CREDIT SCORING & CREDIT CONTROL XIV 26-28 August 2015 Edinburgh

The analysis of the micro enterprise failures using survival models

Aneta Ptak-Chmielewska
Warsaw School of Ecoomics
aptak@sgh.waw.pl

Content

- Background – literature
- Hypothesis
- Data and methods
- Empirical example
- Conclusions

Background

- Small and micro enterprises are the most frequent customers of medium size banks. The risk of failure of micro enterprises is difficult to assess. The main difficulty is caused by the lack of detailed and credible data.

- The assessment of credit risk is based on **scoring-rating model**. This kind of model consists of financial data of the enterprise and personal characteristics of the owner. In case of micro-enterprise the owner is closely related to the enterprise and sometimes cannot be separated.

- In this paper I propose the model for micro-enterprises based on financial ratios and personal characteristics. This kind of model is a hybrid model.

Background

- The estimation is based on a sample of 1000 micro-enterprises (questionnaire survey) and a sample of 900 small and micro enterprises (Financial Data – FS).
- I apply the logistic regression model and semiparametric Cox regression model with time varying variables.
- The basic questions are:
  - (1). What are the main drivers of micro-enterprises' failure (incl. financial ratios and personal characteristics)?
  - (2). What kind of abilities the entrepreneur should have to achieve success? I focus on the accumulation of human and social capital to analyse the risk of enterprises failure.
- Due to unique data those analyses are valuable and useful in further credit scoring-rating model development.

Literature

- E.I.Altman, G.Sabato „Modelling Credit Risk for SMEs: Evidence from the U.S. Market" ABACUS Vol. 43, No. 3, 2007, p. 332-357.
- B.Ibtissem, A.Bouri „Credit Risk Management in Microfinance: The Conceptual Framework" ACRN Journal of Finance and Risk Perspectives, Vol. 2. Issue 1, Nov. 2013, p.9-24.
- I. De Noni, A.Lorenzon, L.Orsi „Measuring and Managing Credit Risk in SMEs: a Quantitative and Qualitative rating Model, ...
- J.Belas, P.Bartos, R.Habanik, V.Hlawiczka „Determinants of credit risk of SMEs in the banking sector of the Czech Republik and Slovakia, ...
- J.Zhu, Z.Huang „Banks' Micro Enterprises Loan Credit Risk Decision-making Model Innovation in the Era of Big Data and Internet Finance, Journal of management and Strategy, Vol. 5, No. 2, 2014, p.63-69.
- A handbook for developing credit scoring systems in a microfinance context, USAID From the American People, 2006

Hypothesis

- Lack of information causes the need of **hybrid** model construction. Scoring-rating model is adequate to apply for micro enterprises.

- Personal characteristics of the owner (entrepreneur) determine the micro enterprise success (survival). As higher **human capital** of the owner as higher chances of survival.

Data and methods

- Two sources of data:
  - **Retrospective survey** on micro and small enterprises sample (1077, employment below 50 people) from one of the regions in Poland, registered in 2006, surveyed in 2011 – liquidation information

Total	Event – liquidation	Censored	% censored
1042	380	662	63.53

Data and methods

- Two sources of data:
  - **Financial information** from one of the Bank's portfolio of small enterprises with simplified accounting – default information
  - The sample consisted of 2,963 Financial Statements from years 2003 – 2014, number of defaulted enterprises 450.
  - **Simplified accounting** – only limited information available.
- Definition of small enterprise:
  - employment below 50 workers,
  - annual turnover below10 million Euro.

Data and methods

- Logistic regression
- Cox regression survival model

Logistic regression

- First type of estimated models is **logistic regression** model with maximum likelihood method.
- Probability is specified as:

$$P(Y=1)=1/(1+\exp\{-(\beta_0+\beta_1 x_1+\ldots+\beta_k x_k)\}),$$

where: $P(Y=1)$ – dependent variable, probability of default,
β_0 – intercept,
β_i for $i = 1, 2, \ldots, k$ – coefficients,
X_i for $i = 1, 2, \ldots, k$ – independent variables (financial ratios).

