A gente trabalha para você crescer.

Insper | Instituto de Ensino e Pesquisa

Inspirar para Transformar

The spatial correlation of credit risk and its gain in credit scoring models

Guilherme Barreto Fernandes, MA

Prof Rinaldo Artes, PhD

Credit Cycle

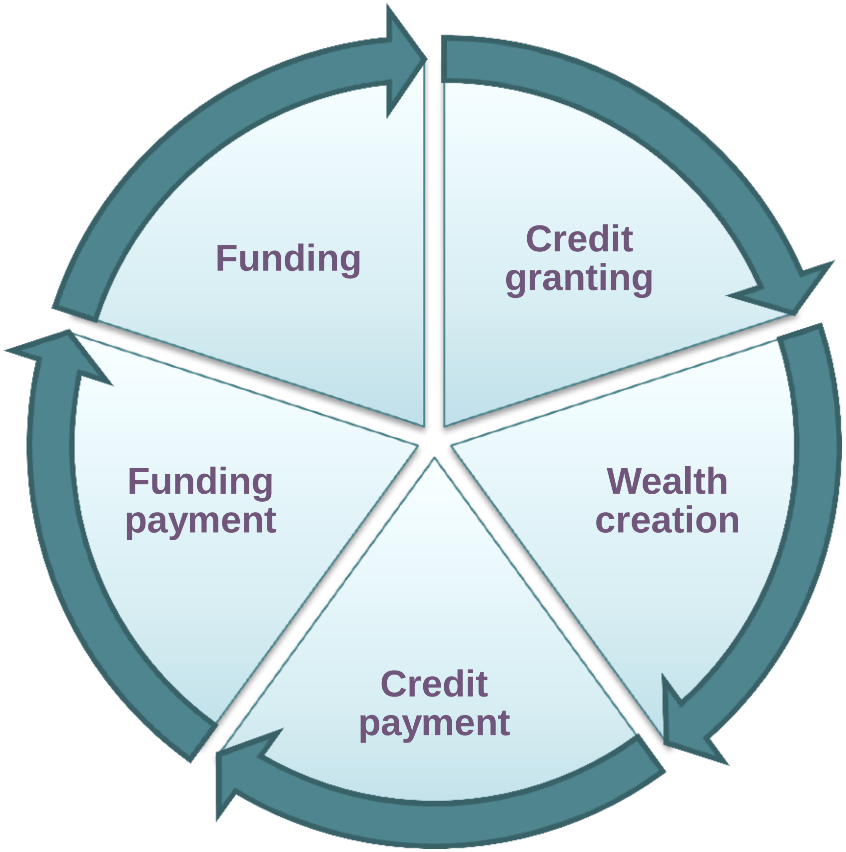
Ciclo virtuoso do crédito

- ✓ The credit cycle
- ✓ Credit Risk: probability of default

Related to many risk drivers (SME):

- ✓ Indebtness
- ✓ Activity sector
- ✓ Payment behaviour
- ✓ Delinquency history
- ✓ Economic information
- ✓ Local information

