



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A comparative analysis of the UK and Italian small businesses using Generalised Extreme Value models

Citation for published version:

Andreeva, G, Calabrese, R & Osmetti, SA 2016, 'A comparative analysis of the UK and Italian small businesses using Generalised Extreme Value models', *European Journal of Operational Research*, vol. 249, no. 2, pp. 506-516. <https://doi.org/10.1016/j.ejor.2015.07.062>

Digital Object Identifier (DOI):

[10.1016/j.ejor.2015.07.062](https://doi.org/10.1016/j.ejor.2015.07.062)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

European Journal of Operational Research

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



1
2
3
4
5
6
7
8
9
10 A comparative analysis of the UK and Italian small
11
12 businesses using Generalised Extreme Value models
13
14
15

16
17 Galina Andreeva¹, Raffaella Calabrese² and Silvia Angela Osmetti³
18

19
20 ¹Business School, University of Edinburgh
21

22
23 29 Buccleuch Place, Edinburgh EH8 9J, UK
24

25
26 `Galina.Andreeva@ed.ac.uk`
27

28
29 ²Essex Business School, University of Essex
30

31
32 Wivenhoe Park, Colchester CO4 3SQ, UK
33

34
35 `rcalab@essex.ac.uk`
36

37
38 ³Department of Statistical sciences, Università Cattolica del Sacro Cuore di Milano
39

40
41 Largo Gemelli 1, 20123 Milano, Italy
42

43
44 `silvia.osmetti@unicatt.it`
45

46
47 Corresponding author: Dr Galina Andreeva
48

49
50 Tel:+44(0)65123293
51

52
53 Fax:+44 (0)131 651 3197
54

55
56 E-mail: Galina.Andreeva@ed.ac.uk
57
58
59
60
61
62

Abstract

This paper presents a cross-country comparison of significant predictors of small business failure between Italy and the UK. Financial measures of profitability, leverage, coverage, liquidity, scale and non-financial information are explored, some commonalities and differences are highlighted. Several models are considered, starting with the logistic regression which is a standard approach in credit risk modelling. Some important improvements are investigated. Generalised Extreme Value (GEV) regression is applied in contrast to the logistic regression in order to produce more conservative estimates of default probability. The assumption of non-linearity is relaxed through application of BGEVA, non-parametric additive model based on the GEV link function. Two methods of handling missing values are compared: multiple imputation and Weights of Evidence (WoE) transformation. The results suggest that the best predictive performance is obtained by BGEVA, thus implying the necessity of taking into account the low volume of defaults and non-linear patterns when modelling SME performance. WoE for the majority of models considered show better prediction as compared to multiple imputation, suggesting that missing values could be informative.

Keywords: Decision support systems, Risk analysis, Credit Scoring, Small and Medium Sized Enterprises, Default prediction.

1 Introduction

Small and Medium Enterprises (SMEs) play a central role in the European Union (EU) economy, as recognised by the Small Business Act of the European Commission in 2008 (http://ec.europa.eu/enterprise/entrepreneurship/docs/sba/SBA_IA). In 2011 SMEs represented 99% of enterprises in Europe, employing more than two thirds of the workforce and contributing 58% of total EU added value. The importance of SMEs varies across the EU. In some countries, e.g. Italy, Spain and Portugal, SMEs have larger shares in employment

1
2
3
4 and added value and higher presence than the EU average. On the contrary, these figures are
5
6 lower than the EU average in other countries, e.g. the UK, Germany and France.
7

8 In this work we compare Italy and the UK since the economies of these countries are
9
10 different, and it is of interest to explore the differences in predictors of SMEs failures, especially
11
12 in the aftermath of the "credit crunch". The literature on SME default prediction is limited,
13
14 in particular in cross-country comparisons, and the main objective of this paper is to fill in this
15
16 gap. This paper contributes to the existing cross-country research by an initial exploratory
17
18 investigation of risk predictors using accounting and some non-financial information that are
19
20 available from public sources.
21

22
23 Several models are considered, starting with the logistic regression which is a standard
24
25 modelling approach in credit risk research (Thomas et al., 2002). Yet in situations with low
26
27 numbers of events (defaults), alternative approaches producing more conservative estimates
28
29 of default probabilities might be of importance. In this paper we concentrate on asymmetric
30
31 link function and non-linearity between the response and predictors. In real applications
32
33 the number of defaults is small, therefore, suggesting the asymmetric link function might be
34
35 beneficial. At the same time the assumption of linearity is not always supported by patterns in
36
37 the real data. An additional contribution of this paper consists in extending the application of
38
39 Generalised Extreme Value (GEV) regression that has been proposed for low default portfolios
40
41 by Calabrese & Osmetti (2013) to two countries. Furthermore, the problem of non-linearity
42
43 is explored through the application of non-parametric additive model (BGEVA).
44

45
46 The public sources often have incomplete data and this problem is particularly relevant
47
48 for SMEs. Another objective and contribution of this paper consists in the exploration of
49
50 two approaches to handling the missing values: multiple imputation and Weights of Evidence
51
52 transformation which is credit industry's preferred approach.
53

54
55 The rest of the paper is structured as follows: Section 2 provides some background in-
56
57 formation on the importance of SMEs to the economy and some differences across the two
58
59 countries. It also summarises previous research on SMEs failure prediction. Section 3 explains
60
61
62

1
2
3
4 the methodology, and Section 4 presents the empirical results, including data description, com-
5 parison of predictive accuracy and comparison of statistically significant risk predictors. The
6
7 final section concludes.
8
9

10 11 12 **2 Background and literature review** 13 14 15

16 There are some notable differences in characteristics of SMEs in the UK and Italy. In Italy,
17 SMEs form 99.9% of the firms. In 2011 they employed around 81% of the workforce and
18 contributed 68.3% of the Italian added value (EC, 2012a). In terms of the number of SMEs,
19 Italy has the largest SME sector in the EU. With 3.813 million SMEs Italy has almost twice as
20 many as UK (1.649 million). However, the vast majority of Italian SMEs are micro-firms with
21 less than 10 employees. In fact, Italy's share of micro-firms, at 94.6%, exceeds the EU-average
22 (92.2%). Hence, the micro-firms' contribution to employment (46.6% against the EU-average
23 of 29.6%) and added value (29.4% against the EU-average of 21.2%) is high.
24
25
26
27
28
29
30
31
32

33 On the contrary, the UK economy is characterised by larger companies. In 2011 more than
34 half of the UK added value was produced by large companies that employed less than half
35 (45.7%) of the workforce and constituted only 0.4% of the UK companies. The percentage of
36 micro-firms in the UK (89.5%) is lower than the EU-average (92.2%), and those employ only
37 20.3% of the workforce and create only 18.5% of the UK added value (EC, 2012b).
38
39
40
41
42

43 Financial crisis has substantially affected SMEs sectors in both countries and recovery has
44 been weaker than in the EU on the whole. The Italian SME sector has reversed to the levels
45 of 2005 (i.e. before the crisis) in terms of the number of firms, employment and value-added
46 creation. In the UK, SMEs have been hit mostly in terms of employment and value-added
47 creation, but the numbers of SMEs are higher than in 2005 and stable. In both countries
48 larger firms suffered less as compared to the smaller ones.
49
50
51
52
53
54

55 Despite an important role that SMEs play in any economy, academic research into SMEs
56 failure prediction is not very extensive. There are some (albeit not numerous) papers inves-
57 tigating success factors or default risk of SMEs in a specific country, e.g. Altman & Sabato
58
59
60
61
62
63
64
65

1
2
3
4 (2007) for the US, Fantazzini & Figini (2009) for Germany, Sohn & Kim (2013) for South Ko-
5
6 rea, Martens et al. (2011) for Flanders - to give some examples, yet literature on international
7
8 comparisons of failure prediction is exceptionally limited.
9

10 The survey by Altman & Narayanan (1997) summarised previous research on the per-
11
12 formance of companies (not only SMEs) in 22 countries that included both developed and
13
14 developing economies. Most studies surveyed found measures of profitability, leverage, liq-
15
16 uidity, cash flow management, growth, efficiency to be important for bankruptcy prediction,
17
18 although specific measures used would vary from country to country. A more recent study
19
20 by Lussier & Halabi (2010) compared performance of SMEs in the USA, Croatia and Chile.
21
22 Among the variables that were found important for business performance were characteristics
23
24 of managers (education, experience) and the quality of business functions (record keeping,
25
26 financial control, planning, staffing).
27
28

29 The most comprehensive study of European SMEs to date is by Michala et al. (2013)
30
31 where a simple hazard model (Shumway, 2001) has been applied to small businesses from
32
33 eight European countries, namely Czech Republic, France, Germany, Italy, Poland, Portugal,
34
35 Spain and the United Kingdom for the period of 2000-2009. The paper has confirmed the
36
37 significance of indicators of profitability, coverage, leverage and cash flow for bankruptcy
38
39 prediction in cross-country setting. In addition, some non-financial company characteristics
40
41 have been investigated and the effect of macroeconomic variables. Pederzoli et al. (2013)
42
43 modelled credit risk of EU innovative SMEs, but the authors did not make cross-country
44
45 comparisons.
46
47

48 There were some comparisons between two countries. Ihua (2009) compared the key factors
49
50 influencing SMEs failure between the UK and Nigeria, and found that economic conditions
51
52 and infrastructure were more significant in Nigeria, whilst in the UK the key factors were due
53
54 to internal company characteristics, including management efficiency.
55

56 Dietsch & Petey (2004) analysed default probabilities and asset correlations for French
57
58 and German SMEs. Yet the focus of their analysis was more on comparison of correlations
59
60
61
62
63
64
65

1
2
3
4 of SMEs as opposed to large corporations, the paper did not look at financial ratios or other
5
6 predictors of default.

