WORKING PAPER SERIES 9

2019

Jiří Panoš, Petr Polák

How to Improve the Model Selection Procedure in a Stress-testing Framework

CZECH NATIONAL BANK

WORKING PAPER SERIES

How to Improve the Model Selection Procedure in a Stress-testing Framework

Jiří Panoš
Petr Polák

9/2019

CNB WORKING PAPER SERIES

Reviewed by: Marco Gross (International Monetary Fund)
Tomáš Adam (Czech National Bank)

Project Coordinator: Dominika Ehrenbergerová

How to Improve the Model Selection Procedure in a Stress-testing Framework

Jiří Panoš and Petr Polák*

Abstract

This paper aims to introduce a contemporary, computing-power-driven approach to econometric modeling in a stress-testing framework. The presented approach explicitly takes into account model uncertainty of satellite models used for projecting forward paths of financial variables employing the constrained Bayesian model averaging (BMA) technique. The constrained BMA technique allows for selecting models with reasonably severe but plausible trajectories conditional on given macro-financial scenarios. It also ensures that the modeling is conducted in a sufficiently robust and prudential manner despite the limited time-series length for the explained and/or explanatory variables.

Abstrakt

Tato práce si klade za cíl představit moderní, na výpočetním výkonu založený přístup k ekonometrickému modelování v rámci zátěžového testování. Navržený přístup explicitně zohledňuje modelovou nejistotu satelitních modelů používaných k projekcím trajektorií finančních proměnných za pomoci techniky Bayesovského průměrování modelů (BMA) s omezením. Technika BMA s omezením umožňuje výběr takových modelů, které vytvářejí sice zátěžovou, ale přitom věrohodnou, daným makro-finančním scénářem podmíněnou trajektorii. Technika BMA současně zajišťuje, že modelování je prováděno dostatečně robustním a obezřetným způsobem navzdory omezené délce časových řad pro vysvětlované a/nebo vysvětlující proměnné.

JEL Codes: C11, C22, C51, C52, E58, G21.

Keywords: Bayesian model averaging, model selection, model uncertainty, probability of default, stress testing.

---

*Jiří Panoš, University of Economics, Prague, and Czech National Bank, Financial Stability Department (e-mail: jiri.panos@cnb.cz);
Petr Polák, Charles University Prague, and Czech National Bank, Monetary Department (e-mail: petr.polak@cnb.cz);
This work was supported by grant VSE IG102029, and we acknowledge support from the Grant Agency of Charles University (grant #1548119).
The authors would like to thank Libor Holub, Dominika Ehrenbergerová, Tomáš Adam, Marco Gross, and Simona Malovaná for their comments and suggestions, which helped improve the paper significantly.

1. Introduction

One of the crucial items included in the mandate of a national supervisory authority (often the central bank) is to supervise financial markets and maintain financial stability. Over the past decade, stress testing has become an increasingly relevant tool heavily used by financial institutions and their supervisors. The importance of stress testing has been growing since the Global Financial Crisis (GFC), when large banks failed in several countries and put the stability of financial systems in those countries in jeopardy. For supervisors to minimize the possibility of recurrence of such an event, a profound understanding of the banking sector is crucial. Stress testing is one of the newly used regulatory and macro-prudential tools that help supervisors to properly evaluate potential threats to financial stability, discover hidden vulnerabilities, and assess the resilience of the banking sector.

We show how our methodology can be applied on the example of one of the key credit risk parameters – the probability of default (PD). The probability of default is most often modeled in stress-testing frameworks using a satellite model. The satellite models employed in various risk areas, such as credit risk, market risk, and interest rate risk, are used to establish a link between the modeled risk parameters and the macro and financial variables included in the macroeconomic scenarios to project scenario-conditional forward paths of these parameters. Credit risk modeling requires parameters such as probabilities of default, loss given default (LGD), and, with the introduction of IFRS 9, also transition probability matrices (TPMs) and 12-month and lifetime loss rates (LRs). These parameters are important drivers of the profit and loss statement (P&L) and risk-weighted assets (RWA) of a financial institution. To calibrate the model, we use data from the Czech Republic as an example. This approach is useful, since we face the same issue as other regulatory and supervisory institutions – short data time series and resulting uncertainty with respect to selecting the most appropriate model.

The presented PD model is used in the macro-prudential solvency stress-testing framework at the Czech National Bank (CNB). In particular, the CNB assesses the resilience of the banking sector as a whole primarily through the impact of the credit risk channel (the market risk and operational risk channels have a much smaller impact on the stability of the banking sector). The tests are performed using the available supervisory data about the Czech banking sector. The results are used to assess macro-prudential aspects of the capital position of the banking sector under a set of macroeconomic scenarios (baseline and adverse) and are published regularly in the annual Financial Stability Report.

The motivation for the new model is based on our experience with common one-equation-based ARDL and ARIMAX models, where the modeled variable is the 3-month PD. The originally selected models were not sensitive enough to macroeconomic shocks and thus it was often necessary to modify the forecasted scenario-conditional path using expert-judgment shocks. Since our time series start at the dawn of the GFC and from the crisis onwards the situation mostly improves, we lack sufficiently long time series to make the one-equation-based models more robust and more sensitive to shocks at the same time.

To our best knowledge, papers focusing specifically on PD satellite models are scarce (examples include Castren et al., 2010, or Gray et al., 2013) and most papers on stress testing present the whole credit risk framework. However, none of these papers account for model uncertainty, meaning that only a single equation is picked out of a large pool of possible equations which may all be acceptable from the econometric and economic standpoint. However, different equations can imply quite different dynamics of the risk parameters modeled and consequently result in different projected capital positions of the banks tested. In this paper, we promote the use of the constrained Bayesian model averaging (BMA) framework combining and adjusting the approaches of Henry and Kok (2013), Gross and Poblacion (2017), and Siemsen and Vilsmeier (2017) to obtain robust and plausible estimated forward PD paths to be used in macro-prudential stress testing. We enrich the existing literature mainly by setting prior constraints in terms of the desired range of scenario-conditional forecasts and by taking the out-of-sample performance of the models explicitly into account.

We propose a modern and robust approach to satellite modeling in any stress-testing framework. The contribution of this paper to the literature on stress testing is supported by the three main advantages of the presented approach. First and foremost, we combine and enhance existing frameworks to obtain a robust and plausible estimate of the scenario-conditional forward path of PD (given the known relationships and the overall scenario, the forecasts produced by the model cannot deviate from ex-ante expert judgment; see the benchmark values constraint in Section 3.3). Second, the model approach presented here can be used without any ex-post expert-judgment adjustments while preserving a conservative approach to stress testing. Third, the framework can be further used to calibrate any stressed variable, such as loan interest rates and funding costs. Using readily available data and our straightforward methodology, supervisory and regulatory authorities can obtain conservative yet reasonably shocked forward paths of the key variables of interest.

The paper is structured as follows. Section 2 summarizes the current literature. Section 3 provides information on the data and an overview of the BMA framework employed. Section 4 gives the sample calibration and discusses the results. Section 5 concludes the paper.

2. Literature Review

Over the past two decades, there has been rising interest among policy-makers and academics in developing and improving new tools for assessing the resilience of the financial sector. According to Vazquez et al. (2012), one of the pioneering works in the area of stress testing is Wilson (1998), who presented a framework for examining a portfolio's credit risk under distressed macroeconomic conditions. The growing number of recent stress-test-related studies can be divided into two groups. First, there is a group of papers providing a broader overview of macro-stress-testing approaches, surveys of methodologies, and general discussions about what stress testing is able to achieve, such as Sorge and Virolainen (2006), Foglia (2009), and Borio et al. (2012). Papers in the second group are primary studies focusing on the development of a coherent framework for stress testing and covering the individual risk areas tested. These papers are mainly based on the applied work of national central banks and international organizations such as the European Central Bank, the International Monetary Fund, and the Bank for International Settlements (e.g. Henry and Kok, 2013; Dees et al., 2017; Cihak, 2007; Schmieder et al., 2011; Gersl et al., 2012; Vazquez et al., 2012; Alessandri et al., 2009; Louzis et al., 2012; Hirtle et al., 2016; Kapinos and Mitnik, 2016).

