Measuring bank contagion using binary spatial regression models

Citation for published version:
Calabrese, R, Elkink, J & Giudici, P 2017, 'Measuring bank contagion using binary spatial regression models', *Journal of the Operational Research Society*, vol. 68, no. 12, pp. 1503-1511.
https://doi.org/10.1057/s41274-017-0189-4

Digital Object Identifier (DOI):
10.1057/s41274-017-0189-4

Published In:
Journal of the Operational Research Society

Measuring Bank Contagion in Europe Using Binary Spatial Regression Models

Abstract

The recent European sovereign debt crisis clearly illustrates the importance of measuring the contagion effects of bank failures. Indeed, to better understand and monitor contagion risk, the European Central Bank is assuming the supervision of the largest banks in each of the member states. We propose a measure of contagion risk based on the spatial autocorrelation parameter of a binary spatial autoregressive model. Using different specifications of the interbank connectivity matrix, we estimate the contagion parameter for banks within the Eurozone, between 1996 and 2012. We provide evidence of high levels of systemic risk due to contagion during the European sovereign debt crisis.

Keywords: Contagion risk, spatial autoregressive models, European banks, binary data.

1 Introduction

The recent banking crises in the United States and Europe have generated frequent comments about the contagion effects of banks in distress—referred to as systemic risk. The collapse of one major US bank, Lehmann Brothers, triggered a cascade of crises among financial institutions in the US and abroad. Similar fears related to the potential collapse of banks that are "too big to fail" has lead to renewed attention to the containment of risk among banks in the Basel Committee deliberations; within Europe in particular by the European Banking Authority and the European Central Bank.

The definition of systemic risk involves a collection of interconnected institutions that have mutually beneficial business relationships through which insolvency can quickly propagate during periods of financial distress (Billio et al., 2012). Sytemic risk can be decomposed into an idiosyncratic and a systematic (or contagion) component. The first part affects only the health of a single financial institution, while the latter affects the banking system as a whole at the same time.

During the latest financial crisis, Eurozone banks have been suffering considerably, especially in southern countries. In recent years provisions and write-offs of loan credits have increased dramatically and an increasing number of banks required capital injections. These are not easy to obtain, as country solvency risk and financial frictions in Europe has increased cost and availability of funding. However, especially after the European Banking Union Resolution and Recovery Directive (BRRD) has come into force, on January 1st, 2016, it makes a lot of sense to study the characteristics of the Eurozone banking system as a whole, with the aim of understanding which are the most likely factors of default of the banks, whether they are country specific or, rather, idiosyncratic and whether there is a feedback contagion effect.

The main aim of this paper is to propose a method that central banks can use for three main purposes:

- predicting the probability of financial distress for a bank;
- measuring the contagion risk of a banking system;
- performing a stress test analysis.

We propose to use a binary spatial autoregressive model (Fleming, 2004; Calabrese and Elkink, 2014) to achieve these aims. The dependent variable in this context is banking default, since the defaulting of a bank implies that other banks which have direct financial relations with the bank in crisis are losing their assets. We build a dependent variable based on banks which end up in bankruptcy, which are dissolved, or which are liquidated. We apply Klier and McMillen (2008)'s estimators on 4,661 European banks from 1999 to 2012. To obtain an early warning model, the independent variables, given by banks' balance sheet data and macroeconomic variables, are evaluated one year in advance with respect to the time in which the bank distress (response variable) is evaluated. We propose to use the autocorrelation parameter of the spatial model as a measure of contagion risk. Once the model is estimated on the data, it can be used to forecast the probability that a specific bank will be in financial distress in the future. Furthermore, choosing appropriate macroeconomic variables, a central bank can use the model to perform a stress testing analysis on the banking system.

The main advantage of our proposal is that a central bank can achieve three important aims, such as predicting banks' default, measuring contagion risk and performing a stress test analysis, using only one model. Central banks usually use two or three models to complete these objectives. The most used model to gain the first and the third purposes is a logistic regression model Calabrese and Giudici (2014) that ignores the interconnectedness between the financial institutions. On the contrary, our approach can represent both the idiosyncratic effect, given by banks' balance sheet characteristics, and the contagion component, given by the autocorrelation parameter. The widely used framework for modelling contagion risk is network analysis Allen and Babus (2008), which cannot include the idiosyncratic effect of systemic risk and cannot be used for stress testing from a central bank.

After describing the literature review in the next section, Section 3 discusses the binary spatial regression model applied in this analysis. Section 4 describes our approximation of the interbank credit network of banks, while Section 5 provides the regression results. A brief discussion with suggestions for future research follows.

2 Literature review

Research studies on bank failures can be classified into two main streams: the analyses principally focused on financial market data or those on balance sheet data.

Financial market models originate from the seminal paper of Merton (1974), in which the market value of a bank's assets, typically modeled as a diffusion process, is insufficient to meet its liabilities. Because of its practical limitations, Merton's model has been developed into reduced form by Vasicek (1984), leading to widespread diffusion of the resulting model, and the related implementation in Basel II credit portfolio models. In order to implement market models, diffusion process parameters and, therefore, bank default probabilities, this model only requires share price data that can be collected almost in real time from financial markets. Market data are relatively easy to collect and publicly available.

The second stream of studies on bank failures is based on financial book values, taken from publicly available balance sheets. Its diffusion has followed the seminal paper by Altman (1968) and further developments are given by the noticeable studies in Sinkey (1975), Tam and Kiang (1992) and Cole and Gunther (1998). The development of the Basel regulation[1] and the recent financial crisis have further boosted the literature on models based on balance sheet data for banking failure predictions. Recent examples include Arena (2008), Davis and Karim (2008*a*) and Klomp and Haan (2012). This stream has been extended in different ways: interesting developments include the incorporation of macroeconomic components (see, e.g., Calabrese and Giudici, 2014; Koopman, Lucas and Schwaab, 2012; Kanno, 2012; Kenny, Kostka and Masera, 2013) and the explicit consideration of the credit portfolio, as in the Symbol Model of De Lisa et al. (2011), that allows stress tests of banking asset quality and capital, as emphasized in the recent paper by Halaj (2013).

