Predictive power of Bayesian CAR models on scale free networks: an application for credit risk

Claudia Berloco [1,2], Raffaele Argiento[3,4] and Silvia Montagna[1,4]

[1] Dipartimento di Scienze Economico-sociali e Matematico-statistiche, Università degli Studi di Torino, Corso Unione Sovietica, 218/bis, 10134 Torino, Italy, (e-mail: claudia.berloco@unito.it, silvia.montagna@unito.it)

[2] Intesa Sanpaolo, Piazza San Carlo, 156, 10121 Torino, Italy

[3] Dipartimento di Scienze statistiche, Università Cattolica Sacro Cuore, Largo A. Gemelli, 1, 20123 Milano, Italy, (e-mail: raffaele.argiento@unicatt.it)

[4] Collegio Carlo Alberto, Piazza Vincenzo Arbarello, 8, 10122 Torino, Italy

ABSTRACT: The monitoring of loans' life-cycle has received the increasing attention of the scientific community after the 2008 global financial crisis. A number of aspects of this broad topic have been addressed by means of several regulatory, statistical and economical tools. However, many issues still require further investigation. In this work, we are interested in the monitoring phase of granted loans to anticipate possible defaults and to investigate whether there is evidence of a liquidity contagion effect within a trade network of firms. To this end, we apply a Bayesian spatial model to a proprietary dataset, and assess its out-of-time predictive performance.

KEYWORDS: Bayesian modelling, spatial modelling, credit risk, CAR model.

1 Introduction

The European Central Bank requires banks to adapt their organization, processes and IT infrastructure in order to give an integrated answer to the non-performing loans problem. Banks can mitigate their credit risk in several steps of the loan life-cycle, for example by foreseeing liquidity problems for those customers which already have a debt to the bank. A timely detection of the transition to financial distress is pivotal, and it will be addressed it in this work leveraging on statistical models and bank data.
Recently, a number of contributions (see, e.g., Dolfin *et al.* , 2019) focused on introducing information on the supply chain connections in credit risk models based on the evidence of trade credit use in European markets. The main idea is that liquidity distress can flow along these connections, and a firm experienc-

ing a period of liquidity distress can delay payments towards its commercial partners, that can consequently experience liquidity distress. The supply chain is seen as a complex network in these studies, but it can also be represented as an adjacency matrix with proper assumptions (Lamieri & Sangalli, 2019).
In this work, we set up a predictive model leveraging Bayesian conditionally auto-regressive (CAR) models for areal data (Banerjee *et al.* , 2003). Specifically, inference is based on a sample of firms from a trade network in a given month, and the predictive performance of a CAR model is tested by estimating the probability of default for both a different sample of firms and for the same sample in the future. Although spatial CAR models have been widely used in ecology, environmental science, biology and medicine, to the best of our knowledge they have not yet been fully exploited in econometrics when dealing with hundreds of thousands of data points interacting in a dynamic complex network (e.g., firms or natural persons).

2 Methodology

With some due simplifications, the monthly goal for a lending bank is to red flag those borrowing firms that have the greatest probability of default (delay in paying their debts to the bank) in the following 3 months. In this paper, we analyse a proprietary dataset of Intesa Sanpaolo collected in a given month, for a total of $n = 944$ firms. Our response variable is a binary indicator such that $Y_k = 1$ if firm k switches to a liquidity distress state in the next 3 months.

The trade network can be represented as a link matrix $W \in \mathbb{R}^n \times \mathbb{R}^n$, with binary entries $w_{kj} = 1$ if $k \neq j$ and k supplier, j customer in the previous year. The link matrix W represents a complex network with a scale free structure (Barabási & Albert, 1999). Further, the Bank database stores several credit and trend information on each specific customer firm, but for the sake of simplicity here we only consider two possible covariates $\boldsymbol{x}_k$ for each firm k. The first covariate, x_k^1, represents the used amount of credit over the granted amount among all Italian financial institutions, while the second, x_k^2, represents the maximum number of days of payment delay recorded in the past 3 months.

We fit a proper CAR specification (Banerjee *et al.* , 2003) to our data as follows:

$$
\begin{aligned}
Y_k &\sim Bernoulli(\theta_k) \\
logit(\theta_k) &= \boldsymbol{\beta}\boldsymbol{x}_k + \phi_k \\
\phi_k | \phi_{-k}, \alpha, \tau, W &\sim N\left(\alpha \frac{\sum_{i=1}^n w_{ki}\phi_i}{\sum_{i=1}^n w_{ki}}, \tau^{-1}\right),
\end{aligned} \tag{1}
$$

Here ϕ_k is a firm-specific spatial random effect incorporating the information contained in the network of relationships W. Conditionally on W, ϕ_k is modelled as a Markov random field, meaning that the value of ϕ_k only depends on the value of its neighbours. Indeed, we expect the probability of default of firm k to increase (decrease) if one of more firms connected with k are (not) in default. Parameters α and τ represent the strength and the precision of the autocorrelation, respectively. The CAR specification is chosen because the information arising from the network (incorporated through ϕ_k) can help explain those default events that are not ubiquitously captured by the linear covariates. Standard priors are placed on α, τ, and $\beta_0, \beta_1, \beta_2$, and estimation of model parameters proceeds via MCMC (Banerjee *et al.* , 2003).