Macroeconomic scenarios serve as the starting point for any stress-testing exercise. Breuer et al. (2009) discuss how to find a stress scenario which is at the same time plausible, severe, and suggestive for designing risk-reducing actions. For this paper, we use a general approach in which the scenarios are designed using a DSGE prediction model and take the period of 2007–2009 as the historical reference for the possible magnitude of a crisis. The macro variables are supplemented with estimated scenario-conditional evolutions of certain additional parameters using satellite models (see, for example, Gersl et al., 2012, for more discussion about the links between the scenario design and the satellite models). While the most probable macroeconomic developments are described by the baseline scenario, the adverse scenario is based on the perceived sources of risks to the economy in the near to medium-term future. Both scenarios are modeled solely for stress-testing purposes and explore the 5-year horizon.

Credit risk is typically the most important part of the macro stress test, and probability of default is arguably the most important credit risk parameter. This paper does not aim to present a new credit risk framework, but rather introduces a modern probabilistic approach to the model selection procedure for the key driver of credit risk, especially when long enough time series are not available and model uncertainty is high. We build upon previous studies, mainly following Dees et al. (2017) and Gross and Poblacion (2017) and their predecessors such as Castren et al. (2010), who analyze the behavior of corporate default probabilities under a series of shocks and proposes employing a GVAR model to link the macro-financial scenarios with micro-level default data, and Gray et al. (2013), who develop a model framework to connect risk indicators and macroeconomic variables using a GVAR model. Our goal is to combine the standard approach, in which risk parameters are linked with macroeconomic variables, with an up-to-date Bayesian approach to model selection.

Bayesian model averaging is a method used to overcome the issues with model uncertainty and model selection procedures. Steel (2019) puts BMA into the overall context of model averaging and its use in econometric modeling and presents not only Bayesian, but also frequentist model averaging methods. He argues that model averaging is a formal and in fact quite natural response to model uncertainty. Model uncertainty is a common feature of many econometric models and occurs especially when there are plenty of plausible ways of explaining the variable of interest. Model uncertainty has a direct impact on model selection, which is a procedure aimed at finding the best model according to some pre-specified criteria (e.g. explanatory power and forecasting performance). The model selection procedure is therefore determined by the modeler's perception of the "best" model.

To deal with model uncertainty it might be useful to employ model averaging methods. Model averaging is not a completely new idea, as the pioneering works of Bayesian and frequentist model averaging include Hoeting et al. (1999) and Hjort and Claeskens (2003). In the area of macro-prudential stress testing, the model uncertainty is huge, as there are many possible explanatory variables and the required time series are usually rather short. The model selection criteria are often driven by attempts to be sufficiently conservative in the adverse scenario in order to make the scenario-conditional projections of PD well-suited for direct use in the stress test or as an input to downstream models. In this paper, we promote the use of the constrained BMA methodology built upon the work of Henry and Kok (2013), Bonti et al. (2006), Gross and Poblacion (2017), and Siemsen and Vilsmeier (2017). This is a natural yet sophisticated approach to reducing model uncertainty, as it emulates the work of an econometrician, but it allows us to analyze and evaluate a large number of potential models at once. Henry and Kok (2013) present a whole credit risk framework for stress testing, with satellite models built using the BMA methodology, which is described in more detail by Gross and Poblacion (2017). We use "constrained BMA" as presented by Gross and Poblacion (2017), but we augment it by introducing a new set of constraints that are particularly useful for stress testing. Siemsen and Vilsmeier (2017) also use BMA and a filtration procedure based on econometric and economic properties (correlation of covariates, autocorrelation, sign restriction, and out-of-sample forecasting). Following Bonti et al. (2006), they also have some benchmark constraints based on quantiles. Our approach is similar, but due to data availability we use historical values and the deviations from them instead of quantile benchmarks.

3. Data and Methodology

This section provides information on the data used for the calibration and presents an overview of the constrained BMA framework.

3.1 Data

Our dataset consists of historical macroeconomic variables with quarterly frequency and historically observed 3-month default rates. The default rates are taken from internal (Central Credit Register;[1] in the future we plan to use data from AnaCredit) and external credit registers[2] and are aggregated across the whole banking system for the three key credit portfolio segments: households for house purchase (HH-HP), consumer credit to households (HH-CC) and non-financial corporates (NFC). Thus, the key assumption here is that the default rate data are realizations of the underlying PD data-generating process. Quarterly data are preferable in our case, as the Czech National Bank's top-down stress-testing framework is also calculated using quarterly time-steps. For the purposes of this paper, we use the HH-HP default rate time series spanning from 2007Q3 to 2017Q4.[3] Moreover, macro and financial data serving as the potential right-hand-side explanatory variables of the model are also incorporated. These variables include real GDP growth, the unemployment rate, real private consumption, CPI inflation, wages, property prices, short and long interest rates, the yield curve slope, foreign demand, investment growth, real export and import growth, and the exchange rate. See Table 1 for the full list for the calibrated HH-HP credit segment.[4]

Figure 1 depicts some of the key variables for the calibrated PD satellite models for both scenarios. GDP is an output of a shocked DSGE prediction model and the rest are calibrated by various satellite models (for more details on the long-term interest rate see Kučera and Szabo, 2019). The red lines show the paths of the macroeconomic variables for a hypothetical crisis that starts immediately and lasts for almost 2 years. The panel also provides an insight into the historical values used to calibrate the PD model. Pre-2000 data are often either not available or not sufficiently reliable, as the Czech banking sector and the whole economy was transformed in

[1] http://www.cnb.cz/en/supervision-financial-market/central-credit-register

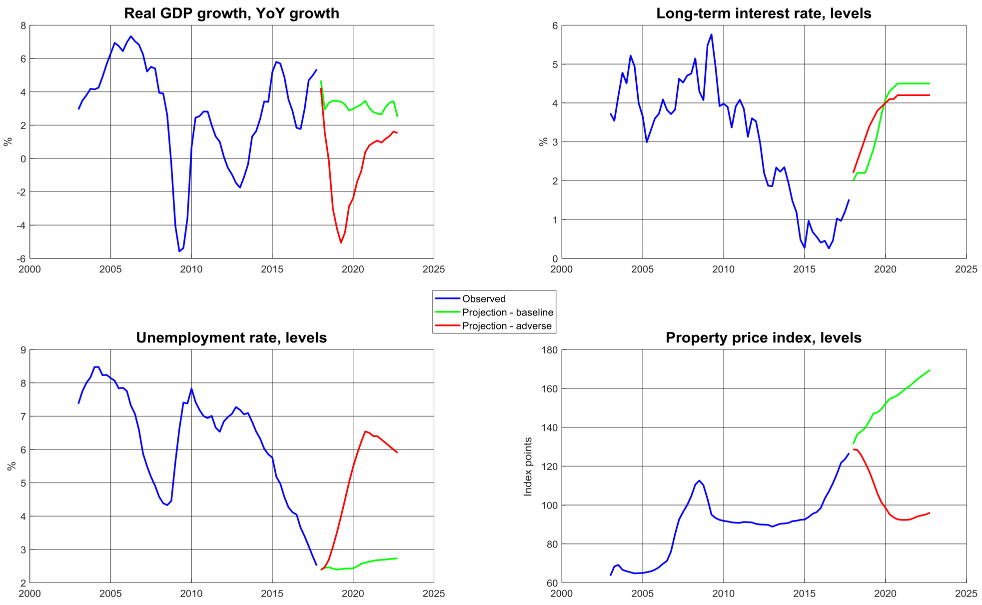
[2] SOLUS (www.solus.cz) and Czech Credit Bureau (www.crif.cz)

[3] For stress-testing purposes, the model is always re-calibrated using the most recent data available for each segment.

