DEUTSCHE BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 09/2019

Model and estimation risk in credit risk stress tests

Peter Grundke
(Osnabrück University)

Kamil Pliszka
(Deutsche Bundesbank)

Michael Tuchscherer
(Osnabrück University)

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 978–3–95729–567–5 (Printversion)
ISBN 978–3–95729–568–2 (Internetversion)

Non-technical summary

Research question

This paper deals with stress tests for credit risk and shows how exploiting the discretion when setting up and implementing the underlying model can drive the results of a quantitative credit risk stress test for default probabilities.

Contribution

We contribute to the scarce literature on model and estimation risk in stress tests. We employ several variations of a CreditPortfolioView-style model using US data ranging from 2004 to 2016 and compare the forecasted default probabilities of these models. Our clear focus on stress tests is the aspect that differentiates our paper from existing studies most. This is particularly relevant against the background of regulatory stress tests which have become more important in recent years.

Results and policy implications

This paper shows that stress forecasts of default probabilities highly depend on the modelling assumptions and that seemingly only minor variations can affect the results of stress tests considerably. That said, our findings reveal that the conversion of a shock (i.e., stress event) increases the (non-stress) default probability by 20% to 80% - this high range can be explained by the sensitivity of stress test models to model and estimation risk. Interestingly, forecasts for non-stress default probabilities are less exposed to model and estimation risk. In addition, the risk horizon over which the stress default probabilities are forecasted and whether we consider mean stress default probabilities or high quantiles seem to play only a minor role for the dispersion between the results of the different model specifications. These findings emphasize the importance of extensive robustness checks for model-based credit risk stress tests, particularly in regulatory stress tests.

Nichttechnische Zusammenfassung

Fragestellung

Dieses Forschungspapier untersucht Modellrisiken bei Stresstests für Kreditrisiken. Es zeigt auf, wie sich der vorhandene Gestaltungsspielraum bei der Durchführung und Implementierung des Modells auf die Ergebnisse von quantitativen Kreditrisiko-Stresstests für Ausfallwahrscheinlichkeiten auswirkt.

Beitrag

Wir erweitern die Literatur zu Modell- und Schätzrisiken in Stresstests. Wir verwenden verschiedene Spezifikationen des Kreditrisikomodells CreditPortfolioView unter Nutzung von Daten für den US-amerikanischen Markt im Zeitraum von 2004 bis 2016 und vergleichen die Spezifikationen hinsichtlich der prognostizierten Ausfallwahrscheinlichkeiten. Unser Schwerpunkt auf Stresstests grenzt unsere Analyse von bisherigen Studien ab; das ist vor allem vor dem Hintergrund der gestiegenen Bedeutung von regulatorischen Stresstests in den letzten Jahren relevant.

Ergebnisse und Politikempfehlungen

Unsere Ergebnisse zeigen, dass Prognosen für gestresste Ausfallwahrscheinlichkeiten stark von den Modellierungsannahmen abhängen und dass sich bereits geringe Modelländerungen stark auf die Ergebnisse von Stresstests auswirken können. Konkret bedeutet das, dass die Berücksichtigung eines Schocks (Stressfall) zu einer Erhöhung der Ausfallwahrscheinlichkeit um 20% bis 80% führen kann – diese große Spannweite erklärt sich durch die hohe Sensitivität von Stresstestmodellen hinsichtlich Modell- und Schätzrisiken. Im Gegensatz dazu zeigt sich, dass nicht gestresste Ausfallwahrscheinlichen in geringerem Maße Modell- und Schätzrisiken ausgesetzt sind. Darüber hinaus spielen die Länge des Risikohorizonts, über den hinweg die Prognose der gestressten Ausfallwahrscheinlichkeiten erfolgt, und die Frage, ob mittlere gestresste Ausfallwahrscheinlichkeiten oder hohe Quantile betrachtet werden, nur eine untergeordnete Rolle für die Unterschiede zwischen den Ergebnissen der einzelnen Modellspezifikationen. Diese Resultate machen deutlich, dass aufwändige Robustheitsüberprüfungen für modellbasierte Stresstests erforderlich sind – vor allem in regulatorischen Stresstests.

Model and estimation risk in credit risk stress tests*

Peter Grundke
Osnabrück University

Kamil Pliszka
Deutsche Bundesbank

Michael Tuchscherer
Osnabrück University

Abstract

This paper deals with stress tests for credit risk and shows how exploiting the discretion when setting up and implementing a model can drive the results of a quantitative stress test for default probabilities. For this purpose, we employ several variations of a CreditPortfolioView-style model using US data ranging from 2004 to 2016. We show that seemingly only slightly differing specifications can lead to entirely different stress test results – in relative and absolute terms. That said, our findings reveal that the conversion of a shock (i.e., stress event) increases the (non-stress) default probability by 20% to 80% - depending on the stress test model selected. Interestingly, forecasts for non-stress default probabilities are less exposed to model and estimation risk. In addition, the risk horizon over which the stress default probabilities are forecasted and whether we consider mean stress default probabilities or quantiles seem to play only a minor role for the dispersion between the results of the different model specifications. Our findings emphasize the importance of extensive robustness checks for model-based credit risk stress tests.

Keywords: credit risk, default probability, estimation risk, model risk, stress tests

JEL-Classification: G21, G28, G32

*Contact addresses: Peter Grundke: Osnabrück University, Chair of Banking and Finance, Katharinenstraße 7, 49074 Osnabrück, Germany, E-mail: peter.grundke@uni-osnabrueck.de.
Kamil Pliszka: Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main, Germany, E-mail: kamil.pliszka@bundesbank.de.
Michael Tuchscherer: Osnabrück University, Chair of Banking and Finance, Katharinenstraße 7, 49074 Osnabrück, Germany, E-mail: michael.tuchscherer@uni-osnabrueck.de.
We thank the participants of the Bundesbank Seminar, the seminars at the Universities of Osnabrück and Düsseldorf, the OR conference (Aachen, 2014), the FEBS conference (Nantes, 2015), the IRMC (Luxembourg, 2015), the EFMA conference (Amsterdam, 2015), the DGF conference (Leipzig, 2015), the CREDIT conference (Venice, 2015) and the SFA conference (Captiva Island, 2015). The views expressed in this paper are those of the authors and do not necessarily coincide with the views of the Deutsche Bundesbank or the Eurosystem.

1 Introduction

Banks are often required to translate the impact of an economic shock onto its risk parameters. Examples include the Basel II credit risk framework where IRB banks have to reflect economic downturns in their risk parameters in Pillar 1 (see article 177 CRR) or the CEBS' guidelines on stress testing (see CEBS (2010, p. 18)) which require banks to consider a severe economic downturn for their internal risk coverage calculations under Pillar 2. More topical examples are the EBA stress tests in 2014, 2016 and 2018 in the euro area, where banks either could translate a prescribed economic downturn scenario into their risk parameters or could directly employ the parameter values provided by the EBA.[1] As failed internal or external stress tests may force a bank to increase its equity and banks usually consider equity to be expensive,[2] banks at least have an incentive to employ those modelling and estimation techniques that yield the stress test results that are most favourable for them. Up to now, there is no empirical evidence whether banks use this discretion to their favour or not when setting up and implementing a stress test model. However, there are some empirical hints that banks use the degrees of freedom within internal ratings-based approaches in such a way that the modelled default probabilities are partly below historical default rates (see BCBS (2014) and Behn et al. (2016)). Thus, at the current stage of research, it at least cannot be excluded that the same effect could be observed in the context of model-based stress tests.

In this paper, we focus on a specific risk type (credit risk) and a specific risk parameter (probability of default, PD) and empirically analyze to which extent multi-period stress PD values can vary depending on the employed modelling assumptions and estimation techniques. To achieve this, we employ several variations of a CreditPortfolioView (CPV)-style model[3] using US data for the period 07/2004 to 08/2016. All variations are statistically sound approaches employed by practitioners and in related literature and it is ex-ante not obvious why one specification or estimation technique should be more adequate than another. Our out-of-sample forecast ability comparison of the specifications also shows that no single specification is dominating the other ones.

[1] For the macro stress tests performed by the EBA, this is exactly what banks had to do (unless they wanted to employ EBA's benchmark PD and LGD values). The corresponding forecasts of the EU commission for a risk horizon of two to three years are employed as the economic baseline and adverse scenario (see EBA (2014), ECB (2014), EBA (2016), EBA (2018a)).

[2] See Admati and Hellwig (2013) for an extensive discussion of supposedly expensive bank equity.

[3] See Wilson (1997a, 1997b).

We show that the chosen model specification and the employed estimation technique can hugely influence the results for the stress default probabilities. Accordingly, the conversion of a shock, i.e., moving from non-stress to stress PDs, exposes banks in relative and absolute terms to model and estimation risk. More specifically, the conversion of a shock (i.e., stress event) increases the (non-stress) default probability by 20% to 80% - depending on the stress test model selected. This dispersion of results shows the importance of extensive robustness checks for the underlying model when interpreting the results of credit risk stress tests. Interestingly, forecasting non-stress PDs is less exposed to model and estimation risk. In addition, the risk horizon over which the stress default probabilities are forecasted and whether we consider mean stress default probabilities or quantiles seem to play only a minor role for the dispersion between the different model specifications.

Our paper contributes to various strands in the literature. First, it is related to statistical approaches for the prediction of default probabilities (see, e.g., the recent papers of Blöchlinger and Leippold (2018), and Jones et al. (2016) as well as the references cited therein). Having models that transform firm-level or macroeconomic predictor variables in forecasts for default probabilities is a necessary prerequisite for doing model-based credit risk stress tests. Second, our study is most closely related to that strand of literature in which CPV-style models (or extensions thereof)[4] are used for carrying out a model-based credit risk stress test. These papers look for macroeconomic variables that can explain the systematic variation of default rates across time and, afterwards, these macroeconomic variables are shocked to compute stress default rates (see, for example, Boss (2002), Sorge and Virolainen (2006), Jokivuolle et al. (2008)). In some cases, feedback effects between the performance of the banking sector and the real economy are considered in these papers (see, for example, Virolainen (2004), Wong et al. (2008)). As an alternative to CPV-style econometric stress test approaches, Schechtman and Gaglianone (2012) apply quantile regressions to estimate the link between macroeconomic variables and credit risk. A systematic analysis of how different modelling assumptions and estimation techniques may influence the stress test results is usually not (or only in a limited way) done in these papers. The fact that this is the clear focus of our paper is an essential difference between our study and the previously mentioned ones. Third, our paper is obviously related to the literature on model risk in risk models. Examples are Danielsson et al. (2016) who evaluate the model risk of models employed for forecasting systemic and market risk, Frey and McNeil (2003) and Hamerle and Rösch (2006) who analyze the model risk of credit

[4] For a more detailed survey on quantitative credit risk stress test methodologies see, for example, Foglia (2009).

portfolio models, and Hayden et al. (2014) who evaluate the influence of the chosen variable selection approach on model-based default probability predictions. Surprisingly, the literature in which credit risk stress tests and the aspect of model risk are combined (as we do) is rather scarce. One notable exception is Hale et al. (2015) who analyze the influence of the aggregation level on the results of macroeconomic credit risk stress tests. Another related paper is Canals-Cerdá and Kerr (2015) who empirically study issues of model specification, sample selection and stress scenario selection for credit card portfolios. With respect to model risk, they focus on the interplay between macroeconomic and account-level variables.[5]

The remainder of the paper is structured as follows: Section 2 presents the methodology of the analysis and Section 3 shows the results. Section 4 concludes.

2 Methodology

In the following, first, we introduce the baseline specification of a CreditPortfolioView-style model for predicting stress default probabilities. Amongst several others, these models are widespread in German savings banks (see S-Rating und Risikosysteme (2018)). Second, various modifications of this specification are described. All modifications are statistically sound, and it is ex-ante not obvious why one specification should be more adequate than another. However, as we show in Section 3.2, the modifications can hugely influence the results for the stress default probabilities.

2.1 CreditPortfolioView-style baseline specification and PD forecasts

For all our specifications, we employ a CPV-style approach that relates macroeconomic variables to sector-specific default rates. The macroeconomic variables are chosen in such a way that they explain a large fraction of the time series variation in default rates. More precisely, it is assumed that for each sector s, $s \in \{1,2,...,S\}$, a macroeconomic index in period t

$$y_{s,t} = \beta_{s,0} + \sum_{i=1}^{I}\sum_{k=0}^{K_i}\beta_{s,i} \cdot x_{i,t-k} + \sum_{k=1}^{K_y}\delta_{s,k} \cdot y_{s,t-k} + u_{s,t} \tag{1}$$

linearly depends on some contemporaneous and/or time-lagged risk factors $x_{i,t-k}$, $i \in \{1,2,...,I\}$ and $k \in \{0,1,2,...,K_i\}$, and time-lagged values of the macroeconomic index

[5] A further recent exception is Siemsen and Vilsmeier (2018) who focussed in parallel but mutually unknown work on a similar topic as we.

