JEL Classification: E44, E47, G21
Keywords: stress testing, credit risk, bank capital

How to Improve the Quality of Stress Tests through Backtesting*

Adam GERŠL—Joint Vienna Institute; Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague (adam.gersl@gmail.com), *corresponding author*
Jakub SEIDLER—Czech National Bank; Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague (jakub.seidler@cnb.cz)

Abstract
This paper describes the stress-testing framework used in the Czech central bank and focuses on the general question of how to calibrate the models and parameters used to stress test the most important risks in the banking system. The paper argues that stress tests should be calibrated conservatively to overestimate the risks so that sufficient buffers are in place for when adverse shocks materialize. However, to ensure that the stress test framework is conservative enough over time, backtesting, i.e., comparison of the actual values of key financial variables with the predictions generated by the stress-testing models, should be a standard part of the stress-testing framework.

1. Introduction

Stress tests are used by commercial financial institutions, regulators, and central banks as a means of testing the resilience of individual portfolios and institutions or the entire sector to adverse changes in the economic environment. This paper is focused on "macro" stress tests of banks, which have become a standard tool among central banks and regulatory authorities for assessing the vulnerabilities of the banking sector as a whole (see, for example, Foglia, 2009, or Drehmann, 2009, and references therein). However, general methodological problems apply also to macro stress tests for other financial industries (pension funds, insurance companies, credit unions, etc.).

The earliest banking sector stress-testing models, which were initially based on simple historical scenarios linking macroeconomic developments with financial sector variables (e.g. Blaschke et al., 2001), have been developed into more sophisticated models integrating market, credit, and interest rate risk and capturing inter-institution contagion and the feedback effect between the financial sector and the real economy. These relatively complex models have become regular tools for analyzing the resilience of the financial sector—see, for example, Danmarks Nationalbank (2010, p. 45), Oesterreichische Nationalbank (2010, p. 51), Norges Bank (2010, p. 49), the RAMSI (Risk Assessment Model for Systemic Institutions) of the Bank of England (Aikman et al., 2009), and European Banking Authority (2011).

Nevertheless, the global financial crisis uncovered deficiencies in the stress-testing methodologies used in many countries. Before the crisis, many tests were

* The authors would like to thank Jan Frait, Michal Hlaváček, Tomáš Holub, Kateřina Šmídková, Christian Schmieder, Claus Puhr, and two anonymous referees for helpful comments and useful recommendations. However, all errors and omissions are those of the authors. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not represent the views of any of the above-mentioned institutions. The authors acknowledge the support of the Czech National Bank (CNB Research Project No. C1/10) and the Czech Science Foundation (GA CR Grant No. 403/10/1235).

wrongly indicating that the sector would remain stable even in the event of sizeable shocks (Haldane, 2009; Borio et. al., 2012). These deficiencies related not only to the configuration of the adverse scenarios used, which had initially seemed implausibly strong but were often exceeded in reality, but also to the shock combination assumed, which had not been adequately anticipated in the scenarios (Ong and Čihák, 2010; Breuer et al., 2009). A role was also played by deficiencies in model calibration and in the assumed behavior of banks and markets, and by the absence of testing of liquidity risk alongside traditional financial risks (in particular credit risk and interest rate risk), since distress after the Lehman failure confirmed the importance of the spiral between market and funding liquidity and its fragile link to the solvency of institutions (Gorton, 2009; Brunnermeier et al., 2009). This problem in stress-testing frameworks is also demonstrated by Ong and Čihák (2010) using the example of Iceland, where the banking sector collapsed in the fall of 2008 even though stress tests conducted in mid-2008 had indicated it was stable.

Consequently, the assumptions and parameters used in stress tests are gradually being re-examined so that the tests can better analyze the impacts of strong shocks to the financial system, and stress tests are becoming a standard tool in the new macroprudential framework (FSB, 2011), though there are some doubts about their ability to serve as an early warning device (Borio et al., 2012). In defense of stress testing, however, it should be mentioned that this is a relatively new tool[1] and hence it still requires methodological development and refinement.[2] And the recent financial turbulence has clearly suggested some possible ways of improving their methodology.

This paper focuses on how to calibrate models used to stress test the most important risks in the banking system. We argue that stress tests should be calibrated conservatively and overestimate the risks. However, to ensure that the stress test framework is conservative enough over time, a process of backtesting, i.e., comparison of the actual values of key banking sector variables with the predictions generated by the stress-testing models, should become a standard part of the stress-testing framework. Direct backtesting of adverse scenarios is not possible in the majority of cases (i.e., non-crisis periods). Thus, the backtesting should be performed on baseline scenarios. However, the whole stress-testing model should be calibrated conservatively in order to take into account the uncertainty related to possible changes in the estimated relationships in the event of adverse economic developments. Hence, ex-post comparison between reality and the predictions generated by the baseline scenarios should indicate systematic risk overestimation.

To illustrate our point, we first present a simple case of how a model estimated in good times leads to underestimation of risk in bad times. After that, the results of a backtesting exercise conducted using the Czech National Bank's (CNB) stress-testing framework are presented. The CNB has been performing macro stress tests of the banking sector since 2003 and has significantly expanded its

[1] Tools based on various types of financial soundness indicators have traditionally been used to assess the resilience of financial institutions (Geršl and Heřmánek, 2008).

[2] The formal obligation of commercial banks to conduct stress tests on their own portfolios was only introduced by Basel II (for banks using advanced methods for calculating capital requirements), which was implemented in the EU in 2006–2007. However, there is now a set of CEBS/EBA guidelines related to stress testing in commercial banks (see Committee of European Banking Supervisors—CEBS, 2009).

methodology over the past few years. The most recent major update was done in mid-2009 and involved the introduction of dynamic features into the system (see Section 2). On this occasion, backtesting of the overall stress-testing methodology was conducted in the context of the aforementioned international debate on the reliability of the predictions of the impacts of shocks to the banking sector. The aims were to demonstrate whether the stress test assumptions were correctly configured and to identify any deficiencies in those assumptions.

The analysis reveals that the current CNB stress-testing system generally errs on the right—i.e., pessimistic—side and slightly overestimates the risks. This leads on average to estimates of key financial soundness indicators (in particular capital adequacy) that are lower (more conservative) than the actual values. Some back-testing results were used to further develop the stress tests. To our knowledge, there is no other study that systematically and transparently presents the backtesting of a stress-testing methodology. With this paper we would like to contribute to the debate on how to develop and calibrate reliable stress-testing frameworks.

The paper is structured as follows. Section 2 briefly describes the CNB's stress-testing methodology that was subsequently verified. Section 3 argues why stress test parameters should be set conservatively. Section 4 summarizes the back-testing methodology and presents summary conclusions of the backtesting for capital adequacy (including its two main constituents, i.e., regulatory capital and risk-weighted assets, RWA) and some other key banking sector variables used in the stress tests. Section 5 concludes by summarizing the backtesting results and proposing a medium-term plan for further development of the stress test methodology.