- $P(Y=1)$ takes value from $<0;1>$, where 0 means non-default, 1 means default.
- **Odds ratio** value – informs as higher/lower chances of default are where the value of independent variable changes about 1 unit.
- Logit model requires strict **assumptions** like: random sample with high number of observations, noncollinearity of explanatory variables and independent observations.

Cox regression model

- For the basic Cox model (without the time dependent variables) the hazard function is expressed by the equation (Allison 2010, p. 127):

$$h_i(t) = \lambda_0(t)\exp(\beta_1 x_{i1} + \ldots + \beta_k x_{ik})$$

- It means that the hazard ratio for the unit *i* in time *t* is the multiplication of two components:
- · nonnegative function $\lambda_0(t)$ left side undetermined,
- · exponential of the linear combination of the set *k* explanatory variables.
- The function $\lambda_0(t)$is so called baseline hazard which can be interpreted as the baseline hazard ratio for the unit for which all explanatory variables assume 0 value.

Cox regression model

- The Cox regression model is described as proportional hazard model because the hazard ratio for any unit remains in constant relations to the hazard ratio for any other unit. For any units i and j this condition can be written as the relationship:

$$\frac{h_i(t)}{h_j(t)} = \exp\left\{\beta_1\left(x_{i1} - x_{j1}\right) + \ldots + \beta_k\left(x_{ik} - x_{jk}\right)\right\}$$

- It means the hazard ratios for any two units will go paralleled in time.

Cox regression model

Including the time dependent variables causes that Cox model becomes the nonproportional hazards model.

However not fulfilling assumption about the proportionality of hazards by variables which are independent of time also causes that the model becomes the nonproportional hazards model.

It is important to verify the **assumption of the proportionality** of the hazards. It is not the only assumption of the model.

It must be remembered that equally important are assumptions to include all significant variables, noninformative censoring or about lack of measuremeant errors which cannot be ignored.

Empirical example (1)

- Financial information model – logistic regression model

V1	Current ratio	$\frac{\text{current assets}}{\text{short term liabilities}}$
V2	Quick ratio	$\frac{(\text{current assets} - \text{stocks})}{\text{short term liabilities}}$
V3	Total stock turvover	$\frac{\text{revenues} * \text{days}}{\text{stocks}}$
V4	Total stock turnover (2)	$\frac{\text{revenues}}{\text{stock}}$
V5	Total stock turnover (3)	$\frac{\text{costs} * \text{days}}{\text{stocks}}$
V6	Receivables collection	$\frac{\text{receivables} * \text{days}}{\text{revenues}}$
V7	Short-term liabilities to sales (in days)	$\frac{(\text{short term liabilities} - \text{short term loans and credits})}{\text{revenues} * \text{days}}$
V8	Equity to total assets	$\frac{\text{equity}}{\text{total assets}}$
V9	Equity to fixed assets	$\frac{\text{equity}}{\text{fixed assets}}$

Empirical example (1)

- Financial information model – logistic regression model

V10	Net profit on sales	$\frac{\text{profit on sales}}{\text{revenues on sales}}$
V11	Gross profit on sales	$\frac{\text{gross profit}}{\text{revenues on sales}}$
V12	Gross profit to equity	$\frac{\text{gross profit}}{\text{equity}}$
V13	Coverage and servicing the debt	$\frac{(\text{net profit} + \text{depreciation} + \text{interest on credits and loans})}{\text{repayment of loans and credits} + \text{interest}}$
V14	Financial leverage	$\frac{\text{interest on credits and loans}}{\text{revenues on sales}}$
V15	Debt servicing capability	$\frac{(\text{gross profit} + \text{interest on credit and loans})}{\text{interest on credit and loans}}$
V16	Financial costs to revenues	$\frac{\text{interest on credit and loans}}{\text{revenues}}$
V17	Cash flow	$\frac{\text{funds (cash)}}{\text{shor term liabilities}}$

Empirical example (1)

