# Model and estimation risk in credit risk stress tests

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

This version: 8 January 2015

**Abstract:**

Since the outbreak of the financial crisis in 2007-2009 and the subsequent European sovereign debt crisis, stress tests have experienced a real boom as a supplementary instrument in the quantitative risk management toolbox. 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, loss given defaults, 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. Our findings show that seemingly only slightly different 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: G 21, G 28, G 32

[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 second pillar of Basel II have significantly been extended (see FSA (2008, 2009), CEBS (2010), BIS (2011)). Second, regulatory authorities, such as the European Banking Authority, 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, for example, twenty percent, 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, 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, it needs a model that translates the increased values for the volatilities and correlations into stressed risk measure values (such as value-at-risk or expected shortfall). In the field of credit risk stress tests, a model is also needed 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]

If a model is needed either to translate 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), it is likely that the stress test results will depend on the modelling assumptions and the applied estimation techniques. From the perspective of the regulatory authorities, it is crucial to know how large the potential is for banks to manipulate the results of stress tests by choosing specific modelling and estimation techniques. 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 modelling assumptions and estimation techniques that are employed. To achieve this, starting from a base model specification, we employ several varia-

---

[4] For the macro stress tests performed by the European Banking Authority (EBA) in the eurozone, this is exactly what banks have to do (unless they want to employ the benchmark PD and LGD values provided by the EBA). 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)).

tions of a CreditPortfolioView (CPV)-style model.[7] All variations are statistically sound approaches and it is not obvious ex-ante 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 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 five 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), Pesaran et al. (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 reali-

[7] See Wilson (1997a, 1997b).
[8] For a more detailed survey on quantitative credit risk stress test methodologies see, for example, Foglia (2009).

zations of these risk parameters in adverse scenarios (see, for example, Rösch and Scheule (2007) or Höse and Huschens (2008)). Third, 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. Fourth, multi-risk stress test approaches are proposed. For example, Drehmann et al. (2010) present an integrated bank model for a simultaneous stress test of credit and interest rate risk. Fifth, 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 (2013)). Literature on reverse stress testing is still relatively sparse (see, for example, Grundke (2011, 2012) or Grundke and Pliszka (2013) and the papers cited therein). Finally, 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 approaches and it is not obvious ex-ante 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] Sec-

[9] The AIC and the BIC did not contradict each other in any case. Thus, prioritization was not necessary.

ond, insignificant parameters $\gamma_{i,j}$ ($i \in \{1,2,...,I\}$, $j \in \{0,1,2\}$) are set to zero ($p$-values $\leq 0.1$) and the significance of the parameters of the remaining parts of the process is checked again.[10]

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:[11]

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

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

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

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

Combining (1) to (5), the distribution of the sector-specific default probabilities for the next $m$ time periods can be computed using the following Monte Carlo simulation algorithm with $D$ simulation runs:[12]

[10] See, for example, Banque de France (2009) for a similar approach.

[11] 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.

[12] See Boss (2002, p. 81-82).

For $d = 1$ to $D$

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

(i) Using the Cholesky decomposition of the variance-covariance matrix $\Sigma$, draw random numbers for the multivariately normally distributed error terms $u_{s,n}^{(d)}$, $s \in \{1, 2, ..., S\}$, and $v_{i,n}^{(d)}$, $i \in \{1, 2, ..., I\}$.

(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. Above, in step (i), a vector of i.i.d. standard normally distributed random variables $z \in \mathbb{R}^{(S+I)\times 1}$ is multiplied by the lower triangular matrix $A$ from the Cholesky decomposition of the variance-covariance matrix $A \cdot A^T = \Sigma$. However, to perform a stress test, one or several components of the vector $z$ are replaced by standardized shocks for the systematic risk factors. In the base setting, only one risk factor is shocked and the shock is set equal to the standardized historical realization of the error term which had the most negative impact on the macroeconomic index.[13] More precisely, we define the shocked component of the vector $z$ as:

$$\frac{v'_{i,t+1}}{\sigma_{v_i}} \qquad \text{with} \qquad v'_{i,t+1} = \begin{cases} \min\limits_{n\in\{1,2,\dots,t\}} v_{i,n}, & \beta_{s,i} > 0 \\ \max\limits_{n\in\{1,2,\dots,t\}} v_{i,n}, & \beta_{s,i} < 0 \end{cases} \tag{6}$$

where $\sigma_{v_i}$ denotes the standard deviation of the error term of the shocked macroeconomic variable $x_i$. When we have $S = 1$ (which 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. The shock (6) has to be placed on the $S+1$th position of the vector $z$ and the following components as well as the variance-covariance matrix have to be rearranged accordingly. This procedure corresponds to using a conditional (on the shock for the error term) multivariate normal distribution for the remaining (not shocked) error terms. In the following, we set $m = 3$ and we nearly always[14] 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 multivariate normal distribution (4). However, of course, the initial shock propagates into the next periods according to the employed AR processes.[15] To achieve high accuracy in the Monte Carlo simulation, we employ $D = 1,000,000$ draws.

[13] See Boss (2002).
[14] The exceptions are models 10 and 11, where the stress scenario is based on the Mahalanobis distance (see Section 3.3.5.2).
[15] Even if the stressed risk factor is only modelled by 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 global yearly default rate data from Standard & Poor's ranging from 1983 to 2010.[16] 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), 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.

– 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.[17] 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

[16] See Standard & Poor's (2011, p. 68). The data was adjusted for rating withdrawals. The correlation between the global yearly speculative-grade default rates and the yearly speculative-grade default rates of US and tax havens obligors is 97%. In 2010, 75.3% of all defaults were defaults of US and tax havens obligors (see Standard & Poor's (2011, p. 68)). 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.

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

potential explanatory variables that are univariately significant.[18] 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 significant, we add the variable with the highest absolute $t$-value.[19] 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 the method.

– insert Figure 1 about here –

To ensure stationarity of the time series of the macroeconomic index as well as the time series 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 the null hypothesis of non-stationarity can be rejected in at least two tests (ADF, PP test) or the null hypothesis of stationarity cannot be rejected (KPSS test). For all three tests, the significance level is 10%.

[18] A detailed description of this approach is provided, for example, in Rawlings et al. (1998, p. 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 model specifications (when it is significant), we do not completely avoid ad-hoc selection elements (see the following description of the selection procedure and Figure 1).

[19] For our data, this was once the case, namely for model 3 (log-returns) (see Section 4.1).

Where there is a unit root in the characteristic equation of a time series model, 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 facilitate 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 of these modifications (with the exceptions of the model with fixed AR(2) processes for the risk factors (model 8) and the models with modified stress scenarios (models 9 to 11)), 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.[20] However, other techniques are statistically equally reasonable, for example, computing returns or log-returns.[21] Changing the time series transformation to achieve stationarity of the data causes some variables in the base model to become insignificant. As only significant ex-

[20] See, for example, Boss (2002, p. 73).
[21] We apply each stationary method to the data of the explanatory variable as well as to the macroeconomic index data.

planatory 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.[22]

### 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 rate.[23] Second, one and two period time-lagged realizations of the macroeconomic index $y_{1,t-1}$ and $y_{1,t-2}$ are introduced as potential explanatory variables in (1). 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. When the Godfrey-Breusch test rejects the null hypothesis of no auto-

[22] 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.
[23] See, for example, Boss (2002) for a similar approach.

correlation (up to order 4), we apply the Newey-West estimator to correct the $t$-statistics.[24] However, this only ensures a consistent but not an efficient estimation.

In the base model, the minimal $p$-value of the Godfrey-Breusch test for the index equation (1) is 0.1321. Thus, the null hypothesis of no autocorrelation is shortly not rejected at a significance level of 10%. Of course, the non-rejection of the null hypothesis is not an approval that it is true. 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:

$$Var\left(u_1 u_1'\right) = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,T} \\ \sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T,1} & \sigma_{T,2} & \cdots & \sigma_T^2 \end{pmatrix} \tag{7}$$

More specifically, we assume that the error term of the macroeconomic index equation (1) follows an AR(1) process without intercept:[25]

$$u_{1,t} = \rho_1 \cdot u_{1,t-1} + \delta_t \tag{8}$$

where the error term $\delta_t$ 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.

[24] However, as the results in Section 4 show, this was only once the case, namely when estimating the AR(2) process for the second differences of Moody's commodities index (model 8).
[25] See McNeil and Wendin (2007) or Miu and Ozdemir (2009) for a similar procedure.

#### 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, however, is not the only possible choice. One alternative is using the probit transformation:[26]

$$p_{1,t} = \Phi\left(-y_{1,t}\right) \Leftrightarrow y_{1,t} = -\Phi^{-1}\left(p_{1,t}\right) \tag{9}$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. The index $y_{1,t}$ gets a negative sign as an argument of $\Phi(\cdot)$ in (9) 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. 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 (see Table 3 in the following). In some cases, only the error term remains, 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.[27]

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

[27] For a further alternative, see, for example, Hamerle et al. (2008), who model their two explanatory risk factors using a VAR process. We also tried to improve the risk factor modelling by adapting a GARCH(1,1)-process to the fitted residuals. However, for the second differences of Moody's commodities index, the corresponding GARCH process was non-stationary due to two extreme outliers in 2009 and 2010 (see also in Section 4.2 the discussion of the effects of these outliers on the forecasted default probabilities in the base model). Thus, we did not pursue this modelling approach any more.

### 3.3.4 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.

#### 3.3.4.1 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.[28]

#### 3.3.4.2 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.[29] The Mahalanobis distance of a random vector $v$ is defined as:

$$Maha(v) = \sqrt{(\mu - v)^T \cdot \Sigma^{-1} \cdot (\mu - v)} \tag{10}$$

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

$$Ell_\tau := \left\{ v \in \mathbb{R}^N \middle| Maha(v) \le \tau \right\}. \tag{11}$$

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

As in our base model we have three explanatory variables ($I = 3$) and as we consider a dynamic three-period stress test ($m = 3$), the dimension of the random vector $v = (v_{1,t+1}, v_{2,t+1}, v_{3,t+1}, v_{1,t+2}, v_{2,t+2}, v_{3,t+2}, v_{1,t+3}, v_{2,t+3}, v_{3,t+3})^T$ is $N = 9$. The random vector $v$ represents a three-dimensional path of the error terms of the macroeconomic risk factors over the three considered periods. We assume $\tilde{v} = (v_{1,n}, v_{2,n}, v_{3,n})^T \sim N(0, \Sigma_{vv})$ for all $n \in \{t+1, t+2, t+3\}$ (see (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 covariance matrix of the random vector $v \in \mathbb{R}^9$.