$y_{s,t-k}$, $k \in \{1,2,...,K_y\}$. The macroeconomic index $y_{s,t}$ is assumed to be related to the sector-specific default probability $PD_{s,t}$ by a logit transformation:

$$y_{s,t} = \ln\left(\frac{1}{PD_{s,t}} - 1\right) \Leftrightarrow PD_{s,t} = \frac{1}{1+\exp\left(y_{s,t}\right)}. \quad (2)$$

Hence, larger values of the macroeconomic index $y_{s,t}$ go along with smaller default probabilities $PD_{s,t}$. The risk factors $x_{i,t}$, $i \in \{1,2,...,I\}$, are modelled by autoregressive processes of k_i-th order (AR(k_i) process):

$$x_{i,t} = \gamma_{i,0} + \sum_{j=1}^{k_i} \gamma_{i,j} \cdot x_{i,t-j} + v_{i,t}. \quad (3)$$

To avoid overfitting, we restrict our search for an adequate time series model to AR(k) processes with a maximum order of $k = 2$. We apply the AIC (Akaike Information Criterion) to choose the appropriate number of lags.

The ordinary least square (OLS) estimator is used to determine the parameters of equation (1) and (3). When the Godfrey-Breusch test indicates that the null hypothesis of no autocorrelation (up to order four) of the error term $v_{i,t}$ and $u_{s,t}$, respectively, can be rejected at a significance level of 5%, the Newey-West estimator is employed to compute the t-statistics and, hence, the p-values of the OLS parameter estimates.[6]

The error terms $u \in \mathbb{R}^{S\times 1}$ and $v \in \mathbb{R}^{I\times 1}$ are assumed to be multivariately normally distributed:[7]

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim N(0,\Sigma) \quad (4)$$

with $0 \in \mathbb{R}^{(S+I)\times 1}$ and

$$\Sigma = \begin{pmatrix} \Sigma_{u,u} & 0 \\ 0 & \Sigma_{v,v} \end{pmatrix} \in \mathbb{R}^{(S+I)\times(S+I)} \quad (5)$$

with $\Sigma_{u,u} \in \mathbb{R}^{S\times S}$, $\Sigma_{v,v} \in \mathbb{R}^{I\times I}$.

[6] For the Newey-West estimations, AR processes with a varying order were employed for capturing the autocorrelation in the error term. However, the coefficient values and significances were relatively stable across the varying orders. Hence, we abandoned higher orders and assumed an order of two for the Newey-West estimations.

[7] The assumed multivariate distribution of the error terms influences the probability distributions of the stress default probabilities. Alternatively, bootstrapping or another distribution could be used. See, for example, Simons and Rolwes (2009), who model the error terms of the index equations as well as the error terms of the risk factor equations by a t-distribution.

Combining (1) to (5), the distribution of the sector-specific default probabilities for the next m time periods (starting from period T) can be computed using the following Monte-Carlo simulation algorithm with D simulation runs:[8]

For $d = 1$ to D

For $n = T + 1$ to $T + m$

(i) Draw random numbers for the error terms $u_{s,n}^{(d)}$, $s \in \{1, 2, ..., S\}$, and $v_{i,n}^{(d)}$, $i \in \{1, 2, ..., I\}$, according to the multivariate normal distribution (4) and (5).

(ii) Calculate forecasts for the macroeconomic variables $x_{i,n}^{(d)}$, $i \in \{1, 2, ..., I\}$, based on $v_{i,n}^{(d)}$ and the historical realizations $x_{i,n-1}^{(d)}$, $x_{i,n-2}^{(d)}$, …, $x_{i,n-k_i}^{(d)}$.

(iii) Calculate forecasts for the sector-specific macroeconomic indices $y_{s,n}^{(d)}$ and default probabilities $PD_{s,n}^{(d)}$, $s \in \{1, 2, ..., S\}$, based on $u_{s,n}^{(d)}$ and the forecasts for the macroeconomic variables $x_{i,n}^{(d)}$.

Based on the realizations $PD_{s,n}^{(d)}$, $d \in \{1, ..., D\}$, we calculate empirical distribution functions for the sector-specific and time period-specific default probabilities $PD_{s,n}$, $s \in \{1, 2, ..., S\}$, $n \in \{T + 1, ..., T + m\}$.

To compute distributions for *stress* sector-specific and time period-specific default probabilities, the algorithm has to be amended slightly. Instead of using the unconditional multivariate normal distribution (4) and (5) in step (i), those error terms that are not stressed have to be sampled from a multivariate normal distribution that is conditioned on the stressed values of the other error terms. If Y is an r-dimensional normally distributed random vector with the following partitioning:[9]

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \text{ with } Y_1 \text{ a } q\text{-dimensional random vector } (q < r),$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \ \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

[8] See Boss (2002, pp. 81-82).
[9] See Greene (2008, pp. 1013-1014).

with $\Sigma_{11} \in \mathbb{R}^{q \times q}$ and $\Sigma_{22} \in \mathbb{R}^{(r-q)\times(r-q)}$, respectively, symmetric positive semidefinite matrices, $\det(\Sigma_{22}) \neq 0$, and $\Sigma_{12} = \Sigma'_{21} \in \mathbb{R}^{q\times(r-q)}$, then the conditional distribution of Y_1 given $Y_2 = y_2$ is a multivariate normal distribution with mean

$$\mu_1\big|_{Y_2=y_2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2) \tag{6}$$

and variance-covariance matrix

$$\Sigma\big|_{Y_2=y_2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \tag{7}$$

In the baseline setting, at any one time only one risk factor is initially shocked for the first three months and the shock is set equal to that historical realization of the error term which had the most negative impact on the macroeconomic index in the past.[10] More precisely, we define the shocked component $Y_2 = v*_{i,T+1} = v*_{i,T+2} = v*_{i,T+3}$ by

$$v*_{i,T+1} = v*_{i,T+2} = v*_{i,T+3} = \begin{cases} \min\limits_{t\in\{1,2,\ldots,T\}} v_{i,t}, & \beta_{s,i} > 0 \\ \max\limits_{t\in\{1,2,\ldots,T\}} v_{i,t}. & \beta_{s,i} < 0 \end{cases} \tag{8}$$

If $S = 1$ (what we assume in the following), the above definition is unambiguous. When, however, we have several sectors $S > 1$ and the sensitivities $\beta_{s,i}$ have different signs, additional criteria have to be introduced to decide whether the largest or smallest historical realization of the standardized error term is chosen. In the following, we set $m = 36$ months and we nearly always[11] assume that there is a univariate shock in the first future quarter and that in the subsequent 33 periods, all error terms are drawn from the unconditional multivariate normal distribution (4) and (5). However, of course, the initial shock propagates into the next periods according to the employed AR processes.[12] To achieve high accuracy in the Monte-Carlo simulation, we employ $D = 1{,}000{,}000$ draws.

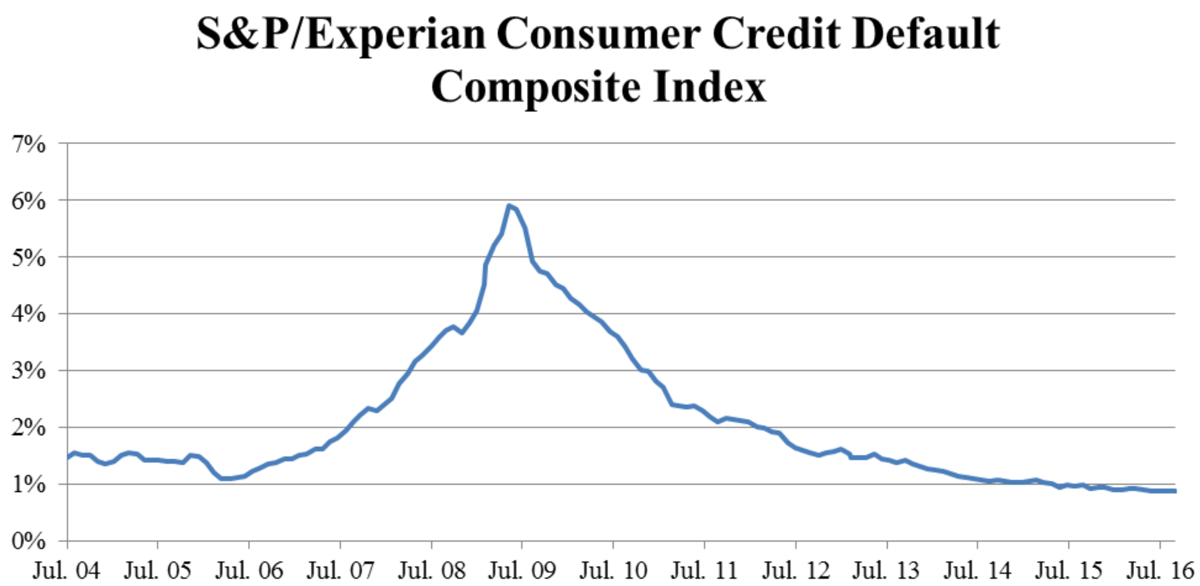
2.2 Data and variable selection

We use monthly S&P/Experian Consumer Credit Default Composite Index data ranging from 07/2004 to 08/2016 for estimating (1) (see Figure 1).[13] This index is a combination of default rates for cars, first and second mortgages and bank cards, and considers 280 Mio. US con-

[10] See Boss (2002, pp. 82-83).
[11] The exception is model 12 where the stress scenario is based on the Mahalanobis distance (see Section 2.3.3).
[12] Due to the correlation of the risk factors, those risk factors that are not explicitly stressed are influenced by the stress realization of the remaining risk factor and this influence propagates into the next periods according to the AR processes employed for modelling the remaining risk factors.
[13] See S&P Dow Jones Indices (2018). This data set has also been used, for example, by Fenech et al. (2015). As an alternative to the S&P/Experian Consumer Credit Default Composite Index, default rates provided by rating agencies, insolvency rates or the fraction of non-performing loans (NPLs) to all loans could be used.

sumers for over 11 trillion USD in loans, credit lines and leases. The S&P/Experian Consumer Credit Default Composite Index is calculated as the sum of all balances that newly defaulted in the last three months divided by the sum of all open good balances and those balances that newly defaulted in the last three months multiplied by twelve months. The definition of default varies by product types: conventional loans default if they are 90 days past due (or worse); unspecified and revolving products default if they are 180 days past due (or worse). In addition, bankruptcy, repossession and a write-off are treated as default. Seasonal variations of the time series are eliminated by using X-13ARIMA-SEATS.[14] The S&P/Experian Consumer Credit Default Composite Index encompasses different subcategories (e.g., first mortgages and bank cards), but we subsume all in one sector and set $S = 1$.

Figure 1: Evolution of the S&P/Experian Consumer Credit Default Composite Index over time



The S&P/Experian Consumer Credit Default Composite Index is a combination of default rates for cars, first and second mortgages and bank cards, and considers 280 Mio. US consumers for over 11 trillion USD in loans, credit lines, and leases. The index is calculated as the sum of all balances that newly defaulted in the last three months divided by the sum of all open good balances and those balances that newly defaulted in the last three months multiplied by twelve months.

Since the S&P/Experian Consumer Credit Default Composite Index is a combination of defaulted balances from the last three months, only risk factors with two periods delay are considered in the variable selection process in order to avoid potential endogeneity issues. As in Kalrai and Schleicher (2002, pp. 71-75) for Austrian data, economic activity indicators, price

[14] See U.S. Census Bureau (2017). SEATS is the acronym for "Signal Extraction in ARIMA Time Series". We use the seas package in R.

stability indicators, household indicators, firm indicators, financial market indicators and further external indicators for the US are considered to be potential explanatory variables for the default rates (see Table 1). The data are taken from Datastream.

From the comprehensive set of candidate explanatory variables, the most relevant ones explaining historical default rates have to be chosen. Some studies select relevant risk factors based on expert judgement and, afterwards, ensure that the chosen variables are (jointly) significant. In these studies, an economic indicator (e.g., GDP) and an interest rate are often employed.[15] To limit ad-hoc elements in the selection procedure for the explanatory variables, we apply the Bayesian model averaging (BMA)[16] where we include only risk factors with a sufficient high likelihood. Simulations and empirical studies show that the BMA delivers a better forecast performance than other approaches which makes this technique popular (see, e.g., Hayden et al. (2014), Raftery et al. (1997) and Traczynski (2017)).

[15] See, for example, Banque de France (2009) or Sorge and Virolainen (2006).

[16] For a robustness check of the selected risk factors, backward regression with robust (Huber-White) standard errors is also used. A detailed description of this approach is provided, for example, in Rawlings et al. (1998, pp. 218-219). For a discussion of alternative variable selection procedures for logistic credit risk models, see Hayden et al. (2014).

