DEUTSCHE
BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 31/2018

On a quest for robustness: about model risk, randomness and discretion in credit risk stress tests

Thomas Siemsen
Johannes Vilsmeier

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

ISBN 978–3–95729–490–6 (Printversion)
ISBN 978–3–95729–491–3 (Internetversion)

Non-technical summary

Research Question

Following the financial crisis, bank-internal and supervisory stress testing has gained importance in risk management and supervisory practice. As in any exercise with scarce data and highly collinear regressors, stress testing is susceptible to model uncertainty. Given the prominent role of stress tests, quantifying the extent of model uncertainty and developing tools to mitigate its effects on stress test results is of ample importance for both banks and supervisors. This paper takes a first step in trying to address both of these challenges.

Contribution

Our contributions are twofold: first, we quantify the impact of model uncertainty on credit risk stress test results using the German banking sector as an example. We define "model uncertainty" as the range of stress test results that are supported by equally plausible model frameworks. Second, we propose a stress testing framework that mitigates the impact of model uncertainty by considering a battery of different models rather then just one model, filtering the models for statistical, economic and "stress testing" plausibility and finally combining the surviving specifications into a single model.

Results

We find that - depending on model specification - the dispersion in predicted increases of default probabilities during a stress horizon can be huge. In our application they lie in a range from $-90\,\%$ to $+7,000\,\%$. This leaves stress test results subject to significant model uncertainty. Our proposed framework eliminates "stress test implausible" predictions (relative to a second benchmark model) at both ends of the stressed forecast distribution. This leads in our application to the German banking sector to a reduction in the median stress effect, compared to the model without the "stress testing plausibility" filter, from -5.0 pp of CET1 capital to -2.5 pp.

Nichttechnische Zusammenfassung

Fragestellung

Im Zuge der Finanzkrise haben bank-interne und aufsichtliche Stresstests im Risikomanagement und in der aufsichtlichen Praxis an Bedeutung gewonnen. Wie alle Übungen, die versuchen Variablen mit wenigen Beobachtungen und mit stark kollinearen Regressoren zu prognostizieren, unterliegen auch Stresstests einer Modellunsicherheit, die die Robustheit der Ergebnisse beeinflussen kann. Vor dem Hintergrund der prominenten Rolle von Stresstests ist die Quantifizierung der Effekte von Modellunsicherheit auf Stresstestergebnisse und die Entwicklung von Methoden, um diese zu reduzieren, sowohl für Banken als auch für die Aufsicht von hoher Relevanz.

Beitrag

Dieses Papier quantifiziert am Beispiel eines Kreditrisikostresstests für das deutsche Bankensystem den Einfluss von Modellunsicherheit auf Stresstestergebnisse. "Modellunsicherheit" ist dabei definiert als die Bandbreite möglicher Stresstestergebnisse, die durch zunächst gleichsam plausible Modelle generiert wird. Wir schlagen einen Modellrahmen vor, der die Modellunsicherheit reduzieren kann, indem er (anstelle von einem) viele Modelle berücksichtigt, diese nach statistischer, ökonomischer und „Stresstest"-Plausibilität filtert und die verbleibenden Modelle zu einem einzigen Modell kombiniert.

Ergebnisse

Unsere Ergebnisse legen nahe, dass der Einfluss von Modellunsicherheit auf Stresstestergebnisse potentiell stark sein kann. Betrachtet man beispielsweise den Anstieg der Ausfallwahrscheinlichkeit über den Stresshorizont, so kann dieser - je nach Modellspezifikation - zwischen $-90\,\%$ und $+7000\,\%$ liegen. Unser Stresstestmodell eliminiert (gemessen an einem strukturellen Vergleichsmodell) "unplausible" Stressprognosen an beiden Enden der Prognoseverteilung. In unserer Anwendung auf den deutschen Bankenmarkt führt dies dazu, dass der mediane Stresseffekt von -5.0 Pp der harten Kernkapitalquote auf -2.5 Pp zurück geht.

Bundesbank Discussion Paper No 31/2018

On a quest for robustness: About model risk, randomness and discretion in credit risk stress tests*

Thomas Siemsen
Deutsche Bundesbank

Johannes Vilsmeier
Deutsche Bundesbank

Abstract

In this paper we study the impact of model uncertainty, which occurs when linking a stress scenario to default probabilities, on reduced-form credit risk stress testing. This type of uncertainty is omnipresent in most macroeconomic stress testing applications due to short time series for banks' portfolio risk parameters and highly collinear macroeconomic covariates. We quantify the effect of model uncertainty on supervisory and bank stress tests in terms of predicted portfolio loss distributions and implied capital shortfalls by conducting a full-fledged top-down credit risk stress test for over 1,500 German banks. Our results suggest that the impact of model uncertainty on predicted capital shortfalls can be huge, even among models with similar predictive power. This leaves both banks and supervisors with uncertainty when calculating stress impacts and implied capital requirements. To mitigate the impact of uncertainty, we suggest a modeling approach which filters the model space by combining the standard Bayesian model averaging (BMA) paradigm with a structural filter derived from the Merton/Vasicek credit risk model. Applying our stress testing framework, the dispersion decreases and the median stress effect is reduced from -5.0pp of CET1 ratio under the BMA model to -2.5pp under the structurally augmented BMA model, while the predicted capital shortfall is reduced by 70 %. The structural filter eliminates extreme outcomes on both sides of the stress forecast distribution, leading in our application to the German banking sector to a reduction in impact compared to the model without the "stress testing plausibility" filter.

Keywords: model uncertainty, stress test, Bayesian model averaging, quantile mapping, credit risk

JEL classification: C11, C52, G21

*We are grateful to Max Sommer for his excellent guidance on data preparation. Contact address: Deutsche Bundesbank, Wilhelm-Epstein Strasse 14, 60431 Frankfurt am Main, Germany, e-mail: *thomas.siemsen@bundesbank.de*, *johannes.vilsmeier@bundesbank.de*.

1 Introduction

Macro-financial stress tests have become a major supervisory tool for assessing the resilience and capital needs of financial institutions. The 2009 Supervisory Capital Assessment Program (SCAP) conducted by the Federal Reserve Board as a response to the 2008 US banking crisis, may be regarded as the starting point for using comprehensive supervisory stress testing exercises as a tool to determine capital needs of banks and to restore financial stability. In the following period, stress tests were used to recapitalize the banking systems of a number of European crisis countries such as Ireland (2011), Spain (2012) and Greece (2015). In 2014, the ECB – in collaboration with EBA – carried out one of the most extensive stress test exercises to date (called the Comprehensive Assessment or CA) which aimed at ensuring an adequate capitalization of the largest banks in the euro area before taking over the responsibility for banking supervision of these banks from the national competent authorities.

More generally, as pointed out in the guidelines for the supervisory review and evaluation process (SREP) (European Banking Authority, 2014b), outcomes of supervisory stress tests should be used to assess and calibrate banks' (Basel II / Basel III) Pillar 2 capital add-ons. The outcomes of supervisory stress tests thus augment the results of stress tests which supervisors request banks to carry out as part of their internal risk management processes. In particular, since Basel II stress tests have had to form an integral part of banks internal capital adequacy assessment program (ICAAP). The outcomes of the ICAAP are assessed by supervisors and constitute an important input for the SREP decision, which can lead – like supervisory stress tests – to additional Pillar 2 capital add-ons.[1] Supervisors also ask banks to use stress test outcomes (including reverse stress tests) to set up their internal recovery and resolution plans (see for example Financial Stability Board, 2013; European Banking Authority, 2014c).

Crucially, beginning 2018, the implementation of the IFRS 9 accounting standards extends the scope of "stress testing" beyond the regulatory and risk-management border to the accounting world. Under IFRS 9, banks are required to switch from the prevailing "incurred loss accounting" model to the more forward-looking "expected credit loss" (ECL) model for provisioning. To compute ECL, banks need to take into account any relevant information including forward-looking and potentially scenario-dependent factors such as adverse macroeconomic developments, which is in a similar spirit to conducting a stress test (see European Banking Authority, 2016b).

Given the importance assigned to outcomes of stress tests for supervisors' and banks' decisions, the question must be raised as to how reliable these outcomes are. In recent years, after supporting the increased usage of internal risk models to calculate capital requirements since Basel II, supervisors have raised some doubts about the general reliability of the models used by banks to calculate capital requirements. Supervisory work streams have been put in place to assess revisions to the internal rating based (IRB) framework due to large discretions found in banks' IRB capital requirements calculations (see for example European

[1] Under Pillar 1 of Basel II/Basel III the minimum capital requirements defined in the Capital Requirements Regulation (CRR) are determined based on a unified framework for all banks. Under Pillar 2 (among others) potentially additional capital needs – using banks internal risk management processes – are assessed.

Banking Authority, 2013, 2014a), and an updated SREP methodology was introduced which asks supervisors to challenge banks ICAAP outcomes more strongly by using supervisory benchmarks (see European Banking Authority, 2014b). More recently, a floor to the calculation of risk-weights under the IRB approach was suggested and, later, agreed on. The floor limits banks' discretion in reducing risk weights too far below those of the standardized approach (see for example Basel Committee on Banking Supervision, 2016).

Supervisory stress tests, like the EBA/ECB CA 2014, aim to cover all relevant risk drivers for banks' profitability and solvency. These are generally decomposed to net interest income (NII), credit risk, market risk and operational risk (see for example European Banking Authority, 2017). Usually the reduction in NII, which forms lending banks' fundamental income source and credit risk, as banks' major source of losses may be regarded as the core risk drivers of supervisory stress test outcomes (see for example European Banking Authority, 2016a).

Regarding the modeling of NII, Bolotny, Edge, and Guerrieri (2015) raise the question about the reliability of macroeconomic stress test forecasts using an array of different modeling choices. They come to the conclusion that forecast uncertainty stemming from model risk is not reflected in the current quantitative capital assessment framework, despite the high level of model risk prevailing in stress test exercises. Following the lead of Bolotny et al. (2015), our paper assesses the reliability and the impact of model risk on stress test outcomes from credit risk. Thereby, we define "model risk" as the range of stress test results (e.g. in terms of CET1 impact) that are supported by ex ante equally plausible model frameworks.

While model uncertainty is a frequently addressed issue in other research areas such as in GDP forecasting, unemployment forecasting, growth theory or the analysis of stock returns, (e.g. Doppelhofer, Miller, and Sala-i-Martin, 2004; Avramov, 2002; Cremers, 2002; Stock and Watson, 2005; Wright, 2008; Garratt, Mitchell, Vahey, and Wakerly, 2011), this has hardly featured so far in the literature on credit risk stress testing. This seems rather surprising, since, in general, the same type of models and set of regressor variables are used for forecasting default probabilities (PDs) as, for example, in the GDP or unemployment forecasting literature (see, for example, Wilson, 1997a,b; Sorge and Virolainen, 2006; Castrén, Dées, and Zaher, 2010; Vazquez, Tabak, and Souto, 2012; Schechtman and Gaglianone, 2012; Schuermann, 2014; Covas, Rump, and Zakrajšek, 2014). Possible solutions suggested in the literature to address model uncertainty are, for instance, the statistical concepts of boosting, shrinkage or forecast/model combination methods.

The model risk or model uncertainty found in macroeconomic forecasts is usually induced by scarce data and a highly correlated set of regressors which lead to non-unique parameter estimates. In such a statistical environment reduced-form models are very sensitive to small changes in the data and/or the inclusion or exclusion of explanatory variables (see, for example, Belsley, Kuh, and Welsch, 1980, among other). Stress tests in addition – as pointed out, for example, by Misina and Tessier (2008) – face the problem that effects at the edge of the "observation space" of variables are assessed, where linear approximations of regression based models, focusing on average observations in the sample period, are likely to perform poorly. This issue is aggravated by the fact that in a scare-data-high-collinearity environment forecasts outside the observation space suffer from high forecast variability. These issues translate into additional sources of model uncertainty, since none of the – usually OLS-based

and linear – models properly describes the true, potentially non-linear relationship.
Credit risk stress test studies which deal explicitly with issues of model uncertainty are – to our knowledge – restricted to Misina and Tessier (2008), who theoretically emphasize the need to account for model risk and non-linearity in credit risk stress tests, Henry and Kok (2013) who use an OLS based model combination approach called Bayesian model averaging (BMA) to carry out their credit risk stress test exercise, and Gross and Poblacion (2016).[2] More implicitly – by attempting to avoid the potential inconsistencies when modeling the impact of stressed macroeconomic conditions on PDs via OLS – the stress testing approach suggested by Bonti, Kalkbrener, Lotz, and Stahl (2006), too, can be classified as a study designed to deal with model risk. Bonti et al. (2006) apply a rather simplistic approach which builds on the Merton-Vasicek factor model for credit risk (Gupton, Finger, and Bhatia, 1997; Vasicek, 2002) underlying for example the Basel II IRB formula to derive risk weighted assets for credit exposure (Basel Committee on Banking Supervision, 2005). Their approach rests on a mapping procedure based on historical distributions of the variables. The historical probability for, say, GDP growth levels assumed in a scenario are translated to a systemic factor with the same historical probability of occurrence, i.e. a perfectly co-monotonic relationship is assumed between the two variables. The systemic factor is then translated into a corresponding PD through the Merton-Vasicek factor model.
Our contributions to the literature on model uncertainty in a stress testing context are twofold: *first*, we quantify the effects of model uncertainty on top-down credit risk stress test results for the German banking sector. To this end, we conduct a comprehensive credit risk stress test for 1,500 German banks and study the influence of uncertainty in model specifications that link macro variables to PD dynamics on final stress test results. *Second*, we propose a framework to address and mitigate the effects of model uncertainty in top-down stress test exercises. This framework combines the state-of-the-art BMA approach as employed, for example, in Henry and Kok (2013) and Gross and Poblacion (2016) with a structural filter derived from a Merton/Vasciek-type model. The structural filter augments the reduced-form perspective of state-of-the-art stress testing frameworks with predicted stress dynamics derived from an economic model rather than from historical correlations. In a stress testing context historical correlations are less informative than for standard business cycle forecasting, since adverse stress scenarios induce dynamics mostly outside of the observation space. For top-down supervisory credit risk stress tests this issue is oftentimes aggravated by short time series of credit risk parameters and collinear macro regressors, which makes predictions outside of the observations space unstable and very sensitive to model specifications. To mitigate the issue of model uncertainty, the structural filter eliminates specifications from the model space which are deemed "stress test implausible" relative to the structural benchmark model prior to applying a BMA aggregation. In addition, our framework extends the baseline BMA model of Gross and Poblacion (2016) to the use out-of-sample model weights. We resort to out-of-sample weights since the lack of sufficient crisis observations in historic PD time-series renders a good out-of-sample forecasting performance

[2]In an application of the general methodology proposed in this paper, Siemsen and Vilsmeier (2017) conduct a top-down stress test of the German residential mortgage market. Pelster and Vilsmeier (2017) apply a similar framework to the market for CDS. See their paper for an complementary discussion of the methodology applied.

relatively more important than a good in-sample performance.[3] Last but not least, to map the macro scenario to the structural benchmark model, we modify the "quantile mapping" (QMap) framework of Bonti et al. (2006) in a way that allows it to consider model risk in a way similar to BMA, i.e. by averaging over the entire set of plausible models in the model space. We show how one can easily map multiple macro variables to one systematic factor and, hence, take into account several scenario variables, and how to use the latent factor approach in a dynamic, multi-period stress test setting.

Our findings have major implications for supervisors as well as for banks. The results of our stress tests suggest that even when filtering the model space for economically plausible and statistically sound models, the range of possible stress test outcomes and implied capital depletions is still huge. For example, in our application of the framework to the German banking sector we find that for an exemplary sector the unfiltered model space predicts PD increases due to the adverse macroeconomic scenario in a very wide range between $-90\,\%$ and $+7,000\,\%$. Filtering this model space for statistically and economically plausible specification reduces this range to between $20\,\%$ and $500\,\%$. This implies that for the initial sectoral PD of $2.5\,\%$ the stressed PD at the end of the scenario horizon could lie in the range between $3\,\%$ and $15\,\%$. Crucially, since the expected loss model that most top-down credit risk stress tests employ to predict impairments during the stress horizon is linear in default probabilities, the model uncertainty in PDs translates directly to dispersion in expected loss and thus stress outcomes (ceteris paribus). This uncertainty is to be kept in mind when supervisors and also banks themselves assess the outcomes from ICAAP and supervisory stress tests. By introducing the additional structural benchmark filter to the model space (benchmark-constrained BMA or BCBMA), we further reduce model uncertainty related to predicted PD increases by informing the model space with stress predictions from the structural PD model. Thereby, model uncertainty is reduced because the benchmark filter is designed such that it does not suffer from the same qualitative and quantitative ambiguity (due to the reasons mentioned above) as the reduced-form models. Coming back to the example sector, the benchmark filter reduces the range of predicted PD increases to between $60\,\%$ and $150\,\%$ such that the sectoral initial PD of $2.5\,\%$ would end up in a region between $4\,\%$ and $6.25\,\%$. Conducting a full-fledged top-down credit risk stress test shows that the reduction in dispersion can have strong effects on the measured capital shortfall at the aggregate level and thus on both the qualitative and quantitative interpretation of stress test results. In our application to the German banking sector, we find the the aggregate capital shortfall is reduced by $70\,\%$ comparing the benchmark constrained with the unconstrained results. These results suggest that, rather than looking at the outcomes of single models, filtered model spaces (model confidence sets or MCS) and their related forecast distributions should be taken into account when assessing banks' internal (and also supervisors') models.[4]

The remainder of this paper is structured as follows. Section 2 sets out our stress testing framework by describing how we link the macro scenario to credit risk parameters. In particular, Section 2.1 introduces the BCBMA model. Section 2.3 shows how we translate

[3] We say "relatively" since the use of out-of-sample weights does not mitigate the issue that a linear model may not be able to capture non-linear crisis dynamics very well.

[4] Note that we use the term "model confidence set" with a slight abuse of notation, as unlike MCS in the Hansen, Lunde, and Nason (2011) sense, we do not conduct a formal testing that the best model is contained in the MCS with a specific probability.

stressed credit risk parameter dynamics to bank-specific capital impacts. Section 3 discusses our data sources. Section 4 elaborates on the effect of model uncertainty on stress test results by comparing the results for the unfiltered BMA model space, the filtered model space and the benchmark- constrained model space. Finally, Section 5 concludes.

2 Methodology

Most frameworks for credit risk stress testing resort to an expect loss (EL) model (see, for instance, Board of Govenors of the Federal Reserve System, 2016; Bank of England, 2017; European Banking Authority, 2017). This model predicts future impairments due to credit risk as the estimated losses on defaulted exposures, i.e.

$$EL_t = LGD_t \times PD_t \times EaD_t \,, \tag{1}$$

where LGD_t denotes the loss given default, PD_t denotes the probability of default over a given time horizon (e.g. one year) and EaD_t denotes the relevant exposure at default. Therefore, $PD_t \times EaD_t$ is the expected default flow over the defined time horizon. While expected losses can readily be computed for standard business cycle fluctuations using banks' internal parameter estimates for PD and LGD, which are, for example, calibrated to historical realizations, computing expected losses in a multi-period *stress test* context is more daunting. Here, the credit risk parameters need to be estimated conditional on an adverse macroeconomic scenario that lies in the very tails of the historical distributions or even beyond. Thus, when estimating the co-movement of macro variables and risk parameters in these crisis times, risk managers and supervisors have to base their estimation on very few, if any observations.

