Solvency stress testing of banks: Current practice and novel options*

Thomas Breuer [†] Martin Summer [‡]

5th July 2017

*We thank Stefan Palmquist, Mathew Pritsker and Jakob Winstrand for helpful comments on an earlier version of this paper. We would also like to thank Mathias Drehmann, Robert Ferstl, Gerald Krenn, Til Schuermann, Miguel Segoviano and Ralph Spitzer for discussing with us various technical and procedural aspects of current stress testing frameworks.

[†]Thomas Breuer, University of Applied Sciences Vorarlberg, Research Center PPE, Hochschulstraße 1, A-6850 Dornbirn, Austria, E-mail: thomas.breuer(AT)fhv.at, Tel: +43-5572-792-7100, Fax: +43-5572-792-9510.

[‡]Martin Summer, Oesterreichische Nationalbank, Economic Studies Division, Otto-Wagner-Platz 3, A-1090 Wien, Austria, E-mail: martin.summer(AT)oenb.at, Tel: +43-1-40420-7200, Fax: +43-1-40420-7299

Contents

Executive Summary

In this paper we discuss the current theory and practice of bank solvency stress testing. It should support the Swedish central bank, Sveriges Riksbank, and the Swedish regulatory and supervisory authority, Finansinspektionen, in thinking about their current stress testing framework. It should help both institutions opening a new perspective on how this framework could be developed.

Our leading questions We organise our paper along a set of five leading questions: What are the different purposes of stress testing? What can we learn from international best practice in this field? What is a useful conceptual frame to think about the bigger picture of stress testing as well as about its most important details? How should the results of a stress test be best represented? If we could design a stress testing framework from scratch, what would be the most important elements, keeping in mind the specific structure of the Swedish banking system?

The different purposes of stress testing We begin our analysis in Section 2, by discussing the different purposes of stress testing. We analyse the common distinction between micro-prudential stress tests, with a focus on the solvency of individual banks, and macro-stress tests, with a focus on financial stability of the banking system as a whole. We believe that a useful stress testing framework must integrate both aspects. It would be unrealistic in a micro-prudential stress test to ignore the effects banks have on each other. Likewise a risk assessment for the banking system as a whole that ignored the details of what happens to individual banks in a stress would also be of limited use. At the end of this section we give a brief and high level discussion of the Swedish Banking System and the current division of responsibilities between the Riksbank and Finansinspektionen.

An overview of international best practice Section 3 gives a brief overview of international best practice in stress testing. It also describes the historical context. We stress the conceptual similarities between the approaches taken at different institutions. This section also contains a few brief case studies. The IMF is interesting because of its pioneering role and its international scope. The Bank of England is of particular interest because of its unique approach of counter-cyclical scenario design. It is also interesting for its integrated approach to micro- and macro stress testing. The US Federal reserves seems to be the most advanced institution in scope and depth of the stress test. Finally we discuss the Austrian central bank, mainly because of its unique and advanced process organisation and its adherence to modern technical design principles.

Stress testing frameworks: From basic to advanced Section 4 contains a conceptual discussion of stress tests. We are interested in the big picture as well as in the most important details. With respect to the big picture we distinguish two generations of stress testing approaches. The first generation works with the concepts of risk factors, scenarios and loss functions. In the stress test only one or two adverse scenarios are used and the plausibility of scenarios is not quantified. All stress tests used in practice today are of this kind. The second generation of stress tests, so far exists mainly in the academic literature but not yet in practice. In this approach the plausibility of scenarios is quantified and many scenarios are analysed in a systematic way. The literature on second generation stress tests is developed far enough to allow for a practical application. With respect to the details we discuss the necessity to take fire sales mechanisms and possibly contractual networks into account when evaluating losses. The focus in this section is descriptive. The discussion is deliberative, balancing trade offs and pointing out different options of modelling.

Interpreting the ouput of a stress test in a meaningful way In Section 5 we discuss the interpretation of stress testing output. In particular we concentrate on the concept of stressed balance sheets, the interpretation and representation of worst case scenarios, the concept of key risk factors and the representation of system wide stress.

Designing a stress testing system from scratch: Framework, priorities and resources In Section 6 we make a specific proposal what in our view would be the core elements of a modern stress testing framework. Our proposal rests on the following key elements:

1. *The automated evaluation of a large number of scenarios* reduces the risk of missing out dangerous but plausible scenarios. All current stress tests use one or two stress scenarios and one baseline scenarios. There are no procedures in place to check whether other, equally plausible scenarios might be more dangerous.

2. *The quantification of the plausibility of scenarios* prevents consideration of highly implausible scenarios. Since all stress tests used by authorities at the moment avoid the quantification of the probability that particular scenarios might occur, they are prone to both false alarm concerning events which have severe consequences but are extremely unplausible.

3. *The use of mixed scenarios* allows for a unified treatment of trading book and banking book.

4. *Modelling fire sales* allows for a realistic quantification of the consequences of bank reactions to stresses. Despite the known fact that the

most important mechanism by which stress manifests itself in a crisis are deleveraging processes, stress testing frameworks in place fail to take account of this fact.

5. *Modelling of network contagion effects* allows for a realistic quantification of consequences of bank defaults. Current stress testing frameworks ignore contractual interconnectedness. This impairs the credibility of the stress tests.

In summary our proposal suggests a modular design, which allows a framework with systematic scenario evaluation and incorporation of systemic risk analysis. Yet a key idea of our proposal is to construct the framework such that the current best practice and the more traditional methods in use at the moment can be embedded as special cases. We consider this property key for allowing both institutions to simultaneously fulfil their concurrent duties while at the same time investing in further development of methodology and improvement of process.

We suggest a priority ranking of the different elements of our proposed framework. Finally we give a rough estimate of resources that would be required to develop the framework both in the setup phase as well as in the steady state.

1 Introduction

Since the crisis of 2008 stress testing has gained a central role in financial risk assessment by authorities. Today the results of stress tests have strong policy impact.

Since stress testing is now so much more important, authorities all over the world face the challenge of rethinking their frameworks. They must judge whether they are well adapted to the new and more prominent role of stress tests. The discussion is still very much in progress. This is documented - for instance - by the fact that such an old and well established institution as the Bank of England, which has always been at the frontier of risk modelling and financial stability analysis, presented a new stress testing framework as recently as 2015 (see [1]).

This study aims at supporting the Swedish central bank, Sveriges Riksbank, and the Swedish regulatory and supervisory authority, Finansinspektionen, in their assessment of their current stress testing frameworks. The study also develops ideas and proposes options how these frameworks might be developed in a new way so that they can meet the future challenges. These challenges are: The integration of bank-oriented (micro) and more system-oriented (macro) perspectives, the challenge to assess plausibility and severity of stress scenarios in a credible way, the challenge to deal with second round effects and systemic risk issues and the challenge to interpret the output of a stress test in a meaningful way.

We begin by a discussion of the different purposes of stress testing in Section 2. In Section 3, we give a brief history of stress testing, discuss the core elements current stress testing frameworks have in common and present a few case studies of current best practice. In our next Section 4, we discuss some key conceptual issues of stress testing models, both as they are currently used and also with regard to key elements that should be included in a modern stress test but are currently missing. In Section 5 we discuss how both traditional and advance stress tests can be interpreted and how their results might be best represented. In the last Section 6, we make a specific proposal for a modern stress testing framework, which addresses some important new issues absent from many current stress tests. Yet our proposal allows for embedding more traditional stress tests as a special case within the new framework.

2 The different purposes of stress testing

2.1 Stress testing vs. statistical risk measurement

Solvency stress testing for banks is an idea stemming from risk management. Its aim is to identify future circumstances that would be extremely damaging to the current portfolio and to quantify losses that might occur under such detrimental scenarios. Let us call stress tests which do not aim at quantifying the plausibility of scenarios, stress tests of the first generation. Stress testing is a special case of risk measurement. It is a thought experiment with the basic question: Which scenarios lead to big losses? The answer to this question suggests risk reducing actions.

By a a stress test, the resilience of a bank or a banking system to severe but plausible shocks can be analysed. Knowledge of the dangerous scenarios can provide a foundation to necessary adoptions to the bank portfolio that might be needed to have a more resilient risk position. Thus stress tests can support both, risk measurement and risk management.

In contradistinction, statistical risk measurement estimates and interprets the profit/loss function. This answers the question which losses can occur with which probability. Statistical risk measurement does not aim at providing information about the circumstances (scenarios) under which these losses may arise.

2.2 Micro- and macro-prudential stress tests

In the current practice stress tests usually have two purposes, micro-prudential and macro-prudential. Micro-prudential stress tests assess the risk of individual banks. A stress test with this purpose supports the analysis of whether banks need to strengthen their capital positions. Macro-prudential stress tests assess the overall stability of the financial system. In this perspective the analysis of system vulnerability to detrimental scenarios, the overall adequacy of the capital framework in place as well as the widespread impairment of the banking system's key functions are in the focus. In crisis management, stress testing can fulfil an additional purpose. In this situation a stress test can support a broad bank restructuring program and help to restore confidence among banks and among banks and the public (see [27]).

The separation between the micro- and the macro-prudential is, however, somewhat artificial. It would be unrealistic in a micro-prudential stress test to neglect the effects banks have on each other. Banks have an impact on each other, both via markets and via cross-holdings (Sections 4.3.1 and 4.3.2). While the two purposes, risk measurement and capital calculation for individual banks and assessment of the financial system, provide a slightly different context for the stress test the differences between these two purposes should not be exaggerated. In fact they are really two sides of the same problem.[1]

The purpose is important to the question where the resources in a given

[1]However, in systematic stress tests (Section 4.2), which try to identify worst case scenarios, an important difference between macro-prudential and micro-prudential stress tests may arise. The objective function, which describes the criterion for what constitutes a worst case, in a macro-prudential stress test may be some property of the banking system, while in a micro-prudential stress test it would be some property of the bank in question.

stress test framework are concentrated and what the key modelling trade offs are. For example a key problem in designing a good stress test is to find plausible but severe scenarios and to make sure that no relevant scenarios are ignored. This might require to analyse many scenarios not just a few.[2] In order to be able to analyse many scenarios it might be advantageous to use a less detailed model. But then the capital calculation for the individual bank might be too coarse. To give another example: If the purpose of the stress test is overall systemic stability assessment, the interactions between different banks in a stress scenario will be a very important aspect of risk measurement. In this case the resources in a stress test that go into the modelling of interactions, amplification mechanisms and systemic effects will become relatively important.

2.3 The Swedish institutional setup

The different responsibilities In Sweden, it is the Riksdag (the parliament) and the government that decide on regulations, laws and statutes and thus have ultimate responsibility for the ground rules of the financial system. The responsibility for safeguarding financial stability and maintaining an effective financial system has, however, been divided between the Government (through the Ministry of Finance), Finansinspektionen, the Riksbank and the Swedish National Debt Office.

To ensure financial stability the roles and responsibilities of the authorities are divided as follows: The Ministry of Finance is responsible for the regulation of financial markets and has the overall responsibility for crisis management. Finansinspektionen is responsible for the supervision of financial companies and macroprudential policy. The Riksbank is responsible for a central payment system and providing liquidity in the system (lender of last resort). The Swedish National Debt Office is responsible for the deposit insurance scheme and the preparedness and the management of institutions in distress (the resolution authority).