7
8 As for SME research in the UK, Lin et al. (2012) compared different definitions of financial
9
10 distress on a sample from 2001 to 2004 and concluded that although each definition changed
11
12 the model composition substantially, the most useful variables in distinguishing between dis-
13
14 tressed and healthy companies, were profit related measures, growth and efficiency ratios.
15
16 Altman et al. (2010) developed a default prediction model using financial indicators of lever-
17
18 age, profitability, working capital and non-financial information (e.g. age, default events in
19
20 the past) using the data from 2000 to 2007. They found the non-financial variables provided a
21
22 notable improvement in predictive performance. Orton et al. (2011) explored the behaviour of
23
24 the UK SMEs from 2007 to 2010 - through the "credit crunch". They demonstrated that there
25
26 was a significant degree of stability and accuracy of credit risk models, despite increases in the
27
28 numbers of SMEs defaults. Similar to Altman et al. (2010) they found company demographics,
29
30 derogatory events and information about directors to be of significant value.
31
32

33
34 Regarding the modelling approaches, the overwhelming majority of studies reviewed above
35
36 used logistic regression. Other models included proportional odds or simple hazard model
37
38 (Michala et al., 2013; Fantazzini et al., 2009), Bayesian and classic panel models (Fantazzini
39
40 et al., 2009), random survival forests (Fantazzini & Figini, 2009), Support Vector Machines
41
42 (Martens et al., 2011).
43

44
45 In Italy Vallini et al. (2009) attempted to model SME defaults on a sample of small
46
47 firms from 2001- 2005 using profitability, liquidity and leverage ratios. Multiple discriminant
48
49 analysis was compared to logistic regression, and the latter was found to produce better
50
51 predictions. Later study by Ciampi & Gordini (2013) applied neural networks to the same
52
53 dataset and reported their superior performance as compared to algorithms used in the earlier
54
55 work. Both studies noted that credit scoring models could be built on accounting information,
56
57 yet predicting default for SMEs was much more difficult as compared to large enterprises, with
58
59 predictive accuracy decreasing in smaller firms segments.
60
61
62

1
2
3
4 Calabrese & Osmetti (2013) and Calabrese et al. (2013) applied GEV and BGEVA models
5
6 to the sample of Italian SMEs from 2006 to 2011 and found superior performance of both
7
8 models as compared to logistic regression. Variables found significant in predicting default
9
10 were again measures of profitability, leverage and liquidity.
11

12 The current paper extends the existing literature by looking at two countries in comparison
13
14 (Italy and the UK), by exploring SMEs failure in a more recent time period and by using more
15
16 comprehensive list of financial measures.
17
18

19 20 21 **3 Methodology** 22 23

24 When constructing a credit scoring model, three common problems are often mentioned: first,
25
26 low numbers of defaults, second, non-linear relationship between the response and predictors,
27
28 and third, missing values in predictor variables.
29

30 Logistic regression is the most commonly used model for credit scoring applications (e.g.,
31
32 Altman & Sabato, 2007; Becchetti & Sierra, 2002; Lin et al., 2012; Zavgren, 1998). As noted
33
34 above, the number of defaults in a sample is often very small (e.g., Kiefer, 2010; Lin et al.,
35
36 2012). King & Zeng (2001) commented on difficulties of obtaining unbiased probability es-
37
38 timates of event occurring in rare events situations. This is due to the fact that the char-
39
40 acteristics of defaults (events) are more informative than those of non-defaults. When there
41
42 is a small number of defaults, there might be insufficient information to produce appropriate
43
44 estimates of the default probability for values close to 1 (Calabrese & Osmetti, 2013). Wang
45
46 & Dey (2010) showed that using an asymmetric link function improved the model fit.
47
48

49 In cases of low default portfolios, the conservative (higher) estimates of default probabilities
50
51 might be preferred, and a flexible asymmetric link function can achieve such higher estimates
52
53 for defaulters in comparison with logistic regression, as shown in Calabrese & Osmetti (2015).
54
55

56 In order to choose the link function, we consider that defaulters' features are represented
57
58 by the tail of the response curve for values close to one. Furthermore, the Generalised Extreme
59
60 Value (GEV) distribution is used in literature (Kotz & Nadarajah, 2000; Falk et al., 2010) to
61
62

1
2
3
4 model the tail of a distribution. Therefore, to focus the attention on defaulters' characteristics,
5
6 Calabrese & Osmetti (2013) propose the quantile function of a GEV random variable as a
7
8 new link function

$$9 \quad \frac{[-\ln(PD_i)]^{-\tau} - 1}{\tau} = \eta_i = \alpha + \sum_{j=1}^p \beta_j x_{ji}, \quad (1)$$

10
11
12
13 where $\tau \in \mathfrak{R}$ is the tail parameter. As discussed, for instance, in Calabrese & Osmetti (2013),
14
15 depending on the value of τ , several special cases can be recovered; e.g., when $\tau \rightarrow 0$ the
16
17 GEV random variable follows a Gumbel distribution and its cumulative distribution is the
18
19 log-log function (Agresti, 2002). In this way, Calabrese & Osmetti (2013) propose the *GEV*
20
21 *regression model*.

22
23
24 Second, the logistic and the GEV (1) models assume a linear relationship between the
25
26 explanatory variables and the response η_i . These models can mask possibly interesting non-
27
28 linear patterns which can help improve our understanding of the underlying covariate-response
29
30 relationships and perhaps improve the prediction accuracy of the scoring model as well (Berg,
31
32 2007; Calabrese et al., 2013; Chuang & Lin, 2009; Gestel et al., 2005; Huang et al., 2006; Lee
33
34 & Chen, 2005; Lin et al., 2012; Ong et al., 2005). Therefore Calabrese et al. (2013) propose
35
36 the BGEVA model, an extension of the GEV model based on penalized regression splines to
37
38 flexibly determine covariate effects from the data.

39
40
41 In the GEV model, the right part of equation (1) is changed to obtain an additive model
42
43 given by

$$44 \quad \frac{[-\ln(PD_i)]^{-\tau} - 1}{\tau} = \alpha + \sum_{j=1}^p \beta_j s(x_{ji}), \quad (2)$$

45
46
47
48 where the $s_j(x_{ij})$ are unknown one-dimensional smooth functions of the continuous covariates
49
50 x_{ji} .

51
52
53 The smooth functions $s(x_{ij})$ in the model are approximated by a linear combination of
54
55 K_j known (e.g., cubic or thin plate regression) spline bases $b_k(x_{ji})$ and unknown regression
56
57

parameters, γ_{jk} (Wood, 2006; Calabrese et al., 2013):

$$s_j(x_{ji}) = \sum_{k=1}^{K_j} \gamma_{jk} b_k(x_{ji}).$$

Calculating $b_k(x_{ji})$ for k and each observation point gives K_j curves with different degrees of complexity which multiplied by some real valued parameters γ_{jk} and then summed to give an estimated curve for the smooth component (Ruppert et al., 2003). Replacing in model (2) the smooth terms with their regression spline expressions yields essentially a classic parametric model. Estimating the β_j parameters and the smooth functions $s(x_{ij})$ we can predict the default probabilities by using the inverse of the equation (2). The smooth functions show the existence of possible non-linear relationships between the response variable and the predictors and allow us to improve on the prediction results obtained using classic alternatives. The model is implemented in the R package *bgeva* (Marra et al., 2013) available for download from CRAN.

Third, SMEs may not provide full details of their financial statements (Sohn & Kim, 2013; Ciampi & Gordini, 2013), for this reason missing values could be a problem for scoring models for SMEs (Lin et al., 2012; Ciampi & Gordini, 2013). In the literature, missing values are classified into three types: Missing Completely At Random (MCAR), Missing At Random (MAR) and Not Missing At Random (NMAR). The missing values are MCAR if the probability that any variable is missing cannot depend on any other variable in the model of interest or on the potentially missing values themselves. If we have a single variable Z with missing data and a set of variables which is always observed X , the MCAR assumption can then be expressed by $P(I_z = 1|X, Z) = P(I_z = 1)$ where I is a dummy variable having a value of 1 if Z is missing and 0 if Z is observed. Therefore, the probability that Z is missing depends neither on the observed variables X nor on the possibly missing values of Z itself.

If the probability that Z is missing may depend on X , but it does depend on Z itself $P(I_z = 1|X, Z) = P(I_z = 1|X)$, the MAR assumption is satisfied. This means that MCAR

1
2
3
4 is a special case of MAR. If the data are MAR, it is possible to get optimal estimates of
5
6 parameters without directly modeling the missing data mechanism since the missing-data
7
8 mechanism is ignorable. Unfortunately the MAR assumption is not testable. Finally, if the
9
10 MAR assumption is violated, missing data are said to be NMAR.

11
12 There are several methods for handling missing values. The first is to delete cases with
13
14 any missing data on the variables of interest. This method often deletes a large fraction of
15
16 the sample and it is particularly suitable if the data are MCAR. When the data are MAR,
17
18 this procedure may introduce bias into parameter estimates, so the use of a different method
19
20 is preferable. The second method is to impute values for the missing covariates and carry
21
22 out the analysis as if the imputed values were observed data. A wide variety of methods falls
23
24 under the general heading of imputation, for example imputations based on the mean, on the
25
26 linear regression or on the maximum likelihood and EM algorithm (see Rubin (1976, 1977,
27
28 1987)).

29
30
31 One of the widely used approach in the latter method is multiple imputation, which was
32
33 proposed by Rubin (1987) and described in detail by Graham (2012). Multiple imputation
34
35 can be described as a three-step process. First, in order to capture the uncertainty in the
36
37 estimates of the missing values, more sets of plausible values for missing observations are
38
39 created. Each of these sets of plausible values can be used to 'fill-in' the missing values and
40
41 create a 'completed' dataset. Second, each of these datasets can be analysed using complete-
42
43 data methods. Finally, the results are combined, which allows the uncertainty regarding the
44
45 imputation to be taken into account. The multiple imputation requires that the missing values
46
47 are MAR. The advantage of these methods is that it can be applied to any type of data and
48
49 it is implemented in the conventional software. Moreover, it has optimal statistical properties
50
51 (see Rubin (1987); Graham (2012)).

