The Power of Transitions: A Smooth Transition Regression Approach for Improved Credit Risk Modeling in a Post-COVID World

ERASMUS UNIVERSITEIT ROTTERDAM

Erasmus School of Economics

Seminar Financial Case Study: Zanders Group

Abstract

In this research paper we tackle the problem of credit risk modeling after the Covid-19 pandemic. We propose a new approach to credit risk modeling that shifts from a Logistic Regression framework to a Smooth Transition Regression framework. The Smooth Transition Regression framework employs a transition function that captures the credit state of each company, enabling us to model credit risk more accurately. To select the most relevant variables, we use the Weight of Evidence (WoE) approach. We evaluate our model's performance using AUC scoring, which measures the ability of the model to distinguish between good and bad credit risks. Our results indicate that our proposed model outperforms the Logistic Regression model, providing evidence for its increased accuracy in predicting credit risk.

Supervisor: Prof. Dr. Chen Zhou

Group B

Gabor de Bergeyck - 656524

Orestis Chatzinas - 476238

Ariel Levi - 653596

Alfonso Llaca Kuri - 659282

October 8, 2024

Table Of Contents

1 Introduction 2

2 Literature review 3

3 Data 4

4 Methodology 6

4.1 Smooth Transition Regression 6

4.2 Variable Selection 7

4.3 Parameter estimation 10

4.4 PD Binning 10

5 Results 11

5.1 Model Specification 11

5.2 Model Performance Comparison 12

5.3 Coefficient Interpretation 14

5.4 Interpretation of the STR transition 16

5.5 Probabilities of Default Calibration 18

6 Conclusion 19

Appendix 22

A1 Confidence Intervals for the STR model 22

A2 WoE IV Transformation 24

A3 PD Binning with different bins 26

1 Introduction

In the past few years, Covid-19 has been ruling everyone's life, changing what is regarded as normal. In the financial world, credit rating models have also been heavily affected, mainly in the form of sub-optimal bankruptcy predictions caused by various government measures such as tax breaks and moratoria (Telg et al. (2023)). This causes a problem, as traditional credit models, such as Logistic Regression (LR), cannot quickly respond to these exogenous shocks, resulting in inaccurate Probability of Default (PD) predictions. In this paper, we attempt to tackle this credit risk modeling challenge caused by Covid-19.

We first analyze a basic LR model as our baseline model. Next, we will propose a Smooth Transtition Regression (STR) model optimized for imbalanced data in the fashion of González et al. (2005). By allowing for different parameter values depending on the state, the smooth states will improve PD forecasting, with the state being determined by a smoothing function. Ullah et al. (2021) and González et al. (2005) show that for times of crisis, regime-switching models have better performance than "static" models. Due to this phenomenon, we expect the STR to incorporate the Covid-19 data much more efficiently into the model, which we additionally expect to lead to significantly better performance.

The aforementioned STR model needs to have a predetermined regime transition function. The most common solution is to use a logistic transition function with a prespecified transition variable. The most important conditions for a valid transition variable are that the transition variable captures enough variation over the years and in the cross-section, and that it closely captures the impact of the Covid-19 pandemic. We compare this method to the LR model based on its ability to transition to the proper state and estimate post-Covid PDs.

Another point of notice is the interpretability of the respective model, which is an essential requirement for credit risk models set by regulators to avoid "black-box" machine learning techniques. The research question of this paper is then: *Can a regime-transition model improve the forecasting of PDs in the context of the Covid-19 pandemic and in the post-crisis period, while maintaining a high interpretability of the model?*

To answer this question, we select the explanatory variables based on the Weight of Evidence (WoE) approach and consequentially estimate the aforementioned model. For the remainder of this paper, we first describe the dataset. After that, the literature review summarizes the relevant literature regarding credit risk and probability modeling for our model framework. Subsequently, the used model is extensively described in the methodology, and the results are stated in the results section. Finally, we wrap up the paper with a conclusion.

2 Literature review

Logistic regression models have been the golden standard since the early days of credit risk modeling. Martin (1977) shows that with a relatively simple maximum likelihood maximization, he is able to get intuitive LR estimation results regarding bank failures. Others give LR their twist, such as Hauser and Booth (2011) using Bianco and Yohai estimation instead of maximum likelihood to better capture outliers for the LR and Jabeur (2017) adjusting the model by solving multicollinearity through Partial Least Squares LR. The popularity of the LR model to this day is therefore not a surprise, especially since Devi and Radhika (2018) show that LR can outperform machine learning techniques in some specific situations, although they also point out that LR can lead to convergence issues arriving from computational problems at the likelihood maximization.

Due to a large number of potential variables, knowing which ones to use for analysis is a significant issue in credit risk modeling. Yang et al. (2015) elaborate on ordering the explanatory variables using an approach called Weight-of-Evidence (WoE). Nehrebecka (2016) uses the same approach. This approach categorizes each variable and measures its information value (IV), allowing it to follow a top-down approach for variable selection based on predictive power. WoE thus provides a different way to approach LR and can be compared to unmodified LR (Lin et al., 2012).

A well-known drawback of LR is its inability to capture different regimes in the data. An early model to capture these regimes through transition is the Smooth Transition AutoRegressive (STAR) model by Terasvirta and Anderson (1992). This model has credit risk applications such as Huang and Hu (2012), which uses the STAR model to identify different credit default swaps (CDS) regimes. They find clear evidence for smooth transitions between low-price and high-price CDS regimes, which correctly identify the Global Financial Crisis of 2008.

An extension of the STAR model that incorporates cross-sectional information of the data is the Panel Smooth Transition Regression (PSTR) model as proposed by González et al. (2005). We use a version of this model to identify two regimes across companies and over time. Our model differs from González et al. (2005) because we use a logistic specification to produce PDs rather than for modeling a variable itself.

The choice of metric for our problem requires some consideration. For binary classification, most metrics consider the confusion matrix as a centerpiece, with derived metrics such as accuracy and sensitivity. These metrics perform well with a predefined threshold and offer intuitive results with no complexity (Hossin and Sulaiman (2015)). However, in our application, a predefined threshold is not attainable as we predict PDs, and therefore do not perform binary classification. Thus, the Receiver Operating Curve (ROC) is a good solution, where in short, the ROC graphically depicts the trade-off between the true positive rate and the false positive rate for every single threshold (Fawcett (2006)). The measure known for assessing the quality of the ROC is called Area Under Curve (AUC), which is simply the entire area captured under the ROC. Ling et al. (2003) formally prove that AUC is a better measure than accuracy for evaluating learning algorithms.

3 Data

To support our model with empirical evidence, we use the dataset containing annual observations provided by Zanders Group. The dataset consists of $2,143,012$ observations and 218 financial and non-financial explanatory variables of $575,266$ companies between 2015 and 2021.

The model's dependent variable is a binary indicator denoting the default status of a company. Table 1 displays the default ratio for each year in the sample. The data contains 15,109 defaults in total.

Table 1: Number of observations and default ratio per year.

	Number of observation	Default ratio
2015	290,117	0.91%
2016	314,459	1.01%
2017	327,010	0.78%
2018	337,617	0.67%
2019	358,855	0.69%
2020	327,614	0.46%
2021	184,640	0.26%

Dotted line displays that we will predict PD's for 2021

To reduce the number of explanatory variables that will be used in our models, we will use the Weight of Evidence (WoE) and Information Value (IV) method as proposed by Yang et al. (2015). We explain this procedure in detail in Section 4.2. We omit redundant and multicollinear variables to prepare the data for this analysis. Hence, we apply the WoE to the resulting 57 explanatory variables.

