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/i20705948v13n1p211

Different causes of closure of small business enterprises: alternative models for competing risks survival analysis

By Caroni, Pierri

Published: 02 May 2020

Electronic Journal of Applied Statistical Analysis
Vol. 13, Issue 01, May 2020, 211-228
DOI: 10.1285/i20705948v13n1p211

Different causes of closure of small business enterprises: alternative models for competing risks survival analysis

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

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

Published: 02 May 2020

We examine the time until closure of Small Business Enterprises in Umbria, Italy between 2008 and 2013, and the factors that influence it. Earlier analysis, using Cox regression, considered “failure” (closure) from any cause. However, there are different reasons for inactivity: voluntary winding-up (1808 of 15184 firms in our data, 59.3% of the 3049 failures); bankruptcy (236, 7.7%); and closure without action by creditors or courts (1005, 33.0%). While the earlier analysis provides a valuable overall picture, it is also interesting to examine the separate causes, their rates of occurrence and which factors influence them separately. We do this using competing risks analyses, employing both of the regression methods that are prominent in the literature, based on cause-specific and sub-distribution hazard functions (Fine-Gray model). Furthermore, a proportional odds model was used to estimate cumulative incidences of failure by cause. Data included the firm’s year of foundation, location, legal form and sector of activity. Financial indexes were constructed from annual balance sheets. The date and reason for closure were recorded if the firm ceased activity. Findings included major differences between types of firm; for example, cooperatives had greatly increased hazards for winding-up (HR of 2.44 and 2.61 in the two approaches) but greatly reduced hazards for closure (0.48 and 0.45) compared to publicly traded companies. All-causes analysis averaged these strong effects into an insignificant one (1.05). Coefficients from the proportional odds model were similar to those from the Fine-Gray model, but have the advantage of interpretability.

*Corresponding author: ccar@math.ntua.gr

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

keywords: bankruptcy, survival analysis, competing risks, cause-specific hazards, sub-distribution hazards, proportional odds.

1 Introduction

There is a substantial literature on statistical modelling to predict the failure of business enterprises (Balcaen and Ooghe, 2006). Starting with Ohlson (1980), logistic regression with the occurrence of the failure event as dependent variable became established as the predominant method. Subsequently, models that had originally been developed in other fields under the name of survival analysis were introduced for predicting business failure or default (Narain, 1992; Banasik et al., 1999; Shumway, 2001). Survival models examine not only whether or not the event occurs, but also the time elapsed until it happens. These models have been developed very extensively in medical statistics (for example: Therneau and Grambsch, 2013; Hosmer et al., 2008; Collett, 2015) and also, under the name of reliability analysis, in engineering and technological fields.

A previous paper has presented an analysis of the time until closure of Small Business Enterprises in a region of Italy, and of the factors that influence it, using the Cox regression method of survival analysis with time-varying covariates (Pierri and Caroni, 2017). That study examined the occurrence of the event of the firm's "failure" (closure), irrespective of its cause.

However, different routes to an enterprise's inactivity exist: voluntary winding-up, dissolution or liquidation; bankruptcy; and closure without action by creditors or the courts. Balcaen et al. (2012) discussed in detail the need to analyse the determinants of different types of exit from the economic system. However, their study was based on a sample of distressed firms, so that exits could only be analysed comparatively between modes of exit (using a nested logit model), and not in relation to healthy firms as is done in a survival analysis.

In the terminology of survival analysis, different causes of failure are "competing risks": it is as if the various processes that can lead to failure are in a race with each other to be the first to cause failure. Although the earlier all-causes analysis is important in itself, because it provides an overall picture of the life-course of firms, it is also interesting to examine the separate causes, the rates at which they operate and the factors that influence them (which may not be the same for every cause). In the present paper, we carry out such an analysis of the separate causes, using competing risks methodology from survival analysis. The application is to Small Business Enterprises in the Umbria Region of Italy over the period 2008-2013.

There have been several previous competing risks analyses in the economics and financial literature: recent examples include Bhattacharjee et al. (2009); Chancharat et al. (2010); Kwon and Hahn (2010); Sohn and Jeon (2010) and Amendola et al. (2015). One approach - the cause-specific hazards analysis which we describe in the following section - predominates. However, another important methodology in the analysis of competing risks, using sub-distribution hazards (which we also describe below), does not seem to have been adopted in this literature although it is familiar in other areas of application.

In this article, we will present and apply these two main approaches to competing risk data and discuss their use in applications of the type examined here. Furthermore, we also present and apply an alternative model which is not based on hazard functions, namely a proportional odds model. Our main purpose is the methodological presentation of these approaches. The discussion emphasises the important differences between the models in terms of which aspect of the survival analysis problem is the focus of interest, and in the interpretability of parameter estimates. Thus the analyst will be better able to choose the appropriate technique in any given problem.

2 Methodology: cause-specific and subdistribution hazards

We start by recalling the basic functions that are most important in the analysis of lifetime data, working first in the context without competing risks. We denote by T the random time until the event and by $\boldsymbol{\beta}$ the vector of coefficients that expresses the effects of the elements of a vector of covariates $\mathbf{x}$ upon T. The values of these covariates for a given firm will here be assumed to be measured at time zero and to be independent of time. Naturally, an item of fundamental interest is the survival function

$$S(t|\mathbf{x}) = P(T > t|\mathbf{x})\,, \tag{1}$$

which expresses the probability that a firm will remain in operation until at least time t. The complement of the survival function is $F(t|\mathbf{x}) = P(T \leq t|\mathbf{x})$ which is the *cumulative incidence function* giving the probability of failure up to time t. In our modeling framework, a central role is occupied by the hazard function $h(t)$ which expresses the instantaneous rate of failure at time t among firms that have survived that long:

$$h(t) = \lim_{\delta t \to 0} \frac{P[t < T \leq t + \delta t | T > t]}{\delta t} = \frac{f(t)}{S(t)}\,. \tag{2}$$

