Model and estimation risk in credit risk stress tests

Peter Grundke[1], Kamil Pliszka[2], Michael Tuchscherer[3]

This version: 30 April 2015

Abstract:

Stress testing is often performed in a model-based (implicit) way, i.e. adverse realizations of risk factors (e.g., macroeconomic factors) derived from a specific scenario need to be translated with the help of a quantitative model into adverse risk parameter realizations (e.g., default probabilities, default correlations). Usually, these are the same models that are also employed by banks for Pillar 2 risk coverage calculations. In this paper, we focus on credit risk and show how exploiting leeway when setting up and implementing the underlying 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 (including topical approaches like Bayesian model averaging). Our findings show that seemingly only slightly differing specifications can lead to entirely different stress test results. This emphasizes the importance of extensive robustness checks.

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

JEL classification: G21, G28, G32

[1] Osnabrück University, Chair of Banking and Finance, Katharinenstraße 7, 49069 Osnabrück, Germany, Phone: +49 (0)541 969 4720, Fax: +49 (0)541 969 6111, E-mail: peter.grundke@uni-osnabrueck.de.

[2] Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main, Germany, Phone: +49 (0)69 9566 6815, E-mail: kamil.pliszka@bundesbank.de. The views expressed in this paper are those of the author and do not necessarily reflect those of the Deutsche Bundesbank or its staff.

[3] Osnabrück University, Chair of Banking and Finance, Katharinenstraße 7, 49069 Osnabrück, Germany, Phone: +49 (0)541 969 6115, Fax: +49 (0)541 969 6111, E-mail: michael.tuchscherer@uni-osnabrueck.de.

1 Introduction

Since the outbreak of the financial crisis in 2007-2009 and the ongoing European sovereign debt crisis, the importance of stress tests for financial institutions has enormously increased. First, standards for bank-individual stress tests as part of the requirements of the Pillar 2 of Basel II have significantly been extended (see FSA (2008, 2009), CEBS (2010), BIS (2011)). Second, regulatory authorities, such as the Federal Reserve or the European Banking Authority (EBA), have carried out system-wide stress tests to analyze the vulnerability of the largest banks in the financial system.

Depending on the risk type and the aim of a stress test, it can include model-based elements. If, for example, one is interested in the loss of an equity portfolio that would occur if all stock values decreased by a fixed percentage, no model is needed. The same is true of computing losses of fixed-income portfolios that would occur when there is a parallel shift in the term structure of risk-free interest rates of, for example, 200 basis points.

However, in several cases a bank is required to translate the impact of an economic shock into its risk parameters. Examples include the Basel II credit risk framework where IRB banks have to reflect economic downturns in their risk parameters even in Pillar 1 (see article 177 in the 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. A more topical example is the EBA stress test 2014 in the eurozone, where banks could either translate a prescribed economic downturn scenario into their risk parameters or could directly employ the parameter values provided by the EBA.

Moreover, a bank is also often interested in the impact of stress scenarios in their internal measurement system. If a bank wants to calculate the economic capital requirement (e.g., measured by the value-at-risk or by the expected shortfall) that is necessary in an adverse scenario with increased stock return volatilities and increased stock return correlations, a model that translates the increased values for the volatilities and correlations into stressed risk measure values is needed. In the field of credit risk stress tests, a model is also required when analysing the effects of increased default probabilities (PD), loss given defaults (LGD) and increased default correlations (or asset return correlations) on the economic capital requirements. If a bank is only interested in the increase in expected credit losses in adverse economic scenarios (e.g. over a risk horizon of one year), correlations do not matter and a model would not be necessary to carry out a stress test when the bank explicitly assumes a specific increase in all PD and LGD values as the stress scenario. A possible strategy might be to take the largest percentage change in these values that have been observed in the past and to apply them to the current levels. If, however, a bank wants to test the effect of a forecasted baseline or adverse scenario for the economy (measured by some economic indicators, such as GDP growth, unemployment rate or inflation rate) on the expected credit portfolio loss, it needs a model to translate the economic indicator forecasts into modified PD and LGD values.[4]

A model for either translating adverse risk factor realizations (corresponding to an assumed adverse scenario) into stressed risk parameters or to translate explicitly stressed risk parameters into stress test results (such as stressed value-at-risks or expected shortfalls) adds a layer of complexity and, thus, provides leeway regarding modelling assumptions and the applied estimation techniques which are likely to affect the stress test results. From the perspective of the regulatory authorities, it is crucial to know how large the potential to manipulate the results of stress tests by choosing favourable modelling and estimation techniques is for banks. As failed internal or external stress tests may force a bank to increase its equity and banks usually consider equity to be expensive,[5] 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.[6]

In this paper, we analyze for a specific risk type (credit risk) and for a specific objective of a stress test (expected losses and partly risk measure values) how large multi-period stressed PD values can vary depending on the employed modelling assumptions and estimation techniques. To achieve this, starting from a base model specification, we employ several variations of a CreditPortfolioView (CPV)-style model.[7] All variations are statistically sound approaches and it is ex-ante not obvious why one specification or estimation technique should be more adequate than another. However, as we show, the chosen model specifications and the employed estimation techniques can hugely influence the results for the stressed default

[4] For the macro stress tests performed by the EBA, this is exactly what banks had to do (unless they want 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 scenario and adverse scenario (see EBA (2014), ECB (2014)).
[5] See Admati and Hellwig (2013) for an extensive discussion of supposedly expensive bank equity.
[6] For example, there are indications that banks use the degrees of freedom within internal ratings-based approaches in such a way that the volume of risk-weighted assets and, hence, the regulatory capital requirements decrease (see BIS (2014) or Behn et al. (2014)).
[7] See Wilson (1997a, 1997b).

probabilities. These results show the importance of extensive robustness checks for the underlying model when interpreting the results of credit risk stress tests.

The remainder of the paper is structured as follows: Section 2 provides a short review of the credit risk stress testing literature. Section 3 presents the methodology of the analysis and Section 4 shows the results. Section 5 discusses potential shortcomings and extensions. Section 6 concludes.

2 Literature Review

We divide the review of the credit risk stress test literature into six fields: First, a large body of the literature deals with implicit stress tests within CPV-style models (or extensions thereof).[8] These papers look for macroeconomic variables that can explain the systematic variation of default rates across time and, afterwards, these macroeconomic variables are stressed to compute stressed 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 macro variables and credit risk. Second, in another strand of literature, (asymptotic) confidence intervals of statistically estimated risk parameters (such as PD or asset return correlations) are used as the base for deriving extreme, yet plausible realizations of these risk parameters in adverse scenarios (see, for example, Rösch and Scheule (2007) or Höse and Huschens (2008)). Third, literature deals with model and estimation risk in credit portfolio models. For example, Frey and McNeil (2003) evaluate the impact of various dependence structures of defaults on risk measures. Hereby, model specifications with larger tail dependencies could be interpreted as some kind of stress analysis for credit portfolio risk measures. Other paper conduct simulation studies and compare risk measures for various credit portfolio models (see, for example, Hamerle and Rösch (2006), Han (2014)). Fourth, stress tests for credit risk concentrations are carried out. For example, Bonti et al. (2006) employ a multi-factor credit portfolio model (similar to CreditMetrics™) for a stress test on sector credit risk concentrations. To achieve this, they restrict the support of the probability distribution of specific systematic risk factors to adverse realizations. Fifth, multi-risk stress test approaches are proposed. For example, Drehmann et al. (2010) present an integrated bank model for a simul-

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

taneous stress test of credit and interest rate risk. Sixth, since a few years, banks have been obliged to carry out so-called reverse stress tests. While in regular stress tests, adverse scenarios are chosen on the basis of historical observations or expert knowledge (or both) and, afterwards, the consequences of these scenarios for the target indicator of the stress test ((expected) losses, regulatory or economic capital, liquidity) are analyzed, reverse stress tests do it the other way round. In reverse stress tests, exactly those scenarios are looked for that make a bank's business plan unviable and cause the bank to cross the frontier between non-default and default. In the next step, the most plausible of these scenarios has to be found (Grundke and Pliszka (2015)). Literature on reverse stress testing is still relatively sparse (see, for example, Grundke (2011, 2012) or Grundke and Pliszka (2015) and the papers cited therein). In a strand of literature related to reverse stress testing, the worst (in the sense of 'expected losses for a given portfolio') scenario from a set of scenarios with a given plausibility (for example, measured by the Mahalanobis distance) is looked for. Examples of this approach include Breuer et al. (2008, 2010, 2012).

3 Methodology

In the following, first, the base model for predicting stressed default probabilities is introduced. Second, various modifications of this base model 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 4.2, the specifications can hugely influence the results for the stressed default probabilities.

3.1 Base model

As the base model, we employ a CreditPortfolioView (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

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

linearly depends on some macroeconomic variables $x_{i,t}$, $i \in \{1, 2, ..., I\}$. The macroeconomic index $y_{s,t}$ is assumed to be related to the sector-specific default probability $p_{s,t}$ by a logit transformation:

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

The macroeconomic risk factors $x_{i,t}$, $i \in \{1, 2, ..., I\}$, are modelled by autoregressive processes of k_i-th order (AR(k_i) process):

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

To avoid overfitting, we restrict our search for an adequate time series model to AR(k) processes with a maximum order of $k = 2$. First, we apply the AIC (Akaike Information Criterion) and the BIC (Bayesian Information Criterion) to choose the appropriate number of lags.[9] Second, insignificant parameters $\gamma_{i,j}$ ($i \in \{1, 2, ..., I\}$, $j \in \{1, 2\}$) are set to zero (significance level of 10%) and the significance of the parameters of the remaining part of the processes is checked again.[10] As we do not center our risk factors, an intercept $\gamma_{i,0}$ is always employed in order to retain unbiased estimates, even if it is not significant. The same is done for the index equation (1).[11]

When the Godfrey-Breusch test indicates that the null hypothesis of no autocorrelation (up to order 4) of the error terms $v_{i,t}$ 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 ordinary least squares (OLS) parameter estimates. The same is carried out for the estimation of the index equation (1).

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:[12]

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

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

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

[9] The AIC and the BIC did not contradict each other in any case. Thus, prioritization was not necessary.
[10] See, for example, Banque de France (2009) for a similar approach.
[11] Only in the model based on Bayesian model averaging (model 9 in the following) no intercept is used.
[12] The assumed multivariate distribution of the error terms influences the probability distributions of the stressed default probabilities. For an alternative, 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.

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

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 Monte Carlo simulation algorithm with D simulation runs:[13]

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 $p_{s,n}^{(d)}$, $s \in \{1, 2, ..., S\}$, based on $u_{s,n}^{(d)}$ and the forecasts for the macroeconomic variables $x_{i,n}^{(d)}$.

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

To compute distributions for stressed 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 one or several other error terms. If Y is an r-dimensional normally distributed random vector with the following partitioning:[14]

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

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

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

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 base setting, only one risk factor is shocked and the shock is set equal to the historical realization of the error term which had the most negative impact on the macroeconomic index.[15] More precisely, we define the shocked component $Y_2 = v*_{i,T+1}$ by

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

When we have $S = 1$ (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 $m = 3$ and we nearly always[16] assume that there is a univariate shock only in the first future period and that in the subsequent two 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.[17] To achieve high accuracy in the Monte Carlo simulation, we employ $D = 1{,}000{,}000$ draws.

[15] See Boss (2002, pp. 82-83).

[16] The exceptions are models 11 and 12, where the stress scenario is based on the Mahalanobis distance (see Section 3.3.5.2).

[17] Even if the stressed risk factor is only modelled by an intercept plus an error term (which is done in some cases; see Table 3), the initial shock propagates into the next periods because the realization of the macroeconomic index in a specific period is the sum of the initial index realization in 2010 and the modelled stressed and unstressed changes in the index in the previous periods. Furthermore, due to the correlation of the risk factors, those risk factors that are not explicitly stressed in the first future period 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.

3.2 Data and variable selection

As the data input for estimating (1), we use yearly US-default rate data from Standard & Poor's ranging from 1983 to 2010.[18] We only employ default rate data for speculative grade obligors because defaults of investment grade obligors are very rare and, hence, default rates are near zero and hardly fluctuate over time. Furthermore, we do not differentiate between various sectors. Thus, we have $S = 1$. In practice, banks could use their internal sector-specific default rate data with a shorter periodicity (e.g., quarterly data). However, to ensure that the data are representative, banks will probably only use data from the most recent years so that short data samples remain as a statistical problem.

As in Kalrai and Schleicher (2002, pp. 71-75) for Austrian data, economic activity indicators, price stability indicators, household indicators, firm indicators, financial market indicators and further external indicators for the US are considered to be potential explanatory variables (see Table 1). The data is taken from Datastream. For each variable, average values per year are computed.[19]

– insert Table 1 about here –

From the comprehensive set of potential explanatory variables, the most important 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 (multivariately) significant. In these studies, an economic indicator (e.g., GDP) and an interest rate are often employed.[20] To essentially avoid ad-hoc elements in the selection procedure for the explanatory variables, we apply the stepwise regression upon those variables out of the set of potential explanatory variables that are univariately significant.[21] In detail, the selection procedure works as follows: First, we include, if univariately significant, the GDP in the model. If the GDP proves not to be univariately significant, we add the variable with the highest ab-

[18] See Standard & Poor's (2011). The data was adjusted for rating withdrawals. As an alternative to default rates provided by rating agencies, insolvency rates or the fraction of non-performing loans (NPLs) to all loans could be used.

[19] See similarly Boss (2002, p. 76).