The previous models focused on the prediction of bank failures usually ignore financial contagion, i.e. financial intitutions are connected through bilateral exposures, so the failure of one financial institution can cause difficulties at the financial institutions with claims on it. These difficulties can propagate through the banking system through chains of interbank flows. One of the aims of this paper is to extend scoring models, taking interconnectedness into account by means of a spatial modelling approach. Analogously to the literature on bank failures, also the empirical studies on financial contagion can be initially classified into two main streams, one using financial market data (see, e.g., Billio et al., 2012; Gropp, Lo Duca and Vesala, 2009) and one using banks' balance sheet data (see, e.g., Boss et al., 2004; Mistrulli, 2011; Upper and Worms, 2004).

The most used approach to model connections among financial institutions is network analysis (see Allen and Babus (2008) for a review). Several studies show that a high network connectivity is both positive and negative for financial stability. For a high level of connectivity, there are risk sharing opportunities in the event of small shocks. However, over a given threshold of the connectivity level, the network enables shock propagation (Acemoglu, Ozdaglar and Tahbaz-Salehi, 2015; Elliot, Golub and Jackson, 2014), thus exacerbating crisis. Minoiu et al. (2013) show that if the financial interconnectedness in a country increases and its neighbours' connectedness decreases, the probability of banking crisis is higher.

In this paper, we suggest to use spatial econometrics instead of network analysis to study the interconnectedness of the banking sector. Spatial econometrics (LeSage and Pace, 2009) incorporates dependence among observations that are in any kind of proximity, not only geographical. The main advantages of the spatial econometrics approach over the traditional network analysis is that central banks can use it both as an early warning model, to forecast the failure of a given bank, and as a stress testing technique, the importance of which is growing, as demonstrated by the analyses of the European Banking Authority in recent years. Therefore, relatively to network-based models, spatial econometrics is not only a descriptive technique, but also predictive. As the network is assumed given and exogenous in our proposal, we cannot use spatial econometrics to identify the characteristics of the network that might improve the financial stability of the banking sector, as Acemoglu, Ozdaglar and Tahbaz-Salehi (2015) and Gai, Haldane and Kapadia (2011) show in their studies. Neither the causes of interconnected-

[1] See `http://www.bis.org`.

ness, that represents the main focus of Allen and Babus (2008)'s analysis, nor the mechanism of propagation, which is the objective of Elliot, Golub and Jackson (2014)'s study, can be investigated using the spatial econometrics approach. On the contrary, these important aspects of financial contagion can be deeply examined using network analysis.

3 Spatial logit

The model used in this paper is that of a binary spatial autoregressive structure, whereby the dependent variable is binary and a spatial autoregressive structure is assumed in the underlying latent variable or utility function. Taking the latent underlying quantity to be represented by a continuous variable Y_i^*, we consider the observation mechanism as

$$Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

with $i = 1, 2, \ldots, n$. We implement the spatial structure with an autoregressive model specification, such that

$$\mathbf{Y}^* = \rho \mathbf{W} \mathbf{Y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2}$$

where $\mathbf{Y}^*$ is a continuous random vector, $\mathbf{X}$ represents an $n \times k$ matrix of explanatory variables, the error term $\boldsymbol{\epsilon}$ follows a multivariate logistic distribution and $\mathbf{W}$ is the spatial lag weight matrix with ρ the associated latent parameter. Note that only the latent variable can be used for the spatial lag, since both the models $\mathbf{Y}^* = \rho\mathbf{W}\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and $\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ are infeasible (Anselin, 2002; Beron and Vijverberg, 2004; Klier and McMillen, 2008).

This model implies heteroskedastic errors $\mathbf{e}$ as follows:

$$\mathbf{Y}^* = (I - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{3}$$

where

$$\mathbf{e} = (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}. \tag{4}$$

Calabrese and Elkink (2014) demonstrate through Monte Carlo simulations that among the estimators for binary spatial autoregressive models, the by far least computationally intensive estimator, proposed by Klier and McMillen (2008), is suitable when the data set is sufficiently large and the intensity of the spatial coefficient sufficiently low. This is precisely the type of data we have here, where we study a large number of banks and the collapse of one bank is likely to have some impact on the probability of default of other banks, but not so dramatically as to undermine the banking sector. We therefore apply this estimator to our data.

Following the notation in Calabrese and Elkink (2014), the variance of the error term is

$$var(\mathbf{e}) = var\left[(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}\right] = \sigma_\epsilon^2\left[(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})\right]^{-1}. \tag{5}$$

Let

$$\mathbf{D} = diag(\boldsymbol{\sigma}_\mathbf{e}) \tag{6}$$

be the diagonal matrix with diagonal elements $\boldsymbol{\sigma}_\mathbf{e}$ that represent the root square of the diagonal elements in the matrix (5) and

$$\mathbf{q} = \mathbf{D}^{-1}(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}. \tag{7}$$