The dataset split used to find the optimal number of variables for forecasting and testing its performance is selected as follows. First, for the variable selection step, we randomly select a validation sample from 2015-2020 that will mirror the imbalanced data for 2021. Then, we create a balanced training subsample with the remaining defaults from the same years. The process is illustrated in Figure 1.

For the forecasting step, we also create a balanced estimating subsample with all defaults in 2015-2020. To create this balanced subsample, we collect all defaults in 2015-2020 and randomly draw the same number of non-defaults. To analyze the robustness of the outcome, we consider multiple random draws. The results using the WoE approach on different random draws are stable and are stated in more detail in section 4.2. A similar procedure is used to create the balanced validation set. Finally, we test on the imbalanced forecast sample with all the observations of 2021. The process is illustrated in Figure 2.

Figure 1: Data selection process, variable selection.

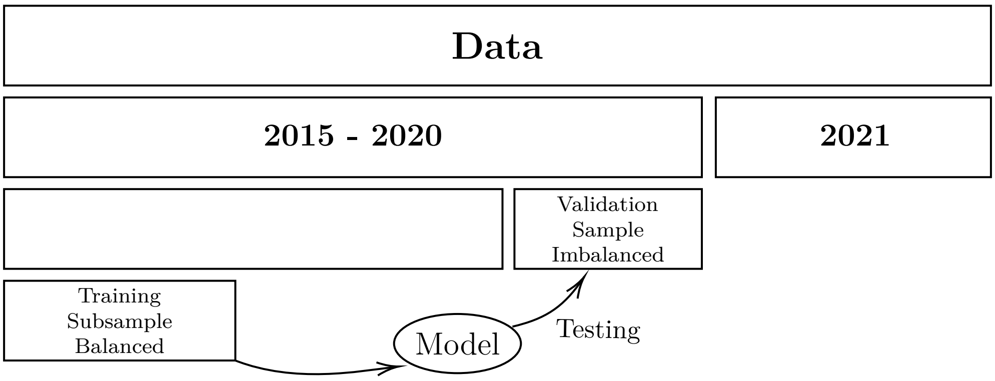
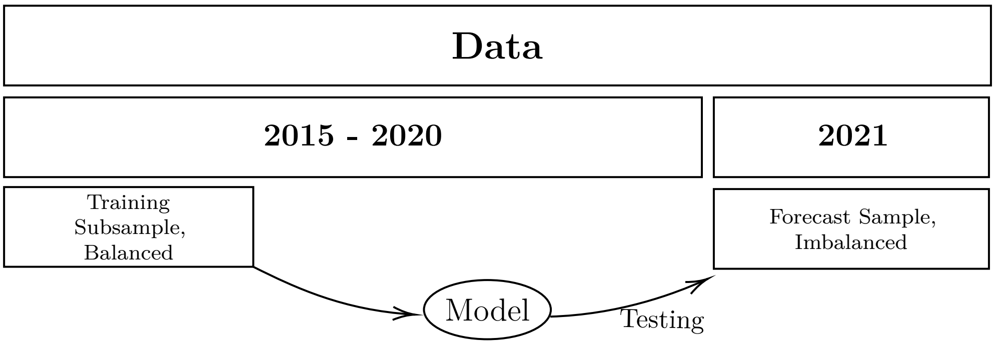


Figure 2: Data selection process, forecast.



There are two main reasons for undersampling to obtain balanced training subsamples. The most important reason is that the WoE approach is not suited for heavily imbalanced datasets. For the specific explanation, we again refer to section 4.2. The second reason is computational efficiency. Our two models are estimated using maximum likelihood. As there is no closed form solution for the maximization problem (see section 4.3), we maximize numerically. Undersampling increases both the speed and likelihood of convergence. Since after undersampling the training subsamples still contain around 30.000 observations, we consider that the use of the WoE and the efficiency increase favor the choice of undersampling.

4 Methodology

In this paper, we predict companies' PD in the next 12 months. We first use a basic logistic model as a benchmark. Formally, we have a binary output variable $Y_{i,t}$, which indicates whether or not a company i defaults in year t. We model the conditional probability $p(x_{i,t}) := \mathbb{P}(Y_{i,t} = 1|X_{1,i,t} = x_{1,i,t}, X_{2,i,t} = x_{2,i,t}, ..., X_{p,i,t} = x_{p,i,t})$ as a function of $x_{i,t} = (x_{1,i,t}, x_{2,i,t}, ..., x_{p,i,t})$, with p characteristics of company i at time t. The logistic regression takes the form:

$$p(x_{i,t}) = \frac{1}{1 + e^{-f(x_{i,t})}}, \tag{1}$$

with $f(x_{i,t}) = a + \beta_1 x_{1,i,t} + \beta_2 x_{2,i,t} + ... + \beta_p x_{p,i,t} = \theta u'_{i,t}$, where $\theta = (a, \beta_1, \beta_2, ..., \beta_p)$ are the coefficients and $u_{i,t} = (1, x_{1,i,t}, x_{2,i,t}, ..., x_{p,i,t})$ collects all explanatory variables.

Nonetheless, as motivated in the introduction, we expect the LR model to perform poorly during the Covid-19 period due to its static structure which does not take transitional periods into account. For this reason, we describe the Smooth Transition Regression model in Section 4.1, which allows for smooth regime-switching and offers greater flexibility.

Additionally, due to the high number of explanatory variables, which can affect both computational speed and interpretability, we perform variable selection to filter out uninformative independent variables from the dataset. We use the Weight of Evidence (WoE) approach as in Yang et al. (2015) and Nehrebecka (2016), which is explained in section 4.2. This approach categorizes each variable and measures its predictive power, allowing us to follow a top-down approach for variable selection based on predictive power.

4.1 Smooth Transition Regression

The main advantage of the STR model compared to the logistic regression model, is that it allows the parameters to vary across a limited number of different regimes. The model allows for a smooth transition between the regimes as opposed to hard-regime-switching models such as Markov Switching models.

In this model, an observable variable $q_{i,t}$ determines the state transition. In our study, since we consider a STR model with two regimes $j = 1, 2$, the model is defined similar to the LR in eq. (1), but the function $f(x_{i,t})$ now takes the following form:

$$f(x_{i,t}) = \theta_1 u'_{i,t}(1 - g(q_{i,t}, \gamma, c)) + \theta_2 u'_{i,t} g(q_{i,t}, \gamma, c), \tag{2}$$

with coefficients $\theta_j = (a_j, \beta_{1,j}, \beta_{2,j}, ..., \beta_{p,j})$ for states $j = 1, 2$. The transition function $g(q_{i,t}, \gamma, c)$

in eq. (2) is a continuous function of the observable variable $q_{i,t}$ and is bounded between zero and one. In other words, this function of $q_{i,t}$ determines the weights for the coefficients in each state $j = 1, 2$. We follow González et al. (2005) and opt for the logistic specification

$$g(q_{i,t}, \gamma, c) = \frac{1}{1 + exp\{-\gamma(q_{i,t} - c)\}}, \gamma > 0, \tag{3}$$

where γ determines the smoothness of the transitions and c stands for the threshold between regimes.