2. Current Banking Sector Stress-Testing Methodology of the CNB

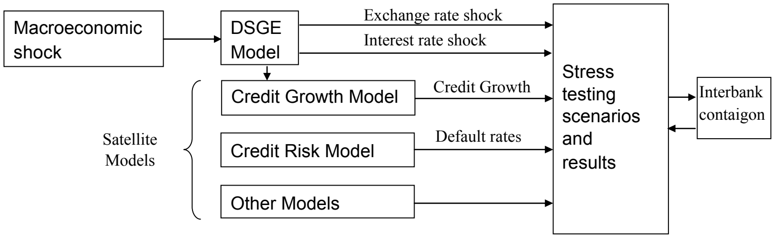
The original banking sector stress-testing methodology applied at the CNB was based on the IMF methodology used for FSAP missions (e.g. Blaschke et al., 2001; Čihák, 2005; Čihák and Heřmánek, 2005).[3] The CNB later switched from testing historical ad-hoc scenarios defined by a combination of shocks (e.g. a 20% rise in non-performing loans, a 15% exchange rate depreciation) to using consistent macroeconomic scenarios generated by the CNB's prediction model and related credit risk and credit growth satellite models (Čihák at al., 2007; Jakubík and Schmieder, 2008; Jakubík and Heřmánek, 2008). This framework was used for the Financial Stability Report 2008/2009 (Czech National Bank, 2009).

In the second half of 2009, the CNB significantly updated its banking sector stress-testing methodology in three respects. First, the tests were "dynamized", in the sense of switching to quarterly modeling of shocks and their impacts on banks' portfolios. This change was described in a box in the CNB Financial Stability Report 2008/2009 (CNB, 2009, pp. 63–64). Second, in the credit risk area there was a change-over to "Basel II terminology", i.e., to capturing the credit risk of several separate portfolios using the standard parameters PD, LGD, and EAD and relating risk-weighted assets to those parameters using procedures specified in the IRB approach to calculating capital requirements.[4] The final major innovation was the extension of

[3] The stress-testing methodology used by IMF FSAP missions has also developed considerably. The current stress-testing framework is described in Schmieder et al. (2011).

[4] PD—probability of default; LGD—loss given default; EAD—exposure at default; IRB— internal ratings based.

Scheme 1 Architecture of Stress Tests



Source: CNB (2007), authors

the shock impact horizon from one to two years (or eight subsequent quarters) and later (in 2011) to three years.

These changes were motivated by the best practices of other central banks and supervisory authorities which had made an effort to develop specific expertise in the field of macro-to-micro linkages in order to assess the banking sector's ability to withstand adverse scenarios.

2.1 Alternative Macroeconomic Scenarios

Alternative macroeconomic scenarios still serve as the starting point for stress testing in the advanced methodological framework. The scenarios are designed using the CNB's official prediction model (currently a dynamic stochastic general equilibrium—DSGE—model; see Brázdik, 2011) supplemented with an estimate of the evolution of some additional variables, which are not directly generated by the model (so-called "satellite models"). Stress scenarios are constructed based on the identification of risks to the Czech economy in the near future. To compare the stress outcome with the most probable outcome, the stress tests use a baseline scenario, i.e., the current official macroeconomic prediction of the CNB.

The predictions for GDP growth, inflation, and other macroeconomic variables enter credit risk, credit growth, and other satellite models, which transform macroeconomic developments into financial sector variables and thereby capture changes in banks' credit portfolios, credit risk, bank income, etc. The stress tests work explicitly with the four main loan portfolio segments by debtor and/or credit type (non-financial corporations, loans to households for house purchase, consumer credit, and other loans), to which the sub-models are also adjusted. The credit risk models are used to predict PDs (default rates) for the individual loan segments, whereas the credit growth models are used to estimate the growth in bank portfolios in relation to the macroeconomic situation and (after certain adjustments) to estimate the evolution of risk-weighted assets. Given the inherent uncertainty in predicting credit risk, the model forecast are often adjusted by expert judgment to reflect all available information about developments in the banking system.

The architecture of the stress-testing framework as a whole is described in the Scheme 1, which illustrates how the CNB's macroeconomic model and satellite models generate alternative scenarios for the banking sector. The part "Other Models" consists of the set of models for property prices, LGD, the yield curve, and bank income (adjusted operating profit).

In the stress tests, the prediction for macroeconomic and financial variables for individual quarters is reflected directly in the prediction for the main balance-sheet and flow indicators of banks. The tests are dynamic, i.e., for each item of assets, liabilities, income, and expenditure there is an initial (the last actually known) stock, to which the impact of the shock in one quarter is added/deducted, and this final stock is then used as the initial stock for the following quarter. This logic is repeated in all quarters for which the prediction is being prepared. Consistency between stocks and flows is thus ensured.

2.2 Credit Risk

Credit risk testing is the most important area of stress testing. This testing is based on the use of PDs for each of the four main segments of the loan portfolio. The second credit risk parameter is LGD, which is estimated by a very simple model in combination with expert judgment, with different amounts being set for different scenarios and different credit segments in line with the regulatory rules, commercial bank practices, the approaches applied by some rating agencies (Moody's, 2009), and existing estimates based on market data (Seidler and Jakubík, 2009). The third parameter is EAD, which is determined as the volume of the non-default part of the portfolio (i.e., excluding non-performing loans) and is influenced mainly by the forecast for credit growth.

An increase in PD and LGD has two main effects on individual banks.

First, the expected loan losses (in CZK millions), against which banks will create new provisions of an equal amount and record them on the expenses side of the profit and loss statement as impairment losses, are calculated as the product of PD, LGD, and EAD for each credit segment and quarter.[5] Total assets are then symmetrically reduced by the amount of these expenses.

The product of PD and the volume of the non-default portfolio forms the volume of new non-performing loans (NPLs) for each quarter. This allows us to generate the volume of total NPLs in the following eight quarters for each bank, and subsequently for the banking sector as a whole, according to the following equation:

$$NPL_{t+1} = NPL_t + \sum_{i=1}^{4} PD_{t+1,i} NP_{ti} - aNPL_t \tag{1}$$

where *NPL* are non-performing loans, *PD* is the probability of default, *NP* is the non-default portfolio in the four segments defined above, and a is an NPL outflow parameter (i.e., write-offs or sales of existing NPLs, i.e., the default part of the portfolio). Parameter a is set by expert judgment at 15% for all segments, i.e., 15% of NPLs are written off/sold each quarter and subsequently disappear from the total volume of NPLs and (gross) assets of the bank. This calibration was chosen on the basis of discussions with commercial banks and estimates conducted as part of the back-testing exercise, which are described in more detail at the end of the next section.