Because $h(t|\mathbf{x})$ and $S(t|\mathbf{x})$ are linked by the equation

$$S(t|\mathbf{x}) = e^{-H(t|\mathbf{x})}\,, \tag{3}$$

where the cumulative hazard $H(t|\mathbf{x})$ is the integral of $h(\tau|\mathbf{x})$ over time τ from zero to t, it follows that the hazard rate and cumulative incidence provide equivalent information. One can be derived from the other. In the proportional hazards (PH) model, which includes Cox regression, the covariates act multiplicatively on a baseline hazard rate $h_0(t)$ through the factor $e^{\boldsymbol{\beta}'\mathbf{x}}$:

$$h(t|\mathbf{x}) = h_0(t) \exp\left\{\boldsymbol{\beta}'\mathbf{x}\right\} \tag{4}$$

(in the usual specification of the PH model). From equation (3),

$$S(t|\mathbf{x}) = S_0(t)^{e^{\boldsymbol{\beta}'\mathbf{x}}}\,, \tag{5}$$

where $S_0(t)$ is the baseline survival function corresponding to the baseline hazard $h_0(t)$. The effect of covariate x_i is expressed by the hazard ratio e^{β_i}: an increase of one unit in the value x_i (with the values of other covariates remaining unchanged) multiplies, from (4), the hazard by this amount while, from (5), the survival function is raised to this power. In this sense, therefore, the coefficients $\boldsymbol{\beta}$ show the effect of each covariate on both the rate (hazard function) and risk (one minus the survival function) of failure.

Now we consider the case of competing risks. There are two principal approaches to the analysis of the data, employing (a) cause-specific hazard functions, or (b) subdistribution hazards (Putter et al., 2007). These alternative approaches exist because the basic identity between hazard functions and survival functions given by equation (3) when there are no competing risks, does not extend to the context of competing risks. Therefore, in the words of Andersen and Keiding (2012), the "one-to-one correspondence between cause-specific hazard and cumulative incidence, between rate and risk, is lost". It becomes necessary to concentrate on the one aspect of the problem or the other.

Simply extending the notation of equation (2), we define the cause-specific hazard for the k^{th} of K causes as

$$h_k(t) = \lim_{\delta t \to 0} \frac{P[t < T \leq t + \delta t \bigcap D = k | T > t]}{\delta t}\,, \tag{6}$$

where D denotes the cause of failure. Thus $h_k(t)\delta t$ is the probability of failure from cause k in the interval $[t, t+\delta t)$ given that the firm has survived until time t. The cause-specific hazard is clearly the appropriate definition of the instantaneous rate of failure from cause k among the surviving firms at time t. It does not, however, lead directly to the cumulative incidence of failure from this cause along the lines of equation (5). This is because survival to time t depends not only on the hazard from this one cause but also on the hazards of all the other causes as well. It is not possible to fail from cause k at time t unless the firm has avoided failure from all causes so far. Therefore, one cause cannot be considered in isolation from the others.

The cumulative incidence of failure cause k can be estimated correctly from the subdistribution hazard function (Fine and Gray, 1999)

$$h_k^s(t) = \lim_{\delta t \to 0} \frac{P[t < T \leq t + \delta t \bigcap D = k | T > t \bigcup (T \leq t \cap D \neq k)]}{\delta t}\,, \tag{7}$$

which gives the instantaneous rate of failure in firms that have not yet experienced an event of type k, although they may have experienced a different event. In contrast, the hazard in equation (6) refers simply to failure in firms that have not yet experienced any event. Thus the practical difference lies in the set of firms that are counted as being at risk at time t. For the cause-specific hazard of equation (6), the risk set consists of the firms that still survive at time t, exactly as in standard survival analysis without competing risks. However, for the subdistribution hazard of equation (7), the risk set for cause k is augmented by the inclusion of the firms that have already failed before time t, but from causes other than cause k (despite the fact that, obviously, they cannot in fact fail again).

Both versions of the hazard function under competing risks, equations (6) and (7), can be extended to include the multiplicative effects of covariates exactly as in the proportional hazards model defined by equation (4). They are, however, different models and the values taken by the regression coefficient β_i of a covariate may be quite different between the two. In fact, it must be borne in mind that it is only the coefficients of the cause-specific model that have a straightforward interpretation; those of the subdistribution model do not. In the former case, the hazard ratio e^{β_i} gives the multiplicative effect of covariate x_i on the cause-specific hazard function, in exactly the same way as for the standard Cox model. Thus it is directly related to a meaningful quantity, namely, the probability of the occurrence of an event of the relevant type in an existing unit. Superficially, the corresponding hazard ratio for the subdistribution hazard model appears to have a very similar interpretation because it acts multiplicatively on the hazard rate given by equation (7); however, the fact that the risk set includes units that no longer exist, means that it is not actually a physically meaningful quantity. For more discussion of the interpretation and misinterpretation of hazard ratios, see in particular Austin and Fine (2017).

