PAPER

Title: The impact of geographical positioning on agri-food businesses' financial failure considering non-linearities

Authors and e-mails of them: Mariluz Maté Sánchez Val. email: mluz.mate@upct.es

Department: Financial and Accounting department

University: Technical University of Cartagena

Subject area: Empresas, emprendimiento y dinámica empresarial en el territorio

Abstract: This study applies spatial econometric techniques to evaluate the effect of external economic agents located in agri-food business environments on the probability of those businesses failing. A logistic regression model that incorporates spatial interactions has been applied to the data. In addition, the article proposes an intermediate step in the modelling process – based on generalized additive models – to determine the specific functional relationships between the probit model and the explanatory variables. The results reveal that geographic location has a significant effect on the probability that agri-food companies fail finding significant non-linearities.

Keywords: business failure, spatial probit, generalized additive model, geographical proximity, urban analysis

JEL codes: G33, Q13,R11

1. Introduction

Do environment elements related to geographical location influence the probability of business failure in the agri-food industry? Answering this question is crucial for maintaining and improving competitiveness within developed countries (Crescimanno et al., 2014). Under adverse economic conditions, companies in this sector mitigate the effects of an economic crisis by stimulating employment and promoting economic growth (Aleksanyan, 2015). In terms of bankruptcy, rates of business failure are higher in sectors other than the agri-food sector (Aleksanyan and Huiban, 2016). Therefore, this sector plays an important role even during challenging economic circumstances. Despite the staple character of the products offered, the last crisis has had a significant negative impact on the agri-food industry, which has reported a dramatic increase in bankruptcy cases. Thus, the identification of the elements which could improve outcomes for the agri-food sector evidently has strong policy implications. The literature on business failure in the agri-food sector is scarce, but a recent French study focuses on this sector. Aleksanyan and Huiban (2016) analyzed the impact of economic and financial factors, at the firm level, on bankruptcy risk in the French food industry and other manufacturing industries. They found that a firm's productivity is an important determinant of its probability of going bankrupt. In addition, they found a positive and significant impact of credit cost on bankruptcy probability; unsurprisingly, lower credit costs reduced this probability.

Regarding other sectors, other studies take into consideration the impact of geography on firms' performance. In particular, there are papers which state that the characteristics of the location where companies are situated – such as the density or the economic development – have an impact on their probability of survival (Fujita and Thisse, 2002). These papers conclude that companies close to other economic agents will have easier access to external resources, such as suppliers and financial providers, and will therefore minimize transportation costs. This, consequently, will make them more efficient and reduce the probability of business failure (Karlsson et al. 2015). In addition, business location triggers different forms of interaction between geographically close firms, and between these firms and their environments. This strengthens the exchange of knowledge and places these companies in more favorable market positions. However, studies focused on the primary sector are scarce and the majority of these primarily concerned with prices (Ham et al., 2012; Lavee and Bahar, 2017).

In order to garner additional understanding of the issue, the purpose of this study is to contrast the role of geography on business failure rates for agri-food companies. In particular, the paper carries out a micro analysis of firms in a specific local environment which evaluates geographical influences of spatial interactions between geographically-close peer-level agri-food companies. It also considers geographical distance from each examined company to its closest external economic agents; the proximity of which could impact on the probability of business failure. Following previous literature, business failure is examined here as a binary classification problem via binary probit regression. This model has been applied extensively in business-failure analysis because of its simplicity and proven performance. In terms of interpretation, the parameters in the binary model are directly explained by the marginal effects. In particular, to take spatial interactions between geographically close agri-food companies into account, we applied a spatial probit model (see Martinetti and Geniaux, 2017, for further details). However, this regression has an important drawback in that it assumes linear relationships between the probit and the explanatory variables. Nevertheless, previous studies in business failure highlight the usefulness of testing for non-linearities. Furthermore, these non-linearities could even be yet more accentuated when geographical variables are included in the model (Khelil, 2016). Therefore, should these not be taken into account, the conclusions drawn from the study may well be biased. In order to overcome this limitation, this study proposes an extension of the probit model by allowing different functional forms between the response and the independent variables. Following Günther et al. (2014), the paper applied generalized additive models (GAMs) (Hastie and Tibshirani, 1990) to determine the functional form between the response in the probit model and the independent variables. As Hastie and Tibshirani propose, we use GAM methodology as an intermediate stage in the modelling of business failure in order to redefine some of the independent variables for a subsequent probit estimation.

The results pinpoint some key geographical variables which may help explain the probability of business failure in agri-food companies highlighting the significant role of spatial interactions between peers. This paper contributes to the current literature in several ways. First, despite the fact that previous literature has found a statistically significant relationship between agri-food companies' failure rate and the characteristics of their environment, few papers to date have examined the relationship between geographic organization and the probability of firms failing in this sector. Second, we use the GAM approach in agri-business failure literature and relax linearity assumptions in the explanatory variables in order to garner richer insights into the impacts of the explanatory variables on agri-food business failure.

This paper is organized as follows: Section 2 presents a set of spatial statistical tools which were used to test the spatial distribution of agri-food business failure and the specific spatial probit model employed. Section 3 lays out the details of the empirical analysis. Section 4 concludes the paper and provides suggestions for further investigation.