[4] The right-hand-side variables used for the NFC and HH-CC segments are included in the Annex (see Table A1 and Table A2, respectively).

the 1990s and large bailouts and portfolio cleanups took place. From the historical perspective, the economy is experiencing some of its best years at the moment, with low unemployment and strong economic growth. On the other hand, property prices are at historical highs and a sudden price correction accompanied by a real GDP contraction, surging unemployment, and rising long-term interest rates could reveal potential vulnerabilities hidden inside the HH-HP portfolio segment of Czech banks.

Figure 1: Overview of the Macroeconomic Scenarios



Note: The figure shows four selected key variables from our dataset (real GDP growth, the long-term interest rate, the unemployment rate, and the property price index). The blue line represents the historical values, the green line the baseline scenario, and the red line the adverse scenario.

3.2 BMA Methodology

The BMA modeling technique explicitly accounts for model uncertainty by working with a large pool of equations and thus helps us find a robust model while avoiding the risks of hand-picking a single equation. Weights in the form of posterior probabilities are assigned to the individual equations, reflecting their predictive performance and other features, and these weights are then used to combine the models into one final posterior equation. The posterior equation is easy to use and communicate.

The pool of equations (the "model space") consists of a large number of equations by considering all possible combinations of explanatory variables and their lags. An autoregressive distributed lag (ARDL) model structure is a convenient basis for forming the model space:

$$Y_t = \alpha + \rho_1 Y_{t-1} + \cdots + \rho_p Y_{t-p} + \sum_{k=1}^{K} \left(\beta_0^{(k)} X_t^{(k)} + \cdots + \beta_{q^{(k)}}^{(k)} X_{t-q^{(k)}}^{(k)} \right) + \varepsilon_t, \quad \forall t. \tag{1}$$

Examining Eq. 1, it is obvious that the dependent variable Y_t is allowed to be a function of its own lags as well as explanatory variables $X^{(k)}$ including their lags. For simplicity, we denote the vector of all model parameters as $\tilde{\boldsymbol{\beta}}$, its elements as $\tilde{\beta}_i$, and their estimates as $\tilde{\boldsymbol{b}}$ and $\tilde{b}_i$, respectively.

The variable Y_t is defined as the probit transform of the 3-month default rate:

$$Y_t \coloneqq \Phi^{-1}(3M\ PD_t), \quad \forall t, \tag{2}$$

where Φ^{-1} is the inverse CDF (i.e., the quantile function) of the standard normal distribution.

Long-run multipliers $\theta^{(k)}$ (LRMs) are coefficients which can be computed for every $X^{(k)}$ and which are defined as

$$\theta^{(k)} \coloneqq \frac{\sigma_{X^{(k)}}}{\sigma_Y} \frac{\partial \mathbb{E}(Y)}{\partial \mathbb{E}(X^{(k)})} = \frac{\sigma_{X^{(k)}}}{\sigma_Y} \frac{\sum_{m=0}^{q_{max}} \beta_m^{(k)}}{1 - \sum_{n=1}^{p_{max}} \rho_n}, \tag{3}$$

where $\sigma_{X^{(k)}}$ and σ_Y are the standard deviations of $X^{(k)}$ and Y, respectively. The LRM for each k can be interpreted as the change in the expected value of Y measured in terms of the number of standard deviations caused by a permanent shift of $X^{(k)}$ by one standard deviation.

Table 1: Explanatory Variables for the HH-HP Segment

Variable	**LRM restriction**	**Options**
Real GDP growth	-	QoQ growth, YoY growth
Unemployment rate	+	Levels, ΔQoQ, ΔYoY
Real private consumption growth	-	QoQ growth, YoY growth
CPI inflation	-	QoQ growth, YoY growth
Real wage growth	-	QoQ growth, YoY growth
Property price index	-	Levels, QoQ growth, YoY growth
Exchange rate CZK/EUR	No sign restriction	Levels, QoQ growth, YoY growth
Long-term interest rates	No sign restriction	Levels, ΔQoQ, ΔYoY
Short-term interest rates	No sign restriction	Levels, ΔQoQ, ΔYoY
Yield curve slope	+	Levels, ΔQoQ, ΔYoY

Note: LRM = Long-run multiplier. Long-term interest rates are measured as the 10-year Czech bond yield and short-term interest rates as the 3-month PRIBOR. The yield curve slope is defined as the difference between the long- and short-term interest rate. "+" means the LRM must be positive and "-" means it must be negative. The LRM restrictions are explained in detail later in this section.

The selected set of right-hand-side variables used for the HH-HP segment includes the macro and financial variables given in Table 1. From this set, which is further augmented by lags of the dependent variable, the model selects in total L variables with the highest explanatory power (either as contemporaneous or in lags), which are then used to form the ARDL structure given by Eq. 1 with the maximum number of K explanatory variables in one equation. All the combinations are considered, so from the L selected variables we can construct in total C models with different right-hand-side variables, where

$$C = \frac{L!}{K!\,(L-K)!}. \tag{4}$$

Moreover, since lags are not forced to be closed (i.e., any variable may appear only as a lag), for each C it is possible to construct $2^{K*(q+1)-1}$ models if the AR component is included, and $2^{K*(q+1)}$ models if the AR component is excluded. Given the computational limits, it is necessary to set a limit on the size of the ARDL structure. For the purposes of this paper, we set the limits as $K = 6$, $L = 10$, and $p_{max} = q_{max} = 2$, which results in almost 20 million configurations considered. We empirically confirmed that the inclusion of higher lags has no significant impact on the model's behavior. This process can also help to limit potential overfitting, which can be an issue for any such method. Adding extra variables to the individual equations would increase the explanatory power on the sample period, but might actually worsen the forecasting performance.

All the macro and financial variables are available in a few options (see the third column in Table 1) and it is left to a step-wise selection algorithm to pick the most appropriate set of size L to be used as the $X^{(k)}$ right-hand-side variables. However, QoQ and YoY options for a given explanatory variable can never be selected simultaneously (but there is no restriction on simultaneously selecting, for example, levels and growths).

3.3 Model Space Constraints

Using the constrained BMA modeling technique for stress-testing purposes has various advantages over the traditional BMA methodology. This is because not all the models in the ARDL structure constructed as described above are econometrically and economically sound, or their predicted PD future paths might not be suitable for stress-testing purposes. Therefore, inspired by Gross and Poblacion (2017), we impose various econometric, economic, and other constraints on the ARDL-based model space (hence the constrained BMA approach) to filter out the unsuitable models. We augment the set of constraints given by Gross and Poblacion (2017). Most importantly, we add benchmark values to reflect prior expectations about the model's forecasting behavior and we specifically take into account the out-of-sample performance of the models.

The econometric and economic constraints are summarized in the following part:

I. Multicollinearity. All the models with a Pearson correlation coefficient above 0.95 between any two explanatory variables are excluded. A certain level of multicollinearity is expected when employing BMA, but very high values can cause matrix invertibility issues when estimating the models.

II. Autocorrelation. All the models with residuals with a Durbin-Watson test p-value below 0.05 are excluded. The AR structure is optimized in the selection process, so the check for autocorrelation is necessary to ensure that only models with a suitable AR structure are selected.

III. Normality. All the models with residuals with a chi-square goodness-of-fit test p-value below 0.05 are excluded. The normality assumption is crucial for the validity of the posterior probability formula (6).

IV. Out-of-sample fit. All the models are re-estimated on a limited sample, and forecasts for a given out-of-sample period are constructed. Models with a worse RMSE-based predictive performance than a naive model given by the simple average of the in-sample observations are excluded. The purpose of this criterion is to limit the potential data-mining problem, which can cause overfitting on the observed dataset while impairing the forecasting performance. The forecasting performance should be tested on the time period that is the most suitable for the model's purpose. In our case, the adverse scenario is usually critical to our stress test and hence we use the part of the data including an economic downturn.[5] Our implementation also allows different out-of-sample periods with different lengths to be specified depending on the stress tester's desire.