The Swedish banking system The Swedish banking system is large in relation to the size of the Swedish economy. At the end of 2015, the size of Swedens banking sector amounted to 340 per cent of Swedish GDP. The Swedish banking system is characterised by four dominating banks that together account for almost 80 per cent of the lending and receive almost 67 per cent of deposits in Sweden. The exposures between these four banks (Handelsbanken, Nordea, SEB and Swedbank) are significant as the cross-ownership of each others securities correspond to approximately

[2]This specific point about scenario selection is for instance discussed in depth in the recent report of the Government Accountability Office in the US to the chairman of the committee of financial services of the house of representatives, reviewing the stress testing framework of the Federal reserve. See in particular [29, p. 63].

30 per cent of their total equity. Another characteristic is that banks to a large extent rely on market funding, partly because Swedish households tend to prefer to save in mutual funds and stocks rather than in deposits. Approximately half of their lending is funded via securities and half via deposits. Furthermore, about two thirds of their market funding is in foreign currency. Thus, the Swedish banking system can be characterized as large, concentrated, interconnected and dependent on market funding.

3 An overview of international best practice

3.1 A brief history of solvency stress testing

Early stress tests of trading books Prior to the financial crisis of 2007/2008 solvency stress testing was mainly used within large international banks as part of their overall quantitative risk management tool kit. This development started in the early 1990ies (see [17] for a detailed historical overview). Initially the only domain of stress tests within these institutions was the bank's trading book. These early stress tests tried to evaluate the profits and losses of the trading book under a set of scenarios formulated in analogy with extreme past market events such as a stock market crash, a sudden big exchange rate movement or similar extreme events. These early stress tests did not follow a standard pattern and varied considerably between different institutions.

With the Basel Committee's amendment to the capital adequacy regime for the trading book of large internationally active banks in 1996, a first step towards standardisation or a best practice benchmark in stress testing emerged. The amendment allowed banks to use their internal risk models to determine the amount of capital held against the exposure in the trading book. The models themselves were examined by the supervisor. Within this regime banks were required to implement a system wide stress testing program for market risk in the trading book.

Credit stress tests of the banking book Stress tests that included also credit risk followed with a considerable lag mainly during the Basel II process in 2004. Following the practice of market risk, banks that followed the internal ratings based approach (IRB) of Basel II were allowed to use their internal credit risk models to calculate capital requirements. Now not only the trading book but also the loan portfolio in the banking book were subject to a stress test. Since banks could choose to assess credit risk alternatively to the IRB approach using the so called standard approach and since Basel II was not universally implemented by advanced economies before the financial crisis, there was up to that point no universally accepted common best practice of stress testing for credit risk.

Financial stability analysis Stress testing of banks developed simultaneously within central banks and international policy institutions as a tool of financial stability analysis. In particular spurred by the Financial Sector Assessment Program of the IMF (FSAP) launched in 1999, stress testing became a key component of financial stability analysis. The focus of these stress tests were a quantification of the vulnerability of a country's banking system to extreme macroeconomic and financial scenarios. The IMF's Financial Sector Assessment Program became a standard setter of stress testing best practice in central banks.

Before the financial crisis, the stress testing models and procedures that evolved out of the FSAP program, usually had no direct policy impact. Their outputs were sometimes included in the central bank's financial stability report. At this stage a rough concept of stress testing had been established. Despite a very active development and many amendments, the basic concept remained in place ever since then.

With the financial crisis of 2007/2008 stress tests gained a much more prominent status: They became huge, systematic risk assessment programs with direct policy responses. The start of this step-change was the US Supervisory Capital Assessment Program (SCAP) conducted by the Fed in early 2009.

At the hight of the crisis both banks and supervisors had to face a severe credibility crisis: Banks lacked credibility because there was general uncertainty about their true financial situations, supervisors lacked credibility because the scenarios used in stress testing exercises before the financial crisis were much more benign than the scenario that materialized during the crisis itself (see [34]). The SCAP stress test was undertaken to restore the credibility of both, banks and supervisors: The largest US banks were subjected to a severe common stress scenario and it was assessed whether they had sufficient capital to absorb the losses in this adverse scenario. The results were disclosed on an individual bank basis. Banks who failed this stress test had to be recapitalized via the capital market within six month or get recapitalised by a backstop from the Treasury. In terms of restoring confidence and creditability the SCAP was very successful.

This success encouraged a similar exercise in the European Union conducted under the Committee of European Bank Supervisors (CEBS) in 2009. This was later followed by stress test exercises conducted under the direction of the European Banking authority (EBA) from 2011 onwards. The success of the European exercise was much less pronounced than in the US. However the EBA stress tests have induced a common stress testing practice both in banks and at the level of supervisory authorities and became a common best practice model.

Through the substantially raised profile of stress testing, there are now considerably more resources devoted to this activity. Also the policy impact is more pronounced and formalised within standard processes. The basic conceptual paradigm of the procedure of scenario formulation and portfolio function evaluation at one or two adverse scenarios is still common practice, which we next discuss in form of a few case studies.

3.2 Some examples of current stress test practice

3.2.1 The IMF

The IMF through its Financial Sector Assessment Program (FSAP) is one of the pioneering institutions of stress testing by public authorities. The stress tests are undertaken to support financial stability assessment within the FSAP in various countries. The IMF usually works with two or three adverse macroeconomic scenarios. The specific risk factors are chosen on the basis of significance for the economy under consideration and data availability.

Initially the IMF-FSAP focussed on bank solvency stress tests. In the meantime it has aimed to broaden stress tests by including other financial institutions beyond the banking system, such as insurance companies or money market funds.

Sometimes, the IMF-FSAP also analyses risks from contractual interconnectedness between banks, between banks and non-banks and between domestic and international economies, see Section 4.3.1. The IMF is also actively developing models for assessing amplification mechanisms and second round effects (see Section 4.3.2). Unlike the analysis of interconnectedness. This further aspect of systemic risk is not yet implemented. The evaluation of losses is produced through a hybrid procedure involving both computations by authorities as well as by the banks themselves. The results are used to provide supporting information both to microprudential- as well as to macroprudential policy.

The IMF is very active in research and development of new modelling approaches an tools in this area. It has recently published a book to document the results of this institutional research effort, see [30].

In terms of our previous discussion the IMF framework can be viewed as a first generation stress test with a rich and sophisticated suite of models to evaluate losses. It usually analyses only few stress scenarios with the exact number varying between countries. In particular the IMF has invested great effort in modelling network and interaction effects in the generation of losses and into a more comprehensive view of the financial system going beyond the banking system.

3.2.2 The Bank of England

The Bank of England ran its first concurrent stress test in 2014. In 2015 it published a new stress testing framework (see [1]), which is applied from 2016 onwards. The stress test is applied to all banks with a volume of more than 50 billion GBP. The stress testing framework at the Bank of England has a process of scenario formulation different from most stress testing frameworks used in practice: The scenario will depend on the state of the financial cycle.

This new approach takes into account one issue with stress testing which has been raised occasionally in the general discussion and has been formulated by the BIS in particular [2]. In that view stress tests are seen as a poor instrument of gauging the resilience of a banking system under stress because it seems to ignore the dynamics of the financial cycle, which is described by the BIS as the cyclical co-movement of credit growth and asset prices. In the BIS view a financial crisis is usually not the result of a big exogenous shock to the real economy. The situations is rather that during a financial boom, periods of strong credit growth and high asset prices create a fragile financial system which precipitates a recession in the real economy. This triggers a dynamics of amplifying crisis feedbacks between the real and financial sector. According to the BIS, a framework that thinks in exogenous shocks hitting the financial system, ignores the financial cycle and the nature of financial instability.

The Bank of England's stress testing framework takes this critique into account by designing scenarios depending on the state of the financial cycle with the declared intention to produce a countercyclical stress test: In the upswing of the cycle, scenarios are more severe than in the downswing. The stress test is designed to be intentionally countercyclical. The scenarios are formulated in terms of a comprehensive macro scenario. In terms of the number of scenarios, however, the framework considers only one to two scenarios like in a more traditional scenario design. The frequency of the stress test is annual.

On top of the cyclical scenario design, the Bank of England uses a biannual so called exploratory scenario to probe other risks which are deemed important but are only indirectly or remotely connected to the financial cycle.

With respect to loss evaluation at a stress scenario, the Bank of England uses a dynamic balance sheet model. This means that bank reactions and asset repricing are taken into account in the modelling and evaluation of losses. This can be seen as an explicit account of second round effects of the stress scenario impact. The process of loss evaluation uses both bottom-up evaluations of banks as well as top down evaluations of banks and authorities.

The stress test is used by the macro-prudential authority, the Financial Policy Committee (FPC) as well as the micro-prudential authority, the Prudential Regulation Authority (PRA) to determine whether the system as a whole is sufficiently resilient as well as whether banks need to adjust their capital positions. A detailed account of how the stress test contributes both to the FPC's as well as the PRA's objectives can be found in [1, Box 1] .

In the scheme suggested previously the Bank of England's framework can be viewed as a first generation stress test. Its approach to scenario formulation is innovative and can be seen as a heuristic attempt to take into account the state of the financial cycle in formulating stress scenarios. It uses sophisticated modelling of losses that take into account bank reactions and repricing of risks. Another interesting aspect of the Bank of England's framework is that it uses an integrated approach for both micro- as well as macro-prudential policy support.

3.2.3 The US Federal Reserve

The US Federal Reserve has two related stress testing programs with different purposes. The Dodd-Frank Act requires statutory stress tests for Federal Reserve-supervised banking institutions with more than 10 billion USD in total consolidated assets. These stress tests are also known as Dodd-Frank-Act-Stress Tests (DFAST). DFAST apply to a broad range of banking institutions and consist of supervisory- and bank-run stress tests that produce capital adequacy information for banks' internal use and for public disclosure.

Additionally the Federal Reserve also conducts the Comprehensive Capital Analysis and Review (CCAR). The CCAR uses information from DFAST stress tests and conducts a quantitative assessment of capital adequacy and capital planning for bank holding companies with total consolidated assets of 50 billion USD or more. CCAR generally does not require additional stress tests and uses the same data, models, and projections used for DFAST.

The primary purpose of DFAST is to produce and disclose information about banks, the CCAR supports supervisory assessment of capital adequacy and capital plans of banks. Its purpose is therefore mainly to support microprudential policy. It may, however, also support the macroprudential policy setting.

The Federal Reserve seems to be the internationally most advanced institution in terms of modelling methodology and model validation. The Fed has a formal quality control and validation process for the models used in the stress test. An independent model validation team reviews all models with a focus on design, estimation and implementation. Model reviewers are drawn from Federal Reserve staff not involved in the model development. Additionally the Federal Reserve has implemented process control procedures surrounding model design and implementation see [29, pp. 66-68].