The point of working with the subdistribution hazard is that it leads to correct estimation of the cumulative incidence of failure from cause k and the effect that the covariates have on this, in a form similar to equation (5) above. From equation (7), the relation is

$$h_k^s(t) = -\frac{d}{dt} \ln\left(1 - F_k(t)\right) , \tag{8}$$

where $F_k(t)$ is the cumulative incidence of this cause, defined by

$$F(t) = P\left(T \leq t \cap D = k\right).$$

(This is the subdistribution function of cause k, so called because, unlike a proper distribution function, it does not tend to one as t tends to infinity: in fact, $F(t) \to P(D = k) < 1,\ t \to \infty$). Note that this definition has the desirable property that the sum of the separate cumulative incidences of the various causes at any time equals the overall cumulative incidence of failure from any cause by that time

$$F(t) = P(T \leq t) = \sum_k F_k(t).$$

In contrast, this requirement is not met by "cumulative incidences" calculated from cause-specific hazards. For example, it is well known that if a separate Kaplan-Meier type estimate of survival is calculated for each cause (treating failures from other causes as censored, as indicated above), then the total of the "incidences" obtained as one minus survival will be greater than one: incidences are overestimated. This can be seen from the fact that the cumulative incidence (subdistribution function) can also be written as

$$F_k(t) = \int_0^t h_k(\tau) S(\tau) d\tau\,, \tag{9}$$

where $S(\tau) = 1 - F(\tau)$, whereas the corresponding quantity obtained from cause-specific hazards is given by

$$\int_0^t h_k(\tau) S_k(\tau) d\tau\,, \tag{10}$$

where $S_k(\tau) = \exp\left(-H_k(\tau)\right)$ and $H_k(\tau) = \int_0^\tau h_k(s)ds$. However, $S(\tau) \leq S_k(\tau)$ for all τ, showing that (10) is greater than (9) in general (Putter et al., 2007).

From the above discussion of the definitions of these hazard functions, it can be seen that a standard program for fitting the Cox model can always be employed in order to calculate a cause-specific hazard: it suffices to fit the Cox model to a data set in which only failure from the cause of interest is treated as a failure event. A failure from any other cause is treated as censored at its failure time, with the result that it is removed from the risk set from that point onwards. On the other hand, the situation is not always so simple for subdistribution hazards, where a firm that has failed from another cause must remain in the study beyond its own failure time. If there is random right censoring, it is necessary to estimate the distribution of censoring times in order to estimate how long the firm would have been under observation in the absence of an event. However, if all censoring is "administrative" or "complete" (Fine and Gray, 1999) - that is, censoring takes place at a pre-determined time (usually the end of the study) without any losses to follow-up - then we know what the potential observation time is and hence for how long the firm must remain in the risk set. Under these circumstances, a standard program for fitting the Cox model can be employed to estimate the subdistribution hazard of failure from cause k simply by constructing a data set in which firms that fail from other causes are treated as censored at the end of their potential (not actual) observation periods. In the more general case, the estimated survival function of time to censoring is used to weight contributions to the risk set for cause k from firms that have already failed from competing causes, employing an inverse probability of censoring weighting technique, implemented in appropriate software such as the R package cmprsk (Gray, 2019).

3 An alternative: the proportional odds model

Given the growing number of articles in the literature that compare the cause-specific and subdistribution hazards approaches, in combination with a Cox-type regression model for the effect of covariates, it appears that these two methods - singly or in combination (Latouche et al., 2013) - will dominate applications of competing risk analysis in the near future, just as Cox regression itself dominates single-cause survival analysis in many fields. But, whereas the cause-specific approach is clear and meaningful, it does not lead to correct estimation of cumulative incidences, as explained above. The subdistribution hazard methodology presented in the previous section, which overcomes this difficulty, suffers from lack of interpretability of the regression coefficients and violates reality by keeping firms in the risk set after they have failed. This is a technical device that some analysts are unwilling to accept (Andersen and Keiding, 2012). Therefore, it is worth exploring alternatives.

Supposing as usual a multiplicative effect of covariates $\mathbf{x}$ as in (4), the cumulative incidence in the Fine-Gray model given by (8) is equivalent to the model

$$\log\left(-\log\left(1 - F_k\left(t|\mathbf{x}\right)\right)\right) = \log\left(A\left(t\right)\right) + \boldsymbol{\beta}'\mathbf{x}$$

where $A(t)$ is an unspecified, non-decreasing function. The appearance here of the complementary log-log link function suggests that this model is a special case within a wider class of transformation models that employ other link functions. (See, for example, Mao and Lin, 2017). In particular, the logit link function, so familiar from logistic regression, could be used. This gives a proportional odds cumulative incidence model for competing risks

$$\text{logit } F_k(t|\mathbf{x}) = \log \frac{F_k(t|\mathbf{x})}{1 - F_k(t|\mathbf{x})} = \log(A(t)) + \boldsymbol{\beta}'\mathbf{x}$$

(Eriksson et al., 2015). One immediate advantage of this compared to the Fine-Gray model is the more direct interpretability of the coefficients $\boldsymbol{\beta}$ as odds ratios. Software for this approach is implemented in the R package 'timereg' (Scheike and Zhang, 2011; Scheike, 2019).

4 Application

The data available to us were drawn from two files provided by the Chamber of Commerce of Perugia: the Business Register of Companies in the Region of Umbria, Italy, and annual balance sheets for capital companies for the years 2008-2013. Background data included the firm's year of foundation, location, legal form and economic sector of activity. Financial indexes were constructed from the firm's balance sheets. If the firm ceased activity, the date of closure and the reason for closure were recorded in the data base. Further details of the data are given by Pierri and Caroni (2017).

We analysed the subsequent survival of firms that were in operation at the start of 2008, using covariates measured at this baseline. Of the 15184 enterprises included in our data set, 3049 (20.1%) became inactive during this period of economic crisis up to the end of the study period in 2013 and were thus counted as failing. The frequencies of the three different routes to inactivity were as follows: voluntary winding-up, dissolution or liquidation (1808 firms, 59.3% of the failures); bankruptcy (236 firms, 7.7%); and closure without action by creditors or the courts (1005 firms, 33.0%).