V. LRM (Long-run multiplier). There are sign restrictions on the LRM values for some of the macro and financial explanatory variables specified in Table 1. All the models with at least one LRM value violating the restrictions are excluded. The reasoning for this constraint is to ensure a certain level of economic validity of the model (however, we are still dealing here with a reduced-form model with all its potential shortcomings; see also Section 4). Levying different restrictions for different scenarios could be considered, as some relations might not work under non-normal economic conditions and developments.

VI. Benchmark values. The maximum value of the adverse-scenario-conditional and both the maximum and minimum values of the baseline-scenario-conditional forward PD paths must fall within certain intervals. These intervals could be specified either using the stress tester's expert judgment or by employing, for example, a structural or non-parametric benchmark model. This ensures that the model produces forecasts which are plausible and within a priori specified limits. This is a key feature especially for the adverse scenario, since one has ex-ante judgments about the severity of the scenario. For the purposes of this paper, we use a non-parametric model[6] to find the maximum value and the last observed fitted value. Afterwards, intervals around these two values are specified employing expert judgment. Since the model's forecasts represent estimates of the conditional expected value of the process given the macroeconomic scenarios, the idea behind using this method is to construct benchmark intervals around the maximum/most recent estimates of the conditional expected value of the process rather than around the maximum/last observation, which also includes random noise.

VII. Occam's window. All the models with an estimated posterior probability lower than $(0.01 \times \ the\ posterior\ probability\ of\ the\ most\ probable\ model)$ are excluded. This helps us to limit the number of models selected and disregard those models with a negligible contribution to the final model. Hence, this criterion has a very limited impact on the characteristics and properties of the final model, but facilitates ex post evaluation of the individual components, as the suitable yet highly unlikely elements of the model space are filtered out.

3.4 Final Model Estimation

The OLS estimation is followed by the filtration process above and results in a set $\mathcal{F}$ of models satisfying all the constraints introduced. Sticking to the notation established earlier, we compute the posterior vector of the estimated parameters as

[5] We use 20 periods (5 years), as this is the usual span of our stress-testing scenarios.

[6] Specifically, we use double-exponential smoothing, which by design captures well both the level and the local trend of the data.

$$\tilde{\boldsymbol{b}}^{(BMA)} = \sum_{i=1}^{|\mathcal{F}|} \pi_{F_i} \tilde{\boldsymbol{b}}^{(F_i)}, \tag{5}$$

where π_{F_i} is the posterior probability of the model F_i. Assuming the non-informative prior (i.e., assuming the prior probability of each model F_i to be the correct one is uniformly distributed) and under the assumption of normality, the posterior probabilities can be estimated as

$$\pi_{F_i} = \frac{e^{-\frac{1}{2} IC\left(\sigma^2_{\varepsilon, F_i}, d_{F_i}\right)}}{\sum_{j=1}^{|\mathcal{F}|} e^{-\frac{1}{2} IC\left(\sigma^2_{\varepsilon, F_j}, d_{F_j}\right)}}, \tag{6}$$

where IC in the equation above is an information criterion which is selected by the stress tester[7] and computed using the out-of-sample residual variance $\sigma^2_{\varepsilon, F_i}$, and d_{F_i} represents the number of parameters estimated in the model F_i. The out-of-sample residual variance $\sigma^2_{\varepsilon, F_i}$ is estimated using the same out-of-sample period and the same approach as specified in constraint IV above. This further limits any potential data-mining issues, as the weights of the individual components of the final model are based on the out-of-sample performance in the selected period of time. The posterior estimated parameters covariance matrix can be analytically computed as follows:

$$\mathbb{C}\mathrm{ov}\left(\tilde{b}_j^{(BMA)}, \tilde{b}_k^{(BMA)}\right) = \sum_{i=1}^{|\mathcal{F}|} \pi_{F_i} \mathbb{C}\mathrm{ov}\left(\tilde{b}_j^{(F_i)}, \tilde{b}_k^{(F_i)}\right) + \sum_{i=1}^{|\mathcal{F}|} \pi_{F_i} \left(\tilde{b}_j^{(F_i)} - \tilde{b}_j^{(BMA)}\right)\left(\tilde{b}_k^{(F_i)} - \tilde{b}_k^{(BMA)}\right), \quad \forall j, k. \tag{7}$$

The first part of equation (7) accounts for parameter uncertainty, while the second part explicitly accounts for model uncertainty. The final BMA estimates of Y can then be computed in the usual way as

$$\hat{\boldsymbol{Y}}^{(BMA)} = \tilde{\boldsymbol{X}}^{(BMA)} \tilde{\boldsymbol{b}}^{(BMA)}, \tag{8}$$

where $\tilde{\boldsymbol{X}}^{(BMA)}$ is a full matrix of the explanatory variables, including their lags, the constant term, and the lags of the dependent variable. Equation (8) is used to get the historical fit. To get the scenario-conditional forecasts, this equation is used as well, but in sequence for the individual rows of $\tilde{\boldsymbol{X}}^{(BMA)}$ to get the final vector of forecasted values $\hat{Y}_t^{(BMA)}$ employing the forecasted values estimated in the previous steps. To get the final 3M PD values, which can then be easily interpreted and communicated, it is necessary to perform the inverse probit transformation

$$3\widehat{M\,PD}_t^{(BMA)} = \Phi\left(\hat{Y}_t^{(BMA)}\right), \quad \forall t. \tag{9}$$

4. Results and Discussion

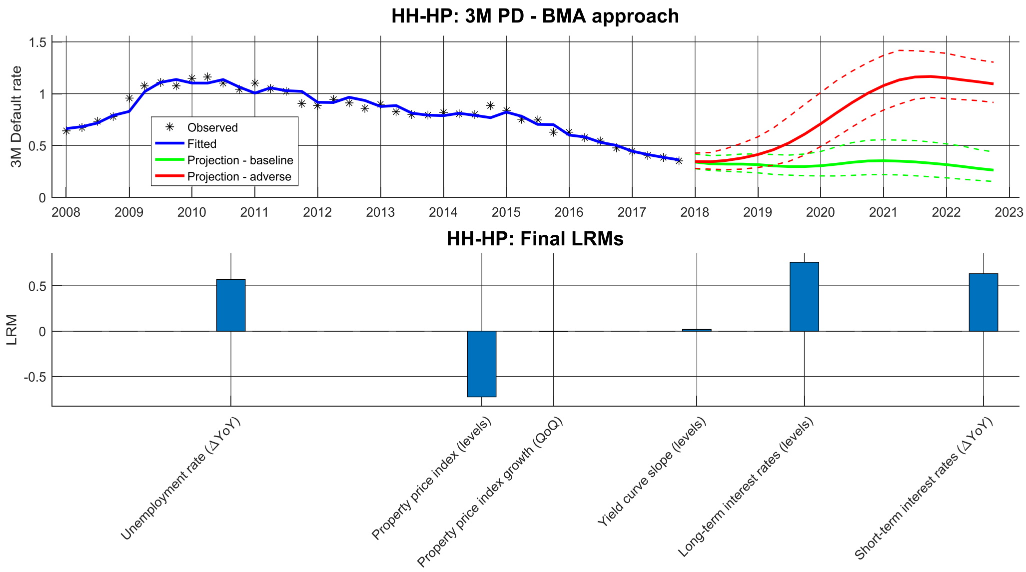
This section shows the application of the framework presented above on the sample calibration for the HH-HP credit segment and briefly discusses the results obtained. The calibration is conducted using the observed historical data as discussed earlier, and two forward PD paths conditional on

[7] The modeler can choose from the Akaike, Schwarz, and Hannan-Quinn information criteria. We observed only a very limited impact of the selected IC on the final model characteristics. For the purposes of this paper, we used only the Hannan-Quinn information criterion in the form $IC\left(\sigma^2_{\varepsilon, F_i}, d_{F_i}\right) = T\ln(\sigma^2_{\varepsilon, F_i}) + 2d_{F_i}\ln(\ln(T))$, where T is the number of observations used in the model estimation.

the hypothetical baseline and adverse scenarios spanning the 5-year period, along with simultaneous confidence intervals, are obtained. The sample calibration results are presented in Figure 2. The sample calibration results for the NFC and HH-CC segments are included in the Annex (see Figure A1 and A2, respectively).