As an illustration, Figure 1 shows the post-stress percentiles of each country participating in the 2018 EU-wide stress test in their respective historical GDP growth distribution. While some countries experience an adverse cumulative GDP growth over the three-year scenario close to or even above their 20^{th} percentile, most countries experience growth very close to the left tail of their historical distribution, i.e. growth which has rarely been observed before. The sparse observation space makes the mapping between the adverse macro scenario and the credit risk parameters particularly prone to model uncertainty, as the co-movement during normal business cycle times has to be extrapolated to severe tail events. In most applications this issue is aggravated by short PD time series, i.e. few, mostly normal times observations, and highly correlated macro covariates on the right-hand side of the mapping. Under strong multicollinearity and especially in interaction with few observations, estimation results become both qualitatively and quantitatively very sensitive to minor changes in model specification, which makes robust model selection and thus, robust forecasting, very difficult. To provide some intuition on the degree of collinearity between macro variables, Table B.3 shows the pairwise correlation between the macro variables (in year-on-year growth rates and deltas) which we use in when defining our macro economic scenarios.[5] It shows that,

[5]Note that in a multivariate setting, pairwise correlation is not a necessary condition for the existence of collinearity. For a discussion on how multicollinearity affects OLS estimators see e.g. Gunst and Webster (1975); Gunst and Mason (1977); Belsley et al. (1980).

not surprisingly, some macro variables are indeed strongly correlated.
However, since stress test results can have both management and supervisory implications, arriving at robust estimates for expected losses in a stress testing context is of ample importance for both banks and supervisors. For example, pursuant Article 32(4d) BRRD solvent banks can receive public bailouts only to close any capital shortfall arising in a supervisory stress testing exercise. We propose a framework to address and mitigate the effect of model uncertainty on stress test results by studying the entire model space of all possible model specification when mapping the adverse macro scenario to credit risk parameters and filter it for specifications that are statistically, economically and "stress testing" plausible' before applying a BMA weighting scheme with out-of-sample weights. Stress testing plausibility is assessed by augmenting the state-of-the-art reduced-form perspective on stress testing with a structural benchmark model. To this end, the predictions of the BMA component models are benchmarked against those of the structural model, and predictions that deviate "too strongly" from the benchmark constraint are deemed to be "stress testing implausible".
While our general approach of combining a reduced-form with a structural perspective can be applied to any of the two risk parameters (PD and LGD) necessary for the expected loss calculation, this paper applies it only to the estimation of stressed PD dynamics. This is due mainly to data availability constraints, since time series for LGDs on a sector- or even bank-level are seldom available in good quality, at least to supervisors, but would be necessary for our benchmark-constrained BMA approach. In principal, any structural model could be used to inform the benchmark constrained including sophisticated partial equilibrium credit risk models (e.g. Corbae, D'Erasmo, Galaasen, Irarrazabal, and Siemsen, 2017) or full-fledged general equilibrium models (e.g. Clerc, Derviz, Mendicino, Moyen, Nikolov, Stracca, Suarez, and Vardoulakis, 2015). In our application of the BCBMA framework, we employ a standard Merton/Vasicek one-factor model, which we believe to be particularly attractive due to its traceability. In addition, it requires no additional data for calibration relative to the reduced-form model. We map the macro scenario to the structural PDs using a multivariate QMap approach.
In the absence of reliable time series, we pursue a more conventional approach for the modeling of stressed LGDs. Here, we apply a constant haircut on stressed LGDs and keep them constant during the stress horizon.

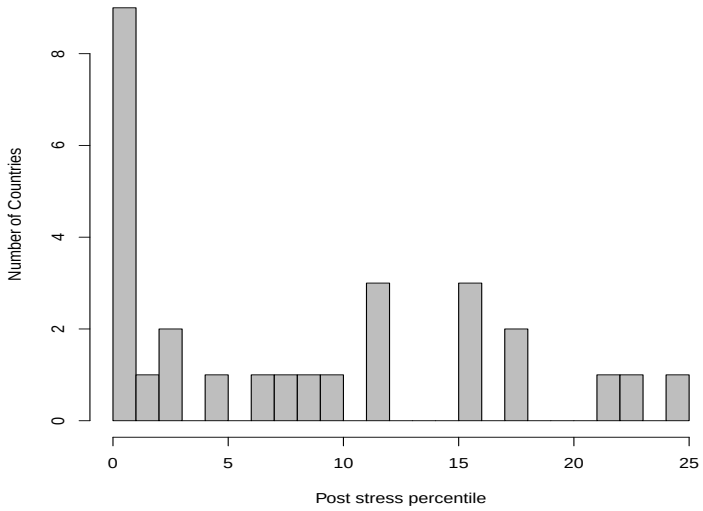


Figure 1: GDP growth post-stress percentiles according to the adverse scenario in the EU-wide stress test 2018

Before we move to a detailed description of the BCBMA framework and an application to the German banking sector, Figure 2 provides a high-level summary of our framework. Our model employs a static balance sheet assumption such that initial exposures are affected only through default flows but not through management actions. As mentioned above, given the data limitations we face, we pursue asimple approach for stressed LGD modeling. For modeling stressed PDs we employ our BCBMA framework. To this end, we first compute the unfiltered model space (MS) for the link between the historical PD time series and the historical time series of macro variables. To be precise, we choose a standard autoregressive distributed lag (ADL) model as standard in for stress testing purposes (see, for example, Henry and Kok, 2013; Gross and Poblacion, 2016) and define the model space as the estimation results of all possible specifications (number of covariates, lag lengths, growth rate definitions) of this ADL model. We then filter the model space for specifications that are statistically plausible, economically plausible and - through the benchmark constraint derived from the structural one-factor model - stress test plausible. All specifications that survive this filtering are then combined through the BMA weighting scheme using out-of-sample weights. The final BCBMA model is then used to map the adverse macro scenario to stressed PD time series for the stress horizon. Together with the constant stressed LGD and the EaD we can compute expected losses according to Equation (1) over a three-year stress horizon. Finally, stress credit risk parameters and expected losses are translated into bank-specific capital effects for the 1,500 German banks.

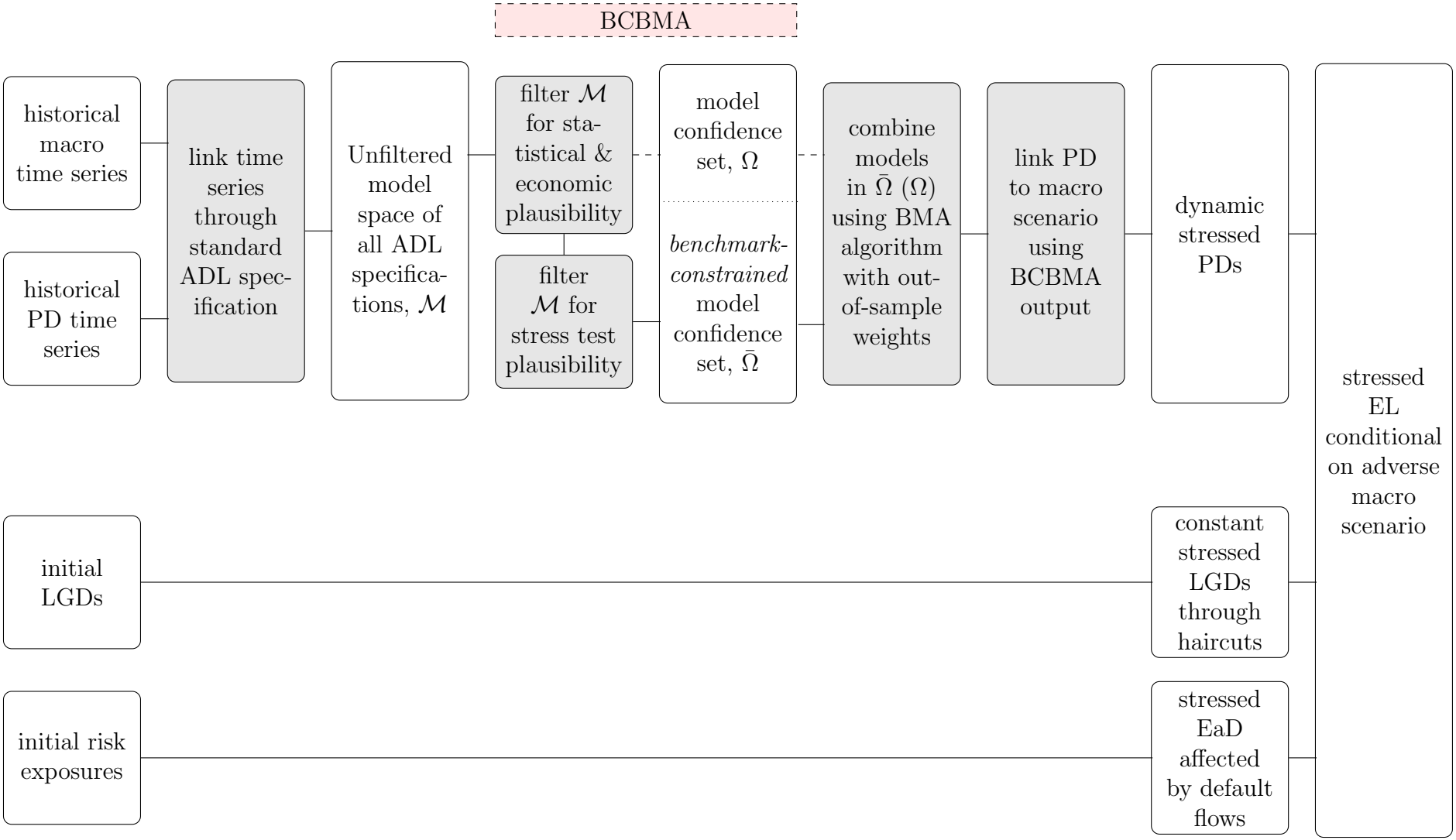


Figure 2: Overview stress testing framework

2.1 BCBMA Framework

The BCBMA model combines the standard BMA approach discussed in Henry and Kok (2013) with a structural benchmark derived from a Merton/Vasicek-type one-factor model (see, for example, Bonti et al., 2006). As discussed above, standard BMA is a common approach to dealing with model uncertainty not only in a stress testing context. For stress testing, the issue of model uncertainty is aggravated by *(1)* mostly short time series for credit risk parameters, i.e. a sparse observation space, *(2)* often highly correlated macro covariates, which make model predictions very sensitive to model specification, and *(3)* very few (if any) crisis observations, which requires to extrapolate normal times correlations outside of the observation space. The interaction of those three issues renders robust model selection and thus robust forecasting conditional on a stress scenario particularily difficult, since under collinear regressors reliable and stable estimates can only be expected within the observation space (see, for example, Belsley et al., 1980).[6]

Since stress test results can have capital and management implications for banks, robust stress testing models are key for both banks and regulators (for a similar discussion, see also Gross and Poblacion, 2016; Siemsen and Vilsmeier, 2017). Against this backdrop, BMA reduces model uncertainty by not only using one particular model specification (mostly the "best" specification according to some statistical criterion), but combines the information of the entire model space. Baele, Bruyckere, Jonghe, and Vennet (2015) show for the US banking sector that by exploiting the full model space, BMA has a superior out-of-sample forecast performance relative to OLS.

To compute the model space, we first estimate all possible specification of a standard ADL model that links the PD time series to the macro variables (Henry and Kok, 2013; Gross and Poblacion, 2016):

$$\Delta \log\left(\frac{PD}{1-PD}\right)_t = \sum_{k=1}^{K} \alpha_k \Delta \log\left(\frac{PD}{1-PD}\right)_{t-k} + \sum_{l=0}^{L} \beta_l' x_{t-l} + \varepsilon_t \,, \tag{2}$$

where Δ denotes the year-on-year change of the PD time series unbanned from the zero-one constraint through the logit transformation and x_t is a vector of macro variables; K and L denote the endogenous and exogenous lag lengths, respectively.[7]

Now, instead of considering one particular specification of Equation (2), we consider various specifiactions from a filtered subset of the model spaces and apply an averaging approach. The model space (MS) is defined as the space of all parameter estimates $\omega = [\{\alpha_k\}_K, \{\beta\}_L]$ for all possible specifications of (2), i.e. combinations of K, $L \in \mathbb{N}$, all possible combinations in the set of macro variables x and different definitions of macro variables (quarterly growth rates, annual growth rates or quarterly and annual deltas).[8]

[6] For a discussion of issues related to the extrapolation of normal times co-movements between macro and bank variables see Corbae et al. (2017).

[7] The estimation of Equation (2) is done at the sector level $s \in \mathcal{S}$. The subscript s is neglected in the following exposition without loss of information.

[8] For example, x could contain only one macro variable, are any possible combination of two, three, ..., variables, each potentially in YoY or QoQ growth rates or deltas. The choice between deltas and growth rates is made under the constraint of stationarity.

Under the standard BMA paradigm one would now apply a weighting scheme to combine all model from the unfiltered model space, $\mathcal{M}$, into a single model (see, for example, Henry and Kok, 2013). However, due to the strictly combinative algorithm of model specification, many specifications may not lead to statistically or economic plausible results. As discussed above, due to the mostly short time series for banks' credit risk parameters, which may in addition be dominated by idiosyncratic characteristics of the financial crisis, together with potentially highly correlated macro covariates, it is well possible, as we show below, that a significant mass of model in $\mathcal{M}$ lies in the statistically and economically implausible region of the model space. To avoid these models biasing stress test results, we will first filter $\mathcal{M}$ to derive a model confidence set, $\bar{\Omega}$, to which the weighting will be applied. The following paragraphs describe the different filters in detail. The first two filters for multicollinearity and autocorrelation in residuals account for statistical plausibility; the third filter, sign restrictions, accounts for economic plausibility, and the fourth filter, the benchmark constraint, accounts for stress test plausibility. Note that the menu of filters could readily be extended.

Multicollinearity filter When defining all possible combination of macro covariates in x (including lags and different frequencies), we only consider variable combinations in each x in which the pairwise correlation between any two candidate variables is below a certain threshold γ, i.e. for a given vector of macro variables $\hat{x} = [\hat{x}_1, \ldots, \hat{x}_n]$ variables $\hat{x}_i$ are filtered out for which

$$corr(\hat{x}_i, \hat{x}_j) > \gamma \,,$$

for at least one $j \in \{1, \ldots, n\}$. Let I_1 denote the set of models $m \in \mathcal{M}$ that are filtered out by the multicollinearity filter, because they include variables $\hat{x}_i$ and $\hat{x}_j$ at the same time.

Autocorrelation filter Quite straightforwardly, we filter for model specifications for which the Durbin-Watson test cannot be rejected at the confidence level d, i.e. models $m \in I_2$ with

$$p^{DW}(m) < d$$

are filtered out.

Sign restrictions We filter out models $m \in I_3$ for which the estimated long-run multiplier (LRM), θ_i, of at least one covariate $x_i \in x$ does not satisfy an exogenously imposed sign restriction, i.e. for a given sign restriction s_i on variable x_i with $i \in \{1, \ldots, n\}$, a model m is filtered out if

$$\exists i . sgn(\theta_i) \neq s_i \,.$$

Note that by imposing the restriction on θ_i instead of β_i we do not constrain the impact response of the endogenous variable, but rather the long-run response after all transitional dynamics have faded off (see Gross and Poblacion, 2016, for a similiar assumption).

When we stop the filtering process after these three filters, without imposing the benchmark constraint, we call the induced MCS Ω *unconstrained*.

Benchmark constraint To derive the benchmark constraint we employ the standard Merton/Vasicek one-factor model of credit risk (Vasicek, 2002). The model allows us to derive a structural dependency between PDs and a systemic factor z, which we interpret as reflecting general macroeconomic conditions. The adverse macro scenario is mapped to the structural PDs using a quantile mapping approach (QMap) (Bonti et al., 2006). The QMap maps the n^{th} quantile of the historical distribution of macro variables to the n^{th} quantile of the systemic factor distribution which is then translated back into a PD. We extend the framework of Bonti et al. (2006) by applying a multivariate QMap, which – similar to BMA – combines all univariate QMaps in to a single model using a weighting scheme.
The intuition for adding the benchmark constraint to the model space is that by augmenting the reduced-form BMA perspective with a structural benchmark, we mitigate potential estimation bias still present in Ω, say, due to the short time series or dominating effects of single crisis events, which cannot be accounted for solely within the ADL regime. QMap is a traceable approach to deriving a non-linear relationship between variables and requires the assumption of a co-monotonic relationship between the variables to be mapped. Using a QMap approach to map the macro scenario to the Merton/Vasicek PDs guarantees plausible stress test results since, through the co-monotonicity assumption, a more adverse scenario, i.e. worse quantiles in the corresponding distributions, always imply higher PDs. Hence, the QMap is particularly suitable within the "stress test plausibility" filter. A further advantage of using QMap to derive benchmark constraints for the ADL-derived model space $\mathcal{M}$ is that the two model classes are fully independent of each other. Thus, the Merton/Vasicek benchmarks are not affected by biased or imprecise estimation of $\mathcal{M}$. Still, we use the Merton/Vasicek QMap predictions only as an additional filter to the ADL-derived model space, instead of basing our entire prediction on it, in order to combine two perspectives on PD forecasts: the reduced-form and the structural perspective. Eventually, the Merton/Vasicek model, as any other structural model, hinges on another set of assumptions that are not necessarily weaker than those assumed for unbiased OLS estimates.
Following Vasicek (2002), we assume that the charter value of borrower j in quarter t, $V_{j,t}$, is driven by a systemic factor z_t and an idiosyncratic factor $u_{j,t}$. By assumption, both factors are standard normally distributed. We assume that the law of motion for $V_{j,t}$ is given by

$$V_{j,t} = \rho z_t + \sqrt{1-\rho^2} u_{j,t}\,,$$

where ρ captures the strength of the dependency between charter value and systemic factor, i.e. the larger ρ, the more the charter value and thus the PD will move together with macro conditions. Borrower j defaults if her charter value drops below an exogenous threshold D. Therefore, the probability of default conditional on the realization of z_t is given by

$$\begin{aligned} PD_{j,t} =& P\left(V_{j,t} < D\right) = P\left(\rho z_t + \sqrt{1-\rho^2} u_{j,t} < D\right) \\ =& P\left(u_{j,t} < \frac{D - \rho z_t}{\sqrt{1-\rho^2}}\right) = \Phi\left(\frac{D - \rho z_t}{\sqrt{1-\rho^2}}\right) = PD_t\,, \end{aligned} \tag{3}$$

where Φ denotes the standard normal cumulative distribution function. Equation (3) gives a structural dependency between the PD, on the one hand, and the systemic factor z and the

two parameters ρ and D on the other hand. Assuming we would observe z_t under stressed macroeconomic conditions, this equation would give us the corresponding PD dynamics. Therefore, in the next step we link the macro variables to the systemic factor in order to be able to compute z_t conditional on the adverse macro scenario. We do so by mapping the quantiles of the historical distributions of the macro variable $x_t = [x_1, \ldots, x_n]$, $\{F_{x_i}\}_{i=1}^n$ to the quantiles of the z_t distribution, F_z, using a QMap model averaging. Given our initial model assumptions $F_z = \Phi$.
Let $\tilde{PD}_t$ and $\tilde{z}_t$ denote the empirical counterparts for PD_t and z_t, respectively. Given an observed PD time series (on the aggregate, sector or bank level) and calibrations for D and ρ (discussed in Section 3.3), we can derive the implied time series for the systemic factor by solving (3) for $\tilde{z}_t$:

$$\tilde{z}_t = \frac{D - \Phi^{-1}(\tilde{PD}_t)\sqrt{1-\rho^2}}{\rho} \tag{4}$$