52
53
54 In this paper we apply a multiple imputation based on an MCMC algorithm known as fully
55
56 conditional specification (Graham, 2012). The basic idea is to impute incomplete variables
57
58 one at time by linear regression, using the filled-in variable from one step as a predictor in all
59
60
61
62
63
64
65

subsequent steps. We have chosen this particular method since Florez-Lopez (2010) showed in application to credit scoring that it is superior to other methods of handling missing values.

Another approach to cope with missing values is based on so-called coarse-classification (Thomas et al., 2002). This procedure consists in dividing the values of a numeric predictor into categories or classes. Normally there are 10-20 fine classes initially produced for the range of ordered values from minimum to maximum. In this paper we divide the numeric predictors into 10 classes of approximately the same size (maintaining exactly the same size is not possible because of the varying numbers of missing values for different variables).

For each fine class a proportion of defaults (or bad accounts or simply Bads) is calculated, and adjacent categories can be further grouped together into coarse classes, if the default rates are sufficiently close. Missing values are entered as a separate category. Categories can be entered into the model as binary dummies or alternatively are transformed into Weights of Evidence (WoE):

$$WoE_i = \ln \left[\frac{b_i/g_i}{B/G} \right] = \ln \left(\frac{b_i G}{g_i B} \right), \quad (3)$$

where b_i is the number of bads (defaults) in category i of a variable, g_i is the number of goods (non-defaults) in category i , B is the total number of Bads, G is the total number of Goods in the sample.

The term (WoE) goes back to early days of computer science and information theory and is defined by Good (1950) as the weight of evidence (or degree of corroboration) in favour of a hypothesis H given by evidence (or information or an experiment outcome) E :

$$WoE = \ln \left[\frac{P(H|E)/(1 - P(H|E))}{P(H)/(1 - P(H))} \right]. \quad (4)$$

Equation (3) above is a generalisation of Equation (4). It has extensively been used in early classification algorithms and specifically in Naive Bayes classifier, please see Good (1985); Greiff (1999); Hand & Adams (2000); Hand et al. (2001). WoE approach can be criticized on the grounds of imposing the ordering of categories observed for each predictor taken separately

1
2
3
4 and not allowing for interactions between predictors (Thomas, 2009). There may also be a
5 concern about using the dependent variable in transforming a predictor. Despite its limitations
6 this transformation is widely used in practice (Anderson, 2007; Baesens, 2014; Siddiqi, 2006;
7 Thomas, 2000, 2009). An alternative to WoE approach consists in partitioning the variables
8 and then turning k partitions into k-1 dummy (0/1) variables. This approach does not impose
9 any ordering or dependency, but has a disadvantage of producing a lot of variables (Thomas,
10 2000, 2009). Banasik et al. (2003) compared WoE and dummy variable approaches and found
11 them similar. Following the latter paper and also Banasik & Crook (2007); Bijak & Thomas
12 (2012); Lin et al. (2012); Malik & Thomas (2010); Orton et al. (2015), and the wide-spread
13 industry practice, we use WoE in this paper.
14
15
16
17
18
19
20
21
22
23
24

25 Given the fact that logistic regression is the most commonly used approach in credit scoring
26 (Thomas et al., 2002), WoE is appealing since this transformation produces log odds measures
27 (same scale as logistic regression). Furthermore, log-odds of each category are compared to
28 that of the whole sample: positive values would indicate riskier classes and negative values -
29 more creditworthy customers.
30
31
32
33
34

35 We use this approach as the benchmark to compare the performance of alternative methods
36 to cope with missing values (multiple imputation) and non-linearity (BGEVA model).
37
38
39
40
41

42 **4 Empirical Analysis**

43 **4.1 Data description**

44
45
46 The empirical analysis is based on explanatory variables from 2010 to predict the default in
47 2011 for 39,785 UK SMEs and 154,934 Italian SMEs. The data are from AMADEUS-Bureau
48 van Dijk (BvD), a database of comparable financial and business information on Europe's
49 public and private companies. The time horizon considered here is of extreme interest as it
50 includes the European sovereign debt crisis of 2011. In summer 2011 interest rates on Italian
51 national debt went out of control.
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 The definition of SME by the European Commission is adopted. That is, a business must
5 have an annual turnover of less than 50 million of Euro, a balance sheet total less than 43 mil-
6 lion of Euro and the number of employees should not exceed 250 ([http://ec.europa.eu/enterprise/](http://ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/sme-definition/index.htm)
7 [policies/sme/facts-figures-analysis/sme-definition/index.htm](http://ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/sme-definition/index.htm)). Furthermore, the number of
8 subsidiaries is capped at 6, in accordance with Lu & Beamish (2001), and the number of
9 directors is 10 maximum, consistent with Gabrielsson (2007); Michala et al. (2013).

10
11
12 In this work, we consider a default to have occurred when a specific SME enters a
13 bankruptcy or a liquidation procedure. Moreover, a SME is classified as default also if it
14 is active and it has not paid a debt (classified as default of payment by BvD) or it is in admin-
15 istration or receivership or under a scheme of arrangement (defined as insolvency proceedings
16 by BvD). On the contrary, non-defaulters include active and dormant SMEs (only 29 for both
17 samples). A dormant company is still registered, but has no significant activity (and no signifi-
18 cant accounting transactions during the accounting period). Consistent with previous studies
19 (Altman & Sabato, 2007; Altman et al., 2010; Pederzoli et al., 2013) we exclude dissolved
20 firms that no longer exist as a legal entity, but the reason for *dissolution* is not specified. This
21 is in line with the objective of this paper that models the probability of going bankrupt using
22 publicly available information. Dissolved category comprises SMEs that may not necessarily
23 experience financial difficulties, they may stop trading because the owner retires or for similar
24 reasons. *The descriptive statistics for dissolved category is shown in Table 8 and Table 9 in*
25 *the Appendix.* Future research can investigate dissolved as a separate category.

26
27
28 The use of the common database has ensured the availability of the common set of variables
29 measured in the same way for both countries. We used financial ratios that have been found
30 important in previous research on SMEs (Altman & Sabato, 2007; Lin et al., 2012; Michala
31 et al., 2013). Adopting the classification of variables suggested in Altman & Sabato (2007)
32 the variables in this research covered all five major groups usually used:

- 33 • Leverage (e.g. Gearing, Solvency ratio);
- 34 • Liquidity (e.g. Current ratio, Liquidity ratio, Shareholder liquidity ratio);

- Profitability (e.g. EBITDA margin, Profit margin, ROCE, ROE);
- Coverage (e.g. Interest cover);
- Activity /Scale/Size (e.g. Total assets, Shareholder funds, No of employees, No of directors, No of subsidiaries).

Following Michala et al. (2013) who found cash flow management significant in predicting default, we also include cash flow based measures (e.g. Cash flow, Cash flow / Operating revenue). The variables have been checked for linear dependence, and highly collinear ones have not been used in the analysis. Table 1 presents short and full names of the variables initially considered and some descriptive statistics on the training sample.

Table 1 around here.

The SMEs in the UK sample are larger as compared to Italian SMEs in terms of Total assets, Operating revenue, No of employees, No of directors. This is consistent with the EU statistics reported in Sections 1-2. The summary statistics for Age and No of subsidiaries are similar for the two countries. The UK businesses have higher liabilities, but profitability is also higher. The Italian companies show better Cash flow and lower debt. Despite using the common source of the data, the percentages of missing values are different across the countries. For Italy, the variable with the highest number of missing is Cash flow / Operating revenue, with 19.5% missing. For the UK, the problem is much more acute, the highest percentage of missing is 59.2% for ROCE. This has an effect on the results, depending on how missing values have been treated, as can be seen from Tables 2 and Table 3 that show the variables that are significant at 10% level or lower across the models.

Table 2 around here

1
2
3
4 Table 3 around here
5
6

7 8 **4.2 Predictive accuracy** 9

10 To avoid sample dependency, the predictive accuracy for the models was tested on control
11 samples, i.e. we used out-of-sample tests. For each country the whole dataset was split into
12 training (70%) and control (30%) samples using a stratified random sampling with stratifica-
13 tion on default indicator. Measures of predictive accuracy used include mean absolute error
14 (MAE), mean square error (MSE) and Area under the ROC curve (AUC). MAE and MSE are
15 standard measures of predictive accuracy in forecasting studies. Obviously, scoring models
16 with lower MSE and MAE should forecast defaults and non-defaults more accurately. For a
17 bank it is much more costly to classify an SME as a non-defaulter when it is a defaulter than
18 the opposite. If a defaulter is classified as a non-defaulter, then it will be accepted for credit,
19 which will subsequently be lost (in part or as a whole). Yet when a non-defaulter is classified
20 as a defaulter, it is only a lost opportunity. Therefore, in this study MSE and MAE are re-
21 ported for defaults only and they are denoted by MSE^+ and MAE^+ . AUC is the most popular
22 measure of model performance in credit scoring (Thomas et al., 2002) that summarises the
23 ability of the model to rank-order the risk correctly over the whole range of predicted PDs.
24 Higher value indicate better performance.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 Table 4 around here
45
46

47 Table 5 around here
48
49

50
51
52
53 Tables 4 and 5 summarise the results¹ for the UK and Italian models for imputed and Weights
54 of Evidence (WoE) data.
55

56
57 Considering WoE approach on the UK data, the GEV model shows better performance on
58

59 ¹To obtain these results we use SPSS for imputed missing values and the package "bgeva" of R-program.
60
61
62