The upper subplot shows the historical observations, the fitted values, and the scenario-conditional forward paths under the two macroeconomic scenarios. We also included the 95% Scheffé's simultaneous confidence intervals for the projected paths, which account for both parameter and model uncertainty. The in-sample fit is solid, with a high R-squared value of over 96%, and the forward PD paths seem well suited to be put to work in the macro-prudential stress test. The adverse scenario PD projection almost reaching the historical sample maxima from the GFC seems justifiable given the severity of the crisis (see Figure 1).

Figure 2: Sample Calibration – HH-HP Segment



Note: The top panel depicts the historical observations (black stars), the fitted values of the model (blue line), and the forecasted conditional expected values of PD under the baseline scenario (green line) and the adverse scenario (red line). The dashed lines represent 95% Scheffé's simultaneous confidential intervals for the projected paths accounting for parameter and model uncertainty. The underlying macroeconomic scenarios are treated as given. The bottom panel shows the final LRM values for all potential explanatory variables, with ticks for the variables with a non-zero LRM only.

Table 2: The Final Posterior Equation Variables and LRMs

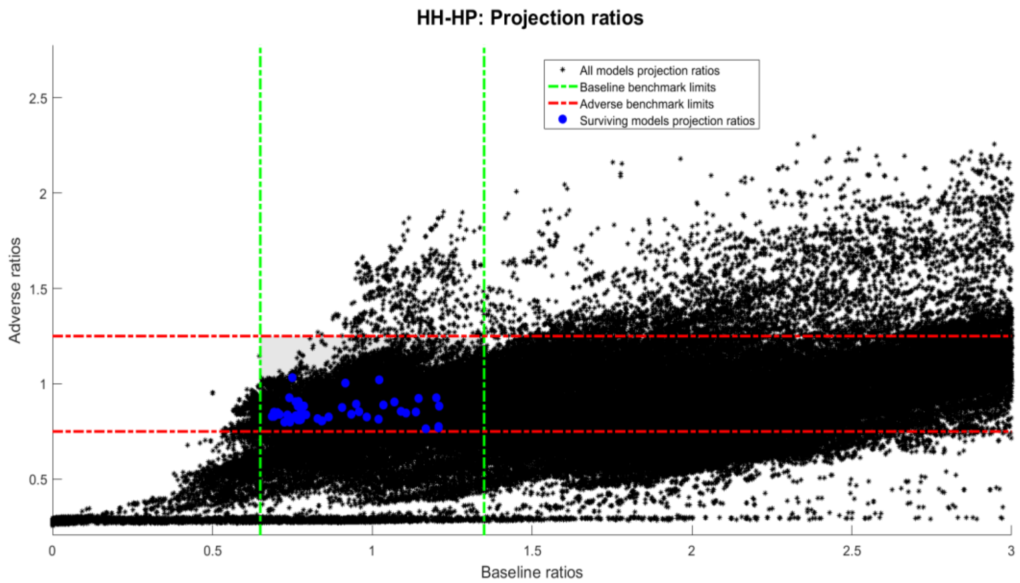
Selected variable	**Option**	**LRM Value**
Unemployment rate	ΔYoY	0.5690
Property price index	Levels	-0.7254
Property price index	YoY growth	-0.0014
Yield curve slope	Levels	0.0189
Long-term interest rates	Levels	0.7593
Short-term interest rates	ΔYoY	0.6334

Note: The table presents the explanatory macroeconomic and financial variables included in the final posterior model equation for the HH-HP segment 3-month PD.

In the lower subplot, the final LRM values are displayed, highlighting that the main risk drivers in the HH-HP segment are changes in the unemployment rate, residential property prices, and interest rates. The LRMs for changes in the unemployment rate and interest rates are positive, suggesting that a positive shock to the explanatory variables leads to a long-term increase in the expected PD. The opposite relation holds true for residential property prices. It should be noted that there may be no actual structural relationship between the variables included in the final model, and the HH-HP PDs and the relationships identified by the model may be just empirical regularities caused by the economic cycle. This is a fundamental shortcoming inherently contained in any universal BMA method or generally any reduced-form econometric model, hence the possible explanatory variables and LRM constraints should be set up with care. The explanatory variables included in the final posterior model equation and their corresponding LRM values are captured in Table 2.

Figure 3 aims to illustrate the filtration procedure (the selection of models that comply with the seven constraints specified in section 3.3). The benchmark values were set as a symmetrical interval of $\pm$ 35% around the most recent value estimated by a non-parametric model (we use double exponential smoothing; see constraint VI) for the baseline scenario, and $\pm$ 25% around the maximum value estimated by the same non-parametric model for the adverse scenario. The horizontal axis shows the baseline scenario ratios of the last value forecasted by the BMA model and the corresponding benchmark value produced by the non-parametric method. Similarly, the adverse scenario ratios (i.e., the ratios of the maximum value forecasted by the BMA model and the corresponding benchmark value produced by the non-parametric method) are shown on the vertical axis. Therefore, all the suitable models must lie within the grey rectangle, with the green lines depicting the benchmark interval for the baseline scenario and the red lines showing the benchmark interval for the adverse scenario. Obviously, many models falling within the grey rectangle are filtered out by the other constraints in Section 3.3. Finally, the blue circles mark the 55 surviving models.

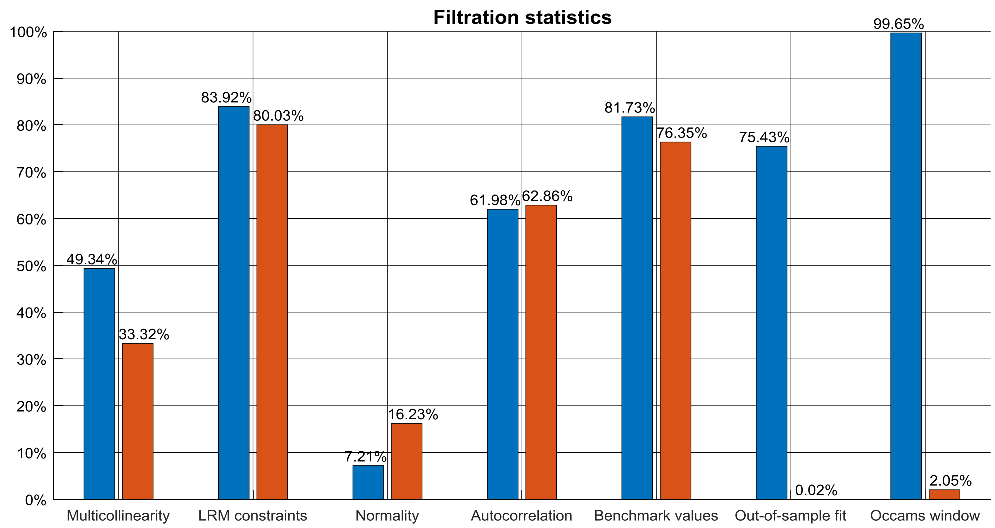
Figure 3: Filtration Process – HH-HP Segment



Note: Illustration of the filtration process for the HH-HP segment. Every black asterisk represents one possible model (not all models are shown, since both axes are limited). Blue dots represent the surviving models, which must fall within the grey rectangle.

