“SPATIAL DIMENSION OF CREDIT RISK: SPATIAL FILTER APPROACH “

ALEKSANDAR PETRESKI, ANDREAS STEPHAN

JÖNKÖPING INTERNATIONAL BUSINESS SCHOOL

CREDIT RISK – GENERAL

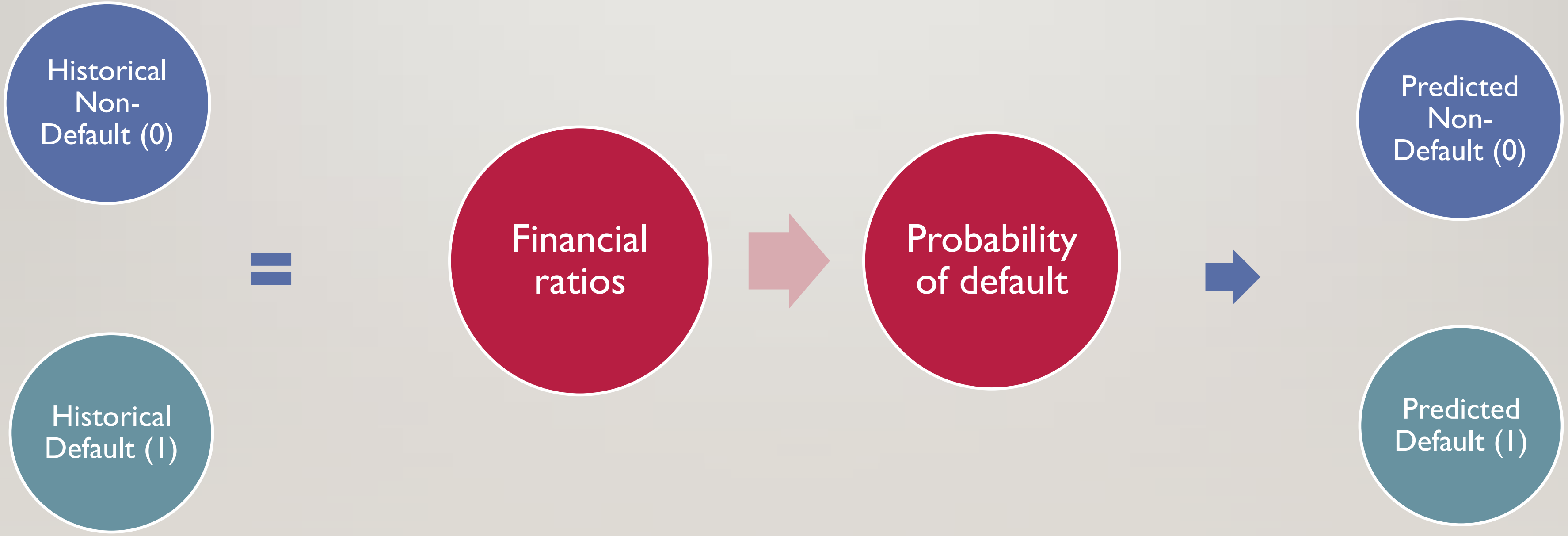
- Credit risk: ”Risk of default by the customer on the obligation”
- Banks are interested in having appropriate credit risk classification techniques, which help them to detect problematic clients and assess the credit exposure and potential losses,
- Central Banks are even more interested in banks having relilible credit risk models and techniques.
- Motivation for the paper:

Current models focus on global credit risk parameters, neglecting possible credit risk clusters and perhaps, underestimating credit risk parameters on the local level.

CREDIT RISK - GENERAL

- EL = PD * EAD * LGD (Basel Accord definition)
- Probability of default (PD): ”What is the probablity that the client will default on his obligation?”
- Exposure at Default (EAD:) ”If the customer defaults, how much is his current obligation at the time of default?”
- Loss given default (LGD): ”Once the customer defaults, how much will he pay from his current obligation?”

BINARY CHOICE MODEL

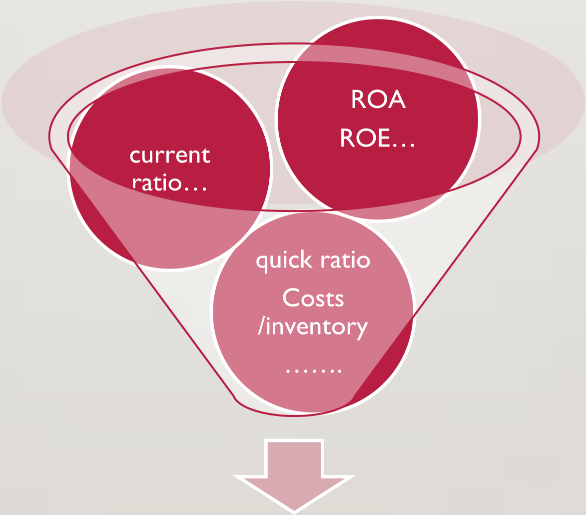


MODEL INPUTS

Financial ratios

liquidity indicators	turnover indicators	profitability indicators	debt indicators	banking credit indicator
current ratio = (current assets / current liability)	sales revenues / accounts receivable	ROE = (net profit / capital + reserves)	leverage = (liabilities / capital + reserves)	value of pledged collateral to outstanding credit (inverse of loan to value)
quick ratio = (current assets - inventory)/current liability	sales revenues / assets	ROA = (net profit / assets)	liabilities / assets	
