Econon Polit Ind
https://doi.org/10.1007/s40812-018-0097-x

Spatial autocorrelation and clusters in modelling corporate bankruptcy of manufacturing firms

M. Simona Andreano[1] · Roberto Benedetti[2] · Andrea Mazzitelli[1] · Federica Piersimoni[3]

Received: 6 September 2017 / Revised: 17 March 2018 / Accepted: 31 May 2018

Abstract The interest in the prediction of firms' bankruptcy is increasing in the recent recession period 2008–2012, when, in Italy, the number of distressed manufacturing firms increased sharply. The most popular model applied by bankruptcy researchers is the logit model (logistic regression model). In the present paper we extend this classical model in two different ways, to take into account the spatial effects that can highly affect bankruptcy probability. We propose to apply the spatial Autologistic model and the Logit Regression Tree, with the aim to find evidence of spatial dependence and spatial heterogeneity in bankruptcy probability, of the manufacturing firms of Prato and Florence (Italy). Our application shows that spatial contagion effects are an important issue when modelling bankruptcy probability. Moreover, the application of the regression tree analysis shows the presence of three different clusters, with heterogeneous behaviours.

Keywords Default probability · Autologistic model · Heterogeneity · Spatial dependence

✉ M. Simona Andreano
s.andreano@unimercatorum.it

Roberto Benedetti
benedett@unich.it

Andrea Mazzitelli
a.mazzitelli@unimercatorum.it

Federica Piersimoni
piersimo@istat.it

[1] Universitas Mercatorum, Rome, Italy

[2] University of Chieti-Pescara, Pescara, Italy

[3] Italian National Statistical Institute, Rome, Italy

Published online: 11 June 2018

JEL Classification C13 · C21 · C54 · R12

1 Introduction

The interest of empirical literature on the firm's performance in terms of survival or exit from the market has recently increased, due to the worldwide negative effect of the recent recession on firms' crises. In Italy, where the enterprise system is characterized by family running small-sized firms, the economic crisis of 2008 significantly worsened the financial situation of the firms (Ferretti et al. 2016), and the number of distressed manufacturing firms increased sharply during this period. Financial structure, like leverage, plays a key role in the solvency of firms and during the crises it is deteriorated, together with the accumulation of the trade debt of the Italian firms. Bonaccorsi di Patti et al. (2015) and Fort et al. (2013) are few references on this topic.

Early attempts to predict corporate failure in a univariate context used the ratio analysis (Beaver 1966), which was later extended by Altman (1968) to the multivariate case. More recently, the most popular model applied by bankruptcy researchers is the logit model (logistic regression model). Since in this case the outcomes are between two discrete alternatives, fail and non-fail, bankruptcy classification is an appropriate application for a binary choice model and logit model is commonly used in such qualitative response studies. For a review on the models applied in the bankruptcy analysis, see Gissel et al. (2007).

In the time domain, an alternative way to model bankruptcy is the use of duration models, that allow to estimate the length of the time until failure (Manjon-Antolin and Arauzo-Carod 2008). The survival of firms depends on several factors, that can be distinguished between internal factors (i.e. firm-specific) and external factors. The latter are related to the environment in which the firm operates and can be summarised in industry, spatial and business-cycle factors. Empirical analysis in different nations [see, e.g., Giovannetti et al. (2011) and Mariani et al. (2013) for Italy, Bernard and Jensen (2007) for the US, Box (2008) for Sweden, Bellone et al. (2006) for France, Disney et al. (2003) for the UK] has produced interesting findings, not always with according results that may bring to stylized facts. The role of firms-specific factors in determining firm failure, like age and size, is central in theoretical firm-survival models, however empirical evidence is not unanimous. Although Esteve-Pérez et al. (2004), Fackler et al. (2013) and Strotmann (2007) show that small firms exhibit a shorter life expectation, Varum and Rocha (2012) assert conversely that smaller enterprises may be more flexible in adjusting to downturns, being more able to exploit market niches and activities characterized by agglomeration economies, and being less reliant on formal credits compared with larger firms (Tan and See 2004). Moreover, Agarwal and Gort (2002) found that the age effect has an inverted-U shape (Esteve-Pérez and Mañez-Castillejo 2008) and size-age influences may also differ across industries (Giovannetti et al. 2011; Lopez-Garcia et al. 2007): large firms that operate in high-tech sectors have higher probability of survival than small firms in traditional sectors.

Spatial factors, as external environmental factor, have been proved to affect firm performance in different way. Strotmann (2007) finds that firms located in rural areas have higher chances of survival than those located in urban areas and Honjo (2000) that agglomeration negatively affects survival. However, Fotopoulos and Louri (2000) show that firms located in greater urban areas have lower hazard rates than those located outside those areas. Moreover, many paper have examined the effect of belonging in a cluster on the survival and performance of firms (Delgado et al. 2014; Folta et al. 2006; McCann and Folta 2011), however attention should be made in discerning between Industrial District and Business Cluster (Ortega-Colomer and Molina-Morales 2016). Although both concepts converge in giving the territory a preeminent role, the first one refers explicitly to the community of people and the context in which knowledge flows and job opportunities occur, whereas Business Clusters (BC) are described by Porter (1990) as being composed by industries connected through vertical (buyer/supplier) and horizontal (common customers, technology, distribution channels, etc.) relationships. Thus, the BC framework is more a theory of the firm, to explain why some firms are more successful than others. Empirical evidence is not concordant on the presence of positive or negative effects in the performance of clustered firms and anyway not all firms may be affected equally (McCann and Folta 2011). Effects also may differ for industries and for the development phases of the clusters (Maine et al. 2010). Ramazzotti (2010) argued that in period of stagnation (and recession) the conditions for the persistence of industrial districts may eventually disappear, leading to different types of local organisation, such as BC.