1
2
3
4 MAE and AUC than the logistic model, although the latter has lower MSE (Table 4). More-
5
6 over, by applying the non-parametric model (BGEVA) the performance and MAE⁺ improves
7
8 further, and on MSE⁺ it becomes the same as for additive logistic. This fact justifies the use
9
10 of a non-parametric credit scoring model that can capture non-linear relationships between
11
12 the accounting characteristics of SMEs and response.
13

14
15 As for imputed values on the UK data, the best MAE⁺ and MSE⁺ are for BGEVA, whilst
16
17 the best AUC is shared between BGEVA and additive logistic model. It can be argued that
18
19 the improvement provided by additive models over GEV and logistic on WoE is modest. This
20
21 is not surprising, since one of the objectives of WoE and coarse-classification is to cope with
22
23 non-linearities (see, e.g. Thomas (2009)). Still it appears there is some benefit from applying
24
25 the semi-parametric approach, albeit it is less pronounced as compared to improvement of
26
27 BGEVA over GEV on imputed data. This further emphasises the advantage of BGEVA in
28
29 forecasting defaults in low default portfolios that performs well on both methods of treating
30
31 the missing values.
32

33
34 Considering WoE approach on Italian data (Table 5), we observe results similar to the UK
35
36 models. BGEVA has the best MAE⁺ and MSE⁺, whilst additive logistic produces slightly
37
38 higher AUC, but the difference is negligible. For Italian imputed values the results are mixed.
39
40 The additive logistic model shows the lowest values of the MAE⁺ and MSE⁺, whilst the GEV
41
42 and logistic models show higher values of the AUC.
43

44
45 The comparison of the predictive accuracy between the countries should be interpreted
46
47 with caution due to the different sample sizes, different proportions of missing values and
48
49 different number of significant variables (as discussed in the next section). Since the UK
50
51 sample size is smaller than the Italian one and the percentage of UK missing values is higher
52
53 than for Italy (see Table 1), one can expect a decrease in the predictive accuracy. However,
54
55 for completeness it could be stated that all models for Italy have better performance than the
56
57 UK models. Moreover, the Italian best model (BGEVA) has also a lowest MAE⁺.
58

59 It should also be noted that WoE coding provides better performance as compared to
60
61
62
63
64
65

1
2
3
4 Imputation with the only exception of MAE⁺ of BGEVA for the UK.
5

6 In conclusion, the empirical results confirm that the BGEVA model performs well for SMEs
7 default forecasting for both countries. This can be attributed to the fact that the linearity
8 assumption is not supported by the data of both countries, as will be discussed in the next
9 section.
10
11
12
13
14

15 16 **4.3 Comparison of risk predictors between Italian and UK SMEs** 17

18 There are differences between the countries in terms of significant variables and their number
19 depending on the model/approach used. Whilst logistic regression for both countries and GEV
20 model for Italy show the same number of variables irrespective of imputation or WoE, there
21 are differences in model composition even in these cases. For example, in logistic regression for
22 the UK - Cash flow, Interest cover and Operating revenue are significant with WoE coding, but
23 not with Imputation; yet with Imputation the following variables become significant: Profit
24 margin, Shareholder funds and Total assets. For the rest of models the numbers of significant
25 variables differ with the extreme cases of GEV and BGEVA for the UK, where WoE coding
26 increases the number of significant variables from 11 to 20. This may be interpreted as
27 suggesting that at least for some variables values cannot be assumed to be missing at random,
28 therefore WoE increase the number of significant variables.
29
30
31
32
33
34
35
36
37
38
39
40
41

42 Only two variables consistently appear across all 16 models for the two countries: No of
43 directors and Solvency ratio (Tables 2 and 3). No of subsidiaries appear in all models, but
44 one. Profit margin and Shareholder funds enter 14 models. Other frequent variables that are
45 significant at 10 per cent level or lower across all 16 models for the two countries are Liquidity
46 ratio (13), Age (12), EBITDA margin (12), No of employees (12), Operating revenue (12),
47 Cash flow / Operating revenue (10), Total assets (10), ROE (10). When looking at most
48 frequent significant variables for each country separately (e.g. common variables that are in
49 more than half of the models for each country) these include No of directors, Solvency ratio,
50 No of subsidiaries, Profit margin, Shareholder funds and Liquidity ratio. This confirms the
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 results from previous research that suggests measures of profitability, leverage and liquidity
5 are important (Altman & Narayanan, 1997; Altman et al., 2010; Michala et al., 2013). Share-
6 holder funds can be interpreted as the interest the shareholders have in the company, and
7 also the ability of the company to raise funds for growth/expansion. Solvency ratio emphasis
8 the importance of the proportion of Shareholder funds in the assets of the company. No of
9 directors and No of subsidiaries may be interpreted as proxies for company size and the scale
10 of the activity, with No of directors also acting as a crude proxy for quality of management
11 (assuming more directors would mean better management).
12
13
14
15
16
17
18
19
20
21
22

23 Table 6 around here
24
25
26

27 Table 7 around here
28
29
30

31 Despite the commonality reported above, there are some interesting differences between the
32 countries. The most notable one is the fact that Gearing is significant in all UK models, whilst
33 not being significant in Italy at all. This suggests the importance of the firm's ability to pay
34 both long-term debt and short-term one in the UK. For Italy measures of profitability are
35 relatively more prominent: EBITDA margin and ROE appear in almost all Italian models, in
36 addition to Profit margin which is common to both countries. Age and No of employees are
37 twice more frequent in the UK models. Age has been previously found important in Altman &
38 Sabato (2007). No of employees indicates the size of the company or its scale. Financial scale
39 for Italy is most frequently represented by Operating revenue, which appears in all Italian
40 models, but only in half of the UK ones. Cash flow/ Operating revenue is also present in all
41 Italian models.
42
43
44
45
46
47
48
49
50
51
52
53

54 As an example of more detailed cross-country differences, consider the estimates of BGEVA
55 model on imputed values presented in Tables 6. The interpretation of WoE is less straightfor-
56 ward since it requires the information on category boundaries and WoE values. This informa-
57
58
59
60
61
62
63
64
65

tion and details of other models are available on request. Financial measures common to both countries include ratios of profitability (Profit margin), leverage (Solvency ratio), liquidity (Liquidity ratio) and scale (Shareholder funds, Total assets). In addition, there are common non-financial variables across the two countries: Age, No of directors, No of employees, No of subsidiaries. This fact emphasises the value of non-financial information in modelling SMEs and confirms some previous research (Altman et al., 2010).

Tables 6 and 7 report the estimation results of the parametric and non-parametric components of the BGEVA model for the two countries and for multiple imputation. Some of the covariate effects are reported in the parametric part of the BGEVA model since their smooth function estimates were linear. Explanatory variables significant deviations from the linearity assumption are reported in the smooth terms part. The variables show different degrees of non-linearity (Edf). The parameter Edf (degrees of freedom) in Tables 6 and 7 controls the smoothness of the curve. The variables with Edfs equal to 1 show linear smooth function so they are reported in the parametric part. The estimated smooth parameters that exhibit Edfs considerably greater than 1 are reported in smooth terms part. Larger Edf allows a very flexible curve, e.g., a curve that can have multiple local maxima and minima. The values of degrees of freedom are estimated from the data. The most interesting smooth terms are displayed in the Figures 1 and 2. In line with the interpretation for the parametric components, if the estimated smooth function of a covariate is decreasing then the estimated *PD* decreases when the explanatory variable increases, and vice versa.

Figure 1 around here

Figure 2 around here

There is some commonality between the countries with Liquidity ratio and Age being non-linear for both countries. No of directors and Total assets exhibit non-linear relationship with

1
2
3
4 the response for the UK, but not for Italy, on the contrary, Cash Flow and No of employees
5
6 show non-linear patterns for Italy only.
7

8 Consider Liquidity ratio that shows a non-linear relationship for both countries (Figures
9
10 1 and 2). For Italy when this variables increases, the PD decreases (although in a non-linear
11
12 way), in accordance with the expectations and prior research by Pederzoli et al. (2013). Yet
13
14 for the UK the relationship is more complex. Up to 30 and from 75 the relationship of this
15
16 covariate to PD is negative (as expected). However, in the middle section it is the opposite:
17
18 increasing values of Liquidity ratio signal increasing chances of default. This may be related
19
20 to difficulties in getting credit for SMEs, if Current Liabilities in denominator are decreasing.
21
22

23 Previous research summarised in Section 2 did not use exactly the same ratio, yet Altman
24
25 et al. (2010) report a negative relationship between a similar variable (Current ratio) and the
26
27 PD. It should be noted though, that the authors did not comment on potential non-linearity.
28
29 For German SMEs Fantazzini et al. (2009) and Figini & Giudici (2011) observed a counter-
30
31 intuitive sign for Liabilities ratio and explained it by the fact that many small business owners
32
33 cover their debts from external sources.
34

35 Examples of variables that show non-linear relationships and are not common for the two
36
37 countries are Total Assets for the UK and No of Employees for Italy, both can be interpreted
38
39 as proxies for SME size. From Figure 2 looking at Total assets we can deduce that the UK
40
41 small and micro enterprises show higher default risk, in line with Fantazzini et al. (2009) for
42
43 German SMEs. Then for companies with Total assets higher than 20 million euros, when this
44
45 variable increases the PD decreases. Altman et al. (2010) also noted the non-linear nature
46
47 of Total assets. Finally, from the plot for Number of Employees (Figure 1) Italian small
48
49 and micro enterprises have higher PD when the number of employees increases. For medium
50
51 enterprises this relationship becomes negative, although the confidence intervals are wide.
52
53 These results highlight some interesting patterns observed from the data, yet further research
54
55 would be beneficial in order to fully understand the implied relations.
56
57
58
59
60
61
62
63
64
65

5 Conclusions and extensions

This paper has compared predictors of SMEs insolvencies across the UK and Italy, using publicly available information from 2010 to model the company status in 2011. The choice of the time period after the credit crisis makes this comparison particularly relevant, due to different economic situations in the two countries. Whilst Italy was experiencing high interest rates for its national debt, that was not the case in the UK despite the latter showing low economic growth. There are also differences across the two countries in the relative importance that SMEs play in the two economies, as discussed in Section 2. Despite these differences, there were some financial measures significant in predicting insolvency. These included measures of profitability, leverage, liquidity and scale. In addition, there was some commonality in non-financial measures, thus highlighting the importance of soft information for analysis of SME performance. As for the differences, profitability measures are significant more frequently for Italy, whilst for the UK Gearing is a significant predictor, not featuring in Italian models.