We examined the impact of the individual criteria from Section 3.3 on the model space in Figure 4. The blue bars show what percentage of models included in the model space would be filtered out if only the particular constraint was used. This is useful to be aware of; nevertheless, these are simple percentages not reflecting the mass of posterior probability associated with the disregarded models. This metric is therefore also provided and it is represented in the figure by the red bars. For instance, using the Occam's window, 99.7% of all the models is filtered out. However, only models with very low posterior probabilities are filtered out by design, so the impact on the final model qualities is mostly limited as all these filtered out models would only account for about 2% of the posterior probability. This is also the case for the out-of-sample fit constraint, where only models with poor out-of-sample fit are filtered. The metrics suggest that LRM constraints, forecasting benchmark values and autocorrelation are the most impactful model space constraints for the HH-HP segment. It is also important to note that in our implementation for stress-testing purposes, the filtration procedure is nested, so a given constraint is only applied to a set of models that have passed all the upstream criteria. This is a logical shortcut greatly reducing the computing time, since performing every check on every model would usually be too costly for the regular use of the model.

Figure 4: Filtration Statistics – HH-HP Segment

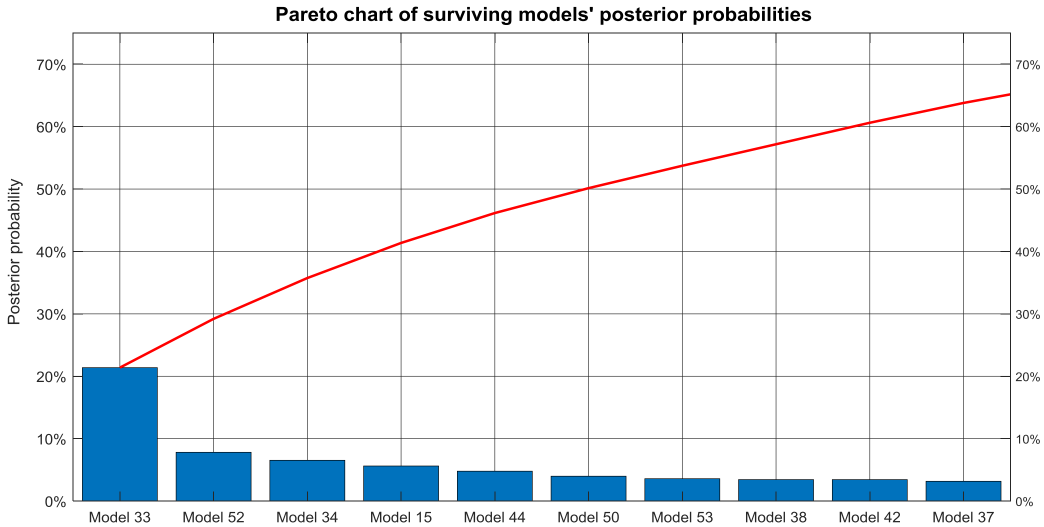


Note: Filtration statistics for the HH-HP segment. Every pair of blue and red bars is associated with one of the constraints from Section 3.3. The blue bars show the percentage of the models affected (filtered out) by the respective constraint where only that particular constraint is applied to the model space. The red bars show the total posterior probability associated with these filtered out models.[8]

Figure 5 focuses on the final stage of the model averaging process. The Pareto chart shows the posterior probabilities of the ten most probable surviving models sorted from the highest to the lowest and shows that the first ten models cover less than 70% of the cumulative posterior probability. This actually promotes the use of the BMA model selection technique, since no single model has a posterior probability above 25%. Hence, we are dealing with several models which check all the criteria and have similar fitting and forecasting qualities. Using BMA directly solves the issue of hand-picking only one equation out of the several dozen models of sufficient quality.

[8] To demonstrate the impact of the individual model space constraints, the posterior probabilities of all models were computed first without any constraint. After that, the constraints were applied one-by-one and the desired filtration statistics were computed.

Figure 5: Posterior Weights of the Surviving Models – HH-HP Segment



Note: Pareto chart showing the posterior probabilities (blue bars) of the surviving models (the blue dots from Figure 3) and their cumulative posterior probability (red line). The models are sorted by posterior probability and those shown in the figure cover more than 65% of the overall posterior probability.

The framework introduced above is very general and, if carefully implemented, could easily be adapted to any financial variable modeled using reduced-form satellite models. However, it is important to note that as with any stress-testing model, the standard quantitative ways of back-testing the adverse scenario forecasts in particular are generally not feasible here and the stress tester must rather focus on the qualitative aspects of the forecast using any available information and expert knowledge.

5. Conclusion

In this paper, we present a straightforward and sufficiently easy-to-implement approach to satellite modeling for macro-prudential stress testing, and present in detail a calibration for the probability of default. We propose the use of a model selection method based on the constrained BMA technique. This technique aims to reduce the risks stemming from model uncertainty by estimating potentially millions of equations given the number of potential explanatory variables and lag structures. Then only the equations fulfilling certain econometric and economic constraints are combined to produce the final posterior model. We employ this framework in scenario-conditional forward-looking projections where we take the paths of the macroeconomic and financial variables as given. The constrained BMA approach addresses several difficulties supervisory authorities face while conducting a stress-testing exercise: (a) the forecasted scenario should be severe yet plausible, (b) reasoning for any additional ex-post expert-based adjustments, and (c) the model specification adjustments with every new data point (i.e. overcoming the challenge of how to select the one single correct equation to be used for creating the projections).

The framework presented in this paper is used in the top-down stress-testing framework at the Czech National Bank, so the data available for the Czech banking sector are used to illustrate the strengths of our approach. This represents a significant improvement compared to the previously used single-equation ARDL and ARIMAX methodology. However, one should keep in mind the model's limitations. Despite the sophisticated model selection procedure, at its core it is still merely a reduced-form model calibrated using a historical data sample. The sample composition is key in determining the scenario-conditional behavior of the variables modeled, since, for instance, if the historical data did not contain any data from the crisis episode, the model's usability for adverse scenario projections would be greatly reduced.

We believe that model averaging techniques like the one we present in this paper could become a vital part of the macro-prudential stress-testing toolkit. Constrained BMA is a robust and powerful tool which can help to overcome risks associated with model uncertainty, improve forecasting performance, and promote a more consistent modeling apparatus. The generality of this framework is potentially another advantage, as it allows for easy adaptation to any financial or macroeconomic variable modeled using basically any appropriate reduced-form satellite model (e.g. loan interest rates, loan flows, bank funding costs). Larger-scale macro-financial model systems can be set up using a BMA method, but in larger-scale equation systems, endogeneity may start to matter, for which BMA in the form it is implemented by this paper is not suited yet. An extended BMA system estimation method accounting properly for endogeneity can be considered in the future. Although our model is calibrated on macro level data, the presented approach also allows for a bank-specific modeling approach providing such data are available.

References

ALESSANDRI P., P. GAI, A. KAPADIA, N. MORA, AND C. PUHR (2009): "Towards a Framework of Quantifying Systemic Stability." *International Journal of Central Banking*, 5(3), pp. 47–81.

BREUER, T., M. JANDACKA, K. RHEINBERGER, AND M. SUMMER (2009): "How to Find Plausible, Severe, and Useful Stress Scenarios." *International Journal of Central Banking*, 5(3), pp. 205–224.

BORIO, C., M. DREHMANN, AND K. TSATSARONIS (2009): "Stress-testing Macro Stress Testing: Does It Live Up to Expectations?" *BIS Working Papers* No 369.

BONTI, G., M. KALKBRENER, C. LOTZ, AND G. STAHL (2006): "Credit Risk Concentrations under Stress." *Journal of Credit Risk*, 2(3), pp. 115–136.

CASTREN O., S. DEES, AND F. ZAHER (2010): "Stress-testing Euro Area Corporate Default Probabilities Using a Global Macroeconomic Model." *Journal of Financial Stability*, 6(2), pp. 64–78.