2. Methodology

2.1 Spatial autocorrelation tests for qualitative data: Joint-Count tests

In previous literature that utilizes spatial techniques, the Join-Count Test (Cliff and Ord, 1981) has been applied to contrast the existence of spatial patterns when binary variables are examined. This analysis compares the spatial distribution of failed (F) versus non-failed (NF) agri-businesses[1]. Thus, in applying these categories, there were three different networks between agri-food companies: F-F, NF-NF and NF-F. The two first cases (F-F and NF-NF) represent geographically close companies with the same category; whereas F-NF measures the number of geographically close companies which fall into different categories. To establish a connectivity criterion between companies, I defined a weight matrix W which is defined as a binary matrix with elements $\mathrm{w_{ij}}$ $(\mathrm{i,j} = 1, \dots, \mathrm{n})$ equal to 1 if the agri-food companies i and j are neighbors, and 0 otherwise. In addition, the elements of the main diagonal in W have a value of 0 by definition. The model employs a neighborhood criterion based on geographical distance considering as neighbors of each company i its k nearest neighbors. This method is exogenous and therefore prevents any undesirable endogeneity. Based on this connectivity criterion, the *Join-Count* tests for each category were defined as follows in equations (1) and (2):

[1] Previous literature tends to use the letters B (Black) and W (White) to name the two possible categories.

$$J_{FF} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} FF_{ij} \quad (1)$$

$$J_{FNF} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} FNF_{ij} \quad (2)$$

w_{ij} are the elements of the weight matrix W; FF_{ij}=1 if the firms i and j are of the same category F, and FF_{ij}=0 in otherwise; FNF_{ij}=1 if the companies i and j are of different categories and FNF_{ij}=0 in otherwise. n is the number of agri-food companies in the sample. The test for the last category J_{NFNF} is calculated from previous Join Count tests (1) and (2) as follows:

$$J_{NFNF} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} - (J_{FF} + J_{NFNF}) \quad (3)$$

The results from these tests can be positive or negative. A positive and significant result indicates that finding spatial concentration of agri-food companies belonging to the same category F or NF in the spatial distribution of these companies is likely. A negative and significant result indicates a high probability of finding geographically close agri-food companies falling into different categories[2].

2.2 Probit models with spatial autocorrelation

Logistic estimations are widely utilized in business-failure literature (Calabresse et al., 2017). These models are based on variations of the formula (4):

$$\mathrm{E(Y)} = \mathrm{f}(\mathrm{X}'\beta) \quad (4)$$

where $E(Y)$ is the expected value of the variable Y, and $X'\beta$ is the linear combination of different explanatory variables X, with unknown β coefficients and f represents the functional form relating both sides of the equation (4). Thus, each outcome of the dependent variable Y is assumed to be generated from a particular distribution function, f. In the case of business failure literature, the dependent variable has a double response giving rise to a binomial distribution and then the relationship between both sides of the equation (4) is interpreted in terms of probabilities as follows:

$$\mathrm{Probit}\{\mathrm{P(X)}\} = \log\left\{\frac{P(X)}{1 - P(X)}\right\} = \alpha + \sum_{i=1}^{k} \beta_i' X_i \quad (5)$$

where $P(X) = \Pr(Y = 1|X)$, X equals the explanatory variables, k the number of explanatory variables and α is a constant term. However, the estimation of this model

[2] The R package *spdep* was used to compute the Joint-Count tests.

without considering the existence of spatial autocorrelation[3] in the dependent variable, in this case, business failure could cause inconsistent and inefficient estimations (McMillen, 1992). Thus, to assuage suspicions about the possible existence of spatial autocorrelation in the process, a spatial-logistic estimation is required (LeSage, 2009). In this case, (5) is transformed as follows:

$$\mathrm{y} = \alpha + \sum_{i=1}^{k} \beta_i' X_i + \rho W y \quad (6)$$

with $y = \log\left\{\frac{P(X)}{1-P(X)}\right\}$; W is a spatial weight matrix defining the neighborhood structure between the i $(i = 1, \ldots, n)$ units (in this case agri-food companies); ρ is the spatial autoregressive parameter, if $\rho=0$ the spatial probit model collapses to the standard logistic model, and if $\rho\neq 0$, then there is spatial interconnection between geographically close companies. In particular, the logistic specification follows a probit model. Thus, I propose the estimation of a spatial probit model by applying the conditional maximum likelihood estimation procedure developed by Martinetti and Geniaux (2017) and which can be calculated with R by applying the package *ProbitSpatial*.

3. Empirical analysis

3.1. Database and variables

3.1.1 Agri-food companies in the municipality of Madrid

In order to develop this empirical methodology, we selected a sample of agri-food companies in the municipality of Madrid (Spain). The analysis of this territory could have interesting policy implications, given Madrid's national weighting when business-default statistics in Spain are examined. In addition, the local statistical office in Madrid provides detailed micro-data on different economic agents located in this municipality, which offers excellent conditions for carrying out a regional analysis. Furthermore, agri-food companies in Madrid have a significant weight representing 5% of the national income in this subsector (DIRCE, 2017)[4].

3.1.2 Database

Financial and accounting information for agri-food companies comes from the SABI (Iberian Balance Analysis System) database. This dataset provides the accounting registers of more than two million Spanish companies. From this database, agri-food

---

[3] Autocorrelation refers to the fact of being connected geographically close variables. In this case, failed(non-failed) agri-companies surrounded by failed(non-failed) agri-companies

[4] Directorio Central de Empresas: *www.ine.es*

companies were selected by applying the criterion already established by the National Classification of Economics Activities (NACE, 2007). In addition, those companies with headquarters in Madrid were selected. Corporate headquarters are the centers where the companies' financial decisions are generally taken and interactions with external agents take place (Pirinsky and Wang, 2010). Furthermore, we chose Small and Medium size (SMEs) Enterprises due to their significant weight in the current economic system; they account for over 99.8% of the Spanish productive system (DIRCE, 2017). The data were cleansed by dropping those submissions with errors in their financial statements. After this filter, a sample of 1354 agri-food SMEs in the municipality of Madrid remained, with information available for the period 2013-2015. This sample represents a coverage rate of approximately 18% of agri-food companies in Madrid (Office of Local Statistics in Madrid, 2016[5]). Figure 1 shows the geographical distribution of these firms.