Related somehow in SME risk assessment

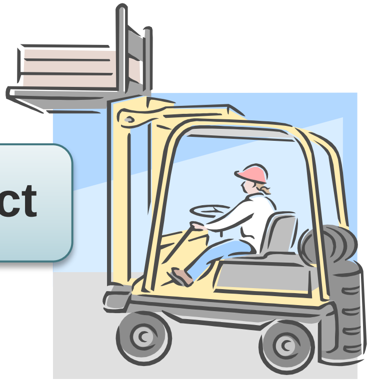


Local economic influence in SME

- ✓ Economic information
- ✓ Local information

Two illustrative explanations

Retailers and other stores



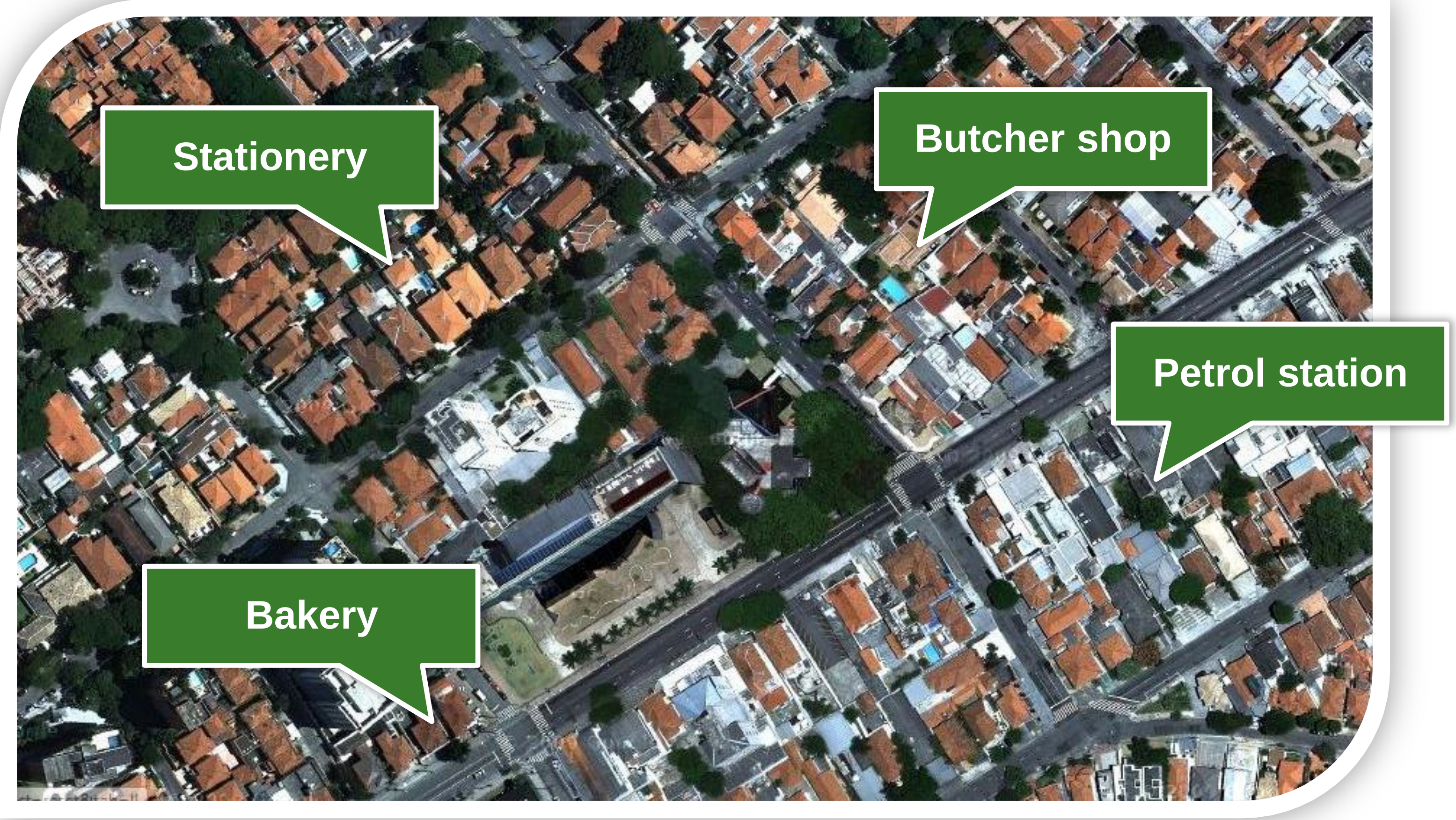
Industrial district

Local economic influence in SME

Retailers and other stores

Local economic influence in SME

Retailers and other stores



Local economic influence in SME

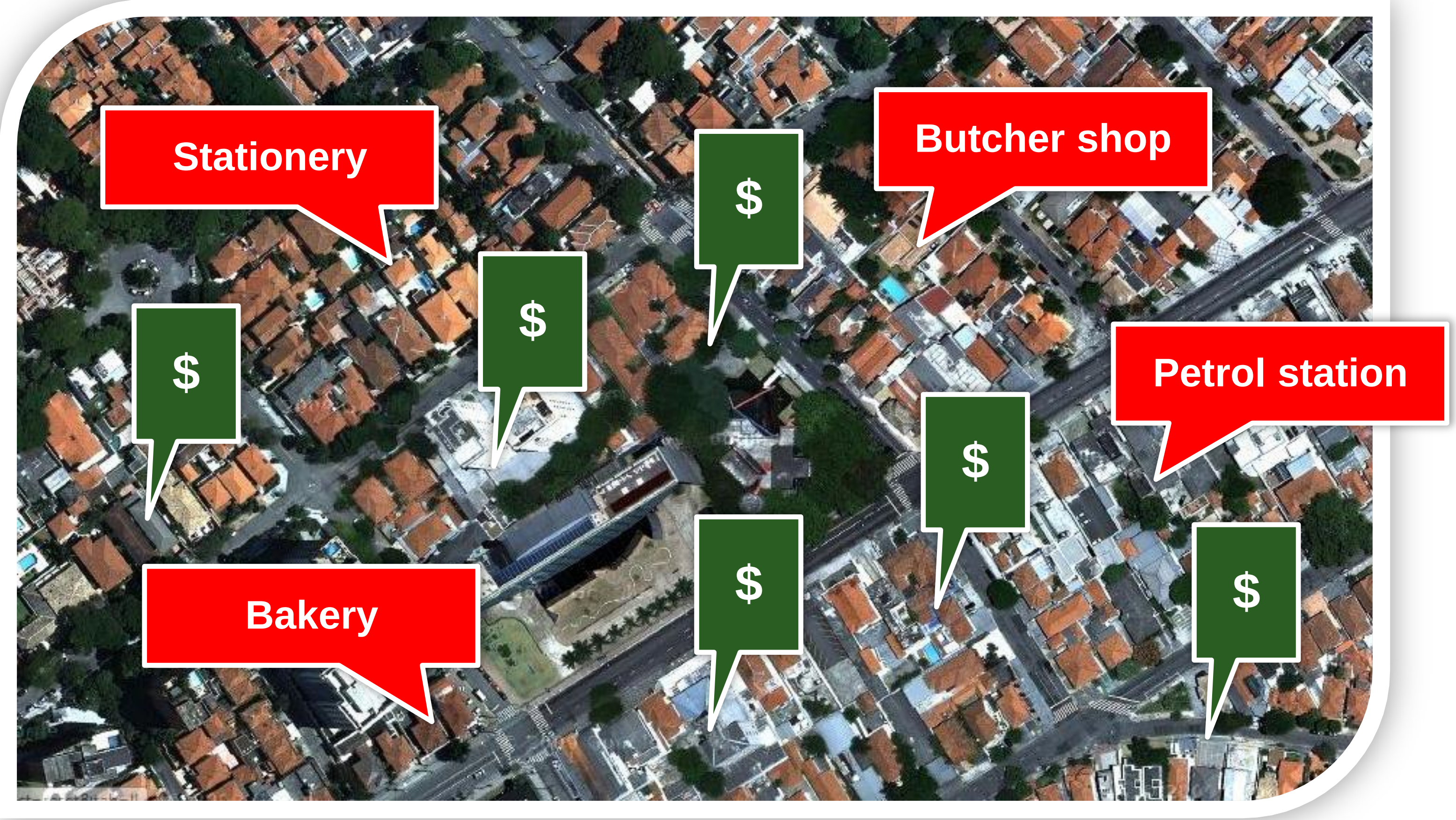
Retailers and other stores

Local economic influence in SME

Retailers and other stores

Local economic influence in SME

Retailers and other stores

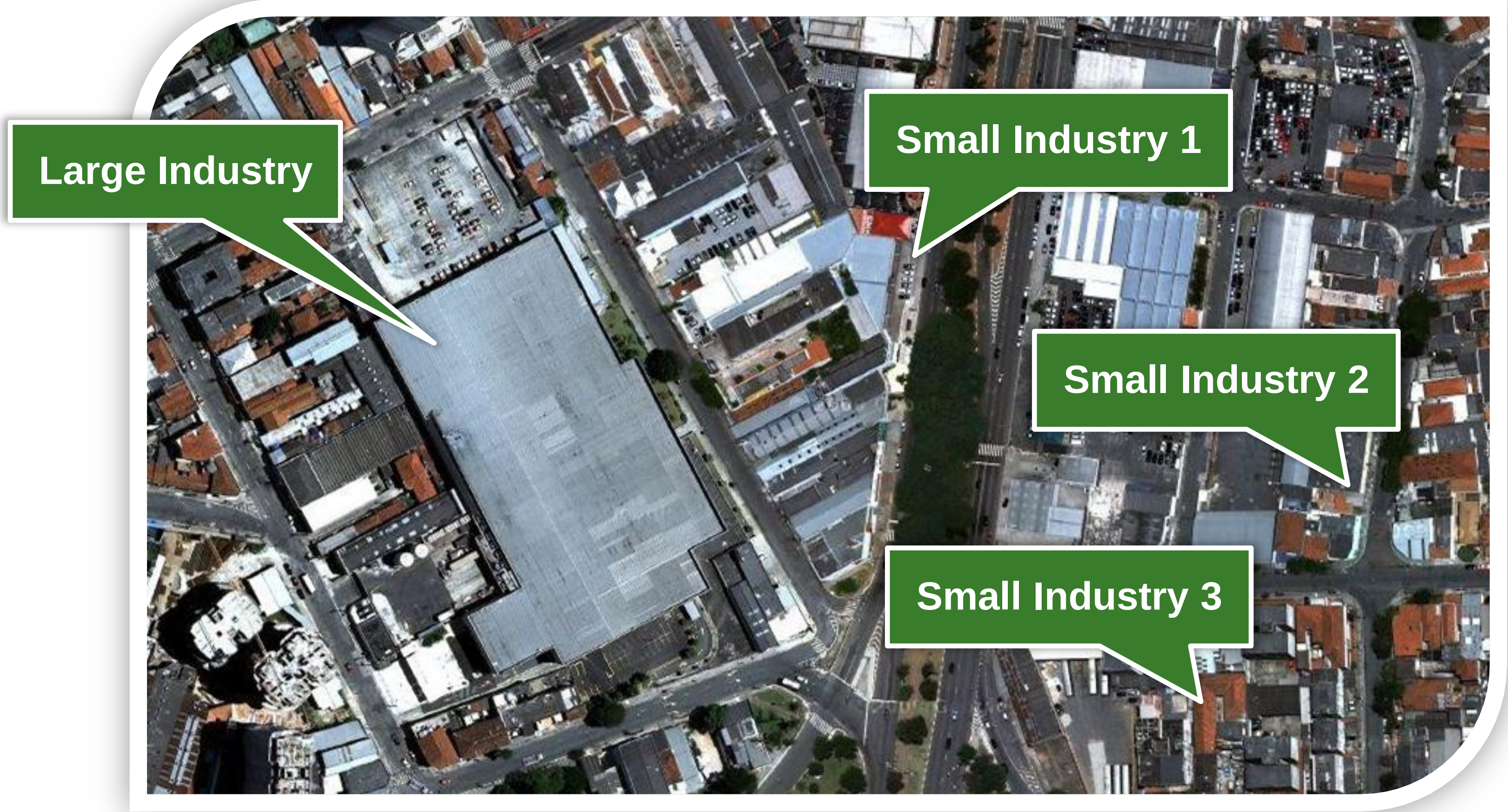


Local economic influence in SME

Industrial district

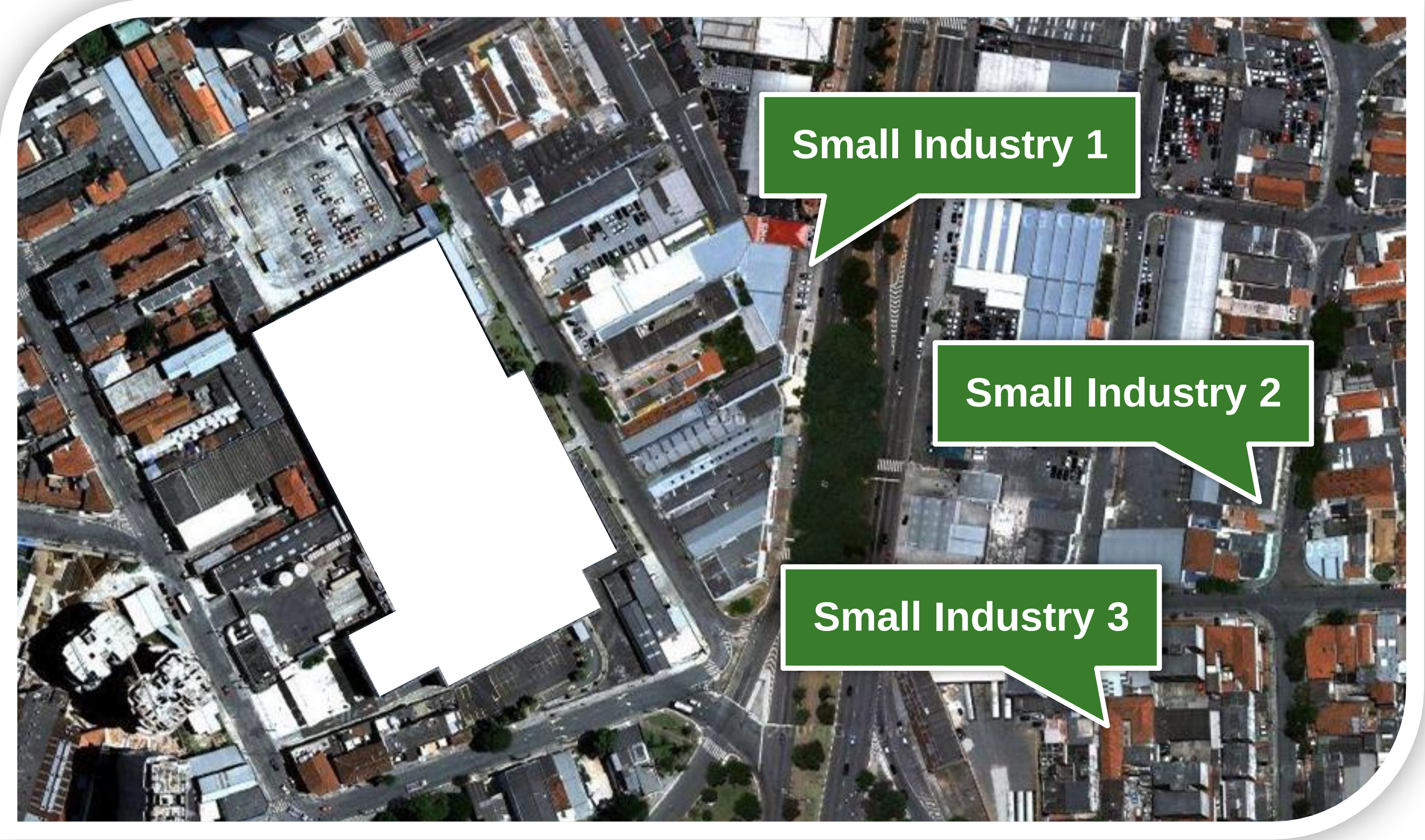
Local economic influence in SME

Industrial district



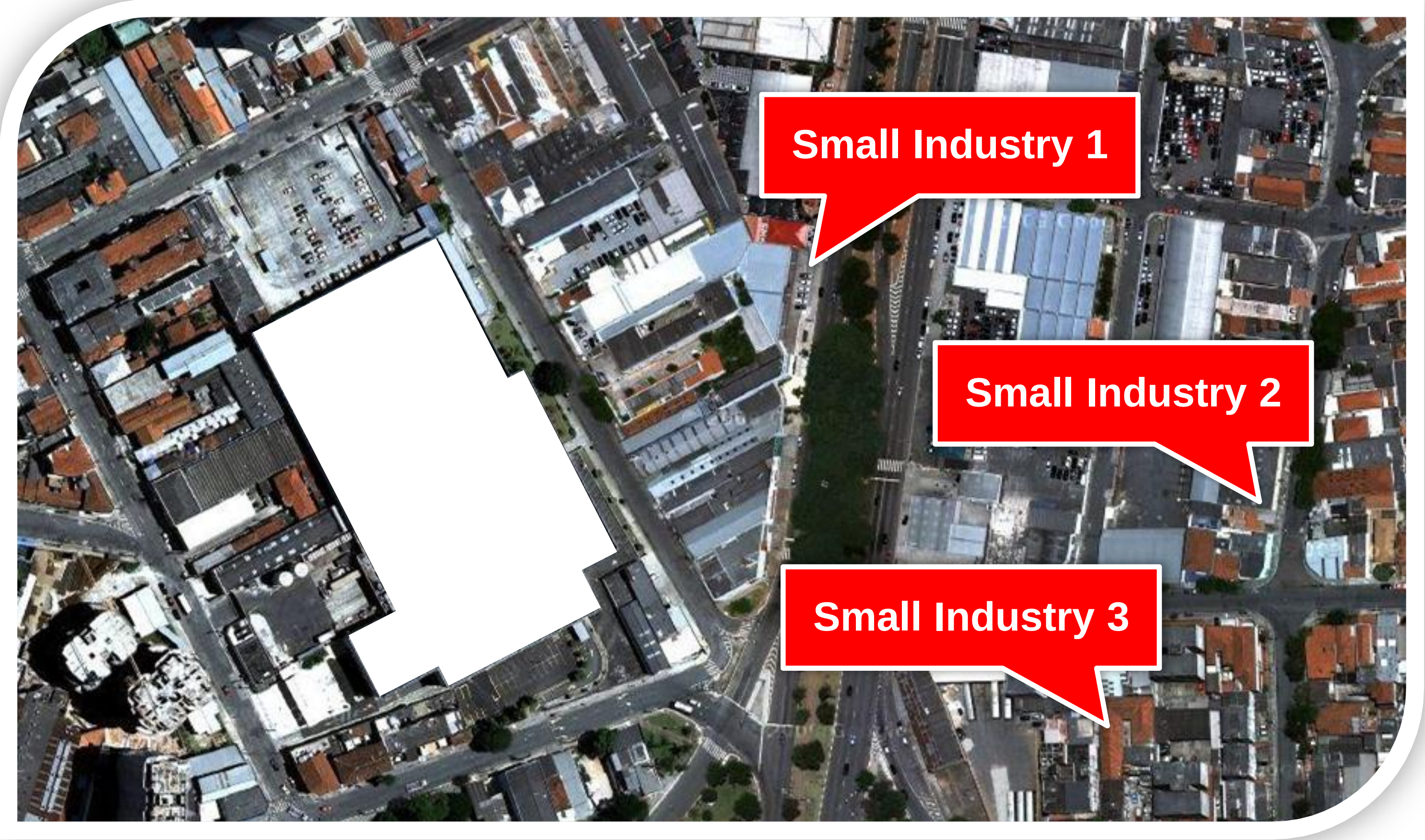
Local economic influence in SME

Industrial district



Local economic influence in SME

Industrial district