To derive $\{F_{x_i}\}_{i=1}^n$, we first standardize and transform x in order to impose a negative co-monotonic relationship between PD and all elements in x (a deterioration in macro conditions induces an increase in PD). The co-monotonicity assumption is required for the QMap in order to preserve the correct ranking of the percentiles between the two distributions (the n^{th} percentile must correspond to the n^{th} *worst* realization in both F_{x_i} and F_z, where, for example, a worse realization implies a higher percentile in the unemployment distribution but a lower percentile in the GDP growth distribution). To be precise

$$\tilde{x}_{i,t} = \mathbb{I}_i \frac{x_{i,t} - \bar{x}_i}{\sigma_{x_i}},$$

where $\bar{x}_i$ denotes the time average of $x_{i,t}$ and $\mathbb{I}_i = \pm 1$ such that a negative co-monotonicity between x_i and the PD can be expected. We estimate $\{F_{\tilde{x}_i}\}_{i=1}^n$ by fitting a univariate skewed t distribution on the time series of each element of $\tilde{x}_t$ using the maximum likelihood estimator (see Theodossiou, 1998). The skewed t distribution is very flexible, nesting the t distribution, the normal distribution and the Cauchy distribution as special cases (see Hansen, McDonald, and Newey, 2010), and is heavy-tailed, which makes it suitable for our application in the benchmark filter since higher probabilities are assigned to adverse macro conditions relative to a standard normal distribution. Crucially, when computing stressed PD predictions, we dynamically update the calibration of ρ and D after each stress period using updated time series for $\{\tilde{x}_i\}_{i=1}^n$ and $\tilde{PD}_t$. This dynamic approach amplifies the stress implied by the QMap as discussed in Section 3.3. Following Bonti et al. (2006) a univariate QMap would now map the percentiles of the of each macro time series $\tilde{x}_{i,t}$ on the corresponding percentile of F_z, i.e.

$$\hat{z}_{i,t} = F_z^{-1}\left(F_{\tilde{x}_i}(\tilde{x}_{i,t})\right),$$

and choose the macro variable $x_i^\star$ that minimizes the RMSE between $\tilde{z}_t$ and $\hat{z}_t$. We pursue a slightly different route by not considering only one macro variable in the QMap but rather combining all univariate QMaps into a single model using a flexible weighting scheme. Thus, the QMap model averaging is similar in spirit to BMA, as no information from the model space is thrown away. Thus, we map the macro variables x_t to the systemic factor z through

$$\hat{z}_t = F_z^{-1}\left(\sum_{i=1}^{n} \pi_i F_{\tilde{x}_i}(\tilde{x}_{i,t})\right) , \tag{5}$$

where the weights π_i of are computed as $\pi_i = \left(1 - \frac{RMSE_i}{\max_i(RMSE_i)}\right)^k$ and $RMSE_i$ is the in-sample root mean squared error of the univariate QMap using x_i and k is chosen such that the RMSE of the combined model is minimized. By plugging the time series of stressed macro variables into Equation (5) we derive the Merton/Vasicek-implied stressed PD time series using Equation (3).

Finally, we derive the benchmark constraint by defining a region in Ω where the scenario-implied PD increases over the stress horizon must lie for them to be "stress test plausible". Let $t \in \{-1, 0, \ldots, T\}$ denote the stress horizon such that $\hat{PD}_{-1}$ denotes the Merton/Vasicek-implied PD one year prior to the stress horizon, $\hat{PD}_0$ denotes the initial Merton/Vasicek-implied PD pre stress and $\hat{PD}_T$ denotes the Merton/Vasicek-implied PD at the end of the stress horizon. Then, we define the benchmark constraint as

$$c = \frac{\Phi\left(\Phi^{-1}\left(\mu\left(\{\hat{PD}\}_{t=1}^{T}\right)\right) \pm \upsilon\right)}{\mu\left(\{\hat{PD}\}_{t=-1}^{0}\right)} , \tag{6}$$

where $\mu(\cdot)$ denotes the mean operator and

$$\upsilon = \bar{\upsilon} \times \left(\Phi^{-1}\left(PD_0(1 + \sigma_{PD})\right) - \Phi^{-1}\left(PD_0\right)\right) , \tag{7}$$

$\bar{\upsilon}$ being a parameter, PD_0 being the last PD observation pre stress and σ_{PD} being the historical PD volatility. υ defines the width of the region around the Merton/Vasicek-implied PD increases in which a model prediction has to lie in order to be "stress test plausible". υ is proportional to observed volatility of the historical (sector-specific) PD time series. We apply a distance-to-default transformation on the PDs in order to first map the PDs to the $]-\infty, +\infty[$ space to ensure that the stressed PDs always lie in the unite interval. The distance-to-default transformations also makes the benchmark constraint sensitive to starting values, since smaller PDs will experience a relatively stronger increases than higher PDs. While this transformation can be derived from the Merton model, in our reduced-form application it simply boils down to applying the standard normal quantile function Φ^{-1} to the PDs. Models $m \in I_4$ that imply a mean PD increase over the stress horizon under the adverse scenario that lies outside the benchmark constraint, i.e. a model for which

$$\frac{\mu\left(\{\hat{PD}(m)\}_{t=1}^{T}\right)}{\hat{PD}_0(m)} \notin c ,$$

are filtered out.

All filters are applied simultaneously, i.e. models $m \notin I_1 \cap I_2 \cap I_3 \cap I_4$ survive the filtering process and define the MCS Ω (without the benchmark constraint, i.e. neglecting I_4) or $\bar{\Omega}$ (with benchmark constraint) which is then combined into a single model using BMA. BMA

uses the posterior model probabilities of model m_i in $\Omega/\bar{\Omega}$ given data $\mathcal{D}$, $p_i = P(m_i|\mathcal{D})$, as model weights. Thus, to apply BMA we need estimations of p_i for each model m_i in the MCS.

The model space $\mathcal{M}$ and thus the corresponding MCS Ω and $\bar{\Omega}$ can be extremely large.[9] To avoid having to estimate the entire model space Madian and Raferty (1994) propose Occam's windows which subsets the model space $\mathcal{M}$ to those models that predict the data far better than models outside the subset and thus focuses only on the relevant region in the model space. Models outside the subset are neglected, i.e. the model space is trimmed. To identify this subset we apply the "Leaps-and-Bounds" algorithm (Furnival and Wilson, 1974). This algorithm efficiently identifies the Q best models in the models space for each model size using the adjusted R^2 as benchmark. If Q is chosen sufficiently high, the subset includes all models in Occam's windows plus many more models, which are then trimmed off. In other words: we approximate the entire model space $\mathcal{M}$ by the Q best models according to adjusted R^2, zoom in on the relevant region of the model space using Occam's windows and then apply the filters to arrive at the MCS Ω and $\bar{\Omega}$. Samuels and Sekkel (2012) show that employing BMA on a trimmed model space reduces the prediction error relative to BMA on untrimmed model spaces. To be precise, the posterior model probabilities are given by

$$p_i = P(m_i|\mathcal{D}) = \frac{P(\mathcal{D}|m_i)P(m_i)}{\sum_{i\in Q} P(\mathcal{D}|m_i)P(m_i)}\,, \tag{8}$$

where $P(m_i)$ denotes the prior model probability which we assume to be uniform. As a consequence $P(m_i) = P(m)\,, \forall i$ and it cancels out in Equation (8). $P(\mathcal{D}|m_i)$ is given by the marginal likelihood $P(\mathcal{D}|m_i) = \int P(\mathcal{D}|\omega_i, m_i)P(\omega_i|m_i)d\omega_i$, where $\omega = [\{\alpha_k\}_K, \{\beta_l\}_L]$. However, instead of deriving the posterior model probabilities and the implied BIC model weights based on this marginal likelihood, we follow Burnham and Anderson (2002) and instead use model weights based on a smoothed AIC, which induces model weights δ_i that are proportional to the probability of being the best Kullback and Leibler (1951) model in repeated samples (see also Pelster and Vilsmeier, 2017, for a discussion). Then,

$$p_i = \delta_i = \frac{\exp(-0.5\Delta_i)}{\sum_{i\in Q}\exp(-0.5\Delta_i)}\,, \tag{9}$$

with $\Delta_i \equiv AIC_i - \min_i\{AIC_i\}$. Assuming that $\epsilon_{i,t} \sim N(\mu_\epsilon, \sigma_{i,\epsilon})$, $AIC_i = O\log(\sigma^2_{i,\epsilon}) + 2N$, where O denotes the number of observations. We calculate σ^2_ϵ on the basis of the out-of-sample estimates of the sum of squared residuals. To this end, we apply the leave-one-out method, where repeatedly one observation from the sample is dropped and the model is estimated using the remaining data. To be precise, for each specification $q \in Q$, we drop observation o with $o \in 1, \ldots, O$ and predict the logit PD using the remaining $O-1$ observations. Let $\hat{y}_o$ denote the predicted logit PD using the parameters estimated on the $O-1$ observations and let y_o denote the realized logit PD in period o. Then $\sigma^2_{i,\epsilon} =$

[9] Computing the entire model space may require the estimation of an enormous number of models, depending on the number of macro covariates in x and lag length K and L. For example, for ten macro covariates, $K = L = 4$ and considering both annual and quarterly changes, the number of models in the unfiltered model space $\mathcal{M}$ is larger than 10^{14}.

$1/O \sum_{o=1}^{O} [\hat{y}_{i,o} - y_{i,o}]^2$. To identify the relevant region of the approximated model space through Occam's window, we compute δ_i for all Q models identified by the "Leaps-and-Bounds" algorithm according to the highest adjusted R^2. Then, we trim off all models i with

$$max_Q \left\{ \{\delta_q\}_{q \in Q} \right\} / \delta_i > o \,,$$

with o being the threshold parameter. This implies that all models which have a probability of being the best Kullback-Leibler model in the model space o times smaller than the model with the highest probability are trimmed off. The remaining subset of the model space defines the sample from which the MCSs are derived.
We combine all models in the MCS $\Omega/\bar{\Omega}$ into a single model applying the weights given by Equation (9). To this end, we re-compute the weights based only on models in the MCS

$$\delta_i = \frac{\exp(-0.5\Delta_i)}{\sum_{i \in \Omega/\bar{\Omega}} \exp(-0.5\Delta_i)} \,, \tag{10}$$

and use these weights to combine the single model parameter estimates

$$\omega^{BMA} = \sum_{i \in \Omega/\bar{\Omega}} \delta_i \omega_i \,, \text{with } \omega_i = [\{\alpha_{k,i}\}_K, \{\beta_{l,i}\}_L] \tag{11}$$

Given the final BMA parameter estimates, the LRM for covariate x_i can be computed as

$$\tilde{\theta}_i \equiv \sum_{s=0}^{+\infty} \frac{\partial \mathbb{E} Y_{t+s}}{\partial x_{i,t}} = \frac{\sum_{l=1}^{L} \beta_{i,l}}{1 - \sum_{k=1}^{K} \alpha_k} \tag{12}$$

To make LRM comparable across covariates, we normalize $\tilde{\theta}_i$ via

$$\theta_i = \frac{\sigma_{x_i}}{\sigma_y} \times \tilde{\theta}_i \,,$$

which allows the interpretation of θ_i having the effect of a permanent increase of x_i by one standard deviation on y (in terms of standard deviations of y). The posterior inclusion probability on x_i in the BMA model is given by

$$poip_i = \sum_{i \in \Omega/\bar{\Omega}} \mathbb{I}_{i \neq 0} \delta_i \,. \tag{13}$$

The posterior inclusion probability summarizes the aggregate evidence of x_i in the MCS by summing the posterior model probability of all models in the MCS in which the covariate coefficient takes a non-zero value. In Bayesian model inference, a variable is said to be significant if the posterior inclusion probability is higher than the prior inclusion probability. Under our assumption of uniform prior model probability, the prior inclusion probability is given by the ratio of average model size over the total number of potential regressors. Below, we let $\bar{N}$ denote the maximum number of right-hand-side variables we allow in Equation (2), including endogenous lags, and N the number of all possible right-hand-side variables, such

that $\bar{N} \leq N$ (equality holds of the maximum number of right-hand-side variables is left unconstrained). Then the prior inclusion probability is given by

$$prip_i = \frac{\frac{\sum_{i=1}^{\bar{N}} i \frac{N!}{i!(N-i)!}}{\sum_{i=1}^{\bar{N}} \frac{N!}{i!(N-i)!}}}{N} . \tag{14}$$

This completes the description of the BCBMA model which will be employed to map the time series of PDs to the adverse macro scenario in the credit risk stress test.

2.2 LGD Modeling

In contrast to the BCBMA-PD model which addresses model uncertainty, our LGD model follows a simpler approach. We assume that stressed LGDs are constant during the stress horizon, such that the stress is imposed only on initial LGDs. This approach reflects the non-availability of robust and long LGD time series for supervisors, which would be required to map the macro scenario to LGDs in a reduced-form manner. For a given uncollateralized and unstressed parametric $L\bar{G}D$ we compute the scenario-implied LGD for loans in sector $s \in \mathcal{S}$ of bank $b \in \mathcal{B}$ as

$$LGD_{b,s} = \max \left\{ 0.10, L\bar{G}D \times \frac{E_{b,s} - C_{b,s}(1+h_s)}{E_{b,s}} \right\} , \tag{15}$$

where E denotes the corresponding exposure, C the corresponding collateral and h a haircut on the collateral value. The haircut is defined by

$$h_s \equiv \begin{cases} \frac{H_s^{-1}(0.2)}{H_s^{-1}(0.5)} - 1 \, , \text{adverse scenario} \\ 0 \qquad\qquad\quad , \text{baseline scenario} \end{cases} ,$$

where H_s^{-1} denotes the quantile function of the collateral value distribution in sector s. Equation (15) transforms the uncollateralized LGD into a collateralized LGD: the homogeneous $L\bar{G}D$ is scaled down by the share of bank-specific collateralized exposure, such that all initial LGDs lie in the [0.10, $L\bar{G}D$] interval. The haircut in the adverse scenario is modeled as a reduction of the collateral value from the median value to the value corresponding to the 20^{th} percentile of the collateral value distribution. [10]

2.3 Stress Impact

At this stage we have derived stressed values for the two credit risk parameters PD and LGD. What is now left to describe is how we link these parameters to banks' capital position at the end of the stress horizon. Generally, we model all stress effects bank-specifically wherever

[10] See Siemsen and Vilsmeier (2017) for a version of the model with dynamics LGDs during the stress horizon. In contrast to the residential mortgage sector, structural dependencies between the collateral value and the macro scenario do not exist in most other credit sectors such as retail business. Therefore, we cannot pursue the same approach as for the residential mortgage stress test here.

possible (see Section 3.1 for a more detailed discussion), i.e. stress effects are heterogeneous at the bank level.
The stressed credit risk parameters affect banks' capital positions through two channels: *first*, they determine the default flow and thus the expected losses over the stress horizon (see Equation (1)). *Second*, they increase unexpected losses through an increase in the variance of the loss distribution under stress. According to regulatory standards (see, for example, Basel Committee on Banking Supervision, 2005) banks are required to hold capital to safeguard the institution against losses in excess of expected losses. *Ceteris paribus*, an increase in unexpected losses reduces banks CET1 ratio as regulatory requirements tighten. Neglecting or underestimating the effect of unexpected losses on banks' capital position may lead to downward-biased stress effects and may thus overestimate the banks' resilience to adverse macroeconomic conditions. In contrast to risk weights computed under the standardized approach (SA) for credit risk, risk weights derived from the IRB formula are theoretically founded and measure *unexpected* losses, defined as the distance between mean and value-at-risk at the 99.9 percentile of the loss distribution. By linking the calculation of risk weights to the dynamics of credit risk parameters, IRB risk weights become sensitive to macroeconomic conditions. The SA, in contrast, has often been criticized of being too risk insensitive as risk weights respond only sluggishly to changes in the risk environment.[11] To get a full picture of expected and unexpected losses in our stress test, we model RWA according to the IRB formula for both IRB *and* for SA loans. Thus, we pursue an economic view instead of a purely regulatory view of stress test losses, since the RWA dynamics we derive for SA exposures would not be observed if the macro scenarios were to materialize. The "true" SA-RWA dynamics would be more sluggish and would likely underestimate unexpected losses (see Siemsen and Vilsmeier, 2017, for a more detailed discussion and a quantification of the effect between IRB and SA risk weights on stress test outcomes).
Below, let $b \in \mathcal{B}$ denote the bank, $s \in \mathcal{S}$ the sector of the relevant exposure, $u \in \mathcal{U}$ one loan, and $t \in T$ a year during the stress horizon. The computation of capital effects is performed in two steps: *first*, the default, impairment and risk weight *factors* are computed and *second*, these factors are scaled to euro amount through multiplication by the relevant exposures. This dichotomy is due to our data sources and will be explained in detail in Section 3.1.

Expected losses Extending Equation (1) we can write the expected loss at the bank-sector-level as

$$\begin{aligned} LGD_{b,s,t} \times PD_{b,s,t} \times EaD_{b,s,t} &= \left[\frac{\sum_{u=1}^{\mathcal{U}} \left(PD_{u,b,s,t} \times LGD_{u,b,s,t} \times EaD_{u,b,s,t} \right)}{EaD_{b,s,t}} \right] \times EaD_{b,s,t} \\ &= Imp_{b,s,t} \times EaD_{b,s,t} \,, \end{aligned} \tag{16}$$

[11] In response Basel Committee on Banking Supervision (2015) suggests more risk sensitive risk weights.

where Imp denotes the impairment factor. Similarly, we can express the default flow as

$$PD_{b,s,t} \times EaD_{b,s,t} = \left[\frac{\sum_{u=1}^{\mathcal{U}} (PD_{u,b,s,t} \times EaD_{u,b,s,t})}{EaD_{b,s,t}} \right] \times EaD_{b,s,t} = Def_{b,s,t} \times EaD_{b,s,t} \,, \tag{17}$$

where Def denotes the default flow factor. Given these definitions of the impairment and default flow factors, the expected loss dynamics over the stress horizon can be written as a simple system of recursive equations:[12]

$$ImpFlow_{b,s,t} = Imp_{b,s,t} \times EaD_{b,s,t} \,, \forall t \in T \tag{18}$$

$$DefFlow_{b,s,t} = Def_{b,s,t} \times EaD_{b,s,t} \,, \forall t \in T \tag{19}$$

$$EaD_{b,s,t+1} = EaD_{b,s,t} - DefFlow_{b,s,t} \,, \forall t \in T \text{ and } EaD_{b,1,s} \text{ given} \tag{20}$$

Unexpected losses To approximate unexpected losses occurring during the stress horizon we use the IRB formula originally intended to compute risk weight for IRB loans. As discussed above, the formula is derived from a standard Merton credit risk model to approximate unexpected losses. It is then calibrated to approximate unexpected losses as the difference between the expected value of the loss distribution (expected losses) and the value at risk at the 99.9 percentile (unexpected loss, see Basel Committee on Banking Supervision, 2005). Thus, independently of the regulatory view of RWA, the IRB formula can be used to measure unexpected losses from an economic perspective.
The IRB risk weight formula, $\mathcal{R}(\cdot)$, computes risk weights as a function of through-the-cycle adjusted (TTC) PDs, downturn LGDs and exposure (to scale risk weights up to RWA)

$$\mathcal{R}(PD^{ttc}_{b,s,t},\, LGD^{dt}_{bs,t},\, EaD_{b,s,t}) \,. \tag{21}$$

Thus, before we can derive risk weights assets, we need to obtain the transformations of the original credit risk parameters derived in Sections 2.1 and 2.2. Here, we take a pragmatic approach constrained by data availability. Since we do not have a time series for LGDs and since stressed LGDs remain constant throughout the stress horizon, we use these constant LGDs directly in Equation (21) without applying a adjustment to condition LGDs on an economic downturn.[13] For TTC PDs we compute a-year rolling windows of the PD time series ranging from the first historical observation to the last quarter of the stress horizon. The a-year rolling windows PDs are used in Equation (21). For the scaling of risk weights to RWA we assume that the relevant exposure stays constant over the stress horizon, i.e. in Equation (21) $EaD_{b,s,t} = EaD_{b,s}$. This assumption reflects a conservative floor on RWA during the stress horizon, as RWA are floored below at their initial value and cannot decrease below this value. This floor is also imposed during the EU-wide stress tests (European Banking

[12]Note that in contrast to the EBA credit risk methodology, we cannot consider impairments to the default stock since these are driven by changes in the LGD over the stress horizon. In our model, LGDs are currently constant.