The Federal Reserve framework can conceptually be regarded as a first generation stress test with sophisticated modelling of losses, since bank reactions and repricing strategies are taken into account in modelling losses. The frequency of the stress test is annual. The number of scenarios used is two. For the evaluation of losses both bottom-up calculations by banks as well as top down evaluations be authorities is used. The stress tests have a direct policy impact because they are the basis for assessing and approving capital adequacy and capital planing of large bank holding companies. The entire stress testing framework was recently assessed by the US Government Accountability Office (GAO) (see [29]). Among the suggestions for improvement are measures to make scenario design more systematic and to develop procedures to deal with model risk.

3.2.4 The Austrian Central Bank

The Austrian Central Bank (OeNB) uses a stress testing framework that is designed such that it can easily be aligned with the stress testing methodology of the European Banking Authority (EBA) [20], the institution running the Europe wide concurrent stress tests since 2011. At the same time the framework strives to include specific risks to the Austrian banking sector. In terms of scenario generation this means that it mainly works with one macro scenario at a biennial frequency. The stress testing model is also used for ad hoc analysis on top of the EBA stress testing cycle.

The Austrian Central Banks Stress Testing framework contains models which are designed to capture systemic risk at the level of contractual interdependencies between banks, see Section 4.3.1. It also provides tools to conduct liquidity stress tests. By a modular design the model is constructed from building blocks of composable modules. Each module can be used or omitted, depending on the particular focus of the stress test. This means that - for instance - the contagion or liquidity module can be switched on or off during a particular run of the model.

The stress testing framework of the Austrian Central Bank avoids the differentiation between a macro- and a microprudential stress test. It implements a bank by bank approach making heavy use of supervisory reporting and avoids projections of aggregates. The Austrian framework can be viewed as a first generation stress test with sophisticated loss modelling, including network and liquidity considerations.

One important practical aspect in which the Austrian framework seems to stand out, is the technology based process organisation of the stress test. Given the complexity of the task and the high frequency of stress testing demands and exercises this is a key aspect of a modern stress testing framework.

The stress testing framework is constructed according to some basic technical design principles. The first principle is automation. All tasks and computations between raw data to the presentation of results are performed by a single script. The second principle is modularity. Each stress test may use the modelling components only partially and each module is independent. The third principle is configurability: Every run of the stress test depends on a configuration which is stored in a separate file. The fourth principle is data abstraction. This means that the stress testing model fills a generic data structure which is stored separately from the raw data and is the sole input source for all computations. The stress test uses a modern version management. Every run of the model is stored separately together with a complete copy of the code, data, configuration and the scenarios used.

Following these principles allows to reduce operational risks and frees up resources needed for the analysis of results.

4 Stress testing frameworks: From basic to advanced

4.1 First generation stress tests: Portfolio value functions, risk factors, scenarios

Most stress tests share a common conceptual structure. This structure is used so widely that it could be termed with some justification, first generation stress tests. The different frameworks which are in place in practical stress testing processes today are all variants of this basic framework. It is a framework with three basic objects: portfolio value functions risk factors and scenarios.

Mark to market The value of tradable positions is determined by the market. The portfolio value function P at some future time horizon, t_1, depends on some variables r which are called risk factors. A scenario is a simultaneous realisation of all risk factors at time t_1. Call the set of scenarios Ω. The risk factors r can be regarded as a coordinate system in Ω, and Ω could be interpreted as a state space. The portfolio value function

$$P_T(r) \tag{1}$$

specifies the profit or loss of the tradable portfolios as a function of the risk factors. The functions P_T encode the decisions the various banks take themselves, whereas the the risk factors r cannot be influenced by the banks, see Fig. 1. The value of tradable positions at a future time t_1 is unknown now, but it will be determined by the markets and in this way will be known at t_1.

Risk factors So the risk factors form a set of random variables fulfilling three conditions: First, their values are observable directly or indirectly, second their probability distribution cannot be influenced by the respective bank. These two conditions ensure that the risk factor distribution can be estimated from past realisations. The risk factor distribution is required if the plausibility of scenarios is quantified. The third condition is that the set of risk factors is the same for all banks in the model. (Still the value function of some bank might be independent of some of the risk factors.) This condition is required for systemic stress tests to be possible. The example of exchange rates illustrates the three requirements: Exchange rates are directly observable, their distribution cannot be influenced by a single institution, and they are risk factors for all banks.

Modelling the controllable and the uncontrollable The two main elements of any risk model are the portfolio value function of each bank modelled and the distribution of risk factors. The two elements separate the controllable and the uncontrollable aspects of reality, see Fig. 1. The risk factors describe the aspects not controllable by the bank, i.e. events in the environment of the bank, including decisions of market participants and also of other banks. The value function of a bank describes the decisions taken by that bank.

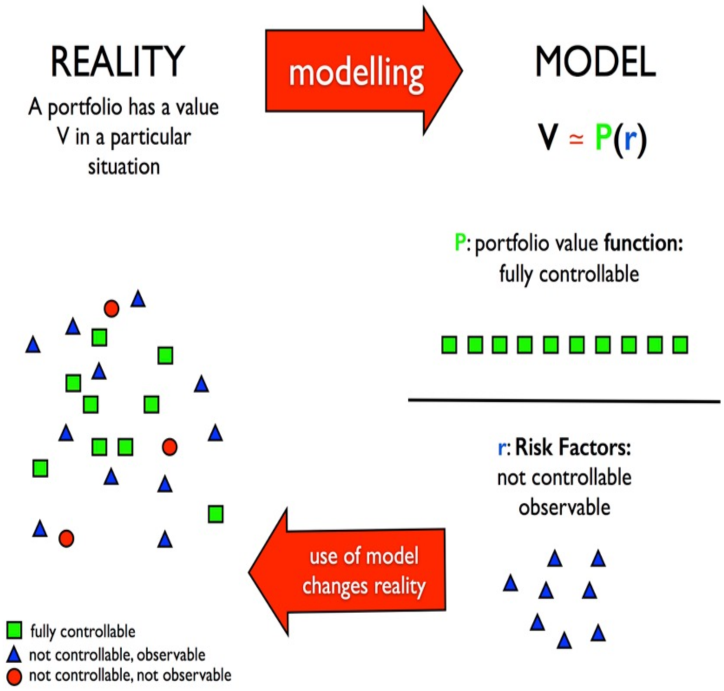

Figure 1: The relation between reality and model. Reality is a compound of three kinds of aspects. Some aspects are controllable by the decision maker. Others are observable but uncontrollable, for example example because they are consequences of events in nature or of decisions by others. Finally there are some unobservable and uncontrollable aspects. A model of reality typically neglects aspects unobservable and uncontrollable, and separates the controllable from the uncontrollable aspects. In the context of financial institutions the controllable aspects are described by the portfolio value function P. The uncontrollable aspects are described by the risk factors r. For economic systems, the models in turn can change reality when market participants use them for decision making. In physical systems typically there is no feedback from the model to reality.

Illiquid positions Apart from the tradable positions, there are positions which mature after t_1 but which are not traded. (Obviously there are shades of grey between the two types, reflecting various degrees of liquidity.) The value of such illiquid positions is not uniquely determined by the observed risk factors, but has to be modelled. In the actuarial framework, the book-value of such positions is modelled as expectation of future payoffs with respect to some probability distribution:

$$E_{Q_I}(P_I), \tag{2}$$

where Q_I is the conditional distribution of risk factors s influencing the value of illiquid positions given the information at t_0. $P_I(s)$ is the total value of the illiquid positions. In actuarial models the value of positions changes if the distribution Q_I changes. We call alternative distributions Q_I *mixed scenarios*. Changes of the distribution may be due to new information arising, or when estimation errors occur, or if the distribution class is misspecified. In any case, the value at t_1 of illiquid positions is not determined by risk factor realisations at t_1 but by a probability *distribution* at t_1.

An example such positions are loans in the banking book. If the loans are not traded, the value of the loan portfolio is modelled as expectation of future payoffs, which in turn is determined by probabilities of default (and LGDs and exposures at default). The actual future payoffs of the loan portfolio can be different from the expected payoffs, which are used in model-based pricing. They are given by the actual profits $P_I(s)$, where the scenario s specifies both the macroeconomic and the idiosyncratic risk factors determining the payoff of each loan.[3]

Mapping of the actuarial framework to the mark to market framework In order to reestablish the standard framework (1), the expected value of illiquid positions (2) is rewritten as a function of risk factors. Starting from a distribution class assumption about the distribution Q_I of risk factors influencing the illiquid portfolio, Q_I is represented as a function of some distribution parameters θ. Different distributions Q_I are then regarded as different realisations of θ. (For example, if $P_I(s)$ is the distribution of profits from an illiquid loan portfolio, one often takes the PDs and LGDs of the loan portfolio as risk factors θ.)

Writing $E_{Q_I(\theta)}(P_I)$ as $P_I(\theta)$ transforms the actuarial evaluation (2) to the traditional representation (1) as a portfolio value function depending on risk factors. But this transformation is subject to distribution class assumptions. If the true distribution Q_I of the risk factors of P_I is not from the assumed distribution class, any representation as $Q_I(\theta)$ is mistaken.

[3] If the loan portfolio is fine-grained (no exposure accounts for more than an arbitrarily small share of the total exposure) and LGDs are independent (conditional on the macroeconomic variable), the distribution of the portfolio loss ratio converges to the expected loss ratio, see Gordy [25]. Furthermore, the q-quantile of the portfolio loss ratio converges to expected portfolio loss ration conditional on the q-quantile realisation of the macroeconomic variable.

First generation stress tests A first generation stress test could be described by two key features: First, it evaluates the portfolio value function $P(r)$ only at very few scenarios. The number of stress scenarios considered is usually two. The set of risk factors describes the relevant risk drivers and their potential future values affecting the value of a given portfolio (or financial institution, or system of financial institutions.)

Second, the statistical distribution of the risk factors plays no formal role. Although stress tests designers take the plausibility of scenarios into account in an informal way, plausibility is usually not quantified rigorously. This has the advantage that results of such stress tests are independent of the risk factor distribution, and immune to the modelling or estimation errors involved in the determination of the risk factor distribution. Most of the stress testing frameworks which are in use today are of this type, see our discussion in Section 3.

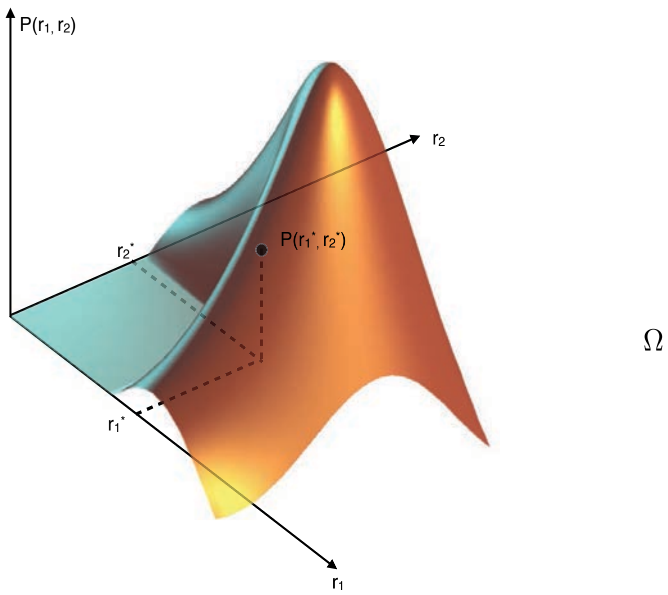

Figure 2: The standard stress testing framework. The space of risk factors Ω in this stylised picture would be the plane spanned by r_1 and r_2. The portfolio values are described by the portfolio function defined on the plane with values along the vertical axis P. If the stress scenario would be described by the coordinates (r_1^*, r_2^*) in the space of risk factors, the corresponding stressed portfolio value would be given by $P(r_1^*, r_2^*)$.