A number of different modelling approaches have been explored in order to improve predictive accuracy. Generalised Extreme Value (GEV) regression with asymmetric link function was applied in comparison to the logistic regression, which is a standard approach in credit risk modelling. The assumption of non-linearity was relaxed through application of BGEVA, non-parametric additive model based on the GEV link function. In addition, two methods of handling missing values were compared: multiple imputation and Weights of Evidence (WoE) transformation. The results suggest that the best predictive performance is obtained by BGEVA, thus implying the necessity of taking into account the volume of defaults and non-linear patterns when modelling SME insolvencies. WoE generally showed better prediction as compared to imputation, suggesting that missing values could be informative.

This study presents an initial attempt to understand the cross-country drivers of SMEs insolvencies, and is exploratory in the general approach adopted. Further extensions could include exploration of additional countries and additional variables, in particular, of non-financial nature, but this depends on the data availability. Causal relations through structural

1
2
3
4 equation models can be investigated. On the practical side, it would be of interest to consider
5
6 predictors significant to both countries and construct a generic model with the objective of
7
8 comparing it to a country-specific model. Finally, different groups of SMEs that go out of
9
10 business can be explored, e.g. dissolved.

11
12 In this paper we perform a cross-sectional analysis. As a possible direction for future
13
14 research, we are planning to extend the BGEVA model to a panel data setting, and compare
15
16 the performance of SMEs for the two countries across time.
17
18

19 20 21 **Acknowledgements**

22
23
24 The Authors gratefully acknowledge the help of Armando Benincasa and Carlotta Prete at
25
26 Bureau Van Dijk in providing the data. They also acknowledge comments and suggestions by
27
28 the anonymous referees, that have helped improving the paper. Dr Andreeva is grateful to
29
30 The Royal Society of Edinburgh for financial support.
31
32

33 34 35 **References**

- 36
37
38 Agresti, A. (2002). *Categorical Data Analysis*. Wiley, New York.
- 39
40
41 Altman, E. & Narayanan, P. (1997). An international survey of business failure classification
42
43 models. *Financial Markets, Institutions and Instruments*, 6, 1–57.
- 44
45
46 Altman, E. & Sabato, G. (2007). Modeling credit risk for smes: Evidence from the us market.
47
48 *ABACUS*, 43(3), 332–357.
- 49
50
51 Altman, E., Sabato, G., & Wilson, N. (2010). The value of non-financial information in small
52
53 and medium-sized enterprise risk management. *The Journal of Credit Risk*, 6, 1–33.
- 54
55
56
57 Anderson, R. (2007). *The Credit Scoring Toolkit*. Oxford University Press: Oxford.
- 58
59
60 Baesens, B. (2014). *Analytics in a Big Data World*. Wiley, Wiley and SAS Business Series.
61
62

- 1
2
3
4 Banasik, J. L. & Crook, J. N. (2007). Reject inference, augmentation, and sample selection.
5
6 *European Journal of Operational Research*, 183(3), 1582–1594.
7
8
9 Banasik, J. L., kin, J. N., & Thomas, L. C. (2003). Sample selection bias in credit scoring
10 models. *Journal of Operational Research Society*, 54(8), 822–832.
11
12
13
14 Becchetti, L. & Sierra, J. (2002). Bankruptcy risk and productive efficiency in manufacturing
15 firms. *Journal of Banking and Finance*, 27, 2099–2120.
16
17
18
19 Berg, D. (2007). Bankruptcy prediction by generalized additive models. *Applied Stochastic*
20 *Models in Business and Industry*, 23, 129–143.
21
22
23
24 Bijak, K. & Thomas, L. C. (2012). Does segmentation always improve model performance in
25 credit scoring? *Expert Systems with Applications*, 39(3), 2433–2442.
26
27
28
29 Calabrese, R., Marra, G., & Osmetti, S. A. (2013). Bankruptcy prediction of small and medium
30 enterprises using a flexible binary generalized extreme value model. *ArXiv*, 1307.6081, 1–28.
31
32
33
34 Calabrese, R. & Osmetti, S. (2015). Improving forecast of binary rare events data: A gam-
35 based approach. *Journal of Forecasting*, Forthcoming.
36
37
38
39 Calabrese, R. & Osmetti, S. A. (2013). Modelling sme loan defaults as rare events: the
40 generalized extreme value regression model. *Journal of Applied Statistics*, 40(6), 1172–1188.
41
42
43
44 Chuang, C. L. & Lin, R. H. (2009). Constructing a reassigning credit scoring model. *Expert*
45 *Systems with Applications*, 36, 1685–1694.
46
47
48
49 Ciampi, F. & Gordini, N. (2013). Small enterprise default prediction modeling through arti-
50 ficial neural networks: An empirical analysis of italian small enterprises. *Journal of Small*
51 *Business Management*, 51(1), 23–45.
52
53
54
55
56 Dietsch, M. & Petey, J. (2004). Should sme exposure be treated as retail or as corporate
57 exposures? a comparative analysis of default probabilities and asset correlation in french
58 and german smes. *Journal of Banking and Finance*, 28, 773–788.
59
60
61
62

- 1
2
3
4 EC (2012a). Sba fact sheet 2012 for italy. *Enterprise and Industry working paper*.
5
6
7 EC (2012b). Sba fact sheet 2012 for uk. *Enterprise and Industry working paper*.
8
9
10 Falk, M., Haler, J., & Reiss, R. (2010). *Laws of Small Numbers: Extremes and Rare Events*.
11 Springer.
12
13
14
15 Fantazzini, D. & Figini, S. (2009). Random survival forests models for sme credit risk mea-
16 surement. *Methodology and Computing in Applied Probability*, 11, 29–45.
17
18
19
20 Fantazzini, D., Figini, S., Giuli, E. D., & Giudici, P. (2009). Enhanced credit default models
21 for heterogeneous sme segments. *Journal of Financial Transformation*, 25(1), 31–39.
22
23
24
25 Figini, S. & Giudici, P. (2011). Statistical merging of rating models. *Journal of the Operational*
26 *Research Society*, 62(6), 1067–1074.
27
28
29
30 Gabriellsson, J. (2007). Boards of directors and entrepreneurial posture in medium-size com-
31 panies. putting the board demography approach to a test. *International Small Business*
32 *Journal*, 25, 511–537.
33
34
35
36
37 Gestel, T. V., Baesens, B., Dijcke, P. V., Suykens, J. A. K., Garcia, J., & Alderweireld, T.
38 (2005). Linear and non-linear credit scoring by combining logistic regression and support
39 vector machines. *Journal of Credit Risk*, 1(4), 31–60.
40
41
42
43
44
45 Good, I. J. (1950). *Probablity and the Weighing of Evidence*. London: Charles Grifin; New
46 York: Hafners.
47
48
49
50 Good, I. J. (1985). Weight of evidence: a brief survey. *Bayesian Statistics*, 27, 249–270.
51
52
53
54
55
56 Greiff, W. R. (1999). *Maximum Entropy, Weight of Evidence and Information Retrieval*. PhD
57 Thesis: University of Massachusetts Amherst.
58
59
60
61
62
63
64
65

- 1
2
3
4 Hand, D. J. & Adams, N. M. (2000). Defining attributes for scorecard construction in credit
5 scoring. *Journal of Applied Statistics*, 27(5), 527–540.
6
7
8
9 Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. MIT Press.
10
11
12 Huang, J. J., Tzeng, J. H., & Ong, C. S. (2006). Two-stage genetic programming (2sgp) for
13 the credit scoring model. *Applied Mathematics and Computation*, 174, 1039–1053.
14
15
16
17 Ihua, U. B. (2009). Smes key failure-factors: a comparison between the united kingdom and
18 nigeria. *Journal of Social Science*, 18(3), 199–207.
19
20
21
22 Kiefer, N. M. (2010). Journal of business and economic statistics. *Journal of Business Finance*
23 *& Accounting*, 28(2), 320–328.
24
25
26
27 King, G. & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9,
28 321–354.
29
30
31
32 Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions. Theory and Applications*.
33 Imperial College Press, London.
34
35
36
37 Lee, T. S. & Chen, I. F. (2005). A two-stage hybrid credit scoring model using artificial neural
38 networks and multivariate adaptive regression splines. *Expert Systems with Applications*,
39 28, 743–752.
40
41
42
43
44
45 Lin, S. M., Ansell, J., & Andreeva, G. (2012). Predicting default of a small business using
46 different definitions of financial distress. *Journal of the Operational Research Society*, 63,
47 539–548.
48
49
50
51
52 Lu, J. W. & Beamish, P. W. (2001). The internationalization and performance of smes.
53 *Strategic Management Journal*, 22, 565–586.
54
55
56
57 Lussier, R. & Halabi, C. (2010). A three-country comparison of the business success versus
58 failure prediction model. *Journal of Small Business Management*, 48(3), 360–377.
59
60
61
62

