Iwona Markowicz*, Beata Stolorz**

APPLICATION OF LOGISTIC REGRESSION FOR FIRMS SURVIVAL ANALYSIS

Abstract. The 90's of the 20th century in Poland it's a period of economic development. Freedom of establishing and carrying on a business caused rising many new, mostly small firms. Yet not all entrepreneurs were able to run their businesses for a longer period of time. The purpose of the paper is to analysis firms survival chance in a determined period of time and to analysis differences between firms established in different years. Logistic regression model for dichotomous dependent variable will be used.
Key words: logistic regression, odds ratio, encoding, firms.

I. LOGISTIC REGRESSION

Model of logistic regression has been used for the first time for modelling the population growth but presently is being used in many fields of science, e.g. in demography, medicine, epidemiology, psychology and also in economy. The logistic function is described as follows [Cramer J.S. (2002)]:

$$P(Z) = \frac{\exp Z}{1 + \exp Z} \qquad (1)$$

and has values from 0 to 1. It can therefore express probability described as a chance or hazard of occurrence of a specific event.

For dichotomous dependent variable the logistic regression can be stated:

* D.Ph., Department of Econometrics and Statistics, University of Szczecin.
** D.Ph., Department of Econometrics and Statistics, University of Szczecin.

$$P(Y = 1 \mid x_1, x_2, ..., x_k) = \frac{\exp\left(\alpha_0 + \sum_{i=1}^{k} \alpha_i x_i\right)}{1 + \exp\left(\alpha_0 + \sum_{i=1}^{k} \alpha_i x_i\right)}. \tag{2}$$

where: Y – dichotomous dependent variable, $x_1, x_2, ..., x_k$ - independent variables, α – regression factor.

In the paper logistic regression model for dichotomous variable will be used, but in the literature one can find logistic models used for categorical data [e.g. Kozyra (2005)]. Instead of the least squares method the maximum likelihood method is being used for estimation of the logistic regression parameters when considering not fulfilling of assumptions. The entry $p = P(Y = 1)$ describes probability of success (or defeat), the expression $\frac{p}{1-p}$ describes chance (or hazard) of occurrence of a specific event. To avoid situation where probability would not contain within the interval [0; 1], the logistic regression is used instead of the linear regression. Expression $\ln\left(\frac{p}{1-p}\right)$ is written down as logit(p) and is used in the logit model notation (converted linear equation):

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \alpha_0 + \sum_{i=1}^{k} \alpha_i x_i. \tag{3}$$

The larger value of regression factor α_i, the larger is influence of explanatory variable for probability of success (or defeat). For interpretation however odds ratio (or risk ratio) e^{α_i} is used. Because:

$$\ln\left(\frac{p}{1-p}\right) = \alpha_0 + \sum_{i=1}^{k} \alpha_i x_i \Rightarrow \frac{p}{1-p} = \exp\left(\alpha_0 + \sum_{i=1}^{k} \alpha_i x_i\right) = e^{\alpha_0} \cdot e^{\sum_{i=1}^{k} \alpha_i x_i} \tag{4}$$

then odds ratio (OR) or risk ratio (RR) following the unitary change of the explanatory variables x_i (remaining explanatory variables are unchanged) are as follows:

$$\frac{P(x_{j+1})}{1 - P(x_{j+1})} \div \frac{P(x_j)}{1 - P(x_j)} = \frac{e^{\alpha_0} \cdot e^{\alpha_i x_{j+1}}}{e^{\alpha_0} \cdot e^{\alpha_i x_j}} = e^{\alpha_i (x_{j+1} - x_j)} = e^{\alpha_i \cdot 1} = e^{\alpha_i} \tag{5}$$

Odds ratio (OR) or risk ratio(RR) following the unitary change of the explanatory variable is called unit odds ratio (OR_u) or unit risk ratio (RR_u). Odds ratio (risk ratio) can also be calculated for changes different than unit. An example can be range odds ratio stated when estimating logistic regression parameters in the *Statistica* programme. It can be noted as follows:

$$OR(RR) = e^{\alpha_i(x_{j+q}-x_j)} = e^{\alpha_i \cdot q} = OR_u^{\ q} \tag{6}$$

II. STATISTICAL DATA AND THEIR ENCODING

Statistical data bought in the Statistical Office in Szczecin have been used in the analysis. Data come from the REGON register and concern economic entities established in Szczecin from the 1990 to the 1999. Observation has been completed on the 31st of December, 2000. The event being the subject of study is unregistering of the firm and it has been encoded as 1. Firms that until the end of observation carried on business activity were encoded as zero. Structure of studied units presents table 1.

