Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v15n1p40

Identifying changes in predictors of business failures during and after the economic crisis in Italy

By Pierri, Caroni

Published: 20 May 2022

Electronic Journal of Applied Statistical Analysis
Vol. 15, Issue 01, May 2022, 40-49
DOI: 10.1285/i20705948v15n1p40

Identifying changes in predictors of business failures during and after the economic crisis in Italy

Francesca Pierri*[a] and Chrys Caroni[b]

[a] *University of Perugia, Department of Economics, Via Pascoli 20, 06123 Perugia, Italy*
[b] *National Technical University of Athens, Department of Mathematics, Zografou Campus, Athens 15780, Greece*

Published: 20 May 2022

Following the prolonged economic crisis of recent years, a new economic shake-up due to the COVID-19 pandemic is under way. We consider whether banks and financial institutions may apply the same models as before for credit scoring and predicting risk. In particular, we investigate the prediction of survival or failure of Small Business Enterprises in Italy between 2008 and 2013, and between 2013 and 2018, using logistic regression models based on baseline balance sheet data. By fitting appropriate models including interaction with the time period, we identify several major differences. Notably, the Investment Rigidity Ratio was very strongly associated with failure probability in the first period but not the second, and a low Tangible Assets Ratio had a much stronger protective effect in the first period than in the second. The effect of the age of the firm also differed between the periods: younger firms were at greater risk of failure than older firms in 2008-2013 but this was not seen in 2013-2018. Especially in times of major changes, it is vital that quantitative aids to decision-making should be valid and up-to-date.

keywords: business failures, prediction, logistic regression, interaction

1 Introduction

There is a very extensive literature on the important practical issue of predicting the failure of business enterprises using quantitative modelling. Various methodological approaches to this and the related topic of credit scoring have been proposed from the

*Corresponding author: francesca.pierri@unipg.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

fields of statistics and artificial intelligence. A recent review of the subject can be found in Bisogno et al. (2018). The main statistical models in use are logistic regression, introduced into this area by Ohlson (1980), and survival analysis techniques such as the Cox regression model (Cox, 1972), following Lane et al. (1986). The latter approach allows consideration of the timing of the failure event, whereas the former examines whether or not it occurs within a fixed interval of time. Pierri and Caroni (2017) applied both a Cox regression model with time-varying covariates and a multi-period logistic regression to balance sheet indicators of Small Business Enterprises and compared their predictive capabilities for bankruptcy (or closure of the firm from any other cause). In this case, due to the structure of the data (D'Agostino et al., 1990), logistic regression performed slightly better than the survival analysis model.

It is obvious that quantitative analysis of business failure continues to be a vital tool during the major economic crisis that began in about 2008 and has now, before complete recovery, been reinforced during 2020 both by the direct effects of the Covid-19 pandemic and by the countermeasures that governments have imposed. During these years, the financial and economic environment has changed, and firms that have managed to survive must in many cases have adapted their mode of operation substantially. Models for predicting business failure have to be fitted and calibrated using retrospective data that record, for each firm in a set of data, the event of failure or survival in the study period together with potential predictors of failure. Therefore, it is interesting to know, especially in this time of substantial change, whether the predictors that worked with these historical data, continue to do their job.

The purpose of the present research is therefore the comparison of models of failure of Small Business Enterprises in two different periods in order to understand if we can apply the same model, that is, if the same factors emerge as the best predictors in different periods. It is possible that the most relevant factors for prediction changed as the crisis progressed. Furthermore, was there a change in our ability to predict failure (with or without changes in the predictors)? It may be that the increased uncertainty of the business climate has decreased the predictive capability of our quantitative modelling. We carry out the analysis using data on enterprises in the same area of Italy, over two periods of equal duration (2008-2013 and 2013-2018), fitting logistic regression models with predictors from the base year (2008 or 2013) to predict failure or survival up to the end of the corresponding period (2013 or 2018, respectively).

2 Data and Methods

The present study analyses data from the balance sheets of capital companies in the Umbria Region in Italy combined with information from the Business Register on whether or not they are currently functioning. Similar data were analysed with a different objective in Caroni and Pierri (2020). We obtained both the present sets of data by applying to the Chamber of Commerce of Perugia. More specifically we used the 2008 balance sheets for all the firms that were operational in January 2008, and established their status (operational or not) at the end of December 2013. Similarly, we built a second database with balance sheets of firms still operating in January 2013 and established their status at the end of December 2018. Thus we obtained two databases each referring to firms' survival, over two periods of equal duration. The potential explanatory variables for failure (through formal bankruptcy or in any other way) or survival in both periods were financial indexes constructed from company balance sheet data, as listed in the Appendix. These variables were also employed in earlier work (Pierri and Caroni, 2017; Caroni and Pierri, 2020).

