

SFB 649 Discussion Paper 2013-037

Default Risk Calculation based on Predictor Selection for the Southeast Asian Industry

Wolfgang Karl Härdle*
Dedy Dwi Prastyo*



* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

Default Risk Calculation based on Predictor Selection for the Southeast Asian Industry

Wolfgang Karl Härdle

Humboldt-Universität zu Berlin, Ladislaus von Bortkiewicz Chair of Statistics
Center for Applied Statistics and Economics (C.A.S.E.)
Unter den Linden 6, 10099 Berlin, Germany
Email: haerdle@wiwi.hu-berlin.de

and

Lee Kong Chian School of Business, Singapore Management University

Dedy Dwi Prastyo

Humboldt-Universität zu Berlin, Ladislaus von Bortkiewicz Chair of Statistics
Center for Applied Statistics and Economics (C.A.S.E.)
Email: prastyod@hu-berlin.de

and

Department of Statistics, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Sukolio, Surabaya 60111, Indonesia
Email: dedy-dp@statistika.its.ac.id

Abstract

Probability of default prediction is one of the important tasks of rating agencies as well as of banks and other financial companies to measure the default risk of their counterparties. Knowing predictors that significantly contribute to default prediction provides a better insight into fundamentals of credit risk analysis. Default prediction and default predictor selection are two related issues, but many existing approaches address them separately. We employed a unified procedure, a regularization approach with logit as an underlying model, which simultaneously selects the default predictors and optimizes all the parameters within the model. We employ Lasso and elastic-net penalty functions as regularization approach. The methods are applied to predict default of companies from industry sector in Southeast Asian countries. The empirical result exhibits that the proposed method has a very high accuracy prediction particularly for companies operating Indonesia, Singapore, and Thailand. The relevant default predictors over the countries reveal that credit risk analysis is sample specific. A few number of predictors result in counter intuitive sign estimates.

Keyword: Default risk, Predictor selection, Logit, Lasso, Elastic-net
JEL Classification: C13, C61 ,G33

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 'Economic Risk', Humboldt-Universität zu Berlin. Dedy Dwi Prastyo was also supported by Directorate General for Higher Education, Indonesian Ministry of Education and Culture through Department of Statistics, Institut Teknologi Sepuluh Nopember (ITS), Indonesia. We thank to Risk Management Institute (RMI) at the National University of Singapore (NUS) for the data used in this study and for the partial financial support under Credit Research Initiative (CRI) project.

1. Introduction

Default is one of the most important events in a company's life which can place stakeholders in financial trouble. The default effect to both economy and society depend on the company's size and the systemic implication of the defaulting company. For a company without any systemic implication, e.g. small company, default event still gives great impact to the stakeholders.

Probability of default (PD) prediction is one of the important tasks of rating agencies in credit risk assessment as well as of banks and other financial companies to measure the default risk of their counterparties. Credit scoring analysis is used to obtain the PD of companies as well as of individual client. Altman and Hotchkiss (2006) mentioned that at least two important factors stimulating the development of credit scoring models: the implications of Basel II's proposed capital requirements on credit assets as well as the enormous amounts and rates of defaults in the United States. Hence, banks and other financial institutions either developed or modified their existing internal credit risk systems. Under the Basel II, the PD and loss given default (LGD) are necessary to qualify for improving the internal rating based (IRB) approach. Therefore, the methods to assess PD and LGD on credit assets are continuously developed by banks and other financial institutions as well as by researchers in the field.

There are many statistical parametric and non-parametric methods used to measure default risk which are mainly grouped into two: reduced-form and market based approach. A market-based model describes capital structure and asset value dynamic of the companies, see Vassalou and Xing (2004), in contrast to reduced-form which directly analyzes financial ratios data in cross-sectional framework. It is difficult to apply the structure models to the most companies unlisted in capital markets since the required data are not available. In such a case, the reduced-form models which rely on cross-sectional data are more appropriate.

Most of the early studies on default prediction were done by comparing financial ratios of default and non-default companies. The univariate discriminant analysis (DA) was introduced by Beaver (1966). He used different threshold points for different single financial ratio to obtain minimum misclassification. In the single financial ratio sense, he found that cash-flow to total debt seemed to provide the most accurate default prediction, followed by net income to total asset ratio. However, default event are too complex to be explained by a single financial ratio. Univariate approach can only apply for a single financial ratio at a time such that the different financial ratio may result in various classification schemes for the same company. Therefore, Altman (1968) introduced multivariate DA to investigate multi financial ratio effect to default prediction. He used five financial ratios as discriminating variables: working capital to total assets, retained earnings to total assets, earnings before interest and taxes (EBIT) to total assets, market value of equity to book value of the total debt, and sales to total assets. The accounting data for the defaulted company are obtained from annual financial report available prior to the default date. The multivariate DA approach is also known as Altman's Z-score analysis since it yields Z score representing ordinal ranking of credit score. The coefficients in DA model are difficult to interpret. Moreover, the DA does not have intuitive interpretation of PD.

The next development of default prediction analysis showed that reduced form approach mostly employ parametric model such as logit and probit regression, see Martin (1977), Ohlson (1980), Lo (1986), Lau (1987), and Platt, Platt and Pedersen (1994). The logit function is bounded between zero and one therefore it suitable to represent PD. The score obtained from the logit model is known as Ohlson's O-score which can be used to rank-order companies. In addition, the coefficients of logit model have natural interpretations, i.e. a positive coefficient implies that an increase in the value of a single financial ratio will increase the PD. In case of there is no collinearity, it is usually not the case in default prediction analysis, the coefficient value represents the relative important of the

corresponding financial ratio. The different scale of attributes, e.g. not financial ratios, makes this advantage is no longer satisfied. Moreover, logit model lends itself to a direct statistical inference of significance test.

The relationship between a company's financial ratios and PD may be nonlinear and too complex for the parametric modeling approaches such as DA and logit model. A nonlinear modeling tool may more appropriate. Recently, the non-parametric classification techniques are proposed such as Classification and Regression Tree (CART) as well as Bayesian Additive Classification Tree (BACT), see Zhang and Härdle (2010), *k*-Nearest Neighbors (Henley and Hand, 1996), Artificial Neural Network (ANN), see Tam and Kiang (1992), Wilson and Sharda (1994) and Altman, Marco and Varetto (1994). ANN is subject to critique since the existence of multiple minima. In such a case, the application of Support Vector Machines (SVM) in credit scoring analysis, see Härdle et al. (2009), Chen, Härdle and Moro (2011) and Härdle, Prastyo and Hafner (2013), typically outperformed the competing models since SVM has a unique global solution, Vapnik (1998) and Steinwart and Christmann (2008).