Using the above notation, the three standard deviation stress scenarios in Section 3.3.4.1 can be represented by $v_1^* = (3\sigma_{v_1}, 0, \ldots, 0)^T$, $v_2^* = (0, 3\sigma_{v_2}, 0, \ldots, 0)^T$, $v_3^* = (0, 0, 3\sigma_{v_3}, 0, \ldots, 0)^T \in \mathbb{R}^9$ with $Maha(v_1^*) = 3.16$, $Maha(v_2^*) = 3.11$ and $Maha(v_3^*) = 3.14$ (in the case of assumed non-existence of serial (cross) correlation). When we employ the empirical (cross) autocorrelations, we get values for the Mahalanobis distance of 3.40, 2.99 and 3.04. To ensure consistency between the univariate stress scenarios as set out in Section 3.3.4.1 and those ones employed in this section, we define three trust regions $Ell_\tau$ by setting $\tau \in \{3.16, 3.11; 3.14\}$ and $\tau \in \{3.40, 2.99; 3.04\}$, respectively. 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 three trust regions $Ell_\tau$, we look for the scenario which minimizes the expected sum of the changes in the macroeconomic index over the three considered periods:

$$v_\tau^{worst} = \arg\min_{v \in Ell_\tau} \left\{ E\left[ \sum_{n=1}^{3} \Delta y_{1,t+n}(u_{1,t+n}, v) \middle| F_t, v \right] \right\}. \tag{12}$$

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 (12) nearly corresponds to the worst case scenario (for a given plausibility) in the sense of expected default 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 (as described in Section 3.1), we effectively obtain a mixture of model specifications (see Table 3). We use first and second-order autoregressive processes both coupled with or without an intercept. In some cases, we even describe the evolution of the risk factors over time only by the error term (with and without an intercept). The specification of the AR processes has an influence on how long it takes until an initial shock vanishes. The coefficient of determination $R^2$ ranges from 15.0% to 74.2% and the values for the adjusted $R^2$ are between 11.5% and 71.7%.

Having applied the stepwise regression approach, we include the first differences in the GDP and in the crude oil price (WTI)[30] as well as the second differences in Moody's commodities index as explanatory variables in the base model (see Table 4). The explained variance of the model is 48.4% and the adjusted $R^2$ is 41.3%. We also tested 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 can usually be observed in economically good times as well as higher prices for commodities and for oil 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 3 (log-returns), GDP is significant in all models. The adjusted $R^2$ ranges from 27.8% to 64.1%. The best fit in terms of the adjusted $R^2$ shows model 5 with time-lagged risk factors and the time-lagged macroeconomic index as additional explanatory variables. However, based on AIC and BIC, this is one of the worst models. In terms of these information criteria, model 3 (log-returns) is the best model.

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 (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 future risk factor realizations are forecasted by (3),

[30] This is the FOB (free on board) price, which does not consider the final delivery costs.

where the error terms are set equal to their means zero.[31] 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 5. Based on 1 million forecasts of the default rates for 2011 to 2013, Table 5 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 ($CMSE$):

$$MD_n = E\left[ p_n - PD_n^{realized} \middle| F_t \right], \tag{13}$$

$$MSE_n = E\left[ \left( p_n - PD_n^{realized} \right)^2 \middle| F_t \right], \tag{14}$$

$$CMSE = \sum_{s=1}^{3} E\left[ \left( p_{t+s} - PD_{t+s}^{realized} \right)^2 \middle| F_t \right], \tag{15}$$

where $F_t$ denotes the available information up to time $t$. The only good news with respect to 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.[32] The best performing models (in terms of (cumulative) mean squared errors) are those with time-lagged explanatory variables (models 4 and 5).

– insert Figure 2 and Table 5 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 (according to the algorithm described in Section 3.1). Depending on the method to make the data stationary, the macroeconomic in-

[31] The realized default rates for the years 2011 to 2013 are taken from Standard & Poor's (2014).
[32] See the following Section 4.2 for a detailed discussion of this observation in the case of the base model 1.

dex $y_n$ (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): $y_n = y_{n-1} + \Delta y_n$ (16)

Model 2 (returns): $y_n = y_{n-1} \cdot (1 + R_n)$ (17)

Model 3 (log-returns): $y_n = y_{n-1} \cdot e^{R_n^{\ln}}$ (18)

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

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 6 to 8 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 between 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 2.9% and that the largest default rate in the whole time period 1983 to 2010 was 11.1%.[33] 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 commodities index in 2009 and 2010. These are very low (-1,279.23 in 2009) and very high (2,559.36 in 2010), respectively. In 2009, this value corresponds to a difference of -2.25 times the standard deviation from the mean. In 2010, this difference is as much as 4.17 times the standard deviation.[34] Through the AR(2) process, by which the second differences of Moody's commodities 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 -2,344.74. This corresponds to a difference of -4.03 times the standard deviation from the mean of the second differences of the index. This explains the very high expected default probability of 16.8% in $t+2$ in the non-stress case of the base model. Without the data points of 2010, the expected default probabilities in the non-stress case of the base model would be 6.12% ($t+1$), 6.13% ($t+2$) and 6.13% ($t+3$). Without the data points of 2009 and 2010, the results would be 3.41% ($t+1$), 4.62% ($t+2$) and 5.33% ($t+3$) and, hence, much less extreme than those results that we get when we use the full data sample to estimate the base model. 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.

[33] The minimum value is 0.89%, the mean default rate is 4.5% and the standard deviation is 2.8%.

[34] 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.

– insert Table 6 and Figure 3 about here –

In model 4, where the two period time-lagged first difference in GDP is used as the explanatory variable in the index equation (1), and in model 5, where the two period time-lagged first difference in the macroeconomic index itself is employed as the explanatory variable, a stress event in these lagged variables in $t+1$ mainly has an effect on the stressed default probabilities in $t+3$. 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 $t+2$.

Table 7 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 7) and smallest (lower part of Table 7), respectively, forecasted stressed default probabilities in the base model 1 and in one of the other model specifications 2 to 8 (separated with respect to the expected forecasted stressed default probability and the 99.9% quantile and with respect to the time period).[35] 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 7 about here –

Table 8 shows the effect of changing the stress scenario definition on the forecasted default probabilities. In models 9 to 11, 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). A comparison of the results for models 10 and 11 (see Table 6) shows that the inclusion of empirical (cross) auto-

[35] The results of models 9 to 11 are separately compared with those of the base model 1 because in these model modifications the definition of the stress scenarios is altered.

correlations in the variance-covariance matrix $\Sigma \in \mathbb{R}^{9\times 9}$ only has a minor effect on the forecasted stressed default probabilities. Comparing the results of model 9 with those of models 10 and 11, the stressed default probabilities of models 10 and 11 in the periods $t+2$ and $t+3$ are larger than those ones of model 9. 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 8 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 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.[36] 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 eight out of eleven model specifications, Figure 4 shows that the extent of the vertical distances can vary considerably with the considered model specification.

[36] For Figure 4, only the first future period $t+1$ has been considered.

– 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 important implications. First, when doing the stress simulations for the future default probabilities, the shock for the error term would have to be placed on the 1st position of the vector $z$ (instead of the $S+1$th position) that is multiplied with the lower triangular matrix of the Cholesky decomposition of the variance-covariance matrix $\Sigma$. 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 estimators 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 use the CPV model the possibility $\Sigma_{u,v} \neq 0$ is not directly excluded, but the issue of endogeneity is rarely explicitly addressed.[37]

As we only *assumed* $\Sigma_{u,v} = \Sigma_{v,u} = 0$ and as an endogeneity problem might exist even if this assumption is true (for example because of missing correlated variables in the index equation), we test for endogeneity of each of the explanatory variables (ΔGDP ($t$), ΔOil price WTI (FOB) ($t$) and ΔΔMoody's commodities index ($t$)) in our base model 1. For this purpose, the Hausman test is employed. To perform this test, we need instrument variables (IV) that are

[37] See, for example, Boss (2002) or Virolainen (2004). An exception is Schechtman and Gaglianone (2012).

strong and exogeneous.[38] First, as in Schechtman and Gaglianone (2012), we try the one period lagged risk factors as IV. However, the lagged variables ΔGDP ($t-1$) and ΔOil price WTI (FOB) ($t-1$) prove to be weak IV, because their $F$-statistics are 4.15 and 5.12, respectively. Only the lagged variable ΔΔMoody's commodities index ($t-1$) can be considered to be strong because its $F$-statistic is 16.85 and, hence, sufficiently above 10. Data from the World Development Indicator 2012 of the World Bank is employed to find strong IV for GDP and the oil price. Out of the more than 100 variables, four contemporary variables are strong IV for ΔGDP ($t$). The variable with the largest $F$-statistic (38.41) is the first difference in 'external balance on goods and services (constant LCU)'.[39] For ΔOil price WTI (FOB) ($t$), the contemporary variable 'portfolio equity, net inflows (BoP, current USD)' is the only strong IV with an $F$-statistic of 22.68.[40] Next, the two latter IV are checked for exogeneity using a further Hausman test.[41] For ΔExternal balance on goods and services ($t$), the one period lagged variable is used as IV ($F$-statistic 14.57) and the null hypothesis of exogeneity cannot be rejected ($p$-value: 0.164). For ΔPortfolio equity, net inflows ($t$), the one period lagged variable is no strong IV, but the one period lagged variables 'net income from abroad (current LCU)' and 'net current transfers from abroad (constant LCU)' (first differences) prove to be strong IV. With these IV, the exogeneity of ΔPortfolio equity, net inflows ($t$) cannot be rejected ($p$-value of at least 0.267).[42]

[38] The IV parameter estimates needed for the Hausman test statistic are computed using Two Stage Least Squares (2SLS).