In the first stage of the analysis, we built and tested two separate logistic regression models, labelled M1 and M2, obtaining results shown in the following section. The first (M1) employed 2008 data to predict status at 2013. The second model (M2) employed 2013 data to predict status at 2018. We obtained these models by successively eliminating non-significant covariates (starting from the model containing all the potential covariates) until the final model that minimised the value of Akaike's Information Criterion was reached. This criterion is defined as $-2\ell + 2k$, where ℓ is the value of the maximised log-likelihood for the model currently under consideration and k is the number of parameters requiring estimation in that model. The second term of the criterion thus constitutes a penalty working against including an unnecessarily large number of covariates in the model. Statistical significance of an individual covariate was assessed by examining the change in deviance upon omitting it from the current model; that is, minus twice the difference between the maximised values of the log-likelihood for the nested models (with and without that covariate) was referred to the chi-squared distribution with one degree of freedom.

Comparison of the significant covariates identified in M1 and M2 suggested that the relevant covariates for prediction were different between the two periods. However, this could be a mistaken impression arising as a result of the variable selection procedures that were applied in building the models. Therefore, in a second stage, we sought to confirm that different models were required for each period by a technically preferable method. To do this, we first combined the two data sets into one, adding a new binary variable which took the value of 0 or 1 if the data pertained to the first (2008-2013) or second (2013-2018) period, respectively. Another logistic model was built for these data (model M3). Next, we verified the possible differences between the covariates' influence within each period by adding and testing interactions between period and each of the significant covariates from the earlier models M1 and M2. For further interpretation of the significant interactions, we refitted the model with significant interactions using an alternative parameterisation of the interactions, as follows.

Let P represent the binary variable denoting period and X another covariate with which it interacts. The automated construction in logistic regression software gives the interaction variable PX which is the arithmetical product of P and X. Although the associated statistical test (e.g. likelihood ratio) for adding PX to the model which already includes P and X among the predictors confirms whether or not an interaction exists between P and X, the resulting parameter estimates for interaction terms expressed in this way are not readily interpretable in logistic regression (Norton et al., 2004). Instead, recognizing that the meaning of interaction is that the effect of X differs between periods

$P = 0$ and $P = 1$, we define two new variables

$$X\_0 = \begin{cases} X, & \text{if } P = 0, \\ 0, & \text{if } P = 1 \end{cases}$$

$$X\_1 = \begin{cases} 0, & \text{if } P = 0, \\ X, & \text{if } P = 1 \end{cases}$$

which show directly the effects of X in the two periods separately. Then we fit the same regression model, but with X_0 and X_1 in place of X and PX (and with P still included). This was model M4. From the coefficients of X_0 and X_1, it is possible to identify different patterns of interaction, as seen in the results below.

Models were fitted in SPSS (Version 26) using the Generalized Linear Models and Logistic Regression procedures, and evaluated using the ROC Curve procedure to calculate the area under the curve.

3 Results

During the period 2008-2013, 24% of the total of 10787 selected firms experienced a failure event, compared to the lower failure rate of almost 10% of 10880 firms in the subsequent period. Events by year within each period are shown in Table 1. The annual number of events became noticeably lower from 2015 onwards.

Table 1: Number of firms active at the start of 2008 or 2013 that ceased operation in the following years.

	Year	0	1	2	3	4	5	Total
2008	N	162	421	445	565	511	511	2615
(N=10787)	%	1.5	3.9	4.1	5.2	4.7	4.7	24.2
2013	N	229	451	111	49	76	109	1025
(N=10880)	%	2.1	4.1	1.0	0.5	0.7	1.0	9.4

The significant covariates identified in M1 and M2 are shown in Table 2. In M1, a higher Investment Rigidity Ratio (OR 3.1, 95% CI 2.3-4.2), higher Liquidity Ratio (OR 1.5, 95% CI 1.3-1.8) and lower Return on Assets (OR 0.28, 95% CI 0.22-0.34) all strongly increased the probability of failure, but did not emerge as significant predictors in M2. Younger age and geographical location in Terni also significantly increased the failure probability in M1 but not in M2. Type of firm was also significant in M1, with lower failure probability for limited liability companies, but not in M2. In contrast, a higher value of the Current Debt Ratio strongly increased failure probability in M2 (OR 1.7, 95% CI 1.5-1.9) but was not significant in M1. The two models had three predictors in common; each of these acted in the same direction in both models, but with different coefficients. Return on Turnover and Turnover both had stronger effects in M2 than in M1, whereas the effect associated with the Tangible Asset Ratio effect was stronger in M1. Higher Turnover, lower Return on Turnover and lower Tangible Assets Ratio all increased the failure probability. Both M1 and M2 had only moderate predictive ability. The area under the ROC curve (AUC) was equal to 0.653 (95% CI 0.639-0.666) for M1 and 0.655 (95% CI 0.634-0.677) for M2.

