Vayanos 2002 and Brunnemeier and Pedersen 2006).[11]

Market participants would argue that such spirals have happened, for example during the LTCM crisis or more recently during the turmoil in the summer 2007. Rather than aiming to embed these spirals into the model themselves, market practitioners use these events as historical stress tests. If these scenarios are indeed representative of market liquidity dry ups and the main objective of the stress test is to assess the robustness of capital, this may be a valid strategy. Said that, such an approach cannot reveal much about the underlying transmission mechanism and hence it is less well suited for FS stress tests.

Therefore, and because PBs are generally illiquid before they are insolvent, some FS stress tests aim to incorporate liquidity risk (see Jenkinson, 2007). However, making empirical progress on these questions remains difficult. First, to measure liquidity risk, not only assets but also liabilities and off-balance sheet items and their respective maturities have to be considered. This expands the universe of necessary data considerably. PBs own approaches rely on vast amounts of confidential data which are changing continuously and rapidly, especially during stress. This limits their use from an FS perspective. Second, data on behavioural responses by depositors and counterparties in the interbank market are also rare. Therefore, liquidity stress tests are based on rules of thumb rather than on empirical relationships. Using bidding data from open market operations, Drehmann and Nikolaou (2008) are so far the only study which measures funding liquidity risk with data available to CBs. They are able to capture the recent turmoil but data restrictions imply that they can measure liquidity risk only over a one week horizon. It remains unclear how their approach could be incorporated into a model with a much longer horizon. The more general problem is that the link between shocks to solvency, modelled by current stress tests, and liquidity is even less clear.

3.2.3 Macro feedbacks

There is strong evidence that system wide solvency and liquidity crises in the banking system lead to significant costs in terms of loss of GDP (e.g. Hoggarth et al., 2002). But so far, linking the real and the financial sector has proven to be difficult. Christiano et al. (2007) for example build on the financial accelerator literature (see Bernanke et al. 1999) and include

[11] The intuition behind these spirals is as follows. Assume a severe drop in asset prices which induces higher margin calls. If the funding liquidity of PBs is constrained, higher margin calls can only be satisfied by selling assets. This lowers asset prices further if many PBs have to raise cash because of a lack of market liquidity. In turn this raises margin calls, leading to increased funding liquidity demands and so forth. Whilst such a spiral is theoretically understood it is very hard to model empirically. For a survey on asset market feedbacks see Shim and von Peter (2007).

an aggregate banking/financial sector in a dynamic stochastic general equilibrium model. They find that this sector is a source of shocks which can account for a significant portion of business cycle fluctuations. But the sector is also an important amplification mechanism.

Whilst this is an interesting macro model, it is of limited use for FS stress tests. As has been vehemently argued by Goodhart in many publications (see e.g. Goodhart et al. 2006) and should also be clear from the discussion on risk measures above, it is important to model heterogeneous actors within the financial system. A single financial sector will mask many of the important relationships. For example, different PBs may have different preferences. Hence, they may take different risks and the most risky ones are likely to fail first. Further, aggregate (funding) liquidity conditions are set by the CB. As long as it does not make any massive policy mistakes, the level of aggregate liquidity is not an issue even in crises. But the distribution of liquidity across institution certainly is, as an institution short of liquidity will fail. Given interbank markets such a failure may induce contagion to other banks – i.e. counter party credit risk– with different ramifications for the real economy depending on how many and which banks fail.

From many private discussions I know that most modellers at CBs intend to extend the model set-up depicted in Figure 2 by implementing a feedback loop from the risk measure – i.e. banks' losses – back to the model of the data generating process. If a large structural macro model is used as a model of the data generating process, an intuitive possibility could be to link the risk premia and availability of investment funds to the capital adequacy of banks. There is indeed empirical evidence that investment is a negatively related to conditions in the banking sector (see Peek and Rosengreen, 2000, or Dell'Ariccia et al. 2005). But I am not aware of any stress testing model which links the banking sector and the real economy. The only successful approaches so far are reduced-form models, which are different in nature than the standard stress testing model set-up depicted in Figure 2.