In the present paper we apply the bankruptcy logit model on the manufacturing firms of Prato and Florence (Italy), to identify the probability of the firms to survive or to fail and exit from the market in the period 2008–2012. More precisely, we consider the firms belonging in the eight Local Market Areas (LMA, Istat) of Prato and Florence. The aim of the paper is to analyse the manufacturing firms and their probability to survive during the most acute period of the crisis 2008–2012. From this point of view, the geographical area of Florence and Prato represents an optimal research workshop, as it is characterized for over 75% by manufacturing firms of small-medium size, essentially concerning textile and clothing industries, for which is possible to achieve accurate microdata. Moreover, our analysis attempts to capture the spatial contagion effects of the crisis between firms. Therefore, it can be useful to see how the sampled firms are distributed over the Local Market Areas (LMA) defined in the area of Florence and Prato. LMAs are sub-regional geographical areas where firms can find the largest amount of the labour force necessary to occupy the offered jobs (Istat 2015) and are characterized by relevant socio-economic relations. LMAs are also the territorial unit on which the Italian National Statistical Institute (ISTAT) defines the Industrial Districts (ID).

Our methodological approach, crucially differs from the previous bankruptcy analysis, because we introduce the spatial dimension of the observed data and this is explicitly added in the model. Potential contagion effects on interconnected firms, generated by chain reactions, are often neglected in literature. Liquidity tensions and internal imbalances is expected to have a positive spatial effect on neighbour firms, and this can increase the probability to become a distressed firm during the analysed period.

Although proximity and agglomeration effects are widely considered in empirical analysis on the performance and organization of cluster firms (Porter 2003; Boari et al. 2003; Tokunaga et al. 2014; Cainelli et al. 2006), no such issue is present in bankruptcy firms studies. It can be interesting to investigate if the spatial bankruptcy propagation effects fallow the local firm connection as defined by the LMAs (and therefore, by IDs).

Standard statistical techniques employed by empirical analysts in the study of regional and local economy assume independence and homogeneity among the observations, neglecting the potential spatial contagion effects. However, the hypothesis of independence and homogeneity is evidently violated in all geographical and territorial studies (Anselin 1988). Empirical models that do not take into account for spatial dependence and/or structural heterogeneities may therefore lead to misspecification problems.

To take into account the propensity for nearby locations to influence each other, a general class of well-known models has been introduced in the statistical literature (Besag 1974; Cressie 1993). In our case, we extend the standard logit model to the autologistic model (Hughes et al. 2011) and our analysis confirms the presence of spatial dependence.

Heterogeneity is the other relevant characteristic highlighted by spatial data. The presence of a not constant relationship between a response variable and the covariates in an area has led to the introduction of spatially varying coefficients (Wheeler and Calder 2007), geographically weighted regression (Fotheringham et al. 2002), or local linear regression models (Loader 1999). Moreover, some partitioning algorithm can be applied on the data, to detect homogenous clusters, with respect to the identified model. In our analysis we apply a modified Regression Tree algorithm with the logistic model and obtain, for the Prato and Florence area, three different homogenous clusters.

Our results are in line with Porter's assertion to introduce geographical proximity in the business cluster analysis, and giving the territory a prominent role in firm's activities. However, our primary aim is not to test a specific theory, but to examine data and relationships between them. The concept of Clusters is constantly on evolution and the interest on it is almost alive, as evidenced, for example, by the European Commission (2008) document: "The Concept of Cluster and Cluster Policies and their Role for Competitiveness and Innovation: Main statistical results and lessons learned".

The layout of the paper is the following. Section 2 is devoted to a brief explanation of the data used in the application. In Sect. 3 we describe the two new methods to analyse bankruptcy in the spatial domain: the Spatial AutoLogistic model and the Regression Trees with the logistic model. Section 4 concerns the presentation of our empirical study of the eight LMA of Prato and Florence and, finally, in Sect. 5 we draw some concluding remarks and we outline a future research agenda.

2 Data and variables

This study is based on a sample of Italian manufacturing bankrupt firms collected by the database AIDA (Analisi Informatizzata delle Aziende Italiane), localized in the eight Local Market Areas (LMA) of Prato and Florence during the recession period 2008–2012. A control group of non-failed firms is also collected. For both, insolvent and "healthy" firms, we observed some economic and financial variables 3 years before the recession began, when all business were not subject to bankruptcy (Ferretti et al. 2016).

Firm level data were drawn from AIDA, an archive that combines corporate financial statements with economic and industrial information. We did not consider companies with missing financial data. First of all, we extracted for each firm the following structural information: (1) Name of the company; (2) Tax code; (3) Localization of the firm (in terms of its geographical coordinates); (4) Year of establishment, year of the crisis and economic activity (based on NACE code); (5) Number of employees; (6) Revenue. Then, we considered additional financial information, like the index of profitability (ROA, Return On Assets), the index of liquidity and the index of debt.