[39] To reach stationarity, first differences have been computed. LCU stands for 'local currency unit'.

[40] Again, first differences have been computed. According to the World Bank, the variable 'external balance on goods and services (% of GDP)' (formerly resource balance) is defined as exports of goods and services minus imports of goods and services (previously nonfactor services). The variable 'portfolio equity, net inflows (BoP, current USD)' includes net inflows from equity securities other than those recorded as direct investment and including shares, stocks, depository receipts (American or global), and direct purchases of shares in local stock markets by foreign investors.

[41] The exogeneity of the lagged variable ΔΔMoody's commodities index ($t-1$) as IV for ΔΔMoody's commodities index ($t$) is just assumed.

[42] The exogeneity of the IV for the IV is just assumed.

Finally, the Hausman test for endogeneity of the explanatory variables of the base model 1 is carried out. As Table 9 shows, the null hypothesis of exogeneity cannot be rejected in any case. 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 26 data points is not very large.

- insert Table 9 about here -

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.[43]

- 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

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

statistical estimation of the parameters $\beta_{s,0},...,\beta_{s,I}$ in (1), but also within the simulation of future (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) could change the density functions of the forecasted stressed default probabilities in Figure 3 (and, hence, the quantiles) considerably.

## 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 approaches and it was not obvious ex-ante why one specification or estimation technique should be more adequate than another specification or technique. We showed that the chosen model specifications and the employed estimation techniques 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. For example, for an assessment of the idiosyncratic risk of a single bank, it may be sufficient to directly employ stressed risk parameters. In contrast, for a standardized system-wide stress test across various jurisdictions, the use of directly stressed risk parameters given by the regulatory authorities appears not to be adequate. This approach might be well-suited to 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, Working Paper, Bonn University, Deutsche Bundesbank.

Bonti, G., M. Kalkbrener, C. Lotz, G. Stahl (2006): Credit risk concentrations under stress, in: 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, in: 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, in: 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? in: 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, in: 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, in: Journal of Banking & Finance, Vol. 34, No. 4, pp. 713-729.

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.
Financial Services Authority (FSA) (2008): Stress and Scenario Testing. Consultant 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, in: International Journal of Central Banking, Vol. 5, No. 3, pp. 9-45.
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.
Grundke, P. (2011): Reverse stress tests with bottom-up approaches, in: Journal of Risk Model Validation, Vol. 5, No. 1, pp. 71-90.
Grundke, P. (2012): Further recipes for quantitative reverse stress testing, in: Journal of Risk Model Validation, Vol. 6, No. 2, pp. 81-102.
Grundke, P., K. Pliszka (2013): 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, in: RisikoManager, No. 8, pp. 1, 8-15.
Höse, S., S. Huschens (2008): Worst-case asset, default and survival time correlations, in: 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, in: 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, S. 64-82.
Maddala, G. (1983): Limited Dependent and Qualitative Variables in Econometrics, Cambridge University Press, New York.
McNeil, A.J., J.P. Wendin (2007): Bayesian inference for generalized linear mixed models of portfolio credit risk, in: 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, in: Journal of Risk Model Validation, Vol. 3, No. 4, pp. 3-38.

Pesaran, M.H., T. Schuermann, B.-J. Treutler, S.M. Weiner (2006): Macroeconomic dynamics and credit risk: a global perspective, in: Journal of Money, Credit and Banking, Vol. 38, No. 5, pp. 1211-1262.

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, in: 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, in: Journal of Financial Stability, Vol. 8, No. 3, pp. 174-192.

Simons, D., F. Rolwes (2009): Macroeconomic Default Modeling and Stress Testing, in: 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, in: 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, in: Risk, Vol. 10, No. 9, pp. 111-117.

Wilson, T.C. (1997b): Portfolio credit risk, Part II, in: 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, in: Journal of Risk Model Validation, Vol. 2, No. 1, pp. 3-23.

## Tables

### Table 1: Endogenous and exogenous variables

| | |
|---|---|
| **Endogenous variable** | |
| Default rates | Standard & Poor’s (2011) |
| **Exogenous variables** | |
| **Economic activity indicators** | |
| Gross domestic product (GDP) | Datastream: USGDP...D |
| Industrial production | Datastream: USIPTOT.G |
| **Price stability indicators** | |
| Inflation | Datastream: USCONPRCE |
| Money supply M1 | Datastream: USM1....B |
| Money supply M3 | Datastream: USYMA013Q |
| Moody’s commodities index (price index) | Datastream: MOCMDTY |
| Reuter’s commodities index (price index) | Datastream: RECMDTY |
| **Household indicators** | |
| Personal consumption expenditure | Datastream: USCNPER.B |
| Disposable personal income | Datastream: USPERDISB |
| New home sales | Datastream: USHOUSESE |
| Unemployment rate | Datastream: USUN%TOTQ |
| **Firm indicators** | |
| NAHB/Wells Fargo housing market index | Datastream: USNAHBMI |
| Consumer confidence | Datastream: USCNFCONQ |
| Consumer sentiment | Datastream: USUMCONSH |
| **Financial market indicators** | |
| 3M-treasury bill rate | Datastream: USGBILL3 |
| S&P 500 | Datastream: S&PCOMP |
| **External indicators** | |
| Exports | Datastream: USEXPGDSB |
| Imports | Datastream: USIMPGDSB |
| Japanese Yen/USD exchange rate | Datastream: JPXRUSD |
| USD/GBP exchange rate | Datastream: STUSBOE |
| Oil price Brent (FOB) per Barrel | Datastream: OILBREN |
| Oil price Brent per Barrel | Datastream: OILBRDT |
| Oil price WTI (FOB) per Barrel | Datastream: OILWTXI |
| Oil price WTI per Barrel | Datastream: CRUDOIL |

**Table 2:**

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

| | **Base model** | **Modifications** |
|---|---|---|
| Stationary method | First differences and, if necessary, second differences | Returns (if necessary second returns), log-returns |
| Time-lagged risk factors | No | Additional to contemporary variables, time-lagged macroeconomic variables and time-lagged macroeconomic index as explanatory variables for the macroeconomic index |
| Estimator | OLS | GLS |
| Time series processes for macroeconomic variables | AR(1)/AR(2) (based on AIC/BIC with only significant parameters) | Fixed AR(2) |
| Transformation between default rate and macroeconomic index | Logit | Probit |
| Stress test scenario | Historical worst case scenario | Hypothetical scenarios based on three standard deviations of the error terms and based on the Mahalanobis distance |

**Table 3: Estimates of the risk factor processes**

| | | Parameters | AIC/ BIC | $R^2$ | Adjusted $R^2$ | Applied specification |
|---|---|---|---|---|---|---|
| **Model 1: Base model** | | | | | | |
| Δ GDP (t) | (Intercept) | 175.7788** | 338.9/ | 0.1529 | 0.1161 | AR(1) |
| | t-1 | 0.392* | 342.6 | | | |
| Δ Oil price WTI (FOB) (t) | t-1 | -0.3937** | 201.5/ 203.9 | 0.1529 | 0.1176 | AR(1) without intercept |
| ΔΔ Moody's commodities index (t) | t-1 | -1.2539*** | 351.4/ | 0.7332 | 0.7089 | AR(2) without intercept |
| | t-2 | -1.5297*** | 355.0 | | | |
| **Model 2: Stationary methods (returns)** | | | | | | |
| Return GDP (t) | | | 195.3/ 196.5 | - | - | Error term |
| Return Moody's commodities index (t) | (Intercept) | 0.07181* | -12.7/ -10.2 | - | - | Intercept |
| **Model 3: Stationary methods (log-returns)** | | | | | | |
| Log-return Money supply M3 (t) | | | 29.4/ 30.7 | - | - | Error term |
| Log-return Imports (t) | (Intercept) | 0.07418*** | -46.5/ -44.0 | - | - | Intercept |
| **Model 4: Time-lagged risk factors** | | | | | | |
| Δ GDP (t-2) | (Intercept) | 155.208* | 309.6/ | 0.1785 | 0.1412 | AR(1) |
| | t-1 | 0.4735** | 313.1 | | | |
| Δ Imports (t) | (Intercept) | 62684* | 671.3/ 673.7 | - | - | Intercept |
| Δ Oil price WTI (FOB) (t) | t-1 | -0.3864* | 193.4/ 195.8 | 0.1541 | 0.1173 | AR(1) without intercept |
| **Model 5: Time-lagged risk factors and time-lagged macroeconomic index** | | | | | | |
| Δ Macroeconomic index (t-2) | | | 43.7/ 45.0 | - | - | Error term |
| Δ GDP (t) | (Intercept) | 176.8452** | 326.6/ | 0.1535 | 0.115 | AR(1) |
| | t-1 | 0.3927* | 330.1 | | | |
| Consumer confidence (t) | (Intercept) | 37.4244*** | 195.0/ | 0.6714 | 0.6385 | AR(2) |
| | t-1 | 1.1975*** | 199.5 | | | |
| | t-2 | -0.5843*** | | | | |
| **Model 6: GLS estimator** | | | | | | |
| Δ GDP (t) | (Intercept) | 176.7004** | 351.3/ | 0.1501 | 0.1146 | AR(1) |
| | t-1 | 0.381* | 355.1 | | | |
| Δ Oil price WTI (FOB) (t) | t-1 | -0.3944* | 208.4/ 210.9 | 0.1533 | 0.1195 | AR(1) without intercept |
| Δ Money supply M3 (t) | | | 112.6/ 113.9 | - | - | Error term |
| **Model 7: Probit transformation** | | | | | | |
| Δ GDP (t) | (Intercept) | 175.7788** | 339.0/ | 0.1529 | 0.1163 | AR(1) |
| | t-1 | 0.392* | 342.6 | | | |
| ΔΔ Moody's commodities index (t) | t-1 | -1.2539*** | 351.4/ | 0.7332 | 0.7089 | AR(2) without intercept |
| | t-2 | -1.5297*** | 355.0 | | | |
| **Model 8: Fixed AR(2) process for risk factors** | | | | | | |
| Δ GDP (t) | (Intercept) | 270.6948*** | 326.3/ | 0.2307 | 0.1575 | AR(2) (fixed) |
| | t-1 | 0.6136** | 331.0 | | | |
| | t-2 | -0.4978 | | | | |
| Δ Oil price WTI (FOB) (t) | (Intercept) | 4.1703 | 194.5/ | 0.2144 | 0.1396 | AR(2) (fixed) |
| | t-1 | -0.511** | 199.2 | | | |
| | t-2 | -0.1801 | | | | |
| ΔΔ Moody's commodities index (t) | (Intercept) | 69.5162 | 352.3/ | 0.7416 | 0.717 | AR(2) (fixed)# |
| | t-1 | -1.2384*** | 357.0 | | | |
| | t-2 | -1.5537*** | | | | |