These models are essentially large scale vector autoregressive models (VAR) and based on the same idea of linking the standard set of macro factors with risk measures of the financial system.[12] Aspaches et al. (2006) use a cross country approach and proxy financial sector risk by bank defaults and bank profitability. Jacobsen et al. (2005) do not explicitly model banks but set-up a panel VAR modelling macro factors and the likelihood of default for Swedish companies. They find that macro feed-

[12] Hoggarth et al. (2005) develop a macro VAR model which includes write-offs by PBs as a risk measure. In contrast to the other papers cited above they find little evidence of feedbacks.

backs can have important implications and that sometimes monetary policy and FS objectives can conflict. De Graeve et al. (2007) follow the approach of Jacobsen et al. (2005) but use PDs for the German banks directly. They show that a negative monetary policy shock impacts significantly on the robustness of the banking sector, but only once the feedback from bank PDs back to the macro economy is allowed for.

Whilst these models are of limited use for PBs, they provide CBs with some quantitative indications about the importance of macro feedbacks. However, given their highly reduced form nature, there are not well suited for *communication.* Furthermore, the endogenous response by the policy maker is already embedded in the model, which is not desirable if the model is used to analyse counterfactual policy experiments.

3.2.4 Non-linearities

Policy makers frequently argue that non-linearities will emerge during stress (e.g. see Haldane et al., 2007, ECB 2006).[13] Conceptually, it is not entirely clear what policy makers really mean by this and it is therefore worth to disentangle various aspects. On the one hand, non-linearities are a result of endogenous behavioural responses. On the other hand, models may not capture them because they are econometrically misspecified.

Standard parametric econometrics generally imposes a log-linear specification on the model of the data generating process. This is also done for macro economic models. Given their objective is to forecast the mean outcome around the equilibrium, results may be acceptable as mistakes made may be not too large.[14] This cannot be expected for extreme stress events. Stress testing modellers therefore have to assess where significant non-linearities can arise. For example, given the binary nature of default the link between systematic risk factors and credit risk is often modelled in a nonlinear fashion for example as a probit specification (see Wilson 1997 a/b) or based on a Merton type model (see Drehmann, 2005, or Pesaran et al., 2006). Drehmann et al. (2006) show that even for probit specifications non-linearities may not be sufficiently well captured. Accounting for the non-linear nature of the data generating process, they show that stress test results are significantly different. It is a question, however, whether tackling this type of non-linearities is of high importance, especially if the objective is *communication*. Modelling non-linearities in the underlying data generating process will not reveal any

13 Mathematically, a model has non-linearities if the impact of a three standard deviation shock is not simply three times the impact of a one standard deviation shock.

14 A log-linear specification can mathematically be interpreted as first order Taylor approximation of true data generating process (see Jordà, 2005). But for severe stress events in the tail of the distribution such an approximation cannot hold.

more about the transmission mechanism from shock to impact. Clearly, it will change the level of different stress test scenarios,[15] which is important if the objective is risk management.

Non-linearities may also arise because of jumps or switching between multiple equilibria. This can be a result of endogenous behavioural reactions and is presumably really what policy makers care about. Jumps and multiple equilibria are central to models about funding and market liquidity discussed above. One possible avenue to explore, is modelling these endogenous non-linear reaction with a structural model. As the previous discussion highlights this is difficult. Another possibility could be to capture them by models with regime shifts. So far, I am only aware of one model. Bruche and Gonzalez-Aguado (2008) use a latent, unobserved factor model to capture the dynamic join distribution of default probabilities and recovery rates in the US. In terms of forecast performance, they find that this model outperforms models based on observable macro factors. But as pointed out before, such a modelling approach is not well suited for *communication* but could be useful for *decision making*. It seems a fruitful avenue for future research to assess how important regime shifts are and whether they could be practically incorporated into risk management and stress testing models.