Local Economic Factor: Spatial dependence

- ✓ Pontentially relevant spatial dependence
- ✓ How to include this information into credit scoring models?

Local economic indexes

Information gathering difficulties (Gerkman, 2011):

Eg. "Neighbourhood" GDP

Post code grouping

+
- ✓ Easy creation and simple implementation
- ✓ Any "Excel-like" software is able to evol such analysis

−
- ✓ Potentially a large number of categories
- ✓ Regions with few SME or defaulters may result in poor risk assessment
- ✓ May result in unstable model or overfitting issue

Local Economic Factor: Spatial dependence

Insper Instituto de Ensino e Pesquisa

Inspirar para Transformar

- ✓ Pontentially relevant spatial dependence
- ✓ How to include this information into credit scoring models?
  1. Local economic indexes
  2. Post code grouping
  3. Out proposal:
     - Spatial dependence effect is a continous measure

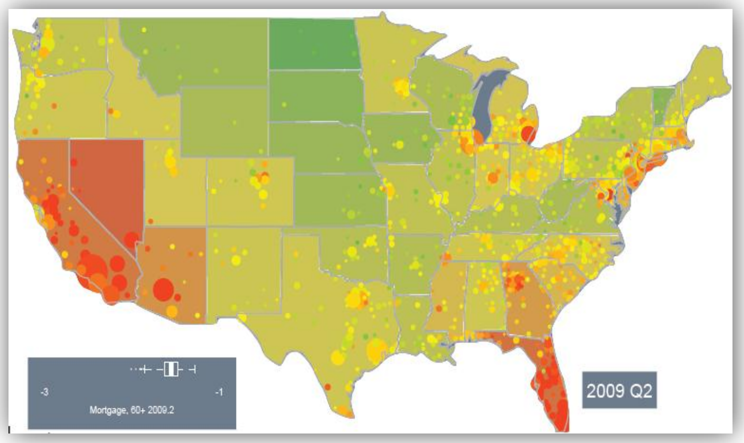
Estimated by Kriging methods

Spatial dependence

- ✓ Probability of default may be conditioned to many risk factors:
  - Indebtness
  - Reference file
  - Payment behaviour
  - Negative statements info
  - Location

→ Evidence that location matters (Stine, 2011)

- Counties level
- Moran's I



Past work

- ✓ Argawal *et al*. (2012) proposed:
  - Neighbourhood information: % of low income people, ethnic mix
  - Not significant in a full model
- ✓ Barro and Barro (2010):
  - Contagion model
  - Combines location (communa) and industry sector
  - Output: a counterparty PD model
- ✓ Fernandes (2012):
  - Correlation between firms is conditioned to distance
  - Kriging method estimates SPATIAL RISK
  - Explanatory variable in the credit scoring model

- ✓ **What is Kriging?**

Kriging Algorithm

Insper Instituto de Ensino e Pesquisa
Inspirar para Transformar

Kriging was first used in geology (soil characteritics), epidemiology (risk areas of certain diseases) and agriculture (nutrients concentration).

