Spatial regression models for SMEs

Raffaella Calabrese University of Essex

joint work with Galina Andreeva and Jake Ansell

Credit Scoring and Credit Control XIV conference, University of Edinburgh

26-28 August 2015

Outline

Outline

Outline

Outline

Credit contagion for SMEs

UK 99.6% Businesses are Small and Medium Enterprises (SMEs), 54.3% of total employment, 49.5% GDP

Since lending to SMEs is riskier than lending to large corporations (Altman and Sabato, 2006), Basel II established that banks should develop credit risk models that are specific to SMEs.

To improve the default forecasting accuracy, we investigate if risk factors leading to SME defaults could depend on the risk factors of other nearby SMEs.

⇓

We suggest to introduce **spatial interdependence** in a scoring model for SMEs.

Credit contagion for SMEs

UK 99.6% Businesses are Small and Medium Enterprises (SMEs), 54.3% of total employment, 49.5% GDP

Since lending to SMEs is riskier than lending to large corporations (Altman and Sabato, 2006), Basel II established that banks should develop credit risk models that are specific to SMEs.

To improve the default forecasting accuracy, we investigate if risk factors leading to SME defaults could depend on the risk factors of other nearby SMEs.

⇓

We suggest to introduce **spatial interdependence** in a scoring model for SMEs.

Credit contagion for SMEs

UK 99.6% Businesses are Small and Medium Enterprises (SMEs), 54.3% of total employment, 49.5% GDP

Since lending to SMEs is riskier than lending to large corporations (Altman and Sabato, 2006), Basel II established that banks should develop credit risk models that are specific to SMEs.

To improve the default forecasting accuracy, we investigate if risk factors leading to SME defaults could depend on the risk factors of other nearby SMEs.

$\Downarrow$

We suggest to introduce **spatial interdependence** in a scoring model for SMEs.

Credit contagion for SMEs

UK 99.6% Businesses are Small and Medium Enterprises (SMEs), 54.3% of total employment, 49.5% GDP

Since lending to SMEs is riskier than lending to large corporations (Altman and Sabato, 2006), Basel II established that banks should develop credit risk models that are specific to SMEs.

To improve the default forecasting accuracy, we investigate if risk factors leading to SME defaults could depend on the risk factors of other nearby SMEs.

$\Downarrow$

We suggest to introduce **spatial interdependence** in a scoring model for SMEs.

Binary spatial autoregressive model

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

$$\begin{aligned} \boldsymbol{Y}^* &= \rho \boldsymbol{W} \boldsymbol{Y}^* + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{\varepsilon} \end{aligned}$$

ε is a multivariate normal distribution in a probit model or a multivariate logistic distribution in a logit model

$$w_{ij} = \begin{cases} 1 & \text{if the } i\text{-th and } j\text{-th observations are contiguous;} \\ 0 & \text{if } i = j \text{ or the } i\text{-th and } j\text{-th observations are not continguous} \end{cases}$$

$\boldsymbol{W}$ is exogenously given.

Binary spatial autoregressive model

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

$$\begin{aligned} \boldsymbol{Y}^* &= \rho \boldsymbol{W} \boldsymbol{Y}^* + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{\varepsilon} \end{aligned}$$

ε is a multivariate normal distribution in a probit model or a multivariate logistic distribution in a logit model

$$w_{ij} = \begin{cases} 1 & \text{if the } i\text{-th and } j\text{-th observations are contiguous;} \\ 0 & \text{if } i = j \text{ or the } i\text{-th and } j\text{-th observations are not continguous} \end{cases}$$

$\boldsymbol{W}$ is exogenously given.

Binary spatial autoregressive model

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

$$\begin{aligned} \boldsymbol{Y}^* &= \rho \boldsymbol{W} \boldsymbol{Y}^* + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{\varepsilon} \end{aligned}$$

ε is a multivariate normal distribution in a probit model or a multivariate logistic distribution in a logit model

$$w_{ij} = \begin{cases} 1 & \text{if the } i\text{-th and } j\text{-th observations are contiguous;} \\ 0 & \text{if } i = j \text{ or the } i\text{-th and } j\text{-th observations are not contiguous} \end{cases}$$

$\boldsymbol{W}$ is exogenously given.