- Financial information model – logistic regression model

V18	Share of short term capital in assets	$\frac{(\text{short term assets} - \text{short term liabilities})}{\text{total assets}}$
V19	Share of working capital in assets	$\frac{(\text{total assets} - \text{long term liabilities} - \text{short term liabilities})}{\text{total assets}}$
V20	Debt ratio	$\frac{\text{total liabilities}}{\text{total assets}}$
V21	Log of fixed assets	$\log(\text{fixed assets})$
V22	Revenues to liabilities	$\frac{\text{revenues}}{\text{total liabilities}}$
V23	Share of financial surplus in total liabilities	$\frac{(\text{net profit} + \text{depreciation} + \text{interest})}{\text{total liabilities}}$
V24	Coverage of short term liabilities to equity	$\frac{\text{equity}}{\text{short term liabilities}}$
V25	Share of equity in liabilities	$\frac{\text{equity}}{\text{total liabilities}}$

Empirical example (1)

- Financial information model – logistic regression model

V26	Receivables turnover (1)	$\frac{\text{total revenues}}{\text{receivables}}$
V27	Receivables turnover (2)	$\frac{\text{total revenues} * \text{days}}{\text{receivables}}$
V28	Assets turnover (1)	$\frac{\text{total revenues}}{\text{total assets}}$
V29	Assets turnover (2)	$\frac{\text{total revenues} * \text{days}}{\text{total assets}}$
V30	Net profitability of equity	$\frac{\text{net profit}}{\text{equity}}$
V31	Operational profit to assets	$\frac{\text{operational profit}}{\text{total assets}}$
V32	Operational profitability of sales	$\frac{\text{operational profit}}{\text{total revenues}}$
V33	Net profit to stocks	$\frac{\text{net profit}}{\text{stocks}}$

Empirical example (1)

- Financial information model – logistic regression model

V34	Stocks to revenues on sales	$\frac{\text{stocks}}{\text{revenues on sales}}$
V35	Stocks to revenues on sales	$\frac{\text{stocks} * \text{days}}{\text{revenues on sales}}$
V36	Current assets to short term liabilities	$\frac{(\text{current assets} - \text{stocks})}{\text{short term liabilities}}$
V37	Receivables to revenues on sales	$\frac{\text{receivables}}{\text{revenues on sales}}$
V38	Equity ratio and long term liabilities to assets	$\frac{(\text{equity} + \text{long term liabilities})}{\text{total assets}}$

Empirical example (1)

- The sample was constructed as balanced sample: all 450 defaults and randomly selected 450 financial statements for good customers.
- Only noncorrelated ratios were selected all together 25 financial ratios.
- Ratios were discretized due to avoid nonlinearities. Discretization was based on WOE (Weight of Evidence).

- Stepwise selection limited the number of ratios in a final model to 6 ratios, p-value for each of them was below $\alpha = 0{,}05$:
- V2 – Quick ratio
- V16 – Financial costs to revenues
- V17 – Cash flow
- V22 – Revenues to liabilities
- V 25 – Share of equity in liabilities
- V34 – Stocks to revenues on sales

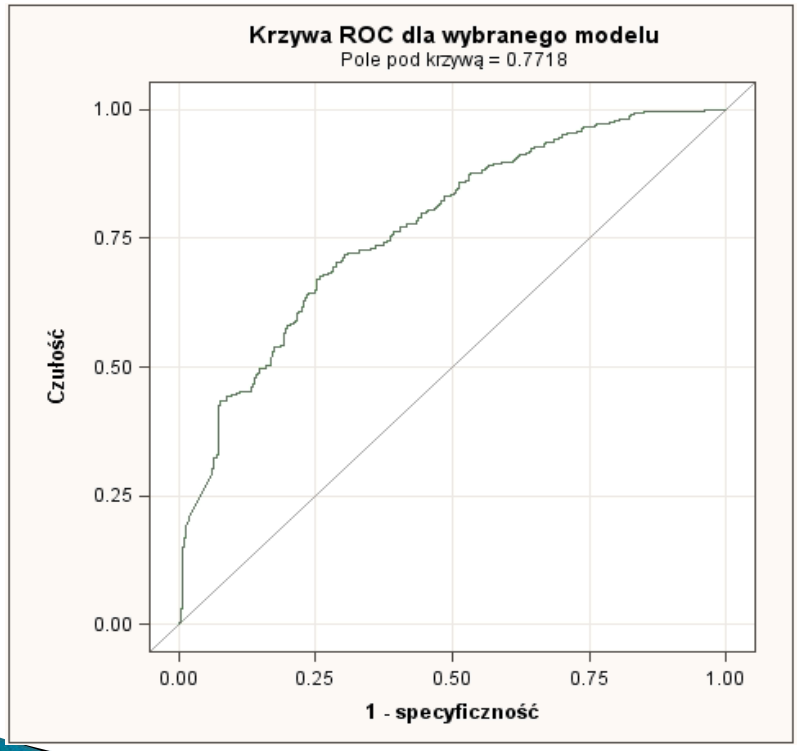
Empirical example (1)