The ROA is an indicator of how profitable is a company in terms of its total assets, and shows how efficient management is using its assets to generate earnings. Calculated by dividing the company's annual earnings by its total assets, ROA is displayed as a percentage. The higher the index, the more efficient the management use its asset base. The liquidity index measures the ability of the firm to pay debt obligations and its margin of safety through the calculation of metrics, including, for instance, the current ratio or the operating cash flow ratio. Current liabilities are analysed in relation to liquid assets, to evaluate the coverage of short-term debts in an emergency. In general, a higher liquidity ratio indicates that a company is more liquid and has better coverage of outstanding debts. In our analysis, we consider as liquidity ratio the current ratio. This index is computed as the ratio between total current assets and total current liabilities. Therefore, current ratio is a rough measurement of a company's financial health. The higher the current ratio, the more the company is able to pay its obligations, as it has a larger proportion of asset value with respect to the value of its liabilities. A ratio under 0.80 indicates that a company's liabilities are greater than its assets and suggests that the company would be unable to pay off its obligations if claimed. Although a current ratio below 0.80 is indicator of bad financial health, it does not necessarily mean that the firm will go bankrupt (Brealey et al. 2011). Finally, we introduce the debt ratio, defined as the ratio of total (long-term and short-term) debt to total assets, expressed as a percentage. It can be interpreted as the proportion of a company's assets that are financed by debt. The higher is this ratio, the more the firm is leveraged, and therefore will have a higher financial risk. A debt ratio greater than 100% tells that a company has more debt than assets. Conversely, a debt ratio less than 100% indicates that a company has more assets than debt. Used together with other financial health measures, the debt ratio can help investors to determine the company's risk level. Typically, a

Table 1 LMA of Florence and Prato and their manufacturing firms

LMA	Bankrupt	Solvent
906	45	185
909	22	145
910	27	119
914	39	150
915	99	632
926	27	93
927	39	300
948	154	567

debt ratio greater than two indicates a risky scenario for the investor, however this rule can vary with respect to the analysed industry (Brealey et al. 2011).

The sample of Italian manufacturing firms are localized in the eight Local Market Areas (LMA) of Prato and Florence. More precisely, the LMAs referring to our geographical area extend over the provinces of Florence and Prato, incorporating also some territories of Pisa, Pistoia and Lucca. LMAs represent a territorial grid whose boundaries are defined using the flows of the daily home/work travel recorded with the Census of the population and housing, not respecting administrative boundary constraints. Home/work travel is used as a proxy for existing relationships in the area. In addition, the LMAs lead to the identification of industrial districts, particularly active in Tuscany, characterized by a strong relational link that should accentuate the contagion effects of survival/bankruptcy of the firms. While in Italy the number of districts has been contracted by about 22% in the last decade, in our analysed area the situation has remained almost unchanged and Tuscany is among the regions with the highest district employment share (Regione Toscana 2017). Moreover, as evidenced by Di Giacinto et al. (2014), spatial concentration exerts favourable effects on local productivity of manufacturing firms, with urban areas generally been more beneficial then industrial clusters. Therefore, the analysed area of Florence and Prato allows having an overview of the survival of the firms, considering different possible spatial agglomerations.

We assign to a firm the value "zero" if it went bankrupt during the observed period, the value "one" otherwise (therefore when it is solvent). The following Table 1 will show how our manufacturing firms of the LMAs of Prato and Florence are distributed with respect to this dichotomous bankrupt—solvent event. The LMA are defined in accordance to the classification provided by ISTAT.

The map in Fig. 1 displays the spatial distribution of these 2643 manufacturing firms, marked with different colours with respect to its business insolvency attribute. We report in the map also the LMA borders and the administrative boundaries at province and region level.

In the sample we observe 452 firms that went bankrupt and 2191 that were solvent. The firms are all manufacturing capital companies that were active at the date of 31/12/2007. The control group was extracted from the same database, by

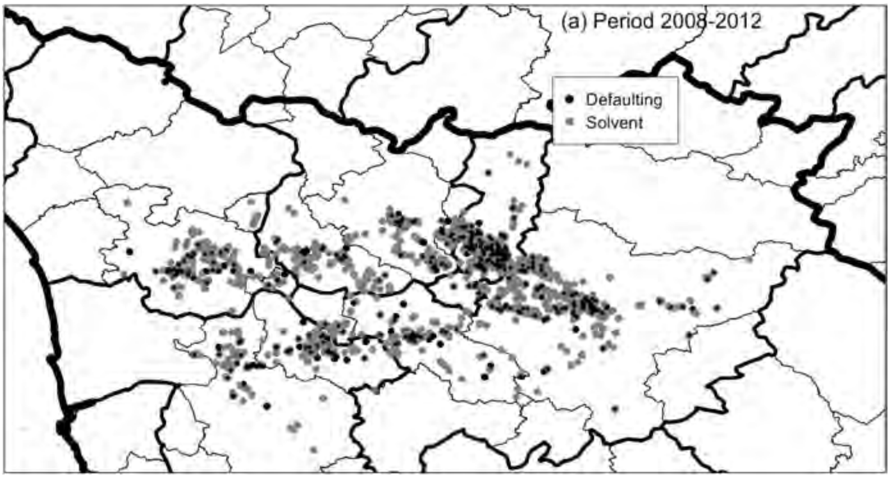

Fig. 1 Localization of the 2643 firms in the LMA of Florence and Prato (Italy)

Table 2 Some characteristic of the sampled firms (mean or modal values)

Variable	All sample	Solvent	Bankrupt
Economic sector	Textiles (49%)	Textiles (48%)	Textiles (52%)
Employees	42	44	30
Age	22	23	16
Revenue	3.600	3.800	2.500

selecting the firms active at the end of 2007 and always active after 2012 (active at least until 2016), according to the Business Register of the Italian Chambers of Commerce. Only manufacturing firms with a revenue at current prices higher than one million Euros were included in the control group. In this way, we are sure that our control firms are not bankrupt just after 2012, and survived after the financial crisis. We do not impose, instead, any restriction on the revenue of the insolvent companies, in order to maintain all distressed firms inside the sample.