- 1
2
3
4 Malik, M. & Thomas, L. C. (2010). Modelling credit risk of portfolios of consumer loans.
5
6 *Journal of the Operational Research Society*, 61(3), 411–420.
7
8
9 Marra, G., Calabrese, R., & Osmetti, S. A. (2013). *bgeva: Binary Generalized Extreme Value*
10
11 *Additive Models*. R package version 0.2.
12
13
14 Martens, D., Vanhoutte, C., Winne, S. D., Baesens, B., Sels, L., & Mues, C. (2011). Identifying
15
16 financially successful start-up profiles with data mining. *Expert Systems with Applications*,
17
18 38, 5794–5800.
19
20
21 Michala, D., Grammatikosa, T., & Filipea, S. F. (2013). Forecasting distress in european sme
22
23 portfolios. *LSF Research Working Paper Series*, (13-02).
24
25
26 Ong, C. S., Huanga, J. J., & Tzeng, G. H. (2005). Building credit scoring models using
27
28 genetic programming. expert systems with applications. *Expert Systems with Applications*,
29
30 29, 41–47.
31
32
33 Orton, P., Ansell, J., & Andreeva, G. (2011). Recent developments in commercial scoring.
34
35 *Credit Scoring & Credit Control XII Conference*, Edinburgh, UK.
36
37
38 Orton, P., Ansell, J., & Andreeva, G. (2015). Exploring the performance of small- and
39
40 medium-sized enterprises through the credit crunch. *Journal of the Operational Research*
41
42 *Society*, 66, 657–663.
43
44
45 Pederzoli, C., Thoma, G., & Torricelli, C. (2013). Modelling credit risk for innovative smes:
46
47 the role of innovation measures. *Journal of Financial Services Research*, 44, 111–129.
48
49
50 Rubin, D. (1976). Inference and missing data. *Biometrika*, (63), 581592.
51
52
53 Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.
54
55
56 Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in
57
58 sample surveys. *Journal of the American Statistical Association*, 72, 538543.
59
60
61
62
63
64
65

- 1
2
3
4 Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge
5 University Press, London.
6
7
8
9 Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Jour-*
10 *nal of Business*, 74(1), 101124.
11
12
13 Siddiqi, N. (2006). *Credit risk scorecards : developing and implementing intelligent credit*
14 *scoring*. Wiley: New Jersey.
15
16
17
18 Sohn, S. Y. & Kim, Y. S. (2013). Behavioral credit scoring model for technology-based firms
19 that considers uncertain financial ratios obtained from relationship banking. *Small Business*
20 *Economics*, 41(4), 931–943.
21
22
23
24
25 Thomas, L., Edelman, D., & Crook, J. C. (2002). *Credit Scoring and Its Applications*. Society
26 for Industrial and Applied Mathematics, Philadelphia.
27
28
29
30 Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk
31 of lending to consumers. *International Journal of Forecasting*, 16(2), 149–172.
32
33
34
35 Thomas, L. C. (2009). *Consumer Credit Models*. Oxford University Press: Oxford.
36
37
38
39 Vallini, C. F., Ciampi, F., Gordini, N., & Benvenuti, M. (2009). Are credit scoring modeld
40 able to predict small enterprise default? statistical evidence from italian small enterprises.
41 *International Journal of Business & Economics*, 8(1), 3–18.
42
43
44
45 Wang, X. & Dey, D. K. (2010). Generalized extreme value regression for binary response
46 data: An application to b2b electronic payments system adoption. *The Annals of Applied*
47 *Statistics*, 4(4), 2000–2023.
48
49
50
51
52 Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall,
53 Boca Raton.
54
55
56
57 Zavgren, C. (1998). The prediction of corporate failure: the state of the art. *Journal of*
58 *Accounting Literature*, 2, 1–37.
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

6 Appendix

Table 8 around here

Table 9 around here

Table(s)

<i>Variables</i>		<i>Italy, n=106967</i>			<i>UK, n=27132</i>		
<i>Short name</i>	<i>Description</i>	<i>% Missing</i>	<i>Mean</i>	<i>Std Dev</i>	<i>% Missing</i>	<i>Mean</i>	<i>Std Dev</i>
Age	Age of the company, months	0.2	207.55	156.072	0.2	206.59	192.983
Capital	Capital, th EUR	0.2	203.46	816.490	2.0	331.96	2689.960
Cash_flow	Cash flow, th EUR	2.4	109.23	617.864	21.4	47.69	40224.523
Cash_flow_oprev	Cash flow / Operating revenue,%	19.5	7.00	8.250	39.4	11.98	14.668
Current_liab	Current liabilities, th EUR	0.2	1606.02	3045.892	1.6	2368.52	6464.733
Current_ratio	Current assets/Current liabilities,%	0.3	1.76	2.767	3.7	4.75	10.125
EBITDA_Margin	EBITDA/Operating revenue, %	3.5	6.79	14.564	23.2	8.64	20.992
Gearing	(Long term liab. + Short term loans)/ Shareholders funds, %	16.5	188.34	220.371	20.7	76.62	149.386
Interest_cover	P(L) before interest/ Interest paid, %	8.2	26.13	96.408	57.6	39.52	119.958
Liquidity_ratio	(Current assets - Stock)/Current liab.	0.3	1.36	2.392	5.0	4.53	10.119
Loans	Loans, th EUR	0.2	437.09	1300.625	3.3	1025.80	4005.364
Net_income	Net income, th EUR	0.2	11.46	635.865	3.2	109.074	4357.018
No_directors	Number of current directors/managers	0.0	1.33	1.810	0.0	4.74	2.555
No_employees	Number of employees	0.2	13.52	22.239	2.2	37.19	47.120
No_subsidiaries	No of recorded subsidiaries	0.0	0.41	0.872	0.0	0.47	0.977
Noncurrent_liab	Non-current liabilities, th EUR	0.2	613.73	1602.776	1.7	1001.87	4763.591
Op_rev	Operating revenue (Turnover), th EUR	0.2	2998.49	5457.546	1.7	6152.03	8545.211
PL_beforetax	Profit (Loss) before tax, th EUR	0.2	56.38	606.305	3.0	179.731	4391.610
Profit_employee	Profit per employee, th EUR	2.1	10.35	60.350	7.0	28.41	182.253
Profit_margin	P(L) before tax/ Operating revenue, %	1.9	1.03	14.395	7.8	7.66	25.868
ROCE	P(L) before tax/ (Total assets - Cur. liab.)	6.5	10.77	57.921	59.2	19.08	81.989
ROE	P(L) before tax/ Shareholder funds, %	7.7	13.36	97.496	18.0	25.85	109.174
Share_liquidity_ratio	Shareholders funds/ Long term liab., %	2.2	7.46	37.293	52.0	36.03	105.713
Sharehold_funds	Shareholders funds, th EUR	0.2	865.00	2381.846	1.6	1495.01	7183.142
Solvency_ratio	Shareholders funds/Total assets, %	1.1	23.35	24.527	6.3	45.43	38.256
Tot_assets	Total assets, th EUR	0.2	3084.70	5422.522	1.8	4860.58	6705.262

Table 1: Descriptive statistics for training samples

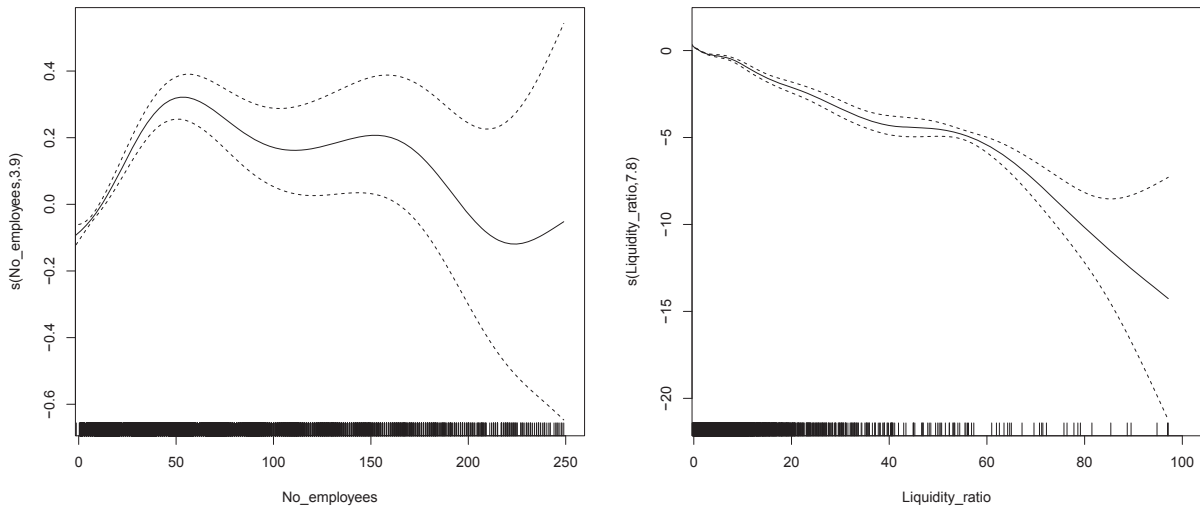


Figure 1: Smooth component estimates of the 2 (out of 4) continuous variables that exhibit a non-linear pattern. These were obtained from applying the BGEVA model on the Italian SME data. Results are on the scale of the predictor. The plot show the 95% confidence intervals. The numbers in brackets in the y-axis captions are the estimated degrees of freedom (*Edf*) of the smooth curves.