4 What's next?

Modellers have to cope with a lack of data, as this cannot be changed at least within the medium run. The endogeneity of risk, however, is an area where progress can be made. So far all stress testing models essentially follow the set-up depicted in Figure 2, except the highly reduced-form models discussed to capture macro feedbacks back to the real economy. The only exception is the work stream by Goodhart et al. (2004, 2005, 2006 a,b)[16] who theoretically derive general equilibrium models with incomplete markets where agents are heterogeneous and default can occur. As stressed by Goodhart et al., both of these features are essential when aiming to model FS. The model by Goodhart et al. goes well beyond the standard stress testing models as all agents, in all markets, in all states of the world are fully optimising over quantities, prices and defaults. The model is therefore able to address some of the problems discussed above, most importantly endogenous behaviour. Modelling defaults in a general equilibrium framework is one of the key challenges for their set up. In the classic Arrow-Debreu model it is implicitly assumed

[15] For univariate models, non-linearities will also not change the rank ordering of different scenarios.
[16] The theoretical model to assess financial stability is based on Tsomocos (2003a,b).

that all agents honour their obligations, and thus there is no possibility of default. Hence, Goodhart et al. follow Shubic and Wilson (1977) and treat default as the repayment rate which is endogenously chosen by agents. In this sense defaults are partial and voluntary. Even though Tsomocos and Zicchino (2005) show that there is a an equivalence between a general equilibrium model with incomplete markets and a model where default is endogenous, this model structure makes it already hard to communicate with senior CB management or PBs as they see default as an exogenous event. *Communication* is also not fostered by the complexity and intractability of the model.

Ultimately, calibrating and finding computational solutions for the model are the major difficulties. So far this has only been tried for the UK (Goodhart et al., 2006b) and Colombia (Saade et al. 2006). In both cases, it was only possible to implement a highly stylised model with three different banks[17], two states of the world (stress and no stress) and two time periods. Even in this case, calibration proves difficult. As Saade et al. (2006) explain, some parameters such as policy variables are observed, some can be calibrated using econometric methods (e.g. the income elasticity) and others such as the occurrence of the stressed state or default penalties are arbitrarily imposed. Saade et al. replicate the Columbian banking crisis in 1997–1999. For some model variables (such as the volume of mortgage loans) the model seems to perform well. For others, such as GDP, projections are far off true developments.

An alternative is therefore to extend the model set-up in Figure 2 to the model depicted in Figure 3. The most ambitious project in this regard is currently undertaken by the Bank of England (see Jenkinson, 2007). As a starting point it takes the structure of the standard Bank of England stress testing model (see Bunn et al. 2005 and Haldane et al., 2007) which covers the macro economy, credit risk and banks net interest profits. Following the Austrian central bank's model (see Boss et al. 2006), the basic structure is extended to include market risk exposures of banks and counter party credit risk. Additionally, interest rate risk is modelled structurally along the lines of Drehmann and Alessandri (2007). The aim is to also cover macro feedbacks as well as market and funding liquidity risk. Because of a lack of data, robust estimates for the latter may not be possible and hence these channels may be very much based on rules of thumb. Nonetheless, the model breaks important new ground and will certainly highlight interesting FS channels. First model results suggest for example that the distribution of systemic risk, if measured as the aggre-

[17] In the model there is a one to one relationship between a class of households and a PB. Hence, there are also three different types of households.

gate loss distribution of the banking system, may be bi-model (see Jenkinson, 2007).[18]

Such a model may be the only possible solution to deal with the limitations of current stress tests. However, an important drawback for this approach is that it does not break with the modular structure inherent in all current stress testing models. This implies that there will be most likely empirical as well as theoretical inconsistencies across modules.[19] Given piece-wise estimation it is also likely that model errors add up with important implications for the robustness of the model. There is also a clear danger that such a model will become so complex and non-transparent, that only a few highly specialised economists are able to understand the dynamics of the model if parameters are changed. Lack of model robustness and high degree of complexity may ultimately limit its usefulness for external *communication*, which is an important consideration as this is the main objective of FS analysis for a non-supervisor CB.