Before considering the effect of covariates on failure and survival, we first look at non-parametric Kaplan-Meier survival estimates. Figure 1 shows the estimated survival for each cause separately and also for failure from any cause. The estimates for each cause separately treat failures from other causes as right-censored at their failure times; in other words, they correspond to the cause-specific method. It can be seen that, from the beginning, voluntary winding-up occurs at a faster rate than closure without action by creditors or courts, which in turn occurs faster than bankruptcy. The cause-specific survival probabilities at 6 years are 0.8445, 0.9120 and 0.9762 in these three curves. Adding the cause-specific failure incidences at each time point gives the upper curve plotted in Figure 2. For example, at 6 years, the "total incidence" appears by this method to be (1 - 0.8445)+(1 - 0.9120)+(1 - 0.9762)=0.2673. However, the true incidence of any failure is obtained as 1 minus the any-cause survival estimate from Figure 1. This is shown in the lower curve of Figure 2, which reaches 0.2481 at 6 years. The difference between the two curves is small up to this time point in these data, in which failures

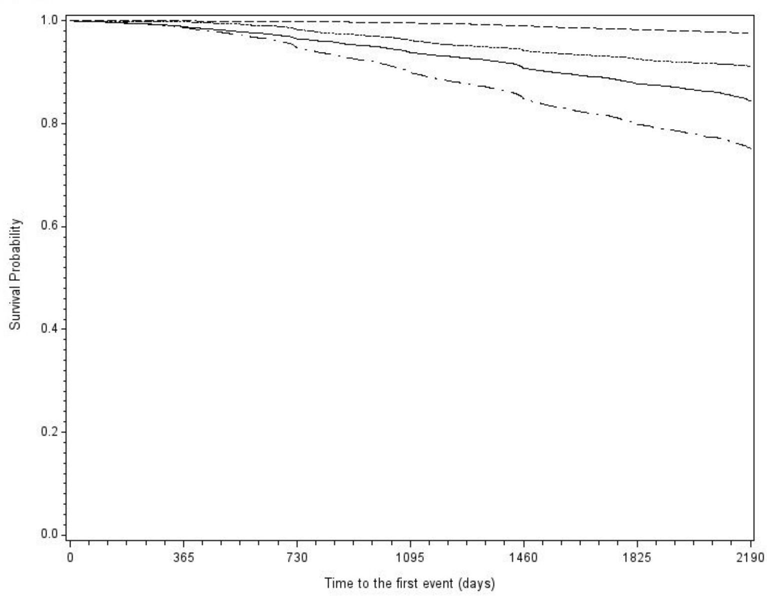

Figure 1: Kaplan-Meier estimates of survival in relation to different causes and any cause of a firm's failure. From top to bottom, the lines refer to: bankruptcy; closure; voluntary winding-up; and any cause.

(especially from the third cause) are occurring quite rarely in comparison to some other applications that one comes across in the literature, but nevertheless illustrates the point made above that the cause-specific approach overestimates the overall incidence of failure. If follow-up were extended, the curves would diverge substantially as the numerical difference between equations (9) and (10) increases.

We now turn to examining the effect of a range of covariates on failure, applying the proportional hazards assumption to both the cause-specific and subdistribution hazard models described above. The covariates, which refer to the firm's activity, location, age and financial performance, are listed in Table 1. Both analyses were carried out by backward elimination of non-significant ($p > 0.05$) covariates. Table 2 shows summary statistics and Table 3 shows the estimated hazard ratios (HR) and 95% confidence intervals (CI) for the significant covariates.

It is apparent that in this particular application there are no great differences between the results of the two approaches. That is, the factors that influence the risk are basically the same as those that influence the rate of failures. However, on comparing results for different causes (within either approach), large differences emerge. We observe some covariates that have effects for one cause but not others (for example, the firm's sector of economic activity and the geographical location), others that affect all causes but with notably unequal HR (for example, debt ratios) and still others that affect different causes in opposite directions - reducing the hazard for one cause, increasing it for another (for example, the type of firm).

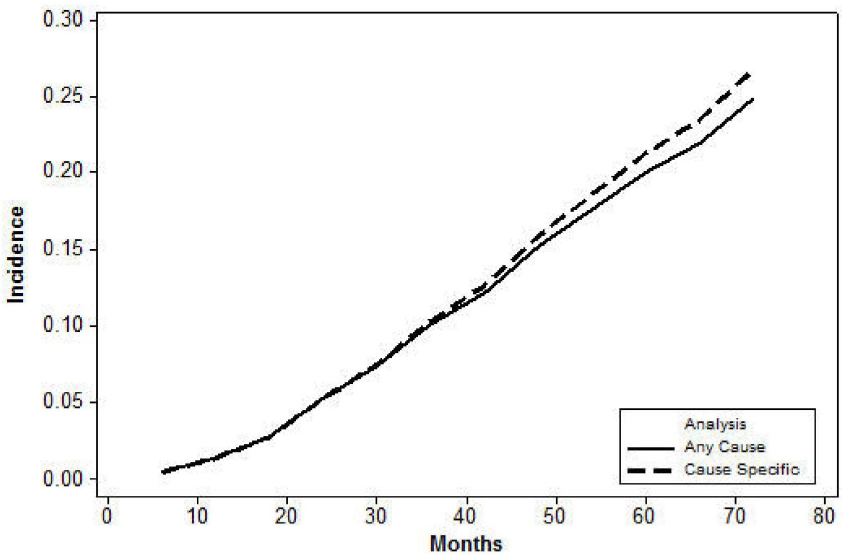

Figure 2: Cumulative incidence estimated from cause-specific hazards (upper curve) and any-cause analysis (lower curve).