The evaluation of losses at only a few scenarios and the abstinence from probabilistic modelling of risk factors is a key characteristics of first generation stress tests. Still within this approach there a lot of details that need attention. We discuss a few important ones in the remainder of this subsection on first generation stress tests.

Multi-period risk In a multi-period model the profit/loss functions are multi-variate, each component describing the profit or loss at a different time. In a dynamic balance sheet context the portfolio function at each time depends on the risk factor realisations at earlier times and on the strategy the bank chooses to react to new information emerging at intermediate time steps. In a static balance sheet approach one assumes that the portfolio remains unchanged over the time horizon; positions maturing before the time horizon are assumed to be replaced by similar new positions. Assuming no reaction by the bank is a strong behavioural assumption, which might be wrong in a stress event.

The dynamic balance sheet assumption has the advantage of offering the more realistic option that banks react to stresses. But it faces the danger that assumptions about reactions may be wrong. Both, static and dynamic balance sheet approaches make explicit or implicit assumption about the behaviour in crisis, which might be wrong.

In a multi-period framework a decision has to be taken for which time horizon a credible quantification of stress impact can be made. Most existing frameworks seem to be quite arbitrary in their choice of event horizon and vary substantially among each other in this respect. This is relevant both with respect to the precision estimates as well as the credibility of behavioural assumptions. While it might be a good approximation to assume a forced deleveraging in a stress event at a short horizon, as the time horizon grows, behavioural options increase and it is increasingly hard to model what banks might do.

Profit function and Balance Sheets Profits and losses of banks are reflected in their balance sheets. Working with balance sheets opens the option of analysing stress test results not just in terms of profits/loss, but also in terms of capital, risk weighted assets, and impact on subcategories of balance sheets. For systemic stress tests the respective quantities can be added up across all banks in the system. Analysing stress test results in terms of balance sheets requires to model the various balance sheet entries as function of the risk factors.

Systemic stress tests A systemic stress test analyses the banking system (or part of it), not just an individual bank. It requires that all banks are subject to the same scenarios, and that all banks evaluate the impact of scenarios in at least roughly comparable ways. Otherwise one cannot combine stress test results of different banks. These two conditions can be naturally fulfilled only in top down stress tests, where the scenarios and the portfolio evaluation function are centrally defined.

Choosing the same set of stress scenarios for all banks in the system might prove politically tricky. A scenario which is severely negative for one bank might be harmless for another bank. Such discussion can be avoided if not only a handful of scenarios is considered, but a large set of scenarios covering the relevant parts of the scenario space. What are the "relevant parts of scenario space"? This issue is addressed in Section 4.2.

The difficulty and at the same time the potential merit of systemic stress tests is to capture to interaction between banks. These second round effects will be discussed in Section 4.3.

Public vs. Supervisory Data Basing stress tests only on public data has the advantage of transparency, if stress test methods are publicly known. Transparent stress tests based on public data which theoretically can be checked by any market agent, probably enhance the market disciplining effect of stress tests.

On the other hand, supervisory data is much richer than public data. This might allow the use of more detailed models which require more detailed data. Whether more detailed models are indeed better models is another question. Simpler models allow for fast automated scenario evaluation, and therefore for the consideration of many scenarios.

Macro variables as risk factors? The choice of risk factors depends on the context of stress tests. For example, some stress test approaches include macroeconomic variables as risk factors, others do not. In case they are included it is necessary to model the influence of macroeconomic variables on the portfolio value.

In the EBA methodology [20] the values of macroeconomic variables over the next three years are included as risk factors. To make explicit the dependence of the future portfolio values on future values of macroeconomic variables, so-called satellite models translate the values of macroeconomic variables into bank-specific risk parameters like PDs, LGDs, exchange rates etc., see Fig. 3. In a second step, the profits or losses in dependence of macroeconomic risk factors can be calculated from the values of the risk parameters.

As an alternative to the EBA approach, one could take the risk parameters directly as risk factors. Then the stress scenarios would be formulated in terms of these risk factors. Macroeconomic variables would not occur explicitly in the model. One challenge faced in this alternative approach is to have the same risk factors for all banks, so that systemic stress tests can be performed.

Another alternative is to model the link from macroeconomic risk factors to PDs and LGDs centrally, as the IMF does in its FSAP. Then all banks

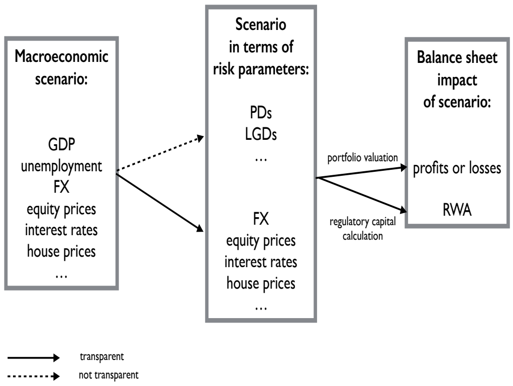

Figure 3: The EBA methodology calculates the balance sheet impact (profit/loss) of a macroeconomic scenario in two steps. First, the change of the macroeconomic factors is translated into changes of the risk parameters; second the changes of the risk parameters determine profits or losses. In the EBA methodology the first step relies on internal models of banks.

are required to use that model for the translation of macro risk factors to the portfolio risk parameters.

There are arguments for and against the use of macro risk factors in stress tests. On the one hand the systematic risk of the banking sector depends on the macroeconomic situations. To describe this it is necessary to have a macro model. Additionally, the use of macro risk factors reduces the number of dimensions, since there are usually three or four macro risk factors compared to PDs and LGDs for many subportfolios.

On the other hand, one could use a fixed model for the link between macroeconomic risk factors and PDs, as e.g. in CreditRisk+, and formulate the macroeconomic situation directly in terms of PDs and LGDs. Both options should ensure that stress test results are comparable across banks.

An additional issue with respect to macro-modelling in stress tests is related to process organisation. In current practice, stress testers have to turn to the central bank's macro model for the generation of macro scenarios. This creates obvious obstacles for stress testing with more than a few scenarios. Furthermore central bank macro models are usually geared towards monetary policy and are not always particularly suitable for stress tests. If macro risk factors are included it might be worthwhile to think about having a suitable macro model tailored to the purpose of stress testing as an integrated part in the stress testing framework.

Separation and integration of market and credit risk Most stress test procedures deal separately with market and credit risk, adding up the capital requirements for market and credit risk. There are claims that this approach is conservative in that it does not take into account possible "diversification effects" between the two risk types. Such claims are not true in general. Detrimental interaction between market and credit risk might lead to a mutual reinforcement of market and credit risk, see [10].

An integration of market and credit risk is most naturally achieved in firm value models of credit risk: Payment obligations may change because of changes in the market risk factors like exchange rates and interest rates. Payment ability is linked to the firm value, which may be an observable market risk factor for publicly quoted companies. Details of such an integrated approach are in [10]. A systemic stress test along these lines would require data about the firm value of each counterparty of each bank in the system. This is not realistic. In an simplified aggregated model one could form pools of counterparties, for example along economic sectors. From the sector PDs one can calibrate "sector values", which serve as risk factors.

Criticism First generation stress tests can be criticised for two reasons. First, the plausibility of scenarios is not quantified. If stress scenarios are highly implausible, it may be hard to justify action based on alarming stress test results. Second, considering only a handful of scenarios might give a false illusion of safety. A bank which fares well in the handful of selected scenarios might still fail in other scenarios, which have not been considered, but which might be equally plausible.

4.2 Second generation stress tests: Plausibility and Completeness

We call stress tests which explicitly provide concepts allowing to quantify the plausibility of scenarios, second generation stress tests. With the exception of some very specialized areas of market risk management in large banks or the academic literature, we are not aware of any stress tests currently in use by central banks or regulatory authorities, which are of this kind. We believe, however, they should conceptually play a larger role in stress testing. We therefore discuss second generation stress tests in some detail here.

The plausibility of Stress Scenarios The main idea of second generation stress test is to consider not only a handful of scenarios, but *all and only scenarios which are sufficiently plausible.*

Plausibility concepts require an additional element in the stress testing framework, in addition to the risk factors and the value function: a probability distribution of the risk factors. The plausibility of a scenario depends on the risk factor distribution. A given scenario might be more plausible under one risk factor distribution than under another. So by the requirement of plausibility the risk factor distribution enters through the back door, in spite of the fact that one of the original intentions of stress tests was to do away with distribution assumptions. But traditional stress tests pay a high price for doing away with the risk factor distribution: it is impossible to quantify the plausibility of scenarios.

A measure of plausibility for point scenarios For mark to market valuations, see (1), we consider point scenarios, i.e. realisations of the risk factor distribution. A natural measure of plausibility of scenarios is the Mahalanobis distance of that point in sample space.

Intuitively, Maha(r) can be interpreted as the number of standard deviations of the multivariate move from the present state to r. Maha takes into account the correlation structure and the standard deviations of the risk factors.

It might not always be sensible to measure the plausibility of point scenario by Maha. Maha takes into account only the expectation and the covariance structure of the risk factor distribution. Higher moments, the dependence structure, and the tail shape of the risk factor distribution are disregarded. Also, Maha is not appropriate for discrete risk factor distributions.

Highly implausible scenarios can safely be excluded from the scenario set. This has two advantages. First it reduces the number of scenarios to be evaluated. Second it supports interpretation of stress test results. If the stress test results in a highly implausible scenario are alarming, banks and supervisors have to face the question, whether such a result justifies or requires action. This issue is addressed by second generation stress tests.

Plausibility of mixed scenarios In the actuarial framework, the book values of non-traded positions (2) depend on the risk factor distribution. We measure the plausibility of a mixed scenario by its relative entropy with respect to some reference distribution, which could be interpreted as a prior distribution. The reference distribution often results from assumptions about a model class and a parameter estimation procedure based on historical data. Distributions close enough to the reference distribution are plausible alternatives. We admit as plausible enough all mixed scenarios for which the relative entropy does not exceed some threshold k.

Completeness of scenario sets One traditional criticism of stress tests is that they might nourish a false illusion of safety. When evaluating only a handful of scenarios it might easily happen that a bank looks safe in all scenarios considered, but that it could fail in other plausible scenarios, which were not considered. Trying to avoid discussion why some plausible scenarios are considered while other scenarios of similar plausibility are not, one would choose the scenario set to be equal to some set of sufficiently plausible scenarios, or at least some set covering the sufficiently plausible scenarios in some representative way.