[13]In any top-down stress test application the computation of downturn LGDs based on regulatory data only is cumbersome and always subject to ad hoc assumptions. See, for example, Rösch and Scheule (2009) for an ad hoc model for downturn LGDs.

Authority, 2017) and prevents fully written down default flows from reducing RWA during the stress horizon. Thus, the RWA factor is computed as

$$RWAf_{b,s,t} = \frac{\sum_{u=1}^{\mathcal{U}} \mathcal{R}\left(, PD^{ttc}_{u,b,s,t},\, LGD_{u,b,s},\, EaD_{u,b,s}\right)}{EaD_{b,s}}\,. \tag{22}$$

Then, RWA are computed as

$$RWAflow_{b,s,t} = RWAf_{b,s,t} \times EaD_{b,s,0}\,, \tag{23}$$

where $EaD_{b,s,0}$ denotes the initial bank- and sector-specific risk exposure. This reflects the assumption that exposures, for the purpose of calculating RWA, stay constant to enforce a floor on RWA.

Capital Impact At this stage, we have a set of variable paths over the stress horizon, $\{Impflow_{b,s,t},\, Defflow_{b,s,t},\, RWAflow_{b,s,t}\}_{t=1}^{T}\,, \forall s \in \mathcal{S}\,, \forall b \in \mathcal{B}$, derived above. Now, we map these paths to banks' capital position. To this end, we first aggregate all flows over all sectors $s \in \mathcal{S}$ by summing up., i.e. we now have $\{Impflow_{b,t},\, Defflow_{b,t},\, RWA_{b,t}\}_{t=1}^{T}\,, \forall b \in \mathcal{B}$. Let the Δ operator denote the difference of variable x between year t and the initial value at date 0, such that $\Delta(x_t) \equiv x_t - x_0$.
In our framework, the stress tests affects interest income (II), impairments (Imp) and profits/losses (π). For interest income, we approximate the average return on one unit of interest bearing assets as

$$r_{b,0} = \frac{II_{b,0}}{A_{b,0}}\,,$$

where A denotes a bank's interest-bearing assets and $t = 0$ corresponds to the initial pre-stress year. We define

$$\Delta^{II}_{b,t} \equiv \begin{cases} r_{0,b} \times \Delta(Defflow_{b,t}), & \text{adverse scenario} \\ 0 & \text{, baseline scenario} \end{cases}$$

as the reduction in interest income due to the default flow relative to the starting period. Note that in the baseline scenario, the reduction in interest income is floored at zero. Then, the interest income level during the stress horizon as given as

$$II_{b,t} = r_{b,0} \times A_{b,0} - \sum_{k=1}^{t} \Delta^{II}_{b,t}\,. \tag{24}$$

The impairment level during the stress horizon is straightforward to derive as

$$Imp_{b,t} = Imp_{i,0} + \Delta(Impflow_{b,t})\,. \tag{25}$$

To map stressed interest income and impairments to profits, we subtract the initial observations of these variables from profits and then add the stressed variables back to it. Thus

$$\pi_{b,t} = \tilde{\pi}_{b,0} - Imp_{b,t} + II_{b,t}\,, \tag{26}$$

with $\tilde{\pi}_{b,0} = \pi_{b,0} + Imp_{b,0} - II_{b,0}$. Therefore, all other P&L components which affect profits, except for interest income and impairments (e.g. interest expenses), are assumed to be constant during the stress horizon. Finally, the level of RWA is computed as

$$RWA_{b,t} = RWA_{b,0} + \Delta(RWAflow_{b,t})\,. \tag{27}$$

Now, we can compute the stress capital level and the stressed capital ratio:

$$C_{b,t} = C_{b,0} + \sum_{k=1}^{t} \pi_{b,t} \tag{28}$$

$$cr_{b,t} = \frac{C_{b,t}}{RWA_{b,t}}\,. \tag{29}$$

3 Data and Calibration

This section provides details of the data sources (and corresponding data limitations) of our top-down stress test application (Section 3.1). Section 3.2 shows the baseline and adverse macro scenario that we use to compute stress effects. Section 3.3 discusses how we calibrate the general framework to our concrete application to the German banking system.

3.1 Data Sources for Stress Test

The data we use to conduct the top-down credit risk stress test relies on two main data sources: the German Credit Register (GCR) and the German Borrower Statistics (GBS). Intuitively, the GCR is used to derive the impacts *factors* discussed in Section 2.3 and the *exposures* in the GBS are used to scale these GCR factors to euro amounts.

German Credit Register The GCR features quarterly loan observations of German banks to borrowers (including foreign branches and off-balance-sheet exposures) starting in 2008Q1. The data are reported at the single borrower level and includes information (among others) on collateral and risk-weighted assets; for IRB loans also the regulatory PD is reported. Moreover, loans can be assigned to sectors according to the NACE code. However, for the GCR only loans with a total volume exceeding €1 million are reported. This constraint clearly reduces the representativeness of the GCR data as *(1)* large parts of the retail exposure is likely to be excluded from the GCR and *(2)* banks that do not provide loans exceeding a total volume of €1 million will not be observed in the GCR.
Below, we will denote banks that are observed in the GCR in the most recent quarter employed for the stress test as *GCR banks*, $g \in \mathcal{G}$, and we will denote banks that are not observed in the GCR in the most recent quarter as *non-GCR banks* $n \in \mathcal{N}$; thus $\mathcal{G} + \mathcal{N} = \mathcal{B}$. Since the GCR does not capture the complete universe of loan exposures, we do not use the GCR exposures for our stress test purposes but only the GCR PD and collateral values. Crucially, for non-GCR banks $\mathcal{N}$ for which no bank-specific impact factors can be derived

from the GCR, *sector-specific* GCR impact factors will be derived. However, we do an ad hoc adjustment to the impairment factor $\{Imp_{n,s,t}\}_{n\in\mathcal{N}}$ in Equation (16) to make it bank-specific for all banks as discussed below. In particular, we use the GCR to derive

- sector-specific time series for PD, $PD_{s,t}$, based on all IRB loans of GCR banks $\mathcal{G}$. These time series are used for the BCBMA algorithm discussed in Section 2.1; in particular they are used in Equations (2) and (4). *Sector*-specific PDs are derived from volume-weighted single IRB loan PDs in each quarter.
- the bank- and sector-specific stressed LGDs in Equation (15). To this end, the IRB loan exposures E and the corresponding collateral values (C) are used. H^{-1} is the empirical quantile function derived from all IRB loan collaterals in a given sector s. For GCR banks LGDs are computed bank-specifically, for non-GCR banks sector-specific LGDs are computed by summing E and C within each sector.
- the calibration of the IRB risk weight in Equation (21) using the reported RWA (see also Section 3.3).

Thus, all impact factors $Imp_{b,s,t}$ (Equation (16)), $Def_{b,s,t}$ (Equation (17)) and $RWAf_{b,s,t}$ (Equation (22)) are based on GCR data. The sector-specific stressed PD time series $PD_{s,t}$ are mapped to bank-specific (GCR banks) and sector-specific (non-GCR banks) starting values using a distance-to-default transformation, such that for bank $g \in \mathcal{G}$

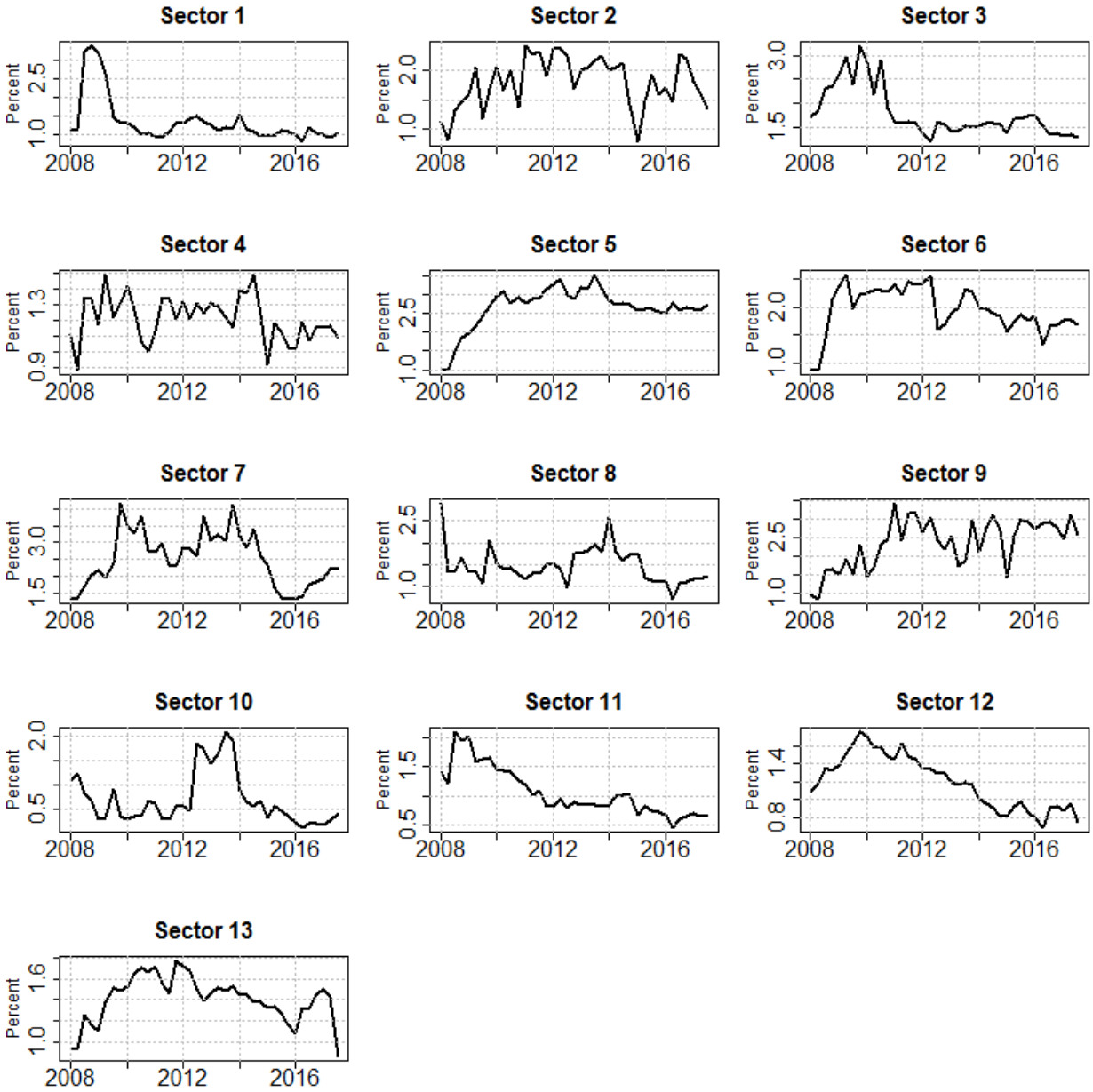
$$PD_{g,s,t} = \Phi\left(\Phi^{-1}\left(PD_{g,s,0}\right) + \left[\Phi^{-1}\left(PD_{s,t}\right) - \Phi^{-1}\left(PD_{s,0}\right)\right]\right)$$

and for $n \in \mathcal{N}$

$$PD_{n,s,t} = \Phi\left(\Phi^{-1}\left(PD_{s,0}\right) + \left[\Phi^{-1}\left(PD_{s,t}\right) - \Phi^{-1}\left(PD_{s,0}\right)\right]\right) = PD_{s,t}\,,$$

where $t = 0$ denotes the last PD observation in the GCR before the stress horizon begins. Generally, there are 23 different NACE sectors in the GCR data.[14] However, we do not estimate the BCBMA model on all of the 23 PD series, since for some sectors erratic time series behavior, induced, for example, by a few large sectoral exposures, does not allow for a robust modeling of these sectors. Thus, we combine these sectors with other sectors to mitigate the erratic behavior. The choice which sectors to combine is made based on pairwise correlation of the sector PD time series. In particular we identify candidate partners as sectors with a pairwise correlation higher than 0.6. We merge sectors with their candidate partners if the median adjusted R^2 in the unconstrained model space is below 0.3. All combined sectors have the same sectoral PD dynamics. Appendix A shows which sectors are combined. Figure 3 shows the historical PD time series for the 13 uniquely modeled sectors.

[14] See Appendix A for a description of the sectors including exposure distributions.



Notes: See Table A.1 for sector descriptions.

Figure 3: Sectoral historical PD time series

German Borrower Statistics The GBS features quarterly observations of all on-balance-sheet loans of German banks to domestic borrowers. In contrast to the GCR, the GBS thus includes *all* loans at the bank level, independently of the volume, but no foreign and off-balance sheet exposures. Also, business sectors are reported, which can be mapped to the GCR's NACE sectors.

We use the GBS exposures to scale the GCR different impact factors to euro amounts and to make sector-specific impairments factors for non-GCR banks bank-specific. We use GBS

exposures, because the exposures reported in the GBS provide a more complete picture of total credit exposures of all German banks than the GCR exposures do, which are floored at a minimum volume of €1 million. The GBS cannot, however, be used to derive thorough impact factors, as PD and collateral are not reported. As mentioned above, the GBS does not include foreign and off-balance-sheet exposures of banks. Since the GCR does include those exposures, we extract them from the GCR data and add them to the GBS exposures for all GCR banks. This provides a better approximation of the total risk position of banks participating in the stress test.

In addition, the GBS includes gross impairments of loans at the bank level. In contrast to the impairment factor derived from the GCR, which is only sector-specific but not bank-specific for the non-GCR banks, the GBS allows us to derive approximations of bank-specific impairments for those banks, too. We use the GBS impairments to compute bank-specific adjustment factors on the sectoral GCR impairment factors for the non-GCR banks; for the GCR banks no adjustment is made. In particular, $\forall g \in \mathcal{G}$

$$Imp_{g,s,t} = \Phi\left(\Phi^{-1}\left(Imp_{s,t}^{GCR}\right) + \left[\Phi^{-1}\left(Imp_{g,s}^{GBS}\right) - \Phi^{-1}\left(Imp_{s}^{GBS,\mathcal{G}}\right)\right]\right) , \tag{30}$$

where $Imp_{s,t}^{GCR}$ denotes the sector-specific impairment factor from the GCR, $Imp_{g,s}^{GBS}$ denotes the non-GCR bank-specific impairment factor derived from the GBS and $Imp_{s}^{GBS,\mathcal{G}}$ denotes the three-year average impairment factor of all GBC banks $\mathcal{G}$ from the GBS. The impairment factors from the GBS are computed as

$$Imp_{g,s}^{GBS} = \frac{\min\left\{0, \text{gross impairments}_{g,s}\right\}}{\text{total exposure}_{g,s}} .$$

The intuition behind the adjustment is the following: both $Imp_{s,t}^{GCR}$ and $Imp_{s,t}^{GBS,\mathcal{G}}$ are based on the same sample of GCR banks, where the latter is a three-year average impairment factor and the former is the contemporary impairment factor. Thus, all non-GCR banks get an adjustment to the sector-specific GCR impairment factor given by how strong their contemporary GBS-implied impairment factor deviates from the average impairment factor of GCR banks in the GBS. To make the adjustment starting value specific, we apply, with a slight abuse of notation, a distance-to-default transformation to these impairment factors. Note that the adjustment factor is time-invariant and is derived from the latest available quarterly observation (except for $Imp_{s}^{GBS,\mathcal{G}}$ which is computed as the average factor over three years). We have to cap the gross impairments since write-ons on previous write-downs can induce positive gross impairments.

Table 1 summarizes the data sources and levels of granularity in modeling when applying the BCBMA framework to the German banking sector, and Table 2 shows briefly some descriptive statistics.

Table 1: Summary of Data Sources for GCR and non-GCR Banks

	GCR Banks	**non-GCR Banks**
Impairment Factors	at bank and sector level from GCR	at sector level from GCR but made bank-specific through adjustment by GBS-approximated bank specific impairment factors
Default Flow Factors	at bank and sector level from GCR	at sector level from GCR
Risk Weight Factors	at bank and sector level from GCR	at sector level from GCR
Exposures	at bank and sector level from GBS with addon from foreign and off-balance-sheet exposures from GCR	at bank and sector level from GBS

Table 2: Some Descriptive Statistics

	GCR	**GBS**
number of loans	102,372	-
total exposure (€ billion)	592.7	2,511.7
number of banks	38	1,672

Notes: All data as of 2017Q2

Regulatory Data In addition to the two data sources discussed above, we require regulatory data at the bank level to derive the final stress impact on banks' capital position. As discussed in Section 2.3, we model stressed P&L dynamics for interest income, impairments and profits. The main metric for assessing the severity of the stress impact will be the effect on the common equity tier 1 ratio (CET1 ratio). The CET1 ratio is defined as the ratio of CET1 capital to RWA. Thus, we also require regulatory data on CET1 capital and aggregate RWA (at the bank level and not at the borrower level as provided in the GCR). The starting values for the P&L items as well as for interest-bearing assets are taken from the Bundesbank's quarterly balance sheet statistics, and the regulatory variables are taken from the Corep reporting. Appendix B provides additional details of the data.

3.2 Stress Scenario

The scenarios we employ to study the effect of model uncertainty on top-down credit risk stress testing closely follow the EBA 2018 scenario.[15] We consider a baseline and an adverse macroeconomic scenario. The scenarios include the following set of ten macro variables at a quarterly frequency: German real GDP growth ($GER.GDP$), change in German ten-year bond yields ($GER.bond$), growth in German consumer price index ($GER.CPI$), growth in German commercial house price index ($GER.CHPI$), growth in German residential house price index ($GER.RHPI$), German unemployment rate ($GER.unemp$), change in the German equity index DAX ($GER.DAX$), change in the one-year EURIBOR rate ($EU.EURIBOR$), growth in EU real GDP ($EU.GDP$) and growth in US real GDP ($US.GDP$). Table B.2 gives an overview of the data sources.

We impose a three-year stress horizon with initial values as of 2017Q2, such that the 12 stress quarters go until 2020Q2. Traditionally, the scenario of the EU-wide stress test is defined annually, such that we employ the annual value values and interpolate them to a quarterly frequency by keeping the quarterly growth rates constant within each year. Table 3 shows the annual path of each variable for the baseline scenario and the adverse scenario, respectively. Figure 4 plots the scenario paths together with one year of the corresponding historical time series.