Table 2 reports some structural characteristics of the sampled firms, separately for solvent and distressed firms.

In all cases, around half of the firms are in the economic sector "Textiles", regardless of being solvent or bankrupt. Not surprising, bankrupt firms are smaller and younger than those survived to the financial crisis (Strotmann 2007).

3 Methodology

In the present paper we apply a nonlinear statistical model, with the aim to verify which economic and financial characteristics observed in the pre-crisis period, could affect the probability of being solvent in the same period.

We start with the application of a standard logit cross sectional model, and analyse the probability that a solvent and active firm in the pre-crisis period is falling during the recession period 2008–2012. As previously mentioned, the bankruptcy event is classified as a binary outcome dependent variable, where a firm that is observed to enter bankruptcy is coded "0" and solvent or non-failed firm is coded "1".

Logistic regression model has been widely used in economic and financial data to explore the risk factors of corporate bankrupt (Bauer and Agarwal 2014; Jones and Hensher 2004; Jones et al. 2017) and is the most common method in a case–control studies.

Consistent with economics and business academic literature (Altman 1968; Altman et al. 1977; Ohlson 1980; Zmijewski 1984; Lau 1987; Jones and Hensher 2004, 2008), we include the following economic and financial covariates (see Sect. 2) as predictors of the model: ROA, Liquidity (i.e. current ratio) and Debt ratio. These financial indices were computed in terms of their averages over the time interval 2005–2007.

In addition, following the literature on this topic, we include some control variables that are related to the corporate failures: firm size, namely the number of employees and the total revenue, and the age of the firm. The number of employees and the revenue are averaged over the period 2005–2007 and opportunely rescaled. The number of employees is augmented by one (the owner of the firm) to avoid obtaining infinity values when the log operator is applied on it. For the insolvent companies, the age variable is given by the difference between the year in which the company fails bankrupt and the year of its establishment. Similarly, for solvent or "healthy" companies, the age is the difference between 2012 (the last year in our time interval) and the year of establishment of the firm.

Therefore, our standard logistic regression model (Agresti 2002; McCullagh and Nelder 1989) is given by the following equation:

$$\begin{aligned}\ln\left(\frac{P_i}{1-P_i}\right) &= \alpha + \beta_1\, ROA_i + \beta_2\, Liquidity_i + \beta_3\, Debt\, ratio_i \\ &\quad + \beta_4\, revenue_i + \beta_5\, \ln(employees)_i + \beta_6\, firm\, age_i + \varepsilon_i,\end{aligned} \tag{1}$$

where P_i is the probability of the occurrence of a solvent firm, namely, the expected value of the dependent variable y_i (so that $y_i = 1$ if the firm is solvent and $y_i = 0$, otherwise); β_i are the regression coefficients and i is the index for the observed firms.

In the model (1) our dependent variable is the logit transformation of P, i.e., $\ln[P/(1-P)]$. The logit of a probability is simply the log of the odds of the response taking the value one. The logit function can take any real value, but the associated probability always lies in the required [0, 1] interval. Our covariates in Eq. (1) are all quantitative continuous variables. The interpretation of the coefficient β_i in (1) can be made with respect to the odds ratio. Therefore, we generally have:

$$\left(\frac{P_i}{1-P_i}\right) = \exp(\beta_i), \tag{2}$$

and, for example, if $exp(\beta_i)=2$, then a one unit change in X_i would make the event twice as likely to occur.

We can extend the logistic model to the Autologistic one, by including the spatial autocovariate that will correct the original model with respect to the effect of spatial autocorrelation. Models that ignore spatial autocorrelation may lead to misspecification problems.

The autologistic regression, is widely used for modelling spatially correlated binary variables (Dormann 2007; Legendre 1993). Corporate bankrupt and localization firms may be positively autocorrelated, such that neighbouring units in space tend to have more similar values than would be expected by random chance. In addition, models that disregarded the spatial autocorrelation may be unsuitable and might overvalue the importance of geographical variables (Gotway and Stroup 1997).

The autologistic regression model is a special case of the general logistic model introduced by Besag (1972, 1974). When Eq. (1) is extended to the Autologistic model, we obtain the following equation:

$$\ln\left(\frac{P_i}{1-P_i}\right) = \alpha + \beta_1\, ROA_i + \beta_2\, Liquidity_i + \beta_3\, Debt\, ratio_i + \\ + \beta_4\, revenue_i + \beta_5\, \ln\left(employees\right)_i + \beta_6 firm\, age_i + \rho Z_i + \varepsilon_i, \tag{3}$$

where ρ is the spatial dependent parameter and Z_i the so-called autocovariate, and is the mean of the dependent variable observed in the neighbouring firms of i. The set of neighbours is defined by all the units that fall within a bandwidth distance from i. In our application, the distance was fixed equal to 0.5 km (distance between latitude and longitude coordinates), a reasonably small distance that guarantees the inclusion of at least one neighbour for each i. The parameter ρ is a measure of Z_i's reactivity to its neighbours. If $\rho=0$, the model reduces to the ordinary logistic one, while $\rho>0$ (<0) corresponds to positive (negative) spatial dependence (Hughes et al. 2011; Sherman et al. 2006). In our application we expect to observe a positive value of ρ, because it is plausible that we are in presence of an attraction rather than a repulsion force between the firms.