Pinkse and Slade (1998) derive Generalised Method of Moments (GMM) moment equations from the likelihood function of a spatial error probit model, for which Klier and McMillen (2008) propose a linearized version based on a logistic distribution with a computationally efficient approximation to estimate the model parameters. Pinkse and Slade (1998) consider the generalized residuals (Cox and Snell, 1968; Chesher and Irish, 1987)

$$\tilde{\mathbf{e}}(\boldsymbol{\theta}) = \mathbf{D}^{-1}E[\mathbf{e}/\mathbf{y},\boldsymbol{\theta}] = \frac{\phi_n[\mathbf{q}(\boldsymbol{\theta})]\{\mathbf{y}-\Phi_n[\mathbf{q}(\boldsymbol{\theta})]\}}{\Phi_n[\mathbf{q}(\boldsymbol{\theta})]\{1-\Phi_n[\mathbf{q}(\boldsymbol{\theta})]\}}, \tag{8}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \rho)'$ is the parameter vector and $\mathbf{D}$ and $\mathbf{q}$ are defined in equations (6) and (7), respectively. The parameter vector $\boldsymbol{\theta}$ is then estimated by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \tilde{\mathbf{e}}'(\boldsymbol{\theta})\mathbf{Z}\mathbf{M}\mathbf{Z}'\tilde{\mathbf{e}}(\boldsymbol{\theta}), \tag{9}$$

where $\tilde{\mathbf{e}}$ is defined in equation (8), $\mathbf{Z}$ is a matrix of instruments, $\mathbf{M}$ is a positive definite matrix and $\boldsymbol{\Theta}$ is the parametric space. In equation (9), Klier and McMillen (2008) let $\mathbf{M} = (\mathbf{Z}'\mathbf{Z})^{-1}$, such that they can propose a nonlinear two-stage least squares method. We define

$$\mathbf{P} = P\{\mathbf{Y} = 1/\boldsymbol{\theta}\} = \frac{\exp[\mathbf{q}(\boldsymbol{\theta})]}{1+\exp[\mathbf{q}(\boldsymbol{\theta})]}, \tag{10}$$

where $\mathbf{q}(\boldsymbol{\theta})$ is defined in equation (7). Taking initial values $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \rho_0)'$ and computing $\mathbf{e}_0$ following equation (4), the gradient terms are computed as

$$\begin{aligned} \mathbf{G}_{\beta i} &= \frac{\partial P_i}{\partial \boldsymbol{\beta}} = \hat{P}_i(1-\hat{P}_i)\mathbf{t}_i \\ G_{\rho i} &= \frac{\partial P_i}{\partial \rho} = \hat{P}_i(1-\hat{P}_i)\left[h_i - \frac{q_i}{\sigma_{ei}^2}\boldsymbol{\Upsilon}_{ii}\right], \end{aligned} \tag{11}$$

where $\mathbf{t}_i$ is the i-th row vector of the matrix $\mathbf{T} = \mathbf{D}^{-1}(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}$, h_i is the i-th element of the vector $\mathbf{h} = (\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{W}\mathbf{q}$, q_i is the i-th element of the vector $\mathbf{q}$ defined in equation (7) and Υ_{ii} is the i-th element of the diagonal of the matrix $\boldsymbol{\Upsilon} = (\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}(\mathbf{I}-\rho\mathbf{W})^{-1}$. At the convenient starting point of $\rho = 0$, it is straightforward to compute the gradients. These gradient terms $\mathbf{G}_{\beta}$ and G_{ρ} are subsequently regressed on $\mathbf{Z}$ and predicted values $\hat{\mathbf{G}}_{\beta}$ and $\hat{G}_{\rho}$ computed. The coefficient estimates of $\boldsymbol{\beta}$ and ρ are then based on regressing $\mathbf{e}_0 + \mathbf{G}_{\beta}\hat{\boldsymbol{\beta}}_0$ on $\hat{\mathbf{G}}_{\beta}$ and $\hat{G}_{\rho}$.

We note that the target of our research are central banks and supervisors. For this reason it is important to derive confidence bands for the ρ parameter.

4 A network of banks

The spatial regression model that we propose is based on an exogenously defined network, where the nodes reflect the individual banks and the ties some value attached to the connection between each pair of banks. The ideal information for this matrix would be information about the claims of any particular bank to any other specific bank, which we denote below as ℓ_{ij}, the liability of bank i towards bank j. This information, however, is not publicly available, and we provide an approximation of this value on the basis of information on the marginals of the matrix.

The Bank for International Settlements (BIS) statistics provides information on the aggregate claims of the entire banking sector in one country to the entire banking sector in another, for a limited number of countries (the informative countries, I) while for some other countries (I^c) only the overall exposure is provided, without details on the country to which the banking sector is exposed. See Giudici and Spelta (N.d.) for a deeper description of BIS statistics. In this paper we use banking sector data available through the BIS, and the individual balance sheet data from Bankscope. All balance sheet data are from the last reporting date.

Define with A the country of bank i, B the country of bank j, and F_{AB} the claims from the banking sector in A to the banking sector in country B, whereby i might be in the same country as j (i.e. $A = B$). This provides a country-to-country connection matrix of the amount of exposures. The connection matrix, which we will denote as W^F, will then be defined as follows:

$$w_{ij}^F = F_{AB}. \tag{12}$$

Where information is unavailable on the specific pair of countries, such that the total exposure of a country's banking sector to sectors abroad is known, but not the detail on the specific dyads, we may assume that the exposure is proportional to the counterpart's market share of the interbank credit market. Doing so, we implicitly assume maximum spread of exposure by banks, as a risk aversion strategy, similar to Upper (2011).

In addition to the international bank credit flows reported by the BIS, data is available from the balance sheets of banks on their exposure to the interbank credit market, for example from the Bureau Van Dijk's Bankscope data base. Such database provides, among other things, the overall liabilities and assets of a bank in the (inter)national credit market, but lacks details on the specific banks or specific countries, where these credit lines are outstanding. In other words, these data provide a reasonable insight into the margins of the full interbank credit network, but not into the individual cells, the specific pairs of banks.