Table 3: Macro scenarios

	Baseline			Adverse		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
$GER.GDP$	1.5	1.6	1.4	-1.9	-2.8	1.4
$GER.bond^{\dagger}$	0.5	0.7	0.9	1.2	1.4	1.5
$GER.CPI$	1.6	1.7	1.9	1.5	0.9	0.3
$GER.CHPI$	4.0	3.8	3.7	-10.6	-7.1	-2.6
$GER.RHPI$	4.8	4	3.8	-8.8	-9.5	0.2
$GER.unemp^{\dagger}$	3.3	3.1	2.9	4.2	5.5	6.1
$GER.DAX$	0.0	0.0	0.0	-30.6	3.8	8.0
$EU.EURIBOR^{\dagger}$	-0.17	0.02	0.3	0.39	0.5	0.7
$EU.GDP$	2.2	1.9	1.8	-1.2	-2.2	0.7
$US.GDP$	2.3	1.9	1.8	-0.3	-0.6	3.1

Notes: † denotes variables in levels instead of annual growth rates.

[15] See European Banking Authority (2018) for a discussion of the corresponding narrative.

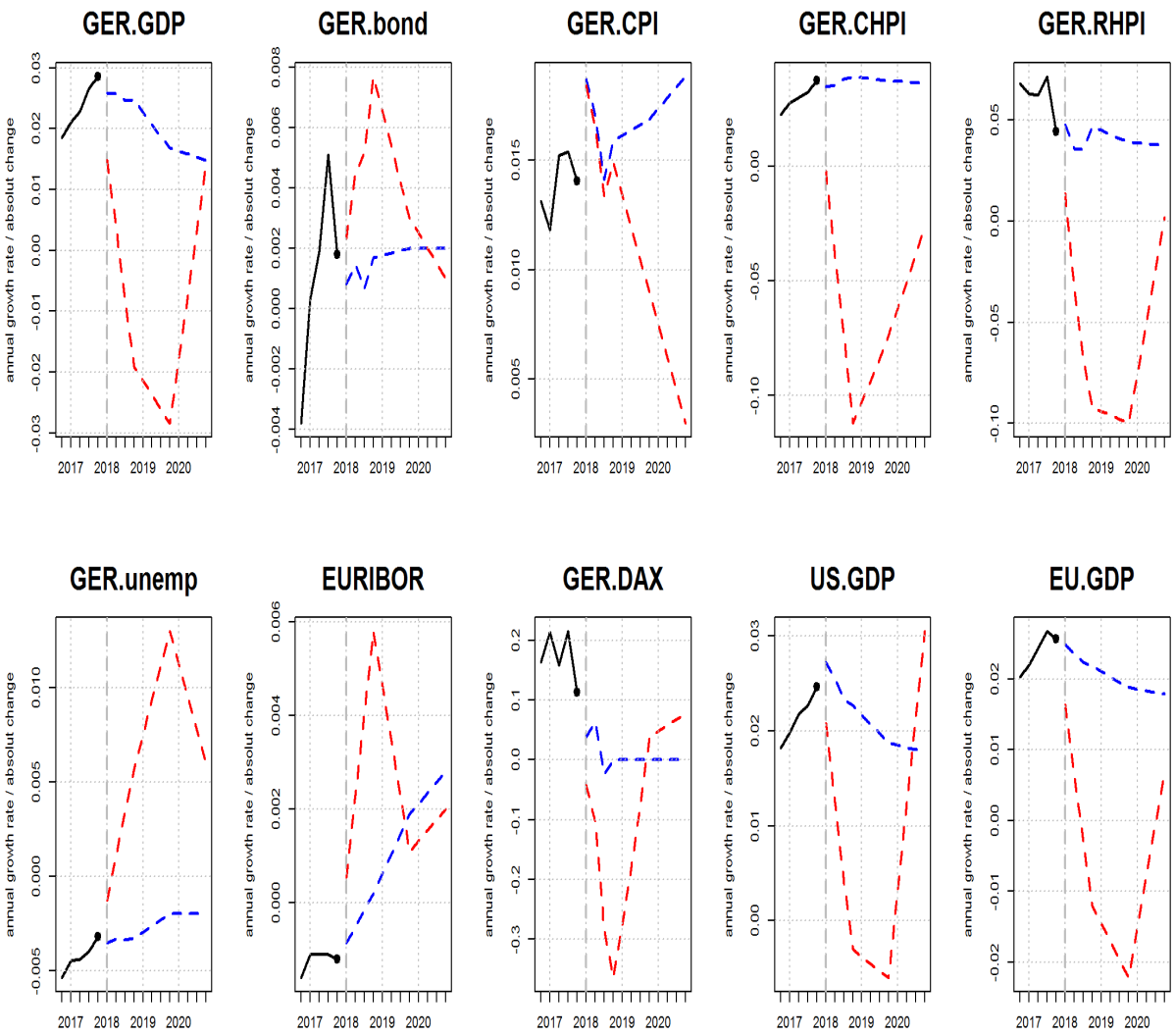


Figure 4: Macro Scenario

3.3 Calibration

This section discusses the calibration of the exogenous parameters, exogenous constraints and exogenous functions in our model. The choice of parameters is strongly driven by data sources, as well as jurisdiction and time period under consideration and is thus very specific to the precise application of the framework. The baseline calibration shown here should thus be understood as exemplary for our application to the German banking sector. An adoption of the framework to other banking sectors or data sources would warrant a thorough re-calibration.

As discussed above, our credit risk stress test, like most top-down stress tests, is constrained by the availability of long time series for the credit risk parameters. The PD time series we use for the BCBMA framework starts only in 2008Q1 which gives us 38 observations up to 2017Q2. To economize on degrees of freedom when specifying the maximum lag lengths in Equation (2) we allow - given the quarterly frequency - for a maximum of $L = 4$ exogenous lags but only a maximum of $K = 2$ endogenous lags.
The unconstrained vector of macro variables, x_t, contains all variables discussed in Section 3.2. To be precise

$$x_t = [GER.GDP, GER.bond, GER.CPI, GER.CHPI, GER.RHPI, GER.unemp, \\ GER.DAX, EU.EURIBOR, EU.GDP, US.GDP]'.$$

Thus, the number of possible variables before the application of filters to be included in x_t, N, is equal to $2 \times [4 \times 10] + 2 = 82$. The BCBMA framework considers model specifications of all sizes (i.e. using 1 up to $\bar{N}$ covariates, including endogenous and exogenous lags, but neglecting the constant). Again, given our short time series, we do not allow all N covariates plus the endogenous lags to be included into one model specification at the same time. Instead, we constrain the maximum number of covariates to $\bar{N} = 4$. Note that, since finally all models in $\bar{\Omega}$ (or Ω) are combined into a single model, this constraint does not affect the number of covariates in the combined model.
When filtering the original model space $\mathcal{M}$, we set the maximum allowed correlation between macro covariates within the same model specification, γ, to 0.8. The p-value threshold for the Durbin-Watson test is set to 0.10, which means that the H0 that there is no autocorrelation in residuals cannot be rejected at the 10 % level. Table 4 shows the sign restrictions that we impose on the LRM, where $+$ denotes a positive sign, $-$ a negative sign and 0 an unconstrained LRM[16]:

Table 4: LRM Sign Restrictions

Variable	LRM Restriction
GER.GDP	-
GER.bondyield	+
GER.CPI	0
GER.CHPI	-
GER.RHPI	-
GER.unemp	+
GER.DAX	-
EU.EURIBOR	0
EU.GDP	-
US.GDP	-

[16]Note that relative to Gross and Poblacion (2016) we relax the sign restriction on CPI and EURIBOR, as the effect of these variables on PDs seems to be ambiguous. For example, if increases in inflation and EURIBOR are driven by strong economic growth rather than supply side shocks, PDs may decrease and vice versa.

The Merton/Vasicek one-factor model outlined in Section 2 requires the calibration of two parameters D and ρ. For the former, we calibrate D to be sector-specific and to correspond to the time-average sector-specific PD in a distance-to-default transformation: $D_s = \Phi^{-1}\left(\bar{PD}_s\right), \forall s \in \mathcal{S}$. The structural PD Equation (3) was derived under the assumption that $z_t \sim N(0,1)$ and is valid only if this assumption holds. Thus, we calibrate ρ such that – given sector-specific PD time series $PD_{t,s}$ – $Var(\tilde{z}_{t,s}) = 1$ using a simple line search procedure. To be precise, we initially set $\rho_s = 0$, compute the time series $\tilde{z}_t$ using Equation (4), compute the variance and then repeat these steps for marginally increased ρ_s. We choose the ρ_s for which $Var(z_{t,s})$ is closest to unity. As mentioned above, we update the calibrations of D_s and ρ_s after each stress period, i.e. we update the time series for $\bar{PD}_s$ and $\tilde{PD}_{t,s}$ with the predicted stressed PD to derive a new set $\{\{\rho_s\}_{s\in\mathcal{S}}, \{D_s\}_{s\in\mathcal{S}}\}$. This dynamic approach amplifies the stressed PD dynamics, since during the stress horizon the increased volatility in the PDs are compensated by an increased ρ_s in order to ensure that $Var(\tilde{z}_{t,s}) = 1$. Thus, in the next period, the macro conditions will have a stronger impact on the PD dynamics than in the previous period. The parameter for the cluster width around the benchmark prediction for the PD increases under the adverse scenario, $\bar{\nu}$, is set to 2, i.e. we allow a cluster of two times the historical volatility-implied PD change around the benchmark estimate (see Equation(7)). This calibration is chosen such that in all sectors at least one ADL model specification survives the benchmark filtering. [17]

When approximating the models space, $\mathcal{M}$, we set $Q = 10,000$. We set the threshold for Occam's windows to $o = 20$, i.e. we filter out models that have a probability of being the best Kullback/Leibler model 20 times lower than the model with the highest probability.[18]

For the LGD model we set homogeneous unscaled initial LGD, $\bar{LGD}$ to 0.45 such that any LGDs in the stress test lie in the $[0.10, 0.45]$ interval. The calibrated value corresponds to regulatory value prescribed in CRR Art. 161(1a) for uncollateralized senior corporate exposures. For the TTC adjustment of the point-in-time PDs (PDpit) derived from the BCBMA model we use an $a = 4$ year rolling windows.

We use the IRB formula to derive dynamics in RWA over the stress horizon for both GCR banks $g \in \mathcal{G}$ (at the bank level) and non-GCR banks $n \in \mathcal{N}$ (at the sector level). However,

[17] While an extensive elaboration on the effect of the calibration of $\bar{\nu}$ is outside the scope of this paper, and potentially of little interest, since the results are rather specific to country, loan sector and stress scenario, our analyses suggest that for $\bar{\nu} = 0.3$ the BCBMA-implied and the Merton/Vasicek-implied stress result distributions are barely distinguishable. In contrast, our baseline calibration of $\bar{\nu} = 2.0$ induces a rather loose benchmark constraint. A more thorough calibration of $\bar{\nu}$ needs to optimize on the trade-off between reducing dispersion in $\bar{\Omega}$, i.e. reducing model uncertainty, and maintaining the dichotomy in the BCBMA by combining both reduced-form and structural perspectives, i.e. not setting the benchmark constraint that tight as no ADL-specification survives the filtering. Also, a more thorough calibration would need to feature a sector-specific $\bar{\nu}_s$ in order to better reflect sector-specific model uncertainty and to avoid that in some sectors only one model specification survives. Alternatively, one could pursue a pragmatic calibration approach by fixing the minimum number of ADL-specification which should survive the benchmark filtering in each sector, e.g. 10, stetting $\bar{\nu} = 1$, which induces the benchmark constrained only to be influenced by historical volatility of the time series and then increasing $\bar{\nu}$ until the minimum number of ADL-specifications survives. While the qualitative and quantitative effect of the benchmark constraint strongly depends on data and calibration the intuition of its effect is straightforward: the tighter the benchmark constraint is set, the more trust is put in the structural filter relative to the reduced-form estimations. Thus, the closer the final stress results will lie to the stress test results implied by the structural model alone.

[18] Madian and Raferty (1994) recommend a threshold between 20 and 100.

there is are several IRB formulas, depending on exposure type (see Articles 153 and 154 CRR). Therefore, we need to calibrate function $\mathcal{R}(\cdot)$ in Equation (21) to reflect the calibration corresponding to the underlying exposure types. Unfortunately, we do not observe exposure type (retail, SME,...) in the GCR. Thus we cannot directly infer the calibration of $\mathcal{R}(\cdot)$ from our data source. However, for the GCR banks we do observe the RWA corresponding to each borrower in the GCR. We use this information to deduce the loan type and approximate the calibration of $\mathcal{R}(\cdot)$ at the borrower level. To this end, we guess a set of different initial loan types for each borrower of a given GCR bank, calibrate $\{\mathcal{R}_{g,s}(\cdot)\}_{\mathcal{G},\mathcal{S}}$ following the CRR, compute the implied RWA and compare these with the RWA given in the GCR. We then choose the loan type from the set that minimizes the difference between IRB-implied and observed RWA. Let $\{\mathcal{R}^{\star}_{g,s}(\cdot)\}_{\mathcal{G},\mathcal{S}}$ denote the optimal calibration at the borrower level. Then, for GCR banks $RWA_{g,s,t} = \sum_{u=1}^{\mathcal{U}} \mathcal{R}^{\star}_{g,s}\left(PD_{u,g,s,t}, LGD_{u,g,s}, EaD_{u,g,s,t}\right)$.[19] For non-GCR banks we compute RWA dynamics at the sector level, i.e. these dynamics depend only on the sector of the loans exposure but not on the bank. To compute sector-specific RWA dynamics we aggregate the optimal calibrations of the GCR banks at the borrower level to the sector level by computing for each of the GCR banks the mean optimal calibration over all loans within each sector, $\{\bar{\mathcal{R}}^{\star}_{g,s}(\cdot)\}_{\mathcal{G},\mathcal{S}}$. We then use $\bar{\mathcal{R}}^{\star}_{g,s}$ to compute RWA using sector-aggregated PD, LGD and exposure over the stress horizon, which gives us $\{RWA_{g,s,t}\}_{\mathcal{G},\mathcal{S},T}$. Finally, we eliminate the $\mathcal{G}$ dimension by weighting these RWA by how often each $\bar{\mathcal{R}}^{\star}_{g}$ is the optimal calibration among all GCR banks relative to all GCR banks in sector s; let $\{w_{g,s}\}_{\tilde{\mathcal{G}},\mathcal{S}}$ denote those weights. Then $RWA_{s,t} = \sum_{g\in\tilde{\mathcal{G}}} w_{g,s} RWA_{g,s,t}$. Table 5 summarizes the calibration.

[19] To be precise, we include six different calibrations into the choice set: corporate exposure with size parameters 5, 25 and 50, retail exposures with size parameter 5, qualifying revolving retail exposures with size parameter 5 and retail exposure secured by immovable property with size parameter 5. This relatively small choice set is due to computational constraints. The IRB formula has to be calibrated for each of the more than $100,000$ loans in the GCR such that an extensive grid search would increase computation time exponentially.

Table 5: Summary parameter calibration

Parameter	Description	Calibration
K	endogenous lags in ADL model	2
L	exogenous lags in ADL model	4
$\bar{N}$	max. number of right-hand-side variables in ADL specification	4
γ	correlation restriction	0.80
d	Durbin Watson p value restriction	0.10
D_s	default threshold in Merton/Vasicek model	$\Phi^{-1}\left(\bar{PD}_s\right)$
ρ_s	systemic impact in Merton/Vasicek model	line search
$\bar{\nu}$	cluster width for benchmark constraint	2
o	Occam's Window threshold	20
Q	best models in untrimmed MCS according to leaps-and-bounds	10,000
$L\bar{G}D$	unscaled initial LGD	0.45
a	rolling windows horizon for TTC PDs (years)	4

4 Quantitative Results

This section studies the quantitative effects of model uncertainty on top-down stress test results using the German banking sector as an example case. In a first step, Section 4.1 analyzes the model space and elaborates on the effects of imposing the filters and the benchmark constraint on predicted PD increases over the adverse macro scenario. In a second, step Section 4.2 then extends the scope to the final stress test results at the bank level by studying the effect of model uncertainty on predicted capital depletion.

4.1 Model Uncertainty and predicted PD increases

As discussed in Sections 1 and 2.1, a scare data environment can make estimation results highly sensitive to model specification. To put some flesh on this thought, Table 6 provides some moments of the forecast distribution in the unfiltered model space $\mathcal{M}$ for the uniquely modeled NACE sectors in the GCR. It shows that the dispersion in predicted PD increases over the adverse macro scenario is – over all sectors – huge. The standard deviation of the forecast distribution ranges from 41 % to 1,782 % PD increase, implying that any stance on the economic and statistical significance of PD effects, let alone on the implied size of the capital effects, is a judgment call. This results clearly shows that, in a stress test application if it is deemed necessary to hand pick a given model specification to link the scenario to credit risk parameters, it is crucial to have at least an understanding of the dispersion in the relevant model space in order to be able to judge whether the chosen specification lies in a dense or very sparse model space region.
Even better, as argued by Henry and Kok (2013) and Gross and Poblacion (2016), in such an environment one should not just pick one specification but rather employ all the information available in the model space, say, by using BMA. However, as Table 6 shows, some sectors

have a significant mass of model that predicts a PD *decrease* in response to the adverse macroeconomic scenario. While this may be plausible from a statistical point of view, for example, owing to strong impacts of idiosyncratic crisis events on the time series, it is hardly plausible or desirable from an applied stress testing perspective.[20] However, the baseline BMA model would assign a positive weight to models predicting a (substantial) PD decrease, which either reduces predicted PD increases or even induces aggregate decreasing PDs at the sector level.