The nonparametric methods are chosen because these methods provide more flexible approach, i.e. they are able to accommodate the non-monotone relations between univariate financial ratio and the PD in various forms. However, this non-monotone relationship, in certain case, does not make sense from the economics point of view. Therefore, the credit officers may do not accept these sophisticated approaches even they outperform the simpler ones. They choose parametric approach in order to get intuitive interpretation. This paper gives an account of default predictor selection using regularization approach in parametric underlying model, i.e. logit model.

The variable selection gives advantages when a sparse representation is required in order to avoid irrelevant default predictors leading to potential over fitting. Most of the reduced-form approaches are supervised classification methods which employ many default predictors as inputs, although

only a subset is relevant. Furthermore, the financial ratios as default predictors exhibit high correlation such that some advantages of logit model is no longer applied, for instance the relative importance of the predictors corresponding to their coefficients value. In addition, over fitting, instability, and large standard error of the estimates may occur. In such a case, variable selection addresses these problems. The regularization approach used in this study discards irrelevant default predictors that are required to prevent over fitting.

We employ a least absolute shrinkage and selection operator (Lasso) and an elastic-net penalty function as regularization term. The analysis is applied on financial report data of companies from industry sector which operate in Southeast Asian countries: Indonesia, Malaysia, Singapore, and Thailand. The data are collected and prepared by Risk Management Institute (RMI) of the National University of Singapore (NUS).

The next section describes the dataset more detail as well as defines the financial ratios used in this study. The third section describes in detail the regularization approach with logit model as the underlying model. The last two sections are intended for empirical study of default predictors selection as well as the conclusion. The empirical result reveals that the subsets of selected default predictors are able to predict the default with very high accuracy.

2. Data and Variables

We apply our method to the data which are collected by Risk Management Institute (RMI) of the National University of Singapore (NUS). The data consist of quarterly, semi annually, and annually financial report of companies in the Southeast Asian region span from 1998 to 2012. We index the quarter financial report in such a way to obtain the financial information on regular monthly basis in order to match the default event reported in other database.

Table 1. Financial Ratios

Variable	Ratio	Explanation
Profitability		
x_1	NI / TA	Net Income / Total Assets
x_2	NI / Sales	Net Income / Sales
x_3	OI / TA	Operating Income / Total Assets
x_4	OI / Sales	Operating Income / Sales
Leverage		
x_5	OF / TA	Own Fund / Total Assets
x_6	CL / TA	Current Liabilities / Total Assets
x_7	TD / TA	Total Debt / Total Assets
Cost Structure		
x_8	INT / TD	Interest payment / Total Debt
x_9	EBIT / INT-paid	EBIT / Interest paid
Liquidity		
x_{10}	STD / TD	Short Term Debt / Total Debt
x_{11}	Cash / TA	Cash / Total Assets
x_{12}	Cash / CL	Cash / Current Liabilities
x_{13}	QA / CL	(Cash and cash equivalent - Inventories) / Current Liabilities
x_{14}	CA / CL	Current Assets / Current Liabilities
x_{15}	WC / TA	Working Capital / Total Assets
x_{16}	CL / TL	Current Liabilities / Total Liabilities
Activity		
x_{17}	TA / Sales	Total Assets / Sales
x_{18}	INV / Sales	Inventories / Sales
x_{19}	AR / Sales	Account Receivable / Sales
x_{20}	AP / C-Sales	Account Payable / Cost of Sales
Dynamics		
x_{21}	Sale-growth	One year growth in Sales
x_{22}	NI-growth	One year growth in Net Income
Size		
x_{23}	Log (TA)	Log of Total Assets
x_{24}	Log (Sales)	Log of Sales

Bankruptcy event code from 100 to 120 in the database are used to define default observation which include bankruptcy filings under Chapter 11, Chapter 15, Chapter 7, liquidation, and restructuring. Moreover, default corporation action code from 300 to 333 also used to define default observation. The i -th company is assigned as default ($y_i = 1$) if it files a credit event report within

one year period from the date of the financial report; otherwise it is assigned as non-default ($y_i = 0$).

We select companies from industry sector number 10011 based on Bloomberg industry classification system. The industry sector covers industry group from aerospace/defense, building materials, electrical component and equipments, electronics, engineering and construction, environmental control, hand/machine tools, machinery-construction and mining, machinery-diversified, metal fabricate/hardware, miscellaneous manufacture, packaging and container, shipbuilding, transportation as well as trucking and leasing.

We group financial ratios into seven categories as described in Table 1: profitability, leverage, cost structure, liquidity, activity as well as dynamics and company size. We used these financial ratios as attributes potentially affecting the default. Based on information from RMI database, notice that the operating income (OI) variable used in our study are similar to EBIT. Therefore, all financial ratio computed from EBIT are already represented by OI and vice versa. We transform the company size attributes in log value in order to make them comparable to the other attributes.

We remove observation containing missing values (after compute financial ratio) such that there is no default observation in Philippines remains in the dataset. Hence, we exclude Philippines from the analysis. After filtering, there are six default companies out of 376 total companies in Indonesia; 45 default companies out of 2198 total companies in Malaysia; nine default companies out of 640 total companies in Singapore; 16 default compnaies out of 1406 total companies in Thailand.

The country names represent in which country the companies operate to ensure that they are within the same country are subject to the same disclosure and accounting rules. We apply our method for each country separately such that the different number of the companies does not affect our analysis. We want to know whether the relevant default predictors across countries are same or

not. Moreover, we want see whether inconsistent sign of the estimates across countries are exist or not. The possible counter intuitive sign of corresponding financial ratios also become our interest.

3. Regularization method

In the context of credit scoring we are given a training data set $\{(x_1, y_1), \dots, (x_n, y_n)\}$, with n is sample size, predictors $x_i \in \mathbb{R}^p$, $i = 1, \dots, n$ associates with response variable $y_i \in \mathcal{Y} = \{0, 1\}$ denoting non-default and default, the classification problem can be formulated as a regularization, or a penalization, of loss function $L\{y, f(x)\}$ as follow

$$\min_{\beta_0, \beta} \sum_{i=1}^n L\{y_i, f(x_i)\} + \lambda R(\beta), \quad (1)$$

where $f(x)$ is a classifier which corresponds to parameters β_0 and $\beta = (\beta_1, \dots, \beta_p)^T$ as well as $R(\beta)$ is regularization term with tuning parameter λ .