Five estimators

- Expectation-Maximization algorithm (McMillen, 1995)
- Gibbs sampler (LeSage, 2000)
- Recursive Importance Sampling (Beron and Vijverberg, 2004)
- Generalised Method of Moments (Pinkse and Slade, 1998)
- Linearized Generalised Method of Moments (Klier and McMillen, 2008)

Five estimators

- Expectation-Maximization algorithm (McMillen, 1995)
- Gibbs sampler (LeSage, 2000)
- Recursive Importance Sampling (Beron and Vijverberg, 2004)
- Generalised Method of Moments (Pinkse and Slade, 1998)
- Linearized Generalised Method of Moments (Klier and McMillen, 2008)

Five estimators

- Expectation-Maximization algorithm (McMillen, 1995)
- Gibbs sampler (LeSage, 2000)
- Recursive Importance Sampling (Beron and Vijverberg, 2004)
- Generalised Method of Moments (Pinkse and Slade, 1998)
- Linearized Generalised Method of Moments (Klier and McMillen, 2008)

Five estimators

- Expectation-Maximization algorithm (McMillen, 1995)
- Gibbs sampler (LeSage, 2000)
- Recursive Importance Sampling (Beron and Vijverberg, 2004)
- Generalised Method of Moments (Pinkse and Slade, 1998)
- Linearized Generalised Method of Moments (Klier and McMillen, 2008)

Five estimators

- Expectation-Maximization algorithm (McMillen, 1995)
- Gibbs sampler (LeSage, 2000)
- Recursive Importance Sampling (Beron and Vijverberg, 2004)
- Generalised Method of Moments (Pinkse and Slade, 1998)
- Linearized Generalised Method of Moments (Klier and McMillen, 2008)

Klier & McMillen (2008)

By linearizing the Generalised Method of Moments estimator (Pinkse and Slade, 1998), the estimates of ρ and $\boldsymbol{\beta}$ are extrapolated from the convenient starting point at $\rho = 0$.

At this point, the derivative of the objective function

$$\tilde{\boldsymbol{e}}'(\rho, \boldsymbol{\beta})\boldsymbol{ZMZ}'\tilde{\boldsymbol{e}}(\rho, \boldsymbol{\beta})$$

with respect to β and ρ, using $\mathbf{M} = (\mathbf{Z}'\mathbf{Z})^{-1}$, significantly simplifies.