Parameter		estiamtion	Stnd error	Wald chi-sqr	p-value
Intercept		0.0225	0.0765	0.0864	0.7688
V2 – Quick ratio		-0.4682	0.1474	10.0950	0.0015
V16 – Financial costs to revenues		-0.4291	0.1691	6.4369	0.0112
V17 – Cash flow		-0.3730	0.1741	4.5887	0.0322
V22 – Revenues to liabilities		-0.7861	0.1120	49.2832	<.0001
V 25 – Share of equity in liabilities		-0.6596	0.1540	18.3493	<.0001
V34 – Stocks to revenues on sales		-0.6277	0.1585	15.6756	<.0001

	Odds ratio	Wald CL 95%	
V2 – Quick ratio	0.626	0.469	0.836
V16 – Financial costs to revenues	0.651	0.467	0.907
V17 – Cash flow	0.689	0.490	0.969
V22 – Revenues to liabilities	0.456	0.366	0.567
V 25 – Share of equity in liabilities	0.517	0.382	0.699
V34 – Stocks to revenues on sales	0.534	0.391	0.728

Empirical example (1)

- ROC =77.2% – good classification accuracy (**AR=54,4%**).



Model Empirical	default	non-default	total
default	312	138	450
non-default	314	136	
total	626	274	900

Assuming p=0.5
All correct 69,6%
Sensitivity 69.3%
Specificity 69.8%
1-specificity 30.2%

Empirical example (1)

- Hosmer & Lemenshow test

decil	total	default = 1		default = 0	
		observed	expected	observed	expected
1	90	8	7.55	82	82.45
2	90	22	20.53	68	69.47
3	90	29	28.32	61	61.68
4	90	41	35.06	49	54.94
5	91	33	42.00	58	49.00
6	90	53	48.64	37	41.36
7	90	56	56.46	34	33.54
8	90	60	65.52	30	24.48
9	90	65	70.72	25	19.28
10	89	83	75.20	6	13.80

Test Hosmer and Lemeshow		
Chi-sqr	DF	p-value
15.3558	8	0.0526

Empirical example (2)

- Qualitative information model – Cox survival model

Legal form	0. company 1. natural person
Employment	0. 10-49 workers 1. below 10 workers
Sector of activity	0. sector – low risk group 1. Sector – high risk group
Change in sector of activity	0. yes; 1.no
Market of activity	0. national or international market 1. local or regional market
Change in market of activity	0. yes; 1.no
Sources for start	0. yes; 1.no
Export of goods and services	0. yes; 1.no
Profit in a first year	0. profit; 1. loss or activity suspended/no data
Sex of the owner	0. male or company 1. female
Education of the owner	0. higher and postgraduate, comapny 1. lower or no data
Age of the owner	0. 25 years and older, company 1. below 25 years, no data
Type of previous job	0. low risk group of professions, company 1. high risk group, no data
Fixed assets investments in a first year	0. yes; 1. no
Barriers in sales of goods and services	0. no barriers reported 1. barriers reported, suspended activity, no data

Empirical example (2)

