International Journal of Forecasting 39 (2023) 1065–1077

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

# Forecasting short-term defaults of firms in a commercial network via Bayesian spatial and spatio-temporal methods

Claudia Berloco [a], Raffaele Argiento [b,c], Silvia Montagna [d,c,*]

[a] Intesa Sanpaolo, Piazza San Carlo, 156, 10121 Torino (TO), Italy
[b] Università Cattolica del Sacro Cuore, Largo Agostino Gemelli, 1, 20123 Milano (MI), Italy
[c] Collegio Carlo Alberto, Piazza Vincenzo Arbarello, 8, 10122 Torino (TO), Italy
[d] Università di Torino, Corso Unione Sovietica, 218/bis, 10134 Torino (TO), Italy

ARTICLE INFO

*Keywords:*
Credit risk
Bayesian spatio-temporal models
Conditional autoregressive models
Complex networks
Contagion effect

ABSTRACT

To protect financial institutions from unexpected credit losses, during the monitoring phase of granted loans it is of primary importance to foresee any evidence of a contagion of liquidity distress across a network of firms. This term indicates a situation of lack of solvency of a firm (e.g., a customer) that propagates to other firms (e.g, its suppliers), which could consequently face challenges in repaying their own granted loans. In this paper, we look for the evidence of contagion of liquidity distress on an Intesa Sanpaolo proprietary dataset by means of Bayesian spatial and spatio-temporal models. Our results indicate that such models can detect cases of distress not yet apparent from covariate information collected on the firms by instead borrowing information from the network, leading to improved forecasting performance on the prediction of short-term default with respect to state-of-the-art methods.



## 1. Introduction

A reliable assessment of a borrower's credit risk has always been of primary importance for financial institutions to ensure that loans are granted to individuals and/or firms that can promptly repay their debts. However, the global financial crisis of 2008 shed a renewed light on the urgency of monitoring credit risk levels and taking adequate measures to face critical situations. Therefore, Basel Accords,[1] national and international banking authorities and regulators now require banks to adapt their organisation, processes and IT infrastructure to give an integrated answer to the non-performing loans issue. In this regard, banks have designed complex roadmaps of improvements to consistently reduce the level of probable defaults among the banks' customers and optimise the risk-return profile.

Banks can mitigate credit risk in several steps of the loan life cycle. One of the most developed tools consists of assigning a credit rating (or credit scoring) to customers before the loan is granted. Indeed, assigning a credit rating is explicitly required by international agreements. There are different techniques to compute this scoring in the econometric literature. See, for example, Duan et al. (2012), Orth (2012) and Thomas (2000). A different tool consists in foreseeing liquidity problems for those customers (individuals or firms) who already have a debt to the bank. A timely detection (also called early warning) of the transition to the financial distress of such customers is pivotal for the banks to act and prevent the default event on a short-term horizon or to limit their losses. This work will address the early warning task that involves loans granted by a leading Italian commercial bank, Intesa Sanpaolo, to both small corporates and small and medium enterprises (SMEs). We will forecast the probability of default in the next three months using payments bank

* Corresponding author at: Università di Torino, Corso Unione Sovietica, 218/bis, 10134 Torino (TO), Italy.
*E-mail address:* silvia.montagna@unito.it (S. Montagna).

[1] https://www.bis.org/bcbs/history.htm

https://doi.org/10.1016/j.ijforecast.2022.05.003
0169-2070/

data, which represents a novelty in this context of the application.

When approaching the task of forecasting short-term default events from a data-driven point of view, the first step consists in exploring the different sources of information and checking whether informative covariates are available. Informative variables are based on the customer's credit status, expressing whether they already have some delays in repaying their debts or if they have overdrafts towards the bank or other creditors. These covariates are expected to be strongly informative of liquidity distress (from Intesa Sanpaolo internal studies, for example). Further, there may be less obvious exogenous factors that could lead rapidly to liquidity distress. Ideally, this information could be translated into covariates to be included (together with the credit status) in a statistical model to predict the probability of default. A reasonable indicator is the economic context or the market sector in which the customer is working (delivering services or producing goods). Indeed, if the market sector of a customer firm is flourishing, we can expect the firm to keep the amount of its cash flow at least on the levels of the previous years, whilst, in the opposite case, the bank might be alerted about the firm's cash flow. More specifically, the information that one would like to summarise with these "mean" predictors on the trend of market sectors is whether the firm's customers will keep their demand high and pay for their purchases on time and whether its suppliers will continue delivering high-quality products on time. In other words, we would like to extract information on the interconnection of firms' cash flows in their specific supply chain rather than the average indicators of the wealth of the whole market sector. To this end, we will use payment data available to Intesa Sanpaolo to reconstruct firms' supply chains.

Recently, several contributions explored firms' interdependence based on the use of trade credit in European markets. See, for example, Bussoli and Marino (2018) and McGuinness et al. (2018). Here, the main idea is that liquidity distress can flow along these connections. A firm experiencing a period of liquidity distress can delay payments towards its suppliers, which can also consequently experience liquidity distress. Further, these studies introduce the concept of trade credit as a "buffer" for small and medium enterprises, that is, the ability that solid suppliers must absorb the liquidity distress of their customers, thus mitigating the default risk of their counterparts. These studies are based on balance sheet data where the amount of unpaid commercial debts (*account payables*) and the amount of unreceived commercial credits (*account receivables*) have to be declared each year. Thus, the interdependence of each firm to its suppliers/customers is studied at an aggregated level and cannot be decomposed into single contributions of a supplier/customer nor with a short-term (e.g., three-month) frequency. A different approach is proposed in Battiston et al. (2007) and Dolfin et al. (2019). The authors reproduce the structure of a financial network via simulations and study the systemic response to the propagation of bankruptcy under various network structures. Here firm-to-firm interconnections are explicitly modelled, but this is done on simulated data and summarised in a systemic risk computation. In Lamieri and Sangalli (2019), the commercial relationships are derived from payments and invoices data. These are studied both as a complex network (e.g., Newman, 2010) and as an adjacency matrix, namely, a matrix recording the existence of a common border between regions (here, firms). This matrix is typically used in spatial models to parameterise the covariance of individual-specific random effects (Gelfand et al., 2010). In Lamieri and Sangalli (2019), the authors use the commercial relationships to model the propagation of financial and liquidity imbalances of firms summarised over the time interval 2008–2013 and understand the importance of network effects in the transmission of liquidity distress during the financial crisis. Further, the authors explain how much the contagion effect can impact firms' default with respect to balance sheet ratios, but they do not propose a forecasting method. Our context of application is similar to Lamieri and Sangalli (2019), but we will use a more recent dataset, and we will produce a firm-level forecast in a short-term horizon leveraging on the evidence of existing network effects.

This paper aims to build a predictive model to forecast short-term defaults on a set of clients of Intesa Sanpaolo. Our model(s) are embedded into a Bayesian spatial statistics framework. We reconstruct the commercial relationship network among firms by using a novel, proprietary payments dataset, and use such a network to build an adjacency matrix. In this way, we can define two firms as spatial neighbours if they are connected, equivalently, if the relevant entry of the adjacency matrix is non-zero. Several different definition have been used in spatial models, such as geographical proximity (e.g., Beck et al., 2006), correlation of financial indicators (e.g., Catania & Billé, 2017; Fernandez, 2011) or cross-border relationships (e.g., Blasques et al., 2016). Borrowing ideas from Bayesian spatial statistics (Banerjee et al., 2003), we consider a conditional autoregressive (CAR) spatial model and two spatio-temporal variations of the CAR model. We treat our data as real data; each firm is interpreted as an area on a map that may share a common border with other areas (firms) on the map. Such models will combine firm-specific linear covariates with the modelling of random effects. The random effect is interpreted as local (that is, firm-level) evidence of the global spatial or spatio-temporal effects and models the impact of the contagion of liquidity distress on the firm coming from the network. Random effects are modelled as a multivariate normal distribution parametrised using the adjacency matrix. We acknowledge that these techniques are not novel, and have been widely used in medicine, mostly in functional magnetic resonance imaging (e.g., Bowman et al., 2008; Ge et al., 2014; Woolrich et al., 2004), or disease mapping (e.g., Adegboye & Kotze, 2012; Alegana et al., 2013; Watson et al., 2017), among other fields. However, to the best of our knowledge, they have not been fully exploited yet with a forecasting perspective in econometrics when dealing with thousands of data points interacting in a complex network. We will show how these methodologies can be useful in our application context to adjust the forecast of the short-term default produced by

strongly informative predictors (such as the credit status known at the time of prediction) by leveraging the network information. Here the adjacency matrix is not used to describe the geographical proximity among firms but rather to indicate the presence/absence of transactions, thus representing the commercial network.