Table 6: $\mathcal{M}$-implied PD increases in adverse scenario (%)

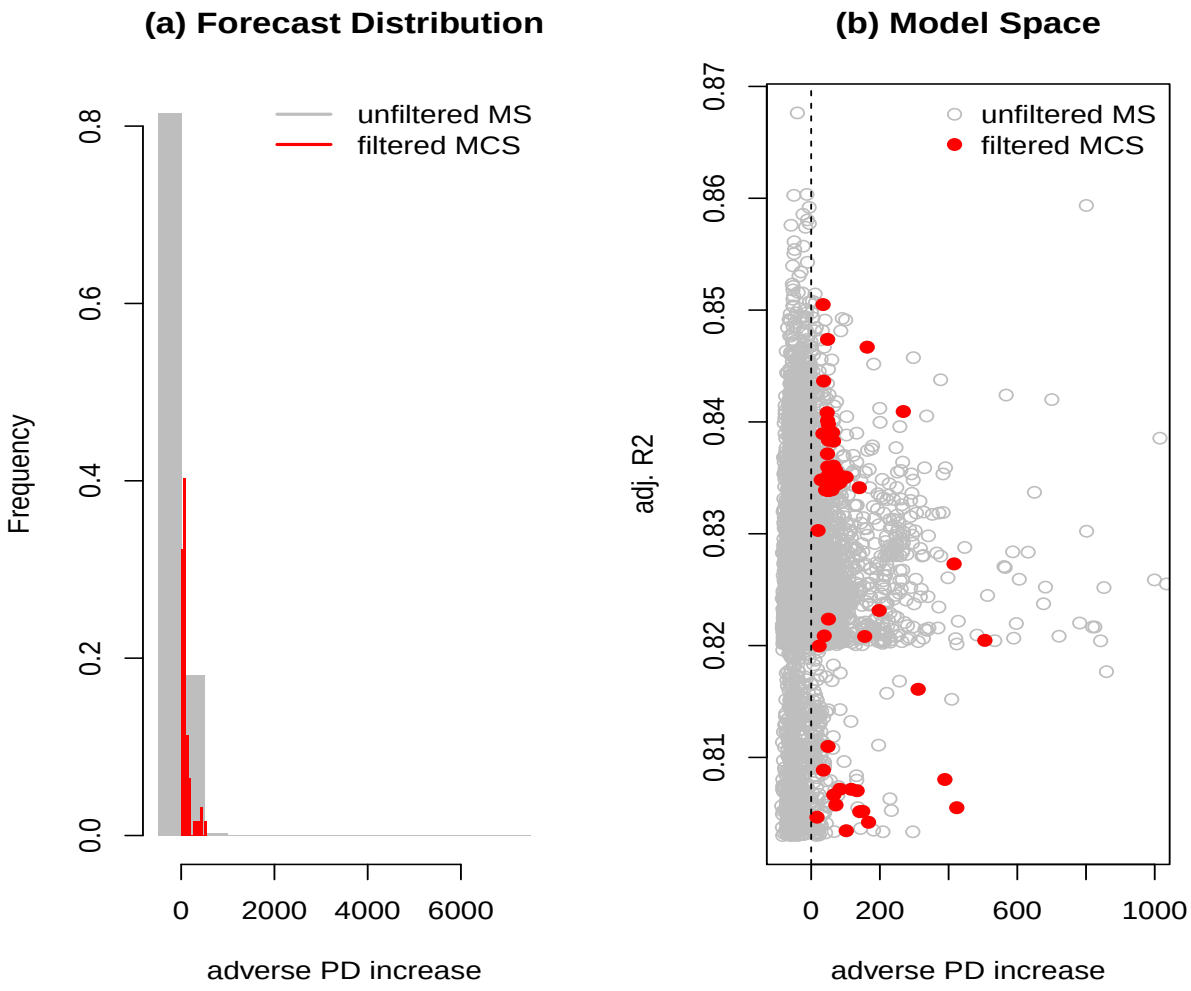
Sector	Min	Median	Max	SD	Median(R^2)
1	-85.7	-36.6	7027.7	146.9	82.5
2	-97.3	133.5	4753.4	909.2	38.5
3	-85.2	144.3	5066.9	543.2	65.6
4	-76.5	-52.8	6763.6	180.3	57.1
5	-73.4	19.8	2131.6	110.5	73.7
6	-91.3	30.7	2237.0	200.8	34.9
7	-98.9	-84.6	810.4	40.7	72.7
8	-98.6	-35.8	4172.0	346.1	68.8
9	-92.0	509.3	2880.1	705.9	29.8
10	-99.9	-80.9	30463.1	1782.1	71.5
11	-96.4	-29.6	9698.6	570.0	48.3
12	-93.3	-23.6	3049.7	62.9	60.5
13	-93.3	28.9	5567.6	509.8	60.5

Notes: PD increase over the three years in adverse scenario. All values as a percentage. Number of models for each sector $Q = 10,000$.

Figure 5 zooms in on Sector 5 and shows the full forecast distribution of the unfiltered model space (Panel a, gray bars) as well as the implied dependency between the adjusted R^2 and the predicted PD increase under the adverse scenario (Panel b, gray circles). The figure emphasizes that the adjusted R^2 – a metric often used when hand picking a given model specification – is not necessarily of good predictor for statistic and economic plausibility. This becomes evident when we impose the filters for statistic and economic plausibility on the unfiltered model space (filters 1-3 in Section 2.1, i.e. all filters except the benchmark constraint). After filtering $\mathcal{M}$ to the unconstrained model confidence set Ω, both the dispersion in the forecast distribution and the adjusted R^2 decrease strongly. In particular, the filtering eliminates any model that predicts a decrease in sector PDs over the horizon of the adverse scenario. Table C.1 in Appendix C shows which fraction of model in the unfiltered model space survives each filtering criterion. This shows that in this sector the most binding filter is the sign restriction filter, which only 8.4 % of models pass. Table C.2 extends Table

[20] For example, a crisis events that originates in the financial sector with less contagion to the real economy may lead, for example, to stable unemployment dynamics. Since unemployment is a mayor determinant of PDs in the retail sector, PDs may remain stable or even decrease in these retail sectors. In Germany such dynamics may have been amplified due to the introduction of short-time work during the recent financial crisis.

6 by comparing moments of the forecast distribution in the unfiltered model space $\mathcal{M}$ and filtered model confidence set Ω.



Notes: Unfiltered MS corresponds to $\mathcal{M}$, filtered MCS corresponds to Ω. Number of model in $\mathcal{M}$: 10,000. X-axis in Panel (b) cut off at $1,000\,\%$.

Figure 5: Impact of Model Specification on predicted PD increases: Sector 5

While filtering reduces the dispersion of the predicted PD increases in Ω and eliminates implausible specification, the overall dispersion of the remaining models remains large in most sectors with sectoral standard deviations between $15\,\%$ and $4,500\,\%$ (see Table C.2). As the adverse PD dynamics strongly affect stress test results at the bank level, obtaining sound estimates for those dynamics in the scarce data environment is of ample importance for both banks and supervisors. Our suggested BCBMA framework aims at reducing the MCS dispersion in a theoretically founded way, thereby limiting the weight of models with implausible predicted PD increases over the stress horizon in the final BMA model by applying the additional benchmark constraint to Ω.

To this end, we map the adverse macro scenario to the structural PDs from the Mer-

ton/Vasicek model using a multivariate QMap approach as discussed in Section 2.1. The QMap-implied PD increases are then transformed into the benchmark-constrained (or stress test plausible) region according to Equations (6) and (7). Table 7 shows the Merton/Vasicek-implied average PD increases over the stress horizon at the uniquely modeled sector level for both a univariate QMap and our multivariate extension. Two results deserve emphasis: *first*, QMap does not rely on the identification of historical correlations but hinges instead on the assumption of a co-monotonic relationship between the macro variables and the systemic factor z.[21] Thus, by assumption, decreasing PD dynamics in response to deteriorating macro conditions are not feasible. This is why the benchmark constraint imposes "stress test plausibility" on the model space. *Second*, the multivariate QMap model that we employ increases the in-sample fit between predicted and observed sectoral PDs. Relative to the lowest normalized root-mean-square error (NRMSE) of a univariate QMap, which only employs the macro macro variable with the highest predictive power for the sectoral PDs, the multivariate QMap approach reduces the NRMSE by on average 20 % and by up to one-third. Thus, we suggest the multivariate QMap discussed in Section 2.1 as an easy and traceable approach to increasing the quality of the fit for quantile mapping models.[22] The PD increases in the first column of Table 7 define the middle of the symmetric benchmark-constrained region in Ω.

Table 7: Multivariate QMap-implied PD increases in adverse Scenario

Sector	Δ PD (%)	NRMSE(MA)	min{NRMSE(Uni)}
1	93.9	0.54	0.63
2	73.4	0.94	1.31
3	64.6	0.63	0.79
4	29.1	0.82	1.01
5	18.8	0.89	1.13
6	44.9	0.88	1.10
7	46.5	0.76	0.83
8	133.2	0.88	1.03
9	17.7	0.94	1.40
10	136.8	0.90	1.22
11	407.1	0.51	0.58
12	91.8	0.74	0.98
13	85.4	0.90	1.12

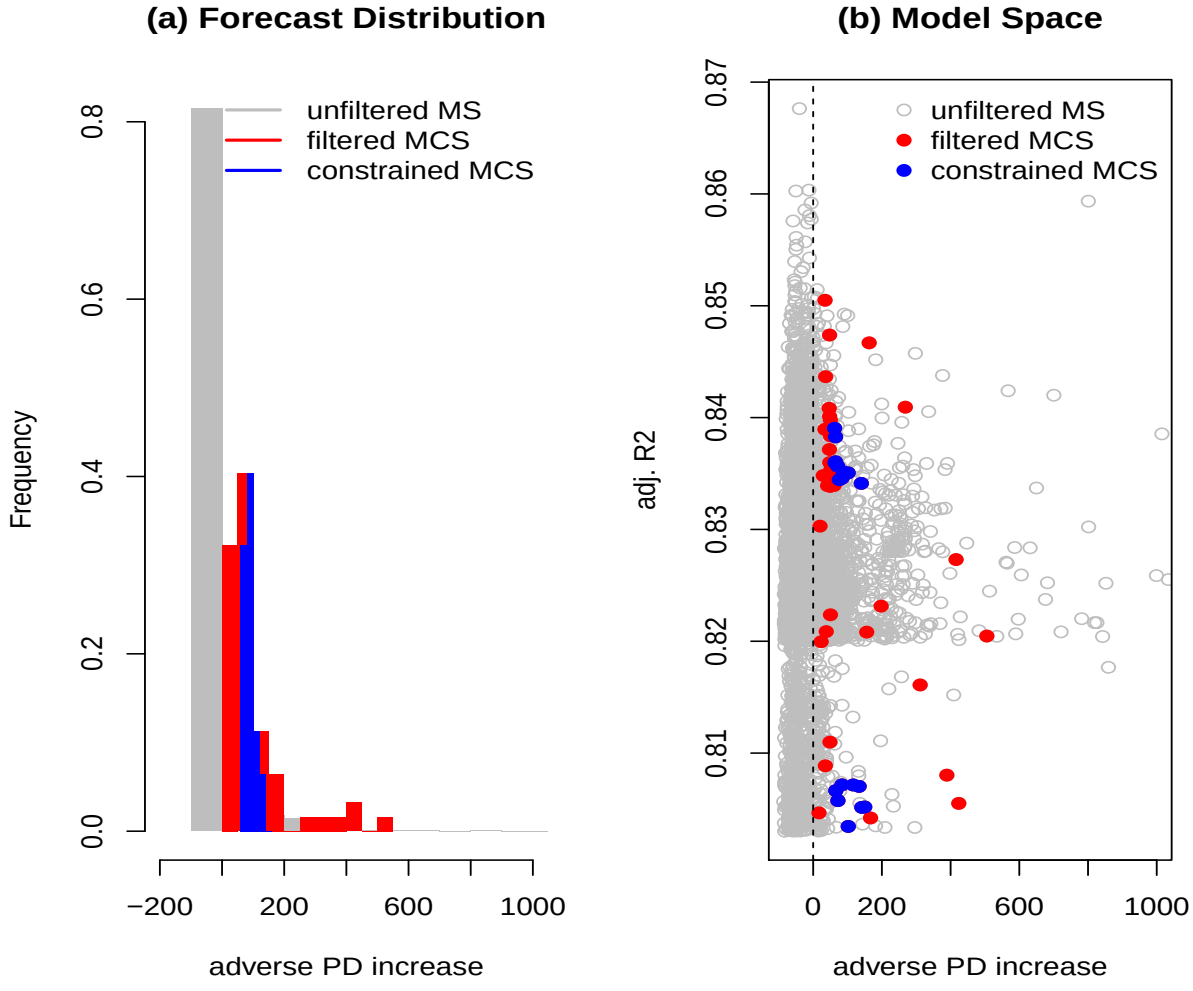
Notes: Δ PD (%) denotes the average relative PD increase over the three-year stress horizon. NRMSE(MA) denotes the normalized RMSE of the multivariate QMap model and min{NRMSE(uni)} denotes the minimum normalized RMSE over all univariate QMap models.

Using Sector 5 again as an example, Figure 6 shows the forecast distribution of $\bar{\Omega}$ relative

[21] As mentioned above, QMap is not affected by a scarce data environment to the same extent as reduced-form models, as long as the skewed-t distribution can be fit thoroughly.

[22] Appendix D shows some additional exemplary model output for Sector 5.

to both $\mathcal{M}$ and Ω.[23] Unlike the unconstrained MCS, the "stress test plausible" models do neither predict very low nor extremely high increases in PDs. Table 8 shows the full details for all sectors. Comparing columns Ω and $\bar{\Omega}$, we see that the benchmark constraint trims off specifications at both tails of the forecast distribution, thereby reducing its standard deviation substantially. In our application the constraint mainly affects very high predicted PD increases. However in some sectors (e.g. sectors 1 and 12) the median PD increases rises due to the benchmark constraint. On average over all sectors the maximum predicted PD increase is reduced from $2,912\,\%$ in Ω to $132\,\%$ in $\bar{\Omega}$. The average minimum predicted PD increase rises from $56\,\%$ to $79\,\%$.



Notes: Unfiltered MS corresponds to $\mathcal{M}$, filtered MCS corresponds to Ω and constrained MCS corresponds to $\bar{\Omega}$. Number of model in $\mathcal{M}$: 10,000. X-axis in both panels cut off at 1,000.

Figure 6: Impact of Model Specification on predicted PD increases: Sector 5

[23] Appendix E shows posterior inclusion probabilities and long-run multipliers for Sector 5.

Table 8: $\mathcal{M}$-, Ω- and $\bar{\Omega}$-implied PD increases in adverse scenario (%)

	$\mathcal{M}$				Ω					$\bar{\Omega}$				
Sector	Min	Median	Max	SD	Min	Median	Max	SD	#	Min	Median	Max	SD	#
1	-85.7	-36.6	7027.7	146.9	16.8	57.5	505.5	104.7	62	62.1	83.8	149.9	30.4	18
2	-97.3	133.5	4753.4	909.2	41.1	562.7	3652.6	862.0	664	41.1	50.2	89.4	21.6	4
3	-85.2	144.3	5066.9	543.2	70.9	204.6	938.6	137.2	439	70.9	70.9	70.9	–	1
4	-76.5	-52.8	6763.6	180.3	9.7	112.2	2025.1	201.4	146	9.7	29.5	34.0	7.6	9
5	-73.4	19.8	2131.6	110.5	0.3	24.1	340.8	33.0	1789	0.3	22.6	67.4	15.1	1665
6	-91.3	30.7	2237.0	200.8	11.8	143.9	1367.0	168.9	850	16.4	56.4	88.2	13.7	323
7	-98.9	-84.6	810.4	40.7	94.1	160.2	546.7	81.1	29	94.1	143.8	150.7	19.2	8
8	-98.6	-35.8	4172.0	346.1	160.2	360.3	637.3	155.5	11	160.2	170.3	200.3	20.8	3
9	-92.0	509.3	2880.1	705.9	71.5	978.0	2607.6	612.6	1514	11.6	12.7	13.8	1.6	2
10	-99.9	-80.9	30463.1	1782.1	196.3	2862.8	13089.1	4520.9	9	196.3	196.3	196.3	–	1
11	-96.4	-29.6	9698.6	570.0	46.6	446.7	7520.8	1541.3	348	318.4	404.4	544.4	61.5	150
12	-93.3	-23.6	3049.7	62.9	0.3	16.0	67.9	14.1	107	39.4	48.9	67.9	10.4	9
13	-93.3	28.9	5567.6	509.8	1.8	72.1	4559.1	665.9	654	9.4	32.1	46.8	9.1	239

Notes: PD increase over the three years in adverse scenario. All values, except #, in percentages. Number of models for each sector $Q = 10,000$. "#" denotes surviving models post applying the respective filters.

Considering the effect when applying the benchmark constraint to the unfiltered model space in terms of the in-sample performance (R^2) but also in terms of the out-of-sample predictive performance, it is important to bear in mind that the BCBMA does not aim at predicting the PD conditional on observed normal business cycle times but rather conditional on rarely observed severe tail events outside the bounds of the observation space. The benchmark constraint filters out some informational content by imposing a plausibility benchmark on the range of PD increases, thereby reducing the forecast performance of the BCBMA-combined model *measured against the observation space*. But since the observation space is less relevant in a stress testing context, where normal business cycle correlations are less representative for expected crisis dynamics, the BCBMA model increases the plausibility of predicted PD increases that lie outside the observation space by assuming a second, structural perspective on PD increases. Table 9 compares the out-of-sample forecast performance measures as the normalized root-mean-squared error (NRMSE) for the filtered but unconstrained MCS Ω and for the benchmark constrained MCS $\bar{\Omega}$. To be precise, we compare the out-of-sample NRMSE between Ω and $\bar{\Omega}$ but also, within each MCS, between the singular model with the lowest NRMSE in the MCS (best component model, BCM) and the combined model. Two points should be emphasized: *first*, in both MCS the combined model induces a (weakly) lower NRMSE compared to the BCM. This result is in line with Baele et al. (2015) and reflects the fact that the combined BMA or BCBMA is founded on a larger observation space than a single model. *Second*, we see that the NRMSE of $\bar{\Omega}$ is indeed higher relative to the NRMSE in Ω. As discussed above, the BCBMA model increases the plausibility of stressed credit risk parameters – benchmarked against the Merton/Vasicek-implied PDs – at the cost of filtering out information which is contained and employed in Ω.[24]

[24]It shall also be mentioned that the BMA also has lower NRMSE compared to a pure QMap approach. In addition, the use of out-of-sample weights reduces the out-of-sample NRMSE relative to in-sample weights. These results are available from the authors upon request.

Table 9: Out-of-sample Performance Ω vs $\bar{\Omega}$

Sector	Ω			$\bar{\Omega}$		
	NRMSE (BCM)	NRMSE (BMA)	Δ (%)	NRMSE (BCM)	NRMSE (BCBMA)	Δ (%)
1	0.42	0.41	-2	0.46	0.44	-4
2	0.75	0.68	-9	0.82	0.74	-10
3	0.47	0.47	-1	0.74	0.74	0
4	0.65	0.54	-18	0.71	0.70	-2
5	0.44	0.44	0	0.44	0.44	0
6	0.76	0.69	-10	0.77	0.76	-2
7	0.53	0.52	-2	0.53	0.52	-2
8	0.60	0.55	-8	0.65	0.64	-3
9	0.76	0.70	-8	0.96	0.95	-1
10	0.57	0.47	-18	0.62	0.62	0
11	0.61	0.60	-1	0.67	0.64	-3
12	0.58	0.56	-4	0.68	0.65	-3
13	0.55	0.48	-13	0.55	0.50	-8
Mean	0.59	0.55	-7	0.66	0.64	-3

Notes: NRMSE denotes the normalized root-mean-squared error; BCM denotes the best component model in the respective model confidence set; Ω denotes the MCS without the benchmark constraint; $\bar{\Omega}$ denotes the benchmark constrained MCS. A Δ of zero implies that only one model survives the benchmark filter. This is induced by our exemplary calibration of $\bar{\nu} = 2$ for all sectors. A more through calibration should ensure that sufficient models survive the benchmark filter in order to keep the dual perspective of reduced form and structural (see also footnote 17).

4.2 Model Uncertainty and predicted capital impact

The previous section discussed the effect of model uncertainty on predicted stress dynamics of sectoral default probabilities. However, from the perspective of applied credit risk stress testing, the effect of model uncertainty on predicted capital depletion at the bank level is of greater importance. Thus, in this section we conduct a full-fledged top-down credit risk stress test for the German banking sector and analyze the impact of PD model choice on final stress test results. This exercise is intended as an illustration and the quantitative results depend on model calibration as well as on data quality. The calibration suggested in Section 3.3 should be considered as a working example, and the model may require re-calibration depending on the macro scenario, data or the banking sector under consideration.

For the credit risk stress test we consider German banks for which we have recent starting value data as of 2017Q2, which leaves us with a total of 1,513 banks.[25] The sample of banks consists mainly of cooperative banks ($\simeq 60\,\%$) and savings banks ($\simeq 30\,\%$).