Heterogeneity is the other relevant characteristic highlighted by spatial data and in our application, we introduce an innovative algorithm that combines logistic regression and tree decision methods to analyse the presence of this heterogeneity. Although it is possible to use recursive partitioning methods for classification corporate insolvency (Cashin and Dattagupta 2008), we deal with the classification problem into a regression logistic context (Frank et al. 1998; Landwehr et al. 2005), by fitting a collection of local regression models to subsets of the data (i.e. a segmented logistic regression model), in order to obtain a better fit and higher predictive accuracy (Rusch et al. 2013). Recursive partitioning algorithms, as classification trees (Zhang and Singer 2010), are generally performed by splitting the data into a number of different possible partitions, by looking to an objective function that should be minimized. This function is usually named the heterogeneity measure, and will be computed for each segment.

Here, our aim is to identify clusters of firms that are homogeneous in terms of parameters of their bankruptcy model or, in other words, we want to identify situation of local stationarity in the bankruptcy probability model of the different firms. Following these considerations, the objective function is not, as usual, the sum of the squared residuals, but the difference among the parameters of the model under investigation: a group of firms are divided into two sub-groups if the parameter estimates are significantly different among them (Postiglione et al. 2008).

More formally, a logistic model tree is a standard decision tree with logistic regression functions at the leaves of the trees. Unlike classical decision trees, the leaves $t \in T$ have associated a logistic regression function l_t instead of a class label (Landwehr et al. 2005). If we define with S the whole space of observation, the tree structure returns a disjoint partition of S into regions S_t, and every region is represented by a leaf in the tree:

$$S = \bigcup_{t \in T} S_t, \; S_t \bigcap S_{t'} = \emptyset \quad for\, t \neq t'. \tag{4}$$

Hence, the model represented by the whole logistic model tree is given by the following equation:

$$l_t = \sum_{t \in T} l_t(x) \cdot I(x \in S_t), \tag{5}$$

where $I(x \in S_t) = \begin{cases} 1, & if\, x \in S_t \\ 0, & otherwise \end{cases}$.

The model tree combines data driven partitioning algorithm, as in standard classification trees, with model-based prediction. In addition, both—standard logistic regression and decision trees—are special cases of the logistic model tree (Landwehr et al. 2003). In particular, if there is only a single segment, the logistic regression tree reduces to a standard logistic regression (Rusch et al. 2013).

4 Empirical results

As mentioned in Sects. 1 and 3, standard bankruptcy prediction uses standard logistic regression model. Therefore, we start with the estimation of such a model on our sample of manufacturing firms of Prato and Florence (Italy), observed over the period 2008–2012. More precisely, we consider the eight LMAs in the Prato and Florence provinces, identified with the numbers: 906, 909, 910, 914, 915, 926, 927, 948. The goal is to verify how some business pre-crisis economic and financial characteristics have affected the probability of getting out of the market during the recession period 2008–2012.

Table 3 reports our empirical evidence for bankruptcy prediction of the firms in the Prato–Florence area: the first column displays the Logistic results for non-spatial

Table 3 Logistic and Autologistic model

	Logistic	Autologistic
(Intercept)	− 1.5233*** (0.2047)	− 4.0877*** (0.7909)
Age	3.9259*** (0.5939)	3.8957*** (0.5929)
Revenue	0.6609 (1.1398)	0.6711 (1.1541)
ROA	8.4993*** (0.8509)	8.2476*** (0.8357)
Liquidity	1.2004*** (0.1650)	1.1892*** (0.1645)
Debt ratio	− 2.1628# (1.2362)	− 2.2278* (1.2179)
Ln(Employees)	0.4275*** (0.0571)	0.4134*** (0.0570)
Z (autocov)	– (–)	3.1949*** (0.9398)
AIC	1890.9	1863.5

Significance codes: ***0.001, **0.01, *0.05, #010

Table 4 Predictions with logistic and autologistic model

	Logistic model		Autologistic model	
	0	1	0	1
0	**111**	339	**116**	336
1	26	**2167**	27	**2166**

In bold the correct occurrences

specification (i.e. Eq. 1), while the second column is devoted to the estimate of the Autologistic spatial version of the model (i.e. Eq. 3). Standard errors are described in parentheses, and significance levels are also highlighted. Finally, we report also the AIC statistics for the two models.

The coefficients are mainly significant for both models. The signs of the coefficients are in line with the literature (Ferretti et al. (2016) defined the probability of occurrence exactly in a complementary way, with P equals the probability to fail). Surprising, the Revenue variable is not significant in both models. Although the estimated coefficient of the two models are rather similar in terms of magnitudes, the presence of spatial dependence is confirmed by our data and the autocovariate Z, that highlights these spatial transmission effects between the firms, is significant. Moreover, the AIC statistics of the Autologistic model is better than that of the standard one. These two last results confirm that the inclusion of the spatial dependence will improve the goodness of fit of the model and avoid estimating spurious regression.

The value of ρ is approximately equal to 3.2: this means that looking at the adjacent firms, the probability to be a solvent firm is three times higher when similar solvent firms are in the neighbourhood. This confirms that contagion effects are very important when dealing with firms' bankruptcy and that the spatial effect in firms collaboration is positive ($\rho > 0$).

The presence of spatial dependence seems to reduce a few the magnitude of all the estimated coefficients, because apparently, they were inflated by indirectly embodying the spatial effects.