That said, engaging in a model building process is already an important step in deepening the understanding of FS and its system dynamics. Limitations of these models will certainly remain. But an overall FS model can already provide useful inputs into the policy debate and help communication efforts, as long as these limitations are made transparent and model results are presented carefully.

5 Conclusions

In this article I look at the objectives, challenges and modelling choices stress tests. I argue that for model builders as well as model users it is essential to understand the main objective of the stress tests. Different objectives highlighted are *validation*, *communication* and *decision making.* Sometimes these objectives imply different and possibly conflicting priorities for model design. The main building blocks of stress tests are then discussed in the light of different objectives and the article concludes with a discussion about the key challenges for current stress testing models and how they could be addressed. The endogeneity of risk is the main issue for standard risk management models as it challenges the fundamentals of the current model set-ups which assume a chain from an exogenous shock via the data generating process to the impact on banks' balance sheets.

Finally I want to point out that a successful stress testing model is not only designed with a clear objective in mind, but suitable instruments

[18] It should be pointed out that this result relies on an extreme assumption for market risk.

[19] An easy to make mistake would for example be to treat interest rates as I(1) variable in one module but I(0) in another.

are identified with which the objective can be achieved. Summer (2007) highlights, that PBs understand their objectives and instruments much better than CBs. PBs ultimate objective is to maximise shareholder value by taking and managing risks so that an optimal risk return trade-off is achieved. Stress tests contribute to this objective by deepening the understanding of risks. And if issues are identified, the PB has clear instruments to address problems for example by raising more capital, setting limits or reducing exposures. Instruments for CBs to address FS issues are more ambiguous. In general, interest rates are set to ensure price stability. During the recent financial turmoil CBs have successfully used open market operations to address FS issues. However, by design these instruments are focused narrowly at providing liquidity for a specific horizon (e.g. one week), to ensure the appropriate implementation of the monetary policy stance. No other issues, such as the building up of FS vulnerabilities, can be addressed. Therefore, the main instrument used to address FS is *communication* with financial institutions.

There is however a question how much this can ultimately achieve. For example, the most recent turmoil should give some pause for thought. Notwithstanding that the asset-backed commercial paper market was not specifically highlighted as possible vulnerability, CBs around the globe had identified complex financial products, high leverage and trading in illiquid markets as FS risks before the turmoil (see IMF 2007; Bank of England 2007; ECB 2007; and Geithner 2007). And publications demonstrate that these calls were acknowledged by the banking industry (see CRMPG 2005; and IFRI-CRO Forum 2007). Nonetheless the crisis occurred. Can we conclude that communication had no impact? Would the crisis have been worse without FS reports? Maybe CB warnings were not acted upon this time. But given that CBs made valid attempts to identify the vulnerabilities, their reputations should be enhanced. But does this mean that the private sector will be more responsive in the future? Finding answers on these questions is not easy. Ultimately, I remain doubtful about how much communication can achieve given considerable uncertainties and the incentives for risk taking by PBs. For communication to become more than “cheap talk” it seems essential to develop reliable measures of FS and link those to policy instruments such as regulations. As Summer (2007) points out, the link between stress tests as a measure of FS and concrete policy instruments is so far uncharted territory. I hope that the continued development and use of stress tests with all their benefits and limitations will foster this debate. This, and capturing the endogeneity of risk in stress testing models, is certainly needed.

Bibliography

Allen, F and D. Gale, (2000), "Financial contagion, *Journal of Political Economy*, Vol. 108(1), pages 1–33.