As for the observable transition variable $q_{i,t}$, we consider four alternatives. The first two are prominent financial ratios, namely the debt ratio and the current ratio. The former represents a company's level of leverage, while the latter measures it's ability to pay obligations due within one year. Hence, both ratios may contain predictive power about the credit state of each company, i.e. if they are more likely to default within next year or not.

The third alternative we explore is the lagged country specific default rate, i.e. the percentage of companies in the country of company i that defaulted at $t - 1$. The rationale behind this transition variable is that it went down for almost all countries in 2020, contrary to the PDs that many models predicted for that period. Therefore, this transition variable should allow the STR model to identify that although the predicted PDs were high, the actual default rate was low and hence there is no necessity of predicting high PDs for 2021 again.

The last transition variable that we consider is a macroeconomic indicator, specifically the country-specific average yearly unemployment rate. This choice is based on the observation that during the Covid-19 pandemic, the unemployment rate in most countries was higher than usual. Therefore, it seems reasonable to incorporate this variable in the analysis with the aim of distinguishing between Covid and non-Covid periods. Furthermore, this indicator is published on a monthly basis, which is much more timely than most other macroeconomic indicators.

An advantage of the first three variables is that they can be calculated using the variables in the provided data set, while the unemployment rate has to be extracted from external sources. An additional advantage of both financial ratios is that they are company-specific, providing more granularity compared to the other two country-specific variables.

4.2 Variable Selection

The following section provides the specifics of the WoE & IV approach as described by Nehrebecka (2016) and Yang et al. (2015). The main idea is that each explanatory variable is ranked based on its predictive power in relation to the dependent variable, the default indicator. The steps to achieve that for each explanatory variable are as follows.

We begin by splitting the data into equally sized bins according to one specific explanatory variable. Specifically, we explore different sizes between 2 and 20. The choice of maximum 20 bins follows from Zeng (2014) as they state that no bin can contain less than 5% of the total number of observations. For each bin we calculate the percentage of non-defaults and the percentage of defaults and the WoE of each bin for each variable is defined as:

$$WoE_{ij} = ln\left(\frac{\%\text{ of non-defaults}_{ij}}{\%\text{ of defaults}_{ij}}\right). \tag{4}$$

where i indicates the variable and j indicates the specific bin of that variable.

The information value is defined as

$$IV_i = \sum_j (\%\text{ of non-defaults}_{ij} - \%\text{ of defaults}_{ij}) \times WoE_{ij}. \tag{5}$$

Consequently, the IV is calculated for the given explanatory variable using 2 to 20 bins. The representative IV value for this variable is then determined as the highest value across all the different bins. This procedure is repeated for every explanatory variable. A ranking of all explanatory variables is computed based on the maximum IV score. The first input is the variable that is deemed to be the most informative in relation to the dependent variable, as determined by the Weight of Evidence and Information Value method.

It is important to remark on how we dealt with the missing values (NaN), following the approach in Zeng (2014) we bin all NaN values separately. The maximum of 20 bins will not be surpassed, in which all NaNs are binned together and its corresponding WoE and IV is calculated in the same manner as previously described. The reasoning behind is that the NaN values could contain some informative power in contrast to most applications, primarily due to their large presence in some specific independent variables.

Finally, each numerical value of the variables is transformed by replacing it with the bin it belongs to. This bin is then mapped to the corresponding Weight of Evidence (WoE) of that particular variable for that specific bin. Table 2 provides a brief illustration of this transformation process. For a more extensive illustration see Tables 6 and 7 in the appendix.

It is also worth mentioning that, as described in Section 3, the size of the observations was significantly reduced by creating balanced subsamples, where the non-default observations were selected randomly. To evaluate the robustness of this undersampling approach, the model was run multiple times using multiple different randomly selected subsets of non-default observations. The resulting Information Value (IV) ranking was consistent across all runs, with only minor

Table 2: WoE Transformation Example

x_1	x_2	x_3	x_4
31677442	19984767	3395201	21765445
917923	867923	129975	1682531
1113502	1077442	NaN	NaN
7936000	7916000	1724000	NaN
11945652	9200820	1114245	9028106

x_1^{WoE}	x_2^{WoE}	x_3^{WoE}	x_4^{WoE}
1.69	1.79	1.80	1.19
0.57	0.58	1.02	0.77
0.57	0.81	1.13	-0.54
1.32	1.64	1.76	-0.54
1.44	1.64	1.69	0.98

Left table represents a random subset of original observations and Right table their respective transformation to the WoE of the bin it belongs to.

variations to a couple of decimal places. This suggests that the use of the Weight of Evidence and Information Value method is robust in the present case.

As for the reason on why WoE was performed on a balanced subsample, it is worth looking at eq. (4) and eq. (5). For the equations to be defined for each bin, each bin has to contain at least some defaults. When using a heavily-imbalanced dataset and considering equally sized bins, some bins may contain no defaults at all, making this approach unfeasible. Although it may be possible to apply WoE on an imbalanced dataset by considering bins of different sizes, it is out of the scope of this study.

After deriving the ranking of all the variables, we can choose the number of variables used based on the model specification approach in the fashion of Nikolic et al. (2013). For this variable selection method the main idea is to successively add the IV-ranked variables to the model, and calculate their respective AUC attained from validation set. When the values start to plateau and the model starts to overfit, the optimal number of variables is attained.

Lastly, we explain the trade-off behind using the WoE approach. Some advantages are straightforward: The method allows to reduce the number of explanatory variables drastically and is suited both for quantitative and qualitative variables. As previously explained, it also allows to use NaN values by assigning them to a separate bin, as they may contain predictive power. One last advantage, is that from eq. (4) we see that the explanatory variables are transformed into a log-scale. This way, we have a linear relationship between $ln(\frac{p(x_{i,t})}{1-p(x_{i,t})})$ and the explanatory variables, an assumption of the LR model.

As for the disadvantages, we lose variation in the data by reducing the number of different values that a variable can take to a maximum of 20. Consequently some multicollinearity is introduced among the explanatory variables by bringing them all to a log-scale, which can be seen again in eq. (4).

4.3 Parameter estimation

For the parameter estimation we use maximum likelihood by assuming that the binary variable $Y_{i,t}$ is Bernoulli distributed conditional on $(X_{1,i,t} = x_{1,i,t}, X_{2,i,t} = x_{2,i,t}, ..., X_{p,i,t} = x_{p,i,t})$ with probability $p(x_{i,t})$ as introduce above. The likelihood is then

$$L(\theta) = \prod_{i=1}^{N}\prod_{t=1}^{T} p(x_{i,t})^{y_{i,t}}(1 - p(x_{i,t}))^{1-y_{i,t}}.$$

As shown in Hastie et al. (2009), the log-likelihood can be written as

$$l(\theta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\{-log(1 + e^{f(x_{i,t})}) + y_{i,t} \cdot f(x_{i,t})\}, \tag{6}$$

with $f(x_{i,t}) = a + \beta_1 x_{1,i,t} + \beta_2 x_{2,i,t} + ... + \beta_p x_{p,i,t}$ for the benchmark LR model and $f(x_{i,t})$ as in eq. (2) for the STR model. There is no closed form solution to maximize eq. (6). Hence, we solve the maximization problem numerically using the L-BFGS-B algorithm, an optimization algorithm that approximates the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS). For the calculation of standard errors see Section A1 in the Appendix.