--------------------Insert Figure 1----------------------

In addition, the Madrid City Council website (http://datos.madrid.es/) provides geographical localization – determined by the geographical coordinates – of external economic agents whose geographical proximity to companies could impact on their probability of survival (such as industrial parks or logistics centers).

3.1.3 Variables

The dependent variable is a binary variable which indicates if the agri-business is surviving or has failed. Following the models from previous literature, this variable is defined through the economic criteria. This economic definition overcomes some limitations of the legal definition, such as the fact that financially healthy companies may file for bankruptcy for strategic reasons rather than as a result of financial distress (Lin et al., 2012). From this perspective, failed companies were defined as those companies whose accountancy registers showed three straight years of negative shareholders' equity or two straight years of negative shareholders' equity and one additional year without available information (Rubio-Misas, 2008, Tascón and Castaño, 2012). In order to provide robustness to our results, the analysis was replicated using the legal definition of failure. In this case, an agri-food company is classified as failed if it has declared bankruptcy in any year of the period analyzed (2013-2015) (Dakovic et al.

[5] NACE: *http://www.madrid.es/portal/site/munimadrid*

2010). The final sample includes approximately 12-13% of failed companies when the number of failed companies is evaluated with each definition.

As explanatory variables, the model uses geographical and control variables. As geographical variables, *density variables* were defined, which evaluate the impact of inter- and intra- specialization on the probability of agri-food companies' failure. In particular, Sectorial Density (*DensSec*) was defined as the number of agri-food companies located inside of a buffer built from each of these agri-food companies i. The variable Density (*Dens*) was also used to evaluate the diversification of the neighborhood where the company i is situated. This variable is computed by the number of companies operating in different sectors within a buffer from each agri-food company. From previous literature, these variables should show a negative sign by reducing the probability of business failure as the values of the density variables increase. In this sense, intra- and inter- specialization of firms' environment will improve the economies of scale and strengthen informational flows between companies situated there (Weterings and Marsili, 2015). In order to define density variables, the size of the buffer from each company i must be calculated. To compute this value, we follow De Silva and Mc Combe's (2012) procedure and consider different size buffers by changing the radii r_i around each company. Subsequently, those radii that best fit the estimated probit model in terms of likelihood function were selected. By applying this method, the space is assumed to be continuous and, therefore, the connectivity criteria between the economic agents are independent from jurisdictional divisions in the territory.

This research also considers geographical variables that represent the distance from the agri-food companies in the sample to external economic agents; the geographical proximity of these agents could be positive and decrease the probability of business failure. Therefore, a negative sign would be also expected for these variables. Thereby, a variable to evaluate the minimum geographical distance from each agri-food company to the largest agri-food companies (*DMinLC*) is defined in terms of Euclidean distance. The largest companies have more information available to make decisions and can therefore react faster to changes in economic conditions. Companies which are geographically close to the largest companies will receive positive externalities that will reduce their probability of failure (Pirinsky and Wang, 2010). In addition, companies located geographically close to industrial estates (*DMinIE*) will also benefit from the diffusion of knowledge between economic agents working in a specialized area and reduce costs through economies of scale (Mota and Castro, 2004). According to the census of industrial estates, there are twenty-two industrial states in the municipality of Madrid. The model also includes the geographical distance to logistics centers (DMinSL)[6] as a geographical variable. From Google maps, we identified thirteen logistics centers in the municipality of Madrid. Shorter distances to these centers should improve the profitability of agri-food companies by reducing transportation costs.

Apart from the geographical variables, control variables were included in order to consider agri-food companies' economic characteristics (Lohmann and Ohliger, 2017). In particular, the study includes financial indicators that are representative of the internal economic situation of the examined companies. In this sense, *Profitability* ratio (PROF) for each company is calculated as net operating income on total assets. *Debt* ratio (DEBT) is computed as total liabilities on total assets. This ratio is representative of the financial structure of the company. Turnover ratio (TURN) was computed as turnover growth over two consecutive periods. TURN evaluates the evolution in the productive activity of agri-food companies in the sample. In accordance with previous studies, a negative signs for PROF and TURN are expected with the probability of failure decreasing as the values of these variables increase. For the DEBT variable a positive sign is expected. Agri-food company size (SIZE) was defined as the logarithm of total assets. Previous literature finds a negative sign for this variable, which highlights that larger companies enjoy such advantages as economies of scale and greater market presence, which result in reduced probability of failure (Back, 2005). The internationalization of the company was also included as a representative variable. This variable (INT) has a value of one if the company has access to international markets, and zero otherwise. Following previous literature, INT should show a negative sign which implies that companies working in international markets have a reduced probability of failure (Esteve et al., 2004). Finally, the age of the company was included. This variable (*AGE*) is defined as the logarithm of the number of years that have passed since the the company was founded. In accordance with previous literature, this variable should show a negative sign which indicates that older firms have more resources available than younger companies and a lower probability of failure (Rubio-Misas, 2008).

-----------------------Insert Table 1-------------------------

[6] Logistic centres represent to specialized buildings in which firms stock their products (in this case agri-food products) to be redistributed to retailers, to wholesalers, or directly to consumers.

3.2 Spatial distribution for agri-business failure

This study differs from an exploratory spatial analysis where Join-Count tests are computed to determine the existence of co-localized patterns in the spatial distribution of failed (non-failed) agri-food companies in the sample. With this aim, a neighborhood criterion between companies was defined in function of the k-nearest neighbors. In particular, we consider different k-values (k=4,6,8,10). Table 2 shows these results:

-------------- Insert Table 2------------------

Join-Count tests (Table 2) confirm a positive and significant spatial co-localization pattern of failed agri-food companies for k=4,6 (first columns in Table 2). In this sense, the number of connected failed companies (J_{FF}) is larger than the expected value (J'_{FF}). For example, when the four nearest neighbors are considered for each company, J_{FF}=18.5 and J'_{FF}=15.3. Thus, for this neighborhood order, it can be concluded that there is a high probability of finding failed agri-food companies surrounded by failed agri-food companies.