CIHAK, M. (2009): "Introduction to Applied Stress Testing." *IMF Working Paper* 07/59.

CNB (2019): "Stress Testing – Czech National Bank." [online] *Czech National Bank*. Available at: https://www.cnb.cz/en/financial_stability/stress_testing/ [Accessed March 1, 2019].

DEES, S., J. HENRY, AND M. REINER (2017): "STAMP€: Stress Test Analytics for Macroprudential Purposes in the Euro Area." *European Central Bank*. Available at: doi:10.2866/86845.

FOGLIA, A. (2009): "Stress Testing Credit Risk: A Survey of Authorities' Approaches." *International Journal of Central Banking*, 5(3), pp. 9–45.

GERSL, A., P. JAKUBIK, T. KONECNY, AND J. SEIDLER (2012): "Dynamic Stress Testing: The Framework for Testing Banking Sector Resilience Used by the Czech National Bank." *Czech National Bank Working Paper* 11/2012.

GRAY, D., M. GROSS, J. PAREDES, AND M. SYDOW (2013): "Modeling Banking, Sovereign, and Macro Risk in a CCA Global VAR." *IMF Working Paper* 13/218.

GROSS, M. AND J. POBLACION (2017): "Implications of Model Uncertainty for Bank Stress Testing." *Journal of Financial Services Research*, 55(1), pp. 31–58.

HENRY, J. AND C. KOK (2013): "A Macro Stress Testing Framework for Assessing Systemic Risks in the Banking Sector." *European Central Bank Occasional Paper* No 152.

HIRTLE B., A. KOVNER, J. VICKERY, AND M. BHANOT (2016): "Assessing Financial Stability: The Capital and Loss Assessment under Stress Scenarios (CLASS) Model." *Journal of Banking and Finance*, 69(S1), pp. 35–55.

HJORT, N. L. AND G. CLAESKENS (2003): "Frequentist Model Average Estimators." *Journal of the American Statistical Association*, 98(464), pp. 879–899.

HOETING, J. A., D. MADIGAN, A. E. RAFTERY, AND C. T. VOLINSKY (1999): "Bayesian Model Averaging: A Tutorial." *Statistical Science*, 14(4), pp. 382–401.

KAPINOS, P. AND O. MITNIK (2016): "A Top-down Approach to Stress-testing Banks." *Journal of Financial Services Research*, 49(2–3), pp. 229–264.

KUČERA, A. AND M. SZABO (2019): "Estimating the Neutral Czech Government Bond Yield Curve." Thematic article on financial stability 3/2019, *Czech National Bank*

LOUZIS, D. P., A. T. VOULDIS, AND V. L. METAXAS (2012): "Macroeconomic and Bank-specific Determinants of Non-performing Loans in Greece: A Comparative Study of Mortgage, Business and Consumer Loan Portfolios." *Journal of Banking and Finance*, 36(4), pp. 1012–1027.

SCHMIEDER, C., C. PUHR, AND M. HASAN (2009): "Next Generation Balance Sheet Stress Testing." *IMF Working Paper* 11/83.

SIEMSEN, T. AND J. VILSMEIER (2017): "A Stress Test Framework for the German Residential Mortgage Market: Methodology and Application." *Deutsche Bundesbank Discussion Paper* No. 37/2017.

SORGE, M. AND K. VIROLAINEN (2006): "A Comparative Analysis of Macro Stress-testing Methodologies with Application to Finland." *Journal of Financial Stability*, 2(2), pp. 113–151.

STEEL, M. F. (2019): "Model Averaging and Its Use in Economics." *Journal of Economic Literature*, forthcoming

VAZQUEZ, F., B. M. TABAK, AND M. SOUTO (2012): "A Macro Stress Test Model of Credit Risk for the Brazilian Banking Sector." *Journal of Financial Stability*, 8(2), pp. 69–83.

WILSON, T. C. (1998): "Portfolio Credit Risk." *Economic Policy Review*, 4(3).

Appendix A: Additional Calibrations

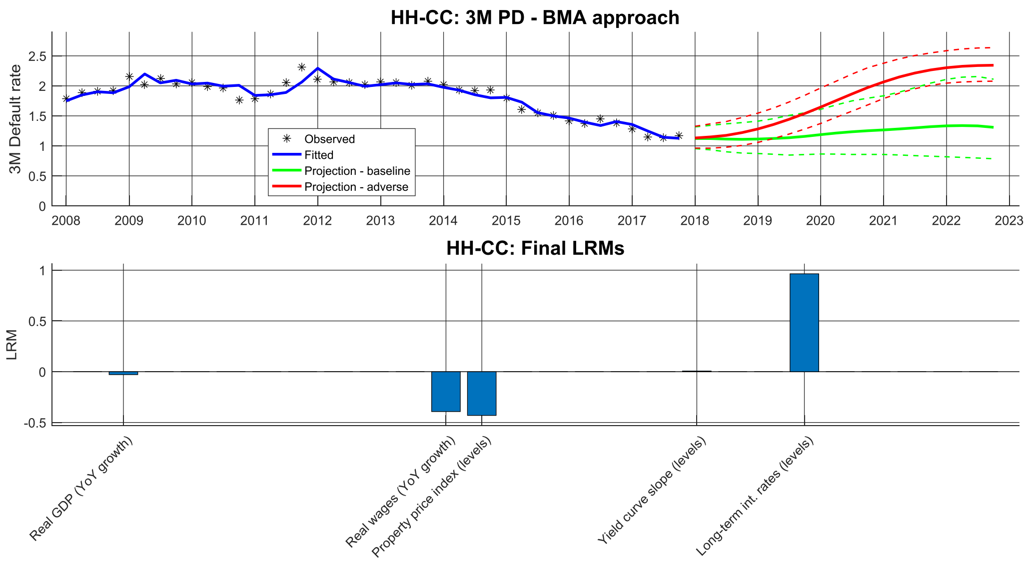
In this part of the paper, we add calibration details for two additional credit segments – consumer credit and loans to non-financial corporations.

Table A1: Explanatory Variables for the HH-CC Segment

Variable	LRM restriction	Options
Real GDP growth	-	QoQ growth, YoY growth
Unemployment rate	+	Levels, ΔQoQ, ΔYoY
Real private consumption growth	-	QoQ growth, YoY growth
CPI inflation	-	QoQ growth, YoY growth
Real wage growth	-	QoQ growth, YoY growth
Property price index	-	Levels, QoQ growth, YoY growth
Exchange rate CZK/EUR	No sign restriction	Levels, QoQ growth, YoY growth
Long-term interest rates	No sign restriction	Levels, ΔQoQ, ΔYoY
Short-term interest rates	No sign restriction	Levels, ΔQoQ, ΔYoY
Yield curve slope	+	Levels, ΔQoQ, ΔYoY

Note: LRM = Long-run multiplier. Long-term interest rates are measured as the 10-year Czech bond yield and short-term interest rates as the 3-month PRIBOR. The yield curve slope is defined as the difference between the long- and short-term interest rate. "+" means the LRM must be positive and "-" means it must be negative.