3.1. Lasso and Elastic-Net Penalties

The regularization method using the Lasso penalization was introduced by Tibshirani (1996) which employ the L_1 -norm penalty to yield a sparse solution, i.e. many estimates shrink to zero. The Lasso regularization approach automatically selects the relevant variables and excludes the non-relevant variables by shrinking their coefficients to zero. Zou and Hastie (2005) stated that the L_1 -norm penalty has two noticeable shortcomings: (i) the number of selected predictors is bounded by the number of samples size as shown in Rosset, Zhu and Hastie (2004), and (ii) the Lasso technique tends to select only one (or a few) predictors from a subset of correlated predictors and shrinks the rest to zero. An elastic-net penalty is proposed to address the drawbacks of the Lasso.

The elastic-net penalty which is introduced by Zou and Hastie (2005) is a mixture of the L_1 -norm and of the L_2 -norm penalties. The two advantages of the elastic-net penalty which address the drawbacks of the Lasso are: (i) the number of selected predictors is no limited by the number of samples size, and (ii) group of correlated predictors can be selected together (group selection). The

L_1 -norm penalty plays the role to allow predictors selection, while the L_2 -norm penalty help to do group selection.

The elastic-net penalty is a compromise between ridge and lasso which is defined:

$$R_\gamma(\beta) = (1 - \gamma) \frac{1}{2} \|\beta\|_2^2 + \gamma \|\beta\|_1 \quad (2)$$

$$= \sum_{j=1}^p (1 - \gamma) \frac{1}{2} \beta_j^2 + \gamma |\beta_j|, \quad (3)$$

with the weight proportion $0 < \gamma < 1$ that should be optimized simultaneously with tuning parameter λ . If $\gamma = 0$ then the penalty $R_\gamma(\beta)$ will be similar to ridge penalty. The Lasso regularization is employed when $\gamma = 1$. The elastic-net with $\gamma = 1 - \varepsilon$ for very small $\varepsilon > 0$ performs do like the Lasso, but removes any degeneracies and wild behavior caused by extreme correlations. As γ increases from zero to one, for a given λ , the sparsity of the solution to equation (1) increase monotonically from zero to the sparsity of the Lasso solution (Friedman, Hastie and Tibshirani, 2010).

3.2. Regularized Logit

The logit model is very popular in credit scoring analysis since it has a probabilistic interpretation embedded in its model. The probability of default and of non-default, respectively, for i -th company given by default predictor x_i is formulated as:

$$\begin{aligned} P(y_i = 1|x_i) &= e^{\beta_0 + x_i^T \beta} / (1 + e^{\beta_0 + x_i^T \beta}) \\ &= \left\{ 1 + e^{-(\beta_0 + x_i^T \beta)} \right\}^{-1}, \end{aligned} \quad (4)$$

and

$$\begin{aligned} P(y_i = 0|x_i) &= 1 - P(y_i = 1|x_i) \\ &= \left(1 + e^{\beta_0 + x_i^T \beta} \right)^{-1}. \end{aligned}$$

The log of odd ratio implies a linear regression model which also known as a score as follow

$$\log \left\{ \frac{P(y_i = 1|x_i)}{P(y_i = 0|x_i)} \right\} = \beta_0 + x_i^T \beta. \quad (5)$$

Instead of minimizing the loss function, the regularization form as in (1) can be represented in an alternative way by maximizing regularized log-likelihood function:

$$\max_{\beta_0, \beta} \{ \ell(\beta_0, \beta) - \lambda R_\gamma(\beta) \}, \quad (6)$$

where the log-likelihood of binary logit model is

$$\begin{aligned} \ell(\beta_0, \beta) &= n^{-1} \sum_{i=1}^n [y_i \log P(y_i = 1|x_i) + (1 - y_i) \log\{1 - P(y_i = 1|x_i)\}] \\ &= n^{-1} \sum_{i=1}^n \{I(y_i = 1) \log P(y_i = 1|x_i)\} + I(y_i = 0) \log P(y_i = 0|x_i) \\ &= n^{-1} \sum_{i=1}^n I(y_i = 1) \left[(\beta_0 + x_i^T \beta) - \log \{1 + e^{\beta_0 + x_i^T \beta}\} \right]. \end{aligned} \quad (7)$$

The log-likelihood function in (7) is a concave function with respect to the parameters, although it is nonlinear. Concave or convex problem is good for computation and theory. The subtraction of the penalty from the log-likelihood function encourages sparse solution. Maximizing the penalized log-likelihood, for a given constant λ , singles out a certain number of non-zero estimates. If we relax the penalty by reducing the value λ , then more number of predictor, denoted as $p^*(\lambda)$, can enter the underlying model. Wu et al. (2009) stated that $p^*(\lambda)$ is basically a decreasing function of λ with jumps of size one, although minor exception occurs. The entry order of the predictors into model tends to be related to its marginal significance. This rule of thumb is violated when the predictors are correlated.

The unpenalized log-likelihood function in (7) is maximized using Newton algorithm which amounts to iteratively reweighted least squares (IRLS). We follow the algorithm of Friedman, Hastie, and Tibshirani (2010) which used the similar approach to IRLS, i.e. using cyclical

coordinate descent which is computed along a regularization path, see also Zou and Hastie (2005). They concluded that the coordinate descent performed best among the several competing algorithms. In cyclic coordinate descent, each estimate of the parameter is updated in turn. Wu and Lange (2008) developed greedy coordinate descent algorithm which update the parameter leading to the largest increase in the objective function. However, the algorithm suffers from excess overhead, although it makes faster initial progress in logit model.

The cyclic coordinate descent algorithm consists of three steps: outer loop, middle loop, and inner loop. In the outer loop step, we set up the value of λ . In the middle loop step, we update the quadratic approximation (Taylor expansion) of the log-likelihood function about current estimates $(\tilde{\beta}_0, \tilde{\beta})$, i.e.:

$$\ell_Q(\beta_0, \beta) = -\frac{1}{2n} \sum_{i=1}^n \omega_i (z_i - \beta_0 - x_i^T \beta)^2 + C(\tilde{\beta}_0, \tilde{\beta}), \quad (8)$$

with working response and weight, respectively, are:

$$z_i = \tilde{\beta}_0 + x_i^T \tilde{\beta} + \frac{y_i - \tilde{P}(y_i = 1|x_i)}{\tilde{P}(y_i = 1|x_i)\{1 - \tilde{P}(y_i = 1|x_i)\}},$$

and

$$\omega_i = \tilde{P}(y_i = 1|x_i)\{1 - \tilde{P}(y_i = 1|x_i)\},$$

where $\tilde{P}(y_i = 1|x_i)$ is evaluated at the current estimates and $C(\tilde{\beta}_0, \tilde{\beta})$ is a constant. Notice that the first term in $\ell_Q(\beta_0, \beta)$ is weighted least square. In the last step, inner loop, the coordinate descent algorithm is used to solve the following penalized weighted least squares (PWLS) problem:

$$\min_{\beta_0, \beta} \{-\ell_Q(\beta_0, \beta) + \lambda R_\gamma(\beta)\} \quad (9)$$

Each inner coordinate-descent loop continues until the maximum change in objective function, equation (9), is less than a very small threshold times null deviance. The threshold used in this study is 1E-7.