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

[21] 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). Based on a bootstrap analysis, they advocate Bayesian model averaging as an alternative to stepwise model selection procedures that are frequently used in practice. As GDP is used as explanatory variable in all of our model specifications based on stepwise regression (when it is univariately significant and the transformed variable is stationary), we do not completely avoid ad-hoc selection elements (see the following description of the selection procedure and Figure 1).

solute t-value.[22] Afterwards, we maximize the adjusted R^2 by adding univariately significant macroeconomic variables to the model. A prerequisite for adding a variable (to avoid (imperfect) multicollinearity) is that the absolute value of its correlation with any of the other variables that have already been included in the model is below 0.8. If the added variable leads to insignificance in some of the earlier added variables, the new specification without the insignificant variables will be compared with the specification before adding the last variable and the one with the higher adjusted R^2 will be used. If the adjusted R^2 cannot be increased further by adding new variables, the stop criterion is reached. Figure 1 visualizes this procedure. As an alternative to the stepwise regression, we also employ the Bayesian model averaging (see Section 3.3.4).

– insert Figure 1 about here –

To ensure stationarity of the time series of the macroeconomic index and of the explanatory variables, we apply various tests. First, based on a t-test, we check the significance of an intercept and a time trend (significance level 10%) for each time series. Then, the Augmented Dickey-Fuller (ADF), Phillips-Perron (PP) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test are applied. An intercept and/or a time trend are only considered within these tests when they proved to be significant in the first step. 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%. In cases where a unit root in the characteristic equation of a time series model is present, we calculate the first differences. If there is still a unit root, we calculate the second differences.

3.3 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 statistically equally reasonable modifications of the base specification influence the results for the stressed default probabilities. Table 2 summarizes the specification of the base model and gives an overview of the considered modifications that are discussed in this section. To facili-

[22] For our data, we have two models without GDP as explanatory variable (see Table 4 in Section 4.1). First, this is model 2 (returns) because in this case the GDP is not univariately significant. Second, it is model 3 (log-returns) because the log-return of the GDP is not stationary and computing the log-log-return is not possible due to negative numbers.

tate comparisons, in each case only a single modification (compared to the base model) is considered, but no simultaneous modifications.

– insert Table 2 about here –

For each modification where it seems sensible,[23] the stepwise regression approach described above has to be applied once again to select the most appropriate explanatory variables in each case.

3.3.1 Stationary methods

To ensure stationarity of the data, we usually compute first differences of the data points in the base model specification. This technique is also frequently applied in the CPV literature.[24] However, other techniques are statistically equally reasonable, for example, computing returns or log-returns.[25] Changing the time series transformation to achieve stationarity of the data causes some variables in the base model to become insignificant. As only significant explanatory variables shall be used, we repeat the stepwise regression for each transformation type. Again, we test for stationarity of the transformed data using the ADF, PP and KPSS test. As in the base case, conflicting test results were possible. Thus, we again applied the “two out of three” rule described above (see Section 3.2). When the return transformation did not yield stationary data, the second return (return of the return) was computed.[26]

3.3.2 Macroeconomic index process

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

3.3.2.1 Time-lagged variables

In this modification, first, we consider one and two period time-lagged macroeconomic variables $x_{i,t-1}$ and $x_{i,t-2}$, $i \in \{1, 2, ..., I\}$, as potential explanatory variables in (1). This approach enables us to take into account a delayed impact of macroeconomic variables on the default

[23] The risk factors from the base model are retained for the model with fixed AR(2) processes for the risk factors (model 8), the model based on Bayesian model averaging (model 9) and the models with modified stress scenarios (models 10 to 12),
[24] See, for example, Boss (2002, p. 73).
[25] We apply each stationary method to the data of the explanatory variable as well as to the macroeconomic index data.
[26] When the log-return transformed time series of a variable exhibits a unit root, this variable is excluded because, due to negative numbers, it is not possible to calculate the log-returns of the log-return transformed time series. In particular, this was the case for the GDP.

rate.[27] Second, one and two period time-lagged realizations of the macroeconomic index y_{t-1} and y_{t-2} are introduced as potential explanatory variables in (1).[28] For both model modifications, the stepwise regression is repeated to select the most appropriate (multivariately) significant explanatory variables.

3.3.2.2 GLS estimator

The OLS estimator is an efficient estimator only in the case of homoscedastic and serially uncorrelated error terms. Even in the case that the Godfrey-Breusch test would not reject the null hypothesis of no autocorrelation (up to a specific order), it may be appropriate to consider autocorrelation as the non-rejection of a null hypothesis is of course not an unambiguous approval of its adequacy, particularly for small data sets. Thus, as a further modification, we apply the generalized least squares (GLS) estimator as an alternative to account for potential autocorrelation of the error term in the index equation (1). The GLS estimator basically assumes a more flexible structure of the variance-covariance matrix of the error terms:

$$Var\left(uu'\right) = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,T} \\ \sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T,1} & \sigma_{T,2} & \cdots & \sigma_T^2 \end{pmatrix}. \tag{9}$$

More specifically, we assume that the error term of the macroeconomic index equation (1) follows an AR(1) process without intercept:[29]

$$u_t = \rho \cdot u_{t-1} + \delta_t \tag{10}$$

where the error term δ_t is normally distributed and uncorrelated with all other error terms of the model. To determine the model specification in this case, again, the stepwise regression is repeated.

3.3.2.3 Probit function

In the base 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, indeed, is not the only possible choice. One alternative is using the probit transformation:[30]

$$p_t = \Phi\left(-y_t\right) \Leftrightarrow y_t = -\Phi^{-1}\left(p_t\right) \tag{11}$$

---

[27] See, for example, Boss (2002) for a similar approach.
[28] As we have $S = 1$, we omit the sector index s in the following.
[29] See McNeil and Wendin (2007) or Miu and Ozdemir (2009) for a similar procedure.
[30] 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 (11) to ensure that – as in the case of the logit transformation – increasing index values cause decreasing default probabilities. Again, the stepwise regression is repeated for this model specification.

3.3.3 Risk factor processes

In the base specification, the evolution of the macroeconomic risk factors over time is explained by a first or second-order autoregressive process, but only statistically significant parameters are employed (except the intercept). 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.[31] In some cases, only an intercept plus the error term remain, meaning that no autoregression at all is considered. In this section, we want to check for the influence of this assumption on the stressed default probabilities. For this, we employ an AR(k) process of fixed order $k = 2$, regardless of the significance of the estimated parameters.[32]

3.3.4 Variable selection procedure

Up to now, we have chosen risk factors upon a combination of expert judgment and stepwise regression. A pure approach of one of those methods might lead us to unfavorable results. On the one hand, if we would have relied solely on the stepwise regression, some of our considered model specifications would not contain economic indicators or other desirable variables. On the other hand, purely trusting in expert judgment would lead us to insignificant risk factors which could barely be justified. In order to ensure reasonable but significant variables, we have applied a mixture of both in this paper: We assumed that the models should contain the GDP (unless it is univariately insignificant or the data transformation to ensure stationarity is not possible), but this assumption was liaised with the stepwise regression.

However, our approach still bears model risk as we end up with a single (base) model and do not consider the impact that we potentially added the wrong risk factors. This can be tackled by the Bayesian model averaging where we consider all risk factors but include only those with a sufficient high likelihood. Simulation studies show that the Bayesian model averaging delivers a better forecast performance than other approaches what makes this technique popular (see, e.g., Raftery et al. (1997), Hayden et al. (2014)).

[31] See Table 3 in Section 4.1.
[32] For a further alternative, see, for example, Hamerle et al. (2008), who model their two explanatory risk factors using a VAR process.

The idea of the Bayesian model averaging is to calculate for a given number O of candidate risk factors, in our case 22 variables as shown in Table 1, all linear models M_l, $l \in \{1,...,2^{22}\}$ consisting of subsets of the risk factors and, then, to calculate the expected value of the coefficient vector $\beta_{BMA} \in \mathbb{R}^{O\times 1}$ as the weighted sum over all models

$$E\left(\beta_{BMA}\middle|y\right) = \sum_{l=1}^{2^{22}} \hat{\beta}_l \cdot P\left(M_l\middle|y\right) \tag{12}$$

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

Instead of including all risk factors, only those which prove to be sufficiently likely will be included. 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

$$PIP := P\left(\beta_h\middle|y\right) = \sum_{l=1}^{2^{22}} P\left(\beta_h\middle|M_l\right) \cdot P\left(M_l\middle|y\right). \tag{13}$$

We follow the suggestion of Raftery (1995) of including only risk factors with a PIP of at least 50%. 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 $P\left(M_l\right)$

$$P\left(M_l\middle|y\right) \propto P\left(y\middle|M_l\right) \cdot P\left(M_l\right). \tag{14}$$

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

$$\beta\middle|g \sim N\left(0, \left(\frac{1}{g}\Gamma'\Gamma\right)^{-1}\right) \tag{15}$$

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:

$$P\left(y\middle|M_l\right) \propto (1+g)^{-\frac{o_l}{2}} \cdot \left(1 - \frac{g}{1+g} \cdot R_l^2\right)^{-\frac{T-1}{2}} \tag{16}$$

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 $g = T$.[33]

Evaluating all models $P\left(M_l \middle| y\right)$, $l \in \{1,...,2^{22}\}$, what means that in our case we would have to conduct 4,194,304 regressions, often proves to be computationally too intricate. In order to overcome this issue, we employ the Markov chain Monte Carlo sampler (see, e.g., Madigan and York (1995)).

3.3.5 Stress test scenarios

The modifications described in this section do not concern statistical issues, but deal with a degree of freedom that risk managers performing stress tests have, namely the choice of the stress test scenario. For these modifications, the base model is employed.

3.3.5.1 Hypothetical scenario based on three standard deviations

In the base specification, the largest historical deviation of the empirical observations from the theoretical model for the macroeconomic risk factors with a univariately negative impact on the macroeconomic index is employed as the stress scenario. Now, alternatively, the impact of a given shock on the error term of three standard deviations is taken into account.[34]

3.3.5.2 Hypothetical scenario based on the Mahalanobis distance

In this modification, a multivariate and multi-period stress test scenario based on the Mahalanobis distance of the error terms v_i, $i \in \{1,2,...,I\}$, is used.[35] The Mahalanobis distance of a random vector v is defined as:

$$Maha(v) = \sqrt{\left(\mu - v\right)' \cdot \Sigma^{-1} \cdot \left(\mu - v\right)} \tag{17}$$

where $\mu = E\left[v\right]$ 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

[33] 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)).
[34] Three standard deviations is a frequent choice (see, for example, Breuer et al. (2012, p. 337)).
[35] See, for example, Breuer et al. (2012) for the use of the Mahalanobis distance for stress testing.

ellipticity – is the respective realization. The Mahalanobis distance is employed to define so-called trust regions of radius τ around $\mu = E\left[v\right]$:

$$Ell_{\tau} := \left\{ v \in \mathbb{R}^{I \cdot m} \middle| Maha(v) \leq \tau \right\} \tag{18}$$

As we consider a dynamic three-period stress test ($m = 3$), 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 $I \cdot m = I \cdot 3$. The random vector v represents an I-dimensional path of the error terms of the macroeconomic risk factors over the three considered 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)). Furthermore, we differentiate between an assumed non-existence of serial (cross) correlation of the error terms and a model specification in which the empirical (cross) autocorrelations are employed to define the variance-covariance matrix of the random vector $v \in \mathbb{R}^{I \cdot m}$.

Using the above notation, the three standard deviation stress scenarios in Section 3.3.5.1 can be represented by $v_1^* = (3\sigma_{v_1}, 0, \ldots, 0)'$, …, $v_I^* = (0,...,0,3\sigma_{v_I}, 0, \ldots, 0)' \in \mathbb{R}^{I \cdot m}$ with corresponding values $Maha(v_i^*)$ ($i \in \{1,...,I\}$). To ensure consistency between the univariate stress scenarios as set out in Section 3.3.5.1 and those ones employed in this section, we define trust regions Ell_{τ_i} by setting $\tau_i = Maha(v_i^*)$ ($i \in \{1,...,I\}$).[36] Thus, the stress scenarios used in this and in the previous section are equally plausible in the sense of the Mahalanobis distance. However, the stress scenario used in this section defines a multivariate and multi-period shock, whereas the other stress scenarios (historical worst case, three standard deviations) only imply a univariate initial shock in $T+1$. Out of each of the trust regions Ell_{τ_i} ($i \in \{1,...,I\}$), we look for the scenario which minimizes the expected sum of the changes in the macroeconomic index over the three forecasted periods:

$$v_{\tau_i}^{worst} = \underset{v \in Ell_{\tau_i}}{\arg\min} \left\{ E\left[\sum_{m=1}^{3} \Delta y_{T+m}(u_{T+m}, v) \middle| F_T, v \right] \right\} \tag{19}$$

where F_T contains all past information up to time T (in particular about the previous realizations of the macroeconomic risk factors). As decreasing values for the macroeconomic index cause increasing default probabilities, the solution of the optimization problem (19) nearly corresponds to the worst case scenario (for a given plausibility) in the sense of expected de-

[36] The trust regions differ depending on whether empirical (cross) autocorrelations are considered or not (see Section 4.2).

fault probabilities. Of course, due to Jensen's inequality and the non-linear transformation between the macroeconomic index and the default rate, this is not exactly true.

4 Results

In this section, first, we present the results for the model specifications and, second, we show the consequences that differing model specifications have for the stress test results.

4.1 Model specifications

Tables 3 and 4 show the estimation results for the time series processes of the risk factors and for the macroeconomic index equation (1).