[5] According the relevant CNB decree and IFRS, banks are not required immediately to create provisions exactly equal to expected losses, but rather they must create provisions equal to realized losses, i.e., for new NPLs. However, if the loans are gradually reclassified during the quarter into the NPL (i.e., default) category to the extent predicted by PD, banks will ultimately create these provisions in the originally estimated amount.

The credit growth model leads to an estimate of the gross volume of loans in individual segments. Using relation (1) for NPL modeling, this allows us to determine for each bank, and subsequently for the banking sector as a whole, the NPL/total loans ratio, a standard indicator of the banking sector's health.

Second, in the case of banks applying the Basel II IRB approach to the calculation of capital requirements for credit risk, the capital requirements (or risk-weighted assets, RWA[6]) for credit risk are a function of PD, LGD, and EAD. Given that the largest banks in the Czech Republic apply this approach, this relation is applied to all banks for the sake of simplicity. Given a constant non-default portfolio volume, i.e., EAD, an increase in PD and LGD thus generally results in an increase in RWA and therefore a decrease in capital adequacy.[7]

2.3 Interest Rate and Currency Risk

The macroeconomic scenarios contain a prediction of the evolution of the simplified koruna and euro yield curves (rates with 3M, 1Y, and 5Y maturities). A change in interest rates has a direct effect on bank balance sheets mainly in the value of bond holdings.[8] The calculation is based on the estimated duration of the bond portfolios, which is calculated by expert judgment on the basis of a more detailed knowledge of the maturity structure. Account is also taken of bond portfolio hedging using IRS (interest rate swaps), which for some banks lessens the impact of interest rate changes.

The quarter-on-quarter change in the CZK/EUR exchange rate is applied to the net open foreign currency position (including off-balance-sheet items), generating either a loss or a profit depending on the sign of the net open position and the direction of the exchange rate change.[9]

2.4 Interbank Contagion Risk

Interbank contagion risk is modeled in two selected periods (in the fourth and eighth quarters). The test uses data on interbank exposures, with the capital adequacy of individual banks being used to determine their probability of default (PD).[10] As interbank exposures are mostly unsecured, LGD is assumed to be 100%. The expected losses due to interbank exposures are calculated for each bank according to the formula PD×LGD×EAD, where EAD is the net interbank exposure. If these losses are relatively high and will lead to a reduction in the bank's capital adequacy and thus an increase in its PD, there follows another iteration of the transmission of the negative effects to other banks through an increase in the expected losses. These iterations are performed until this "domino effect" of interbank contagion stops, i.e., until the rise in PD induced in one bank or group of banks does not lead to a rise in

[6] Risk-weighted assets = capital requirements (in CZK millions) × 12.5.

[7] This channel of the impact of increased PD and/or LGD on banks is one of the main sources of the much criticized procyclicality of Basel II (see Geršl and Jakubík, 2012).

[8] At the same time, however, interest rate changes have an indirect effect on credit risk via their effect on the PD estimate. An additional effect of changes in interest rates is on net interest income, which, however, is captured in the modeling of adjusted operating income.

[9] For example, a positive open foreign currency position and appreciation of the koruna leads to losses.

[10] The PD values in relation to capital adequacy ratios (CAR) are set by expert judgment as follows: PD = 100% for negative CAR; PD = 25% for CAR between 0% and 5%; PD = 15% for CAR between 5% and 8%; PD = 5% for CAR between 8% and 10%; PD = 0.5% for CAR greater than 10%.

the PD of other banks. Despite the relatively advanced calculation, the interbank exposures are relatively small, so this type of risk never plays a large role in the final results of stress test exercises.

2.5 Profit, Regulatory Capital, and Capital Adequacy

The stress test assumes that banks will continue to generate revenues even in the stress period, particularly net interest income (interest profit) and net fee income. For these purposes, an analytical item of the profit and loss account called "adjusted operating profit" has been constructed of which the main items are interest profit plus fee profit minus administrative expenses.[11] The volume of adjusted operating profit was initially determined by expert judgment for the individual scenarios. A model estimate of this item was introduced only in mid-2010 (CNB, 2010), using nominal GDP (+), the slope of the interest rate curve (+), change in the NPL stock (+), and capital adequacy (+) as explanatory variables.

Regulatory capital is modeled in accordance with the applicable CNB regulations. Each bank enters the first predicted quarter with initial capital equal to that recorded in the last known quarter. If a bank generates a profit in the first predicted quarter (i.e., its adjusted operating profit is higher than its losses due to the shocks), its regulatory capital remains at the same level (is not increased). If, however, it generates a loss, its regulatory capital is reduced by the amount of that loss. The impacts of the shocks are thus reflected in a reduction of capital only if they exceed adjusted operating profit and the bank generates a loss.

It is assumed that those banks which generate a profit for the entire financial year will decide on profit distribution and dividend payments in the second quarter of the following year. Here we assume that each bank, when increasing its capital from retained earnings of the previous financial year, will try to get to its initial capital adequacy ratio if its previous year's profits are sufficient.[12] Depending on the change in RWA, several scenarios are thus possible:

– the bank distributes the entire profit and does not strengthen its regulatory capital (in the event of unchanged RWA);
– the bank uses part of its profit to strengthen its capital and distributes the remainder (in the event of an increase in RWA; however, the entire retained earnings of the previous year will not be needed to reach the initial level of capital);
– the bank uses the entire profit to strengthen its capital (in the event of a relatively sizeable increase in RWA); depending on the size of the increase in RWA, however, it may not reach the original capital adequacy ratio;
– the bank pays dividends that exceed the profit generated (in the event of a decrease in RWA) and thereby also distributes part of retained earnings of previous years.

Total capital adequacy is then calculated for the individual quarters as the ratio of regulatory capital to total RWA. The portion of RWA relating to credit risk is

[11] In some CNB Financial Stability Reports this adjusted operating profit was called "net income". Adjusted operating profit is broadly equivalent to the item "pre-provision profit", i.e., operating profit gross of losses on non-performing loans, but differs in that it does not include the impacts of other (interest rate and exchange rate) shocks, whereas pre-provision profit does.

[12] This assumption may not be very realistic at certain times, as banks may decide to pay higher dividends and reduce their capital adequacy ratio below the initial level.

modeled on the basis of the credit risk parameters (see above), while the other components of RWA (or of the capital requirements for other risks) for the individual quarters are determined by expert judgment or kept constant for simplicity.

3. Backtesting of the Stress Tests: The Argument for Conservative Calibration

The objective of backtesting (often also called validation or verification) is to examine to what extent the assumptions and sub-models used in the stress-testing framework are in line with reality. A problematic aspect of this exercise is that the tests use stress—i.e., unlikely—scenarios, which may not occur in reality. Hence, we cannot subsequently compare the predictions based on adverse scenarios with reality. For this reason, only the scenario that represents the most likely evolution of the economic environment, i.e., the no-stress baseline scenario, could be used for the backtesting.