Table 5 Estimation of the logistic model over the three clusters

	Cluster 1	Cluster 2	Cluster 3
(Intercept)	−3.1901*** (0.5189)	−1.0619* (0.4445)	−1.2045*** (0.2707)
Age	4.2304*** (1.4432)	6.2312*** (1.3795)	2.7283*** (0.7566)
Revenue	109.0723# (63.4141)	26.4859 (27.2021)	−7.4368* (3.3753)
ROA	13.7217*** (2.0608)	5.4525*** (1.2959)	8.7013*** (1.2244)
Liquidity	1.7300*** (0.4104)	1.3157*** (0.3433)	0.9543*** (0.2124)
Debt ratio	−4.2925 (3.1890)	−2.8534 (1.7915)	−1.7358 (1.9215)
Ln(Employees)	0.8395*** (0.1595)	0.1419 (0.1384)	0.4228*** (0.0803)
N	607	809	1227

Significance codes: ***0.001, **0.01, *0.05, #010

We can also compare the two estimated models with respect to their prediction power (Table 4), where the Autologistic model confirms its better performance.

Our estimated global models confirm the presence of a positive effect of firms' age and size (given by the variable Employees) on the survival probability, as widely evidenced in literature, but in contrast with the empirical evidence on Tuscany (Italy) of Mariani et al. (2013), who applied a time survival model over the period 2007–2009.

In this paper, our leading idea is that the treatment of spatial dependence is not sufficient for a satisfactory modelling of geographically distributed data: the spatial heterogeneity cannot be neglected. The aim is to identify spatial cluster of firms that are similar in their solvency probability in the analysed period.

The presence of a not constant relationship over all the geographical area study is verified, in the present paper, through the application of the logistic regression tree (see Sect. 3). Performing the analysis using the k-mean distance on the geographical coordinates of the firms, we obtain three different clusters, with noticeable heterogeneity in their definition of bankruptcy probability and the main results are reported in Table 5.

The three local homogenous models seem to appear more reasonable in explaining the bankruptcy-solvency probability of our analysed firms.

The AIC statistics for the regression tree logistic model is equal 1857.7, significantly lower than those computed for the two global homogenous models. This further suggests that the model should not be estimated in a global way, but over geographical zones, where the relationship among variables is defined in a local way. A glimpse at the results reveals that mostly variables have the expected sign. Figure 2 shows the distribution of the firms over the three identified clusters. Looking at the structural characteristics of the three clusters (Table 6), we don't observe significant differences in terms of Age, Employees and Revenue. The prevalent economic activity is always the Textile one, however the first cluster

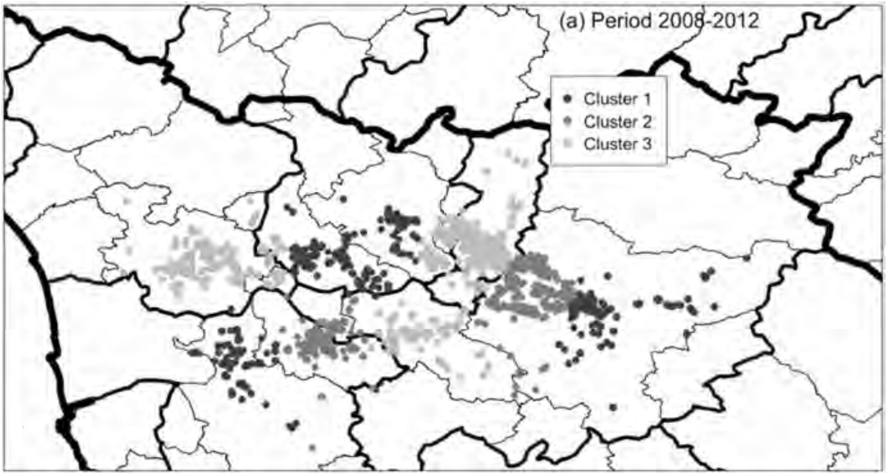

Fig. 2 The spatial distribution of the firms in the three identified clusters

Table 6 Some characteristic of the three identified clusters (mean or modal values)

Variable	C1	C2	C3
Economic sector	Textile (29%)	Textile (51%)	Textile (58%)
Employees	44	43	41
Age	22	23	22
Revenue	3.400	3.800	3.500

shows a lower percentage, with around 18% of the firms belonging also in the Electronic and Mechanic sector. However, the economic sector was not chosen by the CART procedure as a significant splitting variable.[1]

First of all, we note that the magnitude of the three clusters is different, ranging from 607 firms in the first cluster, to 1227 firms in the third. Although the significant variables in the three clusters are the same, the values of the estimated coefficients change greatly from one group to the other. In Cluster 1, the probability to be solvent is highly conditioned by the value of the ROA variable.

When we look at the variables of the financial structure of the firms, e.g., our Liquidity and Debt ratio, we have that only the first one is highly significantly different from zero. The heavily deterioration of liquidity, given by the increasing time length for making the payments and by the trouble to access to credit, seems the most important variable to define the probability to bankrupt.

It is interesting to note that the Revenue variable has coefficients ranging from negative (Cluster 3) to positive (Cluster 1), highlighting a strong heterogeneity of the firms with respect to this variable. However, some more deep analysis should

[1] A logistic model with a dummy variable referring to the economic activities was also estimated, but the dummy was not significant.

be made, in terms of its quantile distribution, to better understand its impact on the probability. For the firms in Cluster 1 the Revenue has a very high impact on the probability to survive (although at a 10% significance level). The Cluster 1 has an inhomogeneous economic sector composition of the firms, therefore this variable seems not to be determinant to explain the different impact of the revenue on the survival probability.

Finally, we can note that the age of the firms is a "guarantee" to remain on the market (the coefficient is always positive).