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. 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 ΔΔ denote first and second differences. When the specification only consists of the error term or an intercept plus error term, calculation of $R^2$ or the adjusted $\bar{R}^2$ 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 | $R^2$ | Adj. $R^2$ | AIC | BIC |
|---|---|---|---|---|---|---|
| **Model 1: Base model** | | >0.1321 | 0.4837 | 0.4132 | 43.1 | 49.3 |
| Intercept | -0.3377* | | | | | |
| Δ GDP (t) | 0.001* | | | | | |
| Δ Oil price WTI (FOB) (t) | 0.0147* | | | | | |
| ΔΔ Moody’s commodities index (t) | 0.0004** | | | | | |
| **Model 2: Stationary methods (returns)** | | >0.2218 | 0.3335 | 0.2779 | -10.3 | -6.5 |
| Return GDP (t) | -0.007* | | | | | |
| Return Moody’s commodities index (t) | 0.5452*** | | | | | |
| **Model 3: Stationary methods (log-returns)** | | >0.568 | 0.4201 | 0.3718 | -16.6 | -11.4 |
| Intercept | -0.0756* | | | | | |
| Log-return money supply M3 (t) | -0.1923** | | | | | |
| Log-return imports (t) | 0.8722** | | | | | |
| **Model 4: Time-lagged risk factors** | | >0.1524 | 0.5984 | 0.541 | 36.3 | 42.4 |
| Intercept | 0.4142* | | | | | |
| Δ GDP (t-2) | -0.0018*** | | | | | |
| Δ Imports (t) | 0.000002*** | | | | | |
| Δ Oil price WTI (FOB) (t) | 0.0192*** | | | | | |
| **Model 5: Time-lagged risk factors and time-lagged macroeconomic index** | | >0.6426 | 0.6858 | 0.6409 | 30.1 | 36.2 |
| Intercept | 1.0193*** | | | | | |
| Δ Macroeconomic index (t-2) | -0.282* | | | | | |
| Δ GDP (t) | 0.0021*** | | | | | |
| Consumer confidence (t) | -0.0164*** | | | | | |
| **Model 6: GLS estimator** | | - | - | - | 74.1 | 80.9 |
| Intercept | -0.4596** | | | | | |
| Δ GDP (t) | 0.0013** | | | | | |
| Δ Oil price WTI (FOB) (t) | 0.0137** | | | | | |
| Δ Money supply M3 (t) | -0.1072* | | | | | |
| **Model 7: Probit transformation** | | >0.2834 | 0.4123 | 0.3612 | 2.3 | 7.3 |
| Intercept | -0.1467* | | | | | |
| Δ GDP (t) | 0.0005* | | | | | |
| ΔΔ Moody’s commodities index (t) | 0.0002** | | | | | |

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. 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.

**Table 5: Out-of-sample performance**

| | t+1 (2011) | t+2 (2012) | t+3 (2013) | |
|---|---|---|---|---|
| **Realized default rates** | 1.80% | 2.52% | 2.23% | |
| | | | | |
| **Mean deviation *$MD_n$*** | | | | |
| Model 1 Base model | 4.28% | 14.28% | 0.87% | |
| Model 2 Stationary methods (returns) | 1.40% | 1.07% | 1.72% | |
| Model 3 Stationary methods (log-returns) | 1.56% | 1.37% | 2.15% | |
| Model 4 Time-lagged risk factors | -0.81% | -1.27% | -0.68% | |
| Model 5 Time-lagged risk factors and time-lagged macroeconomic index | -0.68% | -1.20% | -0.56% | |
| Model 6 GLS estimator | 1.99% | 2.39% | 4.20% | |
| Model 7 Probit transformation | 3.73% | 11.32% | 0.36% | |
| Model 8 Fixed AR(2) process for risk factors | 2.32% | 7.82% | -0.58% | |
| | | | | |
| **Mean squared error *$MSE_n$*** | | | | ***CMSE*** |
| Model 1 Base model | 0.003217 | 0.033652 | 0.001462 | 0.038332 |
| Model 2 Stationary methods (returns) | 0.000910 | 0.002061 | 0.003906 | 0.006878 |
| Model 3 Stationary methods (log-returns) | 0.000714 | 0.001244 | 0.002175 | 0.004132 |
| Model 4 Time-lagged risk factors | 0.000108 | 0.000307 | 0.000451 | 0.000866 |
| Model 5 Time-lagged risk factors and time-lagged macroeconomic index | 0.000087 | 0.000303 | 0.000600 | 0.000990 |
| Model 6 GLS estimator | 0.001157 | 0.003363 | 0.008645 | 0.013164 |
| Model 7 Probit transformation | 0.002343 | 0.019717 | 0.000968 | 0.023028 |
| Model 8 Fixed AR(2) process for risk factors | 0.001168 | 0.012252 | 0.000430 | 0.013851 |

Table 5 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 forecasts of the default probabilities for 2011 to 2013.

**Table 6: Forecasted default probabilities**

| | | t+1 | t+2 | t+3 |
|---|---|---|---|---|
| **Model 1: Base model** | | | | |
| **Mean** | | | | |
| | Non-stress | 6.08% | 16.80% | 3.10% |
| | GDP | 12.49% | 29.50% | 6.97% |
| | Crude oil price | 10.76% | 23.03% | 4.98% |
| | Commodities index | 9.65% | 18.88% | 3.73% |
| **99.9% quantile** | | | | |
| | Non-stress | 27.17% | 70.02% | 33.80% |
| | GDP | 41.40% | 81.97% | 53.01% |
| | Crude oil price | 37.93% | 77.37% | 44.85% |
| | Commodities index | 35.48% | 72.47% | 38.12% |

| | | | | |
|---|---|---|---|---|
| **Model 2: Stationary methods (returns)** | | | | |
| **Mean** | | | | |
| | Non-stress | 3.20% | 3.59% | 3.95% |
| | GDP | 3.77% | 4.21% | 4.61% |
| | Commodities index | 5.00% | 5.52% | 5.96% |
| **99.9% quantile** | | | | |
| | Non-stress | 21.95% | 39.31% | 55.15% |
| | GDP | 23.48% | 42.55% | 58.55% |
| | Commodities index | 27.38% | 48.93% | 65.28% |

| | | | | |
|---|---|---|---|---|
| **Model 3: Stationary methods (log-returns)** | | | | |
| **Mean** | | | | |
| | Non-stress | 3.36% | 3.89% | 4.38% |
| | Money supply M3 | 9.48% | 9.99% | 10.46% |
| | Imports | 6.07% | 6.65% | 7.18% |
| **99.9% quantile** | | | | |
| | Non-stress | 13.50% | 19.60% | 24.20% |
| | Money supply M3 | 20.76% | 27.22% | 31.27% |
| | Imports | 16.41% | 23.15% | 27.78% |

**Table 6 [continued]**

| | | t+1 | t+2 | t+3 |
|---|---|---|---|---|
| **Model 4: Time-lagged risk factors** | | | | |
| **Mean** | | | | |
| | Non-stress | 0.99% | 1.25% | 1.55% |
| | GDP (t-2) | 1.02% | 1.28% | 2.13% |
| | Imports | 3.82% | 4.51% | 5.72% |
| | Crude oil price | 1.89% | 1.93% | 2.55% |
| **99.9% quantile** | | | | |
| | Non-stress | 5.23% | 10.85% | 19.85% |
| | GDP (t-2) | 5.38% | 10.94% | 24.21% |
| | Imports | 13.37% | 27.37% | 46.09% |
| | Crude oil price | 8.17% | 15.28% | 28.25% |

| | | t+1 | t+2 | t+3 |
|---|---|---|---|---|
| **Model 5: Time-lagged risk factors and time-lagged macroeconomic index** | | | | |
| **Mean** | | | | |
| | Non-stress | 1.12% | 1.32% | 1.67% |
| | GDP | 3.27% | 4.36% | 5.12% |
| | Index (t-2) | 1.29% | 1.64% | 2.69% |
| | Consumer confidence | 0.85% | 1.29% | 2.31% |
| **99.9% quantile** | | | | |
| | Non-stress | 4.96% | 11.21% | 23.89% |
| | GDP | 8.64% | 23.28% | 46.27% |
| | Index (t-2) | 5.59% | 13.23% | 31.65% |
| | Consumer confidence | 3.54% | 10.89% | 29.49% |

| | | | | |
|---|---|---|---|---|
| **Model 6: GLS estimator** | | | | |
| **Mean** | | | | |
| | Non-stress | 3.79% | 4.91% | 6.43% |
| | GDP | 9.80% | 14.08% | 18.33% |
| | Crude oil price | 6.78% | 7.92% | 10.67% |
| | Money supply M3 | 5.74% | 7.50% | 9.73% |
| **99.9% quantile** | | | | |
| | Non-stress | 21.52% | 43.75% | 65.63% |
| | GDP | 37.62% | 67.84% | 85.59% |
| | Crude oil price | 30.87% | 55.35% | 77.26% |
| | Money supply M3 | 26.13% | 51.86% | 73.72% |

| | | | | |
|---|---|---|---|---|
| **Model 7: Probit transformation** | | | | |
| **Mean** | | | | |
| | Non-stress | 5.53% | 13.84% | 2.59% |
| | GDP | 9.91% | 22.34% | 5.52% |
| | Commodities index | 8.05% | 14.93% | 2.97% |
| **99.9% quantile** | | | | |
| | Non-stress | 20.96% | 51.48% | 24.99% |
| | GDP | 29.05% | 63.06% | 37.10% |
| | Crude oil price | 25.65% | 53.32% | 27.29% |

