Recovery Rates across Different Industries

Markus Hoechstoetter[1] Abdolreza Nazemi[1]
Svetlozar T. Rachev[1,2,3]

[1]School of Economics and Business Engineering
Karlsruhe Institute of Technology, Germany

[2]Department of Applied Mathematics
Stony Brook University, USA

[3]FinAnalytica USA

Credit Scoring and Credit Control XII Conference, ED, UK, 24-26 August, 2011

Outline

Outline

Introduction

- Structure of paper

Outline

Basel II

Basel II is an international accord between banks to protect the international financial system.

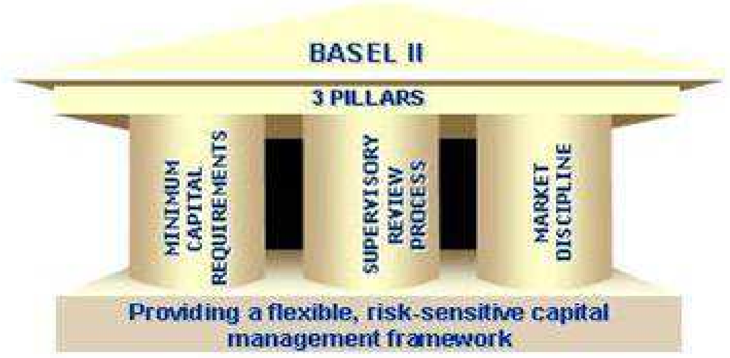
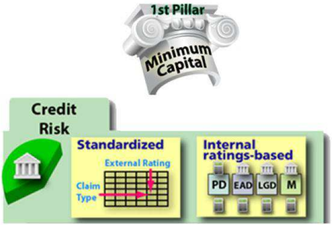


Figure: from http://www.bi.go.id

Credit risk parameters

Key components in Basel II

1. Probability of default (PD)
2. Loss given default (LGD)
3. Exposure at default (EAD)



Credit risk parameters

Internal Ratings Based (IRB) Approach

Foundation
Advanced

	PD	LGD	EAD
Foundation approach	Internal estimate	regulator estimate	regulator's estimate
Advanced approach	Internal estimate	Internal estimate	Internal estimate

But estimation might be difficult.

Credit risk parameters

Internal Ratings Based (IRB) Approach

Foundation
Advanced

	PD	LGD	EAD
Foundation approach	Internal estimate	regulator estimate	regulator's estimate
Advanced approach	Internal estimate	Internal estimate	Internal estimate

But estimation might be difficult.

Outline

Collection Types

Sequencing of recovery process

Recovery process in company
In-house
Third party

- An advantage of internal collection may be that all characteristics concerning the debt are known whereas a third party buyer is lacking important information such as loan details, borrower repayment behavior, or change in score which is a privilege of the original lender according to Fama (1985).

Outline

Data Provider

- Around ten million different unsecured debts
- Purchase between 2001 and 2010
- Arvato infoscore that is one of the largest debt purchasers in Germany
- This company combines collection business, scoring services and factoring
- Factoring

Factoring Type

Normal Factoring

The debt buyer receives all debt from the originator.
The factor is owner as well as collector of the debt after its cession from the originator.
It is the most common form of factoring in Germany.

Selective Factoring

Describes a construct where only selected debt is sold to the third party factor.

Notification Factoring

The debtor is informed about the sale of the debt and can only repay to third party factor.
The default risk remains with the originator.

Factoring Type

Silent Factoring

The debtor is ignorant of the sale of the debt and payment is only possible to the original creditor.

Semi-Factoring

The debtor remains ignorant of the sale of the debt, as well, but payments are to be made exclusively to accounts or addresses that belong to the factor.
In case of arvato infoscore, it is full-service factoring.

Outline

Data Description

- Roughly ten million defaulted or non-performing unsecured debts
- Nine different industries such as mail ordering (MO), business to business (B2B), financial services (FI), energy and utilities (NRGY), miscallaneous (MI), public sector (PS), return debit note (RDN), telecommunication (TC), public transport (PT).
- Each debtor is assigned a unique identification number.
- A payment is characterized by the identification number of the receivable and, thus, can be traced to the corresponding debtor.
- We selected from payment characteristics those that could be most easily be transformed.