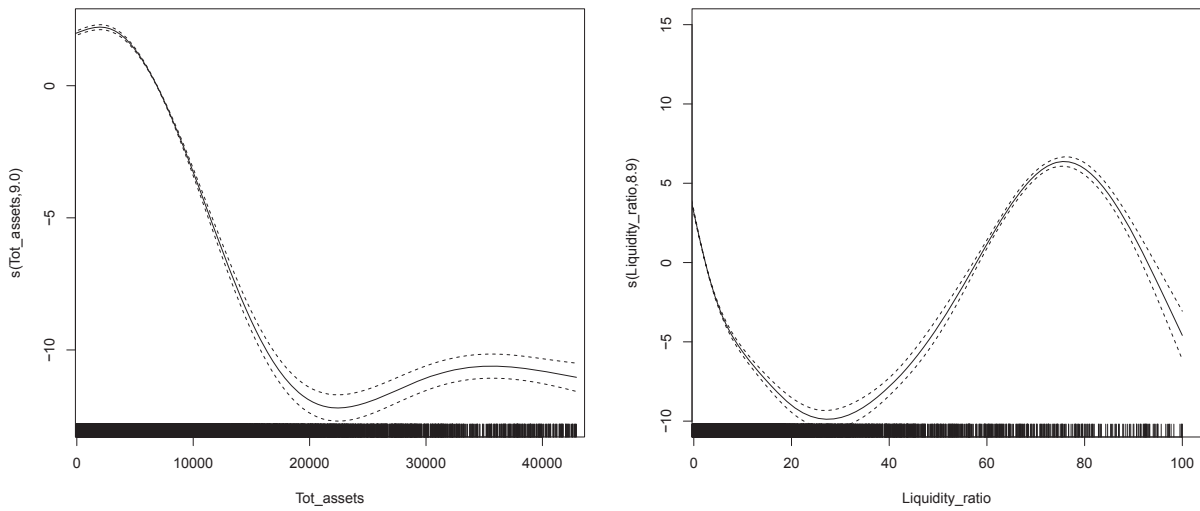


Figure 2: Smooth component estimates of the 2 (out of 4) continuous variables that exhibit a non-linear pattern. These were obtained from applying the BGEVA model on the UK SME data. Results are on the scale of the predictor. The plot show the 95% confidence intervals. The numbers in brackets in the y-axis captions are the estimated degrees of freedom (*Edf*) of the smooth curves.

<i>Variables</i>	<i>Logistic regression</i>				<i>Additive Logistic regression</i>				<i>times in all 16 models</i>
	<i>Italy</i>		<i>UK</i>		<i>Italy</i>		<i>UK</i>		
	<i>Imp</i>	<i>Woe</i>	<i>Imp</i>	<i>Woe</i>	<i>Imp</i>	<i>Woe</i>	<i>Imp</i>	<i>Woe</i>	
<i>Age</i>	X	0	X	X	SX	0	SX	X	12
<i>Cash_flow</i>	X	0	0	X	SX	0	0	SX	8
<i>Cash_flow_oprev</i>	X	X	0	0	X	X	0	0	10
<i>Current_ratio</i>	0	0	0	0	X	0	0	0	2
<i>EBITDA_Margin</i>	X	X	0	X	X	X	0	X	12
<i>Gearing</i>	0	0	X	X	0	0	X	X	8
<i>Interest_cover</i>	0	X	0	X	0	SX	0	X	8
<i>Liquidity_ratio</i>	X	X	X	X	SX	0	SX	X	13
<i>Net_income</i>	0	X	0	0	0	SX	0	0	6
<i>No_directors</i>	X	X	X	X	X	X	SX	X	16
<i>No_employees</i>	X	0	X	X	SX	0	X	X	12
<i>No_subsidiaries</i>	X	X	X	X	X	X	X	X	16
<i>Op_rev</i>	X	X	0	X	X	X	0	SX	12
<i>PL_beforetax</i>	0	X	0	0	0	SX	0	X	7
<i>Profit_margin</i>	X	X	X	0	X	X	X	0	14
<i>ROCE</i>	0	X	0	0	0	SX	0	0	6
<i>ROE</i>	X	X	0	0	X	0	0	SX	9
<i>Sharehold_funds</i>	X	X	X	0	X	X	X	0	14
<i>Shareh_liquidity_ratio</i>	0	0	0	0	0	0	X	0	5
<i>Solvency_ratio</i>	X	X	X	X	X	X	X	X	16
<i>Tot_assets</i>	X	0	X	0	X	0	SX	0	10

Table 2: Significant variables across the countries for logistic and additive logistic models. X - the variable is significant at 10% s.l. or lower; SX - the smooth term of the variable is significant at 10% s.l. or lower

<i>Variables</i>	<i>GEV model</i>				<i>BGEVA model</i>				<i>times in all 16 models</i>
	<i>Italy</i>		<i>UK</i>		<i>Italy</i>		<i>UK</i>		
	<i>Imp</i>	<i>Woe</i>	<i>Imp</i>	<i>Woe</i>	<i>Imp</i>	<i>Woe</i>	<i>Imp</i>	<i>Woe</i>	
<i>Age</i>	X	0	X	X	SX	0	SX	X	12
<i>Cash_flow</i>	X	0	0	X	SX	0	0	SX	8
<i>Cash_flow_oprev</i>	X	X	0	X	X	X	0	X	10
<i>Current_ratio</i>	0	0	0	0	X	0	0	0	2
<i>EBITDA_Margin</i>	X	X	0	X	X	X	0	X	12
<i>Gearing</i>	0	0	X	X	0	0	X	X	8
<i>Interest_cover</i>	0	X	0	X	0	SX	0	X	8
<i>Liquidity_ratio</i>	X	X	X	X	SX	0	SX	X	13
<i>Net_income</i>	0	X	0	X	0	SX	0	X	6
<i>No_directors</i>	X	X	X	X	X	X	SX	X	16
<i>No_employees</i>	X	0	X	X	SX	0	X	X	12
<i>No_subsidiaries</i>	X	X	X	X	X	X	X	X	16
<i>Op_rev</i>	X	X	0	X	X	X	0	SX	12
<i>PL_beforetax</i>	0	X	0	X	0	SX	0	X	7
<i>Profit_margin</i>	X	X	X	X	X	X	X	X	14
<i>ROCE</i>	0	X	0	X	0	SX	0	X	6
<i>ROE</i>	X	X	0	X	X	0	0	SX	9
<i>Sharehold_funds</i>	X	X	X	X	X	X	X	SX	14
<i>Shareh_liquidity_ratio</i>	0	0	X	X	0	0	X	SX	5
<i>Solvency_ratio</i>	X	X	X	X	X	X	X	X	16
<i>Tot_assets</i>	X	0	X	X	X	0	SX	X	10

Table 3: Significant variables across the countries for GEV and BGEVA models. X - the variable is significant at 10% s.l. or lower; SX - the smooth term of the variable is significant at 10% s.l. or lower

<i>Methods for missing values</i>	<i>measure</i>	<i>GEV model</i>	<i>logistic</i>	<i>BGEVA model</i>	<i>additive logistic</i>
<i>Weight of Evidence</i>	MAE ⁺	0.784	0.798	0.782	0.797
	MSE ⁺	0.722	0.705	0.702	0.702
	AUC	0.741	0.731	0.722	0.717
<i>Imputation</i>	MAE ⁺	0.862	0.909	0.761	0.969
	MSE ⁺	0.807	0.838	0.713	0.941
	AUC	0.632	0.632	0.677	0.677

Table 4: Forecasting accuracy measures for out-of-sample exercise obtained from applying the GEV and logistic model and BGEVA and logistic additive models to UK data.

<i>Methods for missing values</i>	<i>measure</i>	<i>GEV model</i>	<i>logistic</i>	<i>BGEVA model</i>	<i>additive logistic</i>
<i>Weight of Evidence</i>	MAE ⁺	0.803	0.804	0.781	0.782
	MSE ⁺	0.679	0.684	0.651	0.662
	AUC	0.813	0.812	0.824	0.825
<i>Imputation</i>	MAE ⁺	0.835	0.814	0.891	0.803
	MSE ⁺	0.730	0.711	0.835	0.701
	AUC	0.806	0.806	0.799	0.801

Table 5: Forecasting accuracy measures for out-of-sample exercise obtained from applying the GEV and logistic model and BGEVA and logistic additive models to Italian data.

<i>Variables names</i>	<i>Italy</i>			<i>UK</i>		
	<i>Estimate</i>	<i>Std.Error</i>	<i>p-value</i>	<i>Estimate</i>	<i>Std.Error</i>	<i>p-value</i>
<i>of parametric model</i>						
Intercept	-1.308e+00	1.526e-02	< 2e-16	3.888e+00	6.149e-02	< 2e-16
Cash_flow_oprev	4.011e-03	1.308e-03	0.002	-	-	-
Current_ratio	1.274e-01	1.473e-03	< 2e-16	-	-	-
EBITDA_Margin	-4.701e-03	1.153e-03	4.56e-05	-	-	-
Gearing	-	-	-	1.086e-02	2.316e-04	< 2e-16
No_directors	-2.935e-01	8.068e-03	<2e-16	-	-	-
No_employees	-	-	-	7.436e-02	1.512e-03	< 2e-16
No_subsidiaries	-1.145e-01	1.150e-02	< 2e-16	-9.365e-01	1.929e-02	< 2e-16
Op_rev	1.563e-05	2.519e-06	5.39e-10	-	-	-
Profit_margin	-3.655e-03	8.980e-04	4.70e-05	-7.075e-02	1.436e-03	< 2e-16
ROE	-2.986e-04	5.703e-05	1.64e-07	-	-	-
Shareh_liquidity_ratio	-	-	-	-1.416e-02	2.864e-04	< 2e-16
Sharehold_funds	-1.137e-04	8.391e-06	< 2e-16	8.769e-04	1.811e-05	< 2e-16
Solvency_ratio	-8.894e-03	3.854e-04	< 2e-16	-8.453e-02	1.724e-03	< 2e-16
Tot_assets	4.043e-05	2.595e-06	< 2e-16	-	-	-
<i>of Smooth terms</i>	<i>Edf</i>	<i>Est.rank</i>	<i>p-value</i>	<i>Edf</i>	<i>Est.rank</i>	<i>p-value</i>
age	2.987	3	0.021	9.000	9	< 2e-16
Cash_flow	8.950	9	<2e-16	-	-	-
Liquidity_ratio	8.084	9	<2e-16	8.914	9	< 2e-16
No_directors	-	-	-	9.000	9	< 2e-16
No_employees	3.898	4	<2e-16	-	-	-
Tot_assets	-	-	-	9.000	9	< 2e-16

Table 6: Parametric and smooth component summaries obtained from applying the semiparametric BGEVA model to the samples of Italian and UK SMEs. The missing values are analysed by imputation method. The values of τ parameters for Italian and UK models are -0.41 and -0.9 , respectively.