3.3 Estimating agri-business failure

3.3.1 A logistic probit model

This research differs from the estimation of a binary probit model that includes explicative variables to explain the probability of agri-food companies' failure. As already shown, the set of representative geographical variables evaluates the geographical proximity from agri-food companies to external agents that could impact on these companies' probability of business failure. The initial model considered here is shown in (8)

$$P(BF=1|_{Xs}) = \beta_0 + \beta_1 CCTO + \beta_2\, Profitability + \beta_3\, Debt + \beta_4\, \mathrm{Age} + \beta_5\, \mathrm{Size} + \beta_6\, \mathrm{Cereals} + \beta_7\, \mathrm{Fruits} + \beta_8\, \mathrm{Meat} + \beta_9\, \mathrm{Int} + \beta_{10}\, \mathrm{Dens} + \beta_{11}\, \mathrm{DensSub} + \beta_{12}\, \mathrm{DMinLC} + \beta_{13}\, \mathrm{DMinIP} + \beta_{14}\, \mathrm{DMinLS} + \varepsilon \qquad (8)$$

To estimate (8) we counter endogeneity by lagging two years financial explicative variables. In addition, density variables (Dens and DensSS) depend on the estimated radii r_i (with $i = 1,2$ for Dens and DensSS respectively). To determine these values, the probit model (8) is estimated by maximum likelihood (ML) which selected those r_i values that maximized the likelihood function. This procedure is based on the studies of Da Silva and McCombe (2012) and those of Rosenthal and Strange (2003). This produces a radius of one kilometer from each company i in each case (Dens and

DensSS). This value is close to the result obtained in De Silva and Mc Combe (2012). Table 3 (first column) shows this estimation.

------------------------Insert Table 3----------------------

The financial variables presented the expected signs, agreeing with previous studies. In this sense, PROF and TURN show a negative and significant sign whereas DEBT has a positive sign. Regarding firms' characteristics, there was a negative relationship between the size of the company and its probability of failure. This coincides with previous studies which stress the positive effects of company size in terms of economies of scale, which reduces the probability of agri-food companies' failure (Rubio-Misas, 2008). Fruits and Cereals also showed negative and significant results. Thus, companies working in these sub-sectors have a lower probability of bankruptcy. In addition, the internationalization of the company plays a relevant role by reducing the probability of business failure. This is also expected from previous studies, which highlight that companies with exposure to international markets diversify their market risk and reduce probability of failure (Esteve et al., 2004). This analysis also indicates the existence of a negative relationship between the age of the company and its probability of business failure (Berger and Udell, 1998).

Geographical variables present the expected signs. In this regard, density variables and geographical proximity of agri-food companies to industrial estates appear to play a significant role by reducing the probability of agri-food companies' failure. In particular, *Dens* measures the number of companies surrounding each company in the sample. This variable presents a negative and significant sign, which confirms that business concentration reduces agri-food firms' probability of failure (Weterings and Marsili, 2015). There is a similar finding when DensSS is examined. This variable evaluates the number of agri-food companies surrounding each agri-food firm; thus, all of them operate in the same sector. This provides a significant and negative value indicating that high sectoral-density will reduce the probability of failure in this sector (Khelil, 2016).

3.3.2 A Spatial probit model for estimating agri-business failure

The identification of significant spatial autocorrelation in the territorial distribution of failed agri-food companies (see Table 2) could be transferred to the residuals of the probit model (8) if not appropriately modelled. In this case, the initial probit estimation (8) would be inconsistent (McMillen, 1992). In order to improve the probit model, a spatial component was included in (8) by introducing a spatial lag to the model and testing its significance (see model (6)). Given the results in Table 2, a binary weighting matrix that considers the six closest neighbors for each company ($W_{k=6}$) was included. In addition, the weighting matrix ($W_{k=6}$) maximizes the log-likelihood of the probit model (Stakhovych and Bijmolt, 2009). Table 3 (second column) shows the spatial probit results. We find coefficients close to those obtained from the probit regression. In addition, this now produces a significant and positive spatial autocorrelation coefficient $\rho=0.2051$; this indicates that companies surrounded by agri-food companies with high probabilities of failure have themselves a higher probability of failure. Finally, there are some interesting differences in the density variables which become non-significant when this spatial coefficient is included in the model.

3.3.3 An intermediate step: Building the functional forms between the explanatory variables and the probit model

In a probit regression model there is a linear relationship between the independent variables and the probit specification, as seen in Equation (6). In the case that this relationship is not actually linear, then the parameter estimation would be biased. In order to overcome this limitation, GAM methodology (Hastie and Tibshirani, 1990) was used to evaluate any possible non-linear relationships between the probit and the independent variables. To this end, the linear term $\beta_i' X_i$ in (6) is substituted for a smooth non-parametric function $s_j(X_i)$. Through this any given independent variable is converted in such a manner that it resembles the curve in any plot of its smooth function s_j (a so-called GAM plot). Based on this method, as an intermediate step in the model building process, a model was fitted with all the independent variables minus the dummy variables that represent subsector and internationalization. Thus, for the continuous variables, smoothing splines were fitted by applying the *gam* command in R with the default choice of smoothing parameters. Figure 2 shows the resulting GAM plots for those variables that showed significant nonlinear relationships. The bold black line represents the estimated spline and the value of the independent variable is plotted on the x-axis, while the effect on the predictor is plotted on the y-axis. In addition, the density of the sample is observed on the x-axis and the y-axis at zero is represented so as to differentiate between positive and negative effects. In this sense, values above y=0 represents positive relationships with the dependent variable, whereas values below zero on the y axis represent negative relationships. The higher the values on the y-axis the higher the probability of bankruptcy will be. The band representing the 95% confidence interval is shaded gray.