– insert Tables 3 and 4 about here –

Using the information criteria AIC and BIC for selecting the order of the AR processes of the risk factors and employing only significant parameter estimates (except the intercept; as described in Section 3.1), we effectively obtain various model specifications (see Table 3). We use first and second-order autoregressive processes and, in some cases, we even describe the evolution of the risk factors over time only by an intercept plus the error term. The specification of the AR processes has an influence on how long it takes until an initial shock vanishes. The goodness-of-fit is satisfying as the coefficient of determination R^2 ranges from 15% to 69% and the values for the adjusted R^2 are between 11.5% and 66%.

Having applied the stepwise regression approach, we include the first differences in the GDP and the second differences in Moody's commodity index as explanatory variables in the base model (see Table 4). The explained variance of the model is 41.8% and the adjusted R^2 is 36.9%. We also tested in addition to the 3-months Treasury bill rate as shown in Table 1 other short-term and long-term interest rates as potential explanatory variables, but they were neither univariately nor multivariately significant. The positive signs of the explanatory variables in the base model 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)). As an increase in GDP as well as higher prices for commodities can usually be observed in economically good times due to the rise in demand, the estimated signs of the explanatory variables are in line with our intuition. In the modified models 2 to 7, this is mostly (but not always) also the case. Apart from model 2 (returns) and model 3 (log-returns), GDP is included in all models based on stepwise regression.[37] The adjusted R^2 ranges from 27.3% to 49.3%. The best fit in terms of the adjusted R^2 shows model 4 with time-lagged risk factors as additional explanatory variables.

Table 5 summarizes the results of the variable selection procedure by Bayesian model averaging. The posterior inclusion probability (PIP) shows the fraction of models in which a variable was included. The mean and the standard deviation refer to the average coefficient values across all evaluated models during the model averaging. The conditional positive sign is the fraction of models in which the variable had a positive sign given that the variable was included in the model. It quantifies the impact direction across all considered models on the macroeconomic index given a risk factor shock and, thus, indicates whether a solely univariate interpretation is valid. For example, an increase in the industrial production had in 99.98% of all considered models a positive impact on the macroeconomic index which would lead to a decreased default probability, accordingly. Moreover, the industrial production outweighs the other variables by far and is included in 80% of all models.[38] As it is the only risk factor with a PIP above 50%, we include no further variables. This is a noteworthy observation as industrial production has neither been chosen in the base model nor in any modified model based on stepwise regression. Moreover, we even had to use the second differences and, thus, lost information for assuring stationarity. It is also interesting that the often applied variable GDP has a relatively low inclusion probability.

– insert Table 5 about here –

Figure 2 shows the realized default rates compared with the in-sample and out-of-sample predictions of the default probabilities (based on (1) and (2)). For the in-sample prediction, the observed risk factor realizations of each model are inserted into the respective index equation (1), the error term is set equal to its mean zero and the calculated realizations of the macroeconomic index are inserted into (2), which yields the predicted default probabilities. For the out-of-sample prediction for the years 2011 to 2013, the mean forecasted default probabilities

[37] In model 2 (returns), the GDP is not included because it is not univariately significant (see our variable selection procedure in Figure 1). However, together with the risk factor consumer confidence, the GDP is multivariate significant (results are not reported). In model 3 (log-returns), the GDP is not included because the log-return of the GDP is not stationary.

[38] We tested several alternative g priors and simulation settings which led qualitatively to the same results, i.e. we only included the industrial production in the model. Moreover, the setting $g = \max\{t; K^2\}$ resulted in a posterior inclusion probability of 91.33% for the industrial production, whereas the variable with the second highest posterior inclusion probability, the 3-months Treasury bill rate, had only a posterior inclusion probability of 13.33%.

in the non-stress case are employed (see Table 7 in the following).[39] As can be seen, for most model specifications, the in-sample prediction is not too bad, but the out-of-sample prediction is poor. This is also confirmed by the results exhibited in Table 6. Based on 1 million forecasts of the default rates for 2011 to 2013, Table 6 shows the mean deviation between the forecasted default probabilities and the realized default rates for each year $n \in \{T+1, T+2, T+3\}$ (MD_n), the mean squared error for each year $n \in \{T+1, T+2, T+3\}$ (MSE_n) and the cumulative mean squared error over all three years ($CMSE$):

$$MD_n = E\left[p_n - PD_n^{realized} \middle| F_T \right], \tag{20}$$

$$MSE_n = E\left[\left(p_n - PD_n^{realized} \right)^2 \middle| F_T \right], \tag{21}$$

$$CMSE = \sum_{m=1}^{3} E\left[\left(p_{T+m} - PD_{T+m}^{realized} \right)^2 \middle| F_T \right] \tag{22}$$

where F_T denotes the available information up to time T. The only good news on the out-of-sample performance of the model is that in most cases, we observe an overestimation of the realized default rates in 2011 to 2013.[40] The best performing models (in terms of (cumulative) mean squared errors) are those with time-lagged explanatory variables (models 4 and 5).[41]

– insert Figure 2 and Table 6 about here –

4.2 Stressed default probabilities

Based on the estimated risk factor processes and the macroeconomic index equations, stressed default probabilities are forecasted three periods ahead. Depending on the method to make the data stationary, the macroeconomic index y_n (and, hence, the default probability p_n) in each period $n \in \{T+1, T+2, T+3\}$ is computed out of the dependent variable as follows:

Model 1 (base model): $y_n = y_{n-1} + \Delta y_n$,

Model 2 (returns): $y_n = y_{n-1} \cdot \left(1 + R_n\right)$,

Model 3 (log-returns): $y_n = y_{n-1} \cdot e^{R_n^{\ln}}$

where $\Delta y_n = y_n - y_{n-1}$ is the first difference, $R_n = \left(y_n - y_{n-1}\right) / y_{n-1}$ is the return and $R_n^{\ln} = \ln\left(y_n / y_{n-1}\right)$ is the log-return of the index values.

[39] The realized default rates for the years 2011 to 2013 are taken from Standard & Poor's (2014).
[40] See the following Section 4.2 for a detailed discussion of this observation in the case of the base model 1.
[41] In model 5, the two period time-lagged first difference in the macroeconomic index is included as an explanatory variable. However, for this, we treated the univariately significant second lag of the index as the GDP in the stepwise regression approach (see Figure 1). Otherwise, it would not have become part of the model.

The main question examined by this paper is whether different empirically reasonable model specifications for credit risk stress tests can yield large differences in the stress test results. Tables 7 to 9 and Figures 3 and 4 give a clear answer to this question: Yes, they can. As we can see, the forecasted expected default probabilities and the 99.9% quantiles of the probability distribution of the forecasted default probabilities can differ considerably across the model modifications. In addition to these differences in the level of the forecasted default probabilities, there are also differences in the variation over time. In some specifications (e.g., in the base model 1), the forecasted default probabilities (expected value as well as 99.9% quantile) increase from period $T+1$ to $T+2$ and decrease from period $T+2$ to $T+3$. In other specifications (e.g., in model 2 (returns) or 3 (log-returns)), an increase in the forecasted default probabilities over all three periods can be observed.

The results for the base model 1 are very surprising. First, the large expected default probabilities even in the non-stress case are remarkable. This is even more evident when one considers that the last observed default rate in 2010 was 3.4% and that the largest default rate in the whole time period 1983 to 2010 was 11.6%.[42] Second, due to the high level of the forecasted default probabilities in the non-stress case, in most model specifications, the forecasted stress default probabilities are smaller than the forecasted default probabilities in the non-stress case of the base model 1. A detailed analysis shows that the main reason for the high level of the forecasted default probabilities in the non-stress case of the base model are the realizations of the second differences of Moody's commodity index in 2009 and 2010. These are very low (-573.64 in 2009) and very high (1,679.16 in 2010), respectively. In 2009, this value corresponds to a difference of -1.67 times the standard deviation from the mean. In 2010, this difference is as much as 4.32 times the standard deviation.[43] Through the AR(2) process, by which the second differences of Moody's commodity index are modelled, these extreme values cause extreme forecasts for the index in later periods. For example, in $T+2$ (corresponding to 2012), the expected forecast of the second difference of the index is -648.13. This corresponds to a difference of -1.87 times the standard deviation from the mean of the second differences of the index. This explains the very high expected default probability of 14.7% in $T+2$ in the non-stress case of the base model. Without the data points of

[42] The minimum value is 1.0%, the mean default rate is 4.6% and the standard deviation is 2.9% (see Table 1).
[43] Of course, this 'problem' could be solved by assuming that this data point is an outlier and by eliminating it from the sample. However, when calibrating models that are used for stress testing, one has to be careful with eliminating presumed outliers, because, otherwise, the calibrated model potentially cannot produce sufficiently harmful events later on.

2009 and 2010, the expected default probabilities in the non-stress case of the base model would be 4.16% ($T+1$), 4.83% ($T+2$) and 6.09% ($T+3$) and, hence, much less extreme than those results that we get when we use the full data sample to estimate the base model.[44] These observations show how sensitive the forecasted (stressed and non-stressed) default probabilities are with respect to the chosen time period upon which the models are calibrated. Of course, this sensitivity is due to the sample length that can usually be used for calibrating the models, which is very short anyway.

– insert Table 7 and Figure 3 about here –

In model 4, where the one period time-lagged first difference in the money supply M1 and the two period time-lagged first difference in the oil price are used as explanatory variables in the index equation (1), and in model 5, where the two period time-lagged first difference in the money supply M1 (beside the time-lagged macroeconomic index itself) is employed as an explanatory variable, a stress event in these lagged variables in $T+1$ mainly has an effect on the stressed default probabilities in $T+2$ and $T+3$.[45] However, due to the correlation between the explanatory variables, those risk factors that are not explicitly stressed in $T+1$ are also influenced by the stress event, which can already have an influence on the stressed default probabilities in $T+1$ and $T+2$. Two observations are striking: First, in model 4, the mean values of the default probabilities in periods $T+2$ and $T+3$ are surprisingly low (compared with the minimum, maximum and mean default rate the data sample; see Table 1). Second, the effect of a shock in a lagged variable is rather low in both models. In case of model 5 and a stress in the two period time-lagged first difference in the money supply M1, the resulting 99.9% quantiles of the default probability are even smaller than in the non-stress case.

Table 8 quantifies how large the stress test results of the different model specifications can be dispersed. It shows the percentage differences between the largest (upper part of Table 8) and smallest (lower part of Table 8), respectively, forecasted stressed default probabilities in the base model 1 and in one of the other model specifications 2 to 9 (separated with respect to the expected forecasted stressed default probability and the 99.9% quantile and with respect to the

[44] However, using the shortened sample, the regression coefficients of the GDP and Moody's commodity index become insignificant in the index equation (1).

[45] For example in case of model 5 and the two period time-lagged first difference of M1, the last two historically observed values for this explanatory variables from $T-1$ and T are inserted into the index equation (1) to compute the index values for $T+1$ and $T+2$. To compute the index value for $T+3$, the shocked value of the explanatory variable is used.

time period).[46] For each model specification, the largest (smallest) forecasted stressed default probability corresponds to a specific stress scenario (GDP shock, oil price shock, etc.).

– insert Table 8 about here –

Table 9 shows the effect of changing the stress scenario definition on the forecasted default probabilities. In models 10 to 12, the base model 1 has been used, but instead of employing the historical worst case as the stress scenario, a stress scenario based on three standard deviations and on the Mahalanobis distance has been assumed (see Section 3.3.5). In the case of an assumed non-existence of serial (cross) autocorrelations, the Mahalanobis distances of a three standard deviation shock in the GDP and in the Moody's commodity index are identical with 3.37. When we employ the empirical (cross) autocorrelations, we get values for the Mahalanobis distance of 4.34 and 3.58. A comparison of the results for models 11 and 12 (see Table 7) shows that the inclusion of empirical (cross) autocorrelations in the variance-covariance matrix $\Sigma \in \mathbb{R}^{6\times 6}$ only has a minor effect on the forecasted stressed default probabilities. When comparing the results of model 10 with those of models 11 and 12, the stressed default probabilities of the latter two models are in the periods $T+2$ and $T+3$ larger than those ones of model 10. This is what one could expect, because, in contrast to the single three standard deviation stress scenario in $T+1$, the Mahalanobis-based stress scenarios distribute the stress over all three periods.

– insert Table 9 about here –

Based on the idea of vertical distances between the tails of the conditional (stress scenarios) and unconditional (non-stress scenarios) cumulative density functions for the default probabilities proposed by Schechtman and Gaglianone (2012), the tail pp-plots in Figure 4 give another possibility to compare the impact of stress scenarios relative to the non-stress scenario for different model specifications. It is assumed that a high quantile (x-axis) of the default probability distribution in the non-stress scenario is the maximum risk a bank is capable to bear. The y-axis visualizes for the non-stress as well as for the stress scenarios the probability of not exceeding this specified default probability quantile. Hence, the blue line is always the identity function which corresponds to the non-stress scenario of each model. The other lines indicate what percentage of the forecasted default probabilities in the stress scenarios is below the respective quantiles in the non-stress scenario. The larger the vertical distance is, i.e. the

[46] The results of models 10 to 12 are separately compared with those of the base model 1 because in these model modifications the definition of the stress scenarios is altered.

more the cumulative density functions of the simulated default probabilities in the non-stress scenario and in the various stress scenarios differ, the more severe is the stress scenario.[47] This corresponds to a low probability of not exceeding the specified default probability quantile in the non-stress scenario. Although the GDP shock proves to be the most severe one in many model specifications, Figure 4 shows that the extent of the vertical distances can vary considerably with the considered model specification.