- Based on the geologist Daniel Krige’s ideas (Krige, 1951)
- Developed by the mathematician Georges Matheron (Matheron, 1963)

- ✓ Kriging: interpolation method based on distance (Matheron, 1963)
- ✓ Prediction method via smoothing weighted averages
- ✓ Ordinary Kriging:

z_i: *observed variable in observation* ***i***

$d_{ij} = h$: *distance between obs* ***i*** *and* ***j***

$\hat{z}_i = \sum_{j \neq i} \lambda_j{}^i z_j$: *average of surroundings of* ***i***

$$\lambda_j{}^i = f(d_{ij})$$

$$\sum_{j \neq i} \lambda_j{}^i = 1$$

Variogram analysis

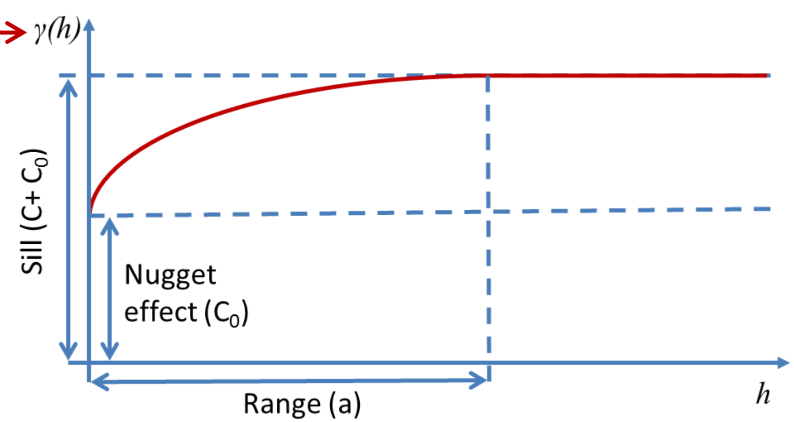
How does $\lambda_j{}^i = f(d_{ij})$ changes with d_{ij}

Insper Instituto de Ensino e Pesquisa

Inspirar para Transformar

Semivariance and semivariogram

$$\gamma(h) = \frac{1}{2}E\left[\left(Z_i - Z_j\right)^2 | d(i,j) = h\right]$$



Theorical variogram model:

Spherical Model of Matheron:

$$\gamma(h) = \begin{cases} c_o + C\left[1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right], & 0 \leq h \leq a \\ c_o + C, & h > a \end{cases}$$

Exponential Model of Formery:

$$\gamma(h) = c_o + C\left[1 - exp\left(-\frac{h}{a}\right)\right]$$

Gaussian Model:

$$\gamma(h) = c_o + C\left[1 - exp\left(-\frac{h^2}{a^2}\right)\right]$$

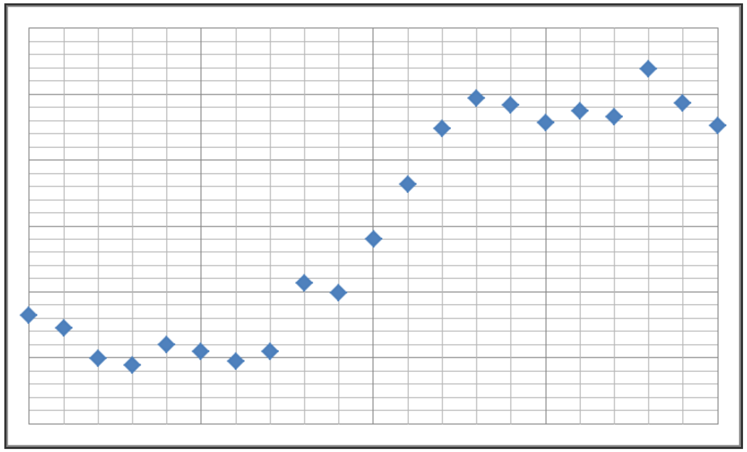
Extension (Hohn, 1989): $\boldsymbol{\gamma_{SUM}(h) = c_0 + \gamma_1(h) + \gamma_2(h),}$

Variogram analysis

Empirical variogram and forecast

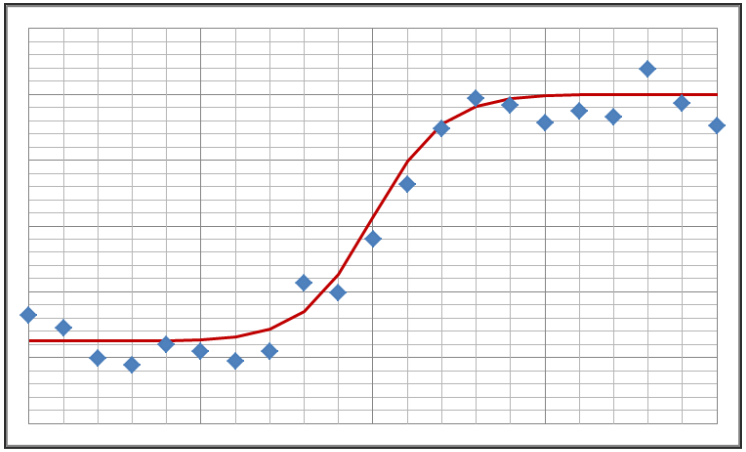
$$\gamma(h) = \frac{1}{2}E\left[\left(Z_i - Z_j\right)^2 | d(i,j) = h\right] \Rightarrow \hat{\gamma}(h) = \underbrace{\frac{1}{2N_h}\sum_{(i,j)|d_{i,j}=h}(z_i - z_j)^2}$$

$\hat{\gamma}(h)$



h

$\tilde{\gamma}(h)$



h

Theorical model estimation

- ✓ How does the semivariogram and semivariance $(\tilde{\gamma}(h))$ connects with $\hat{z}_i = \sum_{j \neq i} \lambda_j{}^i z_j$?

$$\min \sigma_\varepsilon^2(\mathrm{i}) = \min \mathrm{Var}\left[\hat{Z}_\mathrm{i} - Z_\mathrm{i}\right], \text{ subject to } \textstyle\sum_{\mathrm{i}=1}^{\mathrm{n}} \lambda_\mathrm{i} = 1.$$

$$\begin{pmatrix} \hat{\lambda}_1 \\ \vdots \\ \hat{\lambda}_n \\ \hat{\mu} \end{pmatrix} = \begin{pmatrix} \tilde{\gamma}(d_{11}) & \cdots & \tilde{\gamma}(d_{1n}) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{\gamma}(d_{n1}) & \cdots & \tilde{\gamma}(d_{nn}) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\gamma}(d_{10}) \\ \vdots \\ \tilde{\gamma}(d_{n0}) \\ 1 \end{pmatrix}$$