Table 1: Descriptive statistics of the endogenous and exogenous variables

	Mean	Std	Max	Min	Data source	Unit
Endogenous variable						
S&P/Experian Consumer Credit Default Composite Index	2.09	1.24	5.51	0.81	S&P Dow Jones Indices	%
Index (logit)	3.99	0.54	4.81	2.84	-	-
Index (probit)	2.09	0.22	2.41	1.60	-	-
Exogenous variables						
Economic activity indicators						
Industrial production	100.32	4.70	106.69	87.41	Datastream: USIPTOT.G	index
Price stability indicators						
Inflation	219.27	14.97	240.30	189.10	Datastream: USCOPRCE	index
Money supply M1	902.19	230.66	1378.40	642.64	Datastream: USM1....B	billion USD
Money supply M3	4093.73	629.36	5413.29	3304.63	Datastream: USMA013B	billion USD
Moody's commodity index	2059.91	615.24	3320.20	1044.80	Datastream: MOCMDTY	USD per points
Reuter's commodity index	658.59	122.01	932.18	426.24	Datastream: RECMDTY	USD per points
Household indicators						
Disposable personal income	5234.21	285.83	5857.53	4756.96	Datastream: USPERDISB	billion USD
New home sales	600.18	320.34	1389.00	270.00	Datastream: USHOUSSE	thousand
Unemployment rate	6.64	1.84	10.00	4.40	Datastream: USUN%TOTQ	%
Firm indicators						
Consumer confidence	79.22	22.24	111.90	25.30	Datastream: USCNFCONQ	index
Consumer sentiment	79.87	11.15	98.10	55.30	Datastream: USUMCONSH	index
Financial market indicators						
3-month Treasury bill rate	1.25	1.75	5.01	-0.01	US Department of the Treasury	%
Term spread (10-year minus 1-year Treasury bill rate)	1.67	1.03	3.43	-0.48	US Department of the Treasury	%
S&P 500	653.87	131.32	906.38	345.59	Datastream: S&PCOMP	USD
VIX	19.55	8.69	59.89	10.42	Datastream: CBOEVIX	index
External indicators						
Exports	108246.93	21121.17	137512.00	67645.80	Datastream: USEXPGDSB	million USD
Imports	168657.45	22869.28	199284.00	118736.00	Datastream: USIMPGDSB	million USD
USD/JPY exchange rate	0.0046	0.0006	0.0057	0.0034	Datastream: JPXRUSD.	USD
USD/GBP exchange rate	0.27	0.02	0.33	0.23	Datastream: STUSBOE	USD
Oil price WTI (FOB) per Barrel	34.48	9.73	64.49	14.15	Datastream: OILWTXI	USD

The idea of the Bayesian model averaging is to calculate for a given number O of candidate risk factors, in our case 20 variables as shown in Table 1, all linear models M_l, $l \in \{1,...,2^{20}\}$ consisting of subsets of the risk factors and, then, to include only those which prove to be sufficiently likely. The criterion for including a risk factor is the posterior inclusion probability

(PIP) which is given for any component β_h of the parameter vector β_{BMA} as a weighted sum of each model's conditional probability over all models:

$$PIP := P\left(\beta_h \middle| y\right) = \sum_{l=1}^{2^{20}} P\left(\beta_h \middle| M_l\right) \cdot P\left(M_l \middle| y\right) \tag{9}$$

where $y = (y_1, ..., y_T)$ denotes the vector of realizations of the macroeconomic index.

We follow the suggestion of Raftery (1995) of including only risk factors with a PIP of at least 50%.[17] Obtaining a risk factor's conditional inclusion probability $P\left(\beta_h \middle| M_l\right)$ is straightforward as it can be taken from the p-values of the corresponding model. The conditional marginal likelihood $P\left(M_l \middle| y\right)$ is according to Bayes theorem proportional to the product of the conditional distribution of y and a so-called model prior $P\left(M_l\right)$:

$$P\left(M_l \middle| y\right) \propto P\left(y \middle| M_l\right) \cdot P\left(M_l\right). \tag{10}$$

As the priors are initially unknown, commonly g priors (see Zellner (1986)) are assumed for the models' coefficients:

$$\beta \middle| g \sim N\left(0, \left(\frac{1}{g}\Gamma'\Gamma\right)^{-1}\right) \tag{11}$$

where the matrix $\Gamma \in \mathbb{R}^{T \times O}$ contains all T historical observations for the O candidate risk factors. The parameter g allows for considering the degree of certainty, i.e., a smaller value of the parameter goes along with a lower variance. The marginal likelihood is given by:

$$P\left(y \middle| M_l\right) \propto (1+g)^{-\frac{o_l}{2}} \cdot \left(1 - \frac{g}{1+g} \cdot R_l^2\right)^{-\frac{T-1}{2}} \tag{12}$$

where o_l denotes the number of included risk factors in model M_l. It is obvious that this term basically weighs up the goodness-of-fit as measured by model l's coefficient of determination R_l^2 and the term $(1+g)^{o_l}$ for penalizing for the model size. In order to set the parameter g, we apply the popular unit information prior (UIP) which sets $g = T$.[18]

Evaluating all models $P\left(M_l \middle| y\right)$, $l \in \{1, ..., 2^{20}\}$, which means that in our case we would have to conduct over one million regressions, often proves to be computationally too intricate. In

[17] If the PIP is slightly below 50%, we include these variables if they prove to be significant in the regression analysis for equation (1).
[18] Eicher et al. (2011) conclude that the UIP delivers the best performance. Suitable alternative choices would have been $g = \max\{t; K^2\}$ and $g = K^2$ (see, e.g., Fernandez et al. (2001), Feldkircher and Zeugner (2009)).

order to overcome this issue, we employ the Markov chain Monte-Carlo sampler (see, e.g., Madigan and York (1995)).

To test for stationarity of the time series of the macroeconomic index and of the explanatory variables, we apply the Augmented Dickey-Fuller (ADF), Phillips-Perron (PP) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. A time trend is only considered within these tests when it is economically plausible. As the results of these three tests are partly conflicting, we assume stationarity when at least two out of three tests indicate stationarity (null hypothesis of non-stationarity is rejected by the ADF or PP test; null hypothesis of stationarity is not rejected by the KPSS test). For all three tests, the significance level is 10%. We either take the logarithmic-return for exponentially increasing time series or the first difference for time series moving within a limited range. The latter method is employed for the macroeconomic index, new home sales, unemployment rate, VIX, 3-month Treasury bill rate, term spread, USD/JPY and USD/GBP exchange rates, and oil price WTI; the former for the other variables. All transformed time series are stationary. As some risk factors might be prone to multicollinearity, we calculate the variance inflation factor.

2.3 Modifications

Having implemented a reasonable specification for the modelling of the relationship between macroeconomic variables and the default probability (see (1) to (5)), we want to test how economically equally reasonable modifications influence the results for the stress default probabilities. The modifications are obtained from literature on CPV-style models. In addition, we included modifications from other areas if they constituted a technically more accurate approach (e.g., FGLS estimator). Any discrepancies in predictions of default probabilities with our models would, of course, hold also true if we would have included more specifications. The variance inflation factor is calculated for each modification to rule out multicollinearity between the risk factors.

Table 2 summarizes the baseline specification and gives an overview of the considered modifications that are presented in this section. In order to facilitate comparisons, in each modification only a single aspect (compared to the baseline specification) is amended. However, it should be noted that each model is statistically sound and it is ex-ante not obvious why one specification or estimation technique should be more adequate than another.

Table 2: Overview of the specification of the baseline specification and the considered modifications

	Baseline specification (model 1)	Modifications	Model no.
Time-lagged risk factors	Time-lagged macroeconomic variables ($t-2$) and additionally the time-lagged macroeconomic index ($t-2$) are considered as explanatory variables for the macroeconomic index	Without time-lagged macroeconomic index (t-2) as explanatory variable for the macroeconomic index	2
Estimator for the macroeconomic index equation	OLS/Newey-West	FGLS(AR(1)), FGLS(AR(3)) without the time-lagged macroeconomic index ($t-2$) as explanatory variable for the macroeconomic index	3, 4
Transformation between default rate and macroeconomic index	Logit	Probit with BMA and backward regression as method for choosing relevant risk factors	5, 6
Time series processes for macroeconomic variables	AR(1)/AR(2) (based on AIC)	Fixed AR(2), Fixed VAR(1), Fixed VAR(2), SUR	7, 8, 9, 10
Stress test scenario	Historical worst case scenario	Hypothetical scenarios based on three standard deviations of the error terms and based on the Mahalanobis distance	11, 12

2.3.1 Macroeconomic index process

In this section, we describe modifications of the baseline specification that affect the specification and estimation of the macroeconomic index equation (1).

Non-time-lagged macroeconomic index (model 2)

In the base specification, we consider two period time-lagged macroeconomic variables $x_{i,t-2}$, $i \in \{1,2,...,I\}$, and two period time-lagged realizations of the macroeconomic index y_{t-2}, as potential explanatory variables in (1).[19] Within model 2, as in the original CPV-specification (see Wilson (1997a, 1997b)), we do not consider the lagged realizations of the macroeconomic index.[20] For this specification, the BMA is repeated for choosing the multivariately most appropriate risk factors.

FGLS estimator (models 3 and 4)

The OLS estimator is an efficient estimator only in the case of homoscedastic and serially uncorrelated error terms. In our application, the problem of autocorrelation is conceivable due to the methodology of the data preparation for the S&P/Experian Consumer Credit Default

[19] As we have $S=1$, we omit the sector index s in the following.
[20] See also for example Boss (2002), Jokivuolle et al. (2008), and Misina et al. (2006).

Composite Index. In the base specification, we employ the Newey-West estimator to obtain autocorrelation robust standard errors and use the two-period lagged macroeconomic index as an exogenous variable. Another way of considering autocorrelation of the error term u_t in the index equation (1) is to apply the feasible generalized least squares (FGLS) estimator[21]. The FGLS estimator basically assumes a more flexible structure of the variance-covariance matrix of the error terms over time:

$$Var(uu') = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,T} \\ \sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T,1} & \sigma_{T,2} & \cdots & \sigma_T^2 \end{pmatrix}. \tag{13}$$

Based on the autocorrelation-function (ACF) chart, we can observe a significant correlation (up to a significance level of 5%) between the contemporaneous error terms of the macroeconomic index (1) and the error terms of the index equation with a lag of one and three periods. In order to avoid overfitting, we do not consider an AR process of the error terms of equation (1) within the FGLS framework with a lag larger than three. More specifically, we assume an AR(1) (model 3) and an AR(3) (model 4) process (equations (14) and (15)) without intercept for the error term of the macroeconomic index equation (1), respectively:

$$u_t = \rho \cdot u_{t-1} + \delta_t \tag{14}$$

$$u_t = \rho_1 \cdot u_{t-1} + \rho_2 \cdot u_{t-2} + \rho_3 \cdot u_{t-3} + \delta_t \tag{15}$$

where the error term δ_t is normally distributed and uncorrelated with all other error terms of the model. An AR(1) process has also been used for example by McNeil and Wendin (2007) and Miu and Ozdemir (2009). We take the risk factors as selected via the BMA for the baseline specification, but, to avoid the endogeneity problem, we omit the two-period lagged macroeconomic index as exogenous variable.[22]

Probit function (models 5 and 6)

In the baseline model, we employ (as in the original CPV model) a logit transformation to relate the observed default rates to realizations of the macroeconomic index. This is, indeed, not the only possible choice. One alternative is using the probit transformation:[23]

$$p_t = \Phi(-y_t) \Leftrightarrow y_t = -\Phi^{-1}(p_t) \tag{16}$$

---

[21] See Greene (2008, pp. 156-158).

[22] This ensures a higher comparability with our baseline model but neglects that potentially other risk factors might have been included in the model when we would employ a FGLS estimator within the BMA framework.

[23] For further alternatives, see Maddala (1983), Aldrich and Nelson (1984) or Greene (2001).

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. The index y_t gets a negative sign as an argument of $\Phi(\cdot)$ in (16) to ensure that – as in the case of the logit transformation – increasing index values cause decreasing default probabilities. As for model 1, the BMA and the backward regression as a robustness check are repeated for this model specification. Since the selected risk factors of the BMA and the backward regression differ in one risk factor[24], we use both models (models 5 and 6) as specifications of the CPV-style model.

2.3.2 Risk factor processes

Fixed second-order autoregressive processes (model 7)

In the baseline specification, the order of the autoregressive processes by which the risk factors are modelled is selected based on the AIC whereby the order is restricted to a maximum of two. This leads to the situation that for some risk factors an AR(1) process is used and for other risk factors an AR(2) process is implemented.[25] In this section, we want to check for the influence of this assumption on the stress default probabilities. For this, we employ an AR(k) process of fixed order $k = 2$ for all risk factors (model 7).