In the case of the debt ratio, it appears that a high value increases the risk of the firm's failure by any route, but especially heightens the risk of bankruptcy. Regarding the firm's area of economic activity, Table 3 shows significant effects only in relation to bankruptcy, for which every sector has a substantially elevated HR in comparison to the reference sector, which is Agriculture, but only the Manufacturing and Construction sectors have a 95% CI that excludes the value one. Overall, there were no great differences between sectors in the proportion of failures, which ranged from 21.4% in Tourism and Construction to 16.6% in Agriculture (Table 4). However, there were differences between modes of failure, most noticeably with regard to bankruptcy. This is the failure mode presenting least often (only 7.7% of firms went bankrupt in the six-year period), but was twice as common (around 10% of failures) among firms in the sectors of Commerce, Construction and Manufacturing as in Tourism and Other Services (both around 5%), where in turn it was twice as common as in Agriculture, in which it occurred very rarely (only 2 cases, 2.6% of the failures in this sector) (Table 4).

Table 4 also shows failures by type of firm. Cooperatives had a higher failure rate (25.5%) over the six-year period than other types and three-quarters of their failures took the form of winding-up. In contrast, not much more than a quarter of the failures of publicly-traded companies took this form, with limited-liability companies in between. Correspondingly, closure was much more common among publicly-traded companies than among limited-liability companies and cooperatives. This explains the opposite directions of the HR in Table 3. Cooperatives and limited-liability companies tend to

Table 1: Covariates investigated in the survival analyses. In addition, the firm's Activity Sector (6 categories), Legal Form (3 categories) and Geographical Location (2 categories) were represented by indicator variables.

Abbreviation		Variable	
CR		Current Ratio	
QR		Quick Ratio	
L		Leverage	
IRR		Investment Rigidity Ratio	
	TAR		Tangible Assets Ratio
	IAR		Intangible Assets Ratio
	FFAR		Financial Fixed Assets Ratio
	OFAR		Other Fixed Assets Ratio
IER		Investment Elasticity Ratio	
	IIR		Inventories Impact Ratio
	LR		Liquidity Ratio
	STLR		Short Term Liquidity Ratio
	LTLR		Long Term Liquidity Ratio
ER		Equity Ratio	
DR		Debt Ratio	
	PDR		Permanent Debt Ratio
	CDR		Current Debt Ratio
ROA		Return on Assets	
ROE		Return on Equity	
ROT		Return on Turnover	
ROS		Return on Sales	
TUR		Turnover	

agree arrangements with creditors and avoid total closure.

We note that the Cox regression for all-causes failure, also shown in Table 3, gives an HR of 1.05 for cooperatives and 0.84 for limited-liability companies versus publicly-traded companies, neither significantly different from one. The only significant comparison among firm types in that analysis is between cooperative and limited liability companies (HR=1.25, CI=1.10 - 1.42; not shown in table). These results would suggest that the type of firm has only a small impact on failure rates and risks, thereby masking the rather large differences that the more detailed competing risks analysis brings to

Table 2: Robust summary statistics for the covariates that were found to be statistically significant ($p < 0.05$) in the final model: winsorized mean values and standard deviation based on Gini's mean difference, obtained from the SAS procedure PROC UNIVARIATE.

			Cause of failure		
	Survivors (n=12135)	Failure (n=3049)	Winding-up (n=1808)	Closure (n=1005)	Bankruptcy (n=236)
TAR	0.126 (0.206)	0.053 (0.125)	0.043 (0.106)	0.080 (0.162)	0.051 (0.106)
IAR	0.019 (0.056)	0.022 (0.070)	0.024 (0.078)	0.166 (0.059)	0.168 (0.052)
FFAR	0.006 (0.045)	0.010 (0.070)	0.007 (0.056)	0.015 (0.099)	0.015 (0.046)
IIR	0.070 (0.152)	0.048 (0.125)	0.045 (0.118)	0.040 (0.126)	0.091 (0.161)
STLR	0.051 (0.107)	0.037 (0.102)	0.045 (0.117)	0.028 (0.086)	0.017 (0.048)
ER	0.236 (0.321)	0.213 (1.569)	0.201 (2.294)	0.277 (0.505)	0.007 (0.451)
DR	0.700 (0.327)	0.715 (0.496)	0.726 (0.530)	0.670 (0.436)	0.854 (0.444)
PDR	0.170 (0.220)	0.144 (0.211)	0.143 (0.211)	0.149 (0.222)	0.134 (0.159)
CDR	0.508 (0.388)	0.551 (1.489)	0.566 (2.159)	0.507 (0.499)	0.704 (0.508)
ROA	0.004 (0.153)	-0.078 (0.893)	-0.101 (1.279)	-0.043 (0.334)	-0.076 (0.301)
AGE	5.965 (8.396)	4.829 (6.875)	4.698 (6.787)	4.728 (6.845)	6.266 (7.601)

light.

With respect to the geographical location (Table 3) we notice it affects only the voluntary winding-up, with a quite similar intensity in both approaches and in Cox regression for any cause. Overall there were no great differences in the distribution of failures (Table 4) between Perugia and Terni, even though there is a higher percentage of closure in the first (34.1%) than in the second (30.0%), the opposite for winding-up (58.2% and 62.1% respectively).

Looking at the significant financial indexes (TAR, IAR and FFAR), Table 3 shows that a higher value of IAR strongly increases the hazard of voluntary winding-up in both approaches (HR of 2.11 and 2.08). FFAR also affects the hazard of winding-up (HR of 1.59 and 1.46), and acts on the hazard of closure extremely strongly (4.00 and 3.80). In the Cox model for all-causes failure, the HR of 2.31 is a sort of average between these effects. TAR influences hazards in the opposite direction to the other financial indexes, with lower values implying increased hazards.