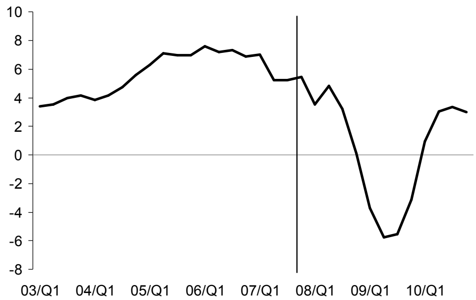
The prediction using the baseline (i.e., likely) scenario should lead to forecasts undervaluing risks compared to those that occur in reality. This is because the whole system should have a "conservative" buffer to offset the uncertainty associated with estimating losses given adverse economic developments, when relations (for example the elasticity between GDP growth and risk parameters such as PD) estimated by standard econometric techniques on data from mainly calm periods can change suddenly for the worse. Being prudent in stress testing is in line with the general macroprudential approach adopted by policymakers and supervisors worldwide, and erring on the conservative side is preferred to possible underestimation of the losses and capital needs of banking systems in crisis, which can have large negative effects on public budgets, on general public perceptions of banks, and back onto the economy.

One dimension of prudent calibration is the decision on whether to set shocks to the banking system as a result of models estimated using available data, or to set the parameters expertly. Clearly, if the data are not sufficiently long and do not include stress periods, the estimated satellite models might not be well suited for stress-testing purposes. On the other hand, for macro stress tests one needs a link between macroeconomic developments and risk factors for banks. Thus, there is a clear trade-off in terms of having all risk factors estimated via models and the possibility of accumulating a large number of errors, which could underestimate the real impact of shocks on the banking system. The option that was selected in the CNB stress-testing framework reflects this trade-off and uses models only for those factors which can be reasonably modeled, with the view that over time, as better and longer data series become available, other factors currently estimated to a large extent by expert judgment could be predicted via models.

The conservative buffer can be imposed in a number of ways, such as applying a conservative add-on to the central predictions (such as adding one standard deviation of the dependent variable), using a prediction from a "conservative" confidence interval, or estimating the elasticity on different sub-samples and taking the most conservative one (usually one estimated over a crisis period if such a period is available). Another possibility is to define some variables (such as the PD) in a conservative way[13] or, for parameters set expertly, just using a very conservative setting.

The requirement for conservative calibration implies that stress test prediction errors should be evaluated differently from the errors of standard macroeconomic

Figure 1 The Adverse Scenario (real GDP growth in %, coincides with the observed outcome)



Source: CNB

predictions, where deviations in either direction are regarded as "equally bad". In backtesting using baseline scenarios, it is appropriate to apply an asymmetric view in the stress tests and tolerate prediction errors toward some overestimation of the risks.

The case for conservative calibration can be illustrated by a simple exercise which uses the data for the Czech economy and assumes the authority (the CNB) running a stress test in early 2008, focusing on forecasting credit losses from corporate portfolios of Czech banks for an adverse scenario. A standard approach would be to estimate the relationship between a credit risk parameter, say the corporate (one-year) default rate, and macroeconomic fundamentals (such as GDP growth), using all available data, which as at early 2008 cover the period 2003–2007 (quarterly data).[14] This relationship would be used to forecast the default rate over the period of the next three years, the current forecasting horizon of the CNB's stress tests, i.e., for the "crisis" period of 2008–2010. If we design the adverse scenario to equal the observed macroeconomic path (a decline in GDP of roughly 4.5% in 2009), we can directly validate the forecast by comparing it to a stress scenario, a unique opportunity that the crisis offers for policymakers.

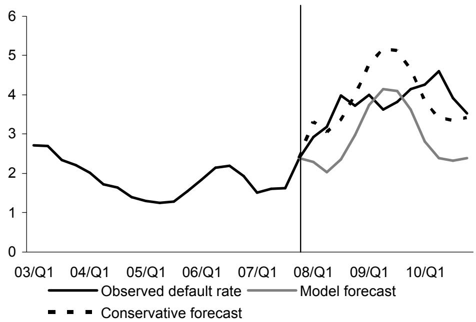
A simple OLS-type model has been estimated to link the corporate default rate and GDP growth in the Czech Republic using quarterly data (other variables proved insignificant).[15] The prediction using the adverse "2008–2010 crisis" scenario correctly indicates the increase in the default rate and its subsequent decrease due to the economic recovery (*Figure 1*). However, mainly due to the fact that the model was estimated over a calm period of economic growth, the model underestimates

[13] For example using a definition of PD that is based on the 30+ days in arrears definition of the default rate, which is generally higher than the standard Basel 90+ days definition. However, given the results of the backtesting as to the large overestimation of credit losses—see later in the text—the CNB changed to the standard 90+ definition of the default rate in June 2010 (CNB, 2010). Currently, the conservative margin is safeguarded via an add-on to the predicted PD.

[14] The 12M default rate for corporate exposures is calculated from the CNB credit register, which started operating in late 2002.

[15] The estimated model looks as follows: default rate = 2.99 + -0.2*GDP growth. No lags were identified as significant, partly due to the fact that the default rate is calculated as forward looking, i.e., we are linking, for example, GDP growth in 2006Q1 with the 2006Q1 default rate, which is calculated over the 4-quarter period 2006Q1–2006Q4. This is a simplified model used to illustrate the point. However, it captures the most important effect, namely, the one from GDP.

Figure 2 Forecasted versus Observed Corporate Default Rate (12M default rate in %)



Source: CNB; authors' calculations

the real outcome (*Figure 2*). As can be seen from the comparison, the observed default rate picked up much earlier than forecasted and started to fall much later. The error is quite high—assuming for simplicity a constant corporate portfolio, the model predicts that over the period 2008–2010, 8.8% of the portfolio would default, while in reality the figure was 11.2%.

Thus, a conservative calibration is needed to properly account for the losses, which were higher than predicted. An ex-post analysis shows that the estimation of the equation over the crisis period would lead to a higher elasticity between GDP and the default rate and would predict higher default rates. Since policymakers do not have the crisis-period data before the crisis, an alternative must be used. If we apply a conservative add-on of one standard deviation of the corporate default rate (which equals roughly 1%), the forecasted path is still different (*Figure 2*). However, as it is at a higher level—the three-year impact (3-year default rate) now amounts to 11.9%, much closer to (and even slightly higher than) the observed rate of 11.2%.

4. Results of the Backtesting of the CNB's Stress Tests

The backtesting was conducted on quarterly data in the period 2004Q4–2009Q2, i.e., for 19 periods in all.[16] The actual values of key variables for the banking sector as a whole are compared with the predictions generated by the current stress-testing methodology for the individual quarters using the relevant baseline scenario of the forecast. As the stress-testing methodology allows us to create a prediction for up to the next 12 quarters, it was necessary to choose a prediction horizon. As most of the time over which the backtesting exercise is conducted the CNB used a one-year horizon, the results presented in this paper are based on a one-year prediction.[17] The predictions for past quarters were therefore created

[16] The first attempt to validate the CNB's stress tests using the baseline forecast scenario was made back in 2007 (Čihák et al., 2007), when the capital adequacy ratio and NPL growth predictions generated by the 2006 stress-testing methodology were compared with their real counterparts.