We consider the distribution of stress effects across the the German banking system (capturing both effects on CET1 capital due to impairments and on RWA due to PD increases)

[25]For a detailed description of starting values see Table B.1.

within both Ω (unconstrained MCS) and $\bar{\Omega}$ (benchmark constrained MCS). Within each of the two MCS, we consider three different estimators for the PD dynamics in each sector:

1. the BMA combination with out-of-sample weights for each sector of all non-filtered model specifications in Ω and $\bar{\Omega}$, respectively,

2. the single model specification in each sector for both Ω and $\bar{\Omega}$, which predicts the *lowest* increase (or highest decrease if occurring) in sectoral PDs at the end of the stress horizon ("lower") and

3. the single model specification in each sector for both Ω and $\bar{\Omega}$, which predicts the *highest* increase in sectoral PDs at the end of the stress horizon ("upper")

Table 10 shows moments for each of these six different stress effect distributions over all banks under the adverse macro scenario. The table features two core insights:
First, consider the two extreme bounds of stress distributions by comparing the "Upper" and the "Lower" single model specifications both within each MCS and between MCSs. We find that the choice of one particular specification of Equation (2) may strongly effect stress test results. In our application to the German banking sector, the most optimistic stress test outcome could actually results in a positive median stress effect of 3.6 pp (2.3 pp) in Ω ($\bar{\Omega}$), while the most pessimistic median stress outcome would lead to a substantial capital depletion of -35.3 pp and -23.5 pp in Ω and $\bar{\Omega}$, respectively. Despite its rather loose calibration of $\bar{\nu} = 2$ the benchmark constraint reduces the range of extreme bounds of stress distributions from 38.9 pp to 25.8 pp, or by roughly 35 %. The large distance between these two extreme stress effect distribution translates to a large variance in implied capital shortfalls. As an illustrative example we measure any bank level capital shortfall against an exogenous hurdle rate of 8 % of CET1 capital.[26] We find that for both Ω and $\bar{\Omega}$ a factor of 100 lies between the most optimistic and the most pessimistic single model specification. Clearly, the range between these bounds would strongly affect the interpretation and consequences of any stress test exercise. These results suggest that prior to choosing one particular specification from the model space of Equation (2) one should at least be aware of the model space dispersion. Better yet, instead of relying on one particular model calibration, one should employ the informational content of various specifications to reduce model uncertainty, for example, by using a BMA weighting scheme.
Second, however, Table 10 also suggests that the pure BMA approach may be susceptible to data issues that cannot be addressed solely within the reduced-form paradigm, as , for example, short available time series and collinear macro covariates. In our application to the German banking sector we find that filtering the model space only for statically and economically plausible specifications leaves us with a substantial mass of specifications that predict very strong PD increases in response to the adverse macroeconomic scenario. For example, the average maximum PD increase within the three scenario years over all 13 sectors is $2,912$ % in Ω (see Table 8). A significant fraction of these strong increases is deemed "implausible" relative to our structural benchmark model, even with a rather loosely calibrated benchmark constraint for most sectors. As a consequence, filtering out these

[26] The results is independent of the particular calibration of the hurdle rate.

implausible specifications reduces the average maximum PD increase over all sectors to 132 % and the median BMA stress effect from −5.0 pp to −2.5 pp, or by 50 % in our application. The rightwards shift of the stress effect distribution due to the benchmark constraint induces a reduction of the measured capital shortfall from €29,144 million to €8,196 million, or by 72 %. Against this quantitative background, we argue that the BCBMA model offers a more precise estimation of stress effects and capital shortfalls by trimming off extreme stress prediction on both sides of the model space. Those specifications may appear plausible from within the reduced-form paradigm but are deemed implausible against the structural benchmark filter.

Table 10: Model Uncertainty and Stress Effects

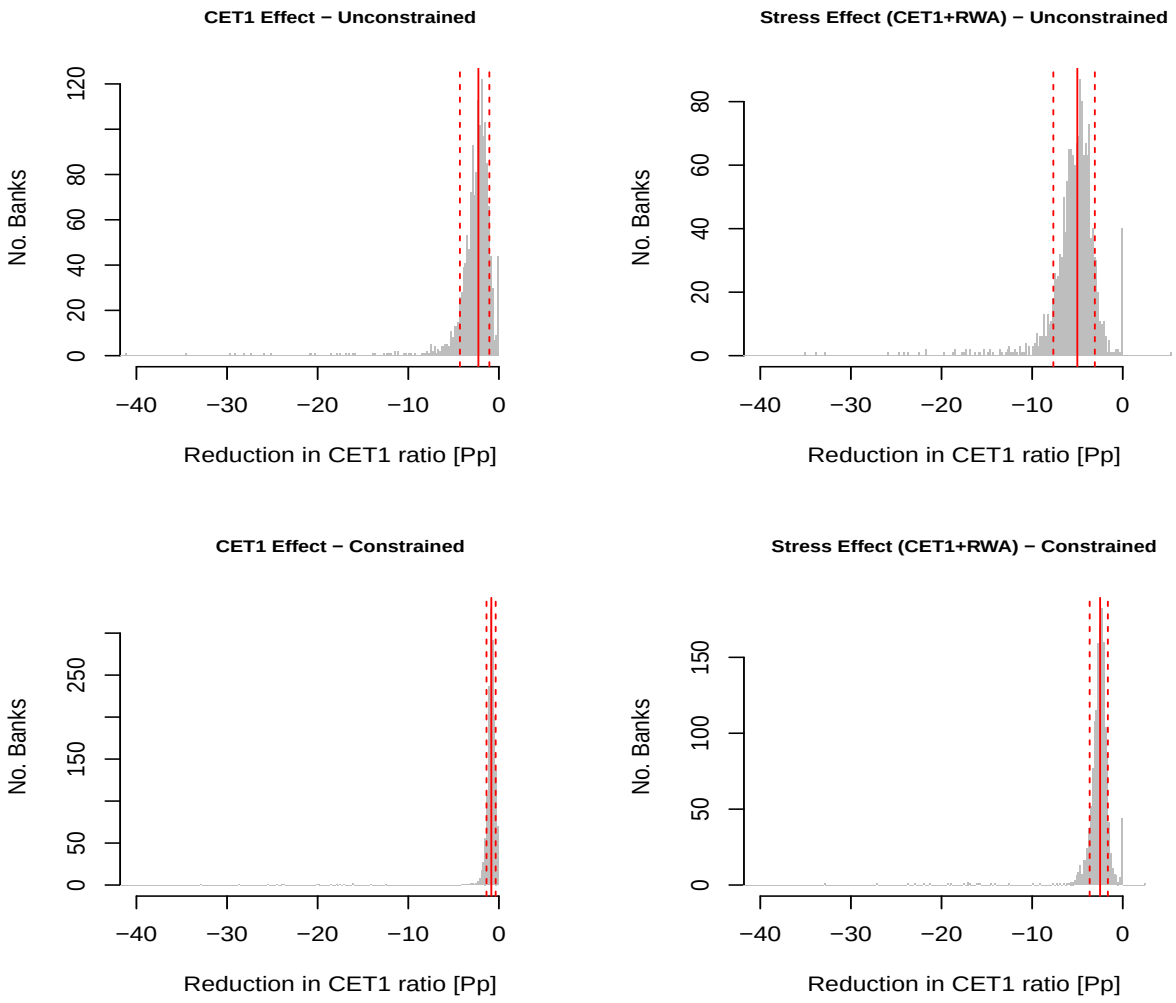
		Unconstrained MCS			Constrained MCS		
		BMA	Upper	Lower	BMA	Upper	Lower
Stress Effect	10^{th} Quantile	-7.7	-52.4	-0.7	-3.7	-34.5	-0.5
	Median	-5.0	-35.3	3.6	-2.5	-23.5	2.3
	90^{th} Quantile	-3.1	-17.2	9.8	-1.7	-11.5	6.3
	Mean	-5.5	-36.4	5.1	-2.8	-23.4	3.1
	Std. dev.	3.7	45.0	28.5	3.0	10.9	18.8
Capital Shortfall		-29,144	-608,558	-5,675	-8,196	-403,816	-3,824

Notes: Stress in percentage points of CET1 capital. Capital shortfall (in € million) indicates the aggregated shortfall in CET1 capital conditional on an 8 % hurdle rate measured in CET1 capital.

Figure 7 shows the distributions of the CET1 effects (only considering the effect on CET1 capital due to impairments and neglecting RWA dynamics) and stress effects for Ω BMA model (unconstrained MCS, first row) and $\bar{\Omega}$ BCBMA model (benchmark-constrained MCS, second row).

Due to the lower sectoral PD increases in the BCBMA framework the distribution of both CET1 effects and stress effects are tighter compared to the BMA framework. In our application to the German banking sector, banks which would suffer a strong capital depletion under the BMA framework experience lower depletions under the BCBMA framework, while banks with low CET1 effects and stress effects are less affected by the benchmark constrained. While this results is of course not an intrinsic feature of the BCBMA model in general but intrinsic to our application, it again emphasizes the potential impact of model uncertainty for microprudential regulation. Data issues in supervisory top-down stress tests, such as short time PD time series or collinear regressors, could potentially be pivotal in a pass-fail exercise.[27]

[27] As a rough indication of how the BCBMA stress testing framework compares to the bottom-up stress testing exercise of the EBA, we feed the 2016 adverse macroeconomic scenario into our model and compute the capital effects for the German EBA 2016 sample. We find that the median deviation for those banks from the EBA 2016 EU-wide capital effects (including RWA) amounts to 0.12% or 0.36 pp CET1 ratio. This deviation is, among others, driven by differences in methodology, data sources and level of aggregation.



Notes: Distributions over all 1,513 participating banks. Solid line indicates median, dashed lines 10th and 90th percentile. CET1 effect is computed holding RWA constant at initial level; stress effects also reflect changes in RWA due to PD dynamic during the stress horizon.

Figure 7: Bank Distributions depending on PD model

5 Conclusion

In this paper we elaborate on the quantitative effects of model uncertainty - when linking the macro scenario to stressed PD dynamics - on stress test results. We show for the case of PDs that aggregated BMA stress predictions may be affected by a sufficient large mass of extreme predictions in the model space. If these extreme outcomes are due to data issues, such as short PD time series or multicollinear regressors, this may systemically bias stress test results. To mitigate the impact of such issues in stress testing, we introduce a filtering criterion to the model space which filters out PD increases which can be deemed "implausible" according to the predicted PD increases of a structural Merton/Vasciek one-factor model. To map the one-factor model to the macro scenario we suggest a multivariate quan-

tile mapping approach which - similar in spirit to BMA - combines various univariate models according to their RMSE and reduces the NRMSE relative to the best component model. Due to the assumption of co-monotonicity, the quantile mapping has the advantage that it guarantees plausible PD dynamics during the stress horizon. The benchmark constraint increases the theoretical coherence of the ADL-predicted PD increases at the cost (contra the BMA paradigm) of neglecting information from the observation space. However, we argue that information from the observation space, while potentially relevant in a standard forecasting context, is of less importance in a stress testing context, which is concerned with predicting dynamics conditional on severe tail events.

In an application of the BCBMA top-down stress testing framework to the universe of 1,513 German banks sector we show that the BCBMA-predicted stress test results suggest substantially weaker stress effects than the standard BMA (which also accounts for statistical and economic plausibility): while, under the standard BMA, the median stress effect is predicted to be -5.0 pp of the CET1 ratio, the median effect under the BCBMA model is only -2.5 pp. These weaker impacts on banks' capital position reduce the predicted capital shortfall against a hypothetical hurdle rate of 8 % CET1 capital by 72 % from €29, 144 million to €8, 196 million. These results highlight the importance of being aware of and accounting for the effects of model uncertainty in stress testing applications, since the interpretation and also the consequences of an internal or supervisory stress test could strongly depend on the choice of model specification.

That said, the exact quantitative effects of the BCBMA model depend the particular application, data and calibration at hand. To further mitigate the effects of model uncertainty on stress test results, the calibration of the benchmark constraint should be based on thorough back testing instead of on an ad hoc calibration as used in this paper. Also, this paper has neglected the role of uncertainty when linking the macro scenario to stressed LGD dynamics. Both tasks are left for future research.

A NACE Sectors

Table A.1: Overview NACE Sectors 2017Q2

				GCR			GBS
Sector	**Description**	**Comined with Sector**	**Sector ID**	**IRB Exp. Share**	**Risk Weight**	**Imp. Factor**	**Exp. Share**
A	Agriculture, Forestry & Fishing	E	1	0.36	0.46	0.0034	1.21
B	Mining & Quarrying		2	1.47	0.52	0.0046	2.02
C	Manufacturing	M, O	3	11.60	0.54	0.0040	4.32
D	Electricity, Gas, Steam and air conditioning supply		4	7.77	0.62	0.0034	1.01
E	Water Supply	A		0.87	0.62	0.0031	1.01
F	Construction	H, L, R	5	2.02	0.87	0.0092	1.66
G	Wholesale and Retail trade	N	6	5.38	0.48	0.0058	3.99
H	Transportation & Storage	F, L, R		8.53	0.34	0.0050	1.40
I	Accomadation & Food Service Activities		7	0.48	0.51	0.0051	0.48
J	Information & Communication Service		8	1.63	0.75	0.0040	1.07
K	Financial & Insurance sector	S	9	4.12	0.94	0.0136	5.64
L	Real Estate Activities	F, H, R		13.95	0.51	0.0055	3.10
M	Professional, Scientific & Technical Activities	C, O		5.34	0.48	0.0040	2.29
N	Adminstrative & Support Service Activities	G		4.31	0.36	0.0033	2.32
O	Public Adminstration & Defence, Compulsary Social Security	C, M		0.48	0.17	0.0023	0.00
P	Education		10	0.23	0.33	0.0009	0.34
Q	Human Health Services & Social Work Activities	T, U	11	1.86	0.32	0.0022	1.71
R	Arts, Entertainment & Recreation	F, L, H		0.44	0.52	0.0060	0.67
S	Other Service Activities	K		0.42	0.78	0.0094	0.51
T	Activities of Households as Employers	Q, U		1.45	0.25	0.0011	15.75
U	Activities of Extraterritorial Organisations	T, Q		-	-	-	0.00
V	Residential Real Estate		12	7.89	0.28	0.0014	30.22
W	Commercial Real Estate		13	19.23	0.18	0.0020	18.90

B Data sources

As discussed in Section 3, we require regulatory data to compute the final stress impact on banks' CET1 capital position. To obtain initial values for each bank's CET1 ratio we draw on the Corep reporting. Corep has generally to be reported at the single bank level. However according to CRR Art. 7(1-3) competent authorities may waive the requirement to report at the single bank level under certain conditions. If a subsidiary waiver (CRR Art. 7(1)) or a parent waiver (CRR Art. 7(3)) is granted to an institute, Corep and Finrep only have to be reported at the group level. Also, for those institutes capital requirements apply only at the group level.
This possibility makes the construction of a consistent data basis slightly more cumbersome as *(1)* for waiver banks there is no Corep reporting on the single bank level available and *(2)* there is no reporting at the group level for P&L items using Bundesbank's quarterly reporting consistent with Corep. Thus, for P&L items of waiver institutes we need to draw on Finrep reporting at the group level to obtain data consistent with the reporting of CET1 capital and RWA. Therefore, three types of institutes participate in our top-down credit risk stress test:

1. single institutes without any group membership,
2. single institutes with group membership which may be either partially consolidated subsidiaries or parents without a parent waiver in the sense of CRR Art. 7(3) or
3. groups which are the fully consolidated unit of parent waiver institute and all fully consolidated subsidiaries.

In Germany, we currently have nine institutes with a parent waiver, eight of those with IFRS reporting and one with nGAAP reporting. Table B.1 summarizes the various data sources of the starting values for the three different types of participating banks together with the exact items used from the reporting sheets.

Table B.1: Data sources for bank-specific starting values

Variable	**Single institute level** (all institutes except without parent waiver)		**Group level** (parent waivers with nGAAP)		**Group level** (parent waivers with IFRS)	
CET 1 capital RWA	Corep, single institute reporting	EC01.00 020.010 EC02.00 010.010	Corep, group reporting	QC01.00 020.010 QC02.00 010.010	Corep, group reporting	QC01.00 020.010 QC 02.00 010.010
Interestbearing assets	quarterly BBk reportings	HV11 040 + 050 + 060 + 070 + 080	FinaRisikoV	QV1 040 + 050 + 060 +070 + 080	Finrep, group reporting	F01.01 080 + 090 + 120 + 130 + 160 + 170 + 173 +174 + 177 + 178 + 190 + 200 + 220 + 230 + 232 + 233 + 236 + 237
Interest Income		GVKI 010		QGV 010		F02.00 010.010
Interest expenses		GVKI 020		QGV 020		F02.00 090.010
Impairments		SAKI 340		QSA1 340		F02.00 460.010
Profit	EGV	EGV 58	KGV	KGV 21		F 02.00 670.010

Table B.2: Data description of variables in macroeconomic stress scenarios

Variable	Description	Source
GER.GDP	real GDP Germany; quarterly; seasonally adjusted	St. Louis FED, `CLVMNACSCAB1GQDE`
GER.bond	long-term interest rate for convergence purposes Germany, unspecified rate type, debt security issued, 10 years maturity, new business coverage, denominated in euro, unspecified counterpart sector	ECB, `IRS.M.DE.L.L40.CI.0000.EUR.N.Z`
GER.CPI	German inflation measured by harmonized consumer price index (CPI), all items	Destatis, `61111-0001`
GER.CHPI	German construction price index for office and industrial buildings including turnover tax	Destatis, `61261-0002`
GER.RHPI	German residential property prices, new and existing flats; residential property in good & poor condition, whole country; neither seasonally nor working-day adjusted	ECB, `RPP.Q.DE.N.TF.00.5.00`
GER.unemp	German standardised unemployment rate, all ages, male and female, seasonally adjusted, not working day adjusted, percentage of civilian workforce	ECB, `STS.M.DE.S.UNEH.RTT000.4.000`
GER.DAX	German blue chip stock market index	Yahoo Finance, `GDAXI`
EU.EURIBOR	Euro area (changing composition) Euribor 1-year historical close rate, average of observations through period	ECB, `FM.M.U2.EUR.RT.MM.EURIBOR1YD_.HSTA`
EU.GDP	European Union (28 countries) GDP at 2010 reference levels, at constant prices	ECB, `AME.A.EU28.1.0.0.0.OVGD`
US.GDP	real GDP US; quarterly; seasonally adjusted	St. Louis FED, `GDPC1`

Notes: All data from 2008Q1 to 2017Q2.

Table B.3: Pairwise Correlation Matrix

	GDP	Bond	CPI	CHPI	RHPI	UE	Swap1Y	DAX	US-GDP
GDP	1.00								
Bond	0.41	1.00							
CPI	0.34	0.33	1.00						
CHPI	0.35	0.35	0.44	1.00					
RHPI	-0.02	-0.08	-0.27	-0.23	1.00				
UE	-0.55	-0.32	-0.37	-0.73	0.28	1.00			
Swap1Y	0.82	0.49	0.50	0.55	-0.22	-0.57	1.00		
DAX	0.36	0.22	-0.11	0.19	0.20	-0.17	0.21	1.00	
US-GDP	0.61	0.12	0.10	-0.06	0.11	0.13	0.54	0.48	1.00
EU-GDP	0.79	0.47	0.37	0.22	-0.14	-0.19	0.78	0.17	0.68

Notes: For a definition of the variables see Table B.2. All variables in year-on-year transformations (growth rates, except for "Bond" and "Swap1Y" which are absolute changes.