– insert Figure 4 about here –

5 Discussion

In Section 3.1, 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 would have an influence on the simulated stressed 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 inconsistent. 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.[48]

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 (ΔGDP (t) and ΔΔMoody's commodity index (t)) in our base model 1. For this purpose, the Hausman test is employed. To perform this test, we need instrument variables that are strong and exogeneous. First, as in Schechtman and Gaglianone (2012), we try the one period time-lagged risk factors as instrument variables. However, the lagged variables ΔGDP ($t-1$) and ΔΔMoody's commodity index ($t-1$) prove to be weak instrument variables because their F-statistics are 4.24 and 7.09, respectively. To find strong instrument variables for the GDP and for the Moody's

[47] For Figure 4, only the first future period $T+1$ has been considered.
[48] See, for example, Boss (2002) or Virolainen (2004). An exception is Schechtman and Gaglianone (2012).

commodity index, data from the World Bank's World Development Indicators as of 2012 is employed. Out of the more than 300 variables, 9 one period time-lagged variables are strong (F -statistics larger than 10) instrument variables for ΔGDP (t).[49] For ΔΔMoody's commodity index (t), 46 one period time-lagged variables could be identified as strong instrument variables. All these strong instrument variables have been used for performing the Hausman test for endogeneity of the explanatory variables of the base model 1.[50] For ΔGDP (t),the null hypothesis of exogeneity could be rejected with none of the 9 instrument variables. For ΔΔMoody's commodity index (t), the null hypothesis of exogeneity could not be rejected with 40 out of 46 instrument variables. Thus, endogeneity and biased parameter estimates seem to be no problem in the base model. However, it has to be considered that the Hausman test is only asymptotically valid and that our sample with only 28 data points is not very large.

In CPV-style stress test models, it is assumed that there is a linear relationship between the macroeconomic index (corresponding to the transformed default rates) and the explanatory risk factors (see (1) and (2)). As the scatter plots in Figure 5 show, for our sample, the relationship between the logit-transformed default rates and the explanatory risk factors is at best rudimentarily linear. There are some severe outliers that are not captured by a linear relationship. These could possibly be explained by the well-known criticism with respect to stress tests that statistical relationships can change in an unpredictable manner in a crisis (see, for example, Alfaro and Drehmann (2009)). This deficiency reduces the suitability of the (linear) approach as a base for a credit risk stress test.[51]

- insert Figure 5 about here -

A further criticism of credit risk stress tests based on CPV-style models concerns the specification of the error terms $u_{s,t}$ for the macroeconomic index in (1) (see Schechtman and Gaglianone (2012, p. 176)). These are assumed to be multivariately normally distributed (together with the error terms $v_{i,t}$ for the risk factors in (3)), serially (cross) uncorrelated and homoscedastic. Unfortunately, the employed data does not always fit with these assumptions. However, a violation of these assumptions would not only have to be considered within a rigorous statistical estimation of the parameters $\beta_{s,0},...,\beta_{s,I}$ in (1), but also within the simulation of fu-

[49] To reach stationarity, first differences have been computed in most cases.

[50] The instrument variable parameter estimates needed for the Hausman test statistic are computed using two stage least squares (2SLS).

[51] Removing these outliers is not appropriate because doing this would destroy stress information.

ture (stressed) default probabilities (see Section 3.1). For example, assuming that the error terms $u_{s,t}$ are fat-tailed t-distributed (as in Simons and Rolwes (2009), for example), the density functions of the forecasted stressed default probabilities in Figure 3 (and, hence, the quantiles) could change considerably.

Our framework is based on yearly US-default rates from Standard & Poor's spanning the period from 1983 to 2010. One may argue that this relatively short time series upon which all models are estimated is the trigger for the high variance in forecasted default probabilities across different model specifications, i.e. for model and estimation risk. Of course, this is a fair point, but, however, banks in practice have to tackle the same issue. Moreover, if a default rate shall be explained by macroeconomic variables, the frequency of the macroeconomic variable might be binding (e.g. for the quarterly published GDP). There exist several alternatives, for example, internal default rates, internal loan loss provisions, external data from database providers, insolvency statistics, etc. Some data sources might not be representative for the current bank portfolio; others might suffer from a poor data quality. In order to consider the potential impact of the number of observations, we have in parallel calibrated a model based on data which is available in a higher frequency. However, the results give an indication that the number of observations is not the main driver for the high variance in forecasted default probabilities across different model specifications.[52]

6 Conclusions

We analyzed to which extent multi-period stressed values of default probabilities within a given framework are affected by modelling assumptions and estimation techniques. To achieve this, starting from a base model specification, we employed several variations of a CreditPortfolioView (CPV)-style model. All variations were statistically sound and it was ex-ante not obvious why one specification or estimation technique should be more adequate than another. We showed that the chosen model specifications and the employed estimation tech-

[52] We used the German insolvency statistic which is available on a monthly basis since 2003 to present. However, we limited the time series to end-2010 to be in line with the US data model. Overall, we had 96 observations instead of 28 as in case of US data. Moreover, the data covered only 8 years and, thus, a structural break was less likely than in longer time series. We compared the prognosis for default probabilities between the whole insolvency data set and three equally divided sub-samples. The results suggest that the predictions did barely differ in the sub-samples compared with the whole data set for various model specifications and that the variation of the default probability predictions across the different model specifications was comparable (low) for short and long data samples. However, we should consider that we forecasted only a period of 3 months instead of 3 years what might lessen model and estimation risk. Thus, conclusions should be drawn carefully. Furthermore, the standard deviation of the German insolvency rates is much smaller than the standard deviation of the realized US-default rates and it can not be excluded that this has an effect on the sensitivity of model and estimation risk in the default probability predictions with respect to the length of the data sample.

niques can hugely influence the results for the stressed default probabilities. These results show the importance of extensive robustness checks for the underlying model when interpreting the results of model-based credit risk stress tests. Considering the close relationship between stress test models and regular risk models, an extensive evaluation of the underlying assumptions is deemed to be necessary in regular risk models, too. Even non-stressed PDs, LGDs or interest rate parameters are likely to bear considerable model and estimation risk.

Furthermore, it should be noted that the transformation of macroeconomic variables into risk parameter realizations is required in many situations. Whereas 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. 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.

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.

Alfaro, R., M. Drehmann (2009): Macro stress tests and crisis: what can we learn? BIS: Quarterly Review, December, pp. 29-41.

Bank for International Settlements (BIS) (2011): Basel III: A global regulatory framework for more resilient banks and banking systems.

Bank for International Settlements (BIS) (2014): Reducing excessive variability in banks' regulatory capital ratios.

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

Behn, M., R. Haselmann, V. Vig (2014): The Limits of Model-Based Regulation, SAFE Working Paper No. 75.

Bonti, G., M. Kalkbrener, C. Lotz, G. Stahl (2006): Credit risk concentrations under stress, Journal of Credit Risk, Vol. 2, No. 3, pp. 115-136.

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

Breuer, T., M. Jandacka, K. Rheinberger, M. Summer (2008): Hedge the Stress: Using Stress Tests to Design Hedges for Foreign Currency Loans, Stress Testing for Financial Institutions, D. Rösch and H. Scheule (eds.), Riskbooks, London, pp. 111-126.

Breuer, T., M. Jandacka, K. Rheinberger, M. Summer (2010): Does Adding Up of Economic Capital for Market- and Credit Risk amount to Conservative Risk Assessment?, Journal of Banking & Finance, Vol. 34, No. 4, pp. 703-712.

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.

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

Drehmann, M., S. Sorensen, M. Stringa (2010): The integrated impact of credit and interest rate risk on banks: A dynamic framework and stress testing application, Journal of Banking & Finance, Vol. 34, No. 4, pp. 713-729.

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

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

Fernández, C., E. Ley, M.F. Steel (2001): Benchmark priors for Bayesian model averaging, Journal of Econometrics, Vol. 100, No. 2, pp. 381-427.

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

Financial Services Authority (FSA) (2008): Stress and Scenario Testing, Consultation Paper 08/24.

Financial Services Authority (FSA) (2009): Stress and Scenario Testing, Feedback on CP08/24 and Final Rules, Policy Statement 09/20.

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.

Grundke, P. (2011): Reverse stress tests with bottom-up approaches, Journal of Risk Model Validation, Vol. 5, No. 1, pp. 71-90.

Grundke, P. (2012): Further recipes for quantitative reverse stress testing, Journal of Risk Model Validation, Vol. 6, No. 2, pp. 81-102.

Grundke, P., K. Pliszka (2015): A macroeconomic reverse stress test, Working Paper, Osnabrück University, Chair of Banking and Finance.

Hamerle, A., R. Jobst, M. Lerner (2008): Mehrjährige makroökonomische Stresstests: Ein ökonometrischer Ansatz, RisikoManager, No. 8, pp. 1, 8-15.

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

Han, C. (2014): Comparative analysis of credit risk models for loan portfolios, Journal of Risk Model Validation, Vol. 8, No. 2, pp. 3-22.

Höse, S., S. Huschens (2008): Worst-case asset, default and survival time correlations, Journal of Risk Model Validation, Vol. 2, No. 4, pp. 27-50.

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.

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.

Kalirai, H., M. Scheicher (2002): Makroökonomische Krisentests - Erste Ergebnisse für Österreich, OeNB, Finanzmarktstabilitätsbericht Nr. 3, pp. 64-82.

Liang, F., R. Paulo, G. Molina, M.A. Clyde, J.O. Berger (2008): Mixtures of g priors for Bayesian variable selection, Journal of American Statistical Association, Vol. 103, No. 482, pp. 410-423.

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.

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.

Rösch, D., H. Scheule (2007): Stress-testing credit risk parameters: an application to retail loan portfolios, Journal of Risk Model Validation, Vol. 1, No. 1, pp. 55-75.

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.

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.

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.

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

Tables

Table 1: Descriptive statistics of the endogenous and exogenous variables

	Mean	Std	Max	Min	Data source	Unit
Endogenous variable						
Default rate	0.0464	0.0288	0.1156	0.0101	Standard & Poor's (2011)	%
Index (logit)	3.20	0.65	4.59	2.03	-	-
Index (probit)	1.75	0.29	2.32	1.20	-	-
Exogenous variables						
Economic activity indicators						
Gross domestic product (GDP)	10620.4	2869.7	14876.8	6443.4	Datastream: USGDP...D	billion USD
Industrial production	881.1	205.2	1200.0	579.4	Datastream: USIPTOT.G	index
Price stability indicators						
Inflation	151.6	40.4	218.1	82.4	Datastream: USCONPRCE	index
Money supply M1	1007.2	358.8	1741.7	395.7	Datastream: USM1....B	billion USD
Money supply M3	6.0	2.6	12.2	1.0	Datastream: USYMA013Q	annual growth rate in %
Moody's commodity index	1662.4	1035.7	5205.4	926.9	Datastream: MOCMDTY	USD
Reuter's commodity index	1798.3	329.3	2555.6	1246.5	Datastream: RECMDTY	GBP
Household indicators						
Personal consumption expenditure	5518.0	2722.2	10215.7	1754.6	Datastream: USCNPER.B	billion USD
Disposable personal income	6082.1	2907.7	11243.7	2017.9	Datastream: USPERDISB	billion USD
New home sales	723.3	234.0	1279.0	321.0	Datastream: USHOUSESE	thousend
Unemployment rate	6.30	1.62	9.70	3.97	Datastream: USUN%TOTQ	%
Firm indicators						
Consumer confidence	93.1	24.3	139.0	45.2	Datastream: USCNFCONQ	index
Consumer sentiment	87.0	12.1	107.5	63.8	Datastream: USUMCONSH	index
Financial market indicators						
3-months Treasury bill rate	5.35	3.30	13.82	0.14	Datastream: USGBILL3	%
S&P 500	694.7	464.5	1476.5	118.8	Datastream: S&PCOMP	USD
External indicators						
Exports	594351.3	329416.6	1287441.0	200662.1	Datastream: USEXPGDSB	million USD
Imports	896072.6	581164.7	2103641.0	243952.4	Datastream: USIMPGDSB	million USD
JPY/USD exchange rate	141.40	49.47	248.92	87.76	Datastream: JPXRUSD	JPY
USD/GBP exchange rate	1.68	0.23	2.32	1.30	Datastream: STUSBOE	USD
Oil price Brent (FOB) per Barrel	33.0	21.5	97.3	13.1	Datastream: OILBREN	USD
Oil price Brent per Barrel	32.9	21.5	97.1	12.8	Datastream: OILBRDT	USD
Oil price WTI (FOB) per Barrel	33.7	22.1	99.6	14.4	Datastream: OILWTXI	USD

Table 2:

Overview of the specification of the base model and the considered modifications

	Base model (no. 1)	**Modifications**	**Model no.**
Stationary method	First differences and, if necessary, second differences	Returns (if necessary second returns), log-returns	2, 3
Time-lagged risk factors	None	Additional time-lagged macroeconomic variables and time-lagged macroeconomic index as explanatory variables for the macroeconomic index	4, 5
Estimator for the macroeconomic index equation	OLS	GLS	6
Transformation between default rate and macroeconomic index	Logit	Probit	7
Time series processes for macroeconomic variables	AR(1)/AR(2) (based on AIC/BIC with only significant parameters)	Fixed AR(2)	8
Method for choosing relevant risk factors	Mixture of expert judgement and stepwise regression	Bayesian model averaging	9
Stress test scenario	Historical worst case scenario	Hypothetical scenarios based on three standard deviations of the error terms and based on the Mahalanobis distance	10, 11, 12