As before, we can assume proportionality of international claims across the banking sector in a particular country, implying that banks avoid risk due to concentration by maximally spreading their interbank credit exposure. For example, if a bank has deposits from other banks that amount to 2% of the total amount of deposits of banks with other banks, we assume that this bank holds 2% of the interbank deposits of each bank—and analogously for loans. This is an unrealistic assumption but a reasonable approximation in the absence of more detailed data on interbank credit exposure. We will denote the resulting connection matrix W^B. In addition, information available from the intercountry F-matrix is juxtaposed such that in the W^B matrix the total flow between countries matches the data available on international bank credits.

In more detail, the following data are available:

- F_{AB} Total claims from the banking sector in country A to country B, F_{AB}, for countries in the set I, for which all bilateral exposures are available.
- $F_{A\cdot}$ Total claims from the banking sector in country A to other countries, $\sum_{B \neq A} F_{AB}$, for countries in set I^c, for which only total exposures are available.
- m_i^c Total claims from bank i to other banks, $\sum_j \ell_{ij}$.
- m_j^l Total liabilities of bank j to other banks, $\sum_i \ell_{ij}$.

We can calculate flows that do not leave the country as

$$F_{AA} = \sum_{i \in A} \sum_{j} \ell_{ij} - \sum_{B \neq A} F_{AB}.$$

We define the marginal flow of claims from country A:

$$M_A^c = \sum_{i \in A} \sum_{j} \ell_{ij},$$

the marginal flow of liabilities to country A:

$$M_A^l = \sum_{i} \sum_{j \in A} \ell_{ij}.$$

We then estimate ℓ_{ij} using the expected value given the marginals as:

$$\hat{\ell}_{ij} = \frac{\sum_j \ell_{ij}}{M_A^c} \cdot \frac{\sum_i \ell_{ij}}{M_B^l} \cdot F_{AB},$$

where $i \in A$, $j \in B$.

For countries in I^c we estimate the overall flows using a similar logic:

$$\hat{F}_{AB} = \frac{(M_A^c - \sum_{C \in I} F_{AC})}{F_{I^c}} \cdot \frac{(M_B^l - \sum_{C \in I} F_{BC})}{F_{I^c}} \cdot \sum_{i \in A} \sum_{j \in B} \ell_{ij} \quad \forall \quad A \in I^c, \quad A \neq B,$$

where

$$F_{I^c} = \sum_{C \in I^c} \sum_{D} F_{CD} + \sum_{D \in I^c} \sum_{C} F_{CD} - \sum_{C \in I^c \wedge D \in I^c} F_{CD}.$$

Time is taken into account through:

$$w_{ij}^B = \begin{cases} \hat{\ell}_{ij} & \text{if } t_j \leq t_i \quad \wedge \quad i \neq j \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

taking t_i to be the year of default of bank i or the last reporting year of bank i in the absence of a default. This implies a constant network structure over time.

—Table 1 about here—

Table 1 shows basic levels of contagion among banks—banks that have failed tend to have more neighbours in their connection matrix that also failed than banks that did not.

We thus provide an alternative approach to the estimation of the interbank credit matrix using a simulation strategy as, for example, in Hałaj and Kok (2013).

5 Data and results

We have decided to concentrate our analysis on the Eurozone banking system for the European sovereign debt crisis. The number of European banks on which we have sufficiently complete data in Bankscope is 4,661 from 1999 to 2012. We consider a bank to be in distress when it is in a legal insolvency status of bankruptcy, or in the stages of dissolution and liquidation. However, we remark that other wider definitions of bank default do exist. For example, some authors (Bongini, Claessens and Ferri, 2001; González-Hermosillo, 1999; Vázquez and Federico, 2012) consider as defaulted banks that have been merged or acquired by other banks. As mergers and acquisitions might have been carried out for strategic aims, rather than for insolvency reasons (Arena, 2008), we consider them as non-defaults. Other authors include state aid and government intervention in their definition of default (see, e.g., Buehler, Samandari and Mazingo, 2009; Brown and Dinc, 2011). However, state aid and interventions are indeed quite heterogeneous and strongly depend on the regulatory framework of the country to which they are applied. As, in our analysis, we consider banks from different countries, we do not include them in our definition of bank failure.

Table 2 provides an overview of the data set with the number of banks in distress and sample size by country and by year.

—Table 2 about here—

In the analysis below, we will separately estimate our models for before 2008 and 2008 onwards, so take account of the different dynamics and different economic context subsequent to the start of the banking crisis. Table 3 gives the frequencies of observed defaults, split by bank size, as measured by equally spaced classes in terms of the logarithm of total assets.

—Table 3 about here—

Table 3 shows that the percentage of defaults decreases as the bank size increases—an instance of the well known "too big to fail" mechanism.

In this analysis we use a combination of balance sheet and macroeconomic variables, in line with earlier work (Calabrese and Giudici, 2014; Calabrese and Osmetti, 2014). As previously discussed, financial ratios associated with the CAMELS rating system can be used to measure bank-level fundamentals related to the asset and liability structure of a bank, assuming that these ratios capture the market, credit, operational, and liquidity risk faced by banks. Taking the balance sheet variables most commonly used in the literature (e.g. Krause and Giansante, 2012) and removing those where high multicollinearity or large amounts of missing data cause significant problems, we propose a model that contains the explanatory variables of leverage, liquidity, loan provisions, return on assets, the loans to assets ratio and (the logarithm of) total assets.