**Table 6 [continued]**

| | | t+1 | t+2 | t+3 |
|---|---|---|---|---|
| **Model 8: Fixed AR(2) processes for risk factors** | | | | |
| **Mean** | | | | |
| | Non-stress | 4.12% | 10.34% | 1.65% |
| | GDP | 8.13% | 20.00% | 3.44% |
| | Crude oil Price | 7.40% | 14.22% | 2.43% |
| | Commodities index | 6.79% | 11.84% | 1.91% |
| **99.9%-quantile** | | | | |
| | Non-stress | 18.80% | 54.06% | 18.98% |
| | GDP | 29.19% | 69.86% | 31.90% |
| | Crude oil price | 27.35% | 62.67% | 25.42% |
| | Commodities index | 25.86% | 57.31% | 21.18% |

| | | t+1 | t+2 | t+3 |
|---|---|---|---|---|
| **Model 9: Three standard deviations stress scenario** | | | | |
| **Mean** | | | | |
| | Non-stress | 6.08% | 16.80% | 3.10% |
| | GDP | 12.71% | 29.88% | 7.11% |
| | Crude oil price | 11.34% | 23.68% | 5.20% |
| | Commodities index | 11.00% | 19.50% | 3.94% |
| **99.9%-quantile** | | | | |
| | Non-stress | 27.17% | 70.02% | 33.80% |
| | GDP | 41.91% | 82.27% | 53.59% |
| | Crude oil price | 39.38% | 78.08% | 46.07% |
| | Commodities index | 39.00% | 73.35% | 39.50% |

| | | t+1 | t+2 | t+3 |
|---|---|---|---|---|
| **Model 10: Mahalanobis-based stress scenario (no (cross) autocorrelation)** | | | | |
| **Mean** | | | | |
| | Non-stress | 6.08% | 16.80% | 3.10% |
| | GDP (equiv.) | 11.68% | 37.60% | 16.60% |
| | Crude oil price (equiv.) | 11.54% | 36.83% | 15.88% |
| | Commodities index (equiv.) | 11.44% | 35.15% | 14.40% |
| **99.9%-quantile** | | | | |
| | Non-stress | 27.17% | 70.02% | 33.80% |
| | GDP (equiv.) | 36.24% | 83.67% | 70.38% |
| | Crude oil price (equiv.) | 35.98% | 83.32% | 69.25% |
| | Commodities index (equiv.) | 35.55% | 81.91% | 66.55% |

| | | t+1 | t+2 | t+3 |
|---|---|---|---|---|
| **Model 11: Mahalanobis-based stress scenario (empirical (cross) autocorrelation)** | | | | |
| **Mean** | | | | |
| | Non-stress | 6.08% | 16.80% | 3.10% |
| | GDP (equiv.) | 12.31% | 39.69% | 18.72% |
| | Crude oil price (equiv.) | 11.26% | 35.88% | 14.98% |
| | Commodities index (equiv.) | 11.21% | 34.41% | 13.73% |
| **99.9%-quantile** | | | | |
| | Non-stress | 27.17% | 70.02% | 33.80% |
| | GDP (equiv.) | 37.66% | 84.96% | 73.65% |
| | Crude oil price (equiv.) | 35.31% | 82.68% | 67.61% |
| | Commodities index (equiv.) | 35.00% | 81.36% | 65.19% |

Table 6 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 10 and 11, the stress test scenarios are characterized by the most harmful (in the sense of (12)) scenarios out of those trust regions $Ell_{\tau}$ that correspond to the respective three standard deviations stress of the macroeconomic variables in the base model (see Section 3.3.4.2).

**Table 7:**
**Percentage differences between forecasted stressed default probabilities in the base model 1 and in model modifications 2 to 8**

| | | t+1 | t+2 | t+3 |
|---|---|---|---|---|
| **Stressed default probabilities (maximum)** | | | | |
| **Mean** | | | | |
| | Max | -20.6% | -24.3% | 163.0% |
| | Min | -73.8% | -85.2% | -50.6% |
| | Mean | -43.5% | -60.9% | 11.8% |
| | Standard deviation | 23.5% | 25.3% | 73.4% |
| **99.9% quantile** | | | | |
| | Max | -9.1% | -14.8% | 61.5% |
| | Min | -79.1% | -71.6% | -41.0% |
| | Mean | -42.7% | -42.9% | -7.4% |
| | Standard deviation | 24.3% | 25.2% | 37.5% |
| **Stressed default probabilites (minimum)** | | | | |
| **Mean** | | | | |
| | Max | -16.6% | -20.9% | 160.7% |
| | Min | -91.1% | -93.2% | -49.0% |
| | Mean | -52.2% | -63.9% | 18.1% |
| | Standard deviation | 29.2% | 27.3% | 80.3% |
| **99.9% quantile** | | | | |
| | Max | -26.3% | -20.9% | 93.4% |
| | Min | -90.0% | -85.0% | -44.4% |
| | Mean | -49.1% | -50.7% | -1.7% |
| | Standard deviation | 27.9% | 28.0% | 53.1% |

Table 7 quantifies the percentage differences between the highest (upper part of Table 7) and lowest (lower part of Table 7) forecasted stressed default probabilities in the base model 1 and in one of the other model specifications 2 to 8 (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 8:**
**Percentage differences between forecasted stressed default probabilities in the base model 1 and in model modifications 9 to 11**

| | | t+1 | t+2 | t+3 |
|---|---|---|---|---|
| **Stressed default probabilities (maximum)** | | | | |
| **Mean** | | | | |
| | Max | 1.8% | 34.5% | 168.5% |
| | Min | -6.5% | 1.3% | 2.0% |
| | Mean | -2.0% | 21.1% | 102.9% |
| | Standard deviation | 4.1% | 17.5% | 88.7% |
| **99.9% quantile** | | | | |
| | Max | 1.2% | 3.7% | 38.9% |
| | Min | -12.5% | 0.4% | 1.1% |
| | Mean | -6.8% | 2.0% | 24.3% |
| | Standard deviation | 7.1% | 1.6% | 20.3% |
| **Stressed default probabilites (minimum)** | | | | |
| **Mean** | | | | |
| | Max | 18.6% | 86.2% | 285.7% |
| | Min | 14.0% | 3.3% | 5.4% |
| | Mean | 16.2% | 57.2% | 186.3% |
| | Standard deviation | 2.3% | 46.8% | 156.9% |
| **99.9% quantile** | | | | |
| | Max | 9.9% | 13.0% | 74.6% |
| | Min | -1.4% | 1.2% | 3.6% |
| | Mean | 2.9% | 8.8% | 49.7% |
| | Standard deviation | 6.1% | 6.6% | 40.0% |

Table 8 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 9 to 11 (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:**
**Hausman test for the base model 1**

| Variable | IV | F-statistic | Correlation | p-value Hausman test |
|---|---|---|---|---|
| ΔGDP ($t$) | ΔExternal balance on goods and services ($t$) | 38.41 | -0.78 | 0.61 |
| ΔOil price WTI (FOB) ($t$) | ΔPortfolio equity, net inflows ($t$) | 22.68 | 0.70 | 0.52 |
| ΔΔMoody's commodities index ($t$) | ΔΔMoody's commodities index ($t-1$) | 16.85 | -0.65 | 0.60 |

Table 9 shows the results of a Hausman test for endogeneity of the explanatory variables in the base model 1. Based on the employed instrument variables (IV), the null hypothesis of exogeneity cannot be rejected for any of the explanatory variables.

**Figures**

**Figure 1: Overview of the stepwise regression approach**

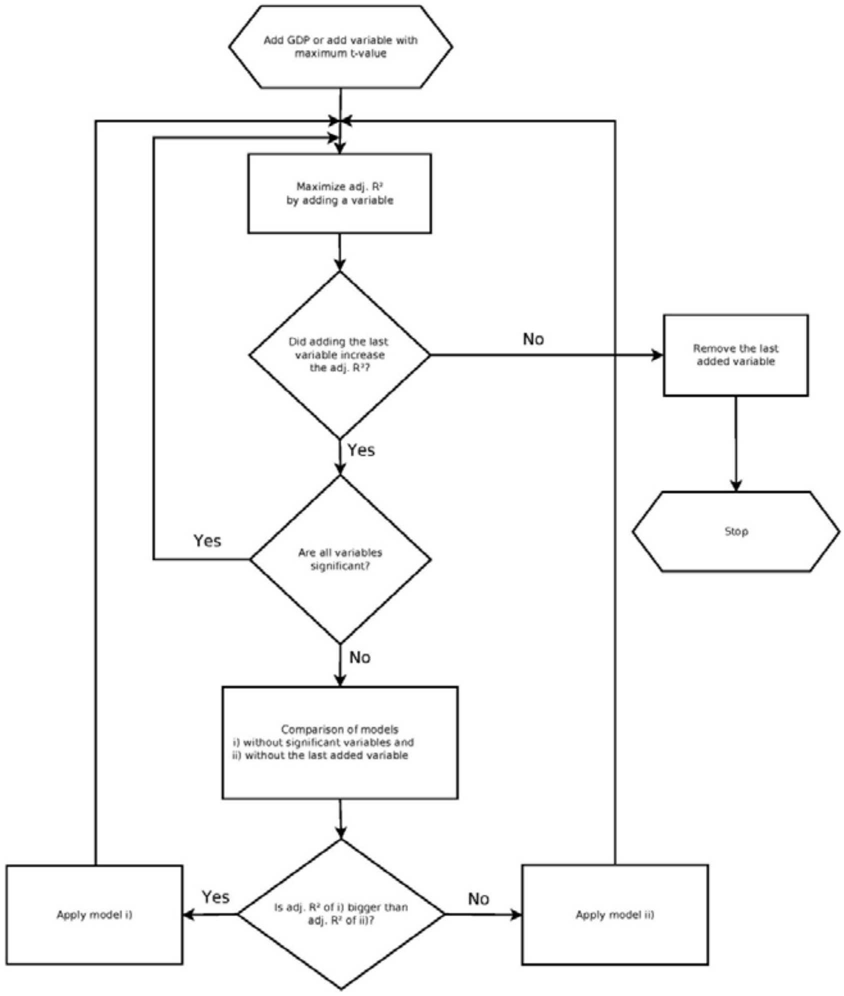



Figure 1 shows the stepwise regression approach for variable selection that is repeated for model specifications 1 to 7. A prerequisite for adding a variable (to avoid (imperfect) multi-collinearity) is that the absolute value of their correlation with any of the other variables that have already been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