4.4 PD Binning

Both the LR and the STR models do not produce accurate PD predictions in their original forms presented in the previous subsections. For instance, it is common for both models to predict PDs exceeding 80%, which is not realistic in practical applications. Therefore, we must calibrate the predicted probabilities before using them in practice. We call these the "calibrated probabilities of default" or calibrated PDs. By calibrating the predicted probabilities, we can ensure that they are consistent with the observed default frequencies in a given population, thus making them more suitable for practical applications.

To do so, we propose a PD-binning method suited for both the LR and the STR models. The first step is to train the models on the (balanced) estimation subsample of 2015-2020 (see Figure 2). We then compute in-sample PDs for the whole imbalanced 2015-2020 dataset and out-of-sample PDs for the forecast sample with these models. We denote the fitted in-sample and out-of-sample PDs by $\hat{p}(x_{i,2020})$ and $\hat{p}(x_{i,2021})$, respectively. Then, we create n equally-sized intervals (the bins) and put all the $\hat{p}(x_{i,2020})$ into their respective interval (bin). From each bin, we compute the actual default rate, i.e. the percentage of companies that defaulted in each bin.

This procedure gives a mapping from the predicted $\hat{p}(x_{i,2020})$ by our model and the actual default rates. Hence, we can finally apply the same mapping to all $\hat{p}(x_{i,2021})$. This way, we assign each $\hat{p}(x_{i,2021})$ to the previously computed bins and map them to the actual default rate observed for the $\hat{p}(x_{i,2020})$. This allows to calibrate the predictions for 2021 by ensuring they are consistent with the observed default frequencies in the 2015-2020 population.

5 Results

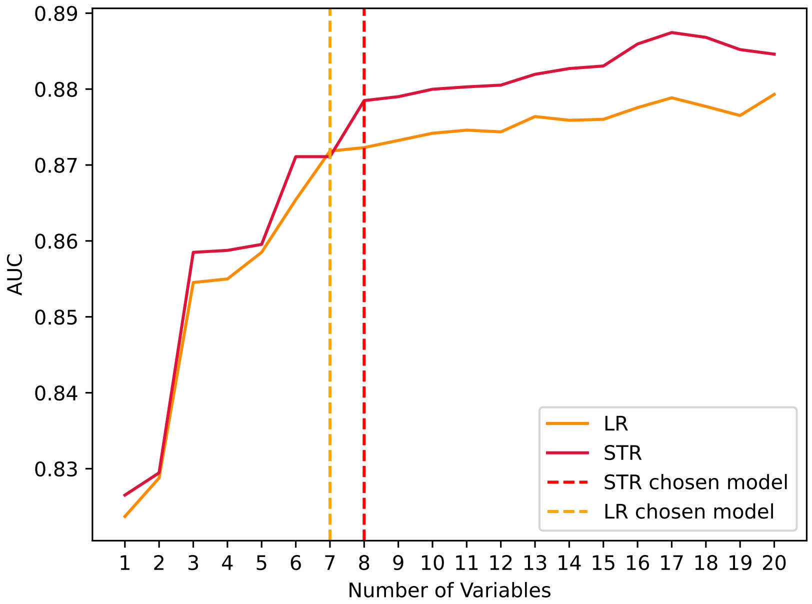
5.1 Model Specification

As mentioned in Section 4.1, we consider different transition variables $q_{i,t}$ for the STR model. The best performing transition variable was found to be the debt ratio, which gave consistently higher AUC's both in-sample and out-of-sample. Hence, in the following we only present the results using this transition variable.

The initial step in both LR and STR models is to determine the number of explanatory variables to be used in the estimation of the final models. As previously noted, the optimal number of variables for each model is compared based on the AUC values attained from training the model on the training subsample and testing on the validation sample with an increasing number of variables. The optimal number of variables is selected when the AUC reaches its peak and begins to level off, indicating that the addition of further variables does not result in an improvement in performance. The procedure for creating the subsample and the validation sample is detailed in section 3 and shown in Figure 1.

The results of this procedure for the LR and the STR are illustrated in Figure 3. For the LR we can see that the chosen optimal number of variables is 7, while for the STR it is 8. However, it is noteworthy that the increase in the AUC is marginal as the number of variables included in the model increases. We contend that while the addition of more variables may lead to a slight improvement in performance, it also significantly undermines model interpretability and computational efficiency. Thus, based on this trade-off between model interpretability, computational power, and marginal gains in performance we choose 7 and 8 variables respectively.

Figure 3: Comparison of optimal number of variables between LR and STR based on AUC



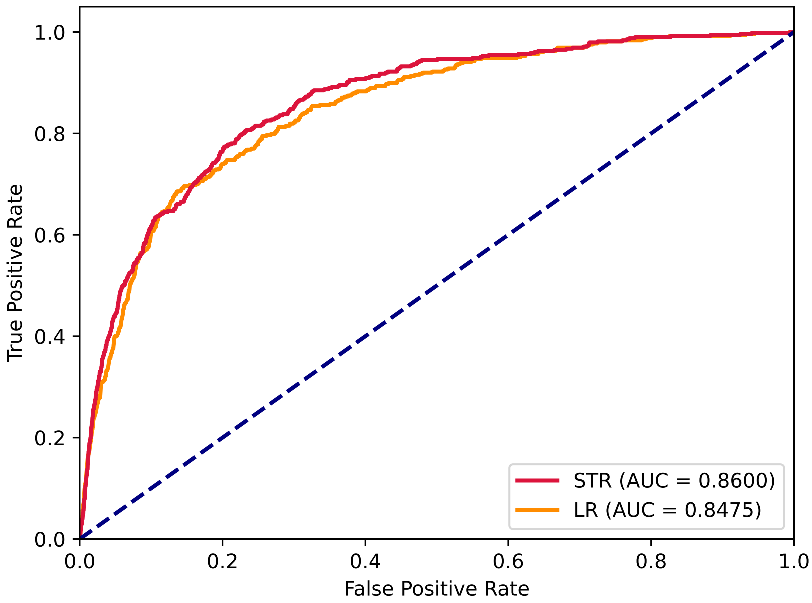
For completeness, for the rest of the transition variables we present the AUC's using only 8 variables, which were computed based on the same procedure as described above: Current ratio with an $AUC = 0.845$, lagged country specific default rate with an $AUC = 0.85$ and lastly unemployment with an $AUC = 0.72$. In Figure 3 it is shown that the debt ratio attains $AUC = 0.88$. This transition variable strictly beats all of the other transition variables, as indicated previously.

5.2 Model Performance Comparison

We train the models specified in the previous subsection on the (balanced) estimation subsample of 2015-2020, and test them on the (imbalanced) forecast sample of 2021 (see Figure 2). As in the last subsection, we report the results using the debt ratio as transition variable $q_{i,t}$ for the STR model.

Figure 4 displays the ROC curves of both models of the selected models from Section 5.1. The performance for forecasting PD's in 2021 is relatively similar with a slight edge in favor of STR.

Figure 4: ROC of the LR and STR



ROC of the logistic regression (Orange) and STR (Red) test set, with 7 and 8 variables respectively. Training on the 2015-2020 data set, and testing on the 2021 data set

For completeness we present Table 3 which provides AUC values for the two models, LR and STR, predicting 2021 out-of-sample. AUC values are shown for different numbers of variables, ranging from 1 to 20. The table indicates that the STR consistently outperforms the LR model for all numbers of variables considered, providing further support for our model.