In addition, our model mixes microeconomic with macroeconomic explanatory factors. To this aim, we have included the most important macroeconomic variables: inflation, growth in GDP per capita, and unemployment rates, analogously to Calabrese and Giudici (2014), Calabrese and Osmetti (2014), Kanno (2012) and Koopman, Lucas and Schwaab (2012). To estimate the spatial regression model, we use the method proposed by Klier and McMillen (2008), explained in Section 3. In a simulation study, Calabrese and Elkink (2014) show that the resulting estimator provides accurate estimates for the autocorrelation parameter ρ. In order to obtain a predictive, early warning model (Davis and Karim, 2008*b*; Squartini, van Lelyveld and Garlaschelli, 2013), our model attempts to predict bank failure one year in advance. Therefore, all explanatory variables are evaluated one year in advance, with respect to the time in which the bank failure response variable is evaluated.

We use the two contiguity matrices W^F and W^B defined, respectively, by equation (12) and (13). The former is based on data on international interbank credit flows, assuming equal interconnectedness of all banks within each country, while the latter is based on more detailed information on interbank loans at the bank level, assuming perfectly proportional allocation of credit across banks, within the constraints provided by the data on international flows. Table 4 provides the results for the models based on W^B, as well as regular logistic regressions without spatial component, while Table 5 provides results for the models based on W^F.

—Table 4 about here—

—Table 5 about here—

From the above tables, we first comment on our main object of interest, the estimation of the intensity of autocorrelation in the interbank credit network. The estimates of ρ are relatively similar, around 0.75, ranging from 0.70 to 0.85, with the exception of the models for 2008 onwards. This indicates a relatively high level of autocorrelation in defaults, despite the low number of defaults in the data. Overall, the models for after the onset of the recent financial crisis are less reasonable, which is likely to be due to the low number of bank failures (visible in Table 6), presumably related to the higher level of government interventions in the banking sector.

Although our main focus is on the contagious effect of bank defaults, it is worth checking whether the signs of the coefficient for other variables are in expected directions, which helps to validate our model specification. In many cases, the significant effects show the theoretically expected signs associated with risk taking behaviour on the part of the banks. Bigger banks tend to take more risks, which is visible in the negative sign for the size of the bank (expressed as a logarithm of the total assets). Higher liquidity suggests that the liquidity is used for more financial trading with higher risks, as it occurred during the financial crisis. High amounts of deposits from banks would have a positive sign by the same logic, while the more conservative and less risky strategy would lead to the negative sign we see for loans to other banks. The positive sign on loan provisions is related to the attempt by the bank to deal with bad loans, which is of course correlated with the risk of a bank failure. The sign for leverage is counterintuitive: a higher leverage means more capital to cover unexpected losses (Arena, 2008). The interpretation we have is that more capitalised banks might, again, demonstrate more aggressive behaviour, thus increasing their risk levels. The sign for the coefficient on return on assets is difficult to interpret, as this is a typical proxy for risky behaviour, and the effect on the probability of failure should therefore be positive, theoretically.

We remark that the extended models (3, 6, 9, 11, 13, and 15) are designed to better capture the network interdependence by including a more comprehensive set of control variables. However, high multicollinearity renders interpretation of the coefficients in these models less reasonable. Similarly, the results in Table 5 are based on the assumption of uniform flows, which may be too restrictive. Although we removed all variables which lead to high multicollinearity, in the extended model specifications the remaining multicollinearity can still lead to coefficients with high standard errors, large magnitudes, and occasionally changes of signs. The purpose of this article, however, is not to provide an accurate estimate of the risk drivers of bank failures, but rather to introduce an approach for predicting bank defaults and for measuring the risk of contagion. Our connectivity matrices W^B and W^F are approximations of the true interbank credit matrices based on limited information and strong assumptions. An applied researcher in, for example, a central bank will have more accurate information on interbank lending and will be able to calibrate a more appropriate model specification.

Concerning the impact of macroeconomic variables, our results suggest that the logarithm of the GDP growth is not significant, in either of the considered periods. This may be due to its high correlation with the logarithm of the unemployment rate, which is instead negatively significant in the second (post-crisis) period, for both W^B and W^F based models, suggesting that, in good economic times (characterised by low unemployment rates) banks take more risk and, therefore, are more likely to default. Last, the inflation rate negatively affects the probability of default, a result in line with what can be expected based on the literature.

6 Conclusion

This paper provides a method, based on binary spatial regression models, to estimate the interbank interdependence of bank failures due to credit ties, both national and international.

The method has been applied to the estimation of the contagion parameter for the banks in the Eurozone, for the period between 1996 and 2012. We have found evidence of a relatively high level of autocorrelation, despite the low number of defaults in the data.

The proposed model provides both a description of the contagion (through the spatial component) and a predictive capability, differently from most existing contagion models, which provide either of the two. The model can be easily implemented, as a modification of a classical logistic regression that includes interconnectedness. We believe the findings which can be derived from the model may be useful, especially for supervisors and central bankers who can use it in an early warning monitoring perspective, for measuring contagion risk and for a stress testing analysis. Particularly, after estimating the proposed model, supervisory authorities could intervene if the level of contagion risk, given by the autocorrelation parameter, is too high, e.g. higher than 0.8. If central banks are interested in performing a stress testing analysis, they can use our proposal and compute the number of predicted banks' failures. If this percentage is too high, for example higher than 5%, the central authority could also intervene to ensure the financial stability of the banking sector.

From an applied viewpoint, further research may involve a discussion of implications of the above finding, partly by visualising the effects in terms of the spatial multiplier as proposed by Franzese and Hays (2008), $(\mathbf{I}-\rho\mathbf{W})^{-1}$, which can demonstrate the expected impact of a particular bank failure on the overall banking sector. From a methodological viewpoint, further research may involve employing a different generalised linear model, such as the generalised extreme value regression models discussed in Calabrese and Giudici (2014) and Calabrese and Elkink (2016). Finally, the dependence structure can be extended to the dynamic case (Arakelian and Dellaportas, 2010).