Finally, Table 5 presents for comparison the effects of the covariates on cumulative incidences in two alternative models: hazard ratios from the Fine-Gray model for subdistribution hazards (already given in Table 3) and the odds ratios obtained by fitting

Table 3: Results from cause-specific and subdistribution hazards regression analyses of competing risks. Results for any-cause survival analysis by Cox regression are shown for comparison. HR=hazard ratio.

	Cause-Specific						Subdistribution Hazard						Any-cause Survival	
	Winding-up		Closure		Bankruptcy		Winding-up		Closure		Bankruptcy			
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Activity Sector *														
Commerce					3.10	0.75-12.78					3.11	0.76-12.69		
Construction					4.14	1.01-16.98					4.10	1.01-16.72		
Manufacturing					4.33	1.06-17.75					4.34	1.07-17.71		
Other Services					1.90	0.46-7.83					1.90	0.46-7.80		
Tourism					2.24	0.52-9.61					2.19	0.51-9.35		
Type of Firm **														
Cooperative	2.44	1.57-3.78	0.48	0.33-0.70			2.61	1.57-4.04	0.45	0.31-0.66			1.05	0.81-1.36
Limited Liability	1.57	1.04-2.39	0.50	0.38-0.67			1.67	1.11-2.52	0.48	0.36-0.65			0.84	0.67-1.06
Location: Terni ***	1.21	1.10-1.34					1.21	1.09-1.33					1.15	1.06-1.24
TAR	0.28	0.21-0.38					0.30	0.22-0.40			0.43	0.21-0.89	0.43	0.35-0.53
IAR	2.11	1.83-2.43	1.71	1.22-2.41			2.08	1.75-2.47			1.55	1.07-2.24	2.02	1.77-2.30
FFAR	1.59	1.12-2.25	4.00	2.89-5.52			1.46	1.03-2.07	3.80	2.73-5.28			2.31	1.83-2.91
IIR	0.67	0.53-0.85					0.70	0.55-0.89					0.75	0.63-0.90
STLR	1.79	1.32-2.42					1.73	1.07-2.81					1.41	1.10-1.80
ER							0.90	0.87-0.94			0.91	0.84-0.98		
DR	1.85	1.67-2.06	1.22	1.03-1.44	3.43	2.71-4.34	1.74	1.53-1.98			3.18	2.46-4.12	1.75	1.61-1.90
PDR	0.52	0.42-0.65	0.56	0.42-0.74	0.23	0.13-0.44	0.48	0.38-0.61	0.69	0.51-0.93	0.24	0.14-0.43	0.51	0.43-0.60
CDR							0.90	0.87-0.94			0.83	0.77-0.90		
ROA	0.995	0.993-0.998					0.995	0.993-0.997			1.06	1.01-1.11		
TUR			0.98	0.96-0.99										
AGE	0.97	0.96-0.98	0.96	0.96-0.97	0.98	0.96-0.99	0.97	0.97-0.98	0.97	0.96-0.97			0.97	0.965-0.974

Reference categories: *Agriculture; **Publicly-traded; *** Perugia

Table 4: Distribution of causes of failure, by Activity Sector, Legal Form and Geographical Location of the firm.

		Closure	Winding-up	Bankruptcy	Total Inactive	Total Active	Total	% Inactive
Activity Sector								
Agriculture	N	30	46	2	78	393	471	16.6
	%	38.5	59.0	2.6				
Commerce	N	170	322	51	543	2412	2955	18.4
	%	31.3	59.3	9.4				
Construction	N	176	331	59	566	2085	2651	21.4
	%	31.1	58.5	10.4				
Manufacturing	N	158	278	59	495	2175	2670	18.5
	%	31.9	56.2	11.9				
Other Services	N	351	592	45	988	3677	4665	21.2
	%	35.5	59.9	4.6				
Tourism	N	120	239	20	379	1393	1772	21.4
	%	31.7	63.1	5.3				
Legal Form								
Cooperative	N	56	193	7	256	747	1003	25.5
	%	21.9	75.4	2.7				
Limited Liability	N	897	1592	224	2713	11063	13776	19.7
	%	33.1	58.7	8.3				
Publicly-traded	N	52	23	5	80	325	405	19.8
	%	65.0	28.8	6.3				
Location								
Perugia	N	748	1275	168	2191	9145	11336	19.3
	%	34.1	58.2	7.7				
Terni	N	257	533	68	858	2990	3848	22.3
	%	30.0	62.1	7.9				
Total	N	1005	1808	236	3049	12135	15184	20.1
	%	33.0	59.3	7.7				

the proportional odds model. We repeat that hazard ratios obtained from the Fine-Gray model do not have a direct interpretation; this is one of the reasons for preferring the proportional odds model, in which an odds ratio has the interpretation familiar from logistic regression. Thus, for example, the odds of winding-up by a given time are 1.21 times greater (95% confidence interval 1.09 - 1.36) for a firm in Terni compared to one in Perugia, all other things being equal. This happens to be equal to the hazard ratio, but this is just a coincidence. However, examining the table, none of the covariate effects differs between the models to the extent that conclusions regarding their relative importance would be changed.