Table 1. Structure of studied firms following year of establishment and occurrence of the event liquidation of firm (1) or not occurrence of the (0)

Year of establishment	Liquidated firms (1)	Not liquidated firms (0)	Together
1990	1641	2758	4399
1991	1925	3907	5832
1992	2237	5000	7237
1993	2559	3434	5993
1994	2007	3014	5021
1995	2717	4145	6862
1996	2574	4507	7081
1997	2291	5568	7859
1998	1614	6201	7815
1999	622	5372	5994
Sum	20187	43906	64093

Source: own study.

For the chance of firm survival study two approaches were used: calculation of survival odds ratio until the 31st of December, 2000 and calculation of odds ratio of survival in the first year of activity, and two ways of variables encoding: 0-1-2-3-4-5-6-7-8-9 (respectively for years 1990–1999) and encoding 0-1 that enables to exchange polychotomous variables to dichotomous variables [Hosmer, Lemeshow (2000)].

III. FIRM SURVIVAL CHANCE UNTIL THE 31ST OF DECEMBER, 2000

Estimating the logistic regression model of dependence of firm survival on polychotomous variable – year of establishment, odds ratio has been determined (table 2). This odds ratio gives information that following the establishment of the firm one year later, its chance to survive increases 10,716% on the average. This is thus mean measures and it does not indicate differences in time. It enables also determination of odds ratio comparing specified year of firm establishment to the 1990 (table 3). The last result is so called range quotient.

Table 2. Results of logistic regression estimation of survival of firm in relation to its year of establishment for firms established in Szczecin within years 1990-1999, observed until the 31.12.2000 (dependent variable-dichotomous, independent variable-polychotomous)

N=64093, χ^2=1105,310, p=0,000000					
Independent variable	Estimator of regression factor	Standard error	p	Odds ratio	Confidence interval 95%
Year of establishment	0,101798	0,003178	0,000000	1,107160	(1,100282-1,114082)

Source: own study - *STATISTICA PL* programme.

Table 3. Odds ratio – comparison with 1990

Compared years of firm establishment	Unitary change of odds ratio	Compared years of firm establishment	Unitary change of odds ratio
90	1,000000	95/90	1,663613
91/90	1,107160	96/90	1,841886
92/90	1,225804	97/90	2,039262
93/90	1,357161	98/90	2,257790
94/90	1,502594	99/90	2,499735

Source: own study - *STATISTICA PL* programme.

More precise results can be obtained exchanging polychotomous explanatory variable to dichotomous variable through application of 0-1 encoding. Year 1990 has been as basic and increase or decrease of survival chances for firms established in the consecutive analysed years is studied in relation to this particular year (table 4). Adaptation of both encoding methods of explanatory variables has been presented on fig. 1 and 2.

Table 4. Results of logistic regression estimation of survival of firm in relation to its year of establishment for firms established in Szczecin within years 1990-1999, observed until the 31st of December, 2000 (variables: dependent and independent- dichotomous)

N=64093, χ^2=2827,771, p=0,000000					
Independent variable (year of establishment)	Estimator of regression factor	Standard error	p	Odds ratio	Confidence interval 95%
1990				1,000000	
1991	0.107907	0.036725	0.003303	1.113944	(1.036545–1.197123)
1992	0.228803	0.036516	0.000000	1.257094	(1.170228–1.350409)
1993	−0.287974	0.036841	0.000000	0.749781	(0.697526–0.805950)
1994	−0.159910	0.044101	0.000288	0.852221	(0.781619–0.929200)
1995	−0.146233	0.036527	0.000063	0.863956	(0.804239–0.928107)
1996	0.039274	0.017346	0.023576	1.040055	(1.005273–1.076041)
1997	0.278432	0.034544	0.000000	1.321057	(1.234536–1.413642)
1998	0.788960	0.037781	0.000000	2.201106	(2.043930–2.370368)
1999	1.589048	0.049219	0.000000	4.899084	(4.448353–5.395485)

Source: own study - *STATISTICA PL* programme.