-----------------------Insert Figure 2---------------------

As we can see, only four explanatory variables present nonlinearities when explaining the probability of business failure. For the rest of the variables, we obtain a linear relationship between the explanatory variable and the probit. As regards non-linearities, breaking points were determined for each of the explanatory variables in Figure 2. In this sense, the *age* of the company has a breaking point at x=3 approximately (15-20 years). Thus, the relationship for age values lower than 3 presents a positive effect and increases the probability of bankruptcy, while it turns negative for values higher than 3. This result coincides with previous literature which states that young companies have higher probability of bankruptcy and the trend changes when mature companies are considered. The same happens with the size of the company, but in this case the breaking point is x= 6.5. Two distance variables were found: The distance between the studied companies and large companies, and the distance to logistics centers also present nonlinear relationships which indicate some breaking points. For the distance to large companies (DMinLC), we get roughly x=4. Although prior to this point, another interval can be seen its significance value is low, and consequently it has not been taken it into account. Finally, the distance to Logistic Centers (DMinLS) also have a breaking point in x= 5. Thus, companies with their closest logistics centers inside a radius of five kilometers will receive a positive effect that decreases their probabilities of bankruptcy, but this relationship changes to negative when the distance exceeds five kilometers.

In order to take into account previous non-linearities, the next step in this study is to estimate the model (8) including multiplicative dummy variables in order to consider the breaking points of the non-linear behavior in the variables showed in Figure 2. Table 4 shows the estimation of (9).

$$\begin{aligned} P(BF=1|_{Xs}) = {} & \beta_0 + \beta_1 \mathit{CCTO} + \beta_2\, \mathit{Profitability} + \beta_3\, \mathit{Debt} + \beta_4\, \text{Age} + \ddot{\beta}_4 \\ & \text{Age}(>3) + \beta_5\, \text{Size} + \ddot{\beta}_5\, \text{Size}\ (>6.5) + \beta_6\, \text{Cereals} + \beta_7\, \text{Fruits} + \beta_8\, \text{Meat} + \beta_9 \\ & \text{Int} + \beta_{10}\, \text{Dens} + \beta_{11}\, \text{DensSub} + \ddot{\beta}_{11}\, \text{DensSub}(>30) + \beta_{12}\, \text{DMinLC} + \ddot{\beta}_{12} \\ & \text{DMinLC}(>4) + \beta_{13}\, \text{DMinIP} + \beta_{14}\, \text{DMinLS} + \ddot{\beta}_{14}\, \text{DMinLS}(>5) + \varepsilon \end{aligned} \tag{9}$$

---Insert Table 4--------------------------------------

These estimation results (first column in Table 4) show non-linearities in the representative-distance variables. In this sense, the research implies that geographical proximity from companies to large companies (DMinLC) presents non-linear behavior that reduces the probability of bankruptcy when companies are surrounded by large companies (at a radius lower than four kilometers); however, when this geographical distance increases then the effect changes to negative and results in increased probability of bankruptcy with a significant result. A similar result is found when the geographical distance from companies to logistics centers is considered (DMinLS). Non-linearities also presented themselves when sectoral density is considered (DensSS). In this case, sectoral density decreases the probability of failure, apparently due to the benefits from the exchange of knowledge between geographically close companies working in similar sectors (Fujita and Thisse, 2002). Nevertheless, there are also studies which propose a congestive effect, suggesting that when the presence of a large number of firms operating in the same industrial sector may increase competition, and thus reduce the probabilities of survival (Khelil, 2016, De Silva and McComb, 2012). These results corroborate this expected non-linear relationship. For the spatial dependence of this model on non-linearities, the spatial lag-term was estimated in the second column of Table 4. It produced a positive and significant result – rho=0.3277 when the weight matrix is defined as the 18 closest neighbors; this is the k value, which provides statistically significant results for spatial interaction.

3.3.4 Marginal effects for the probit model

The spatial probit model is not directly interpretable but its coefficients can be transformed to provide the marginal effects (LeSage and Pace, 2009). In particular, previous literature indicates that marginal effects are computed as in (10)

$$\frac{\partial E[y \mid x_r]}{\partial x_r} = \Phi([I-\rho W]^{-1}\beta_r \bar{x}_r) \circ [I-\rho W]^{-1}\beta_r \qquad (10)$$

where $\bar{x}_r$ is the mean of the r^{th} variable, Φ is a standard normal distribution, and $\circ$ is element-by-element multiplication. Table 5 provides the marginal effects for the spatial probit model shown in Table 4, column 2. These marginal coefficients provide different values (in terms of %) to each of the factors considered in the model evaluating business failure.

-------------------------Insert Table 5--------------------------

The marginal effects indicate that most relevant factors for business survival in agri-food companies are those related to the firms' own characteristics. In this sense, the age of the company (4.7%) and the degree of internationalization (5.9%) – as well as the subsector in which the company carries out its main activity – are factors which reduce the probability of business failure. For example, if the company trades for one additional year, then its probability of failure will be reduced by 4.7%. As regards spatial variables, that the model suggests that the differing geographical proximities of agri-food companies to logistics centers and large companies are the most influential factors in decreasing the probability of business failure for these companies.