Allessandri, P. and M. Drehmann, (2007),"An economic capital model integrating credit and interest rate risk", *mimeo".*

Altman, E., A. Resti, and A. Sironi, (2005), *Recovery risk*, Risk Books.

Akerlof G.E., (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism". *Quarterly Journal of Economics* 84 (3): 488–500.

Aspachs, O.C., A.E. Goodhart, M. Segoviano, D. Tsomocos and L. Zicchino, (2006), "Searching for a metric for financial stability", *FMG Special Paper,* No 167.

Aragones JR., C. Blanco, K. Dowd, (2001), "Incorporating Stress Tests into Market Risk Modelling", *Derivatives Quarterly*.

Bagehot, W., (1873), *Lombard Street: A description of the money market,* re-edited by John Wiley and Sons (1999).

Bank of England (2007), *Financial Stability Report,* June.

Basu, R., N. Choueiri and A.G. Pascual, (2006), "Financial sector projections and stress testing in financial programming: a new framework", *IM Working Paper* 06/33.

BCBS, Basel Committee on Banking Supervision (2006), *Basel II: International convergence of capital measurement and capital standards: A revised framework*.

Berkowitz, J., (2000), "A coherent framework for stress testing", *Journal of Risk,* Vol. 2, No.2 pp. 5–15.

Bernanke, B.S., M. Gertler and S. Gilchrist, (1999), "The financial accelerator in aquantitative business cycle framework," in J.B. Taylor and M. Woodford (eds.), Handbook of Macroeconomics, Volume 1.

Boss, M (2002), "A macroeconomic credit risk model for stress testing the Austrian credit portfolio", OENB, *Financial Stability Review,* 4, October.

Boss, M., G. Krenn, C. Puhr and M. Summer, (2006), "Systemic Risk Monitor: A model for systemic risk analysis and stress testing for Austria", OENB*, Financial Stability Report,* Vol. 11, 83–95.

Bruche, M. and C. Gonzales-Auado (2008), "Recovery rates, default probabilities and the credit cycle", *mimeo*

Brunnemeier, MK. and , LH. Pedersen, (2006), "Market liquidity and funding liquidity', mimeo, Princeton University.

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

Carling K., T. Jacobson, J. Lindé and K. Roszbach, (2007), "Corporate credit risk modelling and the macroeconomy", *Journal of Banking and Finance*, Vol. 31, No. 3, pp 845–868.

Castren O., S. Dees and F. Zaher, (2007), "Global macro-financial shocks and corporate sector expected default frequencies in the Euro Area", *ECB mimeo.*

CGFS, (2005), "Stress testing at major financial institutions: survey results and practice" CGFS publication No. 24.

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

Chen, N-K, (2001), "Bank net worth, asset prices and economic activity", *Journal of Monetary Economics,* Vol. 48, pages 415–36.

Christiano, L., R. Motto and M. Rostagno, (2007), "Financial factors in business cycles", *mimeo.*

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

Credit Suisse (1997), *Credit Risk+: A credit risk management framework.*

CRMPG II, Counterparty Risk Management Group II, (2005), *Towards greater financial stability: A private sector perspective*.

Danielsson, J. (2002), "The emperor has no clothes: Limits to risk modelling." *Journal of Banking and Finance* 26, no. 7, pp. 1273–1296.

Danielsson, J., H.S. Shin, (2003), "Endogenous Risk", In *Modern risk management: A history*, Edited by Field, P. Risk Books.

Davydenko, S. and J. Franks, (2008), "Do bankruptcy codes matter? A study of defaults in France, Germany, and the U.K." *Journal of Finance,* forthcoming.

De Bandt, O. and V. Oung, (2004), "Assessment of "stress tests" conducted on the French banking system", Banque de France, *Financial Stability Review,* No. 5, November.

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

Dell'Ariccia G., E. Detragiache, R. Raghuram Rajan, (2005), "The real effects of banking crises", *IMF Working Paper,* No 05/63.

