Spatial regression models to improve P2P credit risk management

Arianna Agosto, Paolo Giudici and Thomas Leach

University of Pavia

CNECT, Brussels, 3 July 2019

Motivation

- Differently from banks, P2P platforms have limited access to the "classical" historical data on which to base credit risk assessment.
- P2P platforms generate "alternative" data, which consist of the direct (disintermediated) transactions between their customer borrowers and/or lenders. This data can be used to assess credit risk, and investigate contagion effects.
- Transactional data can also be used by banks, especially when evaluating new customers, or to assess contagion effects in networks of borrowers.

Background

- Financial network models (see, eg., Billio et al., 2012, Diebold and Yilmaz, 2014, Hautsch et al., 2015, Giudici and Spelta 2016) allow to study contagion effects but are often merely descriptive.
- Spatial econometric models can incorporate dependence among observations that are in a kind of proximity, not necessarily geographical.
- When applied to credit risk, spatial econometric models can be used both as a credit scoring model, to estimate the default risk of a given company, and as a contagion model (Calabrese et al., 2017).

Contribution

In this paper:

- We simulate transactions between the borrowers of a P2P lending platform using trade flows data between corporate sectors;
- Based on the simulated transactions, we apply a binary spatial model to measure credit risk of SMEs taking contagion effects into account.

The model - I

Let Y a vector of binary dependent variables (default or not) and Y^* a vector of continuous underlying latent variables.
We consider the observation mechanism as

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

with $i = 1, 2, \ldots, n$

while the latent (default risk) mechanism has a spatial autoregressive structure:

$$Y^* = \rho W Y^* + X\beta + \epsilon$$

with

- X a $n \times k$ matrix of explanatory variables
- W the spatial lag weight matrix with ρ the associated coefficient
- ϵ the error term.

The model - II

The model implies heteroskedasticity in the errors:

$$Y^* = (I - \rho W)^{-1} X\beta + e$$

where

$$e = (I - \rho W)^{-1} \epsilon$$

thus

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

Given the connectivity matrix W, the higher the ρ parameter

- the higher the spatial dependence between observations
- the higher the shocks covariance.

It can thus be interpreted as a *contagion* parameter.

Building the connectivity matrix - I

The World Input Output Trade (WIOT) statistics provide information on the aggregate trade volumes of 52 economic sectors in each country with all sectors in all countries.

For a given country, let A the sector of company i, B the sector of company j and f_{AB} the trade flow from sector A to sector B.

Replacing the individual flows with the aggregate (sectorial) ones, we obtain the approximate trade matrix F with entries:

$$f_{ij} = f_{AB} = \sum_{l \in A} \sum_{m \in B} f_{lm}$$

Building the connectivity matrix - II

To proxy the individual companies' flows, we first calculate the ratio between company i turnover $\tilde{x}_i$ and the sector A total turnover:

$$x_i = \frac{\tilde{x}_i}{\sum_{l \in A} \tilde{x}_l}$$

Then we calculate the ratio between company j turnover $\tilde{y}_i$ and the sector B total turnover:

$$y_j = \frac{\tilde{y}_j}{\sum_{m \in B} \tilde{y}_m}$$

The product $x_i y_j$ is a proxy of the proportion of flows from company i to company j on the total flows from sector A to sector B.

Building the connectivity matrix - III

Repeating this calculation for all companies and then computing the entrywise product with the trade matrix F, we get the connectivity matrix:

$$W = \begin{pmatrix} x_1 y_1 F_{1,1} & x_1 y_2 F_{1,2} & \cdots & x_1 y_n F_{1,n} \\ x_2 y_1 F_{2,1} & x_2 y_2 F_{2,2} & \cdots & x_2 y_n F_{2,n} \\ \vdots & \ddots & \ldots & \vdots \\ x_n y_1 F_{n,1} & x_n y_2 F_{n,2} & \cdots & x_n y_n F_{n,n} \end{pmatrix}$$

The ij element can be interpreted as the proxy of the trade flow from company i to company j.
Conversely, the ji element can be interpreted as the proxy of the trade flow from company j to company i.

Application - Data

- Data collected from modeFinance, a European Credit Assessment Institution which supplies credit scoring to P2P platforms specialized in business lending.
- The complete dataset includes $\approx 15,000$ Italian SMEs, for which it is provided:
  - a set of financial ratios relative to accounting year 2015
  - information about the status of the company (0=Active, 1=Default) in 2016.
- From the available financial ratios, we select:
  - the return on equity ratio;
  - the activity ratio, expressed as the ratio between sales and total assets;
  - the solvency ratio, calculated as the ratio between the net income and the total debt.

Results - Estimates

	$n = 800$		$n = 1200$		$n = 2000$		$n = 3000$	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
ρ	0.50	(2.13)	0.78	(3.39)	0.62	(2.33)	0.57	(1.36)
Constant	-0.56	(-0.85)	0.44	(0.96)	-0.17	(-0.25)	-0.53	(-0.55)
ROE	-0.77	(-4.18)	-0.53	(-3.53)	-0.58	(-4.91)	-0.74	(-5.90)
Activity Ratio	-0.12	(-0.65)	0.05	(0.38)	-0.03	(-0.34)	-0.11	(-1.47)
Solvency Ratio	-0.02	(-1.89)	-0.03	(-3.00)	-0.02	(-2.29)	-0.01	(-2.67)

Table: Results of binary SAR model estimation.

Results - Accuracy I

	$n = 800$		$n = 1200$		$n = 2000$		$n = 3000$	
	Logit	Spatial logit	Logit	Spatial logit	Logit	Spatial logit	Logit	Spatial logit
AUC	0.792	0.795	0.798	0.806	0.802	0.806	0.787	0.789

Table: AUC values for the estimated logit and spatial logit models.

Results - Accuracy II

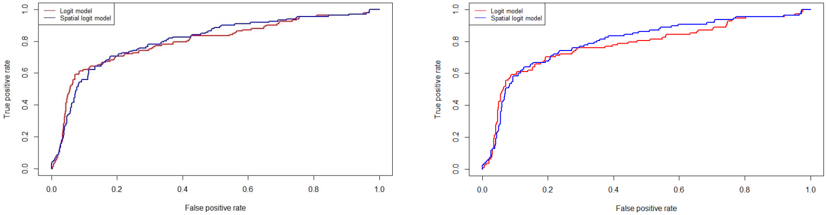

Figure: ROC curves of baseline and spatial logit models ($n = 1200$), in-sample (left) and out-of-sample (right).

Summary and conclusions

- Simulating direct transactions between companies, as if they were in a P2P platform, we have estimated a credit scoring model that includes a systemic component, based on transactions, and an idiosyncratic component, based on financial ratios.
- Our empirical findings show that the contagion parameter ρ is significant and $> 0.5\%$, for different sample sizes.
- The improvement in model accuracy, even with respect to a well performing baseline specification, is positive, and depends on the sample size.
- Direct transactional data between companies could be useful to measure contagion effects and to support credit risk assessment, for both P2P platforms and traditional banks.

Further research

- Sparse covariance structure: it is simplistic to assume that each company has trading relationships with all the others.
- Proximity measure: the distance may be calculated in terms of net flows.
- Dynamic spatial dependence: the interconnections are likely changing with time: need for different model specifications (Poisson Autoregressive models, Tree network models).