Table 2: Odds ratios and 95% confidence intervals of significant covariates in logistic regression, for 2008-2013 using balance sheet data at 2008 (Model 1), for 2013-2018 using balance sheet data at 2013 (Model 2), and for both periods combined (Model 3).

	Model 1	Model 2	Model 3
Period 2013-2018*	N/A	N/A	0.30 (0.27-0.33)
Age (years)	0.977 (0.971-0.982)		0.984 (0.980-0.989)
Type of Firm**			
Cooperative	0.94 (0.69-1.30)		1.10 (0.84-1.45)
Limited Liability	0.68 (0.52-0.89)		0.80 (0.63-1.01)
Location***			
Terni	1.12 (1.01-1.26)		1.12 (1.02-1.23)
Current Debt Ratio		1.67 (1.50-1.86)	1.46 (1.36-1.56)
Investment Rigidity Ratio	3.10 (2.27-4.24)		2.05 (1.61-2.61)
Liquidity Ratio	1.51 (1.28-1.79)		
Return on Assets	0.28 (0.22-0.34)		
Return on Turnover	0.82 (0.71-0.95)	0.59 (0.50-0.70)	0.70 (0.63-0.79)
Tangible Assets Ratio	0.14 (0.09-0.21)	0.51 (0.38-0.68)	0.24 (0.17-0.32)
Turnover	1.17 (1.01-1.35)	1.54 (1.29-1.83)	1.41 (1.26-1.57)

*Reference category: 2008-2013

**Reference category: Publicly traded

***Reference category: Perugia

N/A: not applicable

In model M3 (also shown in Table 2) the new binary variable referring to period (with the first period as reference) had an odds ratio less than one (0.30, 95% CI 0.27-0.33), because the probability of closure for any cause was lower in the second period compared

Table 3: Significant covariates in logistic regression for both periods combined (Model 4). Odds ratios and 95% confidence intervals are shown separately for 2008-2013 and 2013-2018 when the covariate has a significant interaction with period, otherwise a single value is shown applying to both periods.

	2008-2013	Both periods (no interaction)	2013-2018
Period 2013-2018*		0.23 (0.19-0.27)	
Age (years)	0.977 (0.971-0.982)		0.998 (0.991-1.005)
Type of Firm**			
Cooperative		1.07 (0.81-1.40)	
Limited Liability		0.79 (0.62-1.01)	
Location***			
Terni		1.11 (1.01-1.22)	
Current Debt Ratio	1.32 (1.21-1.45)		1.66 (1.49-1.85)
Investment Rigidity Ratio	3.14 (2.30-4.28)		1.06 (0.67-1.66)
Liquidity Ratio		1.38 (1.19-1.61)	
Return on Assets	0.29 (0.23-0.36)		0.38 (0.30-0.47)
Return on Turnover	0.75 (0.67-0.85)		0.59 (0.49-0.70)
Tangible Assets Ratio	0.13 (0.09-0.20)		0.55 (0.33-0.90)
Turnover		1.28 (1.14-1.43)	

*Reference category: 2008-2013

**Reference category: Publicly traded

***Reference category: Perugia

to the first as seen in Table 1. Significant indexes in M1 and M2 were still in M3 with the exception of Liquidity Ratio and Return on Assets. Their absence may be due to their overall effects being weakened due to their absence from M2. The AUC value of M3 was 0.706 (95% CI 0.696-0.716); the increase compared to M1 and M2 must be due to the significant effect of the indicator variable for Period.

Interactions with Period were investigated by first adding to the model all the possible two-way interactions between Period and the other covariates that had been identified in either M1 or M2. The value of Akaike's Information Criterion (AIC) fell from 15574.1 to 15243.0. Removing the interactions of both Location and Type of Firm with Period reduced this further to 15242.2. In addition, removing the interaction with Turnover gave an AIC value of 15243.5, effectively the same as the more complex model that included all the interactions. Finally, the interaction with Liquidity Ratio could possibly be removed, because the change in deviance was 3.965 (P = 0.046) and its removal increased AIC by only 1.97. Results are shown here with this interaction excluded. The AUC of this model was 0.733 (95% CI 0.723-0.742), representing an appreciable improvement over models without interaction.

After thus identifying the significant interactions with Period, the model (M4) was refitted with the alternative parameterisation described above in order to clarify each covariate's impact during different periods. Estimated odds ratios are shown in Table 3. In particular, they confirm that a high Investment Rigidity Ratio greatly increased the failure probability in the first period, but had little or no effect in the second. Similarly, younger firms had a higher failure probability than older ones, but only in the first period. A high Tangible Assets Ratio was strongly associated with lower risk of failure in both periods, but much more in 2008-13 than in 2013-2018. The lack of interaction suggests that the type of firm, its location and the Liquidity Ratio did not have different effects in the two periods, in contrast to the impression given by the simple comparison of models M1 and M2.