Table 3: AUC of LR and STR

# variables	LR	STR
1.	0.7941	0.7969
2.	0.7963	0.8002
3.	0.8207	0.8219
4.	0.8201	0.8215
5.	0.8221	0.8242
6.	0.8443	0.8571
7.	**0.8475**	0.8571
8.	0.8473	**0.8600**
9.	0.8495	0.8619
10.	0.8544	0.8637
11.	0.8547	0.8643
12.	0.8548	0.8651
13.	0.8561	0.8659
14.	0.8534	0.8653
15.	0.8577	0.8695
16.	0.8578	0.8702
17.	0.8599	0.8721
18.	0.8611	0.8712
19.	0.8614	0.8720
20.	0.8629	0.8716

AUC values of logistic regression and STR with increasing variables over the 2021 test data-set.

A Wilcoxon signed-rank test is performed to test whether there is a statistically significant difference between the AUCs of the STR and the LR (Wilcoxon (1945)). This test is primarily used to test whether two samples are derived from the same distribution, which is the null hypothesis. The alternative hypothesis states that the samples are derived from different distributions. Conducting the test leads to a p-value of 0.0063. Considering a significance level of 0.05 we reject the null hypothesis. This is further evidence in favor of the STR model, as we can assume from this test that the improvement of the STR model is statistically significant.

5.3 Coefficient Interpretation

Table 4 below shows the coefficients of the two competing models from last subsection, as well as their significance levels. Since we are regressing on the WoE of each variable and not on their original values (see Table 2), the coefficients are all in a similar range.

It is important to prompt and stress that the aim of this paper is to identify and distinguish between the two different regimes. Therefore, the interpretation of the coefficients in the STR model should be focused on their ability to identify the different regimes rather than on their numerical interpretation. In other words, the emphasis is on understanding which variables perform better in distinguishing between the states, rather than on interpreting their coefficients in a numerical sense. This allows for a more robust analysis of the factors that differentiate the two regimes and provides more insight into the key drivers of default risk.

The coefficients of the STR model can be interpreted as follows: On the one hand, the variables with coefficients that are similar for both states (e.g. Other Shareholder Funds and P&L for Period Net Income) have the same effect on the predicted PDs of companies in both states. On the other hand, variables with coefficients that are different for both states (e.g. EBITDA and Net Current Assets) influence the PDs of companies differently depending on the state. This way, our framework allows to identify if an explanatory variable has a different effect on the PDs depending on the state. This information can help to draw conclusions on which variables do well in distinguishing between the states.

It is also worth mentioning which state is the "Good" (or "Normal") and which one is the "Bad" (or "Abnormal"). The average estimated (uncalibrated) PD for state 1 is 0.26, while for state 2 it is 0.71. Therefore, we can infer that state 1 is the "Normal" state and state 2 is the "Abnormal". This explains the opposing sign of the corresponding constants $\hat{a}_1$ and $\hat{a}_2$.

It is important to note that in our model, the explanatory variables are not used in their raw form, but rather after binning and transforming them to their WoE values, as illustrated in

Table 4: Estimated Coefficients

Variables	LR	STR State 1	STR State 2
Constant	0.104*	-0.370***	1.072***
Shareholder Funds	-0.271***	0.207***	-0.041***
Other Shareholder Funds	-0.219***	-0.140***	-0.119***
Cashflow	-0.339***	-0.319***	-0.286
EBITDA	-0.130***	-0.049**	-0.346*
Tangible Fixed Assets	-0.178***	-0.406***	-0.097***
Net Current Assets	-0.574***	-1.202***	0.104***
P&L for Period Net Income	-0.361***	-0.316	-0.320***
Operating P&L EBITDA	-	-0.061**	-0.213***
$\hat{c}$	-	0.96	
$\hat{\gamma}$	-	8.80	

Estimated coefficients for the two models LR and STR. *,** and *** indicate a significance level of 10%, 5% and 1%, respectively. For the calculation of confidence intervals for the estimates of the STR model see Table 5 and Section A1 in the Appendix. Significance testing is not reported for the estimates $\hat{c}$ and $\hat{\gamma}$ due to the nature of their constraints as outlined in González et al. (2005). Specifically, $\hat{c}$ must lie within the range of $(min\{q_{i,t}\}, max\{q_{i,t}\})$, while the sign of $\hat{\gamma}$ is irrelevant, as a negative sign would merely reflect the transition function in Figure 5(a) about the Y-axis. This way, both estimates are always positive.

Table 2. As a result, when discussing an increase in a variable, we are referring to an increase in terms of WoE value, rather than the variable itself. The relationship between the variable's value and its WoE value is not perfectly linear, but there is a general correspondence, as shown in Figure 6 in the Appendix. Therefore, in interpreting the coefficients of our model, we describe general trends rather than one-to-one relationships. For this reason, we only specifically address the coefficients of Shareholder Funds, as this variable is crucial to our analysis with the highest IV and presents opposing signs in its coefficients.

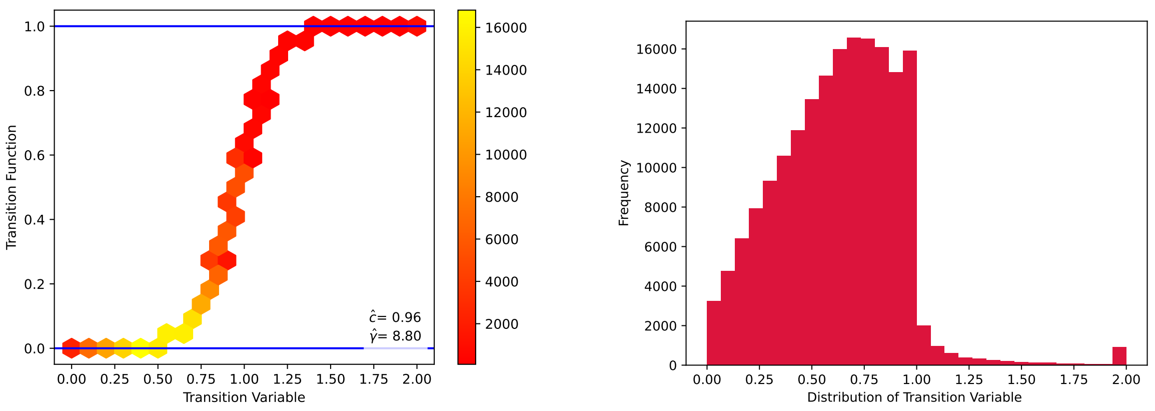
As a general rule, a negative coefficient can be interpreted as that an increase of its respective variable will result in a decrease of the predicted PD (ceteris paribus). Conversely, an increase on a variable with a positive coefficient will lead in a increase of the predicted PD (ceteris paribus). Hence, the sign of the coefficient of Shareholders Funds in state 1 is counter-intuitive, as an increase in this variable leads to a higher PD. One potential explanation for this irregularity is that for a company in the "Good" state, current shareholders may interpret the issuance of additional shares as a dilution of their ownership. Naturally, this does not translate into an increase in the PDs, but it may be creating this effect indirectly. On the contrary, in state 2 an increase in shareholders funds decreases the PD. This comes unsurprising, as later we identify companies in state 2 as leveraged companies (see section 5.4) and this funds could be used to repay some of the debt.