The next step is to decrease the value of λ and repeat the three loops until the estimates is convergent. Once the final estimates are obtained, the sparse representation with selected relevant predictor is used to predict the default as well as PD. Shevade and Keerthi (2003) also developed an algorithm to estimate the logit model with Lasso penalty based on the Gauss-Siedel method using coordinate-wise descent approach.

4. Empirical Study

In our modeling the predictors x_{ij} are standardized and the estimates are always returned on the original scale. We optimize γ and λ in the following way: for a fixed γ , the tuning parameter λ is optimized based on Cross-Validation (CV). We used stratified CV which split the data into $k = 5$ disjoint parts of roughly equal size and each fold contains approximately the same distribution of class labels as the whole data set. The optimum λ minimizes the CV or equivalently maximizing area under the curve (AUC) from receiver operating characteristic (ROC) curve. We increase the value of γ from 0.1 to 0.9 for elastic-net and obtain the optimum corresponding λ for each γ . We choose the best γ and λ which maximize the AUC.

We are more interested in AUC to see the performance of our method since AUC is robust to the unbalance data set. This is the case in default prediction analysis where the proportion of the default is much smaller than the non-default. In case of unbalance data set, the classification method may tend to classify every observation belong to the majority, see He and Garcia (2009) for a comprehensive and critical review of the research development on learning process from unbalance data set. AUC is originally proposed by Sobehart and Keenan (2001). See also Härdle, Prastyo and Hafner (2013) for its application.

Table 2. Model Performance.

Accuracy	γ									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Indonesia										
AUC	0.9914	0.9941	0.9761	0.9937	0.9964	0.9752	0.9941	0.9919	0.9941	0.9941
λ	0.0026	0.0005	0.0186	0.0007	5.5E-5	0.0148	0.0002	0.0014	0.0003	0.0006
$\log \lambda$	-5.9584	-7.4889	-3.9869	-7.1586	-9.8007	-4.2149	-8.4625	-6.5493	-8.1556	-7.4236
$\lambda(1se)$	0.1549	0.0007	0.1087	0.0055	0.0021	0.0451	0.0007	0.0044	0.0006	0.0020
$\log\{\lambda(1se)\}$	-1.8649	-7.3028	-2.2193	-5.2049	-6.1723	-3.0985	-7.2531	-5.4329	-7.5044	-6.2143
p^*	20	21	9	18	20	6	18	13	12	12
Malaysia										
AUC	0.8718	0.8644	0.8510	0.8643	0.8552	0.8692	0.8679	0.8599	0.8701	0.8608
λ	1.2E-5	9.5E-5	0.0006	0.0001	0.0004	5.5E-5	9.1E-5	0.0003	7.7E-5	0.0003
$\log \lambda$	-11.3625	-9.2646	-7.4373	-9.1204	-7.7621	-9.8050	-9.3079	-8.2321	-9.4662	-8.2691
$\lambda(1se)$	0.0009	0.0039	0.0022	0.0012	0.0011	0.0021	0.0010	0.0007	0.0008	0.0007
$\log\{\lambda(1se)\}$	-6.9899	-5.5432	-6.1348	-6.7016	-6.8317	-6.1767	-6.8891	-7.3017	-7.1404	-7.2458
p^*	24	24	22	23	21	24	23	22	22	21
Singapore										
AUC	0.9749	0.9816	0.9582	0.9753	0.9705	0.9733	0.9740	0.9784	0.9739	0.9696
λ	0.0009	0.0013	0.0057	0.0019	0.0026	0.0019	0.0018	0.0012	0.0017	0.0023
$\log \lambda$	-7.0454	-6.6222	-5.1670	-6.2919	-5.9569	-6.2322	-6.2934	-6.7059	-6.3586	-6.0918
$\lambda(1se)$	0.0074	0.0094	0.0120	0.0047	0.0059	0.0041	0.0082	0.0021	0.0058	0.0052
$\log\{\lambda(1se)\}$	-4.9056	-4.6685	-4.4227	-5.3616	-5.1196	-5.4879	-4.8048	-6.1478	-5.1492	-5.2545
p^*	21	20	13	16	13	13	12	14	11	9
Thailand										
AUC	0.9826	0.9700	0.9824	0.9634	0.9346	0.9604	0.9376	0.9622	0.9906	0.9931
λ	0.0026	0.0075	0.0018	0.0066	0.0122	0.0053	0.0059	0.0039	0.0003	0.0001
$\log \lambda$	-5.9622	-4.8877	-6.3165	-5.0226	-4.4085	-5.2420	-5.1171	-5.5297	-8.1594	-8.9160
$\lambda(1se)$	0.0054	0.0091	0.0244	0.0221	0.0233	0.0111	0.0126	0.0146	0.0022	0.0010
$\log\{\lambda(1se)\}$	-5.2179	-4.7016	-3.7116	-3.8132	-3.7572	-4.4978	-4.3728	-4.2272	-6.1127	-6.8693
p^*	21	16	20	12	7	11	8	10	20	19

Table 2 summarizes the performance of our method for different γ and its corresponding optimum λ for all countries. For logit model with Lasso regularization ($\gamma = 1$), the optimum λ of Indonesian, Malaysian, Singaporean, and Thailand, respectively, are 0.0006, 0.0003, 0.0023, and 0.0001. In elastic-net approach we apply parsimonious principle, i.e. we choose smaller selected default predictor when the AUC values only differ very small. For Indonesia we choose $\gamma = 0.4$

instead of $\gamma = 0.5$ since it selects two less predictors. The optimum λ for the corresponding γ is 0.0007. For Malaysia we choose $\gamma = 0.9$ instead of $\gamma = 0.1$ for the same reason with optimum λ is 0.000077. The optimum γ for Singapore and Thailand are 0.8 and 0.9, respectively, with optimum λ are 0.0012 and 0.0003 for each corresponding γ . The number of selected default predictor, p^* , in the Lasso approach are much smaller than the one in elastic-net particularly for Indonesia, Singapore and Thailand.