Odds ratio of survival of firm until the 31st of December,
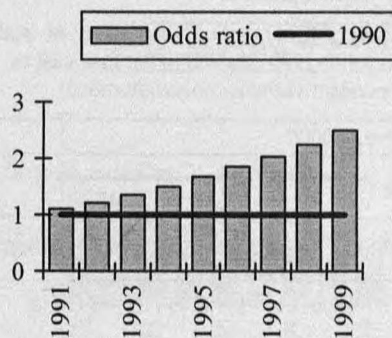

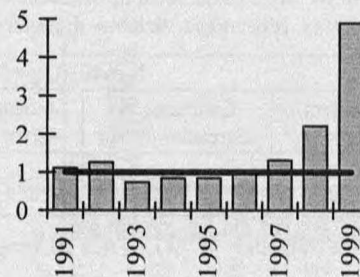Fig. 1. Odds ratio – comparison to 1990 – polychotomous independent variables.

Fig. 2. Odds ratio – comparison to 1990 – dichotomous independent variables.

Figure 1 presents the first approach towards independent variable encoding and indicates systematic increase of firm survival chance following its later establishment. Increase of chances is obvious because survival chance has been studied until the end of observation, i.e. the 31st of December, 2000. Therefore firms established later carried on their activity a shorter time. However, when 0-1 encoding was applied, increase of the chances was not so systematic, even survival chances of firms established in the years 1993, 1994 and 1995 were less than these of firms established in the 1990.

IV. FIRM SURVIVAL CHANCE OF FIRST YEAR OF ACTIVITY

In order to eliminate the influence of time of business activity, odds ratio of firm survival of first year of activity (the worst one, as appears from numerous studies) has been determined. Two encoding methods of independent variable have been applied. Results of logistic regression estimation for polychotomous variable have been presented in the table 5. Determined ratio indicates decrease of survival chance of first year activity 10,8% on average following the delay of firm establishment for one year. Odds ratio of firms established in consecutive years in relation to firms established in 1990 has been presented in the table 6. In the table 7 results of logistic regression estimation with 0-1 encoding for independent variable have been presented. Application of both encoding methods of explanatory variables has been presented on figures 3 and 4.

The first approach toward independent variable encoding indicates systematic decrease of firms survival chance following their later establishment in relation to firms established in 1990 (odds ratio equal to 1). This increase can be explained by increase of competition, therefore more difficult condition of survival on the market.

Table 5. Results of logistic regression estimation of survival of firm in relation to its year of establishment for firms established in Szczecin within years 1990-1999, studied in the first year of activity (dependent variable- dichotomous, independent variable- polychotomous)

N=64093, χ^2=559,518, p=0,000000					
Independent variable	Estimator of regression factor	Standard error	p	Odds ratio	Confidence interval 95%
Year of establishment	−0.114217	0.005098	0.000000	0.892064	(0.883191–0.901026)

Source: own study - *STATISTICA PL* programme.

Table 6. Odds ratio – comparison with 1990.

Compared years of firm establishment	Unitary change of odds ratio	Compared years of firm establishment	Unitary change of odds ratio
90	1,000000	95/90	0,564912
91/90	0,892064	96/90	0,503938
92/90	0,795779	97/90	0,449545
93/90	0,709886	98/90	0,401023
94/90	0,633264	99/90	0,357738

Source: own study - *STATISTICA PL* programme.