4 Discussion

We have compared survival of Small Business Enterprises in one Region of Italy in two different periods: during the depths of the economic crisis in 2008-13, and during the slow economic upturn of 2013-18. In both periods, five-year survival or failure was predicted from balance sheet data at the start of the corresponding interval. Thus the fitted models were static logistic regression models, which have well known limitations compared to the use of dynamic models: see, for example, the superior predictions obtained by Pierri and Caroni (2017) as a result of introducing time-varying covariates and their lagged values. However, many forecasting models in this field, such as credit scoring models, are necessarily of this form, obtained by analyses similar to ours fitted to historical data.

We were particularly interested to learn the answers to two questions: whether the predictability of failure differed between the two periods and whether the covariates that were relevant for prediction changed. The importance of the covariates was identified by appropriate testing of interactions in a single model fitted to the combined data of the two periods. This is superior to a simple comparison "by eye" of the coefficients of the covariates between the two models, and avoids accidental differences that may have appeared more as a result of variable selection procedures than because of genuine differences between periods. Furthermore, appropriate representation of the interaction terms was used in order to obtain results that had the familiar odds ratio interpretation. The answer to our first question was "no": the rather moderate predictive values of our models M1 and M2 for the two periods were equal. The answer to the second question was "yes": examination of interaction terms showed that several covariates had different effects in the two period, some strikingly so. Especially, a high Investment Rigidity Ratio ran a high risk of failure in 2008-13 but not in 2013-18. This may indicate the high risk to a firm's continued existence that a lack of flexibility presents during an acute crisis. Younger firms were at greater risk of failure than older firms in 2008-13 but not in 2013-18. The median age of firms in the database was 5.9 years in the first period but 7.9 in the second. This may suggest that few new firms were founded in the crisis and that the most vulnerable young firms had already failed before the second period. The other marked difference between the two periods was that a low Tangible Assets Ratio had a stronger protective effect in 2008-13 than in 2013-18, indicating the greater relative importance of tangible assets during the main period of crisis. Other significant interaction terms were associated with less dramatic differences between the two periods.

The last dozen years have seen a major shake-up regarding economic activity. The business and financial landscape will change even more as the slow recovery up to 2019 was followed by the major blow of the COVID-19 pandemic in 2020. It is important that banks and financial institutions should use up-to-date and valid quantitative analyses to support their decision-making and be aware of possible modifications that their models may require.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4:71–111.

Bisogno, M., Restaino, M., and Di Carlo, A. (2018). Forecasting and preventing bankruptcy: A conceptual review. *African Journal of Business Management*, 12(9):231–242.

Caroni, C. and Pierri, F. (2020). Different causes of closure of Small Business Enterprises: Alternative models for competing risk survival analysis. *Electronic Journal of Applied Statistical Analysis*, 13(1):211–228.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

D'Agostino, R. B., Lee, M.-L. T., Belanger, A. J., Cuples, L. A., Anderson, K., and Kannel, W. B. (1990). Relation of pooled logistic regression to time dependent Cox regression analysis: The Framingham Heart Study. *Statistics in Medicine*, 9(12):1501–1515.

Lane, W. R., Looney, S. W., and Wansley, J. W. (1986). An application of the Cox proportional hazards model to bank failure. *Journal of Banking and Finance*, 10(4):511–531.

Norton, E. C., Wang, H., and Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal*, 4(2):154–167.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1):109–131.

Pierri, F. and Caroni, C. (2017). Bankruptcy prediction by survival models based on current and lagged values of time-varying financial data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 3(3-4):62–70.

Storey, D. J., Keasey, K., Watson, R., and Wynarczyk, P. (2016). *The performance of small firms: profits, jobs and failures.* Routledge.

5 Appendix: Potential predictors of failure

Using balance statements for each firm and year, we built the economic and financial ratios listed below, as potential predictors of a business's failure. These are variables that are generally used in the literature on bankruptcy prediction analysis in order to analyze the business productivity, and the firm's ability to repay debts and to deal with medium and long term financial obligations (Beaver, 1966; Altman, 1968; Storey et al., 2016). In addition, the firm's Activity Sector (six categories), Legal Form (three categories) and Geographical Location (two categories) were represented by indicator variables.

Current Ratio
Quick Ratio
Leverage
Investment Rigidity Ratio
Tangible Assets Ratio
Intangible Assets Ratio
Financial Fixed Assets Ratio
Investment Elasticity Ratio
Inventories Impact Ratio
Liquidity Ratio
Short Term Liquidity Ratio
Long Term Liquidity Ratio
Debt Ratio
Permanent Debt Ratio
Current Debt Ratio
Equity Ratio
Return on Assets
Return on Equity
Return on Turnover
Return on Sales
Turnover