Goovaerts (1997)

- ✓ New explanatory variable: $\hat{z}_i$ = local average risk for firm ***i*** (SPATIALRISK)

Naïve logistic regression

$$Y_i \sim Bernoulli(p_i)$$

$$p_i = \left(\frac{exp(x'_i\beta)}{1 + exp(x'_i\beta)}\right)$$

Where:

$i = observation$

$y_i = target\ variable\ (1{:}\ default; 0{:}\ non - default)$

$p_i = probability\ of\ default$

$x_i = explanatory\ variables$

$\beta_{naive} = parameter\ vector$

✓ Estimation via MLE

Measurement error logistic model

$$Y_i \sim Bernoulli(p_i)$$

$$p_i = \left(\frac{exp(x'_i\beta + z_i\theta)}{1 + exp(x'_i\beta + z_i\theta)}\right)$$

Where:

$i = observation$

$y_i = target\ variable\ (1{:}\ default; 0{:}\ non - default)$

$p_i = probability\ of\ default$

$x_i = variables\ without\ measurement\ error$

$z_i = variable\ with\ measurement\ error$

$\beta = paramters\ vector$

$\theta = parameter\ of\ the\ spatial\ risk\ variable$

✓ Estimation via SIMEX (Cook and Stefanski, 1994)

Empiral analysis

Steps for analysis

Semivariance and semivariogram estimation

Bureau data

Estimation of spatial risk

Bureau data + specific portfolio

Credit scoring estimation

Specific portfolio

Empiral analysis

Data used in semivariance estimation

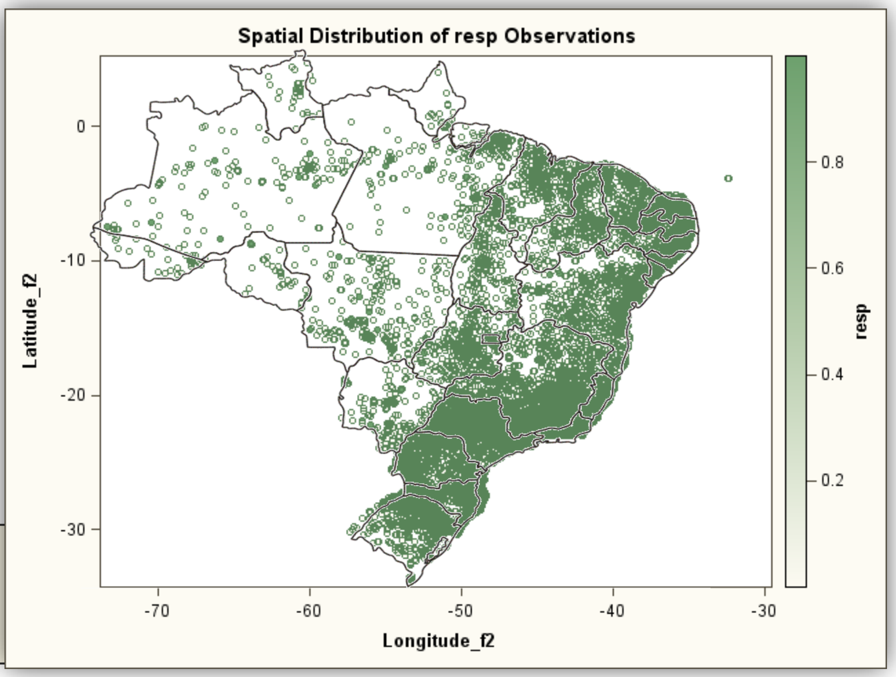
Bureau

9MM SMEs

Information gathered:

- ✓ Market default: 90 days in arrears
- ✓ Latitude and longitude
- ✓ Post code

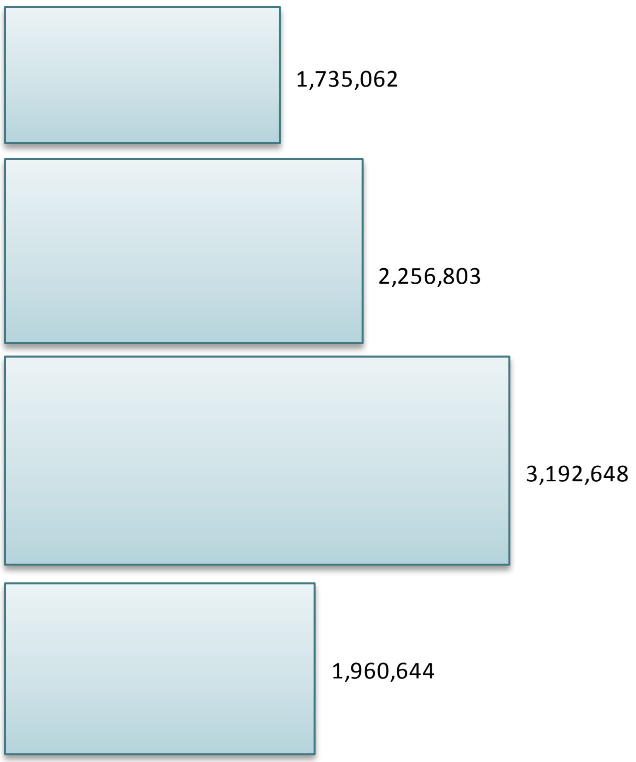
Spatial dependence pattern may
not be equal in every region