Vector-autoregressive regression (models 8 and 9)

Instead of using AR processes, it is also possible to model the risk factors by vector-autoregressive (VAR) processes. For example, VAR models are also employed by Schechtman and Gaglianone (2012). VAR processes are often taken into account if very little is known about the structure or relationships between the variables and, therefore, a dependency between all variables is assumed. This requires the estimation of many parameters and, thus, promotes overfitting. This goes along with a good in-sample fit but leads to less reliable out-of-sample forecasts. Based on these arguments, we limit the number of considered lags and assume VAR(1) (model 8) and VAR(2) (model 9) processes for all risk factors. The general VAR(1) model with the parameter matrix $\gamma \in \mathbb{R}^{I \times I}$ and error terms $v \in \mathbb{R}^{I \times 1}$ is given in equation (17):

$$\begin{pmatrix} x_{1,t} \\ \vdots \\ x_{I,t} \end{pmatrix} = \begin{pmatrix} \gamma_{1,1} & \cdots & \gamma_{1,I} \\ \vdots & \ddots & \vdots \\ \gamma_{I,1} & \cdots & \gamma_{I,I} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ \vdots \\ x_{I,t-1} \end{pmatrix} + \begin{pmatrix} v_{x_1,t} \\ \vdots \\ v_{x_I,t} \end{pmatrix}. \tag{17}$$

[24] Industrial production (t-2) is replaced by logarithmic-return imports (t-2).
[25] See Table 4 in Section 3.1.

In this specification, there exists a correlation between the risk factors due to the dependence of a risk factor to lagged other risk factors. For this reason, the contemporaneous correlations of the error terms of the risk factors $v \in \mathbb{R}^{I\times 1}$ are, in contrast to the baseline specification, not considered. Accordingly, we assume a diagonal variance-covariance matrix $\Sigma_{v,v} \in \mathbb{R}^{I\times I}$ for the simulation algorithm described in Section 2.1.

Seemingly unrelated regression (model 10)

Another possible specification is the seemingly unrelated regression (SUR) methodology (model 10).[26] The difference between the SUR methodology and the usage of a VAR process is that there are no obvious influences between the risk factors in the SUR methodology. The risk factors depend solely on their own time-lagged values as exogenous variables. The correlation is computed contemporaneously via the residuals of the AR processes of the risk factors.[27] In contrast to the baseline specification, in model 10, this assumption is not only used in the simulations to forecast the risk factors, but also in the estimation of the parameters of risk factor processes. For this, the order of the AR processes of the risk factors are set equal to those in the baseline specification.

2.3.3 Stress test scenarios

The modifications described in this section do not concern discretion in setting up a model or in the estimation process, but deal with the degree of freedom that risk managers performing stress tests may have, for example under Pillar 2, namely the choice of the stress test scenario.[28] In supervisory stress tests, a scenario is usually given and, accordingly, discretion in choosing stress macroeconomic/financial variables is limited to cases where these scenarios do not cover all presumed variables of a bank's model. For these modifications, the baseline specification of the CPV-style model is employed.

Hypothetical scenario based on three standard deviations (model 11)

In the baseline specification, we define the stress scenario for a single risk factor as the largest historical deviation of the empirical observations for this risk factor from its theoretical model with a negative impact on the macroeconomic index. In model 11, alternatively, the impact of

[26] See the description of the SUR model in Greene (2008).
[27] Analyses that use a SUR methodology to model and forecast macroeconomic risk factors include Jokivuolle et al. (2008), Trenca and Benyovszki (2008), and Zedginidze (2012).
[28] See similarly, but within another modelling framework, Breuer et al. (2012). The requirements on selecting a scenario are, for example, discussed in EBA (2018b).

a given shock on the error term of three standard deviations is taken into account.[29] However, the assumption that only a single risk factor is stressed (univariate stress scenario) is maintained.

Hypothetical scenario based on the Mahalanobis distance (model 12)

In this modification, a *multivariate* stress test scenario based on the Mahalanobis distance of the error terms v_i, $i \in \{1,2,...,I\}$, is used.[30] The Mahalanobis distance of a random vector v is defined as:

$$Maha(v) = \sqrt{(\mu - v)' \cdot \Sigma^{-1} \cdot (\mu - v)} \tag{18}$$

where $\mu = E[v]$ and Σ is the variance-covariance matrix of the vector components. The smaller the Mahalanobis distance of a realization of the random vector v is, the more likely (plausible) – given the variance-covariance structure of the vector components and assumed ellipticity – is the respective realization. The Mahalanobis distance is employed to define so-called trust regions of radius τ around $\mu = E[v]$:

$$Ell_\tau := \left\{ v \in \mathbb{R}^{I \cdot 3} \middle| Maha(v) \le \tau \right\} \tag{19}$$

As we consider a dynamic three-months stress period, the dimension of the random vector $v = (v_{1,T+1},...,v_{I,T+1},v_{1,T+2},...,v_{I,T+2},v_{1,T+3},...,v_{I,T+3})'$ is $I \cdot 3$. The random vector v represents an I-dimensional path of the error terms of the risk factors over the three considered stress periods. We assume $\tilde{v} = (v_{1,n},...,v_{I,n})' \sim N(0,\Sigma_{vv})$ for all $n \in \{T+1, T+2, T+3\}$ (see (4) and (5)).

Using the above notation, the historical stress scenario for a risk factor i in the baseline specification can be represented by $v*_i = (0,...,0,v*_{i,T+1},0,\ldots,0,v*_{i,T+2},0,\ldots,0,v*_{i,T+3},0,\ldots,0)' \in \mathbb{R}^{I \cdot 3}$ with corresponding values $Maha(v_i^*)$ ($i \in \{1,...,I\}$). To ensure consistency between the univariate stress scenarios as set out in the baseline specification and those ones employed in this section, we define trust regions Ell_{τ_i} by setting $\tau_i = Maha(v_i^*)$ ($i \in \{1,...,I\}$). This ensures that the stress scenarios used in this specification and in the baseline specification are equally plausible in the sense of the Mahalanobis distance. However, the stress scenario used in this section defines a multivariate shock, whereas the other stress scenarios (historical worst case, three standard deviations) only imply a univariate shock. Out of each of the trust regions Ell_{τ_i} ($i \in \{1,...,I\}$), we look for

[29] Three standard deviations are a frequent choice (see, for example, Breuer et al. (2012, p. 337)).
[30] See, for example, Breuer et al. (2012) for the use of the Mahalanobis distance for stress testing.

that scenario during the time period of the next three months (which is identical to the assumed duration of the univariate shocks) that maximizes the expected forecasted default probability in $T+5$:

$$v_{\tau_i}^{worst} = \arg\max_{v \in Ell_{\tau_i}} \left\{ E\left[PD_{T+5}(u,v) \middle| F_T, v \right] \right\} \tag{20}$$

where F_T contains all past information up to time T (in particular about the previous realizations of the risk factors). We choose the risk horizon $T+5$ in the optimization problem (20) because the risk factors affect the macroeconomic index with a lag of two periods as set out in Section 2.2. Hence, the default probability forecasted for $T+5$ is the first one that is influenced by all three stress periods.[31]

3 Results

In this section, first, we present the results for the risk factor processes and for the macroeconomic index equation for all model specifications used. In addition, we conduct an out-of-sample comparison between these models to ensure that one model is not dominating the others in terms of forecast ability for the default rates. Second, we show the impact that differing model specifications have for the stress test results.

3.1 Specification of models

Tables 3 and 4 summarize the estimation results for the macroeconomic index equation (1) and for the time series processes of the risk factors based on the full data sample ranging from 07/2004 to 08/2016.

[31] The macroeconomic index in $T+5$ which determines the probability of default in $T+5$ via (2) is given by: $y_{T+5} = y_T + \Delta y_{T+1}\left(x_{T-1}\right) + \Delta y_{T+2}\left(x_T\right) + \Delta y_{T+3}\left(x^*{}_{T+1}\right) + \Delta y_{T+4}\left(x^*{}_{T+2}\right) + \Delta y_{T+5}\left(x^*{}_{T+3}\right), x \in \mathbb{R}^{I\times(T+3)}$.

Table 3: Estimation results for the macroeconomic index equation

	Parameters		R^2	Adjusted R^2
Model 1: Baseline specification			0.2765	0.2555
Intercept	0.0019			
Industrial Production (t-2)	1.5400	**		
VIX (t-2)	-0.0015	***		
Consumer Sentiment (t-2)	0.1573	**		
Macroeconomic Index (t-2)	0.3381	***		
Model 2: Baseline specification without macroeconomic index (t-2) as explanatory variable			0.1731	0.1553
Intercept	0.0027			
Industrial Production (t-2)	2.2302	***		
VIX (t-2)	-0.0016	**		
Consumer Sentiment (t-2)	0.1544	**		
Model 3: FGLS-estimator (AR(1) process for residuals)			-	-
Intercept	0.0030			
Industrial Production (t-2)	1.3513	***		
VIX (t-2)	-0.0014	**		
ρ	0.4919	***		
Model 4: FGLS-estimator (AR(3) process for residuals)			-	-
Intercept	0.0032			
Industrial Production (t-2)	1.0509	**		
VIX (t-2)	-0.0013	**		
ρ_1	0.4764	***		
ρ_2	0.1971	**		
ρ_3	-0.1776	**		
Model 5: Probit transformation (BMA)			0.2952	0.2748
Intercept	0.0007			
Industrial Production (t-2)	0.6407	**		
VIX (t-2)	-0.0006	***		
Consumer Sentiment (t-2)	0.0664	**		
Macroeconomic Index (t-2)	0.3520	***		
Model 6: Probit transformation (backward regression)			0.2952	0.2747
Intercept	0.0005			
Imports (t-2)	0.1624	***		
VIX (t-2)	-0.0007	***		
Consumer Sentiment (t-2)	0.0596	**		
Macroeconomic Index (t-2)	0.3988	***		
Model 7: Fixed AR(2) process for risk factors				
as model 1				
Model 8: Fixed VAR(1) process for risk factors				
as model 1				
Model 9: Fixed VAR(2) process for risk factors				
as model 1				
Model 10: SUR-process for risk factors				
as model 1				

This table summarizes the OLS parameter estimates with Newey-West autocorrelation robust covariance estimator (except for models 3 and 4: FGLS estimator) of the macroeconomic index equation (1) and their significances for various specifications. The symbols *, ** and *** denote significance at the 10%, 5% and 1% levels. For all specifications, the variance inflation factor has been calculated (not shown in the table). As it is always only slightly above one, multicollinearity between the explanatory variables can be ruled out. For model 3 and 4 we cannot specify the coefficient of determination as it is not well-defined in those models and, thus, cannot be interpreted as the (maximum) fraction of explained variance by systematic risk factors.

Table 4: Estimates of the risk factor processes

		Parameters	R2	Adjusted R2	Applied specification
Model 1: Baseline specification					
Industrial Production (t-2)	Intercept	0.0003	0.0752	0.0686	AR(1)[#]
	t-1	0.2734 **			
VIX (t-2)	Intercept	0.0038	0.0481	0.0343	AR(2)
	t-1	-0.0313			
	t-2	-0.2178 ***			
Consumer Sentiment (t-2)	Intercept	-0.0002	0.0566	0.0429	AR(2)
	t-1	0.0296			
	t-2	-0.2368 ***			
Model 2: Baseline specification without macroeconomic index (t-2) as explanatory variable					
as model 1					
Model 3: FGLS-estimator (AR(1) process for residuals)					
as model 1					
Model 4: FGLS-estimator (AR(3) process for residuals)					
as model 1					
Model 5: Probit transformation (BMA)					
as model1					
Model 6: Probit transformation (backward regression)					
Imports (t-2)	Intercept	0.0014	0.08288	0.0696	AR(2)[#]
	t-1	0.0690			
	t-2	0.2708 **			
VIX (t-2)	as model 1				
Consumer Sentiment (t-2)	as model 1				
Model 7: Fixed AR(2) process for risk factors					
Industrial Production (t-2)	Intercept	0.0002	0.1179	0.1052	AR(2)[#]
	t-1	0.2144 *			
	t-2	0.2152 ***			
VIX (t-2)	as model 1				
Consumer Sentiment (t-2)	as model 1				
Model 8: Fixed VAR(1) process for risk factors					
Industrial Production (t-2)	Intercept	0.0003	0.1000	0.0804	VAR(1)[#]
	Industrial Production (t-3)	0.2873 **			
	VIX (t-3)	0.0002 *			
	Consumer Sentiment (t-3)	0.0153			
VIX (t-2)	Intercept	0.0388	0.0306	0.0095	VAR(1)
	Industrial Production (t-3)	-112.2576 **			
	VIX (t-3)	-0.0228			
	Consumer Sentiment (t-3)	-0.6026			

Table 4: Estimates of the risk factor processes (continued)