<i>Variables names</i>	<i>Italy</i>			<i>UK</i>		
<i>of parametric model</i>	<i>Estimate</i>	<i>Std.Error</i>	<i>p-value</i>	<i>Estimate</i>	<i>Std.Error</i>	<i>p-value</i>
Intercept	-1.334	0.011	< 2e-16	-1.572	0.028	<2e-16
age_w	-	-	-	5.367	0.049	<2e-16
cash_flow_oprev_w	0.128	0.018	3.77e-12	0.398	0.021	<2e-16
EBITDA_Margin_w	0.108	0.020	3.92e-08	1.506	0.017	<2e-16
Gearing_w	-	-	-	-1.617	0.025	<2e-16
Interest_cover_w	-	-	-	0.947	0.009	<2e-16
Liquidity_ratio_w	-	-	-	0.879	0.013	<2e-16
Net_income_w	-	-	-	-0.536	0.040	<2e-16
No_directors_w	0.588	0.013	<2e-16	3.486	0.027	<2e-16
No_employees_w	-	-	-	4.004	0.039	<2e-16
No_subsidiaries_w	0.216	0.031	1.83e-12	6.242	0.058	<2e-16
Op_rev_w	-0.262	0.033	2.43e-15	-	-	-
PL_beforetax_w	-	-	-	0.947	0.033	<2e-16
Profit_margin_w	0.106	0.022	2.08e-06	0.261	0.017	<2e-16
ROCE_w	-	-	-	0.439	0.010	<2e-16
Sharehold_funds_w	-0.154	0.023	5.90e-11	-	-	-
Solvency_ratio_w	0.365	0.018	<2e-16	2.087	0.021	<2e-16
Tot_assets_w	-	-	-	0.750	0.024	<2e-16
<i>of smooth terms</i>	<i>Edf</i>	<i>Est.rank</i>	<i>p-value</i>	<i>Edf</i>	<i>Est.rank</i>	<i>p-value</i>
Cash_flow_w	-	-	-	8.808	9	< 2e-16
Interest_cover_w	8.488	9	<2e-16	-	-	-
Net_income_w	5.592	6	<2e-16	-	-	-
Op_rev_w	-	-	-	8.602	9	< 2e-16
PL_beforetax_w	8.908	9	<2e-16	-	-	-
ROCE_w	8.649	9	<2e-16	-	-	-
ROE_w	-	-	-	8.231	9	< 2e-16
Shareh_liquidity_ratio_w	-	-	-	7.303	8	< 2e-16
Sharehold_funds_w	-	-	-	3.961	4	3.49e-12

Table 7: Parametric and smooth component summaries obtained from applying the semiparametric BGEVA model to a sample of Italian and UK SMEs. The missing values are analysed by Weight of Evidence method. The values of τ parameter for Italian and UK models are -0.41 and -0.42 , respectively

	<i>Default = 0 (healthy), n=151178</i>			<i>Default = 1 (default), n=11525</i>			<i>Default = 2 (dissolved), n= 2272</i>		
	% miss	Mean	St dev	% miss	Mean	St dev	% miss	Mean	St dev
Age	0.00	206.61	155.07	0.00	172.16	137.10	0.00	145.63	103.33
Capital	0.03	248.51	2241.88	0.35	335.40	3089.62	0.61	63.86	356.54
Cash_flow	0.05	137.44	607.13	0.40	-387.08	5249.43	0.6875	-35.01	394.77
Cash_flow_oprev	0.20	7.05	8.23	0.73	6.61	10.25	0.87	8.54	10.84
Current_liab	0.03	1686.67	4204.70	0.35	2606.82	6569.06	0.61	459.50	1910.91
Current_ratio	0.03	1.78	2.80	0.35	1.50	3.33	0.62	2.47	5.64
EBITDA_Margin	0.06	7.48	13.67	0.45	-6.27	24.47	0.72	-7.74	25.87
Gearing	0.17	187.20	218.93	0.66	229.94	248.69	0.78	116.42	193.03
Interest_cover	0.11	27.22	97.25	0.43	4.21	69.97	0.72	8.88	98.82
Liquidity_ratio	0.03	1.36	2.37	0.35	1.15	2.98	0.62	2.29	5.78
Loans	0.03	445.10	1453.38	0.35	888.28	3246.21	0.61	94.76	720.79
Net_income	0.03	33.62	560.09	0.35	-530.82	5254.84	0.61	-56.22	366.75
No_directors	0.00	1.37	1.82	0.00	0.13	0.69	0.00	0.00	0.00
No_employees	0.03	13.73	22.70	0.34	13.29	24.56	0.61	6.23	14.50
No_subsidiaries	0.00	0.39	0.80	0	0.12	0.45	0.00	0.00	0.00
Noncurrent_liab	0.03	720.19	3586.76	0.35	1121.89	10009.62	0.61	141.74	997.82
Op_rev	0.03	3100.14	5649.10	0.35	2378.32	5347.09	0.61	776.59	2811.41
PL_beforetax	0.03	79.92	635.17	0.34	-483.08	3715.36	0.61	-60.77	411.73
Profit_employee	0.05	11.77	66.88	0.44	-6.10	63.39	0.64	-6.80	29.68
Profit_margin	0.04	1.71	13.29	0.43	-12.53	25.13	0.66	-12.92	28.44
ROCE	0.08	12.17	52.98	0.53	-24.31	118.41	0.75	-33.93	156.63
ROE	0.09	14.87	93.09	0.60	-34.53	175.52	0.77	-33.83	173.69
Share_liquidity_ratio	0.05	7.79	38.08	0.40	2.69	30.77	0.72	8.16	51.18
Sharehold_funds	0.03	1068.87	4489.58	0.35	-242.39	7134.65	0.61	113.77	1651.52
Solvency_ratio	0.04	24.26	23.68	0.41	4.75	32.58	0.67	16.87	41.34
Tot_assets	0.03	3475.74	9410.08	0.35	3486.66	9829.77	0.61	714.95	3542.29

Table 8: Some descriptive statistics for healthy, defaulted and dissolved Italian SMEs.

	<i>Default = 0 (healthy), n = 39371</i>			<i>Default = 1 (default), n = 960</i>			<i>Default = 2 (dissolved), n = 1192</i>		
	% miss	Mean	St dev	% miss	Mean	St dev	% miss	Mean	St dev
Age	0.00	206.53	191.72	0.00	234.26	199.60	0.00	152.93	132.51
Capital	0.03	611.88	12748.83	0.34	1081.25	8673.70	0.52	1619.43	33925.85
Cash_flow	0.22	358.28	3999.09	0.44	-243.55	6358.60	0.71	261.57	4742.55
Cash_flow_oprev	0.40	12.18	14.79	0.68	7.94	10.75	0.84	20.27	21.60
Current_liab	0.03	3506.71	28913.74	0.34	5180.11	14967.72	0.52	1053.59	7899.34
Current_ratio	0.05	4.79	10.13	0.35	2.45	6.74	0.57	5.44	11.44
EBITDA_Margin	0.24	9.24	21.43	0.47	0.71	19.48	0.73	7.75	33.79
Gearing	0.21	78.05	151.61	0.56	132.03	182.34	0.71	51.63	136.52
Interest_cover	0.58	39.29	118.91	0.56	10.92	74.55	0.91	18.15	90.00
Liquidity_ratio	0.06	4.55	10.08	0.36	2.30	6.80	0.57	5.34	11.42
Loans	0.04	1825.15	15277.83	0.35	2824.16	11707.98	0.55	761.48	7943.12
Net_income	0.04	130.72	4114.47	0.34	-421.83	6078.29	0.51	106.89	3658.45
No_directors	0.00	4.85	2.47	0.00	1.68	2.53	0.00	0.07	0.50
No_employees	0.03	37.29	47.26	0.33	48.21	53.33	0.50	15.06	34.96
No_subsidiaries	0.00	0.45	0.89	0.00	0.36	0.85	0.00	0.00	0.05
Noncurrent_liab	0.03	2310.76	18034.88	0.34	1608.22	8490.13	0.52	420.97	4789.33
Op_rev	0.03	6310.97	8689.12	0.34	7848.69	9420.69	0.50	1814.56	5632.73
PL_beforetax	0.04	207.81	4179.76	0.34	-370.17	6133.68	0.51	138.49	3686.10
Profit_employee	0.08	36.89	233.55	0.4	31.52	278.23	0.54	25.97	177.92
Profit_margin	0.09	8.04	25.79	0.39	-3.16	26.71	0.57	10.37	38.96
ROCE	0.58	19.17	82.45	0.62	-1.89	101.30	0.93	12.87	162.04
ROE	0.18	25.83	106.85	0.54	5.55	121.55	0.71	13.12	173.76
Share_liquidity_ratio	0.52	35.31	104.34	0.61	23.95	94.05	0.91	18.18	65.22
Sharehold_funds	0.03	2349.38	19766.74	0.34	2981.21	29629.11	0.52	1615.8	19086.83
Solvency_ratio	0.07	45.51	38.11	0.39	23.48	38.55	0.60	46.61	44.40
Tot_assets	0.03	8143.35	41542.95	0.34	9782.48	37483.74	0.53	3142.41	22499.54

Table 9: Some descriptive statistics for healthy, defaulted and dissolved UK SMEs.