- Qualitative information model – Cox survival model

	parameter	p-value	HR	CL HR 95%	
Legal form	-0.45109	0.0402	0.637	0.414	0.980
Employment	-0.82555	0.0366	0.438	0.202	0.950
Sector of activity	-1.28000	0.0051	0.278	0.113	0.682
Change in sector of activity	-0.03092	0.7806	0.970	0.780	1.205
Market of activity	-0.46864	0.0005	0.626	0.481	0.814
Change in market of activity	-0.86360	0.0923	0.422	0.154	1.152
Sources for start	-0.80663	0.0091	0.446	0.244	0.818
Export of goods and services	-0.41892	0.2127	0.658	0.340	1.271
Profit in a first year	-0.76453	<.0001	0.466	0.378	0.573
Sex of the owner	-0.18706	0.0907	0.829	0.668	1.030
Education of the owner	-0.12546	0.2815	0.882	0.702	1.108
Age of the owner	-0.50139	<.0001	0.606	0.473	0.776
Type of previous job	-0.23612	0.0347	0.790	0.634	0.983
Fixed assets investment in a first year	-0.45418	<.0001	0.635	0.516	0.781
Barriers in sales of goods and services	-0.72841	<.0001	0.483	0.390	0.597

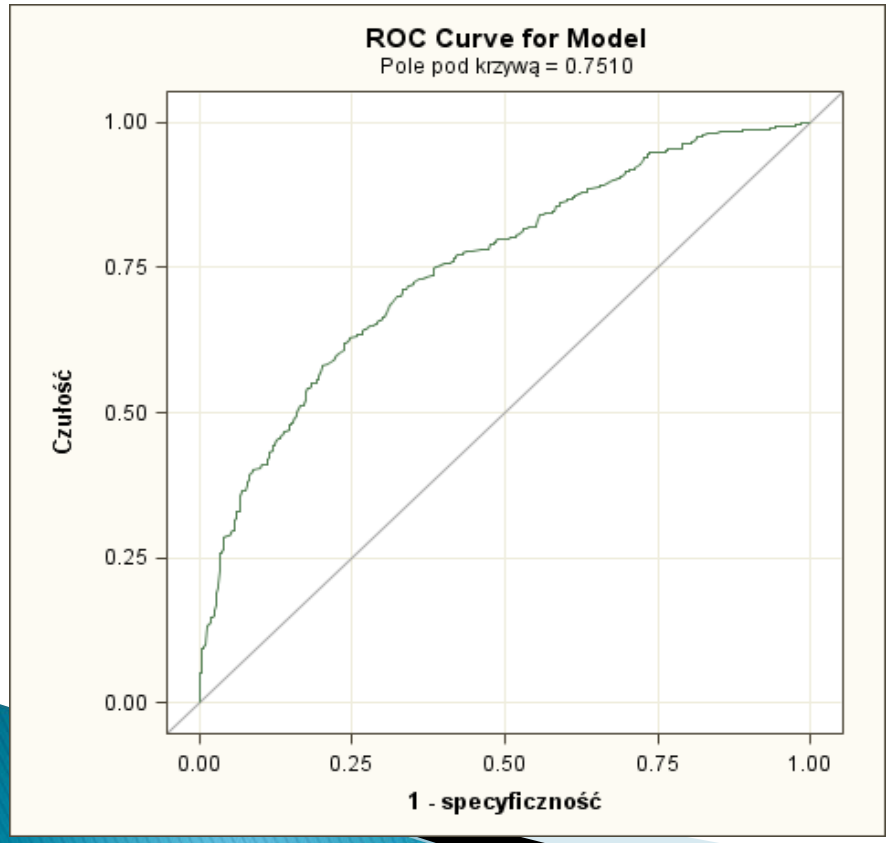
Empirical example (2)

- Qualitative information model – Cox survival model – interactions – PH assumption

	parameter	Std err	Chi-sqr	P-value
Legal form	0.0002836	0.00474	0.0036	0.9523
Employment	-0.00694	0.00770	0.8123	0.3674
Sector of activity	-0.02226	0.00816	7.4352	0.0064
Change in sector of activity	-0.00753	0.00229	10.8345	0.0010
Market of activity	-0.01174	0.00282	17.3190	<.0001
Change in market of activity	-0.00172	0.00898	0.0366	0.8483
Sources for start	-0.01228	0.00582	4.4526	0.0348
Export of goods and services	-0.00610	0.00703	0.7520	0.3858
Profit in a first year	-0.02194	0.00197	123.6913	<.0001
Sex of the owner	-0.02558	0.00212	145.9247	<.0001
Education of the owner	-0.01328	0.00230	33.2228	<.0001
Age of the owner	-0.03297	0.00227	210.5470	<.0001
Type of previous job	-0.00734	0.00215	11.6551	0.0006
Fixed assets investment in a first year	-0.01041	0.00195	28.4543	<.0001
Barriers in sales of goods and services	-0.01534	0.00203	57.3115	<.0001