Data Description

- Variables relating to the debtor include age, gender, residence status and address as well as current credit history. The variables related to the accounts receivables include age of debt, date of purchase by third party, amount outstanding, and last payment date while the original receivable amount is usually unknown. This yields about 15 variables that can be used for the subsequent analysis.
- The total number of receivables is 9,793,590.
- The amount of debt outstanding is 435,864,276.75 euros.
- we decided to use only 100,000 randomly selected receivables from each category or, if the total data set of the industry was less than 300,000, use the complete data set.

Data Description

	Industry (Category)								
	PT[1,a,b,id]	MO[1,id]	TC[1,id]	NRGY[2,id]	RDN[2,id]	B2B[2,id]	MI[2,id]	FS[2,id] Services	PT[2,id] Sector
# debts (original)	5,313,417	2,453,391	1,192,527	270,091	195,155	118,889	88,969	84,965	76,166
# debts (after step 1)	100,000	99,908	99,928	269,740	180,111	118,582	88,771	84,927	75,884
Total debt amount									
(in million euros)	4.98	12.94	21.07	92.91	14.34	24.77	56.64	184.38	23.84
Earliest entry	12-05-1990	08-29-2001	09-05-2001	04-13-1982	02-01-1997	01-04-2004	09-05-1991	02-26-1980	03-09-1994
Last entry	12-29-2010	12-21-2010	12-14-2010	12-28-2010	12-24-2010	12-01-2010	12-24-2010	12-01-2010	12-18-2010
Average debt amount									
(in euro)	49.79	129.51	210.89	344.44	79.61	208.85	638.03	2,171.07	314.11
Mean RR	26%	38%	35%	28%	32%	30%	15%	11%	10%
$RR = 1$									
absolute	22,015	28,154	29,394	46,164	47,677	29,709	8,949	6,317	4,516
percentage	22%	20%	29%	17%	26%	25%	10%	7%	6%
$RR = 0$									
absolute	73,284	60,285	61,255	182,814	121,049	80,415	71,165	72,891	64,397
percentage	73%	60%	61%	68%	67%	68%	80%	86%	85%
$RR \in (0, 1)$									
absolute	4,701	11,464	9,279	40,762	11,385	8,458	8,657	5,719	6,971
percentage	5%	11%	9%	15%	6%	7%	10%	7%	9%
Mean debt age									
(months)	23.07	31.74	35.42	29.30	28.26	28	16.67	59	35.74
(years)	1.92	2.65	2.95	2.44	2.36	2.33	1.39	4.92	2.98
Mean debt age (3rd party)									
(months)	20.68	26.91	30.92	22.56	26.19	23	6.47	16	12.29
(years)	1.72	2.24	2.58	1.88	2.18	1.92	0.54	1.33	1.02
# debts (after step 2)	100,000	72,589	73,258	152,613	88,776	43,054	33,060	76,556	42,006

Table 1: #: denotes count number. RR: denotes recovery rate. 1: subset consists of 100,000 arbitrarily drawn accounts receivables. 2: complete data used. a: age of debt unknown, b: age of debtor unknown, id: identification of debtor impossible.

Outline

As pointed out by Thomas et al. (2011)

Third Party

Seven percent repaid the whole debt, a little over sixteen percent repaid a fraction, and almost eighty-three repaid nothing.

In-house Collection

Thirty percent repaid the whole debt, sixty percent repaid a fraction and only ten percent paid nothing.

- They find that in-house collection yields a large point mass at full recovery, i.e. $RR = 1$ whereas for the third party, the recovery rate has almost mass one at $RR = 0$.
- Models for third party recovery tend to display poor fit with R2 between 8% and 22%.

Findings in the literature

- Calabrese (2010a) analyze 149,378 Italian bank loans.
  - the capitalized recovery amount significantly influences the subsequent recovery rate.
  - report a high concentration of recovery at zero and one.
- Grunert (2009) observe 120 German bank loans.
  - an uni-modal left-skewed distribution is deemed better than beta distribution.
  - average recovery rate of 72.5% and median of 91.8%.
  - the inclusion of macro variables does not improve model quality.
  - a negative correlation between recovery rate and the creditworthiness of borrower is apparent while EAD is a significant covariate in the regression of recovery rates.
- Livingstone and Lunt (1992)
  - sociodemographic factors play a relatively minor role in personal debt and debt repayment.
  - disposable income does not differ between those in debt and not in debt