Consumer Sentiment (t-2)	Intercept	-0.0009	0.0987	0.0791	VAR(1)
	Industrial Production (t-3)	1.3917 **			
	VIX (t-3)	-0.0031 ***			
	Consumer Sentiment (t-3)	0.0152			
Model 9: Fixed VAR(2) process for risk factors					
Industrial Production (t-2)	Intercept	0.0002	0.1704	0.1333	VAR(2) #
	Industrial Production (t-3)	0.2217 ***			
	Industrial Production (t-4)	0.2461 ***			
	VIX (t-3)	0.0002			
	VIX (t-4)	-0.0002			
	Consumer Sentiment (t-3)	0.0035			
	Consumer Sentiment (t-4)	0.0038			
VIX (t-2)	Intercept	0.0504	0.0794	0.0382	VAR(2)
	Industrial Production (t-3)	-103.5955 *			
	Industrial Production (t-4)	1.8514			
	VIX (t-3)	-0.0363			
	VIX (t-4)	-0.2038 **			
	Consumer Sentiment (t-3)	-5.3499			
	Consumer Sentiment (t-4)	7.1600			
Consumer Sentiment (t-2)	Intercept	-0.0008	0.1638	0.1264	VAR(2)
	Industrial Production (t-3)	1.7107 ***			
	Industrial Production (t-4)	-0.5887			
	VIX (t-3)	-0.0032 ***			
	VIX (t-4)	-0.0007			
	Consumer Sentiment (t-3)	0.0258			
	Consumer Sentiment (t-4)	-0.2613 ***			
Model 10: SUR process for risk factors					
Industrial Production (t-2)	Intercept	0.0003	0.0733	0.0666	SUR-AR(1)
	t-1	0.3164 ***			
VIX (t-2)	Intercept	0.0028	0.0469	0.0331	SUR-AR(2)
	t-1	-0.0656			
	t-2	-0.2174 ***			
Consumer Sentiment (t-2)	Intercept	-0.0002	0.0562	0.0425	SUR-AR(2)
	t-1	0.0458			
	t-2	-0.2252 ***			

This table summarizes the OLS parameter estimates (except for model 10: SUR process) of the risk factor processes and their significances. The symbols *, ** and *** denote significance at the 10%, 5% and 1% level. When the minimal p-value of the Godfrey-Breusch test (up to a lag of four) is below 5%, the Newey-West estimator is used instead of the OLS estimator (denoted by #).

After having applied the Bayesian model averaging approach, we include the two-period lagged variables industrial production, VIX and consumer sentiment as well as the two-period lagged macroeconomic index itself as explanatory variables in the baseline specification (see

Table 3).[32] The explained variance of the model is 27.7% and the adjusted R^2 is 25.6%. The signs of the coefficients of the explanatory variables in the baseline specification are economically reasonable. A positive sign implies that increasing risk factor realizations go along with increasing index realizations and, hence, decreasing default probabilities (see (2)), and vice versa. As an increase in industrial production as well as higher values for consumer sentiment can usually be observed in economically good times due to the rise in demand, the estimated positive signs of the regression coefficients of the explanatory variables are in line with our intuition. At the same time, a negative sign for the coefficient of the variable VIX coincides with our intuition, as increased market volatility is due to investor uncertainty. Consequently, a decline in the macroeconomic index at high levels of the VIX is plausible. For the modified models 2 to 10, the signs of the estimated regression coefficients are also in line with our intuition. The adjusted R^2 ranges from 15.5% to 27.5%. The best fit in terms of the adjusted R^2 show model 5 and model 6 with a probit transformation of the S&P/Experian Consumer Credit Default Composite Index.

Using the information criteria AIC for selecting the order of the AR processes for the risk factors,[33] we effectively obtain risk factor processes of order one and two (see Table 4). The specification of the AR processes has an influence on how long it takes until an initial shock vanishes. The R^2 ranges from 3% to 17% and the values for the adjusted R^2 are between 1% and 13.3%.

In Section 2.1, we assumed that the covariances between the error terms of the index equations (see (1)) and the error terms of the risk factor equations (see (3)) are equal to zero ($\Sigma_{u,v} = \Sigma_{v,u} = 0$). Deviating from this assumption would have two implications. First, when doing the stress simulations for the future default probabilities, a non-zero covariance would have to be considered when sampling from the conditional normal distribution (see (6) and (7)) for the remaining error terms. Of course, this could have an influence on the simulated stress default probabilities. Second, the assumption $\Sigma_{u,v} \neq 0$ would directly cause an endogeneity problem in the index equation (1). When the error term u_s of sector s is correlated with the error term v_i of any risk factor i, this implies $Corr(x_i, u_s) \neq 0$. As a consequence, the OLS estimator for the parameters $\beta_{s,0},...,\beta_{s,I}$ of the index equation would be biased and in-

[32] To ensure stationarity, we compute logarithmic returns of the variables industrial production and consumer sentiment and first differences of VIX and the macroeconomic index.
[33] Except for model 7 where we employ a fixed lag number of two.

consistent. In many studies on stress testing that employ the CPV model, the possibility $\Sigma_{u,v} \neq 0$ is not directly excluded, but the issue of endogeneity is rarely explicitly addressed.[34]

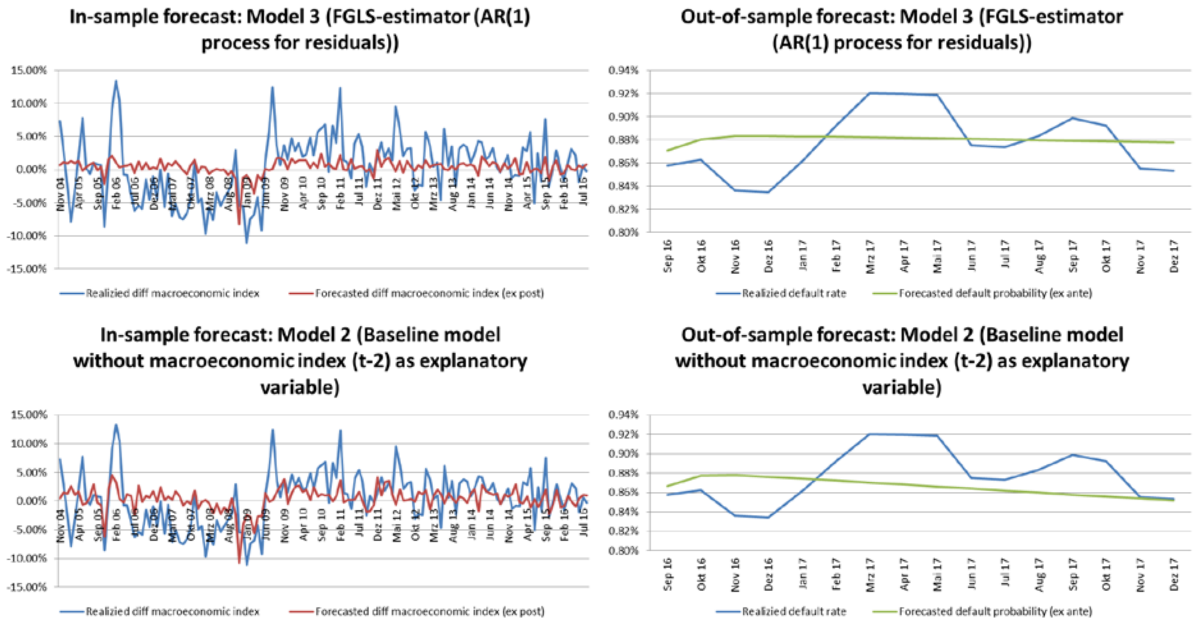
As we only assumed $\Sigma_{u,v} = \Sigma_{v,u} = 0$ and as an endogeneity problem might exist even if this assumption would be true (for example because of missing correlated variables in the index equation), we test for endogeneity of each of the explanatory variables (industrial production ($t-2$), VIX ($t-2$) and consumer sentiment ($t-2$)) in our baseline specification (model 1). For this purpose, the Hausman test is employed. To perform this test, we need instrument variables that are strong and exogenous. First, as in Schechtman and Gaglianone (2012), we use the risk factors itself with a further lag (compared to the baseline specification) as instrument variables. However, the lagged variables VIX ($t-3$) and consumer sentiment ($t-3$) prove to be weak instrument variables because their F-statistics are 0.11 and 0.08, respectively. To find strong instrument variables (that means F-statistics larger than 10) for the VIX ($t-2$) and for the consumer sentiment ($t-2$), (further) lagged and contemporaneous ($t-2$) exogenous variables of Table 1 are tried. Two contemporaneous variables are found to be strong instrument variables for VIX ($t-2$).[35] For consumer sentiment ($t-2$), one two period time-lagged variable ($t-4$) is identified as a strong instrument variable.[36] All these strong instrument variables were used for performing the Hausman test for endogeneity of the explanatory variables of the baseline specification (model 1).[37] For all risk factors in the baseline specification the null hypothesis of exogeneity could not be rejected. Thus, endogeneity and biased parameter estimates seem to be no problem in the baseline specification.

On the left hand side of Figure 2, the realized first differences of the macroeconomic index are compared with the in-sample (07/2004 to 08/2016) forecasted first differences of the macroeconomic index. On the right hand side, the realized default rates are compared with the out-of-sample (09/2016 to 12/2017) predictions of the default probabilities (based on (1) and (2)). We only show the models with the highest (model 3: FGLS estimator (AR(1) process for residuals)) and lowest (model 2: Baseline specification without macroeconomic index (t-2) as explanatory variable) forecasted default probability at the risk horizon of one year. For the in-

[34] See, for example, Boss (2002) or Virolainen (2004). An exception is Schechtman and Gaglianone (2012).
[35] These are the S&P 500 ($t-2$) and the USD/GBP exchange rate ($t-2$).
[36] This is the variable VIX ($t-4$).
[37] The instrument variable parameter estimates needed for the Hausman test statistic are computed using two stage least squares (2SLS). The Hausman test is repeated for the contemporaneous instrument variables S&P 500 ($t-2$) and USD/GBP exchange rate ($t-2$). Following the same procedure, the null hypothesis of exogeneity could not be rejected.

sample predictions of the first differences in the macroeconomic index (monthly changes), the observed risk factor realizations of each model are inserted into (1) and the error term is set equal to its mean zero. As Figure 2 shows, the in-sample performance of the models estimated on the full data sample is not brilliant, but, at least, the downward peak during the crisis is reflected. For the out-of-sample prediction, the mean forecasted default probabilities in the non-stress case are employed.[38]

Figure 2: Realized versus in-sample forecasted first differences in the macroeconomic index and realized versus out-of-sample forecasted default probabilities



On the left hand side, this figure shows the realized first differences of the macroeconomic index compared with the in-sample (07/2004 to 08/2016) forecasted first differences of the macroeconomic index. On the right hand side, the realized default rates are compared with the out-of-sample (09/2016 to 12/2017) predictions of the default probabilities (based on (1) and (2)). We only show the models with the highest (model 3: FGLS estimator (AR(1) process for residuals)) and lowest (model 2: Baseline specification without macroeconomic index (t-2) as explanatory variable) forecasted default probability at the risk horizon of one year. For the in-sample predictions of the first differences in the macroeconomic index (monthly changes), the observed risk factor realizations of each model are inserted into (1) and the error term is set equal to its mean zero. For the out-of-sample prediction, the mean forecasted default probabilities in the non-stress case are employed.

Based on one million forecasts of the default probabilities at a risk horizon of $T+12$ (08/2017), Table 5 shows the mean deviation (MD) between the forecasted default probabilities $PD_{T+12}^{forecasted}$ and the realized default rates $PD_{T+12}^{realized}$ as well as the mean squared error (MSE):

$$MD_{T+12} = E\left[PD_{T+12}^{forecasted} - PD_{T+12}^{realized} \middle| F_T \right], \tag{21}$$

[38] The realized default rates for the periods 09/2016 to 12/2017 are taken from S&P Dow Jones Indices (2018).

$$MSE_{T+12} = E\left[\left(PD_{T+12}^{forecasted} - PD_{T+12}^{realized}\right)^2 \middle| F_T\right], \tag{22}$$

where F_T denotes the available information up to time $T = 08/2016$. Out-of-sample, the best performing models (in terms of mean squared errors) are the baseline specification and the baseline specification without macroeconomic index ($t-2$) as explanatory variable (models 1 and 2). Evaluating the out-of-sample performance based on the mean deviation, models 3 and 8 are leading. However, if we consider the R^2 and the adjusted R^2 as a measure of in-sample forecasting capability, model 2 is the worst performing model. Thus, no specification is clearly dominating the other specifications.