Table 2 also reports the value of $\log\{\lambda(1se)\}$ which express the largest value of $\log \lambda$ such that the CV error is within one standard-error of the minimum. This approach is so called "one-standard-error" rule. This value is represented by right vertical line in Figure 1a and 1b. The left vertical line in each plot corresponds to mean cross-validated error curve or equivalently mean AUC. The fixed values of left and right vertical lines can be checked in Table 2. The top of each plot is annotated with the number of predictors are selected. For $p^* \geq 7$ on Indonesia data set, both Lasso and elastic-net yield very high accuracy where the AUC almost close to one. This is not the case for Malaysia data set where even $p^* \geq 18$ the AUC value are ranging from 0.80 to 0.87. Singapore and Thailand data set result in very high accuracy of default prediction in term of AUC. Singapore has smaller number of default predictors with little higher AUC values than Thailand.

It can be seen from Table 3 that, for each country, the financial ratios selected in Lasso is almost always selected in elastic-net approach particularly for Indonesia in which number of selected default predictors from Lasso is significantly smaller than from elastic-net. For each country the sign of the estimates in both approaches are always same except for own fund ratio (OF/TA) from Thailand data. Across the countries, only nine out of twenty four financial ratios have a consistent sign: OI/TA, OI/Sales, EBIT/INT-paid, Cash/CL, WC/TA, INV/Sales, AR/Sales, NI growth, and log (TA). These financial ratios are highlighted in Table 3. The remaining financial ratios have different

sign across countries. This is in line with Balcaen and Ooghe (2006) who said that the default prediction based on logit model is sample specific, i.e. country specific.

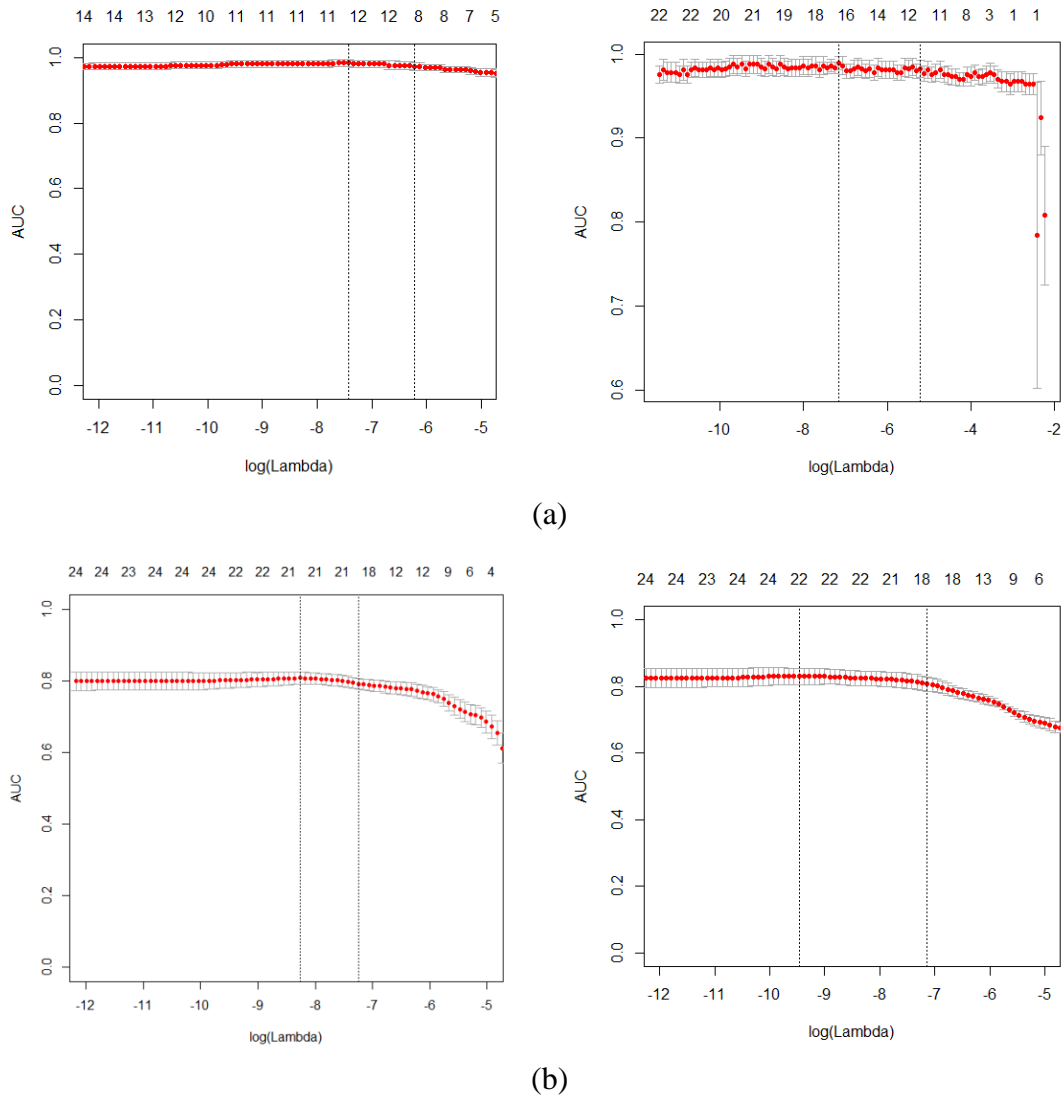


Figure 1a.
AUC values for Logit Model with Lasso (left) and Elastic-Net (right) Regularization:
(a) Indonesia, (b) Malaysia. The γ values for Elastic-Net Penalty Function are:
(a) $\gamma = 0.4$, (b) $\gamma = 0.9$.