References

Acemoglu, Daron, Asuman Ozdaglar and Alireza Tahbaz-Salehi. 2015. "Systemic Risk and Stability in Financial Networks." *American Economic Review* 105(2):564–608.

Allen, Franklin and Ana Babus. 2008. Networks in Finance. Technical report. Wharton Financial Institutions Center Working Paper.

Altman, Edward I. 1968. "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." *The journal of finance* 23(4):589–609.

Anselin, Luc. 2002. "Under the hood. Issues in the specification and interpretation of spatial regression models." *Agricultural Economics* 27:247–267.

Arakelian, Veni and Petros Dellaportas. 2010. "Contagion determination via copula and volatility threshold models." *Quantitative Finance* 12(2):295–310.

Arena, Marco. 2008. "Bank failures and bank fundamentals: A comparative analysis of Latin America and East Asia during the nineties using bank-level data." *Journal of Banking & Finance* 32(2):299–310.

Beron, Kurt J. and Wim P.M. Vijverberg. 2004. Probit in a spatial context: a Monte Carlo analysis. In *Advances in spatial econometrics. Methodology, tools and applications*, ed. Luc Anselin, Raymond J.G.M. Florax and Sergio J. Rey. Berlin: Springer pp. 169–195.

Billio, Monica, Mila Getmansky, Andrew W. Lo and Loriana Pelizzon. 2012. "Econometric measures of connectedness and systemic risk in the finance and insurance sectors." *Journal of Financial Economics* 104(3):535–559.

Bongini, Paola, Stijn Claessens and Giovanni Ferri. 2001. "The political economy of distress in East Asian financial institutions." *Journal of Financial Services Research* 19(1):5–25.

Boss, M., H. Elsinger, S. Thurner and M. Summer. 2004. "Network topology of the interbank market." *Quantitative Finance* 4(6):677–684.

Brown, Craig O. and I. Serdar Dinc. 2011. "Too many to fail? Evidence of regulatory forbearance when the banking sector is weak." *Review of Financial Studies* 24(4):1378–1405.

Buehler, Kevin, Hamid Samandari and Christopher Mazingo. 2009. Capital ratios and financial distress: lessons from the crisis. Technical report Working Paper.

Calabrese, Raffaella and Johan A. Elkink. 2014. "Estimators of binary spatial autoregressive models: A Monte Carlo study." *Journal of Regional Science* 54(4):664–687.

Calabrese, Raffaella and Johan A. Elkink. 2016. "Estimating binary spatial autoregressive models for rare events." *Advances in Econometrics* 37:147–168.

Calabrese, Raffaella and Paolo Giudici. 2014. "Estimating bank default with generalised extreme value models." *Journal of the Operational Research Society,* . forthcoming.

Calabrese, Raffaella and Silvia Osmetti. 2014. "Modelling cross-border systemic risk in the European banking sector: a copula approach." arXiv:1411.1348.

Chesher, Andrew and Margaret Irish. 1987. "Residual analysis in the grouped and censored normal linear model." *Journal of Econometrics* 34:33–61.

Cole, Rebel A. and Jeffery W Gunther. 1998. "Predicting bank failures: A comparison of on-and off-site monitoring systems." *Journal of Financial Services Research* 13(2):103–117.

Cox, D.R. and E.J. Snell. 1968. "A general definition of residuals." *Journal of the Royal Statistical Society B* 30(2):248–275.

Davis, E. Philip and Dilruba Karim. 2008*a*. "Comparing early warning systems for banking crises." *Journal of Financial stability* 4(2):89–120.

Davis, Philip E. and Dilruba . Karim. 2008*b*. "Comparing early warning systems for banking crises." *Journal of Financial stability* 4(2):89–120.

De Lisa, Riccardo, Stefano Zedda, Francesco Vallascas, Francesca Campolongo and Massimo Marchesi. 2011. "Modelling Deposit Insurance Scheme Losses in a Basel 2 Framework." *Journal of Financial Services Research* 40(3):123–141.

Elliot, Matthew, Benjamin Golub and Matthew O. Jackson. 2014. "Financial Networks and Contagion." *American Economic Review* 104(10):3115–53.

Fleming, Mark M. 2004. Techniques for estimating spatially dependent discrete choice models. In *Advances in spatial econometrics. Methodology, tools and applications*, ed. Luc Anselin, Raymond J.G.M. Florax and Sergio J. Rey. Berlin: Springer pp. 145–167.

Franzese, Robert J. and Jude C. Hays. 2008. "Interdependence in Comparative Politics Substance, Theory, Empirics, Substance." *Comparative Political Studies* 41(4-5):742–780.

Gai, Prasanna, Andrew Haldane and Sujit Kapadia. 2011. "Complexity, concentration and contagion." *Journal of Monetary Economics* 58(5):453–470.

Giudici, Paolo and Alessandro Spelta. N.d. "Graphical network models for international financial flows." *Journal of Economics and Business Statistics*. Forthcoming.

González-Hermosillo, Brenda. 1999. *Determinants of Ex-Ante Banking System Distress: A Macro-Micro Empirical Exploration of Some Recent Episodes (EPub)*. International Monetary Fund.

Gropp, R., M. Lo Duca and J. Vesala. 2009. "Cross-border contagion risk in Europe." *Journal of Central Banking* 5(1):367–381.

Halaj, G. 2013. Optimal asset structure of a bank: Bank reactions to stressful market conditions. Technical report European Central Bank. Working Paper 1533.

Hałaj, Grzegorz and Christoffer Kok. 2013. "Assessing interbank contagion using simulated networks." European Central Bank.

Kanno, Masayasu. 2012. "Default Forecasting Considering Correlation Between Business and Credit Cycles." *Journal of Applied Finance & Banking* 2(5):275–305.