Findings in the literature

- Loterman have five bank loan data sets.
  - for all popular models goodness-of-fit of 4% $<$ R square$<$ 43%.
  - SVM and non-linear neural networks have better predictive performance
- Zhang perform analysis on 27,278 UK personal bank loans.
  - An average recovery rate of 42%.
  - The most significant OLS regression variable is EAD.
  - Mixture models are not better than regular linear regression.
- Chen (2010) study 1880 individual residential foreclosed mortgages.
  - the property location is strongly correlated with social, demographic, economic factors and thus is relevant in the explanation of recovery.
- Qi (2009) have 241,293 US high-loan-to-value and insured mortgages.
  - 29.2% $<$ LGD $<$31.7%.
  - LGD and loan size, however, are negatively correlated while LGD and age of loan are positively correlated.

Outline

Models and Findings

- Since our data is from a non-bank third party buyer, we expect rather low recovery rates. From Table 1, we see that this is justifiable given that recovery rates are below 40% and even below 30%, in many cases.
- It is apparent that nearly all probability mass is at RR = 0 and RR = 1. Across all nine industries, the majority of the recoveries is equal to 0, by far.
- We also analyze the recovery rate distributions for the horizons of 12, 24, 36 months. Our conclusion is that, for all, nine industries, the variation in the respective distributions is minimal with mass slightly shifting from RR = 0 to RR = 1 since more debtor pay-o debt as time progresses.

Models and Findings

- We are modeling recovery rate by means of the well-known logistic regression model, i.e. the recovery rate RR is the non-linear transform of the linear model including real and coded categorical numerical data

$$RR = \frac{exp(\beta^{'} x + \varepsilon)}{1 + exp(\beta^{'} x + \varepsilon)}$$

where β is the vector of regression cofficients for the components of the data vector x and ε denotes the residual.

- Model

$$\mu + \alpha.amount + \beta.debtorage + \gamma.debtage + \delta.ageatsale + \ldots$$

$$\ldots + \rho_1.rating_1 + \rho_2.rating_2 + \rho_3.rating_3 + \rho_4.rating_4 + \rho_5.rating_5 + \ldots$$

$$\ldots + \rho_6.rating_6 + \rho_7.rating_7 + \phi.traceability + \theta_1.debtortype_1 + \theta_2 debtortype_2 + \varepsilon$$