Table 3: Estimates of the risk factor processes

		Parameters	R^2	Adjusted R^2	Applied specification
Model 1: Base model					
Δ GDP (t)	Intercept	176.7004**	0.150	0.115	AR(1)
	t-1	0.381*			
ΔΔ Moody's commodity index (t)	Intercept	74.1048	0.412	0.360	AR(2)#
	t-1	-0.9244***			
	t-2	-0.9453***			
Model 2: Stationary methods (returns)					
Return Consumer confidence (t)	Intercept	-0.000231	-	-	Intercept
Return Money supply M3 (t)	Intercept	0.01649	-	-	Intercept
Model 3: Stationary methods (log-returns)					
Log-return Consumer confidence (t)	Intercept	-0.01676	-	-	Intercept
Log-return Money supply M3 (t)	Intercept	-0.05901	-	-	Intercept
Model 4: Time-lagged risk factors					
Δ GDP (t)	Intercept	176.7004**	0.150	0.115	AR(1)
	t-1	0.381*			
Δ Money supply M1 (t-1)	Intercept	30.4192***	0.666	0.636	AR(2)
	t-1	1.3518***			
	t-2	-1.001***			
Δ Oil price Brent (FOB) (t-2)	Intercept	1.82	0.153	0.118	AR(1)
	t-1	0.4767**			
Model 5: Time-lagged risk factors and time-lagged macroeconomic index					
Δ GDP (t)	Intercept	176.8452**	0.153	0.115	AR(1)
	t-1	0.3927*			
Δ Money supply M1 (t-2)	Intercept	24.5236***	0.690	0.660	AR(2)
	t-1	1.1307***			
	t-2	-0.7568***			
Model 6: GLS estimator					
Δ GDP (t)	Intercept	176.7004**	0.150	0.115	AR(1)
	t-1	0.381*			
Δ Money supply M3 (t)	Intercept	-0.36	-	-	Intercept
Model 7: Probit transformation					
Δ GDP (t)	Intercept	176.7004**	0.150	0.115	AR(1)
	t-1	0.381*			
ΔΔ Moody's commodity index (t)	Intercept	74.1048	0.412	0.360	AR(2)#
	t-1	-0.9244***			
	t-2	-0.9453***			
Model 8: Fixed AR(2) process for risk factors					
Δ GDP (t)	Intercept	267.9039***	0.229	0.159	AR(2) (fixed)
	t-1	0.6043**			
	t-2	-0.4745			
ΔΔ Moody's commodity index (t)	Intercept	74.1048	0.412	0.360	AR(2) (fixed)#
	t-1	-0.9244***			
	t-2	-0.9453***			
Model 9: Bayesian model averaging					
ΔΔ Industrial production (t)	Intercept	-1.857	-	-	Intercept

This table summarizes the OLS parameter estimates of the risk factor processes and their significance. The symbols *, ** and *** denote significance at the 10%, 5% and 1% level. The intercept is included in all risk factor process specifications, even if it is not significant. When the minimal p-value of the Godfrey-Breusch test (up to a lag of 4) is below 5%, the Newey-West estimator is used (denoted by #). The symbols Δ and $\Delta\Delta$ denote first and second differences. When the specification only consists of an intercept plus error term, calculation of R^2 or the adjusted R^2 is not possible. Due to differing transformation methods for the time series or differing lengths of the time series, the specification of the AR processes can be different for the same risk factor across the various models.

Table 4: Estimation results for the macroeconomic index equation

	Parameters	p-value Godfrey-Breusch test	R^2	Adjusted R^2
Model 1: Base model		≥0.709	0.418	0.369
Intercept	-0.3131*			
Δ GDP (t)	0.0008916*			
ΔΔ Moody's commodity index (t)	0.0008181***			
Model 2: Stationary methods (returns)		≥0.515	0.334	0.278
Intercept	0.02407			
Return Consumer confidence (t)	0.5642**			
Return Money supply M3 (t)	-0.1808*			
Model 3: Stationary methods (log-returns)		≥0.762	0.480	0.436
Intercept	-0.006225			
Log-return Consumer confidence (t)	0.5208***			
Log-return Money supply M3 (t)	-0.2308***			
Model 4: Time-lagged risk factors		≥0.410	0.551	0.493
Intercept	-0.8087***			
Δ GDP (t)	0.001614***			
Δ Money supply M1 (t-1)	0.006277***			
Δ Oil price Brent (FOB) (t-2)	0.0228*			
Model 5: Time-lagged risk factors and time-lagged macroeconomic index		≥0.294	0.505	0.416
Intercept	-0.6051***			
Δ Macroeconomic index (t-2)	-0.5170**			
Δ GDP (t)	0.001193**			
Δ Money supply M1 (t-2)	0.007133**			
Model 6: GLS estimator		-	-	-
Intercept	-0.3959*			
Δ GDP (t)	0.001179**			
Δ Money supply M3 (t)	-0.1153*			
Model 7: Probit transformation		≥0.561	0.445	0.399
Intercept	-0.1434*			
Δ GDP (t)	0.0004072*			
ΔΔ Moody's commodity index (t)	0.0003834***			
Model 8: Fixed AR(2) process for risk factors				
as model 1				
Model 9: Bayesian model averaging		-	-	-
ΔΔ Industrial production (t)	0.007641			

This table summarizes the OLS parameter estimates of the macroeconomic index equation (1) and their significance for various specifications. The symbols *, ** and *** denote significance at the 10%, 5% and 1% level. The intercept is included in all models, even if it is not significant. For all specifications, the variance inflation factor has been calculated (not shown in the table). As it is always only slightly above 1, multicollinearity between the explanatory variables can be ruled out. For model 6 and 9 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. In model 9 we neglected to use the intercept from the BMA calculation as it captures the mean differences between dependent and independent variables for all candidate risk factors and not only for the industrial production. Moreover, an indication about the significance of the parameter in model 9 cannot be given as the value is calculated as a weighted average of various models (see (12)).

Table 5: Results of the Bayesian model averaging

Variable	PIP	Mean	Std	Cond. pos. sign
ΔΔ Industrial production	80.00%	0.007691	0.0051088	1.0000
Δ New home sales	36.50%	0.000576	0.0010332	0.9998
Δ Personal consumption expenditure	33.45%	0.000589	0.0012465	0.9705
Δ ΔMoney supply M1	32.05%	0.001132	0.0023887	0.9912
ΔΔ Moody’s commodity index	31.12%	0.000159	0.0003407	0.9943
3-months Treasury bill rate	31.07%	-0.016755	0.0375965	0.0263
Δ Consumer confidence	27.55%	0.002757	0.0069672	0.9938
Unemployment rate	26.39%	0.0301	0.0886617	0.8736
Δ Oil price Brent per Barrel	24.87%	0.034776	0.1219770	0.7854
Δ Oil price WTI (FOB) per Barrel	23.30%	-0.018497	0.0706217	0.2175
Δ JPY/USD exchange rate	22.97%	0.001194	0.0037497	0.9790
Δ Reuter’s commodity index	21.69%	0.000098	0.0003632	0.9359
ΔΔ Disposable personal income	21.58%	-0.000128	0.0005329	0.0994
Δ Oil price Brent (FOB) per Barrel	21.39%	-0.014563	0.1040330	0.4002
Δ Gross domestic product (GDP)	21.34%	0.000019	0.0006237	0.5954
Δ S&P500	21.09%	0.00015	0.0005904	0.9188
Δ Consumer sentiment	20.91%	0.001654	0.0123400	0.7724
Δ Imports	20.46%	0.00000004	0.0000013	0.5546
Δ Inflation	20.20%	0.011201	0.0648337	0.7269
Δ Money supply M3	19.98%	0.006329	0.0399630	0.7816
ΔΔExports	19.34%	-0.00000022	0.0000017	0.2769
USD/GBP exchange rate	18.51%	0.018104	0.3534386	0.5431

The table summarizes the results of the Bayesian model averaging models in which the variable was included. The mean and the standard deviation refer to the average coefficient values across all models. The conditional positive sign is the fraction of models in which the variable had a positive sign given that the variable was included in the model. PIP: Posterior inclusion probability.

Table 6: Out-of-sample performance

	T+1 (2011)	T+2 (2012)	T+3 (2013)	
Realized default rates	2.10%	2.58%	2.12%	
Mean deviation MD_n				
Model 1: Base model	6.15%	12.08%	3.65%	
Model 2: Stationary methods (returns)	1.92%	2.11%	3.21%	
Model 3: Stationary methods (log-returns)	1.78%	1.76%	2.64%	
Model 4: Time-lagged risk factors	-0.83%	-2.31%	-1.97%	
Model 5: Time-lagged risk factors and time-lagged macroeconomic index	3.45%	0.02%	-0.61%	
Model 6: GLS estimator	1.85%	2.19%	3.59%	
Model 7: Probit transformation	5.92%	10.72%	3.27%	
Model 8: Fixed AR(2) process for risk factors	4.02%	7.03%	1.60%	
Model 9: Bayesian model averaging	1.81%	1.93%	3.02%	
Mean squared error MSE_n				***CMSE***
Model 1: Base model	0.00600	0.02450	0.00468	0.03518
Model 2: Stationary methods (returns)	0.00136	0.00327	0.00640	0.01103
Model 3: Stationary methods (log-returns)	0.00092	0.00159	0.00269	0.00520
Model 4: Time-lagged risk factors	0.00012	0.00054	0.00039	0.00105
Model 5: Time-lagged risk factors and time-lagged macroeconomic index	0.00202	0.00049	0.00028	0.00278
Model 6: GLS estimator	0.00099	0.00262	0.00597	0.00958
Model 7: Probit transformation	0.00513	0.01728	0.00343	0.02584
Model 8: Fixed AR(2) process for risk factors	0.00291	0.01023	0.00186	0.01499
Model 9: Bayesian model averaging	0.00082	0.00173	0.00358	0.00612

This table shows the mean deviation (in percentage points) between the forecasted default probabilities and the realized default rates for each year, the mean squared error for each year and the cumulative mean squared error over all three years. Expectations are based on 1 million simulated forecasts of the default probabilities for 2011 to 2013.

Table 7: Forecasted default probabilities

		T+1	T+2	T+3
Model 1: Base model				
Mean				
	Non-stress	8.25%	14.66%	5.77%
	GDP	16.61%	24.86%	10.96%
	Moody's commodity index	16.93%	18.96%	7.53%
99.9% quantile				
	Non-stress	33.51%	63.27%	46.04%
	GDP	48.79%	75.96%	62.61%
	Moody's commodity index	47.55%	70.16%	53.02%

		T+1	T+2	T+3
Model 2: Stationary methods (returns)				
Mean				
	Non-stress	4.02%	4.69%	5.33%
	Consumer confidence	8.53%	9.59%	10.49%
	Money supply M3	8.87%	9.90%	10.80%
99.9% quantile				
	Non-stress	25.12%	44.69%	61.35%
	Consumer confidence	37.69%	62.07%	76.58%
	Money supply M3	41.09%	63.99%	78.01%

Model 3: Stationary methods (log-returns)				
Mean				
	Non-stress	3.88%	4.34%	4.76%
	Consumer confidence	8.57%	8.97%	9.32%
	Money supply M3	7.08%	7.51%	7.91%
99.9% quantile				
	Non-stress	14.99%	21.04%	25.66%
	Consumer confidence	20.07%	26.65%	30.84%
	Money supply M3	18.34%	24.90%	29.29%

Model 4: Time-lagged risk factors				
Mean				
	Non-stress	1.27%	0.27%	0.15%
	GDP (t)	3.01%	0.87%	0.54%
	Money supply M1 (t-1)	1.25%	0.38%	0.35%
	Oil price Brent (FOB) (t-2)	1.31%	0.32%	0.27%
99.9% quantile				
	Non-stress	5.41%	2.39%	2.63%
	GDP (t)	9.62%	5.80%	7.71%
	Money supply M1 (t-1)	5.36%	3.16%	4.92%
	Oil price Brent (FOB) (t-2)	5.59%	2.79%	4.43%

Table 7 [continued]

		T+1	T+2	T+3
Model 5: Time-lagged risk factors and time-lagged macroeconomic index				
Mean				
	Non-stress	5.55%	2.60%	1.51%
	GDP (t)	10.13%	6.07%	3.37%
	Money supply M1 (t-2)	5.69%	2.70%	2.01%
99.9% quantile				
	Non-stress	21.02%	18.48%	13.95%
	GDP (t)	30.81%	31.69%	24.81%
	Money supply M1 (t-2)	19.01%	15.16%	12.98%

		T+1	T+2	T+3
Model 6: GLS estimator				
Mean				
	Non-stress	3.95%	4.77%	5.71%
	GDP	8.87%	12.51%	15.14%
	Money supply M3	6.20%	7.71%	9.17%
99.9% quantile				
	Non-stress	19.47%	37.19%	55.49%
	GDP	32.16%	59.93%	77.88%
	Money supply M3	24.50%	47.41%	66.54%

Model 7: Probit transformation				
Mean				
	Non-stress	8.02%	13.30%	5.39%
	GDP	15.09%	21.16%	9.87%
	Moody’s commodity index	15.45%	16.69%	6.95%
99.9% quantile				
	Non-stress	26.94%	47.71%	34.39%
	GDP	37.32%	58.70%	46.20%
	Moody’s commodity index	36.41%	53.41%	39.22%

Model 8: Fixed AR(2) processes for risk factors				
Mean				
	Non-stress	6.12%	9.61%	3.72%
	GDP	12.19%	17.78%	6.87%
	Moody’s commodity index	12.82%	13.05%	4.84%
99.9% quantile				
	Non-stress	26.39%	51.03%	34.45%
	GDP	39.58%	66.26%	49.23%
	Moody’s commodity index	39.17%	59.43%	40.60%

Table 7 [continued]

		T+1	T+2	T+3
Model 9: Bayesian model averaging				
Mean				
	Non-stress	3.91%	4.51%	5.14%
	Industrial production	7.12%	8.07%	9.03%
99.9% quantile				
	Non-stress	16.62%	29.32%	42.06%
	Industrial production	23.16%	40.32%	55.35%

		T+1	T+2	T+3
Model 10: Three standard deviations stress scenario				
Mean				
	Non-stress	8.25%	14.66%	5.77%
	GDP	16.65%	24.90%	10.99%
	Moody's commodity index	18.28%	19.48%	7.75%
99.9% quantile				
	Non-stress	33.51%	63.27%	46.04%
	GDP	48.86%	76.00%	62.67%
	Moody's commodity index	49.95%	70.93%	53.86%

		T+1	T+2	T+3
Model 11: Mahalanobis-based stress scenario (no (cross) autocorrelation)				
Mean				
	Non-stress	8.25%	14.66%	5.77%
	Stress	12.12%	27.55%	21.37%
99.9% quantile				
	Non-stress	33.41%	63.29%	46.50%
	Stress	36.17%	74.42%	75.15%

		T+1	T+2	T+3
Model 12: Mahalanobis-based stress scenario (empirical (cross) autocorrelation)				
Mean				
	Non-stress	8.25%	14.66%	5.77%
	GDP (equiv.)	11.19%	37.04%	33.17%
	Moody's commodity index (equiv.)	10.56%	32.05%	25.53%
99.9% quantile				
	Non-stress	33.41%	63.29%	46.50%
	GDP (equiv.)	34.09%	82.43%	85.66%
	Moody's commodity index (equiv.)	32.82%	78.47%	79.84%

This table shows the mean and the 99.9% quantile of the probability distribution for the forecasted stressed default probabilities in various model specifications. In the case of models 11 and 12, the stress test scenarios are characterized by the most harmful (in the sense of (9)) scenarios out of those trust regions Ell_{τ} that correspond to the respective three standard deviations stress of the macroeconomic variables in the base model (see Section 3.3.5.2). In case of model 11, the parameter τ is identical for both macroeconomic variables (GDP, Moody's commodity index) so that we have only one stress scenario (titled 'Stress').