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

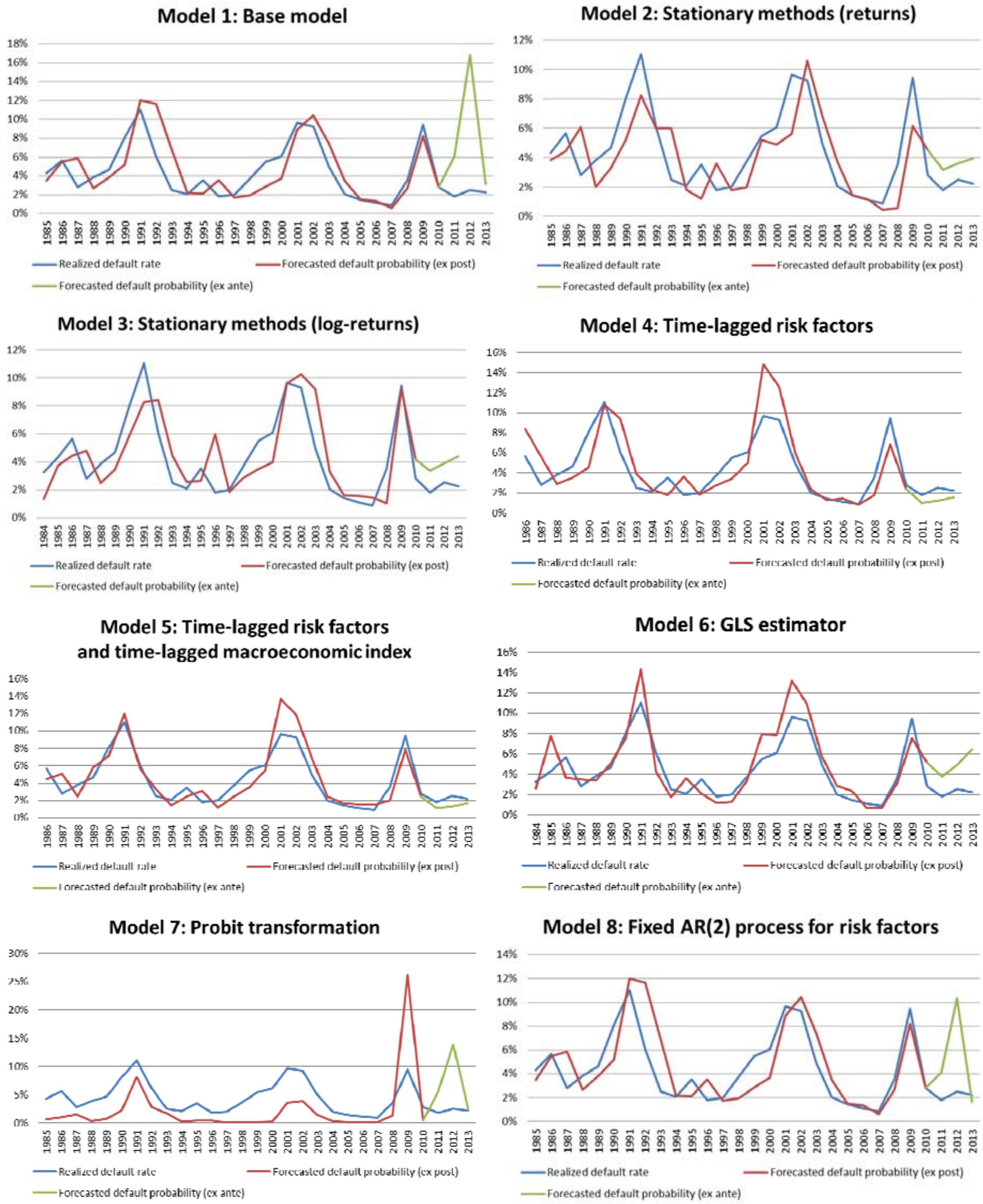



Figure 2 shows the realized default rates compared with the in-sample and out-of-sample predictions of the default probabilities (based on (1) and (2)). For the in-sample prediction, the observed risk factor realizations of each model are inserted into the respective (1), the error term is set equal 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 future risk factor realizations are forecasted by (3), where the error terms are set equal to their means zero.