So in a mark to market framework (1), the scenario set of a systematic stress test could be the ellipsoid of all scenarios with Mahalanobis distance smaller than some threshold, see Fig. 4. In the context of stress testing, this was also proposed by Breuer and Krenn [11] and Čihák [42]. For an application of this approach to stress testing of aggregate credit risk, see [5].

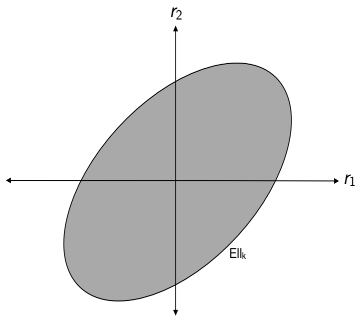

Figure 4: Scenario set: The set of all scenarios above some plausibility threshold formulated terms of Maha forms an ellipsoid in the sample space. The shape of the ellipsoid is determined by the variances and covariances of the risk factor distribution. The size is determined by the plausibility threshold. The figure shows as example a two-dimensional risk factor distribution.

In the actuarial framework (2) of the banking book, the scenario set of a systematic stress test could be the Kullback-Leibler sphere $\{Q : D(Q||\nu) \leq k\}$ of all distributions with relative entropy smaller than some threshold. Here ν is the estimated reference distribution. The "radius" k of the Kullback-Leibler sphere represents the plausibility threshold beyond which scenarios are not considered to be relevant. But within the sphere all scenarios are considered to be relevant.

Severity of scenarios Severity measures the impact of a scenario. For mark to market portfolios (1) we measure the severity of a point scenario r by $P(r)$.

Possible choices for severity are loss arising in a scenario, the resulting CET1 shortfall of a bank, or the capital injection necessary to save the bank in case it fails. For systemic stress test one would measure severity by the value of some function defined on the set all banks in the system. One could consider the sum of the losses of all banks, or the sum necessary to bail out all banks, or reduction in the sum of loans to the real economy made by all the banks which are still operative in the stress scenario, or the systemic impairment considered by Pritsker [33].

In the actuarial framework of the banking book (2) a natural measure of severity of a mixed scenario Q_I, is the expected with respect to that distribution.

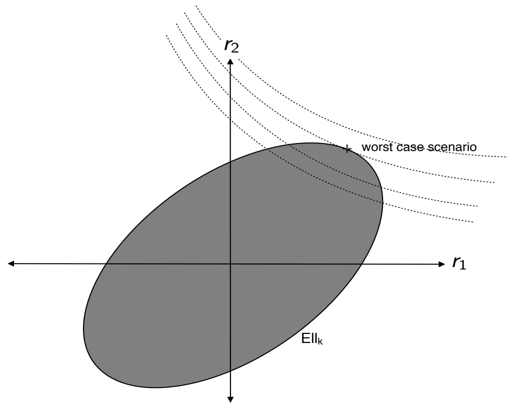

Figure 5: Systematic stress test with point scenarios: determine among the sufficiently plausible scenarios (those in the ellipsoid) the one with the maximal loss. Scenarios with the same loss are indicated by the dashed level lines.

Systematic stress tests: Plausible worst case scenarios A systematic stress test identifies the worst case scenario in the set of sufficiently plausible scenarios [9, 39]. This approach guarantees that two pitfalls of first generation stress tests are avoided. First no highly implausible scenarios are considered, which would impair the credibility of stress tests. Second all scenarios above the plausibility threshold are considered and no dangerous scenario is missed, thus avoiding a false illusion of safety.

For mark to market portfolios (1), finding the most severe in the set of plausible scenarios amounts to determining

$$\max_{r \in \mathrm{Ell}_k} P(r). \tag{3}$$

This is illustrated in Fig. 5.

In the actuarial framework of the banking book (2), a systematic stress test amounts to the worst case problem

$$\sup_{Q: D(Q||\nu) \leq k} E_Q(P) \tag{4}$$

A procedure for the solution of problem (4) can be found in [7]. Instead of relative entropy, other measures of distance between distributions could be used. Such generalisations of problem (4) are solved in [8]. The calculation is based on the evaluation of an integral over Ω, which in turn requires the evaluation of the value function P for a representative set of scenarios. The solution of problem (4) has the same computational demands as problem (3).

A unified framework for systematic stress testing with pure and mixed scenarios Whenever a portfolio contains traded positions and illiquid positions, both pure and mixed scenarios are used in the evaluation, see (1) and (2). For the time being assume that the traded positions are marked to market with a value function $P_T(r)$, where the market risk factors r are distributed according to Q_T, and the illiquid positions are evaluated with a model as $E_{Q_I}(P_I)$, where the risk factors of the illiquid portfolio have a distribution Q_I. The total portfolio value is

$$P_T(r) + E_{Q_I}(P_I).$$

Denote by ν the joint reference distribution of risk factors for the traded and the illiquid portfolio. For any joint distribution Q denote the marginal with respect to the traded risk factors as Q_T and the marginal with respect to the risk factors of the illiquid part as Q_I. A systematic stress test can then be conceived as

$$\inf_{Q:D(Q|\nu)\leq k} \left[E_{Q_T}(P_T) + E_{Q_I}(P_I)\right]. \tag{5}$$

This unified framework for systematic stress tests can be extended to models with overlapping risk factors, which influence both the traded and the illiquid portfolio.

Risk reducing action Stress tests aim at identifying possible risk reducing action. In contradistinction, statistical risk measurement assesses the profit/loss distribution and gives no hint as to possible risk reducing action if risks are unacceptably high, since no information is available as to which circumstances (scenarios) lead to the high losses.

Systematic stress tests are particularly appropriate for identifying possible risk reducing action. Knowledge of the worst case scenario above the plausibility threshold, as provided (3) and (4) allows for the design of risk reducing action. For the worst case scenario, one can identify as key risk factors those with the highest explanatory power, as in [6]: This is the set of risk factors which jointly have the highest contribution to the loss in the worst case scenario, see p. 33. A risk reducing position can then be designed as one with high payoff exactly when the key risk factors have their worst case values.

It can happen that risk reducing positions do not reduce the worst case by much. This will be the case if there are other scenarios, perhaps very different from the worst case scenario, which produce losses almost as bad as the worst case scenario. In order to make sure this does not happen, one can rerun the worst case search on the risk reduced portfolio and then take further risk reducing action against remaining worst case scenarios.

As an example, consider a bank with high loans to the airlines industry, and almost as high loans to the oil industry. A worst case could be a rise in oil prices leading to the bankruptcy of many airlines. Taking insurance against high oil prices reduces this risk. But a scenario with very low oil prices would be almost as bad for the bank, since it leads to many bankruptcies in the oil industry. Reducing that risk would require additional insurance positions against very low oil prices.

4.3 Systemic Risk: Network contagion and second round effects

Systemic stress tests analyse how events in the economy impact the banking system. Via two main channels the losses of each bank may depend on what losses accrue at other banks or in the economy at large: First, banks may hold equity and debt of each other, so that the default or downgrade of one affects the other (Section 4.3.1). Second, banks are linked to each other through the markets. If in a fire sale reaction to a stress event one bank chooses to sell off large amounts of an asset, it depresses the price, which might affect other banks holding that asset (Section 4.3.2).

Clearly the modelling of both kinds of interactions between losses are notoriously difficult to capture in a credible quantitative model. One main challenge is that the interaction between banks modifies the stress event—even if the stress event originated outside the banking system. The other challenge is to predict the actions of banks reacting to stress events. But since both of these mechanisms—contagion through exposures and contagion through stress reactions—is such a key element of a crisis situation, a modern stress testing framework should at least make an attempt to include these mechanisms, even if in a crude and strongly simplified way.

The micro-prudential and the macro-prudential perspective cannot be separated any more. Financial crisis situations and prices that emerge in such a situation are a systemic problem that can only be analysed by understanding the interaction and feedback between individual actions and the pricing of risk in the financial system. The assumption of an exogenously fixed risk factor distribution is inappropriate in such a situation.

Recent accounts of financial crisis such as [12], [35] or [22] stress the cyclical pattern of asset prices and credit, with endogenous feedback mechanisms, feeding these boom and bust patterns. A stress test is a thought experiment in which the quantitative impact of a bust is examined.

A stress testing model that can take the amplification mechanisms discussed in the literature into account would need a significant departure from the current stress testing framework that we have discussed so far. Rather than thinking about a one way impact of exogenous shocks on financial losses, we have to think in a framework that allows shocks, losses plus changes in behaviour with additional feedback to losses.

To better understand how a situation of systemic risk changes the standard stress testing framework we give again a graphical account summarized in Figure 6: With endogenous risk three additional interaction channels enter the standard stress testing model. While in the standard framework the influence of a bank's portfolio choice on the risk factor distribution is ignored this has to be taken into account in an endogenous risk framework. This channel does not appear in the standard framework because it is assumed that the institution which does the stress test is negligible in the financial system as a whole and can treat risk factor distributions as given. Even when it is assumed that the bank's portfolio is fixed and given, in an endogenous risk world the risk factor distribution depends on all the other portfolio decisions in the system. This is also ignored in the standard stress testing framework. Thus there is a feedback loop between the risk factor distribution and the portfolio composition of all the other participants in the financial system.

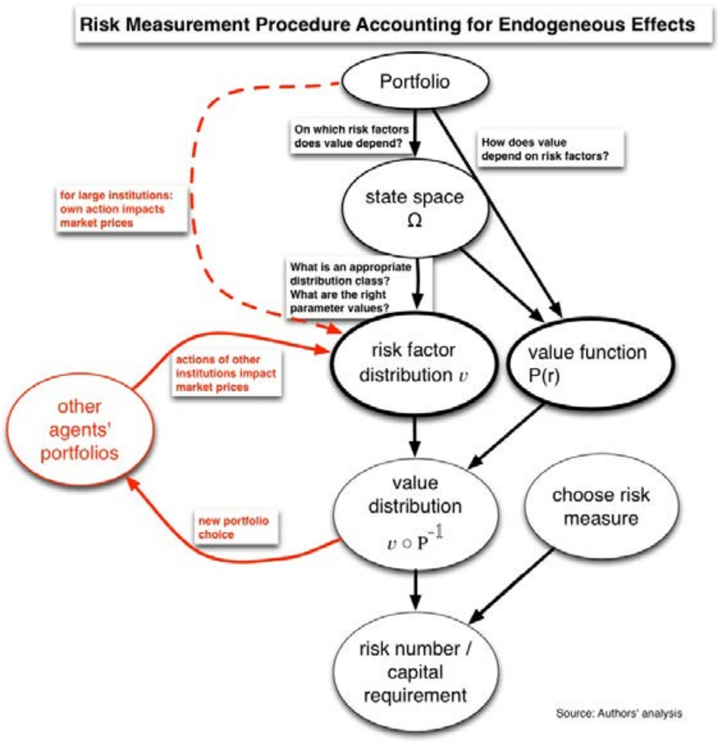

Figure 6: Risk measurement framework with effects of endogenous risk in red. Actions of large instutions or herds might affect expected prices, correlations, and variances and thereby change the risk factor distribution. If large institutions react to changes in the risk factor distribution, this gives rise to feedback effects which are reinforcing the loop if the instutions decide to sell in situations of falling prices.