Empirical example (2)

- Qualitative information model – Cox survival model – ROC
- Value of the predictor as Score in univariate Logistic model
- **AR=60,2.**



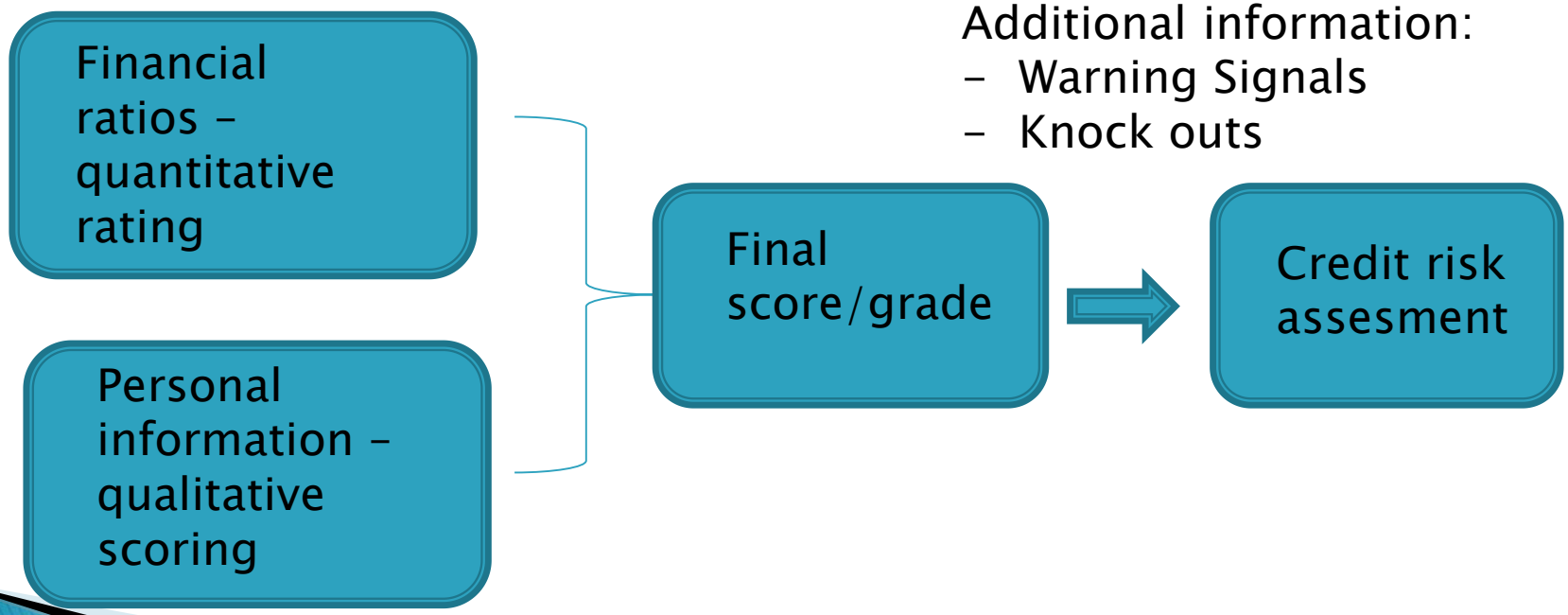
Model Empirical	liquidation	active	total
liquidation	180	200	380
active	103	559	662
total	283	759	1042

assuming p=0.5

Total correct 70.9%
Sensitivity 47.2%
Specificity 84.4%
1- specificity 15.6%

Final model construction

- Two parts: financial part (rating model) and qualitative part (scoring model)
- Combination of both models – hybrid model



Conclusions

- Discrimination between good and bad clients in case of micro enterprises requires two types of information: financial (quantitative) and personal (qualitative).
- Only hybrid model: combination of both parts is sufficient in credit risk assessment
- Financial ratios- very limited financial information
- Qualitative part – human capital of the owner increases the chances of survival.
- Both hypothesis were verified positively.