5 Discussion

The topic of competing risks has a very long history (it is said to go back to Daniel Bernoulli in 1760) and whole books have been written about it (David and Moeschberger, 1978; Crowder, 2001). Yet many general texts on survival analysis and reliability - even one as extensive as Therneau and Grambsch (2013) - barely mention the subject. For this reason, and given the fairly recent development of some of the methodology, quite a large number of expository articles can be found in a variety of journals. The emphasis of these is on the contrast between the cause-specific and subdistribution hazard regression models, usually in a medical context: see, among others, Putter et al. (2007); Bakoyannis and Touloumi (2012); Dignam et al. (2012); Haller et al. (2013) and Austin and Fine (2017). One paper giving a view of reliability and other fields of application is Ma and Krings (2008). Several papers present computational procedures in detail. Kohl et al. (2015) describe a SAS macro for the subdistribution hazards regression model. Others (Putter et al., 2007; Scrucca et al., 2007; Haller et al., 2013) employ the R programming language and in particular the cmprsk package (Gray, 2019). Austin and Fine (2017) describe the application of SAS and R in detail and also mention the facilities available in Stata. Bakoyannis and Touloumi (2012) refer to R and Stata.

The cause-specific hazard regression approach possesses the major advantage of great simplicity, being, in effect, simply a repeated application of the Cox semiparametric proportional hazards model which is so widely employed as to have become almost synonymous with survival analysis for many users. That predominance is probably undesirable, given that even Cox himself in his original publication emphasised that the proportional hazards assumption is essentially an entirely empirical convenience (Cox, 1972) and other authors have stressed that in general there is no physical or biological basis for it (for example: Elsayed and Chan, 1990; Oakes, 2013). Other models deserve wider attention, notably the accelerated failure time model which is generally preferred over the proportional hazards model by reliability engineers but is seen less often in other fields (Hougaard, 1999; Hutton and Monaghan, 2002). Fine and Gray's subdistribution hazard regression approach is less easy to implement in the general case although not when there is only administrative censoring with no losses to follow-up, as in the present application.

The cause-specific and subdistribution hazard regression approaches focus on different aspects of survival, rate and risk, as described earlier in the present paper. Which aspect is the more important may depend on the chief aim of the study. The consensus in the literature seems to be that cause-specific hazards are more appropriate for uncovering the causes (or, at least, the correlates) of failures whereas subdistribution hazards, by producing the correct incidence functions, should be used for developing predictive mod-

Table 5: Effects of covariates on cumulative incidences compared between two models: Fine-Gray subdistribution hazards regression (HR=hazard ratio) and proportional odds (OR = odds ratio). CI = confidence interval.

	Winding-up				Closure				Bankruptcy			
	HR	95% CI	OR	95% CI	HR	95% CI	OR	95% CI	HR	95% CI	OR	95% CI
Activity Sector *												
Commerce									3.11	0.76-12.69	3.00	0.73-12.30
Construction									4.10	1.01-16.72	3.94	0.95-16.28
Manufacturing									4.34	1.07-17.71	4.31	1.05-17.81
Other Services									1.90	0.46-7.80	1.79	0.43-7.46
Tourism									2.19	0.51-9.35	2.06	0.48-8.85
Type of Firm **												
Cooperative	2.61	1.57-4.04	2.80	1.77-4.44	0.43	0.29-0.64	2.34	1.56-3.53				
Limited Liability	1.67	1.11-2.52	1.73	1.13-2.66	0.46	0.34-0.64	1.09	0.82-1.44				
Location: Terni ***	1.21	1.09-1.33	1.21	1.09-1.36								
TAR	0.30	0.22-0.40	0.28	0.20-0.38					0.43	0.21-0.89	0.46	0.22-0.95
IAR	2.08	1.75-2.47	3.03	2.12-4.35					1.55	1.07-2.24	1.51	1.02-2.25
FFAR	1.46	1.03-2.07	1.52	1.04-2.24	3.80	2.73-5.28	4.22	2.92-6.11				
IIR	0.70	0.55-0.89	0.67	0.52-0.87								
STLR	1.73	1.07-2.81	1.76	1.00-3.10								
ER	0.90	0.87-0.94	0.87	0.82-0.93					0.91	0.84-0.98	0.89	0.83-0.96
DR	1.74	1.53-1.98	1.89	1.63-2.20					3.18	2.46-4.12	3.29	2.51-4.31
PDR	0.48	0.38-0.61	0.42	0.32-0.55	0.69	0.51-0.93	0.68	0.50-0.93	0.24	0.14-0.43	0.23	0.13-0.40
CDR	0.90	0.87-0.94	0.87	0.82-0.93					0.83	0.77-0.90	0.81	0.76-0.85
ROA	0.995	0.993-0.997	0.99	0.99-1.00					1.06	1.01-1.11	1.06	1.02-1.10
AGE	0.97	0.97-0.98	0.97	0.96-0.98	0.97	0.96-0.97	0.96	0.96-0.97				

Reference categories: *Agriculture; **Publicly-traded; *** Perugia els (Lau et al., 2009). In particular, the latter is what is required in applications such as the one described here and in credit-scoring, where the interest lies mainly in the prediction of failure over the duration of the study, given a set of covariates recorded at the baseline.

However, some authors recommend, for a full understanding of the issue under study, that both approaches should be applied (e.g. Latouche et al., 2013). This is despite the fact that it is not possible for both of these regression models to possess the proportional hazards property; thus, if the cause-specific hazards is correctly specified as proportional hazards, then the other model would be misspecified as proportional hazards. Nevertheless, there is evidence that the subdistribution hazards model in this form is still useful under these circumstances (Grambauer et al., 2010; Latouche et al., 2013).

The proportional odds model has seen little use so far. The availability of a program in the R language may lead to its wider adoption. It is important for users to be aware that various options are available for competing risks analysis and not be limited to a single choice as has unfortunately happened, in effect, for many users of single-cause survival analysis.

6 Acknowledgment

Francesca Pierri received the support of “Fondo Ricerca di Base, 2018” from Universitá degli Studi di Perugia for the project “Survival Analysis and Competing Risks: a study on firms in Umbria Region”.