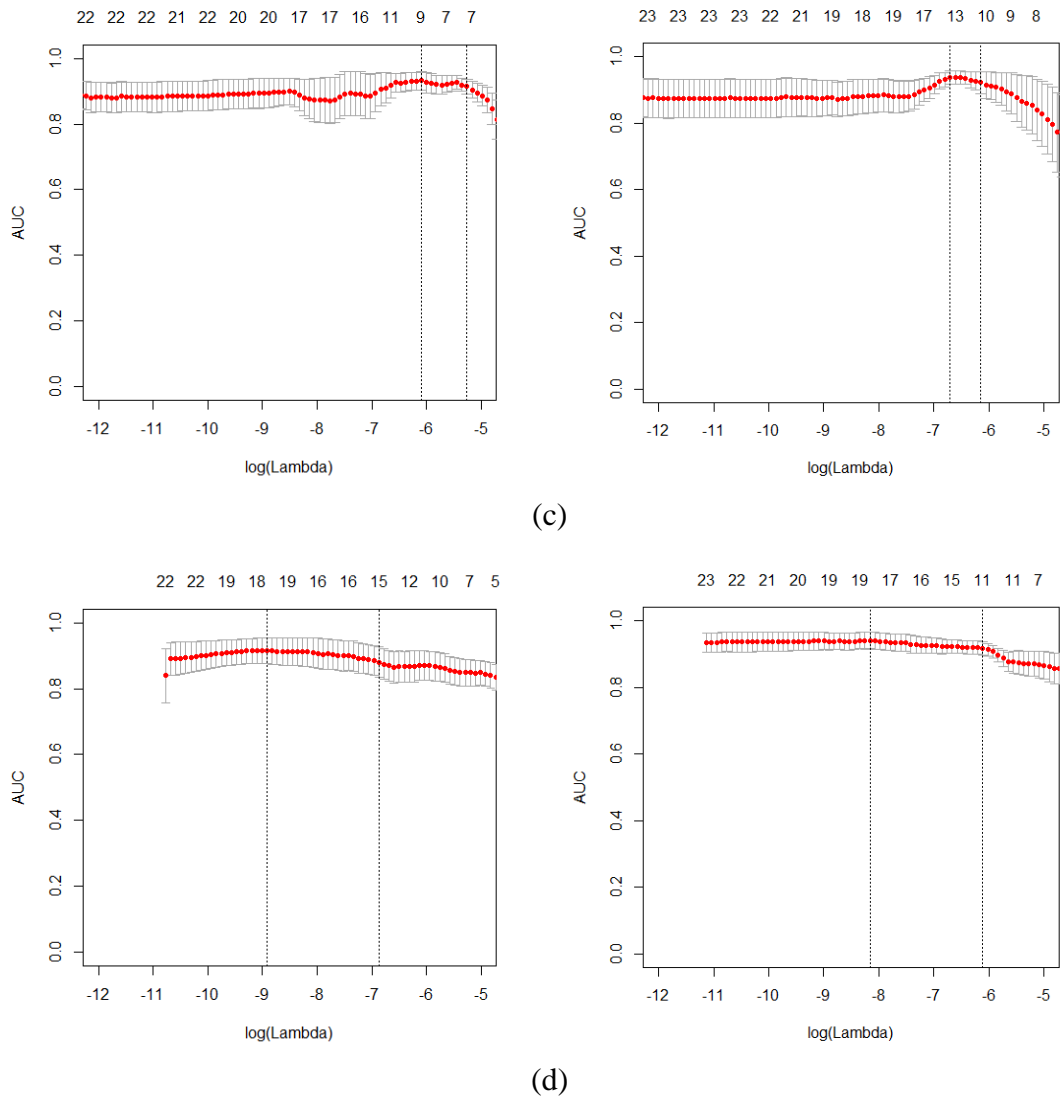


Figure 1b. AUC values for Logit Model with Lasso (left) and Elastic-Net (right) Regularization: (c) Singapore and (d) Thailand. The γ values for Elastic-Net Penalty Function are: (c) $\gamma = 0.8$ and (d) $\gamma = 0.9$.

Table 3. Estimates of the Selected Default Predictors. ID, MY, SG, and TH are for Indonesia, Malaysia, Singapore, and Thailand, respectively.

Variable	Lasso				Elastic-Net			
	ID	MY	SG	TH	ID	MY	SG	TH
	λ	0.0006	0.0003	0.0023	0.0001	0.0007	7.7E-5	0.0012
γ	1	1	1	1	0.4	0.9	0.8	0.9
Intercept	-6.2703	-3.8796	-15.0008	-3.4702	-9.8007	-6.0663	-19.4850	-5.9838
Profitability								
NI / TA	-9.5824	.	.	-17.9160	-10.9869	1.2116	.	-13.4577
NI / Sales	.	0.0232	.	3.3105	.	.	-0.0141	1.7085
OI / TA	.	-5.2335	.	-29.1399	.	-4.9878	.	-12.5319
OI / Sales	1.7700	0.1751	.	2.4896	1.0281	0.2085	.	0.0508
Leverage								
OF / TA	-1.5869	.	.	0.0335	-2.0711	3.2716	.	-0.0547
CL / TA	.	2.5814	.	14.3119	-0.9113	6.2273	-0.4687	9.2847
TD / TA	.	-0.8097	5.8997	-2.3711	0.5077	-0.5356	9.2795	-1.8411
Cost Structure								
INT / D	.	-0.2074	.	0.1703	.	-0.5359	-0.0402	0.1062
EBIT/INT-paid	.	-0.0016	.	.	.	-0.0063	.	.
Liquidity								
STD / D	-0.1601	-0.6132	0.8026	-0.5264	-2.3843	-0.8927	2.0618	-0.1005
Cash / TA	31.7064	-4.1047	.	-40.4293	29.1431	-1.7901	.	-27.3265
Cash / CL	.	-0.3944	.	.	.	-1.2372	.	.
QA / CL	.	-0.2195	0.0636	-5.4752	-0.3774	-0.2638	.	-4.0109
CA / CL	-2.4426	-1.0279	.	-5.0537	-1.3828	-1.2614	0.0882	-1.4935
WC / TA	.	5.2661	4.1490	19.0742	0.2118	6.2477	5.3230	9.4289
CL / TL	-8.4968	1.2489	.	-8.8536	-3.5709	.	0.0287	-6.7756
Activity								
TA / Sales	.	-0.0037	.	.	.	-0.0135	.	0.0018
INV / Sales	.	-0.0179	.	-0.2756	-0.0098	-0.0286	-0.0849	-0.0395
AR / Sales	2.1039	.	0.8242	2.6022	2.0607	0.0019	0.9038	2.0854
AP / C-Sales	-0.7922	0.0020	0.3989	0.8880	-0.9109	0.0059	1.1337	0.7847
Dynamics								
Sale-growth	.	0.0001	0.0077	0.0119	-1.9E-5	0.0001	0.0074	0.0108
NI-growth	-0.0139	-0.0005	.	.	-0.0148	-0.0012	.	.
Size								
Log (TA)	0.0288	0.5337	0.6617	0.6434	0.2237	0.8425	0.6812	0.5283
Log (Sales)	0.3108	-0.7818	0.4386	.	0.2273	-1.0874	0.7911	.
AUC	0.9941	0.8608	0.9696	0.9931	0.9937	0.8701	0.9784	0.9906
p^*	12	21	9	19	18	22	14	20

The inconsistency sign estimates are mostly caused by Singapore, i.e. five ratios: net profit margin (NI/Sales), average cost of debt (INT/D), short term debt (STD/D), quick ratio (QA/CL), and current ratio (CA/CL), followed by Indonesia with four ratios: return on assets ratio (NI/TA), cash to

total assets ratio (Cash/TA), account payable turnover (AP/C-Sales), and Sales growth. Malaysia and Thailand imply inconsistency sign estimate for each one ratio, i.e. Log (Sales) and assets turnover ratio (TA/Sales), respectively. The remaining four ratios are not clear from which country the inconsistency sign is affected: three from leverage group and current to total liability ratio (CL/TL).