Table 8:
Percentage differences between forecasted stressed default probabilities in the base model 1 and in model modifications 2 to 9

		T+1	T+2	T+3
Stressed default probabilities (highest)				
Mean				
	Min	-82.3%	-96.5%	-95.1%
	Max	-8.8%	-14.9%	38.1%
	Mean	-44.8%	-57.1%	-26.0%
	Standard deviation	21.9%	25.9%	41.3%
99.9% quantile				
	Min	-80.3%	-92.4%	-87.7%
	Max	-15.8%	-12.8%	24.6%
	Mean	-40.1%	-41.9%	-26.1%
	Standard deviation	22.3%	28.5%	39.5%
Stressed default probabilities (lowest)				
Mean				
	Min	-92.5%	-98.3%	-96.4%
	Max	-9.2%	-12.0%	39.4%
	Mean	-52.5%	-56.7%	-15.8%
	Standard deviation	25.3%	27.5%	48.5%
99.9% quantile				
	Min	-88.7%	-96.0%	-91.6%
	Max	-17.6%	-11.5%	44.4%
	Mean	-46.5%	-45.6%	-23.4%
	Standard deviation	24.6%	31.0%	47.2%

This table quantifies the percentage differences between the highest (upper part of Table 8) and lowest (lower part of Table 8) forecasted stressed default probabilities in the base model 1 and in one of the other model specifications 2 to 9 (stationary methods, time-lagged variables, GLS estimator, probit transformation, Bayesian model averaging). The values are separated with respect to the forecasted expected stressed default probability (mean) and the 99.9% quantile and with respect to the time period. For each model specification, the largest (smallest) forecasted stressed default probability corresponds to a specific stress scenario (GDP shock, oil price shock, etc.).

Table 9:
Percentage differences between forecasted stressed default probabilities in the base model 1 and in model modifications 10 to 12

		T+1	T+2	T+3
Stressed default probabilities (highest)				
Mean				
	Min	-33.9%	0.2%	0.2%
	Max	7.9%	49.0%	202.6%
	Mean	-18.1%	20.0%	99.3%
	Standard deviation	22.7%	25.6%	101.2%
99.9% quantile				
	Min	-30.1%	-2.0%	0.1%
	Max	2.4%	8.5%	36.8%
	Mean	-17.7%	2.2%	19.1%
	Standard deviation	17.6%	5.6%	18.4%
Stressed default probabilities (lowest)				
Mean				
	Min	-36.4%	2.8%	2.9%
	Max	0.2%	69.1%	239.2%
	Mean	-21.1%	39.1%	142.0%
	Standard deviation	19.0%	33.6%	123.6%
99.9% quantile				
	Min	-31.0%	1.1%	1.6%
	Max	2.7%	11.8%	50.6%
	Mean	-17.4%	6.3%	31.3%
	Standard deviation	17.8%	5.4%	26.1%

This table quantifies the percentage differences between the highest (upper part of Table 9) and lowest (lower part of Table 9) forecasted stressed default probabilities in the base model 1 and in one of the other model specifications 10 to 12 (various definitions of stress scenarios). The values are separated with respect to the forecasted expected stressed default probability (mean) and the 99.9%-quantile and with respect to the time period. For each model specification, the highest (lowest) forecasted stressed default probability corresponds to a specific stress scenario (GDP shock, oil price shock, etc.).

Figures

Figure 1: Overview of the stepwise regression approach

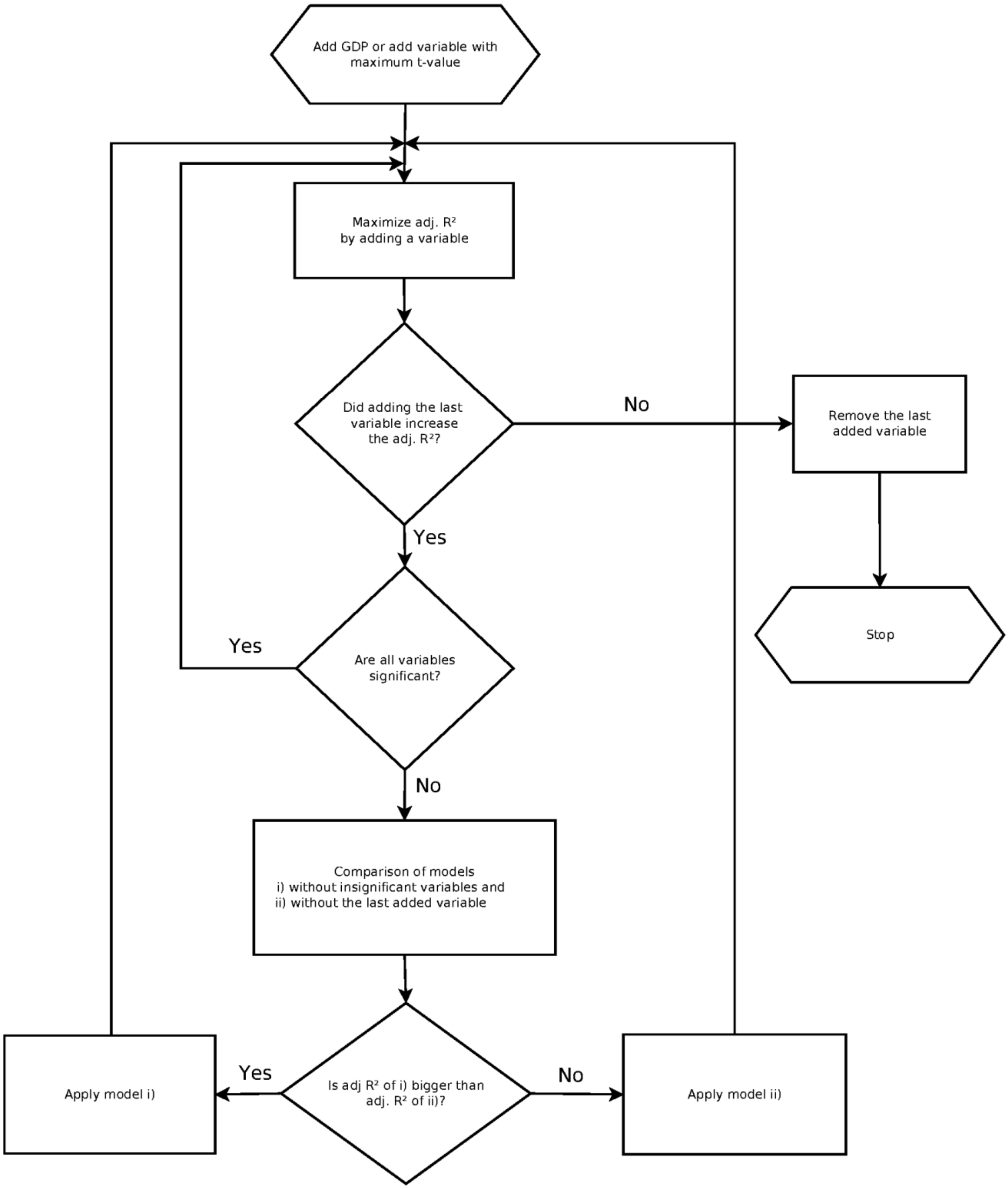


Figure 1 shows the stepwise regression approach for variable selection that is repeated for model specifications 1 to 7. A prerequisite for adding a variable (to avoid (imperfect) multi-collinearity) is that the absolute value of their correlation with any of the other variables that have already been included in the model is below 0.8.

Figure 2: Realized default rates versus in-sample and out-of-sample forecasted default probabilities

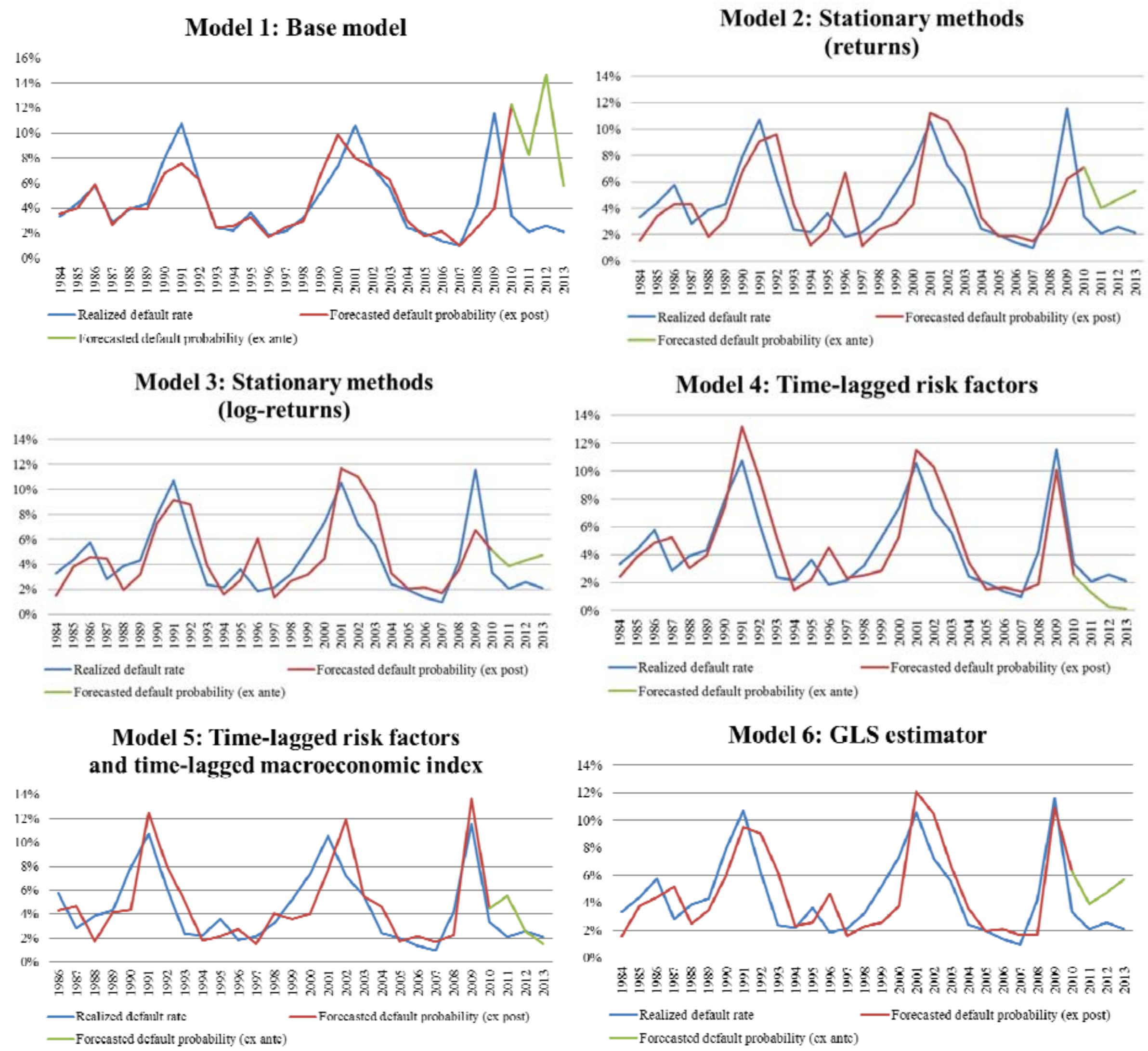


Figure 2 [continued]