4.3.1 Contractual networks and domino effects

Contractual network connections between banks make it possible that insolvency spreads in a contagious manner between banks. The literature on contagion of insolvency through contractual networks of interbank debt, started in the late nineties and had a huge impact within both the central bank and regulatory community as well as in academia. For recent overviews we refer to [41], [40] and [24]. Some of this work has entered also into stress testing methodology (see [3], [13], [15] or [19]) although it has not yet become a common element of concurrent stress tests.

In principle, inclusion of network and insolvency contagion analysis between banks, or between banks and their borrowers and lenders outside the banking system is easy to integrate into the current standard framework of stress testing. The reason for this is that a network contagion analysis can be added to an otherwise standard stress test as an independent module.

Perhaps the best way to understand the potential role of a network contagion model within the overall stress testing framework is to see it as an additional element of the portfolio value function. It is most helpful if the stress test is applied to a large system with many intricate contractual obligations, or for considering particular scenarios formulated in terms of counter-party defaults. In this subsection we give an informal description of how to integrate a network model into the overall stress testing framework. A more technical description can be found in [18].

The first step in constructing a network model of contractual relations between banks is to distinguish assets and liabilities in the balance sheet of banks that refer to other banks. You might think of interbank assets and liabilities between banks as items in the balance sheet which receive a special flag.

In order to construct a particular network of interbank assets and liabilities it is however necessary to have additional information about the actual bilateral exposures, which is not contained in the balance sheet. This information will only be available, if the authority which is in charge of the stress test has access to a credit register or detailed supervisory data which record the bilateral contractual relations between banks in the system. Even if the information about bilateral exposures is incomplete, estimation techniques can give a rough idea of the actual exposures. For a detailed description see, for instance, [18].

If the net value of claims and liabilities outside of the banking system plus the value of claims against other banks is smaller than the liabilities vis a vis the other banks in the system, a bank is insolvent. In case of an insolvency a network contagion model usually computes losses by applying three principles: Limited liability, priority of debt claims and proportionality of payouts. While limited liability and priority of debt claims capture legal properties of debt, the proportional rationing of claims is a simplifying assumption to specify the value of claims in the case the initial promises cannot be fully satisfied. It abstracts from modelling the actual payments that would be made in a bankruptcy procedure as well as from the repricing of assets and liabilities in a default state.

Under the assumptions of limited liability, debt priority and proportionality, potential payouts in a stress scenario can be calculated by computing so called clearing payment vectors. An interesting aspect of this computation is that clearing payment vectors not only encode information about potential payouts and losses from interbank relations. They also encode whether bank defaults are a direct consequence of the stress scenario or rather whether they are an indirect consequence triggered by the default of another bank in the system or a knock on effect.

A network model can be thought of as a module of the portfolio value function. It receives input from the balance sheet and an additional data source on bilateral interbank exposures and gives output to the portfolio value function. The network model can be an independent part of the overall stress test model and can be switched on and off according to convenience. If it is switched off, scenario losses would be evaluated abstracting form potential interbank domino effects, if it is switched on these effects would be included in the evaluation of losses. The stress testing framework of the Austrian Central Bank, for instance, maintains such a modular design. This allows to run EBA stress test, without network contagion considerations, as well as special stress tests, including these effects.

4.3.2 Fire sales and second round effects

Financial crisis in the past, but in particular the recent global financial crisis, have shown the powerful and quantitatively huge effects of loss amplification stemming from deleveraging and fire sale processes, which often accompany stress scenarios. In his post mortem analysis of the crisis, Hellwig ([26]) analyses the repeated (huge) errors made by the authorities in estimating the losses that might be faced as the crisis unfolded. These errors had their root exactly in underestimating second round impacts of losses on asset prices.

Second round effects are defined as effects resulting from the reaction of banks and other market participants to the stress. These reactions might weaken or amplify the consequences of the original shock. Fire sales and forced deleveraging are important examples of second round effects: they depress prices on markets, reduce the sales proceeds for bank engaging in the fire sale, and also lead to losses for other institutions which sell or hold and mark to market that asset. Second round effects can also affect asset classes which have not been hit by the initial shock, since banks may decide to put these assets on fire sale because of stress losses in other asset clasess. Another kind of second round effect is price increase of safe haven assets.

The importance of fire sales has been long recognised in the economic literature. Shleifer and Vishny ([37], [38]), Pedersen ([31]), Brunnermeier ([12]), Shin ([36]), Geanakoplos ([23]) are some important references (among many others), which describe this mechanism theoretically and qualitatively. Quantitative descriptions, which are more easily amenable to stress testing situations have recently been provided by Cont and Wagalath ([15], [16]), Braoucezec and Wagalath [4] and in particular by Cont and Schaaning [14]. Also the stress testing modeling attempts at central banks, in particular the Bank of England's RAMSI model ([13]), made attempts to capture deleveraging effects on the evaluation of losses.

The main difficulty in modelling second round effects come from the fact that they need assumptions on the behaviour of banks and other market participants. The other difficulty is that the theoretical literature on second round effects and deleveraging thinks in terms of equilibrium prices. In order to adapt the modelling of fire sale losses to a stress test, short cuts and simplifications are needed.

Here we want to make a simple new proposal based on Braouezec and Wagalath [4], who analyse the single bank case, and on Cont and Schaanning [14] who analyse the case of a banking system. Both papers show how to incorporate second round effects of deleveraging into a stress test in a practical yet credible way.

In these approaches the assets of a bank are divided into three categories, the illiquid loans, the marketable securities and cash, held as reserves at the central bank as well as into different asset classes. Debt is given and the assets may have risk weights. There is a given hurdle rate of capital a bank has to comply with at all times.

Upon a shock capital may be impaired to such a degree that banks have to take action to comply with the minimum capital standard. A number of assumptions describes the behaviour how banks react to a shock severe enough for them to violate the capital constraint: (1) It is assumed that in the stress scenario capital cannot be replenished by issuing new equity or by other means. Instead banks have to liquidate assets. (2) Banks either sell their liquid assets in proportions independent of their type [14], or in decreasing order of their risk weights [4]. If the assets sold represent a large proportion compared to the market, the sale triggers a price impact. The price impact affects the proceeds from fire sales and the mark to market value of assets remaining in the trading book. The price impact is a linear or concave function. A rough idea about the shape of price impact functions is available from the empirical literature. Basically a price impact function is like an inverse demand function for a given asset class. It's parameters can be empirically estimated from observable financial market data (see [14]).

When deciding about the amount of trading book assets to be sold, the banks take into account the price impact of its planned sale. But the bank does not take into account the price impact of possible fire sales by other market participants. The price impact has to be calculated from the combined fire sale quantities of all banks in the system.

If the bank sells loans it is assumed that it can do so only at a discount $0 < \lambda < 1$. The value of this parameter depends on the illiquidity of loans. It is difficult to assess λ empirically but if there is a secondary market for loans the value of the parameter might be estimated from historical data. If such estimates are not possible, the discount applied in the evaluation of losses has to be explicitly stated. Since loans are not marked to market, the discount does not affect the value of the loans not liquidated in the stress scenario.

Using the discount parameter for loans, the price impact function for tradeable securities, the minimum capital ratio and the assumption that banks sell assets proportionally or alternatively to minimise fire sale amounts, the losses in a fire sale can be evaluated for each bank. In a fire sales scenario losses from the stress scenario get amplified. Despite of many assumptions that have to be taken, the model helps to gauge the order of losses that might be expected if a deleveraging process as a consequence of the stress scenario is taken into account.

In a systemic stress test, some competent authority which has sufficient data about all banks in the stress test, additionally can estimate the combined fire sale amounts of *all* banks. Using the price impact function it can estimate the fire sale prices resulting from the combined fire sales. This allows for an estimate for fire sale losses which are due to the fire sales of all banks. Using these prices it can re-calculate fire sale losses of individual banks. If more than one bank needs to deleverage, the fire sale losses due to the combined action of banks will exceed the fire sale losses calculated by an individual bank based only on its own fire sales.

5 Interpreting the output of stress tests

Since stress testing is a relatively young field, there is yet no standard practice of how to best report results for a stress test. Here we discuss some aspects of the representation of a stress test output depending on whether we use a first or second generation stress testing framework and whether we choose a framework with or without network and second round effects.

Stressed balance sheets In each reported scenario, the state of the banking system is of prime interest. For this purpose one can stick to the established representation of stressed balance sheets of each bank, see e.g. the template [21]. This representation includes the capital requirements for each bank in the scenario reported.

Worst case scenarios The result of the worst case search depends on the objective function P and on the parameter k in (3) and (4).

For the objective function P possible choices for stress tests of individual banks are loss arising in a scenario, the resulting CET1 shortfall of a bank, or the capital injection necessary to save the bank in case it fails. For systemic stress test one would measure severity by the value of some function defined on the set all banks in the system. One could consider the sum of the losses of all banks, or the sum necessary to bail out all banks, or reduction in lending to the real economy.

The parameter k is a positive number representing a "radius". The choice of could be guided in two ways. Either some authority develops an expert scenario, as is current practice, and then one takes k to be equal to the plausibility of that expert scenario. Or one could estimate the plausibility of a historical crisis period and choose this as k.

The result of the worst case search (3), (4), (5) is a worst case scenario satisfying the plausibility constraint. For a point scenario one would report the value of the risk factors in that scenario.

Key risk factors If the number of risk factors is very large, e.g. a few hundred, one could report only the values of the key risk factors with the highest explanatory power, as in [6]: Given the worst case scenario, the search for the key risk factors is a search for a small subset of risk factors which can explain the loss under the worst-case scenario as much as possible. Let us take as the m key risk factors those which jointly have the highest contribution to the loss in the worst case scenario. (For some set of m risk factors the partial scenario is the one in which those m risk factors take their worst case values and the other risk factors have their current values. The contribution of a set of k risk factors to the loss in the worst case scenario is the ratio of the loss in the partial scenario over the full worst case loss.)

As a further result of the worst case search one reports the worst case value of the objective function P or $E(P)$ which is achieved in the worst case scenario, as well as the state of all banks in the worst case scenario as represented by their stressed balance sheets.

Representation of system-wide stresses System-wide stress tests allow for an analysis of network effects (Section 4.3.1) and second round effects (Section 4.3). To report the quantitative impact of these effects, one can compare the effect of an initial shock on the banking system and the effect of the initial shock plus the subsequent network and second round effects.

The size of network and second round effects depends on the initial shock scenario. To reflect this dependence one can search for the worst case initial shock which leads to the worst consequences for the banking system, taking into account network and second round effects. The choice of the objective function and the representation of the worst case scenario proceed along the lines sketched in the previous paragraphs.

The systemic risk contributions of individual banks can be represented by the Indirect Contagion Index of [14]. This representation is based on a matrix of liquidity weighted overlaps between banks. The i, j-entry is the sum over all asset classes of the asset values banks i and j hold in the respective class, divided by the market depth of the asset class. This overlap matrix can be represented as a graph whose nodes are the banks and weight of the edges represents the overlap.