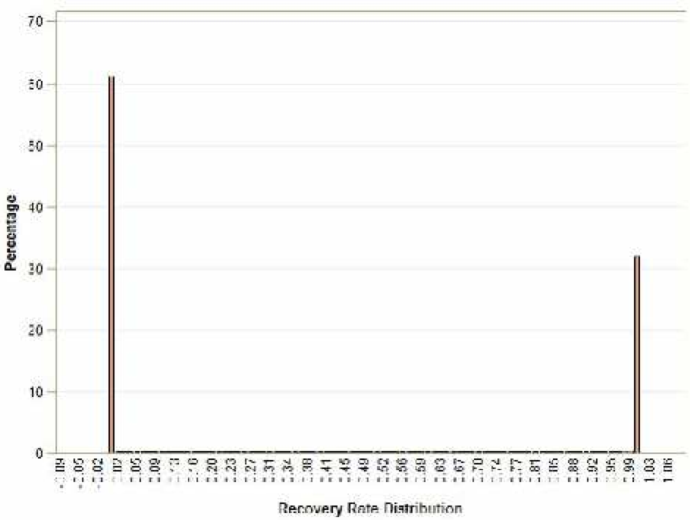
Models and Findings

Variable	symbol	MO		B2B		NRGY		FS		MI	
Intercept	μ	−3.1912	$0.0001^{a,b,c}$	−2.8891	$0.0001^{a,b,c}$	−2.6099	$0.0001^{a,b,c}$	−6.9403	0.0093	−3.8413	$0.0001^{a,b,c}$
debt amount	α	0.0004	$0.0001^{a,b,c}$	0.0003	$0.0001^{a,b,c}$	0.0001	$0.0001^{a,b,c}$	0.0000	0.0584^{c}	0.0001	$0.0072^{a,b,c}$
debtor age	β	0.3140	$0.0001^{a,b,c}$	−0.0691	$0.0152^{b,c}$	−0.0211	0.2962	−0.0226	0.4450	0.0517	$0.0344^{b,c}$
debt age	γ	−0.5335	0.2007	0.9023	0.2855	0.0612	0.6386	−0.0615	0.1118	0.9537	0.4534
age at sale	δ	0.2523	0.2282	−1.0869	0.1254	−0.0743	0.2108	0.3958	$0.0001^{a,b,c}$	−1.2026	0.3144
rating (original)	ρ_1	0.4462	$0.0001^{a,b,c}$	0.5653	$0.0001^{a,b,c}$	0.5020	$0.0001^{a,b,c}$	−9.3665	0.9787	0.7561	$0.0001^{a,b,c}$
	ρ_2	−0.2287	0.4884	0.2150	0.6439	0.3899	0.1813	3.6901	0.9517	0.4738	0.1374
	ρ_3	0.4863	$0.0001^{a,b,c}$	0.3238	$0.0204^{b,c}$	0.1804	0.1142	−9.2320	0.9705	0.3598	$0.0021^{a,b,c}$
	ρ_4	0.3833	0.0771^{c}	−0.0141	0.9651	−0.2133	0.4606	3.3601	0.9560	−0.3462	0.3861
	ρ_5	0.2084	$0.0355^{b,c}$	0.1296	0.3230	0.0694	0.4823	2.7865	0.9685	0.4417	$0.0006^{a,b,c}$
	ρ_6	−0.5507	$0.0026^{a,b,c}$	−0.4065	$0.0424^{b,c}$	−0.1538	0.2360	2.7604	0.9639	−0.6808	$0.0025^{a,b,c}$
traceable	ϕ	0.8585	$0.0001^{a,b,c}$	0.3896	$0.0001^{a,b,c}$	0.7513	$0.0001^{a,b,c}$	0.2470	0.3063	1.0865	$0.0001^{a,b,c}$
debtor type	θ_1	0.0862	0.5204	−0.0199	0.7911	0.2263	$0.0008^{a,b,c}$	0.2592	0.3241	0.3260	$0.0225^{b,c}$
	θ_2	0.3405	$0.0108^{b,c}$	0.1552	$0.0486^{b,c}$	0.3872	$0.0001^{a,b,c}$	−0.4104	0.5542	0.2680	0.0664^{c}
concordance (in %)		70.7		67.5		60.7		75.8		71.6	

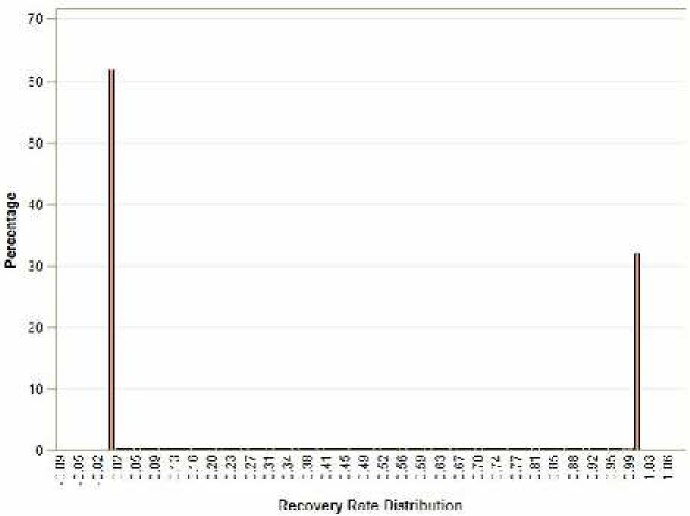
Variable	symbol	PS		RDN		TC		PT	
Intercept	μ	−5.7962	0.7616	−4.2613	$0.0001^{a,b,c}$	−3.5474	$0.0001^{a,b,c}$	−15.7940	$0.0001^{a,b,c}$
debt amount	α	0.0007	$0.0001^{a,b,c}$	0.0001	0.3382	0.0007	$0.0001^{a,b,c}$	−0.0456	0.1874
debtor age	β	0.0469	0.0667^{c}	0.0012	0.9662	−0.1561	$0.0001^{a,b,c}$	−2.4547	$0.0358^{b,c}$
debt age	γ	−6.3716	$0.0001^{a,b,c}$	4.5735	$0.0001^{a,b,c}$	−5.3874	$0.0001^{a,b,c}$	0.7217	$0.0174^{b,c}$
age at sale	δ	4.9486	$0.0001^{a,b,c}$	−0.3137	$0.0028^{a,b,c}$	1.4805	$0.0001^{a,b,c}$	2.8045	$0.0097^{a,b,c}$
rating (original)	ρ_1	2.3454	0.9023	0.4795	$0.0001^{a,b,c}$	0.4861	$0.0001^{a,b,c}$	−0.3147	0.6229
	ρ_2	−9.5470	0.9431	0.6695	0.0885^{c}	−0.1931	0.6139	0.2586	0.2868
	ρ_3	1.8007	0.9249	0.4516	$0.0019^{a,b,c}$	0.5607	$0.0001^{a,b,c}$	0.2908	0.5353
	ρ_4	1.0078	0.9579	−0.6059	0.2489	0.1647	0.5873	0.1109	0.6352
	ρ_5	1.3156	0.9451	0.1084	0.4897	0.3647	$0.0021^{a,b,c}$	−0.7059	$0.0365^{b,c}$
	ρ_6	1.0500	0.9562	−0.6381	0.0545^{c}	−0.7631	$0.0010^{a,b,c}$	−0.3503	$0.0251^{b,c}$
traceable	ϕ	1.6459	$0.0001^{a,b,c}$	0.9256	$0.0001^{a,b,c}$	0.7079	$0.0001^{a,b,c}$	11.7266	$0.0001^{a,b,c}$
debtor type	θ_1	0.1579	0.6204	0.5452	$0.0458^{b,c}$	0.3537	$0.0058^{a,b,c}$	11.7802	$0.0001^{a,b,c}$
	θ_2	0.4431	0.1659	0.7114	$0.0080^{a,b,c}$	0.4399	$0.0015^{a,b,c}$	−35.7288	$0.0001^{a,b,c}$
concordance (in %)		85.6		72.6		69.0		74.6	