**Figure 3: Density functions of forecasted default probabilities**

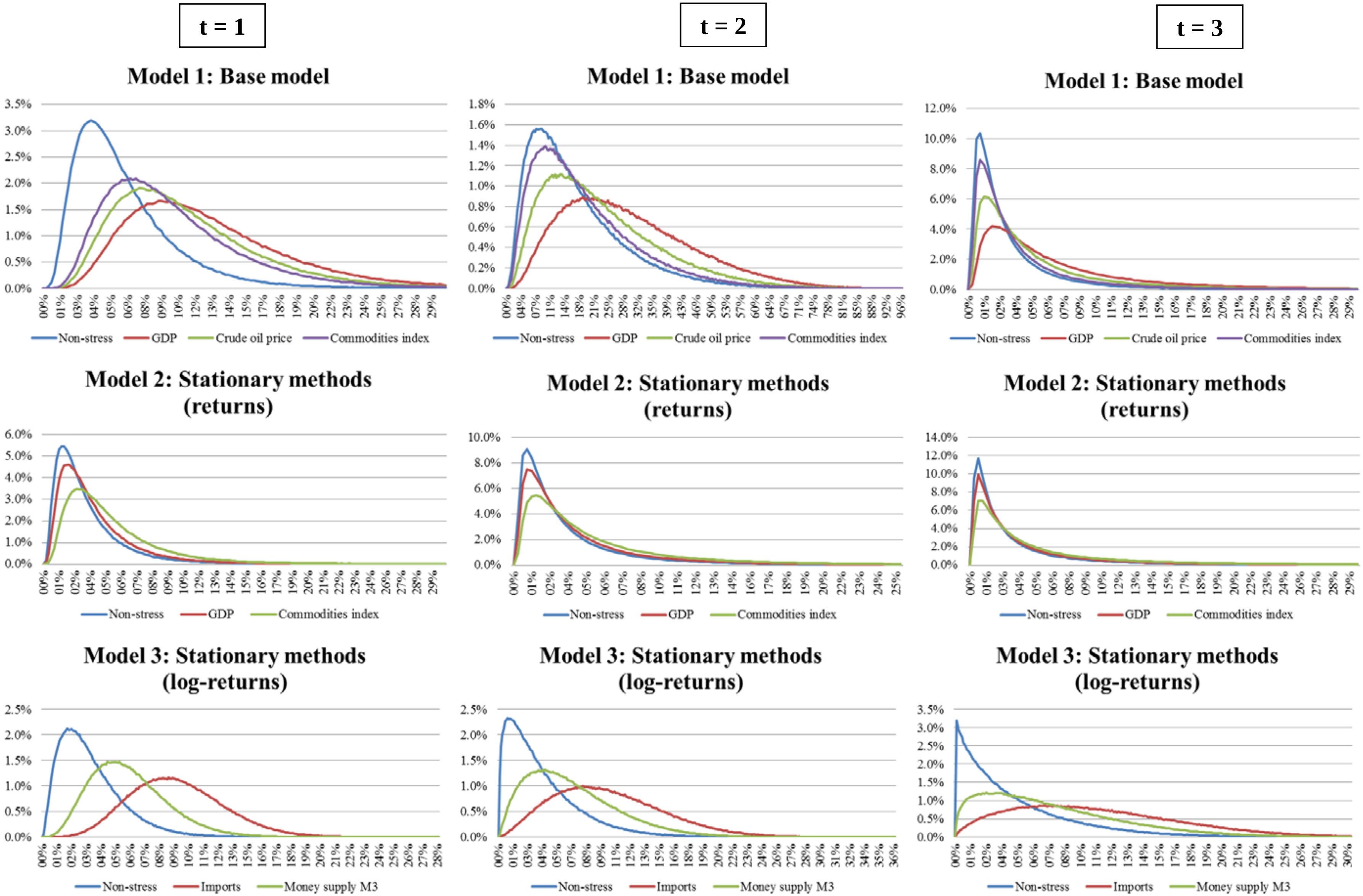


**Figure 3 [continued]**

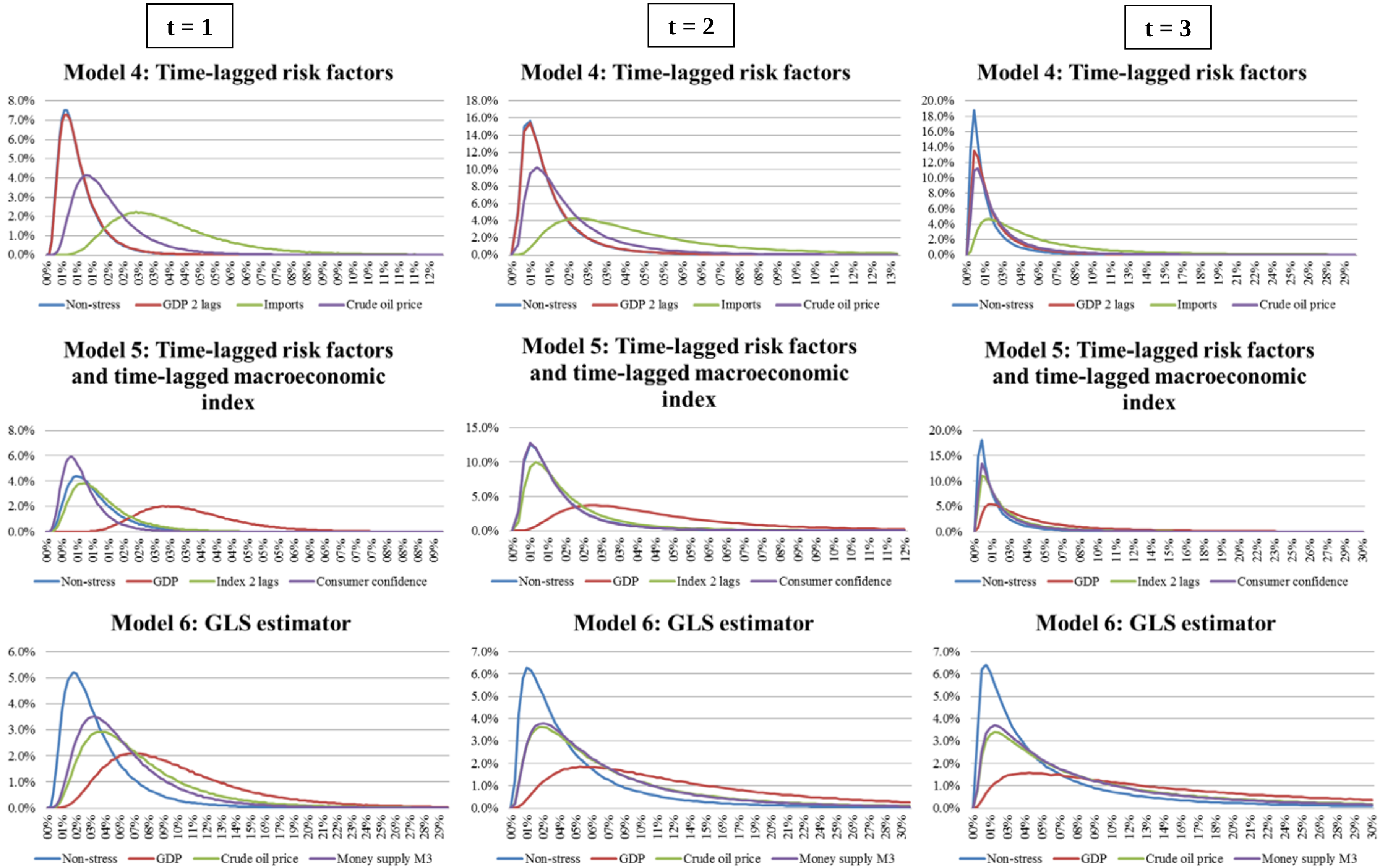


Figure 3 [continued]