5. Discussion and Conclusions

The main objective of this study was to examine the role played by geographical proximity between economic agents on the probability of failure in the agri-food sector. With this purpose in mind, an empirical modelling process was applied to a sample of agri-food companies in Madrid. Spatial econometric techniques and non-parametric GAM methods were used to investigate the potential existence of spatial effects and non-linearities in this context. The results confirm the existence of spatial co-localized patterns for pairs of failed agri-food companies. In addition, proximity to external economic agents – demonstrating the important role played by proximity to logistics centers and large companies – reduces the probability of failure in the agri-food sector. Finally, the analysis also confirms the existence of non-linearities in the probit estimation.

The role of geographical proximity in agri-food sector is an attractive research area which has been barely considered prior to this study. These results confirm the significant role of geography on the probability of business failure for agri-food companies. From this analysis, we can draw some interesting conclusions for managers, policymakers and researchers. Managers and policymakers should consider the important roles played by logistics centers, large companies and industrial estates in decreasing the probability that agri-food companies fail. Thus, regional policies could encourage agri-food companies to concentrate around these economic agents. Researchers can examine the limitations of this study – which are provided in more detail below – to push on the development of this form of analysis by applying the methodology to other zones to confirm the results general applicability or testing the relevance for other business sectors.

As already noted, this research has some limitations, which could indeed be considered as lines of investigation for future research. First, the analysis considers a particular period of analysis, one which could have been influenced by the financial crisis. Thus, further studies should been undertaken that consider alternative time periods in order to confirm the generalizability these results. Secondly, these findings show statistically significant results for a particular sample of agri-food companies, but further studies should analyze other geographical contexts to ensure that these results are not specific to some geographical quirk of the particular business environment in Madrid. Finally, GAM methodology is applied as an exploratory tool to determine non-linearities, but further studies could also consider modelling the probability of business failure in agri-food companies using this methodology.

References

Aleksanyan L (2015). La situation économique et financière des entreprises agroalimentaires françaises (1998–2012). *Economie Rurale* 349–350:125–147

Aleksanyan, L. and Huiban, JP. (2016). Economic and financial determinants of firm bankruptcy: evidence from the French food industry. *Review of Agricultural, Food and Environment Studies,* 97, 89-108.

Back, P. (2005). Explaining financial difficulties based on previous payment behavior, management background variables and financial ratios. *European Accounting Review*, 14(4), 839-868.

Berger, A. and Udell, G. (1998). The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. *Journal of Banking & Finance*, 22 (6-8), 613-673.

Calabresse R., Andreeva G. and Ansell J. (2017). 'Birds of a feather' fail together: exploring the nature of dependency in SME defaults, *Risk Analysis: An International Journal*, 1-14.

Cliff, A.D. and Ord, J.K. (1981). *Spatial processes: models and applications*. Pion, London.

Crescimanno M, Galati A, Bal T. (2014). The role of the economic crisis on the competitiveness of the agri-food sector in the main Mediterranean countries. *Agricultural Economics,* 60: 49–64.

Dakovic, R., Czado, C., and Berg, D. (2010). Bankruptcy Prediction in Norway: A Comparison Study. *Applied Economics Letters*, 17, (17), 1739-1746

De Silva, D. and McComb, R. (2012). Geographic concentration and high tech firm survival. *Regional Science and Urban Economics*, 42(4), 691–701.

Esteve, S., Sanchis, A. and Sanchis, J. (2004). The determinants of survival Spanish manufacturing firms. *Review of Industrial Organization*, 25, 251-273.

Fujita, M. and Thisse, J. (2002). Economics of Agglomeration: Cities, Industrial Location and Regional Growth. *Cambridge: Cambridge University Press*.

Günther, C.; Tvete, I., Kjersti, A.,Sandnes, G. and Borgan, O. (2014). Modelling and predicting customer churn from an insurance company. *Scandinavian Actuarial Journal*, 1, 58-71.

Ham, C.; Champ, P.; Loomis,J. and Reich, R. (2012). Accounting for Heterogeneity of Public Lands in Hedonic Property Models, *Land Economics*, 88(3): 444-456.

Hastie, T. J., and Tibshirani, R. J. (1990). Generalized additive models. London: Chapman and Hall.

Karlsson, C., Klaesson, J. and Öner, Ö. (2015). Regional Characteristics and the Survival of New Firms. *In ERSA conference papers* (No. ersa15p6). European Regional Science Association.

Khelil, N. (2016). The many faces of entrepreneurial failure: Insights from an empirical taxonomy. *Journal of Business Venturing*, 31(1), 72-94.

Lavee, D. and Bahar, S. (2017). Estimation of external effects from the quarrying sector using the hedonic pricing method, *Land Use Policy*, 69,541-549

LeSage, J.P. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. CRC Press.

Lin, SM; Ansell, J. and Andreeva, G. (2012). Predicting default of a small business using different definitions of financial distress. *Journal of the Operational Research*, 63, 539-548.

Lohmann, C. and Ohliger, T. (2017). Nonlinear Relationships and Their Effect on the Bankruptcy Prediction, *Schmalenbach Business Review*, 18(3), 261-287

Martinetti, D. and Geniaux, G. (2017). Approximate likelihood estimation of spatial probit models. *Regional Science and Urban Economics*, 64, 30-45.

McMillen, D. (1992). Probit with spatial autocorrelation. *Journal of Regional Science*, 32(3), 335-348.

Mota, J. and Castro, L.M. (2004). A capabilities perspective on the evolution of firm boundaries: a comparative case example from the Portuguese moulds industry. *Journal of Management Studies,* 41, 295-316.

Pirinsky, C. and Wang, Q. (2010). Geographic location and corporate finance: a review. *Handbook of Emerging Issues in Corporate Governance*.

Rosenthal, S.S. and Strange, W.C., (2003). Geography, Industrial Organization, and Agglomeration. *Review of Economics and Statistics*, 85(2), 377–393.