Table 5: Out-of-sample performance for a risk horizon of one year

	MD	MSE
Model 1: Baseline specification	-0.0144%	3.17E-06
Model 2: Baseline specification without macroeconomic index (t-2) as explanatory variable	-0.0235%	2.00E-06
Model 3: FGLS-estimator (AR(1) process for residuals)	-0.0035%	4.91E-06
Model 4: FGLS-estimator (AR(3) process for residuals)	-0.0145%	4.56E-06
Model 5: Probit transformation (BMA)	-0.0112%	4.09E-06
Model 6: Probit transformation (backward regression)	-0.0087%	4.62E-06
Model 7: Fixed AR(2) process for risk factors	-0.0127%	3.28E-06
Model 8: Fixed VAR(1) process for risk factors	-0.0057%	3.61E-06
Model 9: Fixed VAR(2) process for risk factors	-0.0093%	3.61E-06
Model 10: SUR-process for risk factors	-0.0135%	3.22E-06
Model 11: Three standard deviations stress scenario	-0.0143%	3.18E-06
Model 12: Mahalanobis-based stress scenario (no (cross) autocorrelation)	-0.0143%	3.19E-06

This table shows the mean deviation (in percentage points) between the forecasted default probabilities and the realized default rates and the mean squared error at a risk horizon of one year ($T+12$). Expectations are based on one million simulated forecasts of the default probabilities in each period.

3.2 Stress default probabilities

We simulate paths of PDs for the twelve models considered (see Table 2) and evaluate these paths after one, two and three years. Then, we assess model and estimation risk based on the discrepancies between the forecasted PDs in the different models.

The results of the simulation are presented in Table 6. Specifically, the table shows the *expected* and *unexpected* (99.9% quantile[39]) PDs of all twelve models for the non-stress and

[39] The 99.9% quantile of the empirical distribution function of the forecasted default probabilities is a much more prudent measure for the PD.

stress scenarios (i.e., separate shocks in industrial production, VIX and consumer sentiment) for risk horizons of one, two and three years.

All twelve models yield *expected* non-stress PDs between 0.86% and 0.88% for a risk horizon of one year. This is a plausible range given that the last observed PD in 08/2016 is 0.85%. When we expand the risk horizon to two or three years, the differences in forecasted PDs between the models remain very low. However, the discrepancies between the models become more evident in the stress scenarios. For example, a shock in the industrial production transmits to expected stress PDs at a risk horizon of one year between 1.03% (model 4) and 1.59% (model 9).

When comparing the individual models with each other, we rely on a measure which is similar to the EBA's proceeding in the EU-wide stress tests. More specifically, we focus on the differences of the stress and the non-stress PDs across the individual models. This is similar to computing multipliers for converting non-stress PDs to stress PDs. In this regard, though the absolute difference between the expected stress PDs resulting from the different models may seem to be low, we have - in relative terms - an increase compared with the non-stress PDs between +19% and +82% (for a shock in the industrial production and a risk horizon of one year). This is a substantial dispersion across the models. The results for longer risk horizons corroborate these findings. When transmitting a shock in the industrial production over a risk horizon of three years, the expected default probability ranges between 1.00% (model 11) and 1.59% (model 9) corresponding to relative increases of +20% and +87%. For the other risk factors (VIX, consumer sentiment), the dispersion across the models is smaller.

The results for the *unexpected* stress PDs confirm the high discrepancies between the twelve models in relative terms. For the sake of comparison with the previous results, we exemplarily describe the results of the scenario with a shock in the industrial production which again yields the largest dispersion across the specifications. The models forecast unexpected stress PDs between 1.88% (model 11) and 2.92% (model 9) for a one-year risk horizon - these numbers correspond to an increase between +18% and +76% compared with the non-stress scenario of the same models. When extending the risk horizon to three years, the discrepancies remain substantial. The shock in industrial production leads to unexpected default probabilities between 2.59% (model 2) and 5.03% (model 9) which correspond to relative increases between +39% and +82%. The results for shocks in the VIX or consumer sentiment confirm these results, albeit the discrepancies between individual models are smaller. This underpins the importance of selecting relevant risk factors.[40]

[40] For example, the EBA provides in its stress test methodology shock scenarios consisting of various risk factors. However, banks are not required to include all of these risk factors in their stress test model for credit risk.

Table 6: Forecasted default probabilities for the full data sample

Risk horizon of one year (T+12)

	Mean			
	Non-Stress	**Industrial Production (t-2)**	**VIX (t-2)**	**Consumer Sentiment (t-2)**
Model 1: Baseline model	0.87%	1.21% (39.54%)	0.98% (12.35%)	0.94% (7.74%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	0.86%	1.20% (40.00%)	0.93% (7.84%)	0.89% (3.33%)
Model 3: FGLS-estimator (AR(1) process for residuals)	0.88%	1.09% (24.41%)	0.93% (5.94%)	
Model 4: FGLS-estimator (AR(3) process for residuals)	0.87%	1.03% (18.50%)	0.92% (5.89%)	
Model 5: Probit transformation (BMA)	0.87%	1.27% (45.75%)	1.00% (14.42%)	0.95% (9.20%)
Model 6: Probit transformation (backward regression)	0.87%	1.29% (47.30%)	1.02% (16.89%)	0.99% (12.84%)
Model 7: Fixed AR(2) process for risk factors	0.87%	1.32% (52.10%)	0.97% (11.78%)	0.93% (6.54%)
Model 8: Fixed VAR(1) process for risk factors	0.88%	1.44% (63.94%)	0.99% (12.61%)	1.00% (13.95%)
Model 9: Fixed VAR(2) process for risk factors	0.87%	1.59% (81.75%)	0.99% (13.51%)	0.96% (9.83%)
Model 10: SUR-process for risk factors	0.87%	1.23% (41.80%)	0.98% (12.51%)	0.94% (7.81%)
Model 11: Three standard deviations stress scenario	0.87%	1.04% (19.59%)	0.94% (8.37%)	0.93% (6.63%)
Model 12: Mahalanobis-based stress scenario	0.87%	1.31% (50.27%)	1.19% (36.80%)	1.13% (29.47%)

	99.9% quantile			
	Non-Stress	**Industrial Production (t-2)**	**VIX (t-2)**	**Consumer Sentiment (t-2)**
Model 1: Baseline model	1.59%	2.20% (38.57%)	1.78% (12.22%)	1.71% (7.58%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	1.39%	1.93% (38.20%)	1.50% (7.88%)	1.44% (3.25%)
Model 3: FGLS-estimator (AR(1) process for residuals)	1.83%	2.27% (23.53%)	1.94% (5.77%)	
Model 4: FGLS-estimator (AR(3) process for residuals)	1.77%	2.10% (18.42%)	1.88% (6.10%)	
Model 5: Probit transformation (BMA)	1.68%	2.35% (39.74%)	1.90% (12.91%)	1.83% (8.56%)
Model 6: Probit transformation (backward regression)	1.75%	2.46% (40.87%)	2.02% (15.41%)	1.96% (11.97%)
Model 7: Fixed AR(2) process for risk factors	1.60%	2.40% (49.58%)	1.79% (11.74%)	1.71% (6.50%)
Model 8: Fixed VAR(1) process for risk factors	1.66%	2.65% (59.85%)	1.86% (12.18%)	1.88% (13.17%)
Model 9: Fixed VAR(2) process for risk factors	1.66%	2.92% (76.24%)	1.88% (13.42%)	1.81% (9.44%)
Model 10: SUR-process for risk factors	1.59%	2.22% (39.50%)	1.79% (12.35%)	1.72% (7.79%)
Model 11: Three standard deviations stress scenario	1.59%	1.88% (18.47%)	1.71% (7.86%)	1.70% (6.71%)
Model 12: Mahalanobis-based stress scenario	1.59%	2.35% (47.75%)	2.14% (34.37%)	2.02% (27.20%)

Table 6: Forecasted default probabilities for the full data sample (continued)

Risk horizon of two years (T+24)

	Mean			
	Non-Stress	Industrial Production (t-2)	VIX (t-2)	Consumer Sentiment (t-2)
Model 1: Baseline model	0.85%	1.19% (40.07%)	0.96% (12.50%)	0.92% (7.82%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	0.84%	1.17% (39.94%)	0.90% (7.83%)	0.86% (3.29%)
Model 3: FGLS-estimator (AR(1) process for residuals)	0.87%	1.09% (24.48%)	0.93% (5.95%)	
Model 4: FGLS-estimator (AR(3) process for residuals)	0.86%	1.02% (18.43%)	0.91% (5.89%)	
Model 5: Probit transformation (BMA)	0.86%	1.26% (46.32%)	0.98% (14.48%)	0.94% (9.20%)
Model 6: Probit transformation (backward regression)	0.87%	1.29% (49.23%)	1.01% (16.92%)	0.98% (12.89%)
Model 7: Fixed AR(2) process for risk factors	0.86%	1.32% (54.07%)	0.96% (11.77%)	0.91% (6.41%)
Model 8: Fixed VAR(1) process for risk factors	0.87%	1.43% (64.81%)	0.98% (12.67%)	0.99% (14.01%)
Model 9: Fixed VAR(2) process for risk factors	0.86%	1.61% (86.47%)	0.98% (13.82%)	0.95% (10.02%)
Model 10: SUR-process for risk factors	0.85%	1.22% (42.31%)	0.96% (12.62%)	0.92% (7.81%)
Model 11: Three standard deviations stress scenario	0.85%	1.02% (19.78%)	0.92% (8.41%)	0.91% (6.67%)
Model 12: Mahalanobis-based stress scenario	0.85%	1.29% (50.74%)	1.17% (37.09%)	1.11% (29.65%)

	99.9% quantile			
	Non-Stress	Industrial Production (t-2)	VIX (t-2)	Consumer Sentiment (t-2)
Model 1: Baseline model	2.06%	2.85% (38.15%)	2.31% (11.72%)	2.21% (7.24%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	1.65%	2.30% (39.02%)	1.78% (7.86%)	1.71% (3.70%)
Model 3: FGLS-estimator (AR(1) process for residuals)	2.51%	3.10% (23.69%)	2.67% (6.40%)	
Model 4: FGLS-estimator (AR(3) process for residuals)	2.37%	2.80% (18.16%)	2.50% (5.58%)	
Model 5: Probit transformation (BMA)	2.22%	3.09% (39.17%)	2.48% (11.89%)	2.39% (7.66%)
Model 6: Probit transformation (backward regression)	2.39%	3.34% (39.81%)	2.73% (14.18%)	2.65% (10.85%)
Model 7: Fixed AR(2) process for risk factors	2.12%	3.21% (51.27%)	2.34% (10.34%)	2.24% (5.65%)
Model 8: Fixed VAR(1) process for risk factors	2.22%	3.59% (61.68%)	2.48% (11.73%)	2.50% (12.45%)
Model 9: Fixed VAR(2) process for risk factors	2.25%	4.07% (81.42%)	2.54% (13.08%)	2.46% (9.45%)
Model 10: SUR-process for risk factors	2.07%	2.91% (40.72%)	2.33% (12.60%)	2.24% (8.21%)
Model 11: Three standard deviations stress scenario	2.06%	2.45% (19.17%)	2.22% (8.13%)	2.20% (6.80%)
Model 12: Mahalanobis-based stress scenario	2.06%	3.06% (48.40%)	2.79% (35.36%)	2.63% (27.53%)

Table 6: Forecasted default probabilities for the full data sample (continued)

Risk horizon of three years (T+36)

	Mean			
	Non-Stress	Industrial Production (t-2)	VIX (t-2)	Consumer Sentiment (t-2)
Model 1: Baseline model	0.84%	1.17% (40.07%)	0.94% (12.52%)	0.90% (7.83%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	0.81%	1.14% (39.92%)	0.88% (7.86%)	0.84% (3.27%)
Model 3: FGLS-estimator (AR(1) process for residuals)	0.87%	1.08% (24.33%)	0.92% (5.90%)	
Model 4: FGLS-estimator (AR(3) process for residuals)	0.85%	1.01% (18.40%)	0.90% (5.84%)	
Model 5: Probit transformation (BMA)	0.84%	1.23% (46.16%)	0.97% (14.38%)	0.92% (9.07%)
Model 6: Probit transformation (backward regression)	0.86%	1.28% (48.91%)	1.00% (16.82%)	0.97% (12.81%)
Model 7: Fixed AR(2) process for risk factors	0.84%	1.30% (54.02%)	0.94% (11.73%)	0.90% (6.39%)
Model 8: Fixed VAR(1) process for risk factors	0.85%	1.41% (64.74%)	0.96% (12.66%)	0.97% (13.96%)
Model 9: Fixed VAR(2) process for risk factors	0.85%	1.59% (86.41%)	0.97% (13.78%)	0.94% (10.03%)
Model 10: SUR-process for risk factors	0.84%	1.19% (42.32%)	0.94% (12.62%)	0.90% (7.81%)
Model 11: Three standard deviations stress scenario	0.84%	1.00% (19.76%)	0.91% (8.40%)	0.89% (6.70%)
Model 12: Mahalanobis-based stress scenario	0.84%	1.26% (50.78%)	1.15% (37.16%)	1.08% (29.61%)