Table 7. Results of logistic regression estimation of survival of firm in relation to its year of establishment for firms established in Szczecin within years 1990-1999, studied in the first year of activity (variables: dependent and independent- dichotomous)

N=64093. χ^2=1578.559. p=0.000000					
Independent variable (year of establishment)	Estimator of regression factor	Standard error	p	Odds ratio	Confidence interval 95%
1990				1.000000	
1991	0.501997	0.071502	0.000000	1.652016	(1.435894–1.900667)
1992	1.312838	0.087957	0.000000	3.716705	(3.127904–4.416343)
1993	0.082393	0.065283	0.206924	1.085883	(0.955405–1.23418)
1994	−0.326924	0.057735	0.000000	0.721138	(0.643949–0.807581)
1995	−0.616046	0.051186	0.000000	0.540076	(0.488500–0.597097)
1996	−0.776945	0.049055	0.000000	0.459809	(0.417640–0.506236)
1997	−0.769304	0.048422	0.000000	0.463336	(0.421365–0.509486)
1998	−0.512709	0.050205	0.000000	0.598871	(0.542723–0.660828)
1999	−0.008404	0.003629	0.020568	0.991631	(0.984601–0.998713)

Source: own study - *STATISTICA PL* programme.

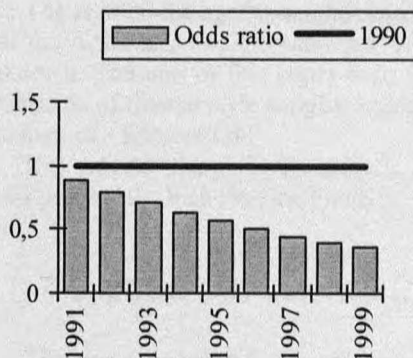Odds ratio of survival of firm at first year of activity

Fig. 3. Odds ratio – comparison to 1990 – polychotomous independent variables
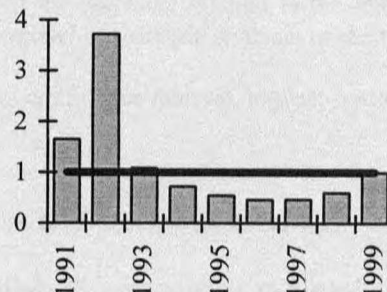
Fig. 4. Odds ratio – comparison to 1990 – dichotomous independent variables

However, following 0-1 encoding, decrease of chances was not so systematic, even survival chances of firms established in 1991, 1992, 1993 were larger than these of forms established in the 1990, i.e. in the beginning of enterprise development in Poland. In succeeding years these chances were smaller, but in the 1999 they were almost equal with those of firms established in the 1990.

Concluding, one can state that determination of firm survival odds ratio based on the logistic regression parameters gives interesting results. These results are dependent in the high degree on the independent variables encoding methods.

REFERENCES

Cramer J.S. (2002), *The Origins of Logistic Regression*, Tinbergen Institute Discussion
    Paper, Faculty of Economics and Econometrics, University of Amsterdam.
Kozyra C. (2005), Zastosowanie regresji logistycznej do analizy danych wielomiano-
    wych dotyczących samooceny zdrowia, *Taksonomia 12*, Prace Naukowe AE we
    Wrocławiu nr 1076, 319–329.
Hosmer D.W., Lemeshow S. (2000), *Applied Logistic Regression*, John Wiley & Sons, Inc.

Iwona Markowicz, Beata Stolorz

ZASTOSOWANIE REGRESJI LOGISTYCZNEJ DO ANALIZY SZANSY PRZETRWANIA FIRM

Lata 90-te XX w. w Polsce to okres rozwoju przedsiębiorczości. Swoboda zakłada-
nia i prowadzenia działalności gospodarczej przyczyniła się do powstawania nowych,
głównie małych, firm. Nie wszystkim przedsiębiorcom udaje się jednak utrzymać dzia-
łalność w dłuższym czasie. Celem artykułu jest analiza szansy przetrwania firmy przez
określony czas oraz różnic dla firm założonych w różnych latach. Zastosowano model
regresji logistycznej dla dychotomicznej zmiennej zależnej. W artykule przedstawiono
wyniki etapu badań prowadzonych w ramach projektu badawczego MNiSW N 111 011
31/1109.