References

Amendola, A., Restaino, M., and Sensini, L. (2015). An analysis of the determinants of financial distress in Italy: A competing risks approach. *International Review of Economics & Finance*, 37:33–41.

Andersen, P. K. and Keiding, N. (2012). Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine*, 31(11-12):1074–1088.

Austin, P. C. and Fine, J. P. (2017). Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Statistics in Medicine*, 36(27):4391–4400.

Bakoyannis, G. and Touloumi, G. (2012). Practical methods for competing risks data: a review. *Statistical Methods in Medical Research*, 21(3):257–272.

Balcaen, S., Manigart, S., Buyze, J., and Ooghe, H. (2012). Firm exit after distress: differentiating between bankruptcy, voluntary liquidation and M & A. *Small Business Economics*, 39(4):949–975.

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(1):63–93.

Banasik, J., Crook, J. N., and Thomas, L. C. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society*, 50(12):1185–1190.

Bhattacharjee, A., Higson, C., Holly, S., and Kattuman, P. (2009). Macroeconomic instability and business exit: Determinants of failures and acquisitions of UK firms. *Economica*, 76(301):108–131.

Chancharat, N., Tian, G., Davy, P., McCrae, M., and Lodh, S. (2010). Multiple states of financially distressed companies: Tests using a competing-risks model. *Australasian Accounting, Business and Finance Journal*, 4(4):27–44.

Collett, D. (2015). *Modelling survival data in medical research*. Chapman and Hall/CRC.

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

Crowder, M. J. (2001). *Classical competing risks*. Chapman and Hall/CRC.

David, H. A. and Moeschberger, M. L. (1978). *The theory of competing risks*. C. Griffin.

Dignam, J. J., Zhang, Q., and Kocherginsky, M. (2012). The use and interpretation of competing risks regression models. *Clinical Cancer Research*, 18(8):2301–2308.

Elsayed, E. and Chan, C. (1990). Estimation of thin-oxide reliability using proportional hazards models. *IEEE Transactions on Reliability*, 39(3):329–335.

Eriksson, F., Li, J., Scheike, T., and Zhang, M.-J. (2015). The proportional odds cumulative incidence model for competing risks. *Biometrics*, 71(3):687–695.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509.

Grambauer, N., Schumacher, M., and Beyersmann, J. (2010). Proportional subdistribution hazards modeling offers a summary analysis, even if misspecified. *Statistics in Medicine*, 29(7-8):875–884.

Gray, B. (2019). "cmprsk": Subdistribution analysis of competing risks. R package version 2.2-9. Available online at: https://cran.r-project.org/web/packages/cmprsk/cmprsk.pdf.

Haller, B., Schmidt, G., and Ulm, K. (2013). Applying competing risks regression models: an overview. *Lifetime Data Analysis*, 19(1):33–58.

Hosmer, D. W., Lemeshow, S., and May, S. (2008). *Applied survival analysis: regression modeling of time-to-event data*. Wiley-Interscience.

Hougaard, P. (1999). Fundamentals of survival data. *Biometrics*, 55(1):13–22.

Hutton, J. and Monaghan, P. (2002). Choice of parametric accelerated life and proportional hazards models for survival data: asymptotic results. *Lifetime Data Analysis*, 8(4):375–393.

Kohl, M., Plischke, M., Leffondré, K., and Heinze, G. (2015). Pshreg: a SAS macro for proportional and nonproportional subdistribution hazards regression. *Computer Methods and Programs in Biomedicine*, 118(2):218–233.

Kwon, S. and Hahn, S. B. (2010). Duration analysis of corporate bankruptcy in the presence of competing risks. *Applied Economics Letters*, 17(15):1513–1516.

Latouche, A., Allignol, A., Beyersmann, J., Labopin, M., and Fine, J. P. (2013). A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *Journal of Clinical Epidemiology*, 66(6):648–653.

Lau, B., Cole, S. R., and Gange, S. J. (2009). Competing risk regression models for epidemiologic data. *American Journal of Epidemiology*, 170(2):244–256.

Ma, Z. and Krings, A. W. (2008). Competing risks analysis of reliability, survivability, and prognostics and health management (PHM). In *2008 IEEE Aerospace Conference*, pages 1–21. IEEE.

Mao, L. and Lin, D. (2017). Efficient estimation of semiparametric transformation models for the cumulative incidence of competing risks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):573–587.

Narain, B. (1992). Survival analysis and the credit granting decision. In: Thomas, L.C., Crook, J.N., Edelman, D.B. (eds.), *Credit Scoring and Credit Control*, Oxford University Press, 109–122.

Oakes, D. (2013). An introduction to survival models: in honor of Ross Prentice. *Lifetime Data Analysis*, 19(4):442–462.

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.

Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26(11):2389–2430.

Scheike, M. T. (2019). "timereg": Flexible regression models for survival data. R package version 1.9.4. Available online at: https://cran.r-project.org/web/packages/timereg/timereg.pdf.

Scheike, T. H. and Zhang, M.-J. (2011). Analyzing competing risk data using the R timereg package. *Journal of Statistical Software*, 38(2):1–16.

Scrucca, L., Santucci, A., and Aversa, F. (2007). Competing risk analysis using R: an easy guide for clinicians. *Bone Marrow Transplantation*, 40(4):381.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1):101–124.

Sohn, S. Y. and Jeon, H. (2010). Competing risk model for technology credit fund for small and medium-sized enterprises. *Journal of Small Business Management*, 48(3):378–394.

Therneau, T. M. and Grambsch, P. M. (2013). *Modeling survival data: extending the Cox model.* Springer Science & Business Media.