	99.9% quantile			
	Non-Stress	Industrial Production (t-2)	VIX (t-2)	Consumer Sentiment (t-2)
Model 1: Baseline model	2.46%	3.44% (39.94%)	2.76% (12.23%)	2.65% (7.85%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	1.86%	2.59% (39.28%)	2.00% (7.66%)	1.94% (4.01%)
Model 3: FGLS-estimator (AR(1) process for residuals)	3.15%	3.87% (22.54%)	3.33% (5.48%)	
Model 4: FGLS-estimator (AR(3) process for residuals)	2.92%	3.44% (17.70%)	3.08% (5.35%)	
Model 5: Probit transformation (BMA)	2.67%	3.69% (38.36%)	3.01% (13.11%)	2.86% (7.29%)
Model 6: Probit transformation (backward regression)	2.94%	4.10% (39.60%)	3.36% (14.57%)	3.26% (10.96%)
Model 7: Fixed AR(2) process for risk factors	2.55%	3.87% (51.68%)	2.83% (10.86%)	2.69% (5.70%)
Model 8: Fixed VAR(1) process for risk factors	2.71%	4.37% (61.38%)	3.04% (12.18%)	3.05% (12.69%)
Model 9: Fixed VAR(2) process for risk factors	2.76%	5.03% (81.99%)	3.14% (13.69%)	3.03% (9.85%)
Model 10: SUR-process for risk factors	2.47%	3.50% (41.36%)	2.80% (13.17%)	2.66% (7.68%)
Model 11: Three standard deviations stress scenario	2.46%	2.94% (19.41%)	2.66% (8.27%)	2.63% (7.06%)
Model 12: Mahalanobis-based stress scenario	2.47%	3.67% (48.88%)	3.34% (35.33%)	3.16% (28.21%)

This table shows the mean and the 99.9% quantile of the empirical probability distribution of the forecasted stress default probabilities for various model specifications. The relative deviation between the stress PDs and the non-stress PDs is indicated in parentheses. For models 3 and 4 (FGLS-estimator with AR(1) (AR(2)) process for residuals), there are no entries in the column 'consumer sentiment' because the variable consumer sentiment is not significant in the index equation (1) for these models. For model 6 (probit transformation with backward regression), the shock in industrial production $(t-2)$ is replaced by a shock in imports $(t-2)$. In the case of model 12, the stress scenarios are characterized by the most harmful (in the sense of (20)) scenarios out of those trust regions Ell_{τ} that correspond to the respective historical worst case stress of the macroeconomic variables in the baseline specification (see Section 2.1). The maximum (dark grey) and minimum (light grey) forecasted values of the PDs are indicated for each (non-)stress scenario.

We have observed large relative discrepancies in the PD forecasts across the models though the absolute differences remained relatively low. This is because PDs decreased considerably after the crisis and remained at a historically low level – the last observed PD is 0.85% in 08/2016 (see Figure 1). In order to demonstrate that the discrepancies in the models' stress PD forecasts are not only existent in relative terms but also in absolute terms, we re-estimate[41] all models using the shorter time period 7/2004 to 12/2009 as these data are dominated by the crisis ("stress period calibration", see Figure 1). The PD observed in 12/2009 is 4.78% which is fundamentally larger than 0.85% in 08/2016. Generally, the models' calibration remains stable. However, it turns out that the VIX only has an insignificant impact for the shorter data sample and, thus, this variable is exempted from the model. The results based on the stress period calibration are provided in Table 7.

[41] This means that we have to re-estimate the macroeconomic equation and the risk-factor processes using only data from the period 7/2004 to 12/2009. The estimation results for these processes are available upon request.

Table 7: Forecasted default probabilities for the subsample up to 12/2009

Risk horizon of one year (T+12)

	Mean		
	Non-Stress	Industrial Production (t-2)	Consumer Sentiment (t-2)
Model 1: Baseline model	5.19%	6.91% (33.03%)	5.70% (9.74%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	5.37%	7.17% (33.46%)	5.67% (5.49%)
Model 3: FGLS-estimator (AR(1) process for residuals)	5.25%	6.87% (30.79%)	5.34% (1.77%)
Model 4: FGLS-estimator (AR(3) process for residuals)	5.39%	7.23% (34.00%)	5.46% (1.23%)
Model 5: Probit transformation (BMA)	5.09%	6.61% (29.84%)	5.53% (8.66%)
Model 6: Probit transformation (backward regression)	4.89%	6.96% (42.43%)	5.30% (8.37%)
Model 7: Fixed AR(2) process for risk factors	5.17%	7.03% (35.98%)	6.05% (17.05%)
Model 8: Fixed VAR(1) process for risk factors	5.15%	7.43% (44.38%)	5.73% (11.32%)
Model 9: Fixed VAR(2) process for risk factors	5.13%	9.07% (76.66%)	5.92% (15.35%)
Model 10: SUR-process for risk factors	5.19%	7.07% (36.22%)	5.70% (9.94%)
Model 11: Three standard deviations stress scenario	5.19%	6.41% (23.36%)	5.72% (10.06%)
Model 12: Mahalanobis-based stress scenario	5.20%	7.37% (41.88%)	6.69% (28.81%)

	99.9% quantile		
	Non-Stress	Industrial Production (t-2)	Consumer Sentiment (t-2)
Model 1: Baseline model	9.64%	12.53% (30.03%)	10.55% (9.43%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	9.10%	11.81% (29.79%)	9.57% (5.20%)
Model 3: FGLS-estimator (AR(1) process for residuals)	12.29%	15.61% (27.01%)	12.52% (1.87%)
Model 4: FGLS-estimator (AR(3) process for residuals)	10.58%	13.83% (30.74%)	10.73% (1.36%)
Model 5: Probit transformation (BMA)	8.76%	10.98% (25.33%)	9.46% (7.96%)
Model 6: Probit transformation (backward regression)	8.77%	11.77% (34.31%)	9.43% (7.60%)
Model 7: Fixed AR(2) process for risk factors	9.66%	12.78% (32.38%)	11.18% (15.74%)
Model 8: Fixed VAR(1) process for risk factors	9.86%	13.78% (39.75%)	10.88% (10.34%)
Model 9: Fixed VAR(2) process for risk factors	10.18%	16.77% (64.83%)	11.52% (13.24%)
Model 10: SUR-process for risk factors	9.69%	12.85% (32.63%)	10.61% (9.48%)
Model 11: Three standard deviations stress scenario	9.65%	11.68% (21.02%)	10.55% (9.23%)
Model 12: Mahalanobis-based stress scenario	9.65%	13.25% (37.36%)	12.10% (25.44%)

Table 7: Forecasted default probabilities for the subsample up to 12/2009 (continued)

Risk horizon of two years (T+24)

	Mean		
	Non-Stress	Industrial Production (t-2)	Consumer Sentiment (t-2)
Model 1: Baseline model	6.46%	8.55% (32.42%)	7.07% (9.56%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	6.66%	8.84% (32.77%)	7.02% (5.41%)
Model 3: FGLS-estimator (AR(1) process for residuals)	6.65%	8.64% (29.85%)	6.77% (1.80%)
Model 4: FGLS-estimator (AR(3) process for residuals)	6.78%	9.02% (33.20%)	6.86% (1.28%)
Model 5: Probit transformation (BMA)	6.15%	7.91% (28.52%)	6.66% (8.27%)
Model 6: Probit transformation (backward regression)	5.91%	8.47% (43.20%)	6.36% (7.63%)
Model 7: Fixed AR(2) process for risk factors	6.38%	8.62% (35.24%)	7.44% (16.73%)
Model 8: Fixed VAR(1) process for risk factors	6.42%	9.21% (43.50%)	7.12% (11.02%)
Model 9: Fixed VAR(2) process for risk factors	6.44%	11.45% (77.84%)	7.42% (15.22%)
Model 10: SUR-process for risk factors	6.45%	8.75% (35.54%)	7.08% (9.71%)
Model 11: Three standard deviations stress scenario	6.46%	7.94% (22.95%)	7.09% (9.85%)
Model 12: Mahalanobis-based stress scenario	6.46%	9.11% (40.97%)	8.29% (28.27%)

	99.9% quantile		
	Non-Stress	Industrial Production (t-2)	Consumer Sentiment (t-2)
Model 1: Baseline model	15.42%	19.79% (28.32%)	16.71% (8.34%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	13.79%	17.73% (28.59%)	14.44% (4.71%)
Model 3: FGLS-estimator (AR(1) process for residuals)	21.79%	26.94% (23.65%)	22.25% (2.09%)
Model 4: FGLS-estimator (AR(3) process for residuals)	16.92%	21.68% (28.13%)	17.15% (1.37%)
Model 5: Probit transformation (BMA)	12.87%	15.78% (22.65%)	13.74% (6.76%)
Model 6: Probit transformation (backward regression)	13.31%	17.67% (32.75%)	14.15% (6.33%)
Model 7: Fixed AR(2) process for risk factors	15.35%	19.99% (30.22%)	17.58% (14.55%)
Model 8: Fixed VAR(1) process for risk factors	16.18%	22.19% (37.14%)	17.69% (9.32%)
Model 9: Fixed VAR(2) process for risk factors	17.30%	27.60% (59.54%)	19.38% (12.02%)
Model 10: SUR-process for risk factors	15.70%	20.33% (29.49%)	16.96% (8.05%)
Model 11: Three standard deviations stress scenario	15.37%	18.44% (19.98%)	16.78% (9.20%)
Model 12: Mahalanobis-based stress scenario	15.41%	20.81% (34.98%)	19.18% (24.45%)

Table 7: Forecasted default probabilities for the subsample up to 12/2009 (continued)

Risk horizon of three years (T+36)

	Mean		
	Non-Stress	Industrial Production (t-2)	Consumer Sentiment (t-2)
Model 1: Baseline model	7.99%	10.50% (31.50%)	8.73% (9.28%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	8.22%	10.84% (31.82%)	8.66% (5.25%)
Model 3: FGLS-estimator (AR(1) process for residuals)	8.35%	10.74% (28.56%)	8.49% (1.67%)
Model 4: FGLS-estimator (AR(3) process for residuals)	8.46%	11.18% (32.12%)	8.57% (1.28%)
Model 5: Probit transformation (BMA)	7.37%	9.36% (27.04%)	7.95% (7.86%)
Model 6: Probit transformation (backward regression)	7.09%	9.99% (40.86%)	7.61% (7.29%)
Model 7: Fixed AR(2) process for risk factors	7.82%	10.50% (34.26%)	9.10% (16.34%)
Model 8: Fixed VAR(1) process for risk factors	7.95%	11.30% (42.07%)	8.81% (10.70%)
Model 9: Fixed VAR(2) process for risk factors	8.03%	14.02% (74.61%)	9.22% (14.80%)
Model 10: SUR-process for risk factors	7.99%	10.74% (34.50%)	8.74% (9.42%)
Model 11: Three standard deviations stress scenario	7.98%	9.77% (22.39%)	8.75% (9.62%)
Model 12: Mahalanobis-based stress scenario	7.99%	11.16% (39.75%)	10.18% (27.47%)

	99.9% quantile		
	Non-Stress	Industrial Production (t-2)	Consumer Sentiment (t-2)
Model 1: Baseline model	22.33%	27.84% (24.69%)	23.87% (6.90%)
Model 2: Baseline model without macroeconomic index (t-2) as explanatory variable	19.37%	24.43% (26.11%)	20.13% (3.93%)
Model 3: FGLS-estimator (AR(1) process for residuals)	32.93%	39.49% (19.90%)	33.24% (0.93%)
Model 4: FGLS-estimator (AR(3) process for residuals)	24.36%	30.64% (25.78%)	24.86% (2.03%)
Model 5: Probit transformation (BMA)	17.25%	20.72% (20.17%)	18.34% (6.36%)
Model 6: Probit transformation (backward regression)	18.22%	23.50% (28.95%)	19.23% (5.51%)
Model 7: Fixed AR(2) process for risk factors	22.05%	28.19% (27.88%)	24.91% (13.00%)
Model 8: Fixed VAR(1) process for risk factors	23.73%	31.48% (32.69%)	25.59% (7.85%)
Model 9: Fixed VAR(2) process for risk factors	25.65%	38.79% (51.22%)	28.46% (10.95%)
Model 10: SUR-process for risk factors	22.76%	28.75% (26.36%)	24.33% (6.94%)
Model 11: Three standard deviations stress scenario	22.30%	26.21% (17.54%)	24.10% (8.09%)
Model 12: Mahalanobis-based stress scenario	22.21%	29.12% (31.10%)	27.11% (22.04%)

This table shows the mean and the 99.9% quantile of the empirical probability distribution of the forecasted stress default probabilities for various model specifications. In contrast to Table 6, use only a stress period for calibration. The relative deviation between the stress PDs and the non-stress PDs is indicated in parentheses. For models 3 and 4 (FGLS-estimator with AR(1) (AR(2)) process for residuals), there are no entries in the column 'consumer sentiment' because the variable consumer sentiment is not significant in the index equation (1) for these models. For model 6 (probit transformation with backward regression), the shock in industrial production ($t-2$) is replaced by a shock in imports ($t-2$). In the case of model 12, the stress scenarios are characterized by the most harmful (in the sense of (20)) scenarios out of those trust regions Ell_{τ} that correspond to the respective historical worst case stress of the macroeconomic variables in the baseline specification (see Section 2.1). The maximum (dark grey) and minimum (light grey) forecasted values of the PDs are indicated for each (non-)stress scenario.