Empiral analysis

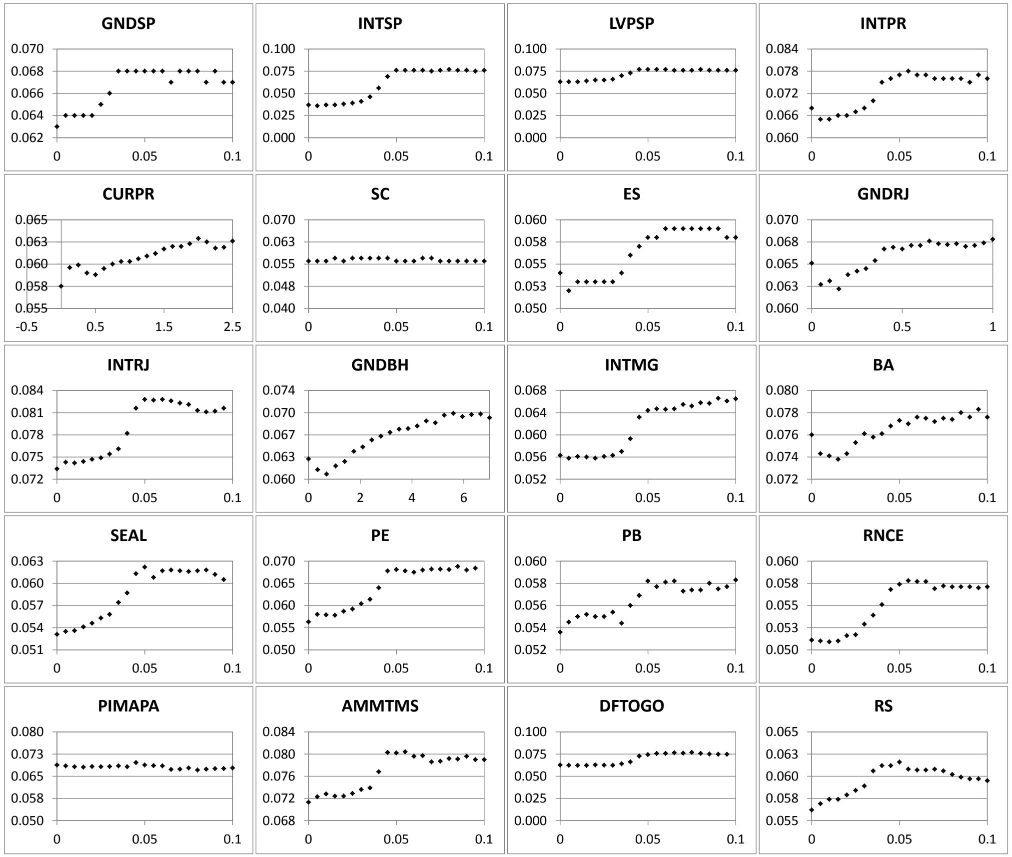
20 regions

Code	SMEs	Description	SMEs
GNDSP	1,269,923	Metropolitan areas	1,960,644
GNDRJ	357,259		
GNDBH	257,975		
CURPR	75,487		
LPVP	284,304	States except the metropolitan areas	3,192,648
INTSP	1,248,828		
INTRJ	365,702		
INTMG	693,458		
INTPR	600,356		
ES	168,582	Entire states	2,256,803
BA	482,030		
PE	264,559		
PB	96,796		
SC	424,232		
RS	820,604		
SEAL	136,853	Group of states	1,735,062
RNCE	376,021		
PIMAPA	393,982		
AMMTMS	448,578		
DFGOTO	379,628		



Empiral analysis

Semivariogram

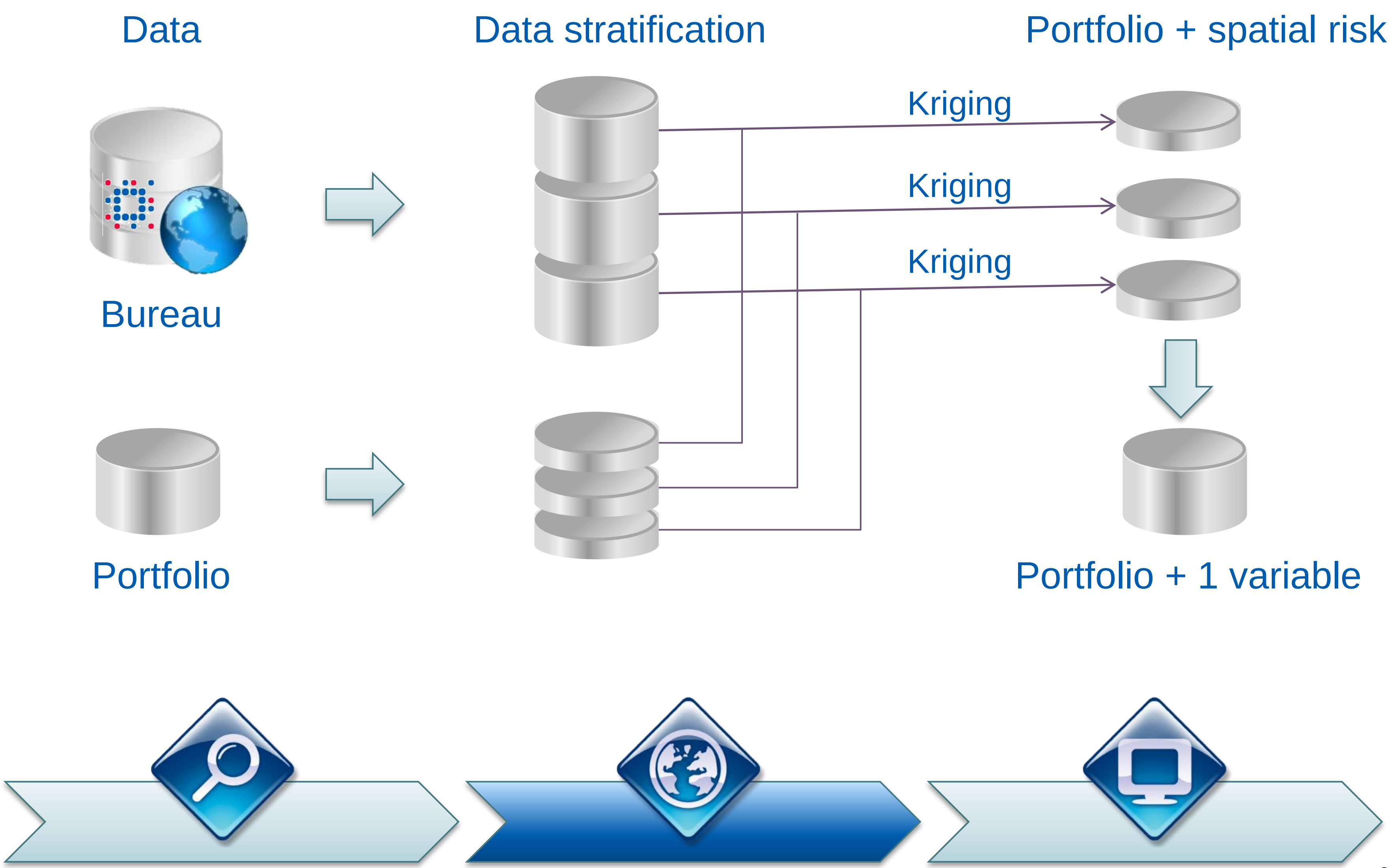


Region	Total number of regions	Strong spatial dependence
Metropolitan areas	4	4
States except the metropolitan areas	5	5
Entire states	6	5
Group of states	5	4