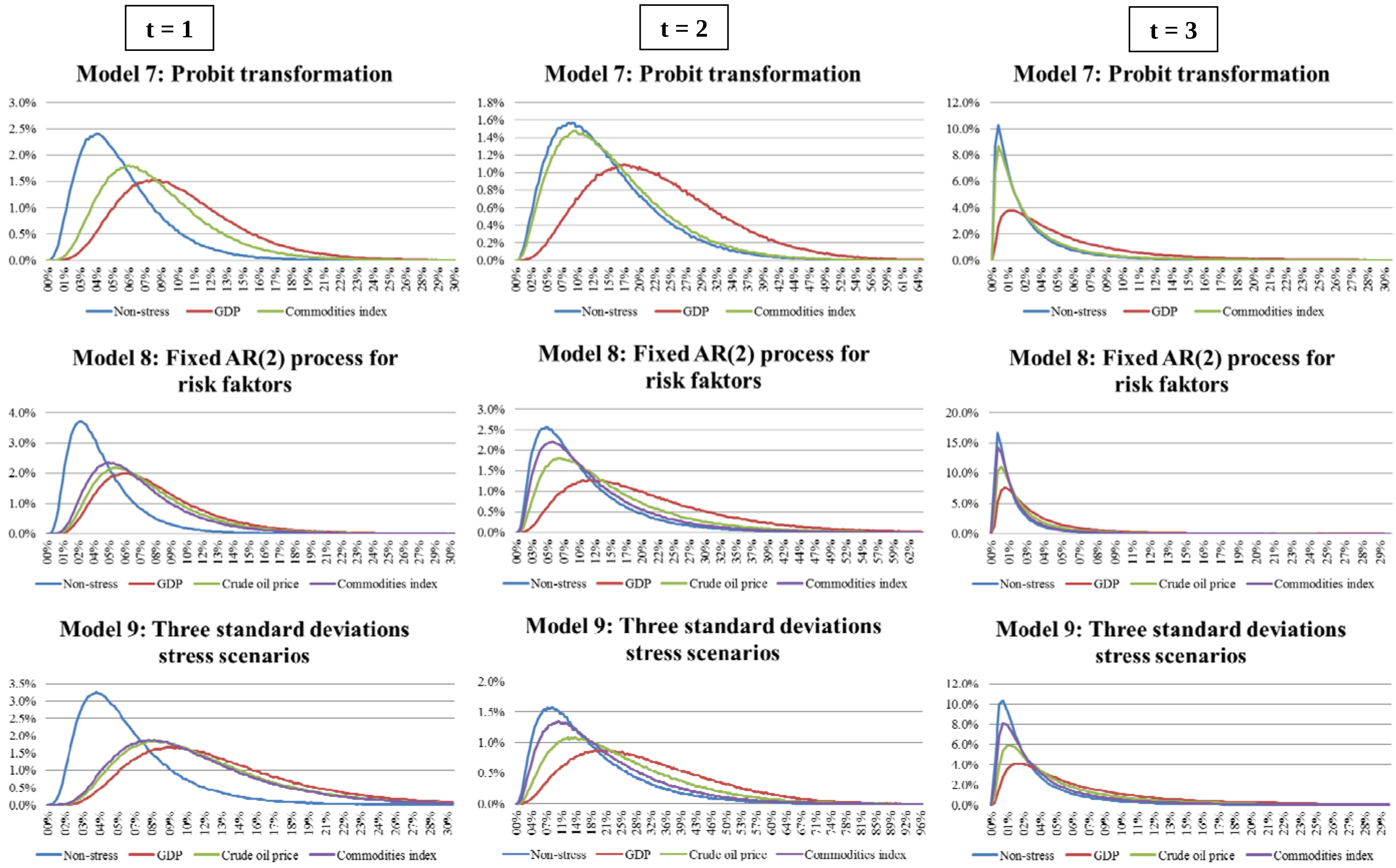


**Figure 3 [continued]**

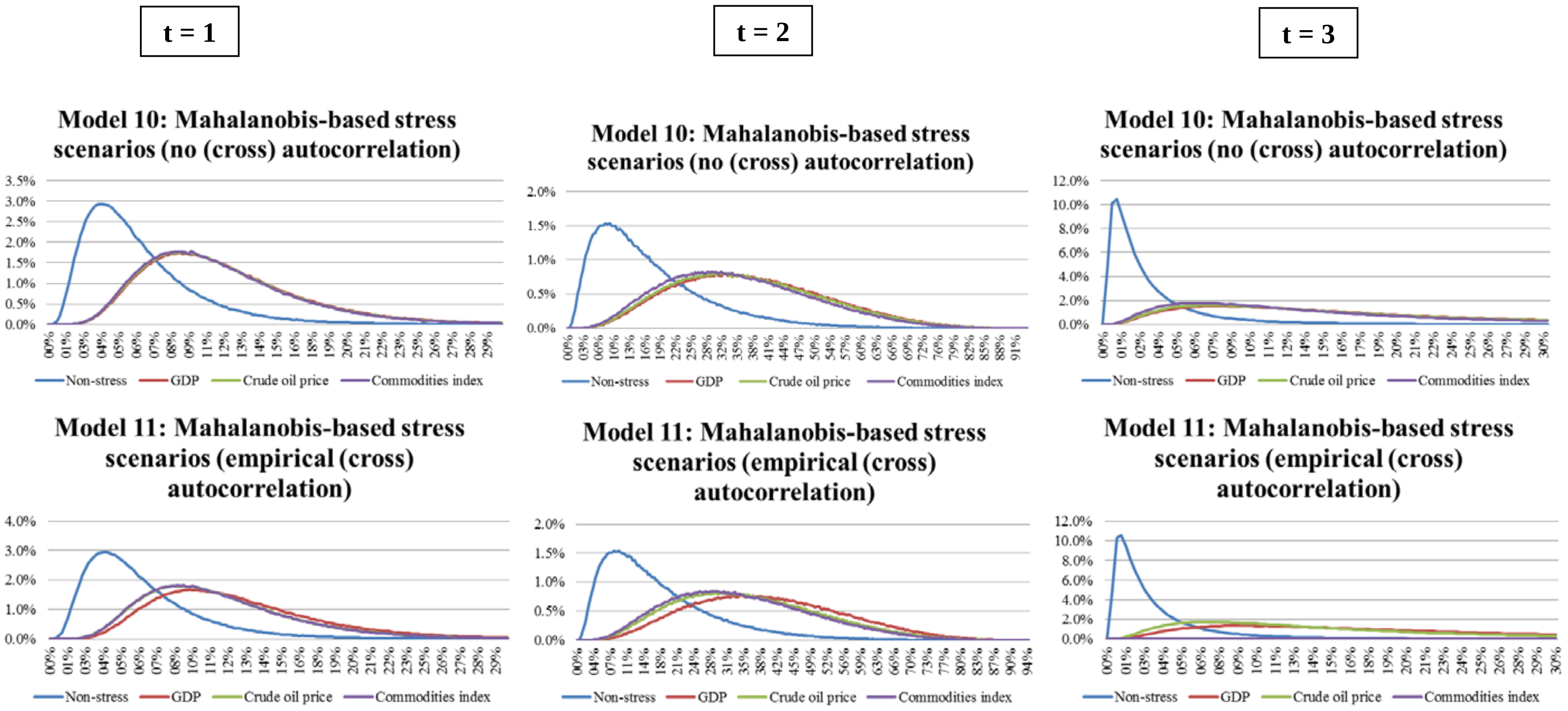


**Figure 4: Tail pp-plots**

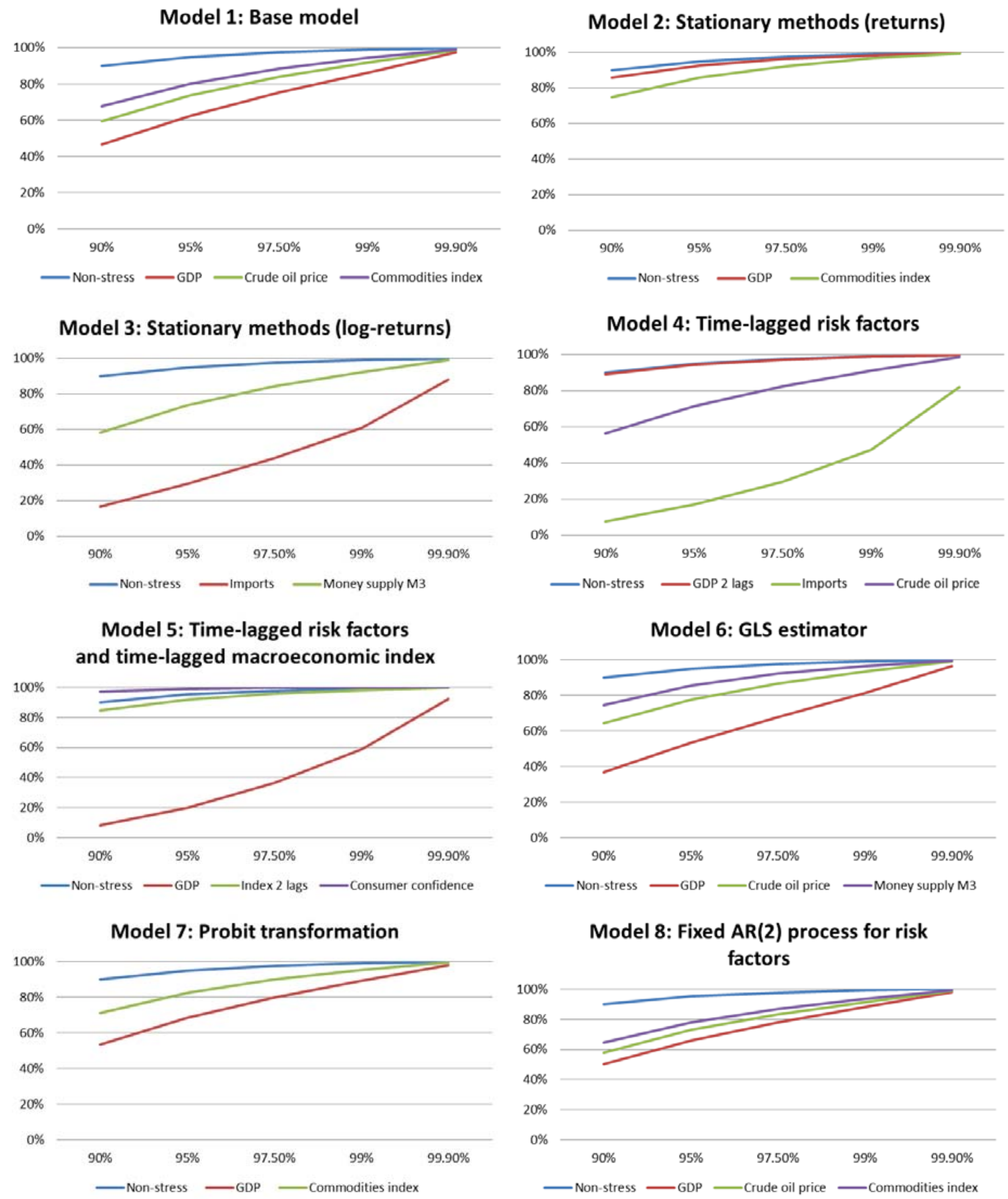
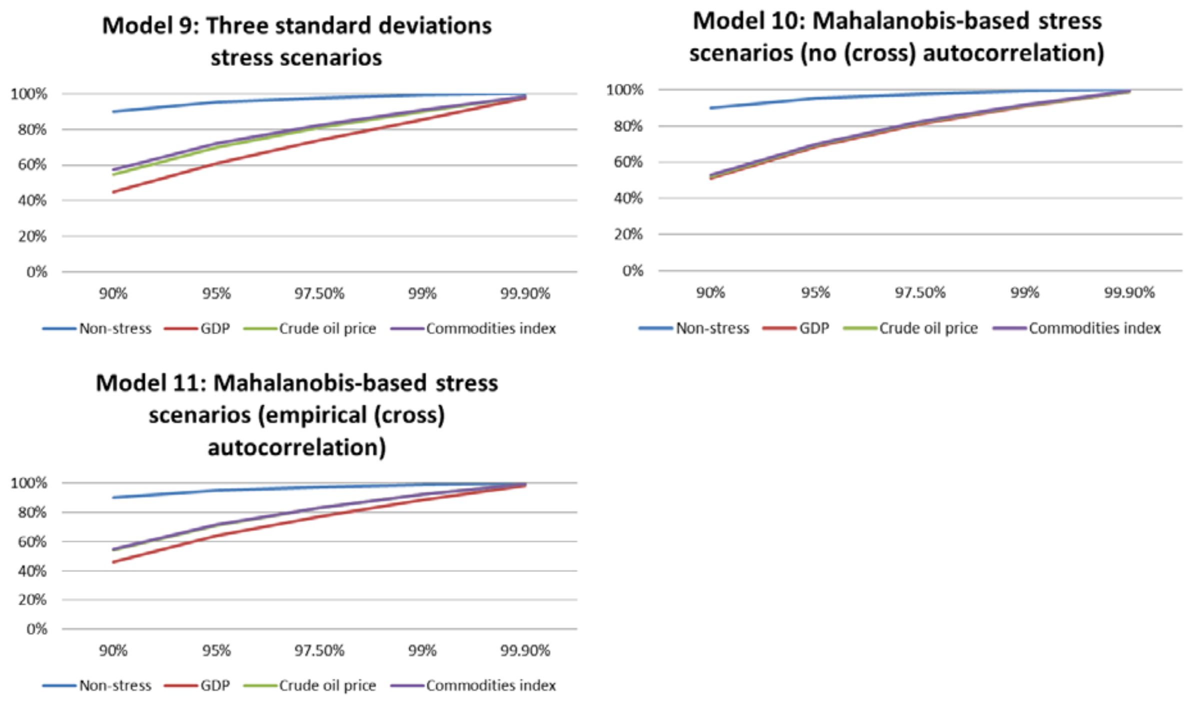


**Figure 4 [continued]**




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**

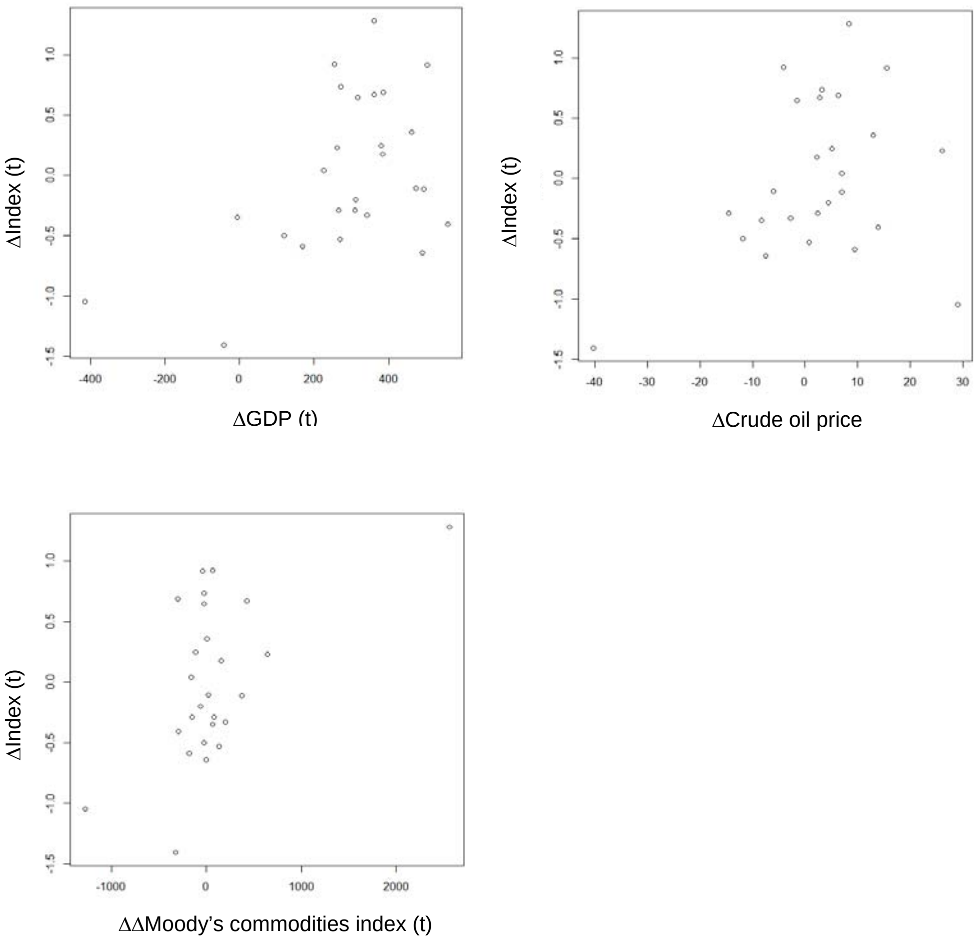



Figure 5 shows that the linear relationship between the logit-transformed realized default rates and the explanatory risk factors that is assumed in CPV-style stress test models is not exactly given for the employed sample.