Kenny, Geoff, Thomas Kostka and Federico Masera. 2013. Can macroeconomists forecast risk? Event-based evidence from the euro area SPF. Technical report European Central Bank.

Klier, Thomas and Daniel P. McMillen. 2008. "Clustering of auto supplier plants in the United States: generalized method of moments spatial logit for large samples." *Journal of Business & Economic Statistics* 26(4):460–471.

Klomp, Jeroen and Jakob de Haan. 2012. "Banking risk and regulation: Does one size fit all?" *Journal of Banking & Finance* 36(12):3197–3212.

Koopman, Siem Jan, André Lucas and Bernd Schwaab. 2012. "Dynamic factor models with macro, frailty, and industry effects for US default counts: the credit crisis of 2008." *Journal of Business & Economic Statistics* 30(4):521–532.

Krause, Andreas and Simone Giansante. 2012. "Interbank lending and the spread of bank failures: A network model of systemic risk." *Journal of Economic Behavior and Organization* 83(3):583–608.

LeSage, James and Robert Kelley Pace. 2009. *Introduction to Spatial Econometrics*. Chapman and Hall/CRC.

Merton, Robert C. 1974. "On the pricing of corporate debt: The risk structure of interest rates*." *The Journal of Finance* 29(2):449–470.

Minoiu, Camelia, Kang Chanhyun, V.S. Subrahmanian and Anamaria Berea. 2013. Does Financial Connectedness Predict Crises? Technical report IFM Working Paper.

Mistrulli, P.E. 2011. "Assessing Financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns." *Journal of Banking & Finance* 35(5):1114–1127.

Pinkse, Joris and Margaret E. Slade. 1998. "Contracting in space: an application of spatial statistics to discrete-choice models." *Journal of Econometrics* 85:125–154.

Sinkey, Joseph F. 1975. "A multivariate statistical analysis of the characteristics of problem banks." *The Journal of Finance* 30(1):21–36.

Squartini, Tiziano, Iman van Lelyveld and Diego Garlaschelli. 2013. "Early-warning signals of topological collapse in interbank networks." *Scientific Reports* 3. 3357; DOI:10.1038/srep03357.

Tam, Kar Yan and Melody Y Kiang. 1992. "Managerial applications of neural networks: the case of bank failure predictions." *Management science* 38(7):926–947.

Upper, C. and A. E Worms. 2004. "Estimating bilateral exposures in the German interbank market: Is there a danger of contagion?" *European Economic Review* 48(4):827–849.

Upper, Christian. 2011. "Simulation methods to assess the danger of contagion in interbank markets." *Journal of Financial Stability* 7:111–125.

Vasicek, Oldrich A. 1984. "Credit valuation.".

Vázquez, Francisco F. and Pablo Federico. 2012. *Bank funding structures and risk: Evidence from the global financial crisis*. International Monetary Fund.

	W^B	W^F
default	7.9	9.3
active	3.6	4.6

Table 1: Percentage of defaults among neighbours in the interbank credit network, by bank status.

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	N
Austria	2	0	0	0	3	0	1	2	2	1	0	1	0	0	387
Belgium	3	2	3	1	3	1	0	1	2	0	0	1	0	0	145
Cyprus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28
Estonia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
Finland	0	1	1	0	0	0	0	0	0	0	0	0	0	0	39
France	5	9	8	10	13	5	10	7	12	8	9	6	0	0	682
Germany	10	7	8	3	4	1	4	2	5	1	2	2	0	0	2552
Greece	0	0	0	0	0	0	0	0	0	0	0	0	1	0	31
Ireland	1	2	1	4	1	1	1	5	1	3	0	1	0	0	98
Italy	3	8	6	7	3	7	0	0	2	4	2	1	2	0	1025
Luxembourg	2	4	1	2	6	1	0	0	2	2	1	0	1	0	186
Malta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17
Netherlands	0	4	1	1	1	5	0	2	2	2	0	0	0	0	141
Portugal	0	1	0	0	0	0	0	0	0	0	0	0	0	0	75
Slovakia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21
Slovenia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27
Spain	2	1	0	1	3	1	0	1	1	0	7	7	3	1	306
N	272	244	225	193	159	123	130	111	189	172	197	203	3030	521	5769

Table 2: Number of defaults by year and by country. Last row and column report overall sample sizes.

	active	default	% default
up to 6 hundred	1	2	66
up to 5 thousand	28	8	22
up to 44 thousand	323	45	12
up to 400 thousand	1887	102	5
up to 3 million	2090	104	4
up to 30 million	824	46	5
up to 266 million	239	10	4
greater than 266 million	59	1	1

Table 3: Number of defaults by bank size in total assets.