If we look at the spatial distribution of the firms over the three clusters, we are not able to find a match with the border defined by the LMAs. The homogeneity seems to cross these borders and the connection between the firms will be defined through new paradigms, not necessary linked to local labour force. The homogeneity in the firms' bankruptcy can be explained through firm-level characteristics and on co-location of establishments positioned vertically along the value-added chains, then by social-community cohesion. Therefore, our empirical results suggest the existence of homogeneous groups more in line with the pragmatic and flexible notion of business clusters (Ortega-Colomer and Molina-Morales 2016).

5 Conclusions

In the present paper we applied the bankruptcy logit model on the manufacturing firms of the Local Market Areas of Prato and Florence (Italy), to identify the probability of the firms to survive or to fail and exit from the market in the period 2008–2012. The potential contagion effects on interconnected firms, generated by chain reactions and liquidity tensions, will produce a positive spatial effect on neighbour firms. Our concern is that the presence of this spatial dependence and spatial heterogeneity is a very important characteristic of spatial data, that cannot be neglected, when studying the propagation effects of bankruptcy. Therefore, we extended the classic Logit model in two ways, for taking into account these two different spatial effects: the spatial Autologistic model and the Logistic regression tree.

Our results highlight that the spatial model outperformed the classical one and the heterogeneity in the relation defining bankruptcy probability suggests the presence of three different clusters. Our spatial model shows that the probability to be a solvent firm is three times higher when similar solvent firms are in the neighbourhood. All our estimated models confirm the presence of a positive effect of age and size on the probability to survive and, instead, that the economic activity is not relevant to the definition of the homogenous groups. Moreover, the spatial distribution of the firms among the three identified clusters doesn't follow the boundaries of the LMAs.

These findings suggest that the negative effect of the recession and its propagation on the small and medium enterprises can be defined by the spatial propagation and the firm's value chains. Therefore, the clusters framework seems to be explained more by the firms' performance, than by the result of the social cohesion within the local community of people (Ortega-Colomer and Molina-Morales 2016). These results are in line with Porter assumptions and should also confirm the assertion of

Ramazzotti (2010) that in period of stagnation or recession, the conditions for the persistence of industrial districts may disappear, leading to different types of local organisation, such as BC.

Future research directions will extend our analysis by merging together both methodologies proposed in the present paper, and applying a spatial autologistic model in the classification tree procedure. Furthermore, the analysis should be extended by including more variables in the estimated model, with the aim to explore how innovation and globalization can affect survival probability of the firms.

Funding This study was not funded by any Institution.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Agarwal, R., & Gort, M. (2002). Firm and product life cycles and firm survival. *American Economic Review, 92*(2), 184–190.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(3), 589–609.

Altman, E. I., Haldeman, R. G., & Narayana, P. (1977). ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance, 1*(1), 29–54.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers.

Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance, 40,* 432–444.

Beaver, W. (1966). Financial rations as predictors of failure. *Journal of Accounting Research, 3,* 71–111.

Bellone, F., Musso, P., Nesta, L., & Quéré, M. (2006). Productivity and market selection of French manufacturing firms in the Nineties. *Revue de L'OFCE, 97*(5), 319–349.

Bernard, A., & Jensen, J. (2007). Firm structure, multinationals, and manufacturing plant deaths. *The Review of Economics and Statistics, 89*(2), 193–204.

Besag, J. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society B, 34,* 75–83.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B, 36,* 192–236.

Boari, C., Odorici, V., & Zamarian, M. (2003). Clusters and rivalry: Does localization really matter? *Scandinavian Journal of Management, 19*(4), 467–489.

Bonaccorsi di Patti, E., D'Ignazio, A., Gallo, M., & Micucci, G. (2015). The role of leverage in firm solvency: Evidence from bank loans. *Italian Economic Journal, 1,* 253–286.

Box, M. (2008). The death of firms: exploring the effects of environment and birth cohorton firm survival in Sweden. *Small Business Economics, 31*(4), 379–393.

Brealey, R. A., Myers, S. C., & Marcus, A. J. (2011). *Fundamentals of corporate finance. International edition*. New York: McGraw Hill Irwin.

Cainelli, G., Iacobucci, D., & Morganti, E. (2006). Spatial agglomeration and business groups: New evidence from Italian industrial districts. *Regional Studies, 40*(5), 507–518.

Cashin, P., & Dattagupta, R. (2008). *The anatomy of banking crises*. Washington DC: IMF WP.

Cressie, N. (1993). *Statistics for spatial data, revised edition*. New York: Wiley.

Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy, 43,* 1785–1799.

Di Giacinto, V., Gomellini, M., Micucci, G., & Pagnini, M. (2014). Mapping local productivity advantages in Italy: Industrial districts, cities or both? *Journal of Economic Geography, 14*(2), 365–394.