The huge impact of model and estimation risk is also evident for the stress period calibration. Again concentrating on a shock in the industrial production, we have comparable expected non-stress PDs ranging from 4.89% to 5.39% across the models for the one-year risk horizon. As before, the discrepancies between the individual models become clearer if we focus on the stress scenarios. More specifically, a shock in the industrial production can lead to expected forecasted PDs between 6.41% (model 11) and 9.07% (model 9) for a risk horizon of one year. This is a relative increase between +23% and +77%. When we expand the risk horizon to three years, the expected stress PDs spread between 9.36% (model 5) and 14.02% (model 9) implying relative increases between + 27% and +75%. The results of shocks in the consumer sentiment corroborate these findings, but, again, the effect is smaller, in particular, when unexpected stress PDs are considered

All models are designed in such a way that it is *a priori* not clear which model is likely to be more or less severe – this is an important prerequisite for our analysis. It turns out that model 9 (VAR(2) model for the risk factors) leads, on average, to the highest stress PDs when shocks in industrial production are assumed. However, model 12 (Mahalanobis-based stress scenario) proves to be in the majority of cases the most severe one for shocks in the VIX or consumer sentiment.[42] For producing particularly low stress PDs, no clear favourite can be identified in case of a shock in the industrial production. For VIX and consumer sentiment shocks, model 2 (baseline specification without the macroeconomic index) leads to the lowest PDs – particularly when unexpected PDs are considered. However, these statements are only true when the models are calibrated on the full data sample. Only using the crisis subsample, the models 5 and 6 (both based on probit transformations) tend to produce the lowest stress PDs for all considered shocks. This high dispersion of models which lead to the most extreme results suggests that our findings are not driven by one or two outlier model specifications but are robust. Furthermore, they show that it is hardly possible to guess *ex-ante* which kind of model will produce the most conservative or least conservative stress test results.

4 Conclusions

The main question examined in this paper is whether different theoretically and empirically reasonable model specifications for credit risk stress tests can provide large differences in the

[42] We also model the Mahalanobis-based stress scenario with empirical (cross) autocorrelation of the error terms of the AR(k) processes in equation (3) instead of including the assumption of no (cross) autocorrelation. The results do not differ qualitatively.

stress test results, i.e., for forecasted stress PDs. In sum, our findings clearly suggest that seemingly minor modifications in modelling assumptions or estimation techniques can have a significant impact (in relative and absolute terms) on the stress test results. More specifically, we find that a shock in a risk factor (i.e., stress event) can materialize in expected and unexpected PD increases between +20% and +80% - depending on the stress test model selected. Furthermore, it is noteworthy that the forecasts for non-stress PDs of various models are closer to each other, i.e., non-stress PD forecasts seem to be less exposed to model and estimation risk.[43] Put differently, the processing of a shock within a model and its transmission to a stress PD seems to be the crucial part. Both, the differentiation between expected and unexpected PDs as well as the length of the risk horizon for which the PDs are forecasted seem to play only a minor role and affect the dispersion of forecasts across the various model specifications only to a limited extent. These findings emphasize the importance of extensive robustness checks and validation processes for the underlying model when interpreting the results of model-based credit risk stress tests.

Furthermore, it should be noted that the transformation of macroeconomic variables into risk parameter realizations is required in many situations. While directly employing stressed risk parameters to assess the idiosyncratic risk of a single bank might be appropriate, a standardized system-wide stress test across various jurisdictions requires more flexibility and the use of directly stressed risk parameters as given by the regulatory authorities appears not to be adequate for each bank. Some directly stressed risk parameters might be well suited for some jurisdictions or some banks, but would be inappropriate for others, for example, because of diverging business models.

[43]However, in contrast, Berg and Koziol (2017) find, using the German credit registry data set from 40 banks and 17,000 corporate borrowers, that the variability of PD estimates for the same borrower across banks is large. This finding diverging from our results might be due to the fact the variety of models and predictor variables employed across banks is much larger than the marginal modifications of our baseline model 1 that we carry out.

References

Admati, A., M. Hellwig (2013): The Bankers' New Clothes: What's Wrong with Banking and What to Do About it, Princeton University Press.

Aldrich, J., F. Nelson (1984): Linear Probability, Logit, and Probit Models, Sage Publications, Beverly Hills.

Banque de France (2009): Macro stress testing with a macroeconomic credit risk model: Application to the French manufacturing sector, Document de travail No. 238.

Basel Committee on Banking Supervision (BCBS) (2014): Reducing excessive variability in banks' regulatory capital ratios, Bank for International Settlements.

Behn, M., R. Haselmann, V. Vig (2016): The Limits of Model-Based Regulation, ECB Working Paper No. 1928.

Berg, T., P. Koziol (2017): An analysis of the consistency of banks' internal ratings, Journal of Banking and Finance, Vol. 78, pp. 27-41.

Blöchlinger, A., M. Leippold (2018): Are Ratings the Worst Form of Credit Assessment Except for All the Others, Journal of Financial and Quantitative Analysis, Vol. 53, No. 1, pp. 299-334.

Boss, M. (2002): Ein makroökonomisches Kreditrisikomodell zur Durchführung von Krisentests für das österreichische Kreditportfolio, Finanzmarktstabilitätsbericht, Oesterreichische Nationalbank.

Breuer, T., M. Jandacka, J. Mencia, M. Summer (2012): A Systematic Approach to Multi-Period Stress Testing of Portfolio Credit Risk, Journal of Banking & Finance, Vol. 36, No. 2, pp. 332-340.

Canals-Cerdá, J., S. Kerr (2015): Forecasting credit card portfolio losses in the Great Recession: a study in model risk, Journal of Credit Risk, Vol. 11, No. 1, pp. 29-57.

Committee of European Banking Supervisors (CEBS) (2010): Guidelines on Stress Testing (GL32).

Danielsson J., K. James, M. Valenzuela, I. Zer (2016): Model Risk of Risk Models, Journal of Financial Stability, Vol. 23, pp. 79-91.

Eicher, T., C. Papageorgiou, A. Raftery (2011): Default priors and predictive performance in Bayesian model averaging, with application to growth determinants, Journal of Applied Econometrics, Vol. 26, No. 1, pp. 30–55.

European Central Bank (ECB) (2014): Comprehensive Assessment Stress Test Manual, August 2014.

European Banking Authority (EBA) (2014): Methodological note EU-wide Stress Test 2014, Version 2.0, 29 April 2014.

European Banking Authority (EBA) (2016): EU-wide Stress Test 2016: Methodological Note, 24 February 2016.

European Banking Authority (EBA) (2018a): 2018 EU-wide Stress Test: Methodological Note 31, January 2018.

European Banking Authority (EBA) (2018b): Final Report on guidelines on institutions' stress testing, 19 July 2018.

Feldkircher, M., S. Zeugner (2009): Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging, IMF Working Paper 09/202.

Fenech, J.P., H. Vosgha, S. Shafik (2015): Loan default correlation using an Archimedean copula approach: A case for recalibration, Economic Modelling, Vol. 47, pp. 340-354.

Foglia, A. (2009): Stress Testing Credit Risk: A Survey of Authorities' Approaches, International Journal of Central Banking, Vol. 5, No. 3, pp. 9-45.

Frey, R., A.J. McNeil (2003): Dependent defaults in models of portfolio credit risk, Journal of Risk, Vol. 6, No. 1, pp. 59-92.

Greene, W.H. (2001): Fixed and Random Effects in Nonlinear Models, Working Paper EC-01-01, Department of Economics, Stern School of Business, New York University.

Greene, W.H. (2008): Econometric Analysis, 6th edition, Pearson, Upper Saddle River, New Jersey.

Hamerle, A., D. Rösch (2006): Parameterizing Credit Risk Models, Journal of Credit Risk, Vol. 2, No. 4, pp. 101-122.

Hale, G., J. Krainer, E. McCarthy (2015): Aggregation Level in Stress Testing Models, Federal Reserve Bank of San Francisco, Working Paper 2015-14.

Hayden, E., A. Stomper, A. Westerkamp (2014): Selection versus averaging of logistic credit risk models, Journal of Risk, Vol. 16, No. 5, pp. 39-52.

Jokivuolle, E., K. Virolainen, O. Vähämaa (2008): Macro-model-based stress testing of Basel II capital requirements, Bank of Finland Research Discussion Papers 17/2008.

Jones, S., D. Johnstone, R. Wilson (2016): Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks, Journal of Business Finance & Accounting, Vol. 44, Nos. 1-2, pp. 3-34.

Maddala, G. (1983): Limited Dependent and Qualitative Variables in Econometrics, Cambridge University Press, New York.

Madigan, D., J.C. York (1995): Bayesian Graphical Models for Discrete Data, International Statistical Review, Vol. 63, No. 2, pp. 215-232.

McNeil, A.J., J.P. Wendin (2007): Bayesian inference for generalized linear mixed models of portfolio credit risk, Journal of Empirical Finance, Vol. 14, No. 2, pp. 131-149.

Misina, M., Tessier, D., Dey, S. (2006): Stress Testing the Corporate Loans Portfolio of the Canadian Banking Sector, Bank of Canada Working Paper No. 2006-47.

Miu, P., B. Ozdemir (2009): Stress testing probability of default and migration rate with respect to Basel II requirements, Journal of Risk Model Validation, Vol. 3, No. 4, pp. 3-38.

Raftery, A.E. (1995): Bayesian Model Selection in Social Research, Sociological Methodology, Vol. 25, pp. 111-163.

Raftery, A.E., D. Madigan, J. Hoeting (1997): Bayesian Model Averaging for Linear Regression Models, Journal of the American Statistical Association, Vol. 92, No. 437, pp. 179-191.

Rawlings, J.O., S.G. Pantula, D.A. Dickey (1998): Applied Regression Analysis - A Research Tool, 2nd edition, Springer, New York.

S&P Dow Jones Indices (2018): S&P/Experian Consumer Default Rates. URL: https://us.spindices.com/documents/additionalinfo/20180619-728351/728351_experian-historical-default-rates.xlsx?force_download=true (07/04/2018).

Schechtman, R., W.P. Gaglianone (2012): Macro stress testing of credit risk focused on the tails, Journal of Financial Stability, Vol. 8, No. 3, pp. 174-192.

Siemsen, T., J. Vilsmeier (2018): On a quest for robustness: about model risk, randomness and discretion in credit risk stress tests, Bundesbank Discussion Paper 31/2018.

Simons, D., F. Rolwes (2009): Macroeconomic Default Modeling and Stress Testing, International Journal of Central Banking, Vol. 5, No. 3, pp. 177-204.

Sorge, M., K. Virolainen (2006): A comparative analysis of macro stress-testing methodologies with application to Finland, Journal of Financial Stability, Vol. 2, No. 2, pp. 113-151.

Standard & Poor's (2011): Default, Transition, and Recovery: 2010 Annual Global Corporate Default Study And Rating Transitions.

Standard & Poor's (2014): Default, Transition, and Recovery: 2013 Annual Global Corporate Default Study And Rating Transitions.

S-Rating und Risikosysteme (2018): Tätigkeitsbericht 2017.

Traczynski, J. (2017): Firm Default Prediction: A Bayesian Model-Averaging Approach, Journal of Financial and Quantitative Analysis, Vol. 52, No. 3, pp. 1211-1245.

Trenca, I., A. Benyovszki (2008): Credit Risk, a Macroeconomic Model Application for Romania, Finance - Challenges of the Future, Vol. 1, No. 7, pp. 118-126.

U.S. Census Bureau (2017): X-13ARIMA-SEATS Reference Manual Accessible HTMLOutput Version, Version 1.1.

Virolainen, K. (2004): Macro stress testing with a macroeconomic credit risk model for Finland, Bank of Finland Discussion Papers.

Wilson, T.C. (1997a): Portfolio credit risk, Part I, Risk, Vol. 10, No. 9, pp. 111-117.

Wilson, T.C. (1997b): Portfolio credit risk, Part II, Risk Vol. 10, No. 10, pp. 56-61.

Wong, J.H.-Y., K.-F. Choi, P.-W. Fong (2008): A framework for stress-testing banks' credit risk, Journal of Risk Model Validation, Vol. 2, No. 1, pp. 3-23.

Zedginidze, Z. (2012): Linking Macroeconomic Dynamics to Georgian Credit Portfolio Risk, EERC Working Paper Series, EERC Research Network No. 12/07e, Russia and CIS.

Zellner, A. (1986): On assessing prior distributions and Bayesian regression analysis with g prior distributions, Bayesian Inference and Decision Techniques – Essays in Honor of Bruno de Finetti, P. Goel and A. Zellner (eds.), Elsevier, Amsterdam, pp. 233-243.