The Indirect Contagion Index of a bank is calculated as the entry of the bank in the normalised principal eigenvector of the contagion matrix. The index ranks banks according to the amount of contagion they could trigger should they engage in a fire sale.

6 Designing a stress testing system from scratch: Framework, priorities and resources

6.1 Framework

After having discussed what institutions at the frontier of stress testing do, what are the conceptual elements of a stress test, what are the pros and cons in each of potential modelling step, in this section we would like to make a more specific proposal for a stress testing framework built from scratch. The key elements of such a framework are represented in the flow diagram of Fig. 7.

The design aims at achieving two main goals. First, in order to implement systematic stress tests with worst case scenarios it should allow for the evaluation of many scenarios in an automated way. The price for evaluating a higher number of scenarios is to restrict to more coarse grained models, which allow for a faster approximation. Second, it should allow for the consideration of systemic risk. This means in particular that second round effects should be taken into account. The most relevant second round effects are price mediated contagion through fire sales and network effects of defaults.

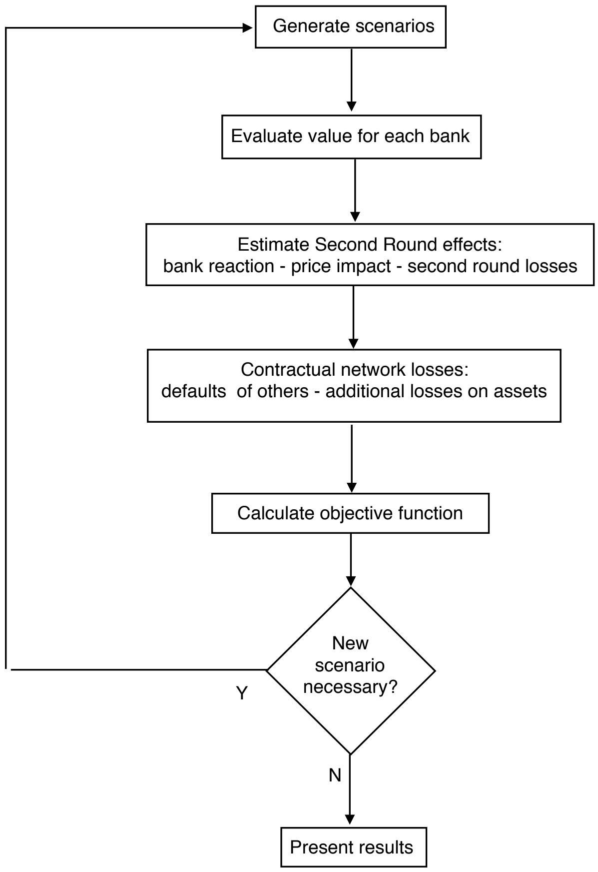

Figure 7: Stress Test Programme Plot

Generate scenarios As risk factors we propose to use the traditional market risk factors, including real estate prices, together with a set of macroeconomic risk factors. Should there be concerns about concentration risks in loan portfolios, one adds idiosyncratic risk factors for important obligors.

Input: Time series of macro and market risk factors; plausibility threshold k.
Output: Estimated joint distribution of macro and market risk factors; scenarios above the plausibility threshold.
Method: The joint distribution of macro and market risk factors is estimated from time series data with standard statistical techniques, often presupposing assumptions about the distribution class. The generation of scenarios above the plausibility threshold implements search algorithms for the worst case scenario [32]. For proposals on the choice of k, see p.33.

In summary we clearly advocate a framework which allows for the evaluation of many scenarios and for the quantification of plausibility. Note however that this approach to scenario generation is able to embed a first generation stress test as follows. A first generation stress test in this context would mean that one works only with one or two stress scenarios, which are constructed in the traditional way and used to calibrate the threshold k. It is like allowing for only a few particular draws from the risk factor distribution. So a first generation stress test may be embedded in a second generation stress test in a natural way within a unified framework.

Evaluate value for each bank The link from a market/macro scenario to stressed PDs can be established with any favoured statistical model, mapping macro risk factors into probabilities of default. One of the most eleborate and data rich model of this kind is the model of Jacobson *et al.* [28], which might be a natural choice for an institution like the Riksbank. Breuer *et al.* [10] provide an alternative, simpler framework. But there are clearly many other models to choose from.
Input: Balance sheet data and portfolio data of each bank; scenario (from 'Generate scenarios').
Output: Portfolio value $P_i(r)$ for each bank i as a function of scenario r.
Method: Standard portfolio evaluation techniques. An aggregation level for the bank portfolio has to be chosen so as to allow for the fast automated evaluation of scenarios.

Let us stress that it is essential that all evaluations of losses are explicitly functions of the scenario. The reason we stress this aspect is that some stress testing frameworks used in practice apply the scenario specification either directly or indirectly to the evaluation of *s*ome balance sheet positions or to some parts of the projected income and expenditure streams but apply ad hoc rules for stressed risk parameters to another set of positions within the same balance sheet. This should in our view be strictly avoided. Since the very rationale of a stress test is to find an answer to the question: "Which scenarios are most detrimental to my current portfolio ?" it is essential that all gains and losses are explicitly linked to stress scenarios. Otherwise the stress test will have consistency issues. Moreover the effort of selecting stress scenarios in a systematic way only makes sense if the mapping between scenarios and losses is made explicit and transparent.

An important aspect here is to be aware that consistent evaluation of scenarios across banks can only be achieved in a top down framework. If losses were evaluated with input from models of banks themselves, a second generation stress test is impossible since this requires the evaluation of many scenarios. Furthermore, reliance on bank internal models makes the losses hard to compare, since banks may use different models.

Estimate second round effects Price mediated seconds round effects are caused by the total fire sales triggered by banks hit by losses bringing their capital down below the required hurdle rate (Section 4.3.2). The price effect of fire sales is described an impact function depending on the market depth of the respective asset class.
Input: Initial loss for each bank caused by the scenario (from 'Evaluate value for each bank'); behavioural rules about bank reaction to losses; data about market depths.
Output: Fire sale losses for each bank.
Method: Price impacts of fire sales on different markets can be modelled as in Cont and Schaaning [14] or in Braouezec and Wagalath [4]. The behavioural rules about bank reaction to losses which is used in [4] requires the risk weights of each asset class.

Again it is important to point out that a second round effect model of price mediated losses can be embedded in an otherwise standard first generation stress testing framework. It can be formulated as an independent module which takes as input the losses in each asset class as a function of the scenario. If the second round module is "switched" on, fire sale amplification of losses will be calculated, if it is "switched off", the evaluation of these effects will be ignored, like for instance in the current EBA stress tests.

Contractual network losses Defaults of banks in the system can lead to losses and possible defaults of other banks via bilateral interbank exposures in debt and equity (Section 4.3.1).
Input: Balance sheet of each bank; bilateral interbank exposures in debt instruments and equity.
Output: Contagion losses and contagion defaults.
Method: Contagion losses and contagion defaults can be calculated as in Elsinger *et al.* [18] or as in [15].

As in the case of second round effects a network model of contractual contagion can be embedded in an otherwise traditional first generation stress test by allowing to switch the network module on an off. As discussed in Section 3.2.4 this is implemented for instance in the modular model of the Austrian Central Bank OeNB.

Calculate objective function After the initial shock, after second round effects, and after contractual network losses, the stressed balance sheets describe the stressed state of the banking system. From this stressed state of the banking system an objective function is calculated, which depends on the purpose of the stress test.
Input: Balance sheet after initial shock (from 'Evaluate value for each bank'); fire sale losses for each bank (from 'Estimate second round effects'); contagion losses for each bank (from 'Contractual network losses')
Output: Value of the objective function
Method: Possibilities for the choice of the objective function are discussed on p.32.

Present results In first generation systemic stress tests, the results are given by the stressed balance sheets in each of a handful of scenarios. In a second generation stress test, the results are the worst case scenario above the plausibility threshold, and the value of the objective function achieved in the worst case scenario. The worst case search involves search algorithms using many scenarios in which the objective function is evaluated.
Input: Value of the objective function in a representative series of scenarios (from 'Calculate objective function')
Output: Worst case scenario and worst case value above the plausibility threshold; worst case distribution and worst case expectation of the objective function; state of each bank in the worst case scenario; key risk factors.
Method: A search algorithm for the worst case scenario above the plausibility threshold is given in [32]. A procedure for calculation the worst case distribution within Kullback-Leibler balls is given in [7]. The identification of key risk factors is discussed on p. 33.

6.2 Priorities

Our proposal contains a number of elements. If there is a need to prioritize among them we would consider the following ranking: Developing a framework which allows the *automated evaluation of a large number of scenarios* is an aspect we consider key and of first priority. At the same level of priority is - in our view - the *quantification of the plausibility of stress scenarios* and the *modelling of fire sales.*

Of second priority to us is the *use of mixed scenarios.* While we consider this as an ultimately desirable feature of a stress testing framework it is sufficient if considered as an intermediate goal. The development of a *network contagion framework* to evaluate losses from mutual contractual exposures is in our view of third order importance. In particular for the highly concentrated Swedish system with four big players, we consider network modelling not the most important feature of the stress testing framework. An overview of our prioritisation is given in Table 6.2.

Issue	Benefit	Reference	Priority
The automated evaluation of large number of scenarios	reduces the risk of missing out dangerous but plausible scenarios. Benefit can only be realised if all losses are strictly linked to scenarios.	p. 23	1
The quantification of the plausibility of scenarios	prevents consideration of highly implausible scenarios	p. 22	1
The use of mixed scenarios	allows for a unified treatment of trading book and banking book	p. 25	2
Modelling fire sales	allows for a realistic quantification of the consequences of bank reactions to stresses	p. 30	1
Modelling of network contagion effects	allows for a quantification of consequences of bank defaults	p. 28	3

Table 1: Issues that need to be addressed in a modern stress testing framework. The first column lists the issues, the second column describes the benefits from addressing this issue, the third column gives a reference to the passage of the text where the topic is discussed in detail. Finally the last column gives a ranking of the issues given the structure of the Swedish banking system.

6.3 Resources

If our proposal would be put into practice a rough resource plan would look as follows: It would be reasonable to distinguish a development phase and the steady state. For the development phase it might be a good idea to set up a team of external interested academics and internal staff preferably in a collaboration with the respective research departments. A team size of four to six people seems to be reasonable for this stage. For the steady state, depending on the frequency of stress testing exercises and the expected number of ad hoc requests for the respective stress testing teams a size of six to ten people seems appropriate. The stress testing group at the Central Bank of Austria has for example seven members.

We believe that there are two key aspects for staffing a stress testing team. First, staff the team sufficiently, so that in the steady state resources for improvement and research are available. It is not optimal if the entire team is fully occupied with concurrent stress testing duties. Second, skill composition is key. A good stress testing framework is not only about modelling, an aspect that was in the center of our paper, but also process organisation. A good skill composition would be for example: One or two people with a computer science background, one or two persons with a data science background. Of course two people with a background in financial economics, macroeconomics and finance with good modelling skills will be essential. Perhaps the team also needs to be amended with one person who knows the institutional, legal and procedural aspects of stress testing very well.