5.4 Interpretation of the STR transition

As mentioned in the Introduction, the interpretability of the model is crucial to abide by regulator's requirements. Here, we demonstrate how our framework has a simple and clear interpretation and allows to track the drivers of the model. For this purpose, in the following we analyze our final STR model from the previous subsection.

The estimated coefficients are shown in Table 4. So far we have discussed how to interpret the coefficients of each variable, now we shift the focus to the parameters c and γ of the transition function $g(q_{i,t}, \gamma, c)$ in eq. (3). Those estimated coefficients are responsible for the shape of the transition function as explained in Section 4.1. Specifically, the constant c stands for the threshold between regimes. This way, companies that have a debt ratio lower than $\hat{c} = 0.96$ are rather in state 1 than state 2 and vice-versa for companies that have a debt ratio higher than 0.96. Additionally, γ determines the smoothness of the transitions.

Figure 5: Transition function shape and Distribution of $q_{i,t}$



(a) Hexbin plot of debt ratio w.r.t. trans. func.

(b) Histogram of the values of the debt ratio.

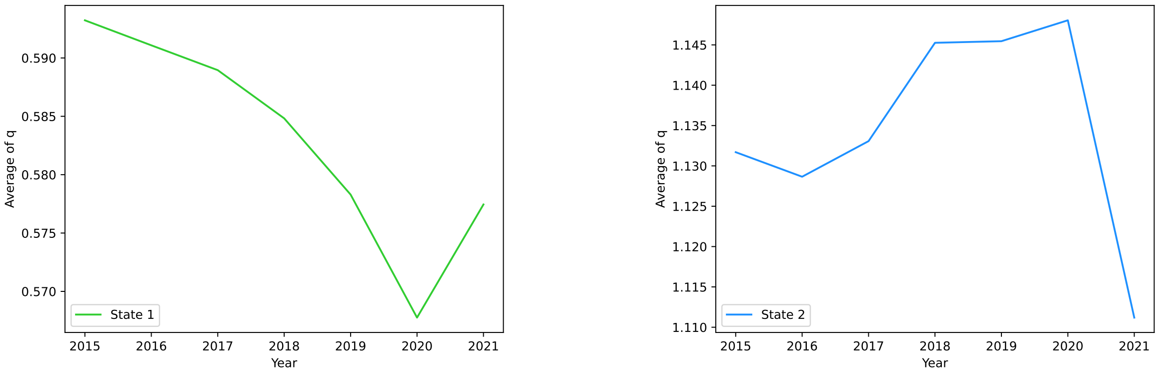
Figure 5(a) represents the shape of the transition function $g(\text{debt-ratio}_{i,t}, \hat{\gamma}, \hat{c})$, while the colors represent the frequency of the values that the transition function takes. As it can be seen the transition is smooth between state 1 and 2, where states are represented by the two blue lines.

We observe from Figure 5(b) that the great majority of companies have a debt ratio smaller than $\hat{c} = 0.96$. This way, we give state 1 the economic interpretation of having a "normal" level of leverage, i.e. a leverage equal or lower than 0.96. Conversely, we give state 2 the economic interpretation of having an "abnormally high" level of leverage. This way, the STR model basically identifies leveraged companies and gives them a special treatment by assigning a higher weight $g(\text{debt-ratio}_{i,t}, \hat{\gamma}, \hat{c})$ to the parameter θ_2 in eq. (2) (corresponding to the coefficients in column "STR State 2" of Table 4). This economic interpretation also explains the results in Figure 4. The slightly improvement comes from the fact of filtering out the few "abnormally high" leveraged companies.

At this point we deviate in order to address the relation of our model to the Covid-19 pandemic. Although our initial aim was to capture the unique Covid-19 situation in one of our states, the limited number of years of available data for some of our companies (maximum of 7 years) has led us to use cross-sectional differences rather than time differences. As a result, the direct link between our study and the impact of Covid-19 on PDs forecasting is limited.

As mentioned above, for the debt ratio transition variable, the different states distinguish the "normally leveraged" firms from the exceptionally "high leveraged" firms. This is thus not directly in link with the Covid-19 situation. However, by looking at Figure 6 we can see that in 2021, the average of the transition variable in both states have a sudden positive and negative jump, respectively. During the Covid period, there is a downwards shift of the highly leveraged firms from state 2 to state 1, which explains the change in average of state 2. As these firms drop just under the cutoff point $\hat{c}$, this explains the increase in the average of debt ratio of the state 1.

Figure 6: Average distribution of the debt ratio $q_{i,t}$ for the two States with estimated $\hat{c}$



(a) State 1 ($q_{i,t} < \hat{c} = 0.957$)

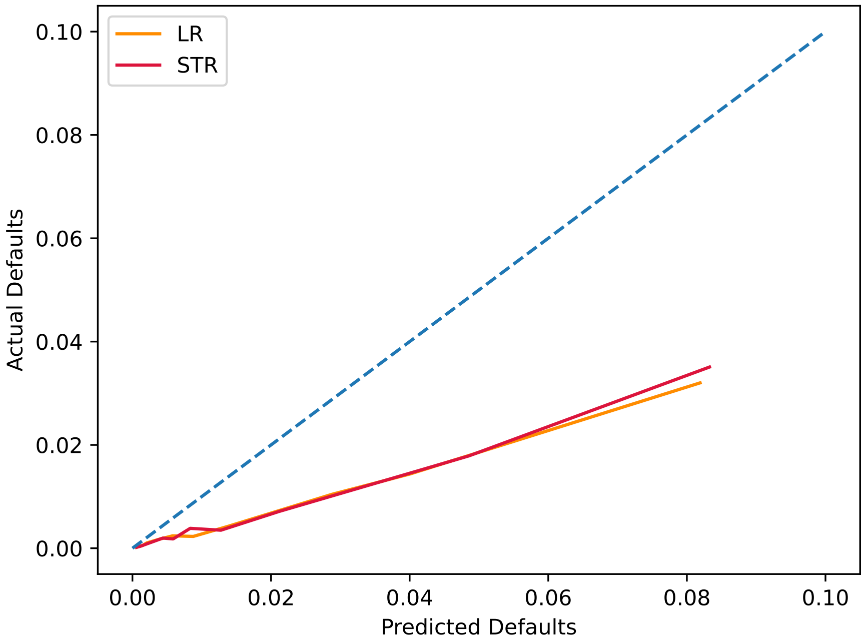
(b) State 2 ($q_{i,t} > \hat{c} = 0.957$)

With these sudden jumps, we can see that the model is able to capture some changes in the data due to Covid-19 and we believe that the model has a performance increase by finding an appropriate transition variable.

5.5 Probabilities of Default Calibration

Using the approach specified in Section 4.4, we calibrate the predicted PDs. Figure 7 displays the calibration curves of the LR and STR model using 10 different bins. We can see that the STR-calibrated probabilities are slightly better than the LR probabilities, but both are still close to each other. This modest improvement further confirms the results obtained for the AUC performance measure. In Figure 8 located in the Appendix we display calibration curves for different numbers of bins. These additional graphs with higher number of bins add a lot of variability to the curve, which could be attributed to overfitting the predicted probabilities of defaults.