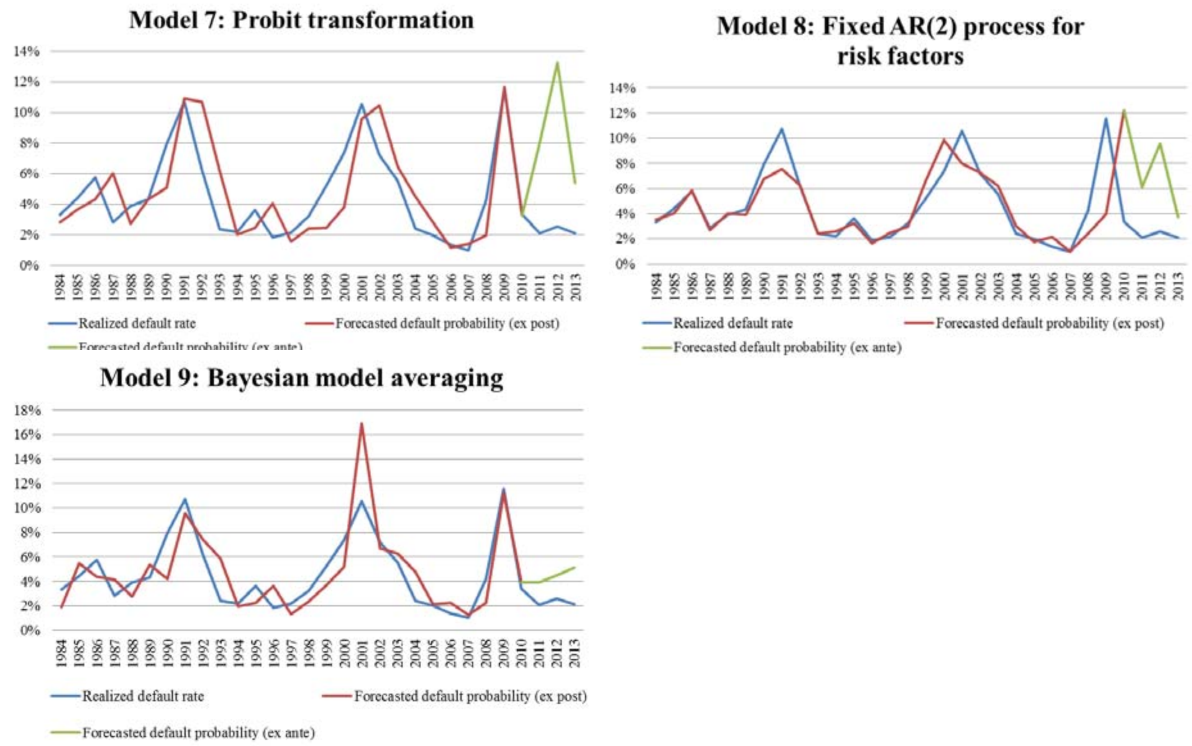


Figure 2 shows the realized default rates compared with the in-sample and out-of-sample predictions of the default probabilities (based on (1) and (2)). For the in-sample prediction, the observed risk factor realizations of each model are inserted into (1), the error term is set equal to its mean zero and the calculated realizations of the macroeconomic index are inserted into (2), which yields the predicted default probabilities. For the out-of-sample prediction, the mean forecasted default probabilities in the non-stress case are employed (see Table 7 in the following).

Figure 3: Density functions of forecasted default probabilities

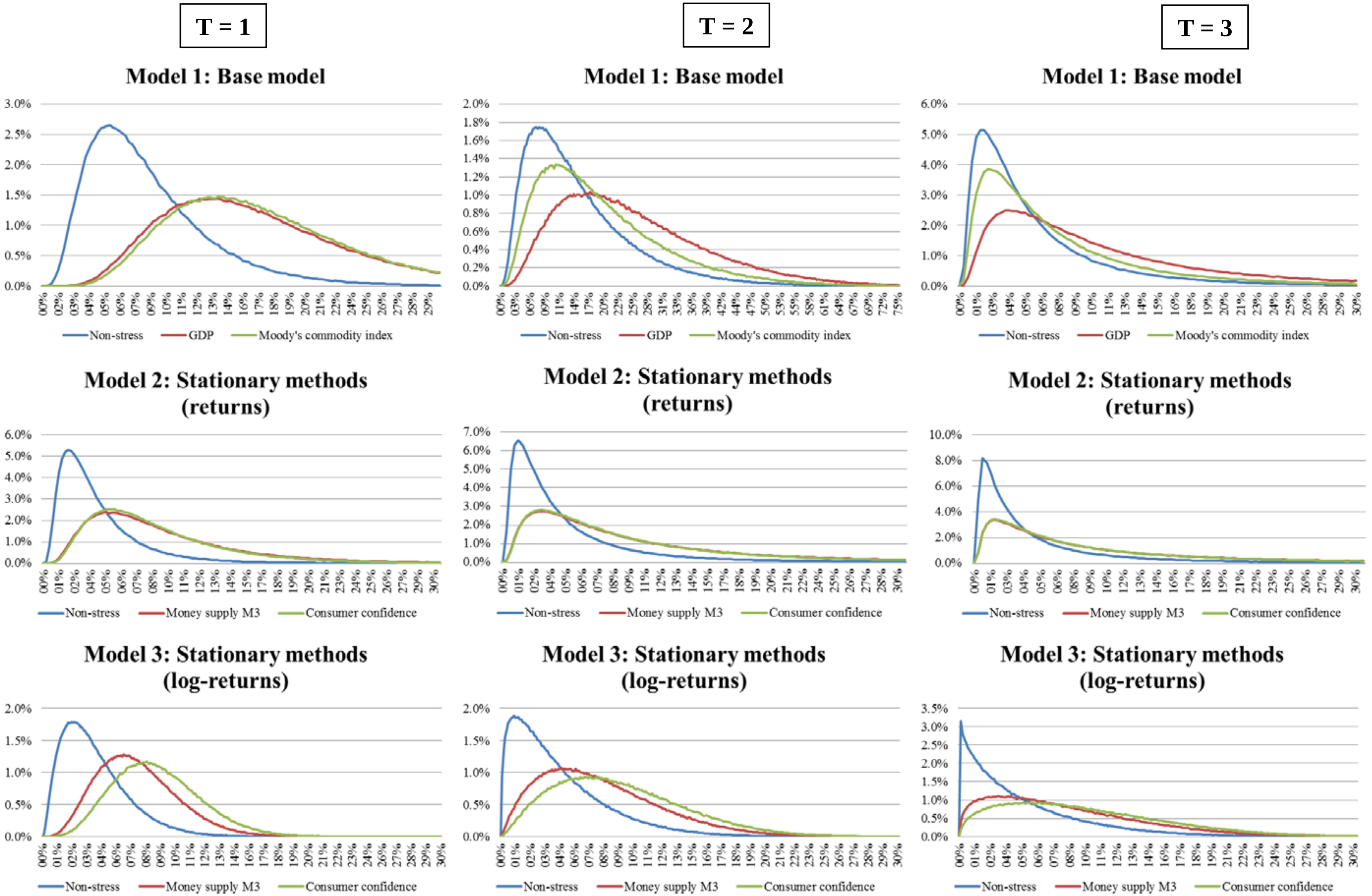


Figure 3 [continued]

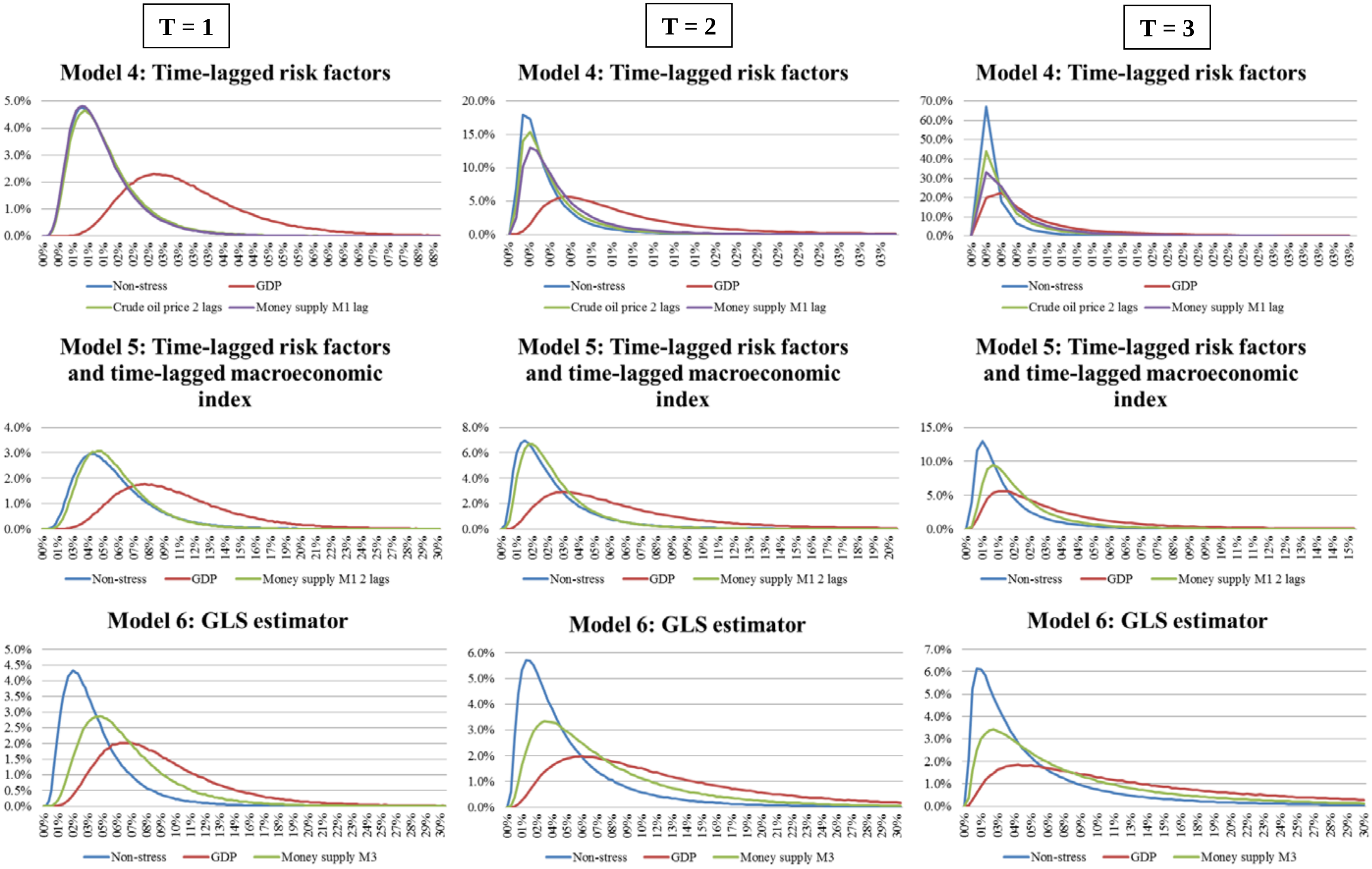


Figure 3 [continued]

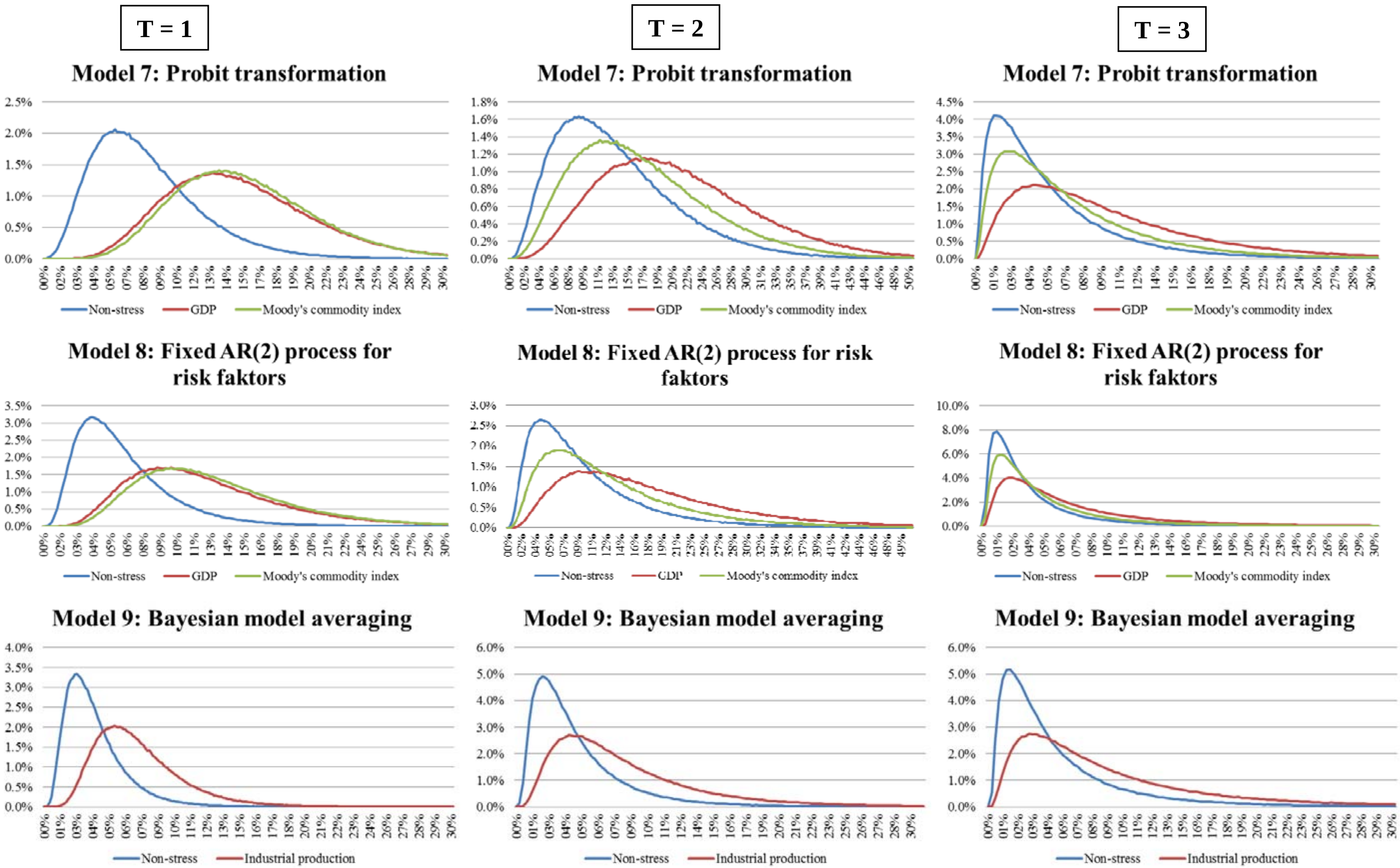


Figure 3 [continued]