	1999–2012			1999–2007			2008–2012		
	1	2	3	4	5	6	7	8	9
Leverage	0.40	0.36	0.58	2.29 **	2.20 **	2.30 **	-4.01	-3.30	-6.00 *
	(0.63)	(0.82)	(1.00)	(.074)	(0.87)	(1.10)	(2.60)	(4.50)	(3.20)
Liquidity	0.06	0.02	-0.03	0.08	0.08	0.02	0.09	0.21	23.00 **
	(0.10)	(0.10)	(0.14)	(0.12)	(0.13)	(0.16)	(0.32)	(0.24)	(9.40)
Loan provisions	0.33	0.36 *	0.30	4.18 **	4.00 **	4.80	-0.24	3.40	-29.00 **
	(0.29)	(0.20)	(0.22)	(1.42)	(1.30)	(3.00)	(2.60)	(5.80)	(14.00)
Return on assets	-2.52	-2.50	4.10	-3.69 *	-3.70	-0.70	-10.11 **	-9.50 *	-0.23
	(1.80)	(3.20)	(2.70)	(1.94)	(2.50)	(5.30)	(3.40)	(5.50)	(0.32)
Loans to assets	-1.18 **	-1.30 **	-1.80 **	-1.18 **	-1.20 **	-1.20 **	-0.09	0.24	24.00 **
	(0.31)	(0.35)	(0.52)	(0.40)	(0.46)	(0.58)	(0.71)	(0.74)	(9.90)
log Total assets	-0.12 **	-0.17 **	-0.29 **	0.02	-0.04	-0.06	0.07	0.10	-1.10
	(0.05)	(0.06)	(0.09)	(0.06)	(0.08	(0.10)	(0.10)	(0.13)	(1.60)
log Deposits	0.002	0.01	0.14 **	0.005	0.02	0.09 **	0.006	0.02	0.01
from banks	(0.01)	(0.04)	(0.03)	(0.01)	(0.04)	(0.03)	(0.04)	(0.11)	(0.07)
log Loans and	-0.03 *	-0.03	-0.01	-0.02	-0.02	-0.004	-0.07	-0.07 **	-1.70 *
advances to banks	(0.02)	(0.02)	(0.03)	(0.02)	(0.02	(0.03)	(0.04)	(0.02)	(0.97)
log Net income			-0.02 **			-0.02 **			1.20
			(0.01)			(0.009)			(1.30)
log Gross loans			-0.01			-0.04			-0.31
			(0.02)			(0.03)			(1.10)
Loans to deposits			0.14			-0.004			-3.00 **
			(0.11)			(0.03)			(1.00)
log Inflation			-1.40 **			0.008			15.00 **
			(0.38)			(0.32)			(5.00)
log GDP growth			0.19			-0.09			-0.53
			(0.22)			(0.18)			(1.20)
log Unemployment			-1.00			0.82			-7.80 **
			(0.70)			(0.68)			(2.00)
Intercept	-1.49 **	-0.79	1.90	-1.06	-0.95	-2.10	-20.52	-4.30	4.50 *
	(0.70)	(3.20)	(1.80)	(0.86)	(2.30)	(2.30)	(1000)	(8.50)	(2.40)
ρ		0.76	0.70 **		0.74	0.74		0.96	0.42
		(0.73)	(0.34)		(0.86)	(0.67)		(2.60)	(0.41)
N	5103	5103	3802	1394	1394	1194	3709	3709	2608

Table 4: Models (1), (4) and (7) are logistic regression models without spatial component. Models (2), (5), (8) are logistic spatial autoregressive models based on W^B as approximation of the interbank credit network and models, and a limited set of explanatory variables, while (3), (6),(9) refer to a wider set. Standard errors in parentheses. All models include country fixed effects. Signif. codes: * $\alpha = 0.1$, ** $\alpha = 0.05$

	1999–2012		1999–2007		2008–2012	
	10	11	12	13	14	15
Leverage	0.40	0.67	2.30 **	2.00 *	-4.00	1.00
	(0.81)	(1.00)	(0.87)	(1.10)	(3.90)	(0.74)
Liquidity	0.06	0.04	0.08	-0.02	0.09	-0.02
	(0.10)	(0.14)	(0.13)	(0.15)	(0.17)	(1.30)
Loan provisions	0.33 *	0.30	4.20 **	6.90 **	-0.24	0.50
	((0.19)	(0.21)	(1.30)	(2.70)	(2.20)	(1.60)
Return on assets	-2.50	5.00 *	-3.70	3.50	-10.00 **	0.27 **
	(3.20)	(2.70)	(2.50)	(4.90)	(4.80)	(0.13)
Loans to assets	-1.20 **	-1.30 **	-1.20 **	-1.10 *	-0.09	1.90
	(0.35)	(0.50)	(0.46)	(0.58)	(0.70)	(1.40)
log Total assets	-0.12 **	-0.15	-0.03	-0.007	0.07	0.58
	(0.05)	(0.10)	(0.07)	(0.10)	(0.07)	(0.74)
log Deposits	0.002	0.05	0.005	0.05	0.006	0.04
from banks	(0.01)	(0.06)	(0.02)	(0.05)	(0.02)	(0.03)
log Loans and	-0.03 *	-0.02	-0.02	-0.006	-0.07 **	0.02
advances to banks	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)
log Net income		-0.02 **		-0.02 **		0.06
		(0.008)		(0.008)		(0.45)
log Gross loans		-0.01		-0.04		-0.94 **
		(0.01)		(0.03)		(0.48)
Loans to deposits		-0.02		-0.02		0.10
		(0.03)		(0.03)		(0.45)
log Inflation		-1.20 **		-0.10		9.30 **
		(0.18)		(0.30)		(4.30)
log GDP growth		-0.06		0.01		-1.20
		(0.17)		(0.14)		(0.82)
log Unemployment		-0.12		0.68		-6.10 **
		(0.70)		(0.63)		(1.70)
Intercept	-1.50 **	-0.03	-1.10	-2.20 *	-3.30 **	2.20
	(0.57)	(1.40)	(0.69)	(1.10)	(0.63)	(1.50)
ρ	0.85 **	0.71 **	0.80 **	0.76 **	0.89 **	0.007
	(0.17)	(0.17)	(0.18)	(0.31)	(0.17)	(0.04)
N	5103	3802	1394	1194	3709	2608

Table 5: Models (10), (12), (14), (15) are logistic spatial autoregressive models based on W^F as approximation of the interbank credit network, on a limited set of explanatory variables, while (11), (13),(15) are based on a wider set. Standard errors in parentheses. All models include country fixed effects. Signif. codes: * $\alpha = 0.1$, ** $\alpha = 0.05$