Disney, R., Haskel, J., & Heden, Y. (2003). Entry, exit and establishment survival in UK manufacturing. *Journal of Industrial Economics, 51*(1), 91–112.
Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecological Biogeography, 16,* 129–138.
Esteve-Pérez, S., Sanchis, Llopis A., & Llopis, J. A. (2004). The determinants of survival of Spanish manufacturing firms. *Review of Industrial Organization, 25*(3), 251–273.
Esteve-Pérez, S., Mañez-Castillejo, J. A. (2008). The resource-based theory of the firm and firm survival. *Small Business Economics*, *30*(3), 231–249.
Fackler, D., Schnabel, C., & Wagner, J. (2013). Establishment exits in Germany: The role of size and age. *Small Business Economics, 41,* 683–700.
Ferretti, S., Filippone, A., & Micucci, G. (2016). *Le imprese uscite dal mercato nel corso della crisi* (p. 317). Rome: Banca d'Italia, Occasional Papers.
Folta, T. B., Cooper, A. C., & Baik, Y. S. (2006). Geographic cluster size and firm performance. *Journal of Business Venturing, 21,* 217–242.
Fort, T. C., Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2013). *How firms respond to business cycles: The role of firm age and firm size*. Cambridge: NBER WP.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: Wiley.
Fotopoulos, G., & Louri, H. (2000). Location and survival of new entry. *Small Business Economics, 14,* 311–321.
Frank, E., Wang, Y., Inglis, S., Holmes, G., & Witten, I. H. (1998). Using model trees for classification. *Machine Learning, 32*(1), 63–76.
Giovannetti, G., Ricchiuti, G., & Velucchi, M. (2011). Size, innovation and internationalization: A survival analysis of Italian firms. *Applied Economics, 43*(12), 1511–1520.
Gissel, J. L., Giacomino, D., & Akers, M. D. (2007). *A review of bankruptcy prediction studies: 1930–present*. Milwaukee: Marquette University WP.
Gotway, C. A., & Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agriculture, Biological and Environmental Statistics, 2,* 157–178.
Honjo, Y. (2000). Business failure of new firms: An empirical analysis using a multiplicative hazards model. *International Journal of Industrial Organization, 18,* 557–574.
Hughes, J., Haran, M., & Caragea, P. C. (2011). Autologistic models for binary data on a lattice. *Environmetrics, 22,* 857–871.
Istat (2015) The new geography of the labor market areas. Istat, Simplicissimus Book Farm, Rome
Jones, S., & Hensher, D. A. (2004). Predicting firm financial distress: A mixed logit model. *The Accounting Review, 79*(4), 1011–1038.
Jones, S., & Hensher, D. A. (Eds.). (2008). *Advances in credit risk modelling and corporate bankruptcy prediction*. Cambridge: Cambridge University Press.
Jones, S., Johnstone, D., & Wilson, R. (2017). Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting, 44*(1), 3–34.
Landwehr, N., Hall, M., & Frank, E. (2003). Logistic model trees. In *Proceedings of the 14th European Conference on Machine Learning* (pp. 241–252).
Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning, 59,* 161–205.
Lau, A. H. (1987). A five-state financial distress prediction model. *Journal of Accounting Research., 25*(1), 127–138.
Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology, 74,* 1659–1673.
Loader, C. (1999). *Local regression and likelihood*. New York: Springer.
Lopez-Garcia, P., Puente, S., & Gómez, Á. L. (2007). Firm productivity dynamics in Spain. Banco de España Working Paper No. 0739.
Maine, E. M., Shapiro, D. M., & Vining, A. R. (2010). The role of clustering in the growth of new technology-based firms. *Small Business Economics, 34,* 127–146.
Manjón-Antolín, M., & Arauzo-Carod, J. M. (2008). Firm survival: Methods and evidence. *Empirica, 35*(1), 1–24.
Mariani, M., Pirani, E., & Radicchi, E. (2013). La sopravvivenza delle imprese negli anni della crisi: Prime evidenze empiriche dalla Toscana. *Economia e Politica Industriale, 1*(1), 25–52.
McCann, B. T., & Folta, T. B. (2011). Performance differentials within geographic clusters. *Journal of Business Venturing, 26,* 104–123.
McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). New York: Chapman & Hall.

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

Ortega-Colomer, F. J., Molina-Morales, F. X. (2016). Discussing the concepts of cluster and industrial district. *Journal of Technology Management and Innovation, 11*(2), 139–147.

Porter, M. (1990). *The competitive advantage of nations*. New York: Free Press.

Porter, M. (2003). The economic performance of regions. *Regional Studies, 37*(6&7), 549–578.

Postiglione, P., Benedetti, R., & Lafratta, G. (2008). A regression tree algorithm for the identification of convergence clubs. *Computational Statistics & Data Analysis, 54,* 2776–2785.

Ramazzotti, P. (2010). Industrial districts, social cohesion and economic decline in Italy. *Cambridge Journal of Economics, 34,* 955–974.

Regione Toscana (2017). I distretti industriali toscani secondo i principali indicatori di contesto, Ufficio Regionale di Statistica, May 2017.

Rusch, T., Lee, I., Hornik, K., Jank, W., & Zeileis, A. (2013). Influencing elections with statistics: Targeting voters with logistic regression trees. *The Annals of Applied Statistics, 7*(3), 1612–1639.

Sherman, M., Apanasovich, T. V., & Carroll, R. J. (2006). On estimation in binary autologistic spatial models. *Journal of Statistical Computation and Simulation, 76,* 167–179.

Strotmann, H. (2007). Entrepreneurial survival. *Small Business Economics, 28*(1), 87–104.

Tan, H., & See, H. (2004). Strategic reorientation and responses to the Asian financial crisis: the case of the manufacturing industry in Singapore. *Asia Pacific Journal of Management, 21,* 189–211.

Tokunaga, S., Kageyama, M., Akune, Y., & Nakamura, R. (2014). Empirical analysis of agglomeration economies in the japanese assembly-type manufacturing industry for 1985–2000: Using agglomeration and coagglomeration indices. *Review of Urban and Regional Development Studies, 26*(1), 57–79.

Varum, C. A., & Rocha, V. C. B. (2012). Do foreign and domestic firms behave any different during economic slowdowns? *International Business Review, 20*(1), 48–59.

Wheeler, D., & Calder, C. (2007). An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems, 9,* 145–166.

Zhang, H., & Singer, B. (2010). *Recursive partitioning and applications*. New York: Springer.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research, 22*(1), 59–82.