Figure 7: Calibration Curve, N = 10



6 Conclusion

In conclusion, the Weight of Evidence (WoE) approach proved to be a powerful variable selection method in the context of credit risk modelling. Parsimonious models with 7 or 8 variables achieve a high AUC when predicting PDs, maintaining high levels of model interpretability as required by regulators.

The Logistic Regression benchmark model confirmed to be a viable option due to its simplicity and yet powerful predictive performance. However, our proposed STR model improves the LR in terms of AUC, achieving an AUC of 0.8600 on the forecasting sample compared to the AUC of 0.8475 attained by the LR.

The choice of the transition variable $q_{i,t}$ showed to be fundamental, as the results of the STR are sensitive to the choice of this variable. We found that for our credit risk application, the best transition variable is the debt ratio. We give the model using this variable the economic interpretation of a filter for highly leveraged firms, that allows to give these firms a special treatment. This leads to the slight improvement compared to the LR model.

Ultimately, this is where our model exhibits its main strengths. Firstly, depending on the application, the user may search for the optimal $q_{i,t}$, i.e. a variable that allows to filter out firms that show an abnormal value for this variable. This holds while keeping a simple and interpretable model. Secondly, even though its forecasting performance could sometimes be relatively similar to a LR model, the STR can be used to extract important information, e.g. the possible presence of different states, and allow to get insights about the models and its variables.

Lastly, different extensions to our model could be considered. For example, considering more than two states could be an option. Furthermore, transition variables of a different nature could be considered. These could be linear combinations of the already considered transition variables, or variables extracted from another model. Doing so, could potentially improve the model's performance, while at the cost of model interpretability.

References

Adjei, I. A., & Karim, R. (2016). An application of bootstrapping in logistic regression model. *Open Access Library Journal*, *3*(9), 1–9.

Devi, S. S., & Radhika, Y. (2018). A survey on machine learning and statistical techniques in bankruptcy prediction. *International Journal of Machine Learning and Computing*, *8*(2), 133–139.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861–874.

González, A., Teräsvirta, T., van Dijk, D., & Yang, Y. (2005). Panel Smooth Transition Regression Models. (604).

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction* (2nd ed.). Springer.

Hauser, R. P., & Booth, D. (2011). Predicting bankruptcy with robust logistic regression. *Journal of Data Science*, *9*(4), 565–584.

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, *5*(2), 1.

Huang, A. Y., & Hu, W.-C. (2012). Regime switching dynamics in credit default swaps: Evidence from smooth transition autoregressive model. *Physica A: Statistical Mechanics and its Applications*, *391*(4), 1497–1508.

Jabeur, S. B. (2017). Bankruptcy prediction using partial least squares logistic regression. *Journal of Retailing and Consumer Services*, *36*, 197–202.

Lin, S. M., Ansell, J., & Andreeva, G. (2012). Predicting default of a small business using different definitions of financial distress. *Journal of the Operational Research Society*, *63*(4), 539–548.

Ling, C. X., Huang, J., Zhang, H., et al. (2003). Auc: A statistically consistent and more discriminating measure than accuracy. *Ijcai*, *3*, 519–524.

Martin, D. (1977). Early warning of bank failure: A logit regression approach. *Journal of banking & finance*, *1*(3), 249–276.

Nehrebecka, N. (2016). Approach to the assessment of credit risk for non-financial corporations. Evidence from Poland (Bank for International Settlements, Ed.). *41*.

Nikolic, N., Zarkic-Joksimovic, N., Stojanovski, D., & Joksimovic, I. (2013). The application of brute force logistic regression to corporate credit scoring models: Evidence from Serbian financial statements. *Expert Systems with Applications*, *40*(15), 5932–5944.

Telg, S., Dubinova, A., & Lucas, A. (2023). Covid-19, credit risk management modeling, and government support. *Journal of Banking & Finance*, *147*, 106638.

Terasvirta, T., & Anderson, H. M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, *7*(S1), S119–S136.

Ullah, A., Zhang, Q., Raza, S. A., & Ali, S. (2021). Renewable energy: Is it a global challenge or opportunity? Focusing on different income level countries through Panel Smooth Transition Regression Model. *Renewable Energy*, *177*, 689–699.

Wilcoxon, F. (1945). Individual comparisons by ranking methods.

Yang, X., Zhu, Y., Yan, L., & Wang, X. (2015). Credit Risk Model Based on Logistic Regression and Weight of Evidence, 810–814.

Zeng, G. (2014). A Necessary Condition for a Good Binning Algorithm in Credit Scoring. *Applied Mathematical Sciences*, *Vol. 8*, 3229–3242.

Appendix

A1 Confidence Intervals for the STR model

In this section, we compute confidence intervals of the estimates of the STR model using the non-parametric bootstrap in Adjei and Karim (2016). The procedure is as follows:

1. We create a new data set for binary response with covariates $(y_{i,t}, x_{i,t})$ from the balanced training subsample (for n=N+T=29244 observations).

2. We sample the pairs with replacement from the new data set for a total of $B = 1000$ bootstrap samples $(y_{i,t}, x_{i,t})^*_b = ((y_{1_{i,t}}, x_{1_{i,t}})^*, ..., (y_{n_{i,t}}, x_{n_{i,t}})^*)$, for $b = 1, ..., B$.

3. For each $b = 1, ..., B$ estimate the bootstrap coefficients $(\hat{\theta}_1, \hat{\theta}_2)^*_1$,...,$(\hat{\theta}_1, \hat{\theta}_2)^*_B$, where $(\hat{\theta}_1, \hat{\theta}_2)^*_b$ are the estimates by refitting an STR model on each bootstrap sample.

4. Estimate $(1 - \alpha)100\%$ bootstrap confidence interval by finding empirical quantiles of bootstrap replicates:
$$(\hat{\theta}_L, \hat{\theta}_U) = (\hat{\theta}^{(\frac{\alpha}{2})}_{(b)}, \hat{\theta}^{(1-\frac{\alpha}{2})}_{(b)}).$$

Table 5 below displays the results of the described procedure for a significance level of $\alpha = 0.05$. Alternatively (and out of the scope of this study), we could use Maximum Likelihood inference to estimate standard errors for our estimates. For this purpose, consider the information matrix, defined as

$$\mathbf{I}(\theta) = -\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{\partial^2 l(\theta; x_{i,t}, y_{i,t})}{\partial\theta\partial\theta'}.$$

Its expectation $\mathbb{E}_\theta\left[\mathbf{I}(\theta)\right]$ is called the Fisher information.

A standard result says that under some regularity conditions, the sampling distribution of the maximum likelihood estimator has a limiting normal distribution (Hastie et al. (2009)). This means that $\hat{\theta} \to \mathcal{N}(\theta,\, \mathbb{E}_\theta\left[\mathbf{I}(\theta)\right]^{-1}),\ as\ \ N \to \infty$, where θ is the vector containing the true parameters. Hence, we approximate the sampling distribution of $\hat{\theta}$ by $\mathcal{N}(\hat{\theta},\, \mathbf{I}(\hat{\theta})^{-1})$ and finally obtain estimates for the standard errors of $\hat{\theta}$ with $\sqrt{\mathbf{I}(\hat{\theta})^{-1}_{k,k}}$.