The inconsistency sign estimate may depend on the quality of the data available. The quality can be distorted by means of creative accounting, i.e. accounting practices that follow required laws and regulations, but deviate from what those standards intend to accomplish. This practice is intended to falsely portray a better image of the company. Defaulting companies adjust their income upwards and provide a more positive financial report particularly when the default event is coming. However, when companies indulge in creative accounting they often distort the value of the information that their financials provide.

We now focus on the nine financial ratios with consistent sign estimates across countries. The financial ratios with negative sign are OI/TA, EBIT/INT-paid, Cash/CL, INV/Sales, and NI growth whereas OI/Sales, WC/TA, AR/Sales, and log (TA) have positive sign. Negative sign estimates imply the higher the corresponding financial ratio the lower PD. It is not a surprise that the higher operating income on asset ratio (OI/TA) will reduce the PD. The PD decrease for the higher interest coverage ratio (EBIT/INT-paid). An increase in cash ratio (Cash/CL) will reduce PD. Cash constitutes a substantial portion of current liabilities such that increase of the numerator is followed by a likewise increase of the denominator. Inventory turnover ratio (INV/Sales) alike has a counter intuitive sign. Keeping inventories as non-productive assets relatively high to sales should increase PD. This should provide positive sign coefficient. However, it is not totally true since too low inventory turnover (INV/Sales) lead to a higher PD represented by negative sign estimate. The relationship between inventory turnover and PD has a U-shaped therefore the magnitudes of the

estimates across countries are close to zero. The higher one year net income growth not surprisingly decreases the PD.

The financial ratio with positive sign estimates means the higher corresponding financial ratio the higher PD. It is a surprise that the higher operating profit margin (OI/Sales) ratio increase the PD. It should be in other way around. This estimates seems has a counter intuitive sign. The same problem happens for working capital (WC/TA) ratio. The higher account receivable turnover (AR/Sales) increases the PD. This ratio can be attributed to the inability of a company to collect payments from its clients on time and show a weak position of a company. It is a surprise that the larger the company the higher the PD since $\log(TA)$ have a positive sign estimate. The relationship between company size and PD may not monotone where the increasing assets of small and medium size company may does not decrease PD. There are non-accounting and qualitative failure indicators affect the default.

5. Conclusion

Regularization approach on logit model is able to simultaneously estimate and select default predictors with very high accuracy particularly for Indonesia, Singapore, and Thailand. For the same level of accuracy, the default predictors selected by Lasso on Indonesia and Singapore data are significantly smaller than are selected by elastic-net. Almost all ratios selected in the Lasso are also selected in the elastic-net. The relevant default predictors vary over the countries. This is in line with related studies which conclude that the default prediction analysis is sample specific. Nine out of twenty four financial ratios have consistent sign estimates where a few of them are counter intuitive.

References

- Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4): 589-609.
- Altman, E., Marco, G., and Varetto, F. (1994). Corporate Distress Diagnosis: Comparison using Linear Discriminant Analysis and Neural Network (the Italian Experience). *Journal of Banking and Finance*, 18: 505-529.
- Altman, E.I. and Hotchkiss, E. (2006). *Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt. 3rd Ed.*, John Wiley & Sons, Inc., Hoboken, New Jersey.
- Balcaen, S. and Ooghe, H. (2006). 35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and their Related Problems. *The British Accounting Review*, 38: 63-93.
- Beaver, W. (1966). Financial Ratios as Predictors of Failures. *Journal of Accounting Research. Empirical Research in Accounting: Selected Studies Supplement to vol. 4*: 71-111.
- Chen, S., Härdle, W., and Moro, R. (2011). Modeling Default Risk with Support Vector Machines. *Quantitative Finance*, 11: 135-154.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Path for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1): 1-22.
- Härdle, W., Lee, Y.-J., Schäfer, D., and Yeh, Y.-R. (2009). Variable Selection and Oversampling in the use of Smooth Support Vector Machine for Predicting the Default Risk of Companies. *Journal of Forecasting*, 28: 512-534.
- Härdle, W., Prastyo, D.D., and Hafner, C. (2013). Support Vector Machines with Evolutionary Model Selection for Default Prediction. In: Racine, J.S., Su, L., and Ullah, A. (Eds.), *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford University Press Inc., New York, NY.
- He, H. and Garcia, E.A. (2009). Learning from Imbalanced Data. *IEEE transaction on Knowledge and Data Engineering*, 21(9).
- Henley, W.E. and Hand, D.J. (1996). A k-Nearest Neighbour Classifier for Assessing Consumer Credit Risk. *Journal of the Statistical Society. Series D (The Statistician)*, 45(1): 77-95.
- Lau, A.H-L. (1987). A Five-State Financial Distress Prediction Model. *Journal of Accounting Research*, 25(1): 127-138.
- Lo, A.W. (1986). Logit versus Discriminant Analysis: A Specification Test and Application to Corporate Bankruptcies. *Journal Econometrics*, 31(2): 151-178.
- Martin, D. (1977). Early Warning of Bank Failure: A Logit Regression Approach. *Journal of Banking and Finance*, 1: 249-276.
- Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1): 109-131.
- Platt, H., Platt, M., and Pedersen, J. (1994). Bankruptcy Discrimination with Real Variables. *Journal of Business Finance and Accounting*, 21(4): 491-510.

- Rosset, S., Zhu, J., and Hastie, T., (2004). Boosting as a Regularized Path to a Maximum Margin Classifier. *Journal of Machine Learning Research*, 5: 941-973.
- Shevade, S.K. and Keerthi, S.S. (2003). A Simple and Efficient Algorithm for Gene Selection using Sparse Logistic Regression. *Bioinformatics*, 19: 2246-2253.
- Sobehart, J. and Keenan, S. (2001). Measuring default accurately. *Risk*, 14: 31-33.
- Steinwart, I. and Christmann, A. (2008). *Support Vector Machine*. Springer, New York, NY.
- Tam, K. and Kiang, M. (1992). Managerial Applications of Neural Networks: the Case of Bank Failure Predictions. *Management Science*, 38: 926-947.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58: 267-288.
- Vapnik, V. (1998). *Statistical Learning Theory*. John Wiley, New York, NY.
- Vassalou, M. and Xing, Y. (2004). Default Risk in Equity Returns. *The Journal of Finance*, 19(2): 831-868.
- Wilson, R.L. and Sharda, R. (1994). Bankruptcy Prediction using Neural Network. *Decision Support System*, 11: 545 - 557.
- Wu, T.T., Chen, Y.F., Hastie, T., Sobel, E., and Lange, K. (2009). Genome-wide Association Analysis by Lasso Penalized Logistic Regression. *Bioinformatics*, 25(6): 714-721.
- Wu, T.T. and Lange, K. (2008). Coordinate Descent Algorithms for Lasso Penalized Regression. *The Annals of Applied Statistics*, 2(1): 224-244.
- Zhang, J.L. and Härdle, W. (2010). The Bayesian Additive Classification Tree Applied to Credit Risk Modelling. *Computational Statistics and Data Analysis*, 54: 1197-1205.
- Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B*, 67(2): 301-320.