Figure A1: Sample Calibration – HH-CC Segment



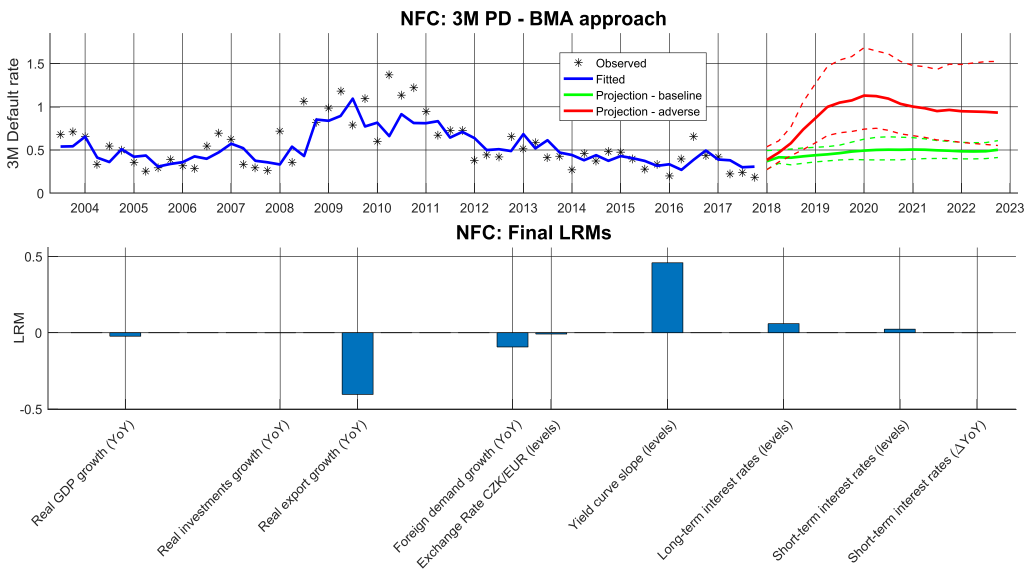
Note: The top panel depicts the historical observations (black stars), the fitted values of the model (blue line), and the forecasted conditional expected values of PD under the baseline scenario (green line) and the adverse scenario (red line). The dashed lines represent 95% Scheffé's simultaneous confidence intervals for the projected paths accounting for parameter and model uncertainty. The underlying macroeconomic scenarios are treated as given. The bottom panel shows the final LRM values for all potential explanatory variables, with ticks for the variables with a non-zero LRM only.

Table A2: Explanatory Variables for the NFC Segment

Variable	LRM restriction	Options
Real GDP growth	-	QoQ growth, YoY growth
CPI inflation	-	QoQ growth, YoY growth
Real investment growth	-	QoQ growth, YoY growth
Real export growth	-	QoQ growth, YoY growth
Real import growth	No sign restriction	QoQ growth, YoY growth
Foreign demand growth	No sign restriction	QoQ growth, YoY growth
Exchange rate CZK/EUR	No sign restriction	Levels, QoQ growth, YoY growth
Long-term interest rates	No sign restriction	Levels, ΔQoQ, ΔyoY
Short-term interest rates	No sign restriction	Levels, ΔqoQ, ΔyoY
Yield curve slope	+	Levels, ΔqoQ, ΔyoY

Note: LRM = Long-run multiplier. Long-term interest rates are measured as the 10-year Czech bond yield and short-term interest rates as the 3-month PRIBOR. The yield curve slope is defined as the difference between the long- and short-term interest rate. "+" means the LRM must be positive and "-" means it must be negative.

Figure A2: Sample Calibration – NFC Segment



Note: The top panel depicts the historical observations (black stars), the fitted values of the model (blue line), and the forecasted conditional expected values of PD under the baseline scenario (green line) and the adverse scenario (red line). The dashed lines represent 95% Scheffé's simultaneous confidence intervals for the projected paths accounting for parameter and model uncertainty. The underlying macroeconomic scenarios are treated as given. The bottom panel shows the final LRM values for all potential explanatory variables, with ticks for the variables with a non-zero LRM only.

CNB WORKING PAPER SERIES (SINCE 2018)

9/2019	Jiří Panoš Petr Polák	*How to improve the model selection procedure in a stress-testing framework*
8/2019	Sebastian Gechert Tomáš Havránek Zuzana Iršová Dominika Kolcunová	*Death to the Cobb-Douglas production function? A quantitative survey of the capital-labor substitution elasticity*
7/2019	Alexis Derviz	*Coexistence of physical and crypto assets in a stochastic endogenous growth model*
6/2019	Dominika Ehrenbergerová Simona Malovaná	*Introducing macro-financial variables into semi-structural model*
5/2019	Martin Hodula	*Monetary policy and shadow banking: Trapped between a rock and a hard place*
4/2019	Simona Malovaná Žaneta Tesařová	*Banks' credit losses and provisioning over the business cycle: Implications for IFRS 9*
3/2019	Aleš Bulíř Jan Vlček	*Monetary policy is not always systematic and data-driven: Evidence from the yield curve*
2/2019	Dominika Kolcunová Simona Malovaná	*The effect of higher capital requirements on bank lending: The capital surplus matters*
1/2019	Jaromír Baxa Tomáš Šestořád	*The Czech exchange rate floor: Depreciation without inflation?*
19/2018	Jan Brůha Jaromír Tonner	*Independent monetary policy versus common currency: The macroeconomic analysis for the Czech Republic through lens of an applied DSGE model*
18/2018	Tomáš Adam Filip Novotný	*Assessing the external demand of the Czech economy: Nowcasting foreign GDP using bridge equations*
17/2018	Kamil Galuščák Jan Šolc Pawel Strzelecki	*Labour market flows over the business cycle: The role of the participation margin*
16/2018	Martin Hodula	*Off the radar: Exploring the rise of shadow banking in the EU*
15/2018	Lukáš Pfeifer Martin Hodula Libor Holub Zdeněk Pikhart	*The leverage ratio and its impact on capital regulation*
14/2018	Martin Gürtler	*What influences private investment? The case of the Czech Republic*
13/2018	Václav Hausenblas Jitka Lešanovská	*How do large banking groups manage the efficiency of their subsidiaries? Evidence from CEE*
12/2018	Simona Malovaná	*The pro-cyclicality of risk weights for credit exposures in the Czech Republic*
11/2018	Tibor Hlédik Karel Musil Jakub Ryšánek Jaromír Tonner	*A macroeconomic forecasting model of the fixed exchange rate regime for the oil-rich Kazakh economy*
10/2018	Michal Franta Tomáš Holub Branislav Saxa	*Balance sheet implications of the Czech National Bank's exchange rate commitment*

9/2018	Dominika Kolcunová Tomáš Havránek	*Estimating the effective lower bound on the Czech National Bank's policy rate*
8/2018	Volha Audzei Sergey Slobodyan	*Sparse restricted perception equilibrium*
7/2018	Tibor Hlédik Jan Vlček	*Quantifying the natural rate of interest in a small open economy – The Czech case*
6/2018	Václav Brož Michal Hlaváček	*What drives distributional dynamics of client interest rates on consumer loans in the Czech Republic? A bank-level analysis*
5/2018	Lukáš Pfeifer Martin Hodula	*Profit-to-provisioning approach for setting the countercyclical capital buffer: The Czech example*
4/2018	Ivan Sutóris	*Asset prices in a production economy with long run and idiosyncratic risk*
3/2018	Michal Franta	*The likelihood of effective lower bound events*
2/2018	Soňa Benecká Ludmila Fadejeva Martin Feldkircher	*Spillovers from euro area monetary policy: A focus on emerging Europe*
1/2018	Jan Babecký Clémence Berson Ludmila Fadejeva Ana Lamo Petra Marotzke Fernando Martins Pawel Strzelecki	*Non-base wage components as a source of wage adaptability to shocks: Evidence from European firms, 2010-2013*

CNB RESEARCH AND POLICY NOTES (SINCE 2018)

2/2019	Jan Filáček Ivan Sutóris	*Inflation targeting flexibility: The CNB's reaction function under scrutiny*
1/2019	Iveta Polášková Luboš Komárek Michal Škoda	*The contemporary role of gold in central banks' balance sheets*
1/2018	Mojmír Hampl Tomáš Havránek	*Central bank financial strength and inflation: A meta-analysis*

CNB ECONOMIC RESEARCH BULLETIN (SINCE 2018)

November 2018	*Interest rates*
May 2018	*Risk-sensitive capital regulation*

Czech National Bank
Economic Research Division
Na Příkopě 28, 115 03 Praha 1
Czech Republic
phone: +420 2 244 12 321
fax: +420 2 244 12 329
http://www.cnb.cz
e-mail: research@cnb.cz
ISSN 1803-7070