Region	Semivariogram model				
	Gaussian	Exponential + Spheric	Spheric	Gaussian + Gaussian	Gaussian + Spheric
Metropolitan areas	2			1	1
States except the metropolitan areas	5				
Entire states	5		1		
Group of states	5	1			

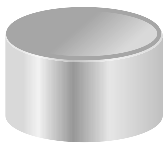
Empiral analysis

Spatial risk component



Empiral analysis

Credit scoring



Portfolio

- ✓ 8.800 firms
- ✓ 12 months performance window
- ✓ Application scoring model

Variables	Model 1	Model 2	Model 3
	Naive LR	Naive LR	SIMEX LR
2 reference file	V	V	V
6 credit demand	V	V	V
4 default history	V	V	V
3 negative statement	V	V	V
2 past credit use	V	V	V
1 payment capacity	V	V	V
1 payment method	V	V	V
1 Serasa bureau score	V	V	V
Spatial risk component	X	V	V

Empiral analysis

Credit scoring

Variable	Parameter estimates			Altman's scaled vector		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Intercepto	0.5978*	-0.2563*	-0.2563*			
IND_RESTR	-0.1007**	-0.1238*	-0.1237*	4%	4%	4%
RESTR2	-0.3801*	-0.4104*	-0.4103*	11%	9%	9%
RESTR3	-0.3020*	-0.3573*	-0.3573*	19%	16%	16%
SERASA	0.7857*	0.6597*	0.6596*	52%	33%	33%
DEMCRED6	0.0291*	0.0207*	0.0207*	6%	3%	3%
DEMCRED1/ DEMCRED2	-0.3434*	-0.3051**	0.3051**	8%	5%	5%
SPATIALRISK	-	-0.3322*	0.3322*	-	30%	30%

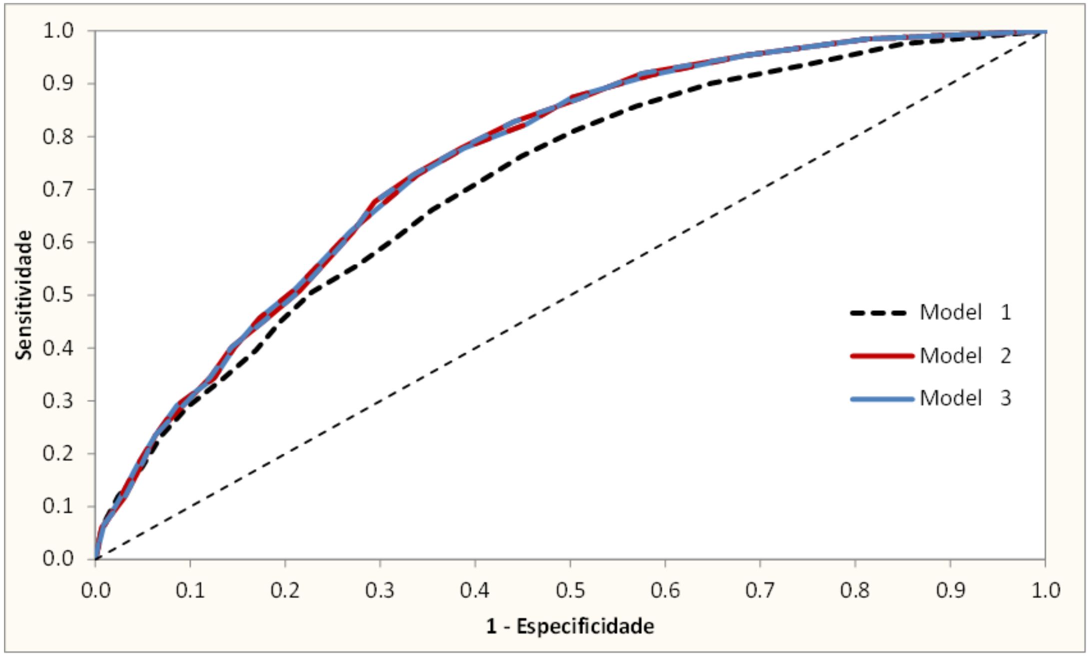
- ✓ 3 negative statement variables
- ✓ 3 credit demand variables
- ✓ Serasa bureau score

- ✓ Bureau score is the most important var.
- ✓ Spatial risk is 2nd most important var.
- ✓ Model 2 and 3 are much similar

Measure	Model 1	Model 2	Model 3
KS	29,53%	36,03%	36,02%
Gini	39,98%	47,13%	47,12%

Hand's KS/ROC test:

- ✓ Model 2 = Model 3
- ✓ Model 1 ≠ Model 2
- ✓ Model 1 ≠ Model 3



Conclusions

- ✓ There is evidence from past studies about the spatial dependence
- ✓ We proposed a proxy for capturing this spatial dependence based on neighbourhood
- ✓ 20 regions were found to have different spatial dependence patterns
- ✓ 20 spatial models were estimated for those
- ✓ The ordinary kriging methodology estimated the proxy for spatial risk
- ✓ The spatial risk component adds 6.5 p.p. in KS statistic and 7 p.p. in Gini
- ✓ There was no significant difference found between the naive logistic regression and the logistic model with measurement error

Reference

Agarwal, S.; Ambrose, B.W; Chomsisengphet, S. and Sanders, A.B. (2012): Thy Neighbor's Mortgage: Does Living in a Subprime Neighborhood Affect One's Probability of Default?, *Journal of Real Estate Economics*, *American Real Estate and Urban Economics Association,* vol. 40, No. 1, pp. 1-22.

Barro, D. and Basso, A. (2010): Credit contagion in a network of firms with spatial interaction*. European journal of Operational Research*. Vol. 205, pp. 459-468.

Cook, J. and Stefanski, L. (1994): Simulation-Extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Society*, Vol. 89, No. 428, pp. 1314-1328.

- **Fernandes**, G.B. (2012): Spatial credit risk measurement and its incorporation in credit scoring (In portuguese). *Master thesis, Insper-SP.*

Gerkman, L. (2011): Empirical spatial econometric modeling of small scale neighbourhood. *Journal of Geographical Systems*, 14(3), 283-298.

Goovaerts, P. (1997): *Geostatistics for natural resources evolution*. 1ª Ed., Oxford University Press, New York, USA.

Matheron, G. (1963): Principles of geostatistics, *Economic Geology*, Vol. 58, pp. 1246–1266.

Stine, R. (2011): Spatial temporal models for retail credit. *Credit Scoring and Credit Control Conference 2011*, Edinburgh, UK.