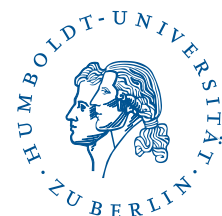
SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Functional Data Analysis of Generalized Quantile Regressions" by Mengmeng Guo, Lhan Zhou, Jianhua Z. Huang and Wolfgang Karl Härdle, January 2013.
- 002 "Statistical properties and stability of ratings in a subset of US firms" by Alexander B. Matthies, January 2013.
- 003 "Empirical Research on Corporate Credit-Ratings: A Literature Review" by Alexander B. Matthies, January 2013.
- 004 "Preference for Randomization: Empirical and Experimental Evidence" by Nadja Dwenger, Dorothea Kübler and Georg Weizsäcker, January 2013.
- 005 "Pricing Rainfall Derivatives at the CME" by Brenda López Cabrera, Martin Odening and Matthias Ritter, January 2013.
- 006 "Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing" by Markus Bibinger and Per A. Mykland, January 2013.
- 007 "Crossing Network versus Dealer Market: Unique Equilibrium in the Allocation of Order Flow" by Jutta Dönges, Frank Heinemann and Tijmen R. Daniëls, January 2013.
- 008 "Forecasting systemic impact in financial networks" by Nikolaus Hautsch, Julia Schaumburg and Melanie Schienle, January 2013.
- 009 "'I'll do it by myself as I knew it all along': On the failure of hindsight-biased principals to delegate optimally" by David Danz, Frank Hüber, Dorothea Kübler, Lydia Mechtenberg and Julia Schmid, January 2013.
- 010 "Composite Quantile Regression for the Single-Index Model" by Yan Fan, Wolfgang Karl Härdle, Weining Wang and Lixing Zhu, February 2013.
- 011 "The Real Consequences of Financial Stress" by Stefan Mittnik and Willi Semmler, February 2013.
- 012 "Are There Bubbles in the Sterling-dollar Exchange Rate? New Evidence from Sequential ADF Tests" by Timo Bettendorf and Wenjuan Chen, February 2013.
- 013 "A Transfer Mechanism for a Monetary Union" by Philipp Engler and Simon Voigts, March 2013.
- 014 "Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?" by Nikolaus Hautsch, Lada M. Kyj and Peter Malec, March 2013.
- 015 "Cyclical Variation in Labor Hours and Productivity Using the ATUS" by Michael C. Burda, Daniel S. Hamermesh and Jay Stewart, March 2013.
- 016 "Quantitative forward guidance and the predictability of monetary policy – A wavelet based jump detection approach –" by Lars Winkelmann, April 2013.
- 017 "Estimating the Quadratic Covariation Matrix from Noisy Observations: Local Method of Moments and Efficiency" by Markus Bibinger, Nikolaus Hautsch, Peter Malec and Markus Reiss, April 2013.
- 018 "Fair re-valuation of wine as an investment" by Fabian Y.R.P. Bocart and Christian M. Hafner, April 2013.
- 019 "The European Debt Crisis: How did we get into this mess? How can we get out of it?" by Michael C. Burda, April 2013.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 020 "Disaster Risk in a New Keynesian Model" by Maren Brede, April 2013.
- 021 "Econometrics of co-jumps in high-frequency data with noise" by Markus Bibinger and Lars Winkelmann, May 2013.
- 022 "Decomposing Risk in Dynamic Stochastic General Equilibrium" by Hong Lan and Alexander Meyer-Gohde, May 2013.
- 023 "Reference Dependent Preferences and the EPK Puzzle" by Maria Grith, Wolfgang Karl Härdle and Volker Krätschmer, May 2013.
- 024 "Pruning in Perturbation DSGE Models - Guidance from Nonlinear Moving Average Approximations" by Hong Lan and Alexander Meyer-Gohde, May 2013.
- 025 "The 'Celtic Crisis': Guarantees, transparency, and systemic liquidity risk" by Philipp König, Kartik Anand and Frank Heinemann, May 2013.
- 026 "State Price Densities implied from weather derivatives" by Wolfgang Karl Härdle, Brenda López-Cabrera and Huei-Wen Teng, May 2013.
- 027 "Bank Lending Relationships and the Use of Performance-Sensitive Debt" by Tim R. Adam and Daniel Streitz, May 2013.
- 028 "Analysis of Deviance in Generalized Partial Linear Models" by Wolfgang Karl Härdle and Li-Shan Huang, May 2013.
- 029 "Estimating the quadratic covariation of an asynchronously observed semimartingale with jumps" by Markus Bibinger and Mathias Vetter, May 2013.
- 030 "Can expert knowledge compensate for data scarcity in crop insurance pricing?" by Zhiwei Shen, Martin Odening and Ostap Okhrin, May 2013.
- 031 "Comparison of Methods for Constructing Joint Confidence Bands for Impulse Response Functions" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, May 2013.
- 032 "CDO Surfaces Dynamics" by Barbara Choroś-Tomczyk, Wolfgang Karl Härdle and Ostap Okhrin, July 2013.
- 033 "Estimation and Inference for Varying-coefficient Models with Nonstationary Regressors using Penalized Splines" by Haiqiang Chen, Ying Fang and Yingxing Li, July 2013.
- 034 "Robust Estimation and Inference for Threshold Models with Integrated Regressors" by Haiqiang Chen, July 2013.
- 035 "A new perspective on the economic valuation of informal care: The well-being approach revisited" by Konstantin Kehl and Stephan Stahlschmidt, July 2013.
- 036 "Herding in financial markets: Bridging the gap between theory and evidence" by Christopher Boortz, Simon Jurkatis, Stephanie Kremer and Dieter Nautz, July 2013.
- 037 "Default Risk Calculation based on Predictor Selection for the Southeast Asian Industry" by Wolfgang Karl Härdle and Dedi Dwi Prastyo, August 2013.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