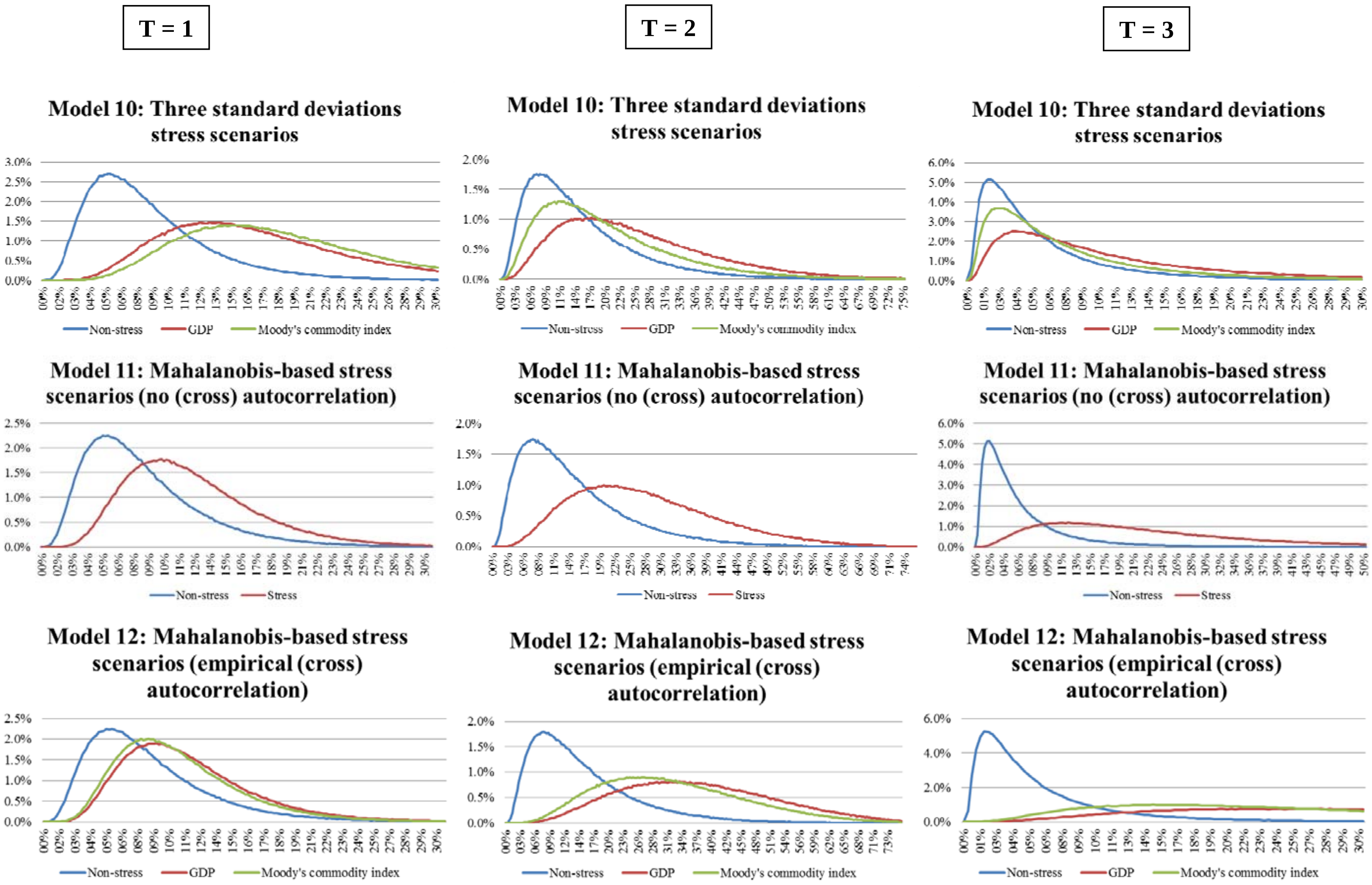


Figure 4: Tail pp-plots

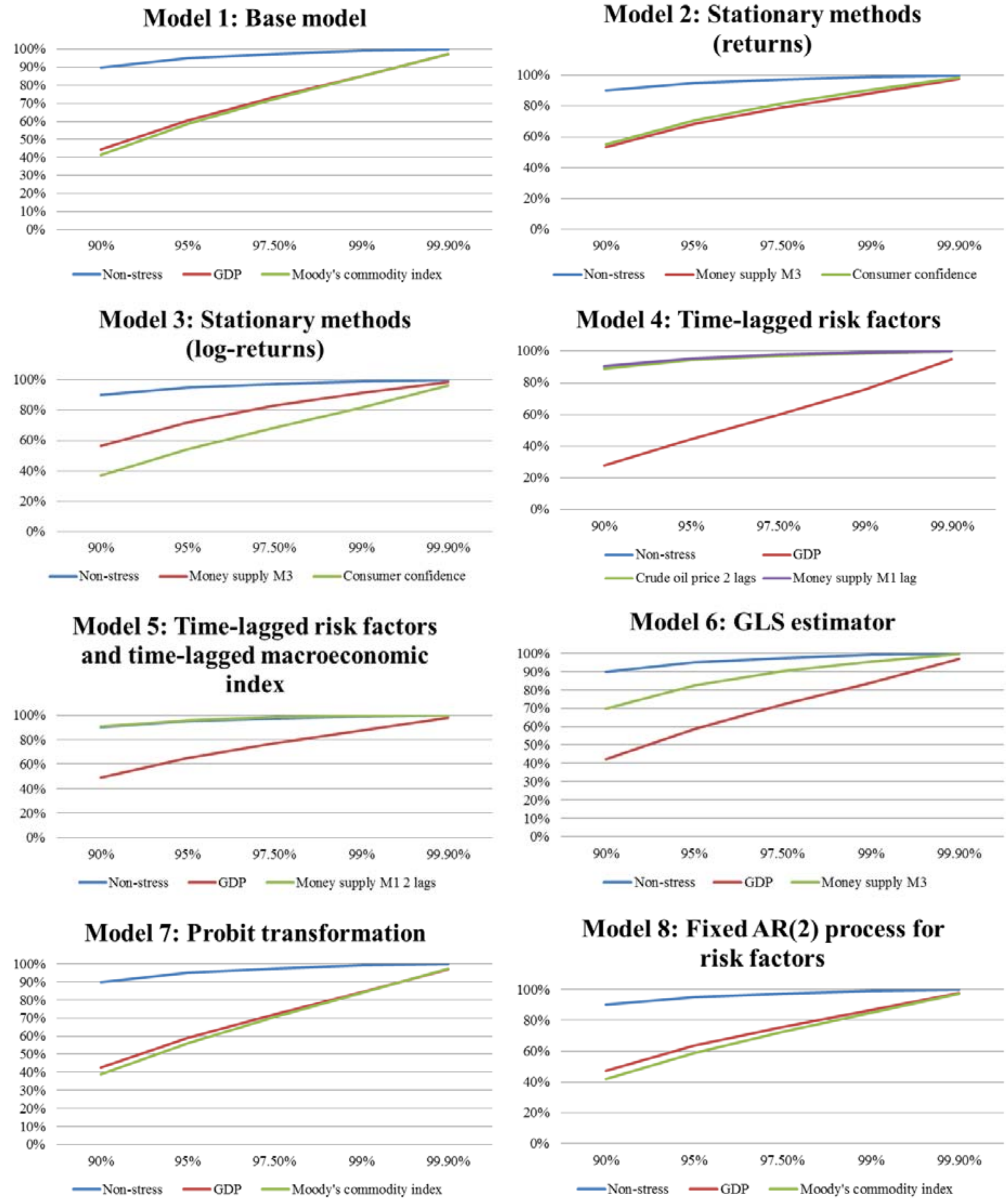
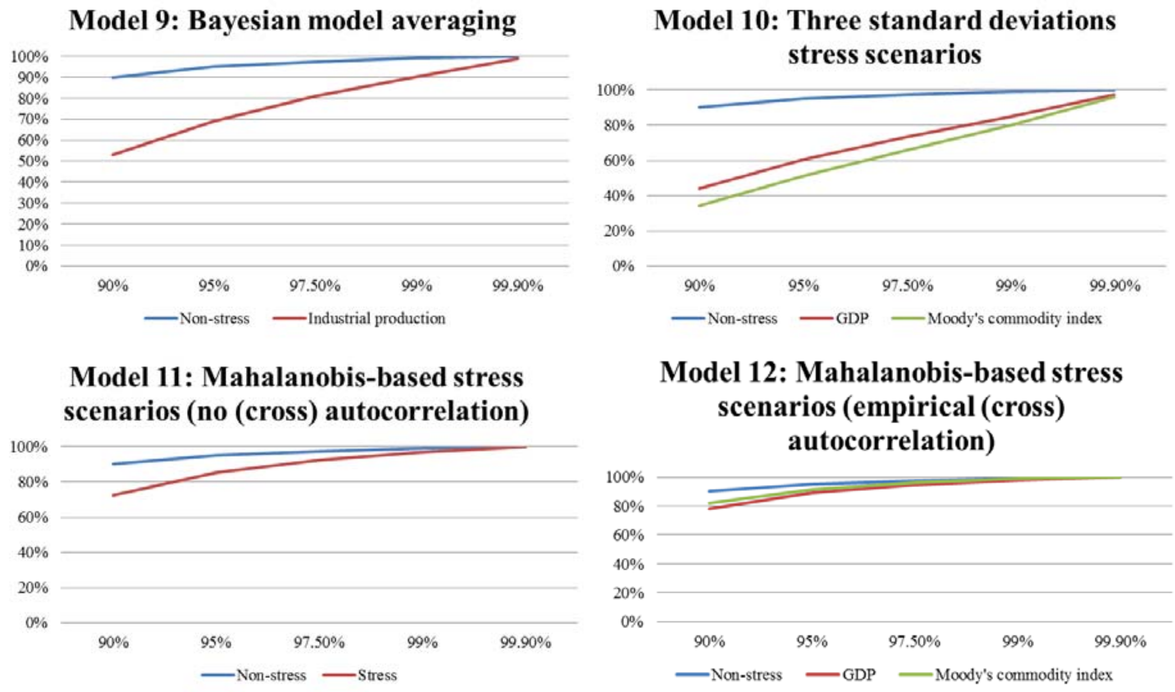


Figure 4 [continued]



Based on the idea of vertical distances between the tails of the conditional (stress scenarios) and unconditional (non-stress scenarios) cumulative density functions for the default probabilities proposed by Schechtman and Gaglianone (2012), Figure 4 shows the tail pp-plots for the various model specifications. It is assumed that a high quantile (x-axis) of the default probability distribution in the non-stress scenario is the maximum risk a bank is able to bear. The y-axis visualizes for the non-stress as well as for the stress scenarios the probability of not exceeding this specified default probability quantile. Hence, the blue line is always the identity function which corresponds to the non-stress scenario of each model specification. The other lines indicate what percentage of the forecasted default probabilities in the stress scenarios is below the respective quantiles in the non-stress scenario. The larger the vertical distance is, i.e. the more the cumulative density functions of the simulated default probabilities in the non-stress scenario and in the various stress scenarios differ, the more severe the stress scenario. This corresponds to a low probability of not exceeding the specified default probability quantile in the non-stress scenario.

Figure 5: Scatter plots of the macroeconomic index and the explanatory variables in the base model 1

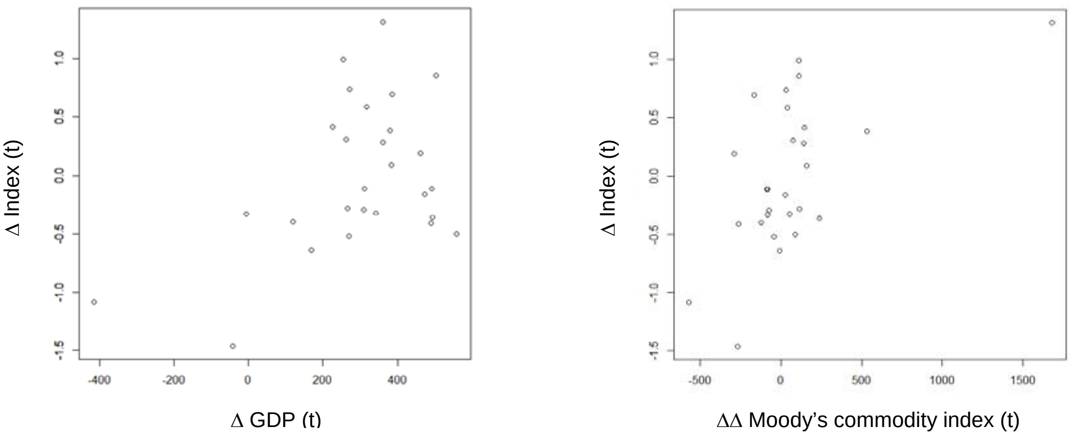


Figure 5 shows that the linear relationship between the logit-transformed realized default rates and the explanatory risk factors that is assumed in CPV-style stress test models is not exactly given for the employed sample.