3 Results and conclusions

Testing model (1) on real data, we notice that the posterior distributions of the linear parameters obtained with the CAR model are coherent with those of a standard GLM, which considers covariates $\boldsymbol{x}_k$ only. The overlap between the credible intervals of the linear parameters from the two models implies that the spatial random effects estimated by the CAR model contribute to explain a part of the default phenomenon not entirely captured by firm-specific information. Further, we record very good in-sample performance in terms of area under the curve (AUC), as the GLM has a 0.79 AUC while the CAR specification reaches a 0.89 AUC. Furthermore, model (1) helps in identifying defaulted firms through the spatial random effects. Indeed, Figure 1 (left panel) shows that, for most truly defaulted firms (red dots), the estimated probability that the spatial effect is positive, computed as $\widehat{\mathbb{P}}(\phi_k > 0) = \frac{1}{T-B} \sum_{g=B+1}^{T} \mathbb{1}(\phi_k^g > 0)$, is strictly greater than 50%. Here T is the total number of MCMC iterations, and B denotes the burn-in.

Further, we test the predictive power of the model on a disjoint sample drawn from the network seen at the same timestamp of the training sample (out-of-sample set composed of unseen firms), and on the training dataset but seen six months later (out-of-time set composed of future observations of the same firms used in training). In line with the original aim of spatial CAR models, which are intended to fit data referring to static maps, the model does not generalise in the out-of-sample case. This is an unfortunate result for our credit risk application, as one can instead expect the liquidity distress contagion dynamics to spread with similar strength (α) and precision (τ) in different areas of the trade network. In the out-of-time case, the CAR model shows slightly better predictive performance with respect to the simple GLM, as shown in

Figure 1 (right panel).

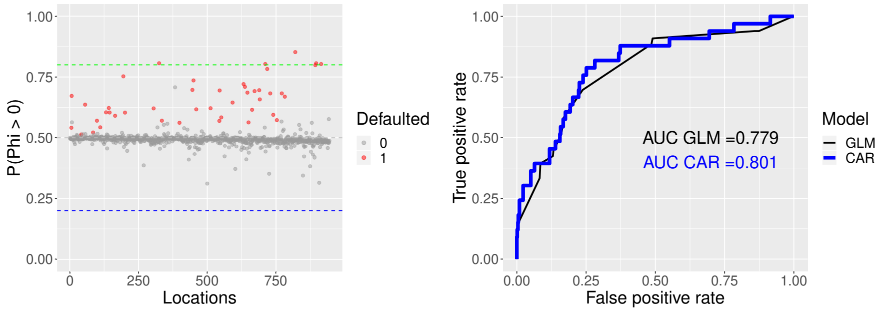

Figure 1. *Left: Estimated probability of a strictly positive spatial effect (i.e., $\widehat{\mathbb{P}}(\phi_k > 0)$) for each firm. Red dots are defaulted firms ($Y_k = 1$) with estimated probability of strictly positive spatial effects greater than 50%. Black dots indicate all other firms. Right: ROC curves and AUC for a GLM considering only covariates $\mathbf{x}_k$ (black) and CAR model (blue) for the prediction six-months ahead with respect to training.*

To conclude, the application of disease mapping methods to a scale free network represents a novelty at present. The encouraging results on the out-of-time set suggest to further investigate spatial modelling of trade networks.

References

BANERJEE, SUDIPTO, CARLIN, BRADLEY P, & GELFAND, ALAN E. 2003. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.

BARABÁSI, ALBERT-LÁSZLÓ, & ALBERT, RÉKA. 1999. Emergence of scaling in random networks. *Science*, **286**(5439), 509–512.

DOLFIN, MARINA, KNOPOFF, DAMIAN, LIMOSANI, MICHELE, & XIBILIA, MARIA GABRIELLA. 2019. Credit risk contagion and systemic risk on networks. *Mathematics*, **7**(8), 713.

LAMIERI, MARCO, & SANGALLI, ILARIA. 2019. The propagation of liquidity imbalances in manufacturing supply chains: evidence from a spatial auto-regressive approach. *The European Journal of Finance*, **25**(15), 1377–1401.