Rubio-Misas, M. (2008). Análisis del fracaso empresarial en Andalucía. Especial referencia a la edad de la empresa. *Cuadernos de Ciencias Económicas y Empresariales*, 1(54), 35-56.

Stakhovych, S. and Bijmolt, T.H.A. (2009). Specification of spatial models: A simulation study on weights matrices. *Papers in Regional Science*, *88*(2), 389-408.

Tascón, M. and Castaño, F. (2012). Variables y modelos para la identificación y predicción del fracaso empresarial: revisión de la investigación empírica reciente. *Spanish Accounting Review*, 15(1), 7-58.

Weterings, A. and Marsili, O. (2015). Spatial concentration of industries and new firm exits: does this relationship differ between exits by closure and by M&A?. *Regional Studies*, 49(1), 44-58.

Table 1. Dependent and independent variables

Variable	Description	
Dependent variable		
BF(Economic Definition)	1 if firm fails, 0 otherwise	12.18%
BF(Legal Definition)	1 if firm fails, 0 otherwise	13.98%
Independent variables		
	Financial variables	Mean (std)
Debt Equity ratio	Total Liabilities to Total Assets.	50.44(25.33)
Profitability ratio	Profits to Total Assets	-0.09(5.52)
TURN ratio	Annual growth rate of revenues ((Revenues (t)/Revenues(t-1))-1)	-0.06(0.3)
Age	Age of company (years from foundation)	23.38 (13.77)
	Firm´s characteristic factors*	n (%)
Micro*	1 if firm size is small (<2 millions Total Assets), 0 otherwise.	968 (71.94%)
Small	1 if firm size is small (2-10 millions Total Assets), 0 otherwise.	305 (22.5%)
Medium	1 if firm size is medium (10-43 millions Total Assets), 0 otherwise.	81 (5.9%)
Fruits	1 if firm belongs to the Fruit Subsector, 0 otherwise.	224(16.5%)
Cereals	1 if firm belongs to the Cereals Subsector, 0 otherwise.	335(24.7%)
Meat	1 if firm belongs to the Meat Subsector, 0 otherwise.	267 (19.7%)
Wine	1 if firm belongs to the Wine Subsector, 0 otherwise.	143 (4.9%)
Milk	1 if firm belongs to the Milk Subsector, 0 otherwise	29(2.1%)
Olive	1 if firm belongs to the Olive Subsector, 0 otherwise	17(1.2%)
Support	1 if firm belongs to the Support Subsector, 0 otherwise	93(6.8%)
Mix	1 if firm belongs to the Mix Subsector, 0 otherwise	246(18.1%)
Local*	1 if firm without internationalisation (not import and/or export), 0 otherwise.	1171 (86.5%)
Imp	1 if firm only import, 0 otherwise.	42 (3.1%)
Exp	1 if firm only export, 0 otherwise.	53 (3.9%)
ImpExp	1 if firm import and export, 0 otherwise.	81 (5.9%)
	Firm´s geographic factors	Mean (std) or %
Dens	Number of total firms of all sectors within a radius r_1.	16.47 (3.79)†
DensSub	Number of firms of the same subsector (NACE-2007, 2 digits) within a radius r_2.	1.18 (0.49)†
DMinLC	Minimum distance from the company to the closest large firm (kilometers)	1.18 (1.30)
DMinIE	Minimum distance from the company to the closest industrial estate (kilometers)	4.40 (1.97)
DMinLS	Minimum distance from the company to the closest to logistic service. (kilometers)	4.52(1.83)

* The reference category in the probit model is a micro firm with low technological intensity and without internationalisation.

†Mean of number firms around each one in sample. r_i =1 represents the radii for defining density variables that maximizes the significant of the model.

Table 2. Join-Count statistics for agri-business failure

k-nearest neighbours	Panel A join-count FF test (Failed-Failed)			Panel B join-count F-NF test (Failed-Non-Failed)			Panel C join-count NF test (Non-Failed-Non-Failed)		
	J_{FF}	J'_{FF}	p-value[1]	J_{FNF}	J'_{FNF}	p-value[1]	J_{NFNF}	J'_{NFNF}	p-
2	10	9.18	0.505	201.6	201.52	0.585	1142	1139.2	0.475
4	18.5	15.3	0.003***	430.5	411.2	1.834	2278	2259.3	2.142
6	17.8	15.0	0.021***	308.5	307.4	0.104	1728.8	1708.5	0.036**
8	36.5	32.7	0.795	848	822.4	1.599	4555.5	4536.7	1.492

[1] p-value of the Z statistic.

*** significant at 1%, ** significant at 5% and * significant at 10%

Table 3. Probit and Spatial probit model estimations

Variable	Probit		Spatial probit	
	Coef.	p-value	Coef.	p-value
Revenues growth (TURN)	-0.0827**	(0.039)	-0.0765*	(0.098)
Profitability ratio (PROF)	-0.0462***	(0.000)	-0.0417***	(0.000)
Indebtedness ratio (DEBT)	0.0361***	(0.000)	0.0280***	(0.000)
Age	-0.0139***	(0.007)	-0.0205***	(0.004)
Size	-0.1334***	(0.000)	-0.1568**	(0.000)
Cereals	-0.4537**	(0.015)	-0.5422**	(0.023)
Fruits	-0. 7308***	(0.000)	-0.7324***	(0.000)
Meat	-0.0827	(0.631)	-0.1522***	(0.372)
Internacionalization (INT)	-0.4644**	(0.044)	-0.3941*	(0.074)
Dens	-0.0755*	(0.089)	-0.0545*	(0.049)
DensSub	-0.1089***	(0.001)	-0.0013***	(0.207)
DMinLC	0.0286	(0.604)	-0.1784**	(0.000)
DMinIE	-0.0761**	(0.035)	-0.0671**	(0.043)
DMinLS	-0.0306	(0.465)	-0.0984***	(0.000)
Spatial effect (ρ)	--	--	0.2051***	(0.006)
Log Likelihood	-222.49		-251	
LR probit vs spatial probit			28.51***	

(*) significant at 10% (**) significant at 5% (***) significant at 1%.