Table 5: Confidence interval with a 95% Confidence Level

Coefs	Lower	Upper
α_1	0.586	1.589
α_2	-5.012	-0.745
$\beta_{1,1}$	-0.432	-0.287
$\beta_{1,2}$	-0.283	-0.071
$\beta_{2,1}$	-0.273	-0.149
$\beta_{2,2}$	-0.244	-0.096
$\beta_{3,1}$	-0.818	-0.547
$\beta_{3,2}$	-0.223	0.259
$\beta_{4,1}$	-0.287	-0.031
$\beta_{4,2}$	-0.203	0.021
$\beta_{5,1}$	-0.325	-0.199
$\beta_{5,2}$	-0.297	-0.067
$\beta_{6,1}$	-0.219	-0.049
$\beta_{6,2}$	-1.555	-1.001
$\beta_{7,1}$	-0.081	0.175
$\beta_{7,2}$	-0.903	-0.509
$\beta_{8,1}$	0.214	0.783
$\beta_{8,2}$	0.898	3.722

A2 WoE IV Transformation

Table 6: Bins of the six variables with highest IV computed on the balanced estimating subsample

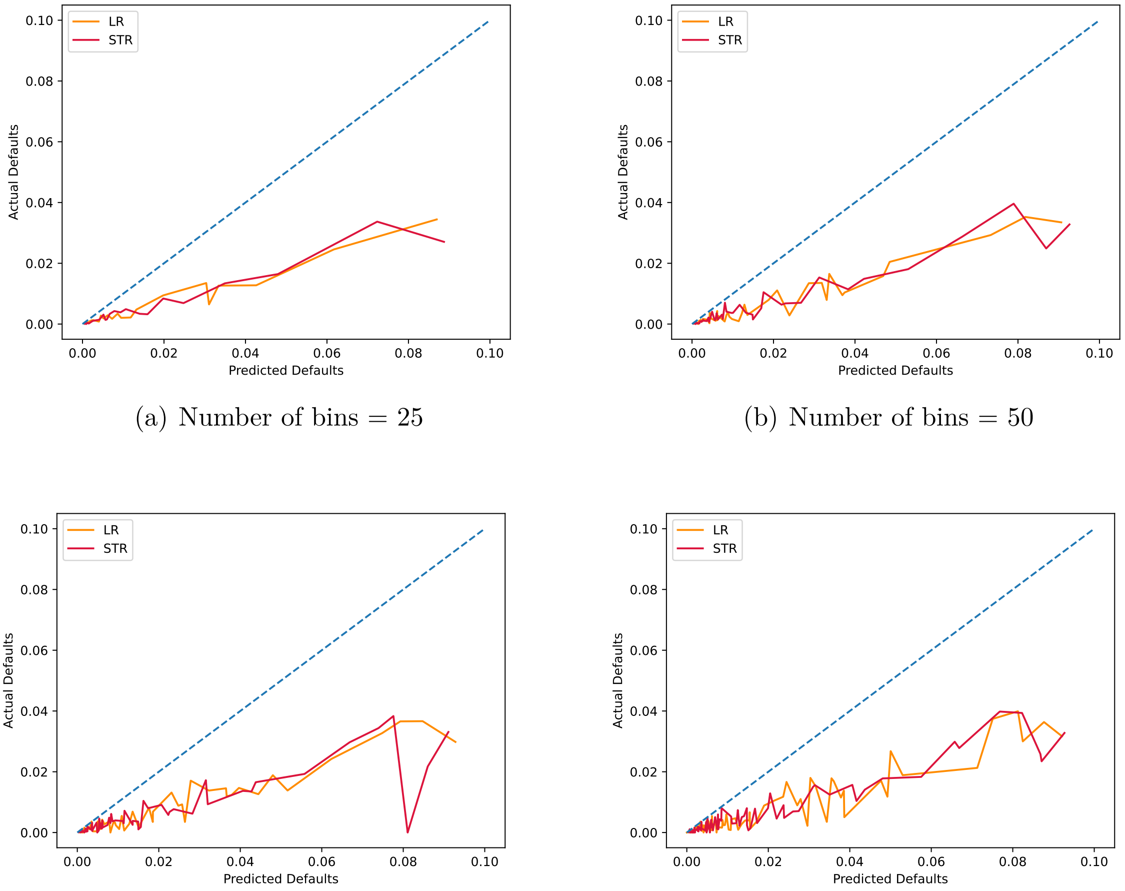
variable	bin	WoE	variable	bin	WoE
shareholders funds	0	-0.5276	ebitda	0	-0.7932
shareholders funds	1	-1.6015	ebitda	1	-1.3409
shareholders funds	2	-1.7000	ebitda	2	-1.4107
shareholders funds	3	-1.5731	ebitda	3	-1.0959
⋮	⋮	⋮	⋮	⋮	
shareholders funds	18	1.6949	ebitda	15	1.5517
shareholders funds	19	1.7732	ebitda	16	1.4501
other shareholders funds	0	-1.3487	tangible fixed assets	0	-0.4891
other shareholders funds	1	-1.1985	tangible fixed assets	1	-1.9526
other shareholders funds	2	-1.2585	tangible fixed assets	2	-1.0790
⋮	⋮	⋮	⋮	⋮	⋮
other shareholders funds	13	1.6601	tangible fixed assets	7	0.8653
other shareholders funds	14	1.8935	tangible fixed assets	8	1.0416
cash flow	0	-0.7872	net current assets	0	0.9992
cash flow	1	-1.2822	net current assets	1	-0.6054
cash flow	2	-1.5544	net current assets	2	-0.7131
⋮	⋮	⋮	⋮	⋮	⋮
cash flow	18	1.6257	net current assets	15	0.9757
cash flow	19	1.4958	net current assets	16	1.2305

Table 7: Bins of the six variables with highest IV computed on the balanced estimating subsample

variable	IV rank	bin	min. value	max. value	variable	IV rank	bin	min. value	max. value
shareholders funds	1	0	nan	nan	ebitda	4	0	nan	nan
shareholders funds	1	1	-4352540990	-3791657	ebitda	4	1	-4122954567	-3910144
shareholders funds	1	2	-3783414	-591719	ebitda	4	2	-3910004	-1354502
shareholders funds	1	3	-591144	344	ebitda	4	3	-1352536	-452442
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
shareholders funds	1	18	16303308	42157011	ebitda	4	15	5978000	13522818
shareholders funds	1	19	42198091	38367000000	ebitda	4	16	13547609	9812178597
other shareholders funds	2	0	nan	nan	tangible fixed assets	5	0	nan	nan
other shareholders funds	2	1	-5681771641	-7380021	tangible fixed assets	5	1	-48983	0
other shareholders funds	2	2	-7359000	-1729614	tangible fixed assets	5	2	1	344
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
other shareholders funds	2	13	7981960	21486390	tangible fixed assets	5	7	2565656	9469210
other shareholders funds	2	14	21520208	36840000000	tangible fixed assets	5	8	9473040	59741862863
cash flow	3	0	nan	nan	net current assets	6	0	nan	nan
cash flow	3	1	-1296164093	-4393985	net current assets	6	1	-4667885676	-6937088
cash flow	3	2	-4392000	-1519275	net current assets	6	2	-6933261	-1744455
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
cash flow	3	18	5706003	13034475	net current assets	6	15	6505215	16265603
cash flow	3	19	13038880	9382233730	net current assets	6	16	16266049	30457485489

A3 PD Binning with different bins

Figure 8: Calibration curve for different number of bins.



(a) Number of bins = 25

(b) Number of bins = 50

(c) Number of bins = 75

(d) Number of bins = 100