[17] This means, for example, that the actual outcome in 2007Q4 was compared with the prediction for that quarter made one year earlier, i.e., on bank portfolios as of 2006Q4 using the January 2007 baseline scenario. Internally, however, the backtesting was performed for more prediction horizons. Additional results using 2Q and 6Q horizons are qualitatively similar (see the *Appendix*).

subsequently using the updated stress-testing methodology in order to verify that methodology and do not match the values published in CNB Financial Stability Reports.

Two statistics based on the mean prediction errors were used to validate the selected variables: the mean absolute error (MAE) defined by equation (2):

$$\frac{1}{n}\sum_{t=1}^{n}\left|P_t - A_t\right| \tag{2}$$

and the mean error in direction (MED) defined as:

$$\frac{1}{n}\sum_{t=1}^{n}\frac{P_t - A_t}{\left|A_t\right|} \tag{3}$$

where P_t denotes the value of the prediction of the estimated variable for the given quarter. A_t denotes the actual value, and t represents the quarter for which the prediction is being made.[18]

MAE serves for simple presentation of the mean prediction error in the units in which the given variable is expressed, while MED expresses whether the given variable was overestimated or underestimated on average and thus gives the degree of "conservatism".

The prediction error of the capital adequacy ratio and other key banking sector variables can be split into two main factors. The first is the potential prediction error caused by inaccuracy in the estimates of the macroeconomic variables entering the stress-testing mechanism (interest rates and the exchange rate), and the second concerns the assumptions and sub-models used in the stress test itself (e.g. the assumptions about how the bank raises its regulatory capital, what interest and non-interest yields it achieves. and how sensitive it is to interest rate risk). The macroeconomic prediction error can be eliminated in the backtesting by using the actual (ex-post) values of macroeconomic variables. The residual error is then due to inaccuracies in the assumptions and sub-models of the stress-testing framework and the intentional conservative buffer.

The most important output variable of the tests is the estimate of the capital adequacy ratio (CAR). The mean absolute deviation (MAE) for CAR equates to roughly 1.6 p.p. of the capital adequacy ratio (see *Table 1*). This means, for example, that the test predicts CAR of 11.4% instead of 13%.

This prediction error equates to roughly 1.8 standard deviations. In the individual shorter periods this error gradually shrinks to 0.8 p.p. (i.e., 1 standard deviation) but then grows again slightly from 2007 onwards. Only a small part of the error is due to errors in the macroeconomic forecast, as the MAE statistic decreases only modestly with knowledge of actual macroeconomic developments.

The negative MED statistic of -10.8% shows that the real values were higher on average in the period as a whole and the stress tests thus tended to generate

[18] As part of the backtesting we also computed other prediction error statistics, e.g. the mean percentage error, the mean weighted percentage error, the mean quadratic error, and the mean percentage quadratic error. The backtesting results using these statistics, however, did not differ significantly from the results using MAE and MED. which are easier to interpret.

Table 1 Deviation of Capital Adequacy Ratio Estimate
(Estimate for 1-year horizon)

	2004–2009	2004–2005	2005–2006	2006–2007	2007–2008	2008–2009
Mean absolute error (MAE)						
Prediction—stress test	1.6	1.0	0.8	1.6	2.1	1.9
Prediction—known macro	1.5	0.9	0.6	1.1	2.0	2.5
Mean error in direction (MED) in %						
Prediction—stress test	-10.8	-1.7	-6.5	-13.1	-17.2	-15.3
Prediction—known macro	-8.8	1.9	-1.3	-7.1	-16.3	-20.0

Source: authors' calculations

overvalued CAR estimates (see *Table 1*). This fact is also demonstrated by *Figure 1*, which reveals that a lower-than-actual CAR is predicted from the end of 2006 onwards. The resulting CAR was thus underestimated for most periods, in line with the conservative design of the tests. This conclusion remains valid even when the predictions are adjusted for the error in the prediction of macroeconomic variables. Similar results are obtained even for different prediction horizons (see the *Appendix*, *Table A1*, where two-quarter and six-quarter horizons are also compared).

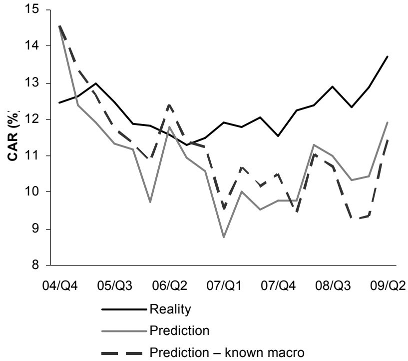
The estimate of a lower-than-actual CAR is due to inaccuracy in the estimate of both RWA and regulatory capital. With few exceptions the stress test overestimated RWA (see *Figure 3*) and simultaneously tended to underestimate regulatory capital (see *Figure 4*). The decomposition of the error in the CAR estimate into the part caused by inaccurate prediction of RWA and the part caused by inaccurate prediction of regulatory capital shows that both variables contribute to the error, but the contribution of the RWA is higher (on average, 65% of the error is due to the RWA and 35% due to capital) (*Figure 5*; see *Figures A1* and *A2* in the *Appendix* for the detailed decomposition of the error). The overestimation of risk-weighted assets has two sources: first, the credit growth model tends to predict higher credit volumes than the ex-post outcome. While at first glance underestimation of credit growth seems to be the conservative calibration, the opposite is true at least from the point of view of risk-weighted assets. Second, the framework uses the estimates of PDs and LGDs as the base for the risk weights (IRB approach), which are also overestimated.

Regulatory capital is regularly increased out of after-tax profits, so the estimate of profits is an important parameter for the evolution of capital. Profits are calculated as the difference between adjusted operating profit and losses due to the individual shocks tested (see section 2). The backtesting of this variable revealed that the stress test systematically underestimates after-tax profit (*Figure 6*). This is due to two factors. First, the test systematically underestimates adjusted operating profit directly through the assumption about its level (for the baseline it was assumed that adjusted operating profit will be 90% of the average for the previous two years).

This is also in line with the more conservative approach to risk assessment (*Figure 7*). The second cause is that the stress test tends to overestimate the impact of the main risk tested, i.e., credit risk, in the form of higher-than-actual PD and related higher provisioning for NPLs (recorded in the "losses from impairment" category), partly also due to too conservative expert estimates of LGD (*Figure 8*).