Diamond, D. and P. Dybvig, (1983), "Bank runs, deposit insurance, and liquidity, *Journal of Political Economy*, Vol. 91, pages 401–19.

Drehmann, M., (2005), "A market based stress test for the corporate credit exposures of UK banks", www.bis.org/bcbs/events/rtf05Drehmann.pdf.

Drehmann, M., M. Stringa and S. Sorensen, (2007), "The integrated impact of credit and interest rate risk on banks: An economic value and capital adequacy perspective", *Bank of England Working Paper.*

Drehmann, M., A. Patton and S. Sorensen, (2006) "Corporate defaults and macroeconomic shocks: non-linearities and uncertainty", *mimeo,* Bank of England.

Duffie D., L. Saita, K. Wang, (2007), "Multi-period corporate default prediction with stochastic covariates", *Journal of Financial Economics* , Vol 83.

ECB, (2006), "Macro stress testing practices", ECB*, Financial Stability Review*, June.

ECB (2007), *Financial Stability Review,* June.

Eisenberg, L. and TH. Noe (2001), "Systemic risk in financial systems", *Management Science*, Vol. 47(2), p 236–49.

Elsinger, H., A. Lehar and M. Summer, (2006), "Risk assessment for banking systems", *Management Science,* Vol. 52, No. 9, pp. 1301–1341.

Geithner, T., (2007), "Liquidity and financial markets", Keynote address at the 8th annual risk convention and exhibition, Global Association of Risk Professionals, New York.

Goldstein, I. and A. Pauzner, (2005), "Demand-deposit contracts and the probability of bank runs", *Journal of Finance*, Vol. 60(3), pages 1293–327.

Goodhart, C.A.E., P. Sunirand and D.P. Tsomocos, (2004), "A model to analyse financial fragility: applications", *Journal of Financial Stability,* Vol. 1(1), pp 1–30.

Goodhart, C.A.E., P. Sunirand and D.P. Tsomocos, (2005), "A risk assessment model for banks", *Annals of Finance*, Vol. 1, pp 197–224.

Goodhart, C.A.E., P. Sunirand and D.P. Tsomocos, (2006 a), "A model to analyse financial fragility", *Economic Theory*, Vol. 27, pp 107–142.

Goodhart, C.A.E., P. Sunirand and D.P. Tsomocos, (2006 b), "A time series analysis of financial fragility in the UK banking system", *Annals of Finance*, Vol. 2(1), pp 1–21.

Gromb, D. and D. Vayanos (2002), "Equilibrium and welfare in markets with financially constrained arbitrageurs", *Journal of Financial Economics*, 66, 361–407.

Haldane, A., S. Hall and S. Pezzini, (2007), *A new approach to assessing risks to financial stability* Bank of England *Financial Stability Paper* No 2.

Hanson, S. and T. Schuermann, (2006), "Confidence intervals for probabilities of default" *Journal of Banking and Finance,* Vol. 30:8, 2281–2301.

Hoggarth, G., R. Reis and.V. Saporta, (2002), "Costs of banking system instability: some empirical evidence", *Journal of Banking and Finance*, Vol. 26(5), pages 825–55.

Hoggarth, G., S. Sorensen and L. Zicchino, (2005), "Stress tests of UK banks using a VAR approach" *Bank of England Working Paper,* No 282.

IFRI – CRO Forum (2007), *Survey on economic capital practice and applications*, The Institute of the Chief Risk Officers (CRO) and International Financial Risk Institute.

IMF (2007), *Global Financial Stability Report,* April.

IMF and World Bank, (2003), *Analytical tools of the Financial Sector Assessment Programme.*

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 308–341.

Jenkinson, N., (2007), "The Bank of England's approach to top-down systemic stress testing", ECB conference on Simulating Financial Instability.

Jobst, N.J., M. Gautam and SA. Zenios, (2006), "Integrating market and credit risk: A simulation and optimisation perspective", *Journal of Banking and Finance*, Vol. 30, pages 717–42.