7 Conclusions

Stress testing is an important tool for risk assessment of banks. It has taken a central role in supporting authorities with safeguarding financial stability. Yet current practice has still a considerable potential for improving methods and processes. This paper points out some of these potential improvements.

A figurative summary of our key points Let's look at the key points of our paper once again from a different perspective. Imagine you have to steer a ship on an open, infinite ocean without a map or other sophisticated navigation instruments. Once in a while obstacles appear in the ocean ahead of the ship: An iceberg, a riff or the sea monster Scylla. They can cause a shipwreck when not spotted early enough. If they are recognized in time the ship can be stirred safely past these fatal obstacles.

Assume that the only way you can watch out for such obstacles is to charge someone with looking ahead of the ship. To see obstacles early enough optical instruments will be needed. There are data on the past occurrence of obstacles that might be statistically analysed and that can help forming expectations of what might be ahead or what are particularly dangerous directions.

In this picture a stress test is a method to watch out for obstacles and assess their distance to the ship. It can be the basis for changing course and steer the ship into another, safer direction.

Now assume the rules of watching out for obstacles would stipulate that there are only two directions in which you must look ahead. They might be chosen based on detailed deliberations based on sophisticated statistical analysis of past obstacles data but in the end they would be only two directions.

This would be very dangerous as a method of risk assessment. If there is an iceberg ahead in a different direction as the one that happened to be chosen, the ship will be wrecked and sink. Clearly it would be far more prudent to look in many more directions. This is our main criticism of first generation stress tests and something we would like to see improved.

Why would such a careless restriction of watch-out-directions be considered in the first place? Of course the argument for the restriction is that the process operating the optical instruments with which we look ahead in the various directions is so costly that one can practically only look into a very few directions. But this does not make the approach any less problematic. If this is really the case, we would need to change the instruments in a way that makes looking into many directions less costly. This is the essence of all the arguments we put forward with respect to process organisation. We need to arrive at a situation which allows automated evaluation of many scenarios.

If it comes to instruments that search our imaginary ocean for obstacles, we obviously need instruments that serve our purpose well or perhaps even in an optimal way. We need to look ahead far enough to be able to navigate the ship in the right direction but not further than that. If we use instruments that look extremely far ahead, the picture might get more and more blurred and unclear. This may cause lots of unnecessary steering, because we think we see obstacles where there in fact aren't any. We must also avoid a too narrow radius of observation because otherwise we might not see the obstacles in time to steer the ship safely past them. Thus with respect to the instruments used to look ahead there are trade offs. All our arguments for including systemic risk and in particular modelling of deleveraging processes are ultimately about choosing the right instruments.

What needs to be done? In the paper we take a stance on what should be done, if a stress testing framework can be designed from scratch. First: A good stress testing framework needs systematic evaluation of extreme but plausible scenarios. The tool we recommend to achieve this is the use of second generation stress tests. Second: A good stress testing framework must consider the most important interactions between banks that occur during financial distress. The tools we recommend to achieve this are two particular frameworks for modelling deleveraging effects and contractual networks. While the literature and some models developed at central banks and institutions provide suggestions how to do this we believe that they are not all of them are equally useful. So the particular models we describe in this paper are the models which would be our first choice for taking into account systemic risk.

Priorities We gave a ranking of the most important elements in our proposal in terms of priorities. *Automated evaluation of a large number of scenarios* is an element we consider of first order importance. We also believe that the *quantification* of the plausibility of scenarios is a key element of high importance. We have proposed to also use the concept of *mixed scenarios* to allow for an integrated treatment of banking and trading book.

We think, however, that this is an element of second order importance in a first development step. With respect to systemic risk and the evaluation of losses, we consider the *modelling of fire sales* as a key element of first order importance. In contrast the *modelling of network contagion* is an element we do not consider as equally important, in particular given the relatively concentrated structure of the Swedish banking system.

Resources We gave a rough estimate of the resources needed to develop and maintain a new stress testing framework. We think that in the development stage it would be reasonable to work with a temporary collaboration with external researchers but also drawing on the available in-house research capacity. For the steady state it is important that the group is sufficiently staffed to develop the framework further, to learn and adapt to new developments and progress in research and to continuously improve models as well as processes. Skill composition is a second key element. It should be borne in mind that modelling and professional process organisation are both key elements of a viable and modern stress testing framework.

References

[1] Bank of England. *The Bank of England's approach to stress testing the UK banking system*, 2015.

[2] Claudio Borio, Mathias Drehmann, and Kostas Tsatsaronis. Stress-testing macro stress testing: does it live up to expectations? *BIS Working Papers*, (396), 2012.

[3] Michael Boss, Gerald Krenn, Claus Puhr, and Martin Summer. Systemic risk monitor: risk assessment and stress testing for the austrian banking system. *OeNB Financial Stability Report*, pages 83–95, 2006.

[4] Yann Braouezec and Lakshithe Wagalath. Risk based capital requirements and optimal liquidation in a stress scenario. *Review of Finance*, forthcoming.

[5] T. Breuer, M. Jandacka, J. Mencía, and M. Summer. A systematic approach to multi-period stress testing of portfolio credit risk. *Journal of Banking and Finance*, 36(2):332–340, 2012.

[6] Thomas Breuer. Providing against the worst: Risk capital for worst case scenarios. *Managerial Finance*, 32:716–730, 2006.

[7] Thomas Breuer and Imre Csiszár. Systematic stress tests with entropic plausibility constraints. *Journal of Banking and Finance*, 37:1552–1559, 2013.

[8] Thomas Breuer and Imre Csiszár. Measuring distribution model risk. *Mathematical Finance*, 26:395–411, 2016.

[9] Thomas Breuer, Martin Jandačka, Klaus Rheinberger, and Martin Summer. How to find plausible, severe, and useful stress scenarios. *International Journal of Central Banking*, 5:205–224, 2009.

[10] Thomas Breuer, Martin Jandačka, Klaus Rheinberger, and Martin Summer. Does adding up of economic capital for market- and credit risk amount to conservative risk assessment? *Journal of Banking and Finance*, 34:703–712, 2010.

[11] Thomas Breuer and Gerald Krenn. Stress testing. Guidelines on Market Risk 5, Oesterreichische Nationalbank, Vienna, 1999. Also available at `http://www.oenb.at/en/img/band5ev40_tcm16-20475.pdf`.

[12] Markus Brunnermeier. Deciphering the liquidity and credit crunch 2007-2008. *Journal of Economic Perspectives*, 23(1):77–100, 2009.

[13] Oliver Burrows, David Learmonth, and Jack McKeown. Ramsi: a top-down stress-testing model. *Financial Stability Report*, 2012.

[14] Rama Cont and Eric Schaanning. Fire sales, indirect contagion and systemic stress testing. Technical report, Imperial College London, 2016.

[15] Rama Cont and Lakshithe Wagalath. Running for the exit: Distressed selling and endogenous correlation in financial markets. *Mathematical Finance*, 23(4):718–741, 2013.

[16] Rama Cont and Lakshithe Wagalath. Fire sales forensics: Measuring endogenous risk. *Mathematical Finance*, 26(4):835–866, 2016.

[17] Kieran Dent, Ben Westwood, and Miguel Segoviano. Stress testing of banks: An introduction. *Bank of England Quarterly Bulletin*, 56(3):130–143, 2016.

[18] Helmut Elsinger, Alfred Lehar, and Martin Summer. Network models and systemic risk assessment. In J.-P. Fouquet and J. Langsam, editors, *Handbook of Systemic Risk*. Cambridge University Press, 2012.

[19] Marco A. Espinosa-Vega and Juan Sole. *A guide to IMF Stress Testing: Methods and Models*, chapter Introduction to the network analysis approach to stress testing, pages 205–209. International Monetary Fund, 2014.

[20] European Banking Authority. *Methodological Note EU wide Stress Test 2014*, 2.0 edition, 2014.

[21] European Banking Authority. *Example of 2016 EU-wide Stress Test templates*, 2016.

[22] John Geanakoplos. *NBER Macroeconomics Annual 2009*, chapter The Leverage Cycle, pages 1–65. University of Chicago Press, April 2010.

[23] John Geanakoplos. *The Leverage Cycle*, volume 24, chapter 1, pages 1–65. University of Chicago Press, 2012.

[24] Paul Glassermann and Peyton Young. Contagion in financial networks. *Journal of Economic Literature*, 54(3):779–831, 20136.

[25] Michael B. Gordy. A risk-factor model foundation the ratings-based capital rules. *Journal of Financial Intermediation*, 3:199–232, 2003.

[26] Martin Hellwig. Systemic risk in the financial sector: An analysis of the subprime-mortgage financial crisis. *De Economist*, 157(2):129–207, 2009.

[27] International Monetary Fund. *Macrofinancial Stress Testing – Principles and Practices*. Monetary and Capital Markets Department, 2012.

[28] Tor Jacobson, Jesper Lindé, and Kasper Roszbach. Firm default and aggregate fluctuations. International Finance Discussion Papers 1029, Board of Governors of the Federal Reserve System, 2011.

[29] United States Government Accountability Office. Federal reserve additional actions could help ensure the achievement of stress test goals. Technical report, United States Government Accountability Office, October 2016. Online at http://gao.gov/assets/690/681020.pdf.

[30] Li Lian Ong, editor. *A Guide to IMF Stress Testing. Methods and Models*. 2014.

[31] Lasse Pedersen. When everyone rund for the exit. *International Journal of Central Banking*, 5(4):177–199, 2009.

[32] Filip Pistovčák and Thomas Breuer. Using Quasi Monte Carlo-Scenarios in risk management. In Harald Niederreiter, editor, *Monte Carlo and Quasi-Monte Carlo-Methods 2002*, pages 379–392. Springer, 2004.

[33] Matthew Pritsker. Choosing stress scenarios for systemic risk through dimension reduction. Technical report, Federal Reserve Bank of Boston, 2016.

[34] Till Schuermann. *Stress Testing and Macroprudential Regulation: A Transatlantic Assessment*, chapter Stress Testing in wartime and in peacetime, pages 125–141. CEPR Press, 2016.

[35] Hyun Song Shin. *Risk and Liquidity*. Oxford University Press, 2010. Clarendon Lectures.

[36] Hyun-Song Shin. *Risk and Liquidity*. Oxford University Press, 2010.

[37] Andrej Shleifer and Robert Vishny. Liquidation values and debt capacity: A market equilibrium approach. *Journal of Finance*, 47(4):1343–1366, 1992.

[38] Andrej Shleifer and Robert Vishny. Fire sales in finance and macroeconomics. *Journal of Economic Perspectives*, 25(1):29–48, 2011.

[39] Gerold Studer. Market risk computation for nonlinear portfolios. *Journal of Risk*, 1(4):33–53, 1999.

[40] Martin Summer. Financial contagion and network analysis. *Annual Review of Financial Economics*, 2013.

[41] Christian Upper. Simulation methods to assess the danger of contagion in interbank networks. *Journal of Financial Stability*, (7):111–125, 2011.

[42] Martin Čihák. Stress testing: A review of key concepts. CNB Internal Research and Policy Note 2, Czech National Bank, 2004. www.cnb.cz/en/pdf/IRPN_2_2004.pdf.