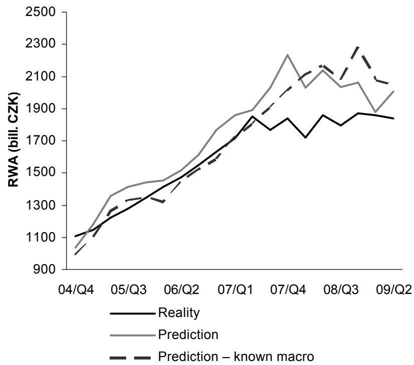
The NPL ratio is a closely monitored financial stability indicator. We therefore present detailed backtesting results for this variable as well. A comparison of

Figure 3 Backtesting of CAR Estimate
(estimate for 1-year horizon)



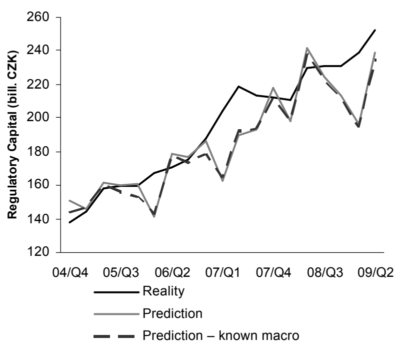
Source: authors' calculations

Figure 4 Backtesting of RWA Estimate
(estimate for 1-year horizon)



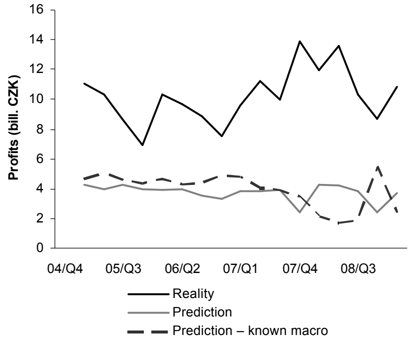
Source: authors' calculations

Figure 5 Backtesting of Regulatory Capital Estimate
(estimate for 1-year horizon)



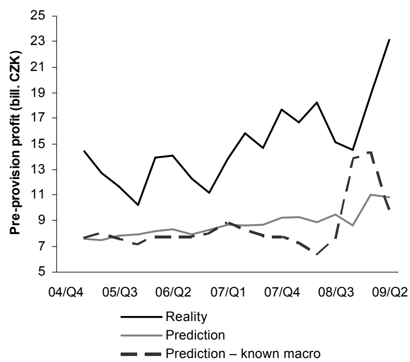
Source: authors' calculations

Figure 6 Backtesting of Profits Estimate
(quarterly values, estimate for 1-year horizon)



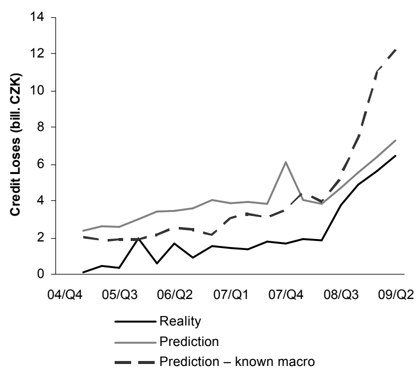
Source: authors' calculations

Figure 7 Backtesting of Pre-Provision Profit Estimate
(quarterly values, estimate for 1-year horizon)



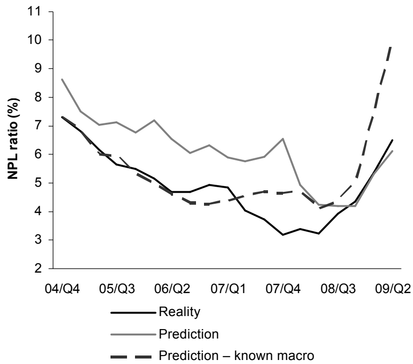
Source: authors' calculations

Figure 8 Backtesting of Credit Losses Estimate
(quarterly values, estimate for 1-year horizon)



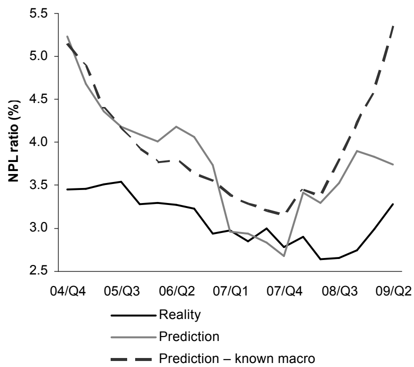
Source: authors' calculations

Figure 9 Backtesting of NPL Ratio—Corporations
(estimate for 1-year horizon)



Source: authors' calculations

Figure 10 Backtesting of NPL Ratio—Households
(in %, estimate for 1-year horizon



Source: authors' calculations

the actual NPL ratios with their predicted values reveals overshooting of the estimates, especially since the end of 2007, for both non-financial corporations (see *Figure 9*) and households (see *Figure 10*).

Table 2 shows that MAE was around 1.3 p.p. for non-financial corporations and 0.7 for households. While the NPL estimates for corporations improve significantly with knowledge of the macroeconomic environment, the opposite is true for households in some periods. In overall comparison, however, the household NPL estimate is more accurate. This conclusion applies even for different prediction horizons (see the *Appendix*, *Tables A2* and *A3*). Also, we observe that the NPL prediction error for known macro increases significantly in the time period 2008–2009, owing to the worse-than-expected development of macro variables at the beginning of the financial crisis. This further illustrates the suitability of a conservatively calibrated stress-testing framework, as the prediction of non-financial corporations' NPL ratio underestimated the actual ratio only negligibly, and in most cases the NPL ratio was prudentially higher.

Table 2 Deviation of NPL Ratio Estimate for Corporations and Households
(in %, estimate for 1-year horizon)

NPL ratio—corporations						
	2004–2009	**2004–2005**	**2005–2006**	**2006–2007**	**2007–2008**	**2008–2009**
Mean absolute error (MAE)						
Prediction—stress test	1.3	1.1	1.4	1.9	1.4	0.6
Prediction—known macro	0.8	0.1	0.2	0.6	0.8	1.5
Mean error in direction (MED) in %						
Prediction—stress test	27.8	18.3	26.2	45.5	38.5	12.1
Prediction—known macro	12.3	-0.1	-3.2	6.1	20.6	31.0
NPL ratio—households						
Mean absolute error (MAE)						
Prediction—stress test	0.7	1.1	0.8	0.5	0.4	0.8
Prediction—known macro	0.9	1.1	0.7	0.4	0.7	1.3
Mean error in direction (MED) in %						
Prediction—stress test	21.6	30.7	25.6	12.1	13.9	26.7
Prediction—known macro	27.7	30.5	21.0	14.2	24.1	43.5

Source: authors' calculations

The overestimation of the NPL ratio is due both to the aforementioned conservative calibration of the PD risk parameter and, to some extent, to underestimation of outflow parameter a from equation (1). To determine the optimum value of a, numerical minimization of the MAE error statistic was performed in various time intervals of 2004–2009. The optimum outflow a for the entire period under review was 20% on average. Owing to the deliberate overestimation of the potential risks this parameter was conservatively set at 15% in the tests.[19]

Despite the relatively positive message of the backtesting results, further gradual refinement of the predictions is desirable. The main problem in the credit risk area is with the sub-models and assumptions used, as they excessively overestimate the impact of credit risk in the form of losses on impaired loans. While the direction toward overestimation is correct, the degree of overestimation should be held in a reasonable range.[20] At the same time, the conservative prediction of adjusted operating income (and, as a result, overall profits) seems to be too far from the ex-post reality, so adjustments in this area are also needed. Following the backtesting exercise, the CNB has started to further recalibrate the stress-testing framework in order to bring the estimates closer to reality, while still preserving some degree of conservatism (CNB, 2010). This has mainly involved recalibration of the satellite models and better prediction of risk-weighted assets.