Jordà, O., (2005), "Estimation and inference of impulse responses by local projections", American Economic Review, vol 95 (1), 161–182.

Kahneman, D., and A. Tversky, (1972). "Subjective probability: A judgment of representativeness". *Cognitive Psychology*, *3*, 430–454.

Kupiec, P. H., (2000), "Stress tests and risk capital", *Journal of Risk,* Vol. 2, No.4 pp. 27–39.

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

Peek, J. and E.S. Rosengren, (2000), "Collateral damage: Effects of the Japanese bank crisis on real activity in the United States" *The American Economic Review*, Vol. 90, No. 1. (Mar., 2000), pp. 30–45.

Pesaran, MH, T. Schuerman, BJ Treutler and SM Weiner, (2006), "Macroeconomic dynamics and credit risk: a global perspective", *Journal of Money Credit and Banking*, Vol. 38, No. 5, pp. 1211–1262.

Saade A., D. Osorio and S. Estrada, (2007), "An equilibrium approach to financial stability analysis: The Columbian case", *Annals of Finance,* Vol. 3, pp 75–105.

Shim, I and G. von Peter, (2007), "Distress selling and asset market feedback: a survey", *BIS Working Paper* No 229.

Shubik, M., and C. Wilson, (1977), "The optimal bankruptcy rule in a trading economy using fiat money," *Journal of Economics* Vol. 37, 337–354.

Smith, R.P., (2007), "Relating DGSE and GVAR models", *mimeo.*

Smith, R.P., (1998), "Emergent policy making with macro-econometric models", *Econometric Modelling,* Vol. 15, pp. 429–442.

Sorge, M. and K. Virolainen, (2006), "A comparative analysis of macro-stress testing with application to Finland", *Journal of Financial Stability,* Vol. 2, pp 113–151.

Summer, M., (2007), "Modelling instability of banking systems and the problem of macro stress testing", ECB conference on Simulating Financial Instability.

Taylor, M and D. Peel, (2000), "Non-linear adjustment, long-run equilibrium and exchange rate fundamentals", *Journal of International Money and Finance,* Vol 19 (1), pp 33–53).

Tsomocos, D.P., (2003 a), "Equilibrium analysis, banking and financial instability", *Journal of Mathematical Economics*, 39 (5-6), pp 619–655.

Tsomocos, D.P., (2003 b), "Equilibrium analysis, banking, contagion and financial fragility", *Bank of England Working Paper*, No. 175.

Tsomocos, D.P. and L. Zicchino, (2005), "On modelling endogenous default",*Financial Markets Group Discussion Paper*, No. 548.

Upper, C., (2007), "Using counterfactual simulations to assess the danger of contagion in interbank markets", *BIS Working Paper,* No 234.

Van den End, W., M. Hoeberrichts and M. Tabbae, (2006), "Modelling scenario analysis and macro stress-testing", *DNB Working Paper* No. 119.

van Lelyveld, Iman P.P. (editor), (2006), *Economic Capital Modelling: Conceptual Issues, Measurement and Implementation*, London: Risk Publishing.

Vayanos, D. (2004), "Flight to Quality, Flight to Liquidity, and the Pricing of Risk", *mimeo.*

Virolainen, K. (2004), "Macro Stress Testing with a Macroeconomic Credit Risk Model for Finland", *Bank of Finland Discussion Papers*, 18.

Wilson, T.C. (1997a), "Portfolio Credit Risk (I)". *Risk*, September.

Wilson, T.C. (1997b), "Portfolio Credit Risk (II)". *Risk*, October.

Working Group of Economic Capital Models (2005) *"A survey into the practical implementation of economic capital models".*

Zenios, SA. and WT Ziemba, (2007), "Volume 1: Theory and methodology". *Handbook of Asset and Liability Management,* Elsevier, Amsterdam.