C Descriptive Statistics for Model Spaces and Model Confidence Sets

Table C.1: Fraction of Surviving Models in Unfiltered Model Space (%)

Sector	Multicoll.	Autocorr.	Sign Res.	Benchmarkt Constr.	Occams's Windows
1	71.7	47.3	0.9	12.8	0.1
2	63.5	92.8	11.9	69.6	0.6
3	63.7	98.1	8.4	81.6	0.0
4	66.9	21.0	5.5	14.6	0.2
5	93.9	44.0	57.2	71.7	0.1
6	64.7	51.5	21.1	63.2	0.4
7	76.6	30.5	0.7	2.3	0.3
8	68.4	99.9	0.4	11.6	0.1
9	62.8	93.6	26.2	88.4	0.7
10	75.3	94.0	0.2	10.8	0.0
11	69.0	94.5	5.0	23.2	0.1
12	78.3	35.4	9.3	11.8	0.1
13	76.5	15.3	34.0	73.3	0.2

Notes: All values expressed relative to models in unfiltered model space $\mathcal{M}$ except for Occam's Windows which is imposed once the previous filters have been applied. Number of models for each sector $Q = 10,000$.

Table C.2: $\mathcal{M}$- and Ω-implied PD increases in adverse scenario (%)

	$\mathcal{M}$				Ω			
Sector	Min	Median	Max	SD	Min	Median	Max	SD
1	-85.7	-36.6	7027.7	146.9	16.8	57.5	505.5	104.7
2	-97.3	133.5	4753.4	909.2	41.1	562.7	3652.6	862.0
3	-85.2	144.3	5066.9	543.2	70.9	204.6	938.6	137.2
4	-76.5	-52.8	6763.6	180.3	9.7	112.2	2025.1	201.4
5	-73.4	19.8	2131.6	110.5	0.3	24.1	340.8	33.0
6	-91.3	30.7	2237.0	200.8	11.8	143.9	1367.0	168.9
7	-98.9	-84.6	810.4	40.7	94.1	160.2	546.7	81.1
8	-98.6	-35.8	4172.0	346.1	160.2	360.3	637.3	155.5
9	-92.0	509.3	2880.1	705.9	71.5	978.0	2607.6	612.6
10	-99.9	-80.9	30463.1	1782.1	196.3	2862.8	13089.1	4520.9
11	-96.4	-29.6	9698.6	570.0	46.6	446.7	7520.8	1541.3
12	-93.3	-23.6	3049.7	62.9	0.3	16.0	67.9	14.1
13	-93.3	28.9	5567.6	509.8	1.8	72.1	4559.1	665.9

Notes: PD increase over the three years in adverse scenario. All values in as percentages Number of models for each sector $Q = 10,000$.

D Multivarite QMap: Exemplary Output

This section shows some additional output of the multivariate QMap model for sector 5. Figure D.1 shows the fitted and the observed sectoral PD time series. Figure D.2 shows the model weights of all models that are candidates to be included in the MA-QMap and Figure D.3 shows the MA-QMap-implied sectoral PD predictions conditional on the baseline and the adverse macroeconomic scenario.

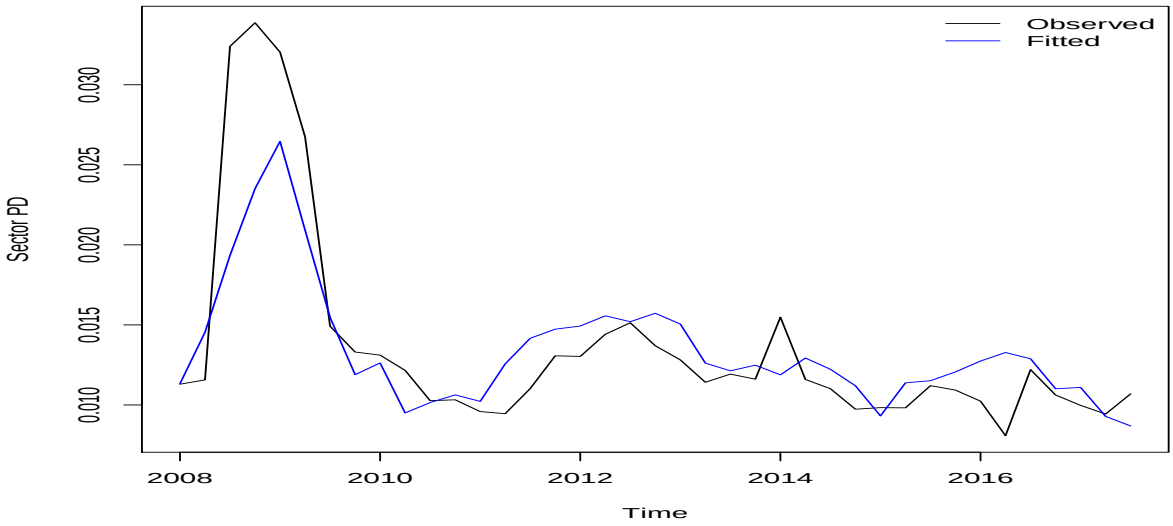


Figure D.1: Multivariate QMap - Fitted VS Observed Values: Sector 5

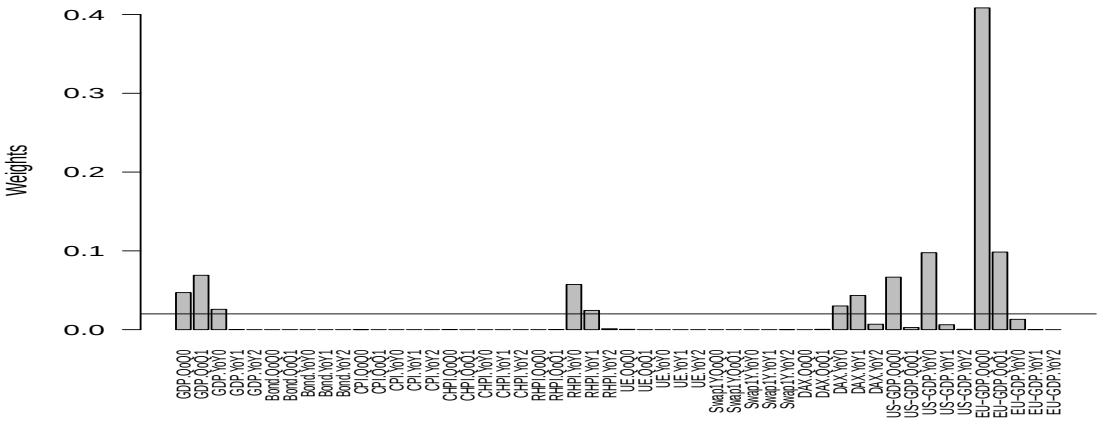


Figure D.2: Multivariate QMap - Covariate Weights: Sector 5

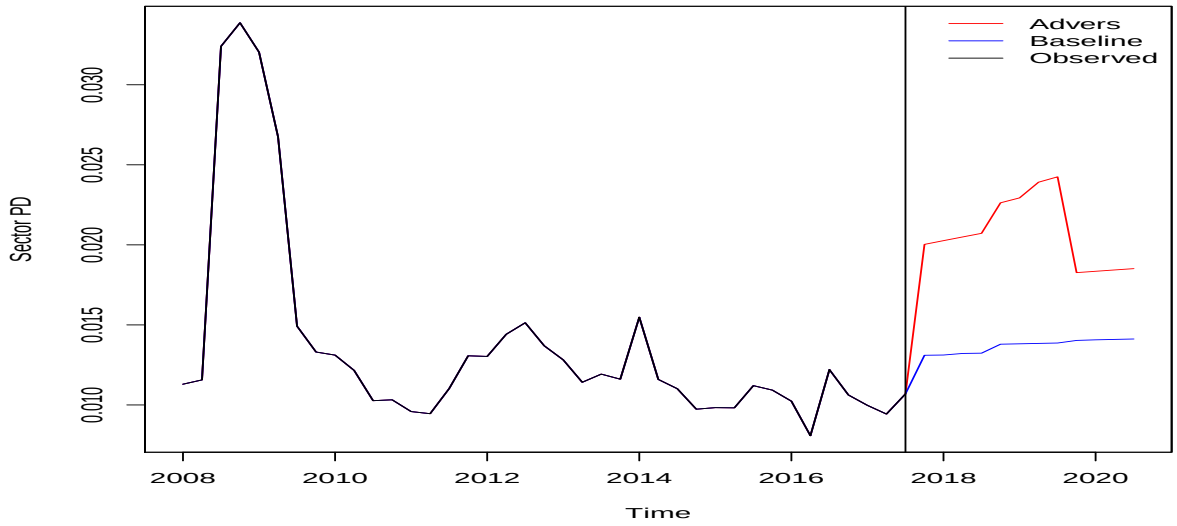


Figure D.3: Multivariate QMap - Forecast: Sector 5

E BCBMA Model: Exemplary Output

This section shows some additional BCBMA model output for Sector 5. Figure E.1 shows the posterior inclusion probabilities of all variables included in the final BCBMA model. For the calculation of the posterior inclusion probabilities, see Equation (13). In the BMA context a variable is judged to be significant if the posterior inclusion probability exceeds the prior inclusion probability (see Equation (14)).

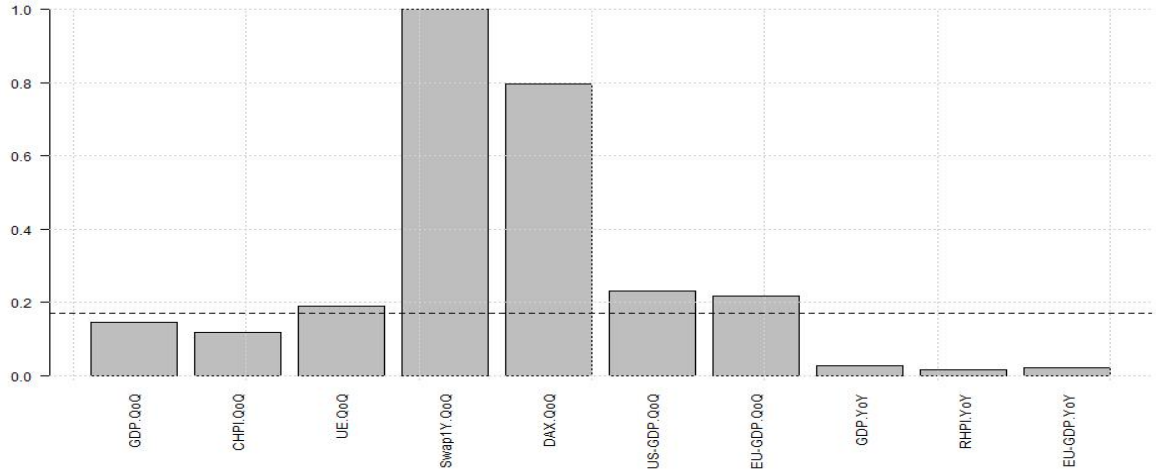


Figure E.1: BCBMA - Posterior Inclusion Probabilities: Sector 5

Figure E.2 shows the LRM of all models included in the final BCBMA model. The LRM (except for CPI and EURIBOR) are subject to the sign restrictions in Section 3.3.

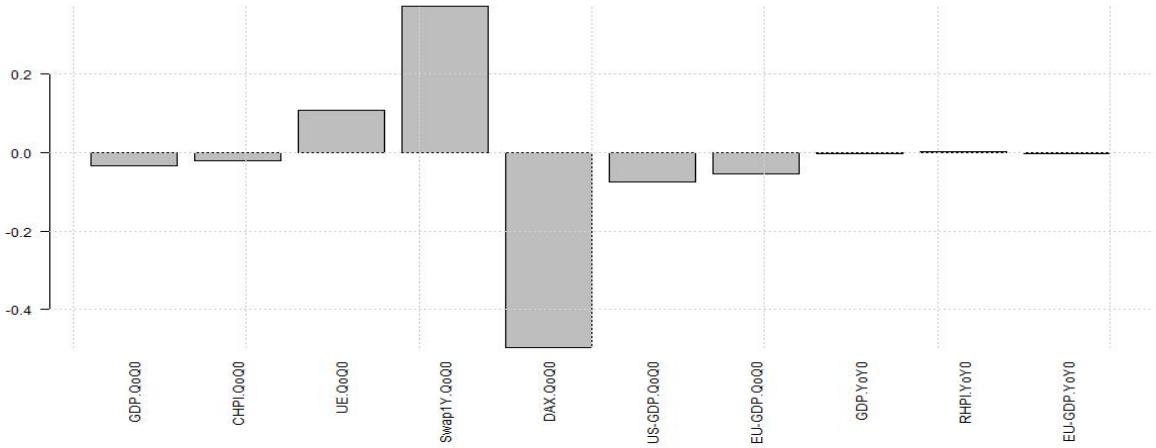


Figure E.2: BCBMA - Long-Run Multipliers: Sector 5

References

Avramov, D. (2002). Stock Return Predictability and Model Uncertainty. *Journal of Financial Economics 64*(3).

Baele, L., V. D. Bruyckere, O. D. Jonghe, and R. V. Vennet (2015). Model uncertainty and systematic risk in US banking. *Journal of Banking & Finance 53*, 49–66.

Bank of England (2017). Stress testing the UK banking system: 2017 guidance for participating banks and building societies.

Basel Committee on Banking Supervision (2005). An Explanatory Note on the Basel II IRB Risk Weight Functions.

Basel Committee on Banking Supervision (2015). Consultative document - Revisions to the Standardised Approach for credit risk. *Standards*.

Basel Committee on Banking Supervision (2016). Reducing variation in credit risk-weighted assets – constraints on the use of internal model approaches. *Consultative Document*.

Belsley, D. A., E. Kuh, and R. E. Welsch (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, Inc.

Board of Govenors of the Federal Reserve System (2016). Dodd-Frank Act Stress Test 2016: Supervisory Stress Test Methodology and Results.

Bolotny, V., R. M. Edge, and L. Guerrieri (2015). Stressing Bank Proftability for Interest Rate Risk. *Louvain School of Management Working Paper Series 2015/15*.

Bonti, G., M. Kalkbrener, C. Lotz, and G. Stahl (2006). Credit risk concentrations under stress. *Journal of Credit Risk 2*(3), 115–136.

Burnham, K. and D. Anderson (2002). Model Selection and Multimodel Inference . *Springer Science & Business Media*.

Castrén, O., S. Dées, and F. Zaher (2010). Stress-testing euro area corporate default probabilities using a global macroeconomic model. *Journal of Financial Stability 2*, 64–78.

Clerc, L., A. Derviz, C. Mendicino, S. Moyen, K. Nikolov, L. Stracca, J. Suarez, and A. P. Vardoulakis (2015). Capital Regulation in a Macroeconomic Model with Three Layers of Default. *International Journal of Central Banking 11*(3), 9–63.

Corbae, D., P. D'Erasmo, S. Galaasen, A. Irarrazabal, and T. Siemsen (2017). Structural stress tests. *University of Wisconsin Working Paper*.

Covas, F. B., B. Rump, and E. Zakrajšek (2014). Stress-testing US bank holding companies: A dynamic panel quantile regression approach. *International Journal of Forecasting 30*(3), 691–713.

Cremers, K. J. M. (2002). Stock Return Predictability: A Bayesian Model Selection Perspective. *The Review of Financial Studies 15*(4), 1223–1249.

Doppelhofer, G., R. I. Miller, and X. Sala-i-Martin (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review 94*, 813–835.

European Banking Authority (2013). Report on the pro-cyclicality of capital requirements under the Internal Ratings Based Approach.

European Banking Authority (2014a). Fourth report on the consistency of risk weighted assets - Residential mortgages drill-down analysis.

European Banking Authority (2014b). Guidelines on common procedures and methodologies for the supervisory review and evaluation process (SREP).

European Banking Authority (2014c). Guidelines on the range of scenarios to be used in recovery plans.

European Banking Authority (2016a). 2016 EU-Wide Stress Test Results.

European Banking Authority (2016b). Draft Guidelines on credit institutions' credit risk management practices and accounting for expected credit losses.

European Banking Authority (2017). 2018 EU-Wide Stress Test Methodological Note.

European Banking Authority (2018). Adverse macro-financial scenario for the 2018 eu-wide banking sector stress test. http://www.eba.europa.eu/documents/10180/2106649/Adverse+macroeconomic+scenario+for+the+EBA+2018+Stress+Test.pdf.

Financial Stability Board (2013). Recovery and Resolution Planning for Systemically Important Financial Institutions: Guidance on Recovery Triggers and Stress Scenarios.

Furnival, G. M. and R. W. Wilson (1974). Regressions by Leaps and Bounds . *Technometrics 16*(4), 499–511.

Garratt, A., J. Mitchell, S. P. Vahey, and E. C. Wakerly (2011). Real-time inflation forecast densities from ensemble Phillips curves. *North American Journal of Economics and Finance 22*, 77–87.

Gross, M. and J. Poblacion (2016). Implications of Model Uncertainty for Bank Stress Testing. *Journal of Financial Service Research*.

Gunst, R. and J. Webster (1975). Regression analysis and problems of multicollinearity. *Communications in Statistics 4*(3).

Gunst, R. F. and R. L. Mason (1977). Advantages of examining multicollinearities in regression analysis. *Biometrics 33*(1), 249–260.

Gupton, G. M., C. C. Finger, and M. Bhatia (1997). CreditMetrics-Technical Document. *J.P. Morgan & Co. Incorporated*.

Hansen, C., J. B. McDonald, and W. K. Newey (2010). Instrument Variables Estimation With Flexible Distributions. *Journal of Business & Economic Statistics 28*(1), 13–25.

Hansen, P. R., A. Lunde, and J. M. Nason (2011, March). The Model Confidence Set. *Econometrica 79*(2), 453–497.

Henry, J. and C. Kok (2013). A macro stress testing framework for assessing systemic risks in the banking sector. *Occasional Paper Series, No 152, ECB*.

Kullback, S. and R. Leibler (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics 22*(1), 79–86.

Madian, D. and A. Raferty (1994). Model Selection and Accounting for Model Uncertainty in Graphical Models using Occam's Windows. *Journal of American Statistical Association 89*, 1535–1546.

Misina, M. and D. Tessier (2008). Non-Linearities, Model Uncertainty, and Macro Stress Testing. *Bank of Canada, Working Paper No. 2008-30*.

Pelster, M. and J. Vilsmeier (2017). The determinants of CDS spreads: evidence from the model space. *Review of Derivatives Research (forthcoming)*.

Rösch, D. and H. Scheule (2009). Credit Portfolio Loss Forecasts for Economic Downturns. *Financial Markets, Institutions & Instruments 18*(1), 1–26.

Samuels, J. D. and R. M. Sekkel (2012). Forecasting with Large Datasets: Trimming Predictors and Forecast Combination. *Working Paper*.

Schechtman, R. and W. P. Gaglianone (2012). Macro stress testing of credit risk focused on the tails. *Journal of Financial Stability 8*(3), 174–192.

Schuermann, T. (2014). Stress testing banks. *International Journal of Forecasting 30*(3), 717–728.

Siemsen, T. and J. Vilsmeier (2017). A stress test framework for the German residential mortgage market – methodology and application. *Deutsche Bundesbank Discussion Paper 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), 113–151.

Stock, J. H. and M. W. Watson (2005). An Empirical Comparison of Methods for Forcasting using many Predictors. *Working Paper*.

Theodossiou, P. (1998). Financial Data and the Skewed Generalized T Distribution. *Management Science 44*, 1650–1661.

Vasicek, O. (2002, December). Loan Portfolio Value. *Risk*, 160–162.

Vazquez, F., B. Tabak, and M. Souto (2012). A macro stress test model of credit risk for the Brazilian banking sector. *Journal of Financial Stability 8*(2), 69–83.

Wilson, T. (1997a). Credit portfolio risk: Part I. *Risk*.

Wilson, T. (1997b). Credit portfolio risk: Part II. *Risk*.

Wright, J. H. (2008). Bayesian Model Averaging and exchange rate forecasts. *Journal of Econometrics 146*(2), 329–341.