In general, the development of the stress tests conducted in any authority should be based on regular backtesting. This should be an integral part of the banking

[19] The sensitivity of the NPL ratio estimate to change in a reveals that an increase in a of 5 p.p. (i.e., from the 15% used to the optimum value of 20%)—i.e., a faster outflow of NPLs from banks' balance sheets—causes on average a decline in the NPL ratio of one-tenth (e.g. from 10% to 9%).

[20] The results of backtesting of other key variables (not reported here, but available from the authors upon request) indicated that besides large overestimation of credit losses, market losses (FX and bond revaluations) are also overestimated to some extent.

sector stress-testing framework to enable ongoing assessment of whether the assumptions are realistic and a conservative buffer is being maintained in the risk predictions.[21]

5. Conclusion

This paper focused on how to calibrate the parameters used in banking sector stress tests. It argued that the parameters should be calibrated conservatively and should slightly overestimate the risks in order to take into account the uncertainty related to possible changes in the estimated elasticities in the event of adverse economic developments. This means that the ex-post comparison between reality and the predictions generated by baseline scenarios should indicate systematic risk overestimation.

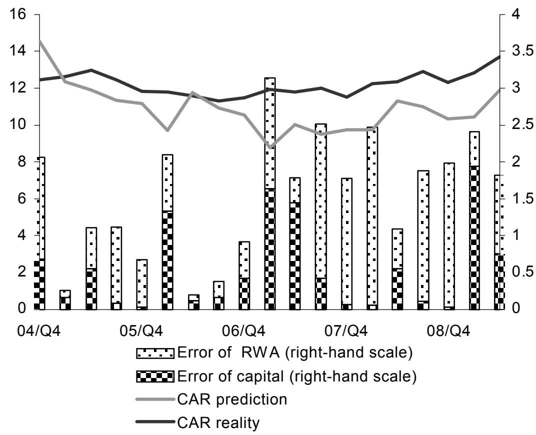
We used a case study of the CNB's banking sector stress-testing methodology and presented the results of a backtesting of that methodology. Such backtesting is a tool that should be used regularly as a guide for refining the assumptions and models used. The results of the backtesting conducted in 2010 reveal that the CNB's stress tests err on the right—i.e., pessimistic—side and slightly overestimate the risks. This leads on average to capital adequacy estimates that are lower (more conservative) than the actual values. This is consistent with the design of the stress tests, which should be built on conservative assumptions. However, account should be taken of the fact that the level of conservatism, i.e., the degree of overestimation of the risks, in the methodology can only be fully assessed after the effects of the current recession disappear. Also, more attention should be focused on probability assessment and precise quantification of the stress-testing conservatism needed. However, this issue is left for other research.

The backtesting results also indicated areas where further refinement of the stress tests is desirable. The main such areas are credit risk (more accurate estimates of PD and LGD), modeling of bank income in relation to the macroeconomic scenario, better estimation of risk-weighted assets, and certain enhancements in calculating the impacts of market risks. These areas have already been tackled to some extent in the CNB's recent stress-testing framework as presented in the CNB Financial Stability Reports published since 2010.

[21] Regular backtesting—i.e., retrospective assessment of prediction performance—is also routinely performed as part of the creation of predictions for monetary policy purposes—see, for example, CNB (2008).

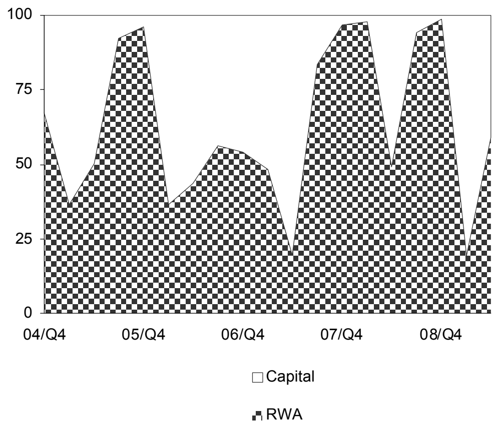
APPENDIX

Figure A1 Decomposition of the Error in CAR Estimate into RWA and Capital (% on left-hand scale, p.p. on right-hand scale)



Source: authors' calculations

Figure A2 Percentage Decomposition of the Error in CAR Estimate into RWA and Capital (%)



Source: authors' calculations

Table A1 Detailed Deviation of Capital Adequacy Ratio Estimate
Estimate for different horizons

	2004–2009	2004–2005	2005–2006	2006–2007	2007–2008	2008–2009
Mean absolute error (MAE)						
Prediction—stress test						
two-quarter horizon	1.4	1.5	1.7	1.5	1.2	1.0
four-quarter horizon	1.6	1.0	0.8	1.6	2.1	1.9
six-quarter horizon	1.6	0.6	0.5	1.1	2.1	2.7
Prediction—known macro						
two-quarter horizon	0.9	0.9	0.8	0.9	1.0	0.9
four-quarter horizon	1.5	0.9	0.6	1.1	2.0	2.5
six-quarter horizon	1.7	0.8	0.5	0.9	2.1	3.3
Mean error in direction (MED) in %						
Prediction—stress test						
two-quarter horizon	-9.8	-11.0	-14.1	-13.1	-9.2	-4.0
four-quarter horizon	-10.8	-1.7	-6.5	-13.1	-17.2	-15.3
six-quarter horizon	-11.2	1.0	-1.3	-8.7	-16.9	-20.7
Prediction—known macro						
two-quarter horizon	-6.5	-4.8	-6.5	-7.4	-8.4	-7.2
four-quarter horizon	-8.8	1.9	-1.3	-7.1	-16.3	-20.0
six-quarter horizon	-11.7	3.0	0.9	-7.0	-17.1	-25.4

Note: Four-quarter horizon corresponds to values in *Table 1.*
Source: authors' calculations

Table A2 Deviation of NPL Ratio Estimate for Corporations
Estimate for different horizons