Table 1: *a*: significant at the 0.01 level. *b*: significant at the 0.05 level. *c*: significant at the 0.10 level.

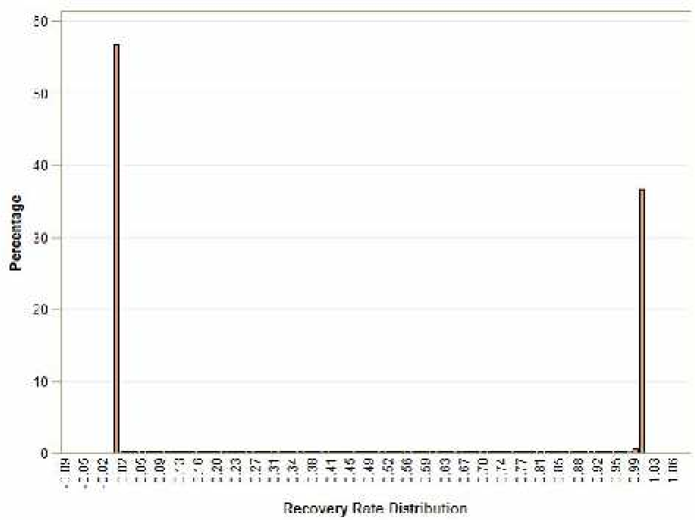
Models and Findings



Models and Findings



Models and Findings



Support Vector Machine

Data	SVM Accuracy Rate	Log. Reg. Accuracy Rate	R-squared
TC	67.12%	65%	0.1404
NRGY	64.54%	65%	0.1439
RDN	69.41%	68%	0.1889
B2B	66.43%	62%	0.1193
MI	64.71%	63%	0.141
FS	63.64%	69%	0.2235
PS	75.96%	75%	0.3875

Conclusion

- We presented the different possible actors in the collection process generally achieving different results in the collection and recovery of defaulted debt.
- it was a result of the different information available to the actors as well as quality of debt that they have access to.
- The distribution of repayment of debt was virtually similar across all industries even for various recovery horizons.
- From the lower levels of the average recovery rates in comparison to the literature on consumer bank loans, it became obvious that recovery achieved by a third party purchaser is less fruitful then when carried out by the original lender, especially in case of a bank.

Conclusion

- The resulting variables yielded no better goodness-of-fit than found in the literature.
- However, the estimated maximum likelihood model generated acceptable precision.
- In our next step, we will apply other statistical and data mining models to cope with the complexity of data and also consider including macro-economic variables.