Table 4. Probit and Spatial probit model estimations with nonlinearities

Variable	Probit		Spatial probit (n=18)	
	Coef.	p-value	Coef.	p-value
Revenues growth	-0.0517	(0.450)	-0.0427*	(0.536)
Profitability ratio	-0.0428***	(0.001)	-0.0430***	(0.000)
Indebtedness ratio	0.0358***	(0.000)	0.0348***	(0.000)
Age	-0.3271***	(0.000)	-0.3502***	(0.000)
Age (>3)	0.1186	(0.707)	0.0980	(0.759)
Size	-0.1261***	(0.000)	-0.1281**	(0.000)
Size(>6.5)	-0.0192	(0.867)	-0.0173**	(0.880)
Cereals	-0.4203**	(0.024)	-0.4516**	(0.015)
Fruits	-0.6571***	(0.000)	-0.6736***	(0.000)
Meat	-0.1350	(0.457)	-0.1303***	(0.473)
Internacionalization	-0.4137**	(0.075)	-0.4367*	(0.060)
Dens	0.0002	(0.525)	0.0001	(0.308)
DensSub	-0.0245***	(0.003)	-0.0248***	(0.002)
DensSub(>30)	0.0243***	(0.006)	0.0249***	(0.004)
DMinLC	-0.2643***	(0.001)	-0.1100***	(0.008)
DMinLC(>4)	0.4585**	(0.015)	0.1807	(0.229)
DMinIE	-0.1731**	(0.009)	-0.0677**	(0.037)
DMinLS	-0.2529***	(0.000)	-0.1261***	(0.001)
DMinLS(>5))	0.3181**	(0.021)	0.1536	(0.134)
Spatial effect (ρ)	--	--	0.3277**	(0.010)
Log Likelihood				
LR probit vs spatial probit				

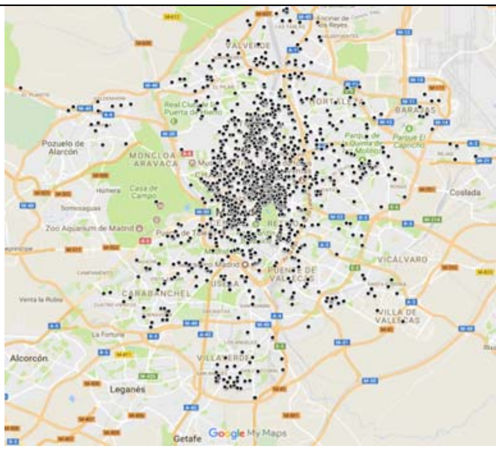
(*) significant at 10% (**) significant at 5% (***) significant at 1%.

Table 5. Marginal effects. Spatial Probit model with non-linearities

Revenues growth	-0.0058**
Profitability ratio	-0.0058**
Indebtedness ratio	0.0047**
Age	-0.0476**
Age (<3)	0.0133**
Size	-0.0174**
Size(>6.5)	-0.0023**
Cereals	-0.0614**
Fruits	-0.0917**
Meat	-0.0177**
Internacionalization	-0.0594**
Dens	-0.0038**
DensSub	0.0039**
DensSub(>30)	0.0004
DMinLC	-0.0149**
DMinLC(>4)	0.0246**
DMinIP	-0.0092*
DMinLS	-0.0245**
DMinLS(>5)	-0.0340**

** Significant at 5% * Significant at 10%

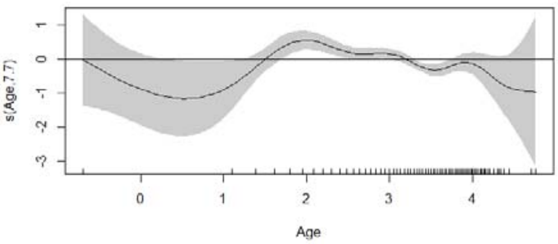
Figure 1



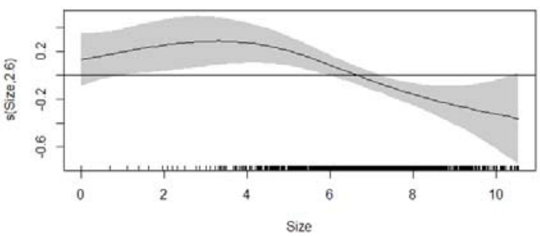
Agri-businesses in Madrid, produced with GoogleMyMaps

Figure2: Spline patterns for the model estimation GAM

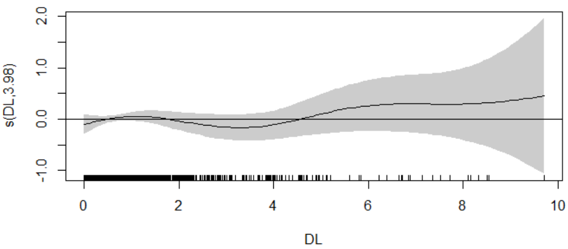
Age in logarithms



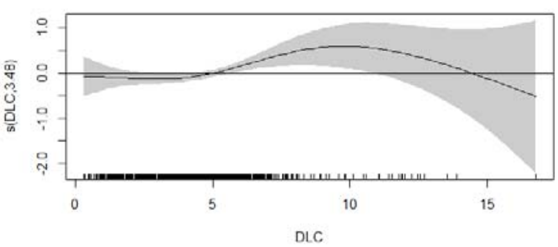
Size



Distance to Large Companies



Distance to Logistic Centers



The other explanatory variables present linear relationships, with k=1