	2004–2009	2004–2005	2005–2006	2006–2007	2007–2008	2008–2009
Mean absolute error (MAE)						
Prediction—stress test						
two-quarter horizon	0.7	0.8	0.8	0.9	0.7	0.4
four-quarter horizon	1.3	1.1	1.4	1.9	1.4	0.6
six-quarter horizon	1.8	1.2	1.6	2.1	2.1	1.6
Prediction—known macro						
two-quarter horizon	0.6	0.2	0.3	0.6	0.5	1.3
four-quarter horizon	0.8	0.1	0.2	0.6	0.8	1.5
six-quarter horizon	0.9	0.2	0.3	0.5	1.0	1.9
Mean error in direction (MED) in %						
Prediction—stress test						
two-quarter horizon	14.8	12.0	14.6	22.0	18.3	8.4
four-quarter horizon	27.8	18.3	26.2	45.5	38.5	12.1
six-quarter horizon	39.3	21.4	31.4	49.6	59.6	34.7
Prediction—known macro						
two-quarter horizon	8.6	-1.2	-3.9	6.1	13.8	23.5
four-quarter horizon	12.3	-0.1	-3.2	6.1	20.6	31.0
six-quarter horizon	15.8	-1.9	-4.2	4.2	24.0	40.0

Note: Four-quarter horizon corresponds to values in *Table 2.*
Source: authors' calculations

Table A3 Deviation of NPL Ratio Estimate for Households
Estimate for different horizons

	2004– –2009	**2004– –2005**	**2005– –2006**	**2006– –2007**	**2007– –2008**	**2008– –2009**
			Mean absolute error (MAE)			
Prediction—stress test						
two-quarter horizon	0.4	0.4	0.3	0.2	0.3	0.6
four-quarter horizon	0.7	1.1	0.8	0.5	0.4	0.8
six-quarter horizon	0.9	1.4	1.2	0.9	0.6	0.6
Prediction—known macro						
two-quarter horizon	0.4	0.4	0.2	0.1	0.4	0.9
four-quarter horizon	0.9	1.1	0.7	0.4	0.7	1.3
six-quarter horizon	1.0	1.2	0.9	0.7	0.9	1.5
			Mean error in direction (MED) in %			
Prediction—stress test						
two-quarter horizon	10.8	12.4	6.1	2.6	10.9	20.0
four-quarter horizon	21.6	30.7	25.6	12.1	13.9	26.7
six-quarter horizon	27.5	39.9	37.5	26.8	18.8	22.2
Prediction—known macro						
two-quarter horizon	13.3	10.7	4.2	3.2	13.8	29.8
four-quarter horizon	27.7	30.5	21.0	14.2	24.1	43.5
six-quarter horizon	34.4	35.1	28.1	22.0	31.0	50.5

Note: Four-quarter horizon corresponds to values in *Table 2.*
Source: authors' calculations

REFERENCES

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

Alessandri P, Gai P, Kapadia S, Mora N, Puhr C (2009): Towards a Framework for Quantifying Systemic Stability. *International Journal of Central Bankin*g, 5(3):47–81.

Bank of England (2009): The role of macroprudential policy. *Discussion Paper, Bank of England*, November 2009.

Borio C, Drehmann M, Tsatsaronis K (2012): Stress-testing macro stress testing: does it live up to expectations? *BIS Working Paper*s, no. 369.

Breuer T, Jandacka M, Rheinberger K, Summer M (2009): How to Find Plausible, Severe and Useful Stress Scenarios. *International Journal of Central Banking*, 5(3):205–224.

Brázdik F, Hlaváček M, Maršál A (2011): Survey of Research on Financial Sector Modeling within DSGE Models: What Central Banks Can Learn from It. *Czech National Bank, Research and Policy Notes*, no. 3/2011

Brunnermeier M, Crockett A, Goodhart C, Persaud A, Shin H (2009): *The Fundamental Principles of Financial Regulation.* Geneva Reports on the World Economy 11.

Čihák M (2005): Stress Testing of Banking Systems. *Finance a úvěr- Czech Journal of Economics and Finance*, 55:418–440.

Čihák M, Heřmánek J (2005): Stress Testing the Czech Banking System: Where Are We? Where Are We Going? *Czech National Bank, Research and Policy Notes*, no. 2005/02.

Čihák M, Heřmánek J, Hlaváček M (2007): New Approaches to Stress Testing the Czech Banking Sector. *Finance a úvěr- Czech Journal of Economics and Finance*, 57: 41–59.

Committee of European Banking Supervisors—CEBS (2009): *Guidelines for Stress Testing.*

Czech National Bank (2008): *Evaluation of the Fulfilment of the CNB's Inflation Targets 1998––2007.*

Czech National Bank (2009): *Financial Stability Report 2008/2009*.

Czech National Bank (2010): *Financial Stability Report 2009/2010*.

Danmarks Nationalbank (2010): *Financial Stability 2010.*

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

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

Financial Stability Board—FSB (2011): *Macroprudential policy tools and frameworks, Report on the G20*.

Foglia A (2009): Stress Testing Credit Risk: A Survey of Authorities' Approaches. *International Journal of Central Banking*, 5(3):9–46.

Geršl A, Heřmánek J (2008): Indicators of financial system stability: towards an aggregate financial stability indicator? *Prague Economic Papers*, (2):127–142.

Geršl A, Jakubík P (2012): How Important Is the Adverse Feedback Loop for the Banking Sector? *Journal of Economics*, 60(1):32–49.

Gorton G (2009): Information, liquidity, and the (ongoing) panic of 2007. *American Economic Review*, 99:567–572.

Haldane A (2009): Why Banks Failed the Stress Test. Paper presented at the Marcus Evans Conference on Stress-Testing, 9–10 February 2009. Available at:
http://www.bankofengland.co.uk/publications/Documents/speeches/2009/speech374.pdf

Jakubík P, Heřmánek J (2008): Stress Testing of the Czech Banking Sector. *Prague Economic Papers*, (3):195–212.

Jakubík P, Schmieder Ch (2008): Stress Testing Credit Risk: Is the Czech Republic Different from Germany? *Czech National Bank Working Paper,* no. 9/2008.

Majnoni G, Martinez Peria M, Blaschke W, Jones M (2001): Stress Testing of Financial Systems: An Overview of Issues, Methodologies, and FSAP Experiences. *IMF Working Papers*, no. 01/88.

Moody's (2009): Approach to Estimating Czech Banks' Credit Losses. *Moody's Global Banking,* July 2009.

Norges Bank (2010): *Financial Stability 1/2010*.

Oesterreichische Nationalbank (2010): *Financial Stability Report 19*.

Ong L, Čihák M (2010): Of Runes and Sagas: Perspectives on Liquidity Stress Testing Using an Iceland Example. *IMF Working Papers*, 156:1–28.

Schmieder Ch, Puhr C, Hasan M (2011): Next Generation Balance Sheet Stress Testing. *IMF Working Papers*, no. 11/83.

Seidler J, Jakubík P (2009): Estimation of Expected Loss Given Default. In: Czech National Bank: *Financial Stability Report 2008/2009*, pp. 102–109.