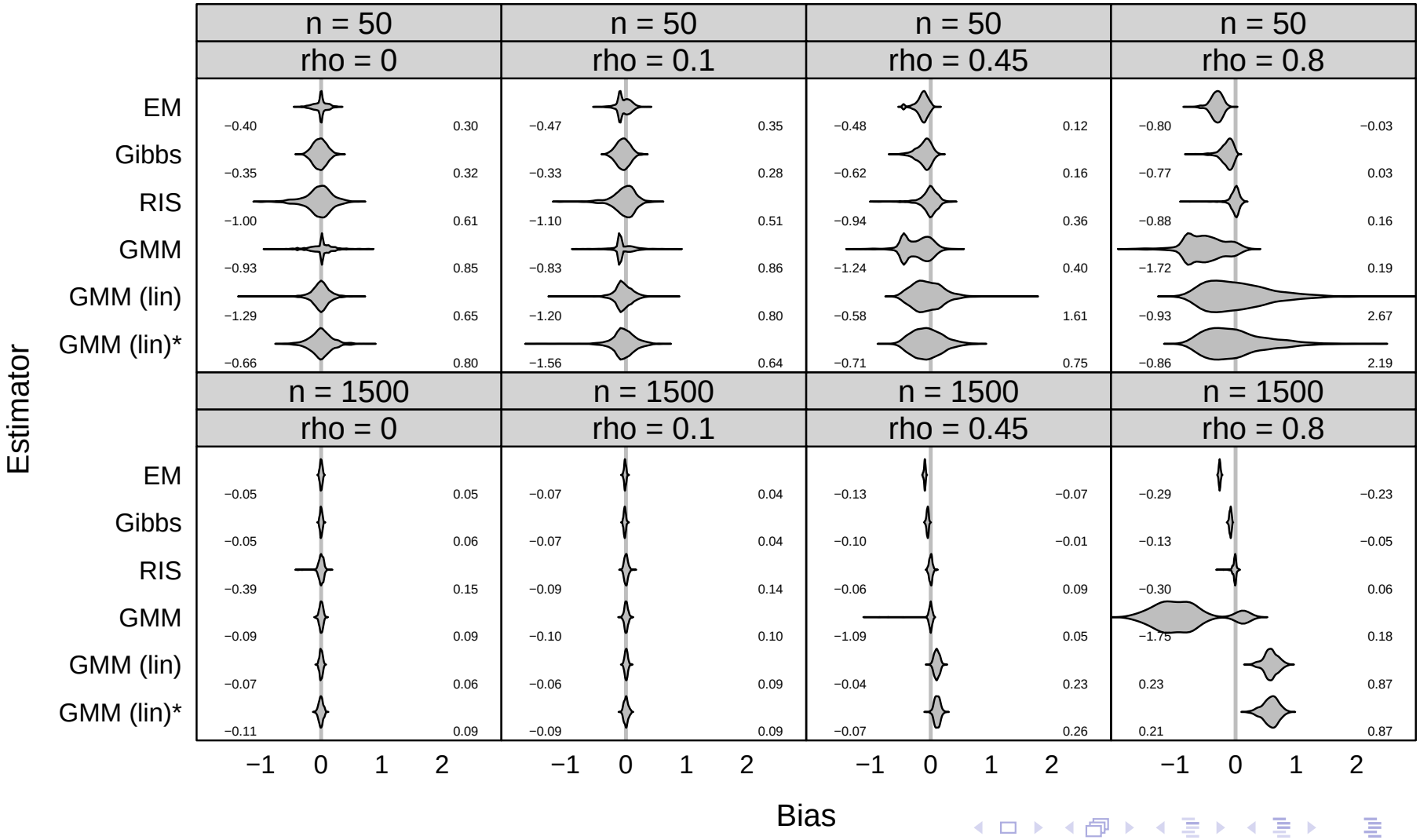
Klier & McMillen (2008)

By linearizing the Generalised Method of Moments estimator (Pinkse and Slade, 1998), the estimates of ρ and $\boldsymbol{\beta}$ are extrapolated from the convenient starting point at $\rho = 0$.
At this point, the derivative of the objective function

$$\tilde{\boldsymbol{e}}'(\rho, \boldsymbol{\beta})\boldsymbol{ZMZ}'\tilde{\boldsymbol{e}}(\rho, \boldsymbol{\beta})$$

with respect to $\boldsymbol{\beta}$ and ρ, using $\mathbf{M} = (\mathbf{Z}'\mathbf{Z})^{-1}$, significantly simplifies.

Monte Carlo simulations (Calabrese and Elkink, 2014)



Data

- Over 2 million enterprises
- Recorded April 2007
- *first two letters of the postcode*
- Risk factors:
  - General Information (legal form, region, SIC, Employees, Age of Company);
  - Directors Information (Directors, Ownership, Changes etc);
  - Previous Credit history (DBT, judgements etc);
  - Accounting Information (Common financial variables and financial ratios).

Data

- Over 2 million enterprises
- Recorded April 2007
- *first two letters of the postcode*
- Risk factors:
  - General Information (legal form, region, SIC, Employees, Age of Company);
  - Directors Information (Directors, Ownership, Changes etc);
  - Previous Credit history (DBT, judgements etc);
  - Accounting Information (Common financial variables and financial ratios).