The rest of the paper is organised as follows. In Section 2, we present the dataset and the peculiarities of the forecasting task. In Section 3, the models used to predict short-term default are described. In Sections 4 and 5, we present an extensive simulation study and the results on the case study, respectively. Conclusions are presented in Section 6.

## 2. Default data

Our data consists of firm-level structured information collected by Intesa Sanpaolo as monthly data between February 2018 and August 2019 on 2592 Italian firms from small to corporate size. Let us first focus on the response variable. Each month, our goal is to forecast the transition from a regular-payment status to a high-level distress status for each firm in the next three months. This condition is defined to be true if one of the following is true: *(i)* the firm is in *regulatory default*, which Basel Accords define as a delay in payments by 90 days or more (with some due simplifications); or *(ii)* it has been assigned a rating score corresponding to either of the two worst rating classes. For simplicity, the transition to high-level distress status will be referred to as a transition to "default" in the following, regardless of the cause. Formally, we set the dependent variable $Y_{tk}$ to be:

$$Y_{tk} = \begin{cases} 1 & \text{if firm } k \text{ switched to default} \\ & \text{between months } t-2 \text{ and } t, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

for each month $t$ and firm $k$. We underline that the three-month gap in defining the response $Y_{tk}$ is driven by the definition of *regulatory default* in Basel Accords. Thus, the response is available on each firm at 16 timestamps from May 2018 to August 2019. The proportion of firms that switches to default at least once over the considered time span is 55%. Moreover, we have 120 new firms in default each month on average (standard deviation = 25). Given that the data is collected monthly, it is possible to recover whether firm $k$ switches to default at a particular month $t$. Table 1 shows the number of new defaults each month over the considered time span. We clarify, however, that this information is provided to the reader for illustrative purposes only. From a modelling perspective, we will retain the three-month gap in the definition of default as indicated by Basel Accords and requested by Intesa Sanpaolo. The average default rate (average of $Y_{tk}$'s) across months and firms is 4.9%, and it displays a moderate increase along the time interval, as shown in Fig. 1. We recall that each decimal point of increase in the default rate may correspond to hundreds of thousands of Euros (or more) of losses for the bank.

Further, firm-specific information is available. Credit and financial information are recorded in the bank databases via various indicators, but in this study, we will consider two covariates for each firm $k$ at time $t$. We will focus on *(a)* the maximum number of days of payment delay recorded in the past three months, $x^1_{kt}$; and *(b)* the used amount over the granted amount among all Italian financial institutions, $x^2_{kt}$. Covariate $x^1_{kt}$ derives from information monitored by the bank daily, whereas $x^2_{kt}$ is a summary of the credit status of the firm over all Italian financial institutions. All Italian financial institutions are required to share the credit status of their customers with the Central Bank of Italy (Banca d'Italia). They gather the Central Credit Register (CR) information and return the statuses of a bank's borrowers enriched with their status towards the other financial institutions. This gathering service provides data with a month of delay; that is, each month the bank receives the CR summary of the previous month. This means that the two covariates are not temporally aligned at the time of prediction. Moreover, both variables are transformed through the weight of evidence (WOE) process. This process is frequently used in credit scoring (see, for instance, Chen et al. (2020), Raymaekers et al. (2021), Soloshenko (2015)), and for continuous variables it consists of two steps: *(i) binning*, namely, dividing the values into non-overlapping intervals (classes); and *(ii) log proportion*, namely, comparing the proportion of "good" subjects ($Y = 0$) to "bad" subjects ($Y = 1$) in each class. Omitting the temporal index for simplicity, the second step is defined mathematically as the logarithm of the ratio of the odds of Bad-to-Good in each class to the odds of Bad-to-Good in the entire sample:

$$\log\left(\frac{b_i/(b_1+b_2+\cdots+b_L)}{g_i/(g_1+g_2+\cdots+g_L)}\right), \tag{2}$$

where $b_i$ and $g_i$, $i \in \{1,\ldots,L\}$ are, respectively, the number of defaulted and non-defaulted corporates in class $i$ and $L$ is the total number of classes. Equivalently, we can write:

$$\log(b_i/g_i) - \log(B/G), \tag{3}$$

where $B$ and $G$ are the total numbers of defaulted and non-defaulted firms in the sample, and Eq. (3) represents the value of the transformed covariate. We do not provide further details on the algorithms to select the optimal binning thresholds (e.g., Zeng, 2014) since we ultimately inherit the bank's standard WOE process here. The resulting values for our dataset are shown in Table 2. For instance, the first value in the Table (−0.86) represents the less risky class for covariate $x_1$. The sign indicates that the difference between the logarithm of the proportion of defaulted and non-defaulted firms in the first class and

**Table 1**
The number of new defaults each month with respect to the previous month between June 2018 to August 2019.

| New defaults | | | | | |
|---|---|---|---|---|---|
| Month | 2018/06 | 2018/07 | 2018/08 | 2018/09 | 2018/10 |
| New defaults | 87 | 103 | 154 | 78 | 118 |
| Month | 2018/11 | 2018/12 | 2019/01 | 2019/02 | 2019/03 |
| New defaults | 134 | 170 | 127 | 106 | 109 |
| Month | 2019/04 | 2019/05 | 2019/06 | 2019/07 | 2019/08 |
| New defaults | 126 | 146 | 132 | 98 | 114 |

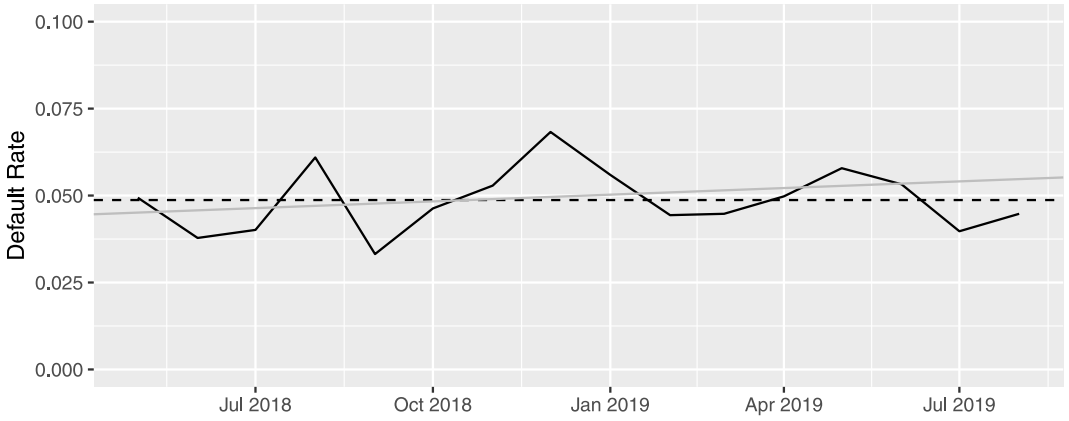



**Fig. 1.** Average (across firms) default rate between May 2018 and August 2019. In grey, the linear regression fit on the default rate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**
The resulting WOE classes for covariates $x_1$ (6 bins in total) and $x_2$ (3 bins in total). We remind that larger values are associated to riskier classes.

| | Weights of evidence | | | | | |
|---|---|---|---|---|---|---|
| $x_1$ | −0.86 | −0.60 | 0.16 | 0.45 | 0.99 | 1.44 |
| $x_2$ | −0.99 | −0.01 | 0.63 | | | |

the logarithm of the same proportion in the full dataset is negative (−0.86). Therefore, the first class has relatively fewer defaulted firms than the full dataset. This number grows up to 1.44 for the riskiest class. We remember that larger values correspond to riskier classes, so we expect positive coefficients linking these covariates to a default status in any generalised linear model.

Finally, information about the trade network is available. Retrieving this information is usually not trivial. Therefore, the analysis of the commercial network and inter-firm dependences is typically made either on simulated networks or on the flattened information of trade credits and debts recorded in firms' balance sheets. Lamieri and Sangalli (2019) provides an exception, and their trade network is also based on a proprietary dataset. Here, we approximate commercial relationships using the cash flow recorded by the bank through payments and invoice discounts. The accuracy of the observed sample increases with the number of payments and invoices the bank can track. In this regard, Intesa Sanpaolo owns the most representative sample in Italy, with market shares no lower than 12% in most Italian regions.[2] The set of commercial interconnections can be represented as a link matrix $W \in \mathbb{R}^n \times \mathbb{R}^n$, with elements:

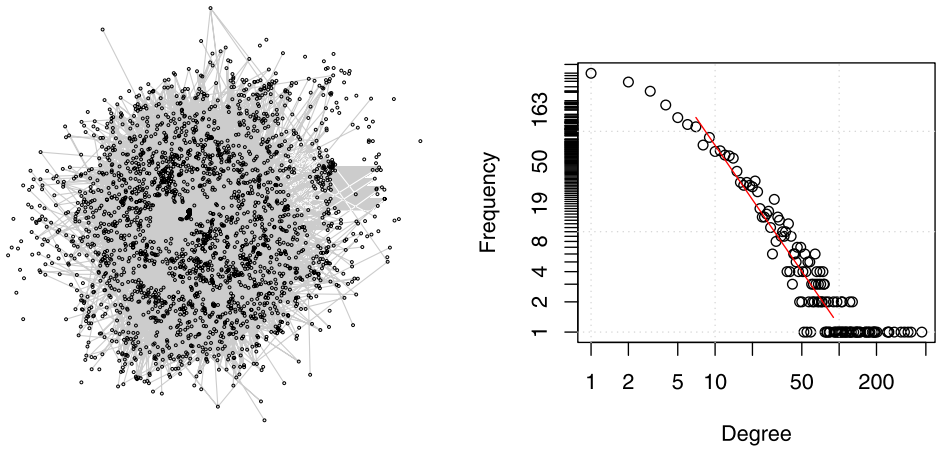
$$W[k,j] = \begin{cases} w_{kj} & \text{if there is a cash flow from } k \text{ to } j \\ & \text{or vice versa, with } k \neq j \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $n$ is the number of firms. $w_{kj}$ is the weight of the link between firm $k$ and another firm $j$, $k, j = 1, \ldots, n$ and $k \neq j$. In this work, they are indicator functions with a value equal to 1 if there is a connection between firms $k$ and $j$ in the past 12 months (we defer to Web Appendix F for a discussion (and assessment) on the use of a non-binary link matrix $W$). The temporal aggregation is chosen to be aligned with the yearly frequency of balance sheet publication. The assumption of undirected connections induces a symmetric adjacency matrix enabling the model to estimate suppliers' dependence on customers and vice versa. Furthermore, we assume that interconnections are static, that is, weights $w_{kj}$'s do not depend on time $t$ as the interconnections between firms do not appear and disappear during the period of observation. This assumption is reasonable in our application because our 16-month time span almost equals the 12-month temporal aggregation to define a connection.

We show our network in Fig. 2(a), omitting only a few firms disconnected from the biggest connected component. This is a typical complex network with a scale-free structure and degree distribution following a power-law $f(k) \propto k^{-\alpha}$ (see, e.g., Easley & Kleinberg, 2010), with coefficient $\alpha = 2.92$ as shown in Fig. 2(b) ($p$-value of the Kolmogorov–Smirnov test is $p = 0.95$, thus failing to reject the null that the data arise from the fitted power-law distribution). The coefficient is computed as the opposite of the slope of a linear fit of a log–log transformation of degrees of nodes (number of neighbours) and their empirical frequency, and it is supposed to be estimated as between 2 and 3 (see, e.g., Easley & Kleinberg, 2010)).

In this paper, we will adopt the link matrix as a building block of a spatial model for areal data (Banerjee et al., 2003). In this framework, areal models can describe the supply chain relationships with the following abstraction: each firm $i$ can be interpreted as a region having a finite set of neighbours $j = 1, \ldots, n$ (commercial partners). Moreover, our firms are embedded in a complex network rich in interconnections and influences. In terms of spatial analysis, we say that areal data exhibit spatial autocorrelation, with observations from areal units close together tending to have similar values. A portion of this spatial autocorrelation may be modelled by including known covariate risk factors in a regression or

[2] See https://group.intesasanpaolo.com/en/about-us.




(a) (b)

**Fig. 2.** Two representations of the trade network. (a) The biggest connected component of the trade network. A few firms disconnected from this biggest component are omitted for simplicity. (b) Degree distribution of the network. The red line shows a linear regression smoother between the first and third quartiles of the data.

classification model. However, it is common for some spatial structures to show in the residuals after accounting for these covariate effects. A remedy for this residual autocorrelation is to augment the model with spatially correlated random effects as part of a Bayesian hierarchical model (Hoff, 2009). Thus, a spatial approach helps us in modelling these observations without the assumption of independence between one another, which is a too strong assumption in (Bayesian and not) linear models (as observed in Lee (2013)).

## 3. Methodology

In this section, we introduce the proposed Bayesian hierarchical spatio-temporal models, and briefly discuss their implementation.

Let us first consider a simple logistic regression model (GLM hereafter) to predict the probability of default of a firm given its financial condition up to time $t-3$:

$$\begin{aligned} Y_{tk} &\sim \text{Bernoulli}(\theta_{tk}) \\ \text{logit}(\theta_{tk}) &= \boldsymbol{\beta x}_{t-3,k} \end{aligned} \tag{5}$$

where $t$ and $k$ are the discrete temporal and firm's indexes, respectively, $\theta_{tk}$ is the probability that $Y_{tk}=1$, and $Y_{tk}$ is the binary target variable indicating whether firm $k$ switched to default between $t-2$ and $t$, with $t=4,\dots,T$ and $k=1,\dots,n$.

While the GLM above is a popular model for credit risk evaluation (Hand & Henley, 1997), it is natural to expect the probability of default of firm $k$ at time $t$, $\theta_{tk}$, to increase if one or more firms trading goods or services with $k$ are in default. To this end, we augment the GLM in Eq. (5) with a firm-specific "spatial" random effect $\phi_k$, which incorporates the information contained in the network of relationships $W$ (Box et al., 2015). We model the process $\{\phi_k\}$ conditionally to $W$ as a Markov random field; the value of $\phi_k$, conditionally to all other firms, only depends on the values of its neighbours (see, e.g., Kinderman & Snell, 1980). We decide first to evaluate each timestamp independently. Thus, we remove the temporal index $t$ and specify the model as follows:

$$\begin{aligned} \text{logit}(\theta_k) &= \boldsymbol{\beta x}_{t-3,k} + \phi_k \\ \phi_k|\boldsymbol{\phi}_{-k},\alpha,\tau,W &\sim N\left(\alpha\frac{\sum_{i=1}^n w_{ki}\phi_i}{\sum_{i=1}^n w_{ki}},\tau^{-1}\right), \end{aligned} \tag{6}$$

where $W$ denotes the adjacency matrix based on the existence of commercial transactions among firms. The $n$-dimensional vector $\boldsymbol{\phi}=(\phi_1,\dots,\phi_n)^\top$ is given a conditional autoregressive (CAR) prior (Banerjee et al., 2003), and parameters $\alpha$ and $\tau$ represent the strength and the precision of the autocorrelation, respectively. The CAR model has proven very successful in disease mapping, see Sun et al. (1999) and references therein. The contagion of default across firms bears some similarities with disease mapping, where default represents the condition of being infected with a disease. Thus, the CAR model may capture a contagion of liquidity distress across firms not explained by the linear covariates (Pace & LeSage, 2010).

The CAR model in Eq. (6) still lacks of time dependence, and each timestamp is analysed separately. However, the probability of default at time $t$ cannot realistically be assumed to be independent of that at time $t-1$. Following (Gelfand, 2003), we thus extend Eq. (6) to account for temporal autocorrelation explicitly as follows:

$$\begin{aligned} \text{logit}(\theta_{tk}) &= \boldsymbol{\beta x}_{t-3,k} + \psi_{tk} \\ \psi_{tk} &= \phi_k + \chi_t \\ \phi_k|\boldsymbol{\phi}_{-k},\alpha,\tau,W &\sim N\left(\alpha\frac{\sum_{i=1}^n w_{ki}\phi_i}{\sum_{i=1}^n w_{ki}},\tau^{-1}\right) \\ \chi_t &\sim AR(1,\rho,\sigma). \end{aligned} \tag{7}$$

The random effects $\psi_{tk}$ are specified as the sum of two separate autoregressive processes: $\phi_k$, modelling the spatial dependence among adjacent firms (constant over time) and $\chi_t$, describing the temporal autocorrelation. The

$\chi_t$'s constitute *temporal* random effects, and are modelled as a stationary autoregressive process $AR(1, \rho, \sigma)$ with 0-mean. Parameters $\rho$ and $\sigma$ are estimated, representing the temporal autocorrelation and noise scale, respectively.

Finally, we consider a different specification for the logit of $\theta_{tk}$:

$$\text{logit}(\theta_{tk}) = \boldsymbol{\beta x}_{t-3,k} + \beta_3 t + \psi_{tk}, \tag{8}$$

where the linear component of the logit has been enriched with a temporal deterministic trend, $\beta_3 t$, to account for (the possible) non-stationarity in the temporal process, and $\psi_{tk}$ is modelled as in Eq. (7). The *deterministic* trend component has been chosen among other more sophisticated hypotheses (e.g., a *stochastic* trend, see Box et al., 2015)) because we only have 16 timestamps in our dataset, and a more complicated model could result in being over-parametrised.

### 3.1. Prior elicitation and posterior inference

We complete the Bayesian specification of the spatio (-temporal) models via prior elicitation for the remaining model parameters. We rely on non-informative priors as in Jin et al. (2005), and specify: $\beta_i \sim$ Normal(0, 100), i = 0, …,3, for the linear regression coefficients; $\alpha \sim$ Uniform(0, 1) and $\tau \sim$ Gamma(1, 0.1) for the spatial random effects (for all $t$'s for the CAR model in Eq. (6)), where the Gamma distribution is parametrised in terms of the shape and rate parameters, respectively; $\chi_1 \sim$ Normal$(0, \sigma^2/(1+\rho^2))$, $\rho \sim$ Uniform$(-1, 1)$, and $\sigma^2 \sim$ Inverse-gamma(1, 0.1) for the temporal autoregressive process. Refer to Banerjee et al. (2003) for a discussion on the choice of non-informative priors for the spatio-temporal models considered here. As suggested by one reviewer, we also assessed sensitivity of results to the choice of priors by specifying informative prior distributions on the parameters for which prior information was available, whilst shrinking the spatio-temporal component towards the GLM a priori. To this end, guided by the linear fit to the average default rate in Fig. 1, we placed an informative prior on $\beta_3$ and chose priors for $\alpha$ and $\rho$ inducing shrinkage towards the GLM (i.e., priors centred on zero and with narrow variance). Out-of-sample forecasting performance, reported in Section 5 for the non-informative case, was unaffected by different prior specifications, and therefore results for the informative case were not included in this manuscript.

Posterior samples of the model parameters are obtained via Markov Chain Monte Carlo (MCMC). However, the full conditional posterior distributions are not available in closed form for (most of) the parameters of the models above. To ensure faster convergence to the high-dimensional posterior distribution than simpler methods (e.g., Metropolis within Gibbs), we rely on Hamiltonian Monte Carlo (Neal, 2011) for posterior sampling, and on STAN (Stan Development Team, 2018) for its implementation. We defer a discussion on STAN and HMC to Web Appendix A, which also includes code to sample from the CAR model (Eq. (6)) in STAN.

To evaluate the forecasting performance (Section 3.3), we fit the models to a dataset comprising of timestamps 1 to 8 (training set) and predict the default response on the 12th timestamp (January 2019 data, hereafter). The four-month gap in prediction is due to the three-month delay in reconstructing the response, plus an additional one-month delay in collecting all the data. We then repeat the forecasting exercise by considering a bigger training set of default data collected from timestamps 1 to 12 and a test set at the 16th timestamp (May 2019 data, hereafter). In Web Appendix E, we repeat the analysis considering longer time horizons for the prediction of the case study.

### 3.2. Competing models

We compare the forecasting performance of the models in Eq. (6), (7) and (8) to that of two competitors. The first competing model is the standard GLM given by Eq. (5). The GLM acts as a baseline model. It can be seen as the degenerate form of the spatio-temporal models if the spatial and temporal components were estimated to be null (i.e., with $\alpha = 0$ and $\rho = 0$). For fitting the GLM, a non-informative prior is placed on $\boldsymbol{\beta}$ as outlined in Section 3.1.

The second comparative method is taken from the theory of information cascades in complex network studies (Roukny et al., 2013). In this context, some network nodes are observed to fail at the beginning of the process, and their failures increase the distress of the neighbouring nodes. When this load exceeds the individual robustness, the node fails, and this process can replicate iteratively across the network in time. The failure (also called activation) probability of a node is often represented by a parameter $p$. This type of propagation dynamics has been studied and modelled with several variants in both social and economic contexts (Hurd, 2016; Kempe et al., 2003), and here we apply it with an activation probability parameter $p = 1$. The method can be thus regarded as a naïve or random walk method (Hyndman & Athanasopoulos, 2019), which prescribes to forecast a process state through its own lagged value. Here the lag is not computed in time but rather in space, and out-of-sample prediction will be made as follows:

$$Y_{tk} = \begin{cases} 1 \text{ if } W_k Y_{t-h} > 0 \\ 0 \text{ otherwise,} \end{cases} \tag{9}$$

where $W_k$ is the $k$th row of the adjacency matrix $W$, and $Y_{t-h}$ are the last known default states of the firms at the time of prediction ($h = 4$ for our case study). Model in Eq. (9) will be called NetNaïve (as Network Naïve) hereafter.

### 3.3. Performance measures

We compare both the in-sample and forecasting performance of the methods above via the Receiver Operating Characteristic (ROC) curve (and related Area Under the ROC Curve (AUC)) and the recall index (REC). The REC is defined as the number of correctly predicted defaults over actual defaults. However, a clarification is necessary here. The spatio (-temporal) models in Section 3 and the GLM are probabilistic methods, therefore providing an estimate

of the probability of default to be used for constructing the ROC and computing AUC and recall. However, NetNaïve is a hard classifier; it only provides a binary prediction of default. To ensure a fair comparison of the models, we adopt the business perspective of the bank, which organises the workload generated by the analysis of all possible predicted default alerts. Therefore, once we can analyse a fixed percentage of files, such as the worst (or *riskiest*) 3, 5, 10 or 50% of the population, we are also able to compute the proportion of truly defaulted firms in these subsets of files. We name these proportions *quantile* RECs, and will be denoted as $QREC_{0.03}$, $QREC_{0.05}$, $QREC_{0.1}$ and $QREC_{0.5}$ hereafter. For probabilistic models, the ranking of firms is done by sorting the estimated posterior probability of default from largest to smallest, whereas we rank firms based on the number of nearest neighbours in default (from a maximum of 4 to a minimum of 0) for the NetNaïve method.

A quantile REC has a natural upper bound, which is given by the minimum between 100% and the ratio between the selected population proportion (e.g., the riskiest 50%) and the actual default rate. These upper bounds will be expressed explicitly in parenthesis in the following tables. For example, $QREC_{0.03}(0-5)$ indicates that we are analysing the riskiest 3% of the population, and $QREC_{0.03}$ can range from a minimum of 0 to a maximum of 5 (the larger the value, the better). Quantile RECs can be computed whenever we predict default (possibly monthly). For out-of-sample forecasting performance, we predict default at a particular month (e.g., May 2019), and the quantiles RECs will refer to that month. If we evaluate performance over multiple months (e.g., in-sample forecasting performance), the average *QREC* will be reported.

Finally, we clarify that while it is possible to compare models according to other performance criteria (e.g., precision), monitoring the recall is most important in our application. Therefore, motivated by our business perspective and for simplicity, we only report (quantile) recall results hereafter.

## 4. Simulation study

In this section, we investigate the performance of the models above in three different simulation scenarios. Before describing the scenarios, we define common elements across settings. We generate data on 2500 firms and for a total of $T = 18$ timestamps. These 2500 firms are supposed to be collected on a 50-times-50 grid, and the adjacency matrix $W$ is a squared 2500-times-2500 symmetric matrix whose elements $w_{kj}$ are equal to 1 if firms $k$ and $j$ have a connection (common border) in the grid. Further, two covariates are generated for every firm $k$ and timestamp $t$ by sampling from a uniform distribution, $x^j_{tk} \sim \text{Uniform}(0, 1)$, for $j = 1, 2$. We set priors as in Section 3.1, but choose here informative prior distributions on $\tau$ and $\sigma^2$. Posterior sampling proceeds as described in Section 3.1 and Web Appendix A. In particular, we run four chains for each model with 2500 iterations, a thinning parameter of 2, and 1000 warmup steps.[3] We train the models on all the data up to the 10th (14th) timestamp and forecast the 14th (the 18th). The in-sample-performance reported below refers to the 10th timestamp, whereas the out-of-sample performance is the average performance on the two forecasts.

### 4.1. Three simulated settings

Three data generating processes are set in line with the three proposed models in Eq. (6), (7) and (8). Specifically, the three scenarios are the following:

1. Data with spatial autocorrelation but without a temporal autocorrelation according to Eq. (6), with $\boldsymbol{\beta} = (1, 4, -6)^\top$. The spatial parameters are set to $\alpha = 0.8$ and $\tau = 0.1$. The drawn samples of $\theta_{tk}$ at timestamp $t = 1$ are displayed in Figure 1 in Web Appendix B.
2. Data with both spatial and temporal autocorrelation in line with Eq. (7), with all parameters set as in the previous scenario, and with $\rho = 0.7$ and $\sigma = 1$. Realisations of $\chi_t$ and $\theta_{tk}$ are displayed in Figure 2 in Web Appendix B.
3. Data with both spatial and temporal autocorrelation and with a linear increasing trend in time as per Eq. (8), with all parameters set as in the previous scenario and the new parameter $\beta_3 = 0.1$, representing an increasing trend. Realisations of $\chi_t + \beta_3 t$ and $\theta_{tk}$ are displayed in Figure 3 in Web Appendix B.

Further, we performed an additional simulation under setting 1 but considering as true $W$ a portion of the real commercial network described in Section 2. We discuss this simulation in Web Appendix C.

### 4.2. Model performance comparison

Tables 3–5 report the posterior mean estimates of the model parameters along with 95% credible intervals. Here, we see that the three proposed models estimate coefficients $\beta_0$, $\beta_1$, and $\beta_2$ are better than the GLM. As discussed in Pace and LeSage (2010), we confirm that the spatial random effects can capture some exogenous conditions that are uncorrelated with the explanatory variables in the linear component. If such exogenous conditions are not modelled explicitly (as in the GLM), there is a lack of accuracy in estimating the coefficients of the linear covariates. Furthermore, the models, including the temporal components (named "AST" and "ASTtrend" in the Tables), can estimate the spatial effects more accurately than the mere spatial model ("CAR") in the second and third scenarios. Indeed, the credible intervals for $\alpha$ are narrower about the true value for both spatio-temporal models. This may be because the spatio-temporal models do not try to attribute all the residual variation to the spatial random effects but also explain the variability in the data via the temporal components. Further, the

[3] The execution time is less than 2 hours for each of the models on a personal computer with the following characteristics: Intel(R) Core(TM) i5-7300U CPU @ 2.60 GHz 2.71 GHz (4 cores, 8 GB RAM).

**Table 3**
Mean and 95% credible intervals of the posterior distributions of the model parameters for the spatio (-temporal) models in Section 3 and the GLM (Eq. (5)) along with their true values in the first simulation scenario of spatially autocorrelated data.

| Param (real) | CAR **mean** (2.5%, 97.5%) | AST **mean** (2.5%, 97.5%) | ASTtrend **mean** (2.5%, 97.5%) | GLM **mean** (2.5%, 97.5%) |
|---|---|---|---|---|
| $\beta_0$ (1.0) | 1.03 (0.63, 1.59) | 1.04 (0.39, 1.67) | 1.01 (−0.05, 2.08) | 0.73 (0.66, 0.81) |
| $\beta_1$ (4.0) | 3.59 (2.86, 5.19) | 3.88 (3.73, 4.04) | 3.88 (3.73, 4.02) | 2.62 (2.52, 2.73) |
| $\beta_2$ (−6.0) | −5.57 (−7.95, −4.60) | −5.93 (−6.11, −5.75) | −5.93 (−6.10, −5.76) | −4.05 (−4.17, −3.93) |
| $\beta_3$ (0.0) | – | – | 0.01 (−0.23, 0.26) | – |
| $\alpha$ (0.8) | 0.81 (0.54, 0.95) | 0.74 (0.66, 0.81) | 0.74 (0.66, 0.81) | – |
| $\rho$ (0.0) | – | 0.27 (−0.90, 0.97) | 0.47 (−0.81, 0.99) | – |

**Table 4**
Mean and 95% credible intervals of the posterior distributions of the model parameters for the spatio (-temporal) models in Section 3 and the GLM (Eq. (5)) along with their true values in the second simulation scenario of spatially and temporally autocorrelated data.

| Param (real) | CAR **mean** (2.5%, 97.5%) | AST **mean** (2.5%, 97.5%) | ASTtrend **mean** (2.5%, 97.5%) | GLM **mean** (2.5%, 97.5%) |
|---|---|---|---|---|
| $\beta_0$ (1.0) | 0.47 (0.15, 0.90) | 0.95 (−0.16, 2.67) | 1.78 (0.17, 3.54) | 0.35 (0.28, 0.42) |
| $\beta_1$ (4.0) | 3.52 (2.76, 5.37) | 3.87 (3.71, 4.02) | 3.87 (3.71, 4.03) | 2.46 (2.35, 2.57) |
| $\beta_2$ (−6.0) | −5.52 (−8.50, −4.49) | −5.95 (−6.14, −5.77) | −5.96 (−6.14, −5.78) | −3.85 (−3.97, −3.74) |
| $\beta_3$ (0.0) | – | – | −0.20 (−0.53, 0.11) | – |
| $\alpha$ (0.8) | 0.86 (0.63, 0.97) | 0.79 (0.72, 0.85) | 0.79 (0.72, 0.85) | – |
| $\rho$ (0.7) | – | 0.43 (−0.32, 0.97) | 0.28 (−0.71, 0.97) | – |

**Table 5**
Mean and 95% credible intervals of the posterior distributions of the model parameters for the spatio (-temporal) models in Section 3 and the GLM (Eq. (5)) along with their true values in the third simulation scenario of spatially and temporally autocorrelated data with a trend.

| Param (real) | CAR **mean** (2.5%, 97.5%) | AST **mean** (2.5%, 97.5%) | ASTtrend **mean** (2.5%, 97.5%) | GLM **mean** (2.5%, 97.5%) |
|---|---|---|---|---|
| $\beta_0$ (1.0) | 4.37 (3.52, 5.96) | 1.39 (0.11, 2.83) | 0.34 (−1.82, 2.54) | 0.80 (0.73, 0.88) |
| $\beta_1$ (4.0) | 3.29 (2.48, 4.58) | 3.85 (3.70, 4.00) | 3.85 (3.69, 4.01) | 2.38 (2.28, 2.48) |
| $\beta_2$ (−6.0) | −5.69 (−7.68, −4.59) | −5.86 (−6.04, −5.67) | −5.86 (−6.05, −5.68) | −3.63 (−3.76, −3.54) |
| $\beta_3$ (0.1) | – | – | 0.23 (−0.22, 0.69) | – |
| $\alpha$ (0.8) | 0.69 (0.19, 0.96) | 0.73 (0.65, 0.81) | 0.73 (0.65, 0.81) | – |
| $\rho$ (0.7) | – | 0.31 (−0.69, 0.95) | 0.33 (−0.62, 0.97) | – |

lower precision of the CAR model may also be because it is evaluated on a single timestamp rather than on the 10 timestamps used by the other models. Concerning the temporal autocorrelation $\rho$ of models AST (Eq. (7)) and ASTtrend (Eq. (8)), we notice wide credible intervals. However, this uncertainty seems to be justified by looking at possible realisations of the data generating processes (see Figure 2 in Web Appendix B with an apparent decreasing trend, and Figure 3 in Web Appendix B). Thus, we do not explain this uncertainty as a limitation of the models but rather as a possible interpretation of the data. To investigate this aspect, we produced further extensive simulations (not reported here) with a longer dataset of 100 timestamps and noticed the ambiguity in estimating $\rho$ vanished.

Table 6 reports the models' in-sample performance averaged across the timestamps in the training set. The spatio (-temporal) models achieve a good improvement in terms of AUC in all scenarios over the GLM and the NetNaïve model (which performs poorly but consistently above the random choice). In terms of QRECs, there is a smaller improvement, but this aligns with a smaller margin for improvement according to the maximum value attainable by these proportions.

Table 7 reports the models' out-of-sample predictive performance. Here we see that the AUC increases by 2 (CAR) to 9 points (spatio-temporal models) over the GLM, and the QRECs almost reach their upper bounds (e.g., see the $REC_{0.03}$ and $REC_{0.05}$ of the spatio-temporal models in the first and second scenarios). Although there is an (expected) drop in performance out-of-sample, the spatio (-temporal) models do not lose their advantage over GLM and NetNaïve. This phenomenon is expected for methods that try to model unexplained effects, including them as parameters, thus adding degrees of freedom and stretching the balance in the bias–variance trade-off.

## 5. Case study

In this section, we present estimation and forecasting performances of the spatio (-temporal) models on real data from the proprietary data set described in Section 2. Parameter estimation is done as described in Section 3.1. We defer to Web Appendix D for diagnostics and checks, ensuring convergence of the MCMC algorithm.

### 5.1. Model estimation

Estimated parameters are shown in Table 8. We see that the posterior means of $\beta_0$, $\beta_1$, $\beta_2$ of model AST (Eq. (7)) are the closest to the GLM ones. The posterior of $\beta_0$ under the AST model is more skewed negatively than that of ASTtrend (Eq. (8)). In contrast, the posterior means and credible intervals of $\beta_1$ and $\beta_2$ under the two models

**Table 6**
In-sample predictive performance measured via AUC and QREC (Section 3.3) of CAR (Eq. (6)), AST (Eq. (7)), ASTtrend (Eq. (8)), GLM (Eq. (5)) and NetNaïve (Eq. (9)) in the three simulated scenarios described in Section 4. Ranges into parentheses indicate the values attainable by QREC considering the worst (riskiest) 3, 5, 10 and 50% of the population of synthetic firms, respectively.

| Scenario | Measure | CAR | AST | ASTtrend | GLM | NetNaïve |
|---|---|---|---|---|---|---|
| Space | AUC | **95.96** | 92.84 | 92.84 | 80.71 | 54.22 |
| | $QREC_{0.03}$ (0–5.94) | 5.91 | **5.92** | 5.92 | 5.66 | 3.55 |
| | $QREC_{0.05}$ (0–9.9) | 9.86 | **9.88** | 9.88 | 9.38 | 5.90 |
| | $QREC_{0.1}$ (0–19.8) | **19.72** | 19.68 | 19.68 | 18.27 | 11.59 |
| | $QREC_{0.5}$ (0–99.02) | **87.78** | 84.50 | 84.56 | 72.72 | 52.84 |
| SpaceTime | AUC | **95.25** | 93.03 | 93.03 | 80.93 | 54.34 |
| | $QREC_{0.03}$ (0–7.22) | 6.94 | **7.17** | 7.17 | 6.50 | 3.55 |
| | $QREC_{0.05}$ (0–12.04) | 11.57 | **11.91** | 11.91 | 10.79 | 5.90 |
| | $QREC_{0.1}$ (0–24.07) | 23.15 | 23.50 | **23.53** | 20.67 | 11.53 |
| | $QREC_{0.5}$ (0–100) | **91.30** | 88.01 | 88.06 | 76.01 | 54.11 |
| Trend | AUC | **96.69** | 92.96 | 92.96 | 81.09 | 54.10 |
| | $QREC_{0.03}$ (0–6.18) | 3.51 | 6.00 | **6.00** | 5.66 | 3.54 |
| | $QREC_{0.05}$ (0–10.3) | 5.85 | 9.98 | **9.98** | 9.33 | 5.69 |
| | $QREC_{0.1}$ (0–20.6) | 11.69 | 19.85 | **19.85** | 18.15 | 11.06 |
| | $QREC_{0.5}$ (0–100) | 58.33 | **81.94** | 81.88 | 71.67 | 53.12 |

**Table 7**
Out-of-sample performance measured via AUC and QREC (Section 3.3) of CAR (Eq. (6)), AST (Eq. (7)), ASTtrend (Eq. (8)), GLM (Eq. (5)) and NetNaïve (Eq. (9)) in the three simulated scenarios described in Section 4. Ranges into parentheses indicate the values attainable by QREC considering the worst (riskiest) 3, 5, 10 and 50% of the population of synthetic firms, respectively.

| Scenario | Measure | CAR | AST | ASTtrend | GLM | NetNaïve |
|---|---|---|---|---|---|---|
| Space | AUC | 85.17 | **89.75** | 89.75 | 82.35 | 53.44 |
| | $QREC_{0.03}$ (0–5.84) | 5.65 | **5.84** | 5.84 | 5.73 | 3.43 |
| | $QREC_{0.05}$ (0–9.74) | 9.47 | **9.62** | 9.62 | 9.35 | 5.74 |
| | $QREC_{0.1}$ (0–19.48) | 18.67 | 19.21 | **19.25** | 18.07 | 11.42 |
| | $QREC_{0.5}$ (0–97.39) | 75.98 | **80.75** | 80.67 | 73.57 | 52.04 |
| SpaceTime | AUC | 84.78 | **90.18** | 90.17 | 81.76 | 55.40 |
| | $QREC_{0.03}$ (0–7.53) | 7.25 | **7.47** | 7.47 | 7.21 | 3.61 |
| | $QREC_{0.05}$ (0–12.55) | 12.03 | **12.37** | 12.37 | 11.44 | 6.24 |
| | $QREC_{0.1}$ (0–25.09) | 23.13 | **24.22** | 24.22 | 21.23 | 12.19 |
| | $QREC_{0.5}$ (0–100) | 80.45 | **86.68** | 86.68 | 77.19 | 54.66 |
| Trend | AUC | 83.98 | **89.33** | 89.32 | 81.79 | 52.51 |
| | $QREC_{0.03}$ (0–4.69) | **4.69** | 4.69 | 4.69 | 4.69 | 3.02 |
| | $QREC_{0.05}$ (0–7.82) | 7.75 | **7.81** | 7.81 | 7.75 | 5.24 |
| | $QREC_{0.1}$ (0–15.62) | 15.47 | **15.59** | 15.59 | 15.19 | 10.61 |
| | $QREC_{0.5}$ (0–78.11) | 68.01 | 71.45 | **71.52** | 66.83 | 51.63 |

**Table 8**
Posterior mean and 95% credible intervals of the model parameters for CAR (Eq. (6)), AST (Eq. (7)), ASTtrend (Eq. (8)) and GLM (Eq. (5)) fitted to the real data with non-informative priors.

| Param | CAR mean (2.5%, 97.5%) | AST mean (2.5%, 97.5%) | ASTtrend mean (2.5%, 97.5%) | GLM mean(2.5%, 97.5%) |
|---|---|---|---|---|
| $\beta_0$ | 1.48 (0.88, 2.05) | 1.72 (1.30, 2.14) | 1.44 (0.88, 2.04) | 1.72 (1.49, 1.95) |
| $\beta_1$ | 0.89 (0.73, 1.04) | 0.90 (0.84, 0.96) | 0.90 (0.84, 0.96) | 0.88 (0.82, 0.94) |
| $\beta_2$ | 2.49 (1.65, 3.30) | 4.52 (4.20, 4.85) | 4.53 (4.21, 4.86) | 4.54 (4.21, 4.86) |
| $\beta_3$ | – | – | 0.07 (−0.05, 0.19) | – |
| $\alpha$ | 0.48 (0.02, 0.95) | 0.48 (0.03, 0.95) | 0.46 (0.02, 0.94) | – |
| $\rho$ | – | 0.29 (−0.81, 0.97) | 0.11 (−0.90, 0.96) | – |

are almost identical. For the CAR model, we highlight that the credible interval of $\beta_2$ does not overlap with the ones of the other models and does not contain the value estimated by the GLM for this parameter. We recall that $\beta_2$ is the coefficient for the covariate built on the proportion of used and granted credit across all Italian financial institutions. The significant difference in its posterior distribution can be read as if the CAR model gives lower importance to the possible default of the firm towards other banks than the GLM and the other models. Furthermore, this difference may be because the CAR model is estimated on a single timestamp while the other models leverage the full dataset for the estimation. The posterior distribution of the spatial autocorrelation parameter $\alpha$ is coherent across all three spatio (-temporal) models (Figure 6 in Web Appendix D). Concerning the temporal autocorrelation for models AST and ASTtrend, Table 8 shows wide credible intervals around mean values 0.29 and 0.11, respectively. However, by looking at the proportion of positive posterior samples over the full

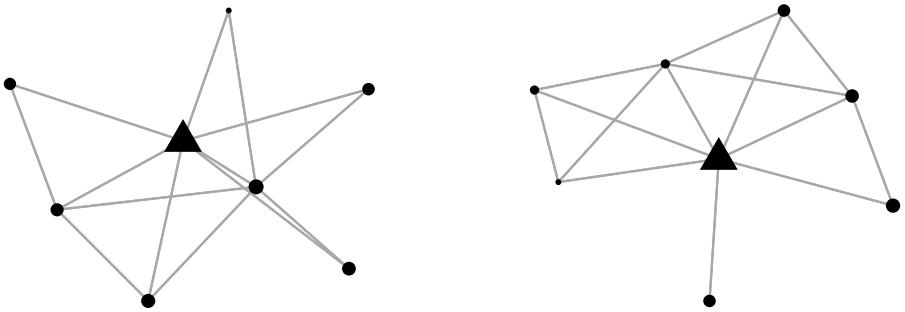


**Fig. 3.** Two ego-networks of order 1 around two truly defaulted nodes (triangles) at the 8th timestamp with the following characteristics: strictly positive mean spatial random effect $\phi_k$ from the CAR model and GLM in-sample predicted probability of default below 0.1. Circular nodes indicate non-defaulted firms. Bigger nodes represent larger (positive) posterior mean of the spatial random effect $\phi_k$.

**Table 9**
In-sample predictive performance measured via AUC and QREC (Section 3.3) of CAR (Eq. (6)), AST (Eq. (7)), ASTtrend (Eq. (8)), GLM (Eq. (5)) and NetNaïve (Eq. (9)) on real data. Ranges into parentheses indicate the values attainable by QREC considering the worst (riskiest) 3, 5, 10 and 50% of the population of firms, respectively, as described in Section 3.3.

| Measure | CAR | AST | ASTtrend | GLM | NetNaïve |
|---|---|---|---|---|---|
| AUC | 79.95 | **83.86** | 83.83 | 80.81 | 51.27 |
| $QREC_{0.03}$ (0–65.2) | 27.12 | 41.45 | **41.71** | 36.22 | 2.31 |
| $QREC_{0.05}$ (0–100) | 27.12 | **52.92** | 52.68 | 42.15 | 4.56 |
| $QREC_{0.1}$ (0–100) | 44.07 | **59.50** | 59.38 | 56.46 | 12.10 |
| $QREC_{0.5}$ (0–100) | 85.31 | **88.87** | 88.72 | 85.89 | 50.10 |

posterior distribution for $\rho$, we can assess that the temporal autocorrelation parameter is estimated to be positive with a probability above 0.58 for both models.

Fig. 3 shows two randomly selected, truly defaulted firms at the 8th timestamp (triangular nodes). The (true) network built around the triangular nodes shows no defaulted firms in their respective first-order neighbours (circular nodes indicate no default). For these truly defaulted firms (triangles), the GLM estimates a low probability of default (below 0.1). Instead, the CAR model estimates a positive mean spatial random effect $\phi_k$, thereby inflating these firms' estimated posterior probability of default. Therefore, the CAR successfully pins the truly defaulted firms despite the absence of defaulted first-order neighbours. A similar result is obtained via the spatio-temporal models.

Finally, we look at the in-sample performance presented in Table 9. The proposed models have higher AUC and QRECs than the GLM (and NetNaïve), except for the CAR model.

### 5.2. Out-of-sample forecasting performance

This section investigates the forecasting performance on two out-of-sample sets: April 2019 is forecasted based on model estimation over February 2018–January 2019, and August 2019 is forecasted based on model estimation over February 2018–April 2019. The temporal gaps between the training and test sets are defined according to the delay in data availability. They are required to compute both the target variable and the linear covariates. In Web Appendix E, we repeat the analysis presented here, considering different time horizons for prediction. Specifically, we target the prediction of default six, nine, and twelve months ahead, holding-out August 2019 data to investigate how the predictive ability of the models behaves as a function of the temporal gap in prediction. We defer to Web Appendix E for these results.

Models' performance on the two three-month ahead out-of-sample tests is presented in Table 10. We see that the predictive performance of the CAR model is now higher than the GLM's in both out-of-sample tests, and its performance is in line (April 2019) or better (August 2019) than the two spatio-temporal models. The improvement is mainly in terms of AUC, but we also see that the CAR model can distinguish among firms whose linear covariates do not explicitly show a critical situation ($QREC_{0.5}$). From an economic point of view, the spatial model can predict defaults in the absence of overdrafts or delays in payments towards the bank or other financial institutions. This is a powerful feature in the application context as it enables the bank to proactively act before the distress happens instead of reacting to a critical event. The same conclusion can be drawn by looking at the ROC in Fig. 4, where we see that the ROC of the CAR outperforms the GLM in the central and right range of the graphs, that is, where the GLM is predicting a low probability of default.

Looking at the spatio-temporal models, we see that the forecasting performance is good in the first out-of-sample test predicting April 2019 data (Table 10), where the AUC is slightly higher than that of both CAR and GLM. In the second case (August 2019 data), we see a drop in their predictive performance as the AUC is smaller than the GLM's. This may be because the marginal distribution

**Table 10**
Out-of-sample forecasting performance measured via AUC and QREC (Section 3.3) of CAR (Eq. (6)), AST (Eq. (7)), ASTtrend (Eq. (8)), GLM (Eq. (5)) and NetNaïve (Eq. (9)) on the held-out real data of January 2019 and May 2019, respectively. Ranges into parentheses indicate the values attainable by QREC considering the worst (riskiest) 3, 5, 10 and 50% of the population of firms, respectively, as described in Section 3.3.

| | *January 2019* | | | | |
|---|---|---|---|---|---|
| Measure | CAR | AST | ASTtrend | GLM | NetNaïve |
| AUC | 68.58 | 68.58 | **68.61** | 67.92 | 49.04 |
| $QREC_{0.03}$ (0–60.47) | 28.68 | 35.66 | 37.21 | **37.98** | 3.88 |
| $QREC_{0.05}$ (0–100) | 39.53 | 41.86 | 41.86 | **41.86** | 4.65 |
| $QREC_{0.1}$ (0–100) | 47.29 | 48.06 | 48.06 | **48.06** | 9.30 |
| $QREC_{0.5}$ (0–100) | **59.69** | 58.91 | 58.91 | 58.91 | 49.61 |
| | *May 2019* | | | | |
| AUC | **74.26** | 69.91 | 69.95 | 70.74 | 50.88 |
| $QREC_{0.03}$ (0–67.24) | 7.76 | **8.62** | 8.62 | 7.76 | 2.59 |
| $QREC_{0.05}$ (0–100) | 12.93 | 21.55 | 21.55 | **22.41** | 6.90 |
| $QREC_{0.1}$ (0–100) | 44.83 | 43.10 | 42.24 | **44.83** | 10.34 |
| $QREC_{0.5}$ (0–100) | **79.31** | 70.69 | 70.69 | 70.69 | 50.86 |

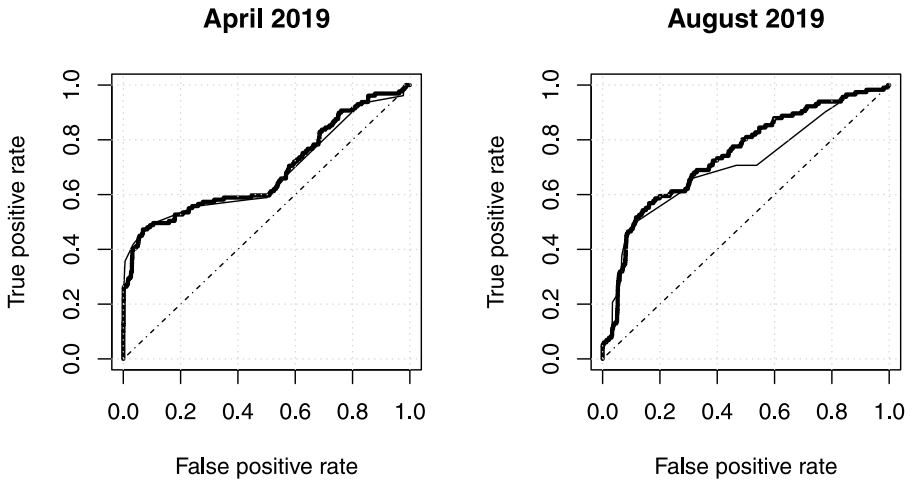



**Fig. 4.** The ROC curves of the CAR model (thick) and the GLM (thin) on the two out-of-sample sets, respectively, April and August 2019. The 45 degree line (dashed) corresponds to random choice.

of the probability of default per month does not increase or decrease consistently over time for all the network, as required from the separable form of our models. Indeed, in Eq. (7) and (8) the spatial and temporal random effects are combined additively, $\phi_k + \chi_t$. Suppose we visualise the spatial effects as a surface over the network. In that case, we should see the surface moving up and down without the ability to increase in certain regions while decreasing in different regions of the network. Thus, the difference between the probability of default of two firms at two different timestamps can only change in the presence of a change in their covariates and not to a change in the local risk of the network (since the $\phi_k$'s are constant over time and the $\chi_t$'s are not firm-dependent). Rather, performance results suggest that, for real data, the surface of the general probability of default may increase or decrease over time differently in different parts of the network. To this end, we have further considered two modelling extensions: (1) an interaction between the spatial effects, $\phi_k$, and temporal effects, $\chi_t$; and (2) non-separable spatio-temporal effects $\phi_{kt}$ modelled via a multivariate AR(1) process with *iid* CAR updates. Preliminary results are presented in Web Appendix G.

## 6. Conclusions

In this work, we have approached the task of short-term forecasting a firm's default by leveraging the information of firms' interconnections to detect potential contagion effects. The most interesting component in the dataset of our case study is the network of commercial relationships among firms, which is not often available in credit risk studies. We modelled the data using Bayesian spatial and spatio-temporal techniques, adapting them to our loan management scenario. We proposed a CAR model on the adjacency matrix built on

the links of the commercial network and further considered two extensions, including additive temporal effects. After a simulation study on several scenarios, we fit the models to the real data and observed an improvement in forecasting performance brought by these approaches in characterising non-trivial distress situations. Indeed, from the ROC curves, we saw that the general improvement in AUC is concentrated on those firms for which the most explicit linear covariates (i.e., payment delays, access to huge amounts of credit) do not give any information. Thus, the baseline GLM cannot highlight any difficulty or distinguish sound firms from distressed ones. Such a feature is interesting from a credit risk perspective because it enables coordinated credit strategies on the firms of a segment of the supply chain detected through the spatial (inter-firm) effects before the distress propagates and before it is converted into traditional explicit distress information.

The choice of a hierarchical setting with a CAR prior for the spatial effects is one of the possible choices to consider for network-linked data (see, e.g., a different precision matrix as in Datta et al., 2019 or penalisation, as introduced in Li et al., 2019). These model definitions may be considered competing specifications to the proposed models in future work. Moreover, by testing the algorithms out-of-sample over eight months, we found that the dynamics of the complex network of commercial interactions among firms are not completely described by the temporal and spatial effects expressed in a separable additive form. This latter finding encourages the exploration of non-separable spatio-temporal effects. At present, the topic of non-separable spatio-temporal modelling constitutes an unexplored field for the complex network of thousands of nodes due to computational issues. It will be explored further in future research.

Finally, we acknowledge that the dataset may contain firms of any market sector with different commercial dynamics, with different speeds in acquiring and losing customers. Thus, removing the assumption of a static adjacency matrix may also be a direction for future research. From a methodological point of view, this is an open field of study (Billé et al., 2019).

## Declaration of competing interest

The research was conducted during the industrial PhD of the first author jointly at Intesa Sanpaolo and Università degli Studi di Torino. The first author is currently employed at Intesa Sanpaolo; however, this industrial affiliation does not alter the author's adherence to IJF policies. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The work presented in this manuscript represents the views of the authors and not necessarily that of Intesa Sanpaolo.

## Availability of data and materials

The data for the case study are part of a propriety dataset owned by Intesa Sanpaolo and cannot be made publicly available. Simulated data and the code implementing the methodologies presented in this paper are available at https://github.com/claudiaber/BayesianSpatialSpatiotemporal.

## Appendix A. Supplementary materials

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2022.05.003.

## References


Adegboye, O. A., & Kotze, D. (2012). Disease mapping of leishmaniasis outbreak in Afghanistan: spatial hierarchical Bayesian analysis. *Asian Pacific Journal of Tropical Disease*, *2*(4), 253–259.

Alegana, V. A., Atkinson, P. M., Wright, J. A., Kamwi, R., Uusiku, P., Katokele, S., Snow, R. W., & Noor, A. M. (2013). Estimation of malaria incidence in northern namibia in 2009 using Bayesian conditional-autoregressive spatial–temporal models. *Spatial and Spatio-Temporal Epidemiology*, *7*.

Banerjee, S., Carlin, B., & Gelfand, A. (2003). Hierarchical modeling and analysis for spatial data. In *Chapman & Hall/CRC Monographs on Statistics & Applied Probability*, CRC Press.

Battiston, S., Gatti, D. D., Gallegati, M., Greenwald, B., & Stiglitz, J. E. (2007). Credit chains and bankruptcy propagation in production networks. *Journal of Economic Dynamics & Control*, *31*(6), 2061–2084.

Beck, N., Gleditsch, K. S., & Beardsley, K. (2006). Space is more than geography: using spatial econometrics in the study of political economy. *International Studies Quarterly*, *50*(1), 27–44.

Billé, A. G., Blasques, F., & Catania, L. (2019). Dynamic spatial autoregressive models with time-varying spatial weighting matrices. Available at SSRN 3241470.

Blasques, F., Koopman, S. J., Lucas, A., & Schaumburg, J. (2016). Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics*, *195*(2), 211–223.

Bowman, F. D., Caffo, B., Bassett, S. S., & Kilts, C. (2008). A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage*, *39*(1), 146–156.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control* (5th ed.). Wiley.

Bussoli, C., & Marino, F. (2018). Trade credit in times of crisis: evidence from European SMEs. *Journal of Small Business and Enterprise Development*, *25*(2), 277–293.

Catania, L., & Billé, A. G. (2017). Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics*, *32*(6), 1178–1196.

Chen, K., Zhu, K., Meng, Y., Yadav, A., & Khan, A. (2020). Mixed credit scoring model of logistic regression and evidence weight in the background of big data. In A. Abraham, A. K. Cherukuri, P. Melin, & N. Gandhi (Eds.), *Intelligent systems design and applications* (pp. 435–443). Cham: Springer International Publishing.

Datta, A., Banerjee, S., Hodges, J. S., & Gao, L. (2019). Spatial disease mapping using directed acyclic graph auto-regressive (DAGAR) models. *Bayesian Analysis*, *14*(4), 1221–1244.

Dolfin, M., Knopoff, D., Limosani, M., & Xibilia, M. G. (2019). Credit risk contagion and systemic risk on networks. *Mathematics*, *7*, 713.

Duan, J.-C., Sun, J., & Wang, T. (2012). Multiperiod corporate default prediction—A forward intensity approach. *Journal of Econometrics*, *170*(1), 191–209.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: reasoning about a highly connected world*. Cambridge University Press.

Fernandez, V. (2011). Spatial linkages in international financial markets. *Quantitative Finance*, *11*(2), 237–245.

Ge, T., Müller-Lenke, N., Bendfeldt, K., Nichols, T. E., & Johnson, T. D. (2014). Analysis of multiple sclerosis lesions via spatially varying coefficients. *The Annals of Applied Statistics*, *8*(2), 1095–1118.

Gelfand, A. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, *4*(1), 11–15.

Gelfand, A., Diggle, P., Fuentes, M., & Guttorp, P. (2010). *Handbook of Spatial Statistics* (1st ed.). CRC Press.

Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *160*(3), 523–541.

Hoff, P. D. (2009). *A first course in bayesian statistical methods* (1st ed.). Springer Publishing Company, Incorporated.

Hurd, T. R. (2016). *Contagion! systemic risk in financial networks*. Springer.

Hyndman, R. J., & Athanasopoulos, G. (2019). *Forecasting: principles and practice* (3rd ed.). OText, https://otexts.com/fpp3/.

Jin, X., Carlin, B. P., & Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics*, *61*(4), 950–961.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In *KDD '03*, *Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 137–146). New York, NY, USA: Association for Computing Machinery.

Kinderman, R., & Snell, S. (1980). *Markov random fields and their applications*. American Mathematical Society.

Lamieri, M., & Sangalli, I. (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.

Lee, D. (2013). CARBayes: An r package for Bayesian spatial modeling with conditional autoregressive priors. *Journal of Statistical Software, Articles*, [ISSN: 1548-7660] *55*(13), 1–24.

Li, T., Levina, E., & Zhu, J. (2019). Prediction models for network-linked data. *The Annals of Applied Statistics*, *13*(1), 132–164.

McGuinness, G., Hogan, T., & Powell, R. (2018). European trade credit use and SME survival. *Journal of Corporate Finance*, *49*, 81–103.

Neal, R. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo* (pp. 116–162). CRC Press.

Newman, M. (2010). *Networks: an introduction*. Oxford University Press.

Orth, W. (2012). The predictive accuracy of credit ratings: measurement and statistical inference. *International Journal of Forecasting*, *28*(1), 288–296, Special Section 1: The Predictability of Financial Markets Special Section 2: Credit Risk Modelling and Forecasting.

Pace, R. K., & LeSage, J. (2010). Spatial econometrics. In A. Gelfand, P. Diggle, M. Fuentes, & P. Guttorp (Eds.), *Handbook of spatial statistics* (pp. 245–260). CRC Press.

Raymaekers, J., Verbeke, W., & Verdonck, T. (2021). Weight-of-evidence 2.0 with shrinkage and spline-binning. URL https://arxiv.org/abs/2101.01494.

Roukny, T., Bersini, H., Pirotte, H., Caldarelli, G., & Battiston, S. (2013). Default cascades in complex networks: Topology and systemic risk. *Scientific Reports*, *3*, 2759.

Soloshenko, O. M. (2015). Generalizations of logistic regression, weight of evidence, and the gini index for a continuous target variable taking on probabilistic values. *Cybernetics and Systems*, *51*(6), 992–1004.

Stan Development Team (2018). Stan modeling language users guide and reference manual. URL http://mc-stan.org.

Sun, D., Tsutakawa, R. K., & Speckman, P. L. (1999). Posterior distribution of hierarchical models using CAR(1) distributions. *Biometrika*, *86*(2), 341–350.

Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, *16*(2), 149–172.

Watson, S. C., Liu, Y., Lund, R. B., Gettings, J. R., Nordone, S. K., McMahan, C. S., & Yabsley, M. J. (2017). A Bayesian spatio-temporal model for forecasting the prevalence of antibodies to borrelia burgdorferi, causative agent of lyme disease, in domestic dogs within the contiguous United States. *PLoS One*, *12*(5), 1–22.

Woolrich, M. W., Jenkinson, M., Brady, J. M., & Smith, S. M. (2004). Fully Bayesian spatio-temporal modeling of fMRI data. *IEEE Transactions on Medical Imaging*, *23*(2), 213–231.

Zeng, G. (2014). A necessary condition for a good binning algorithm in credit scoring. *Applied Mathematical Sciences*, *8*(65), 3229–3242.