Data

- Over 2 million enterprises
- Recorded April 2007
- *first two letters of the postcode*
- Risk factors:
  - General Information (legal form, region, SIC, Employees, Age of Company);
  - Directors Information (Directors, Ownership, Changes etc);
  - Previous Credit history (DBT, judgements etc);
  - Accounting Information (Common financial variables and financial ratios).

Results for 27,648 start-up SMEs in London without spatial interdependence

Variables	*Estimate*	*Std. Error*	*z value*	*p-value*
Intercept	1.1775	0.01207	97.549	$<$ 2e-16 ***
Legal Form	0.92541	0.06477	14.287	$<$ 2e-16 ***
Age of Company	0.72024	0.03513	20.504	$<$ 2e-16 ***
Current/Previous Directors	0.36143	0.04847	7.456	8.91e-14 ***
PP Worst DBT	0.34627	0.07495	4.620	3.84e-06 ***
Number of Previous Searches	0.56500	0.05911	9.558	$<$ 2e-16 ***
Time since last derogatory	0.40753	0.01323	30.806	$<$ 2e-16 ***
Unsatisfied mortgages	0.65688	0.09683	6.784	1.17e-11***
Lateness Of Accounts	0.31760	0.02764	11.493	$<$ 2e-16 ***
Years Accounts Available	-0.31217	0.09759	-3.199	0.00138 **
Current Liabilities	0.68012	0.02010	33.832	$<$ 2e-16 ***
Time Since Last Annual Return	0.59170	0.09248	6.398	$<$ 2e-16 ***

Results for 27,648 start-up SMEs in London with spatial interdependence

Variables	*Estimate*	*Std. Error*	*z value*	*p-value*
Intercept	-1.09123	0.39130	-2.78876	0.00529
Legal Form	0.95431	0.09566	9.97554	$<$ 2e-16 ***
Age of Company	0.72209	0.02753	26.22549	$<$ 2e-16 ***
Current/Previous Directors	0.36544	0.02979	12.26801	$<$ 2e-16 ***
PP Worst DBT	0.34813	0.05776	6.02666	$<$ 2e-16 ***
Number of Previous Searches	0.58166	0.04174	13.93651	$<$ 2e-16 ***
Time since last derogatory	0.40785	0.01069	38.16718	$<$ 2e-16 ***
Unsatisfied mortgages	0.64675	0.06167	10.48711	1.17e-11***
Lateness Of Accounts	0.31435	0.02269	13.85280	$<$ 2e-16 ***
Years Accounts Available	-0.30532	0.06784	-4.50065	0.00001
Time Since Last Annual Return	0.68038	0.01507	45.15916	0.00000
Current Liabilities	0.59889	0.06207	9.64864	$<$ 2e-16 ***
W	0.85025	0.26041	3.265044	0.00054

Missclassification rates for start-up SMEs in London

scoring model without spatial component	0.3225
scoring model with spatial component	0.2913

Conclusions

- Spatial interdependence has an impact on the parameter estimates of a scoring model for SMEs.
- Adoption of interdependent scoring models can aid in the prediction of default.
- An extension of the model which enables to introduce the interdependence between economic sectors of SMEs is a possible direction for further research.

Bibliography

- Beron, Kurt J. and Wim P.M. Vijverberg. 2004. Probit in a Spatial Context: A Monte Carlo Analysis, in Luc Anselin, Raymond J.G.M. Florax, and Sergio J. Rey (eds.), Advances in Spatial Econometrics. Methodology, Tools and Applications. Berlin: Springer, pp. 169195.
- Calabrese, Raffaella and Elkink, Jos. 2014. "Estimators of Binary SPatial Autoregressive Models: A Monte Carlo Study," Journal of Regional Science, 54 (4), 664-687.
- 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.
- LeSage, James P. 2000. Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models, Geographical Analysis, 32(1), 1935.
- Pinkse, Joris and Margaret E. Slade. 1998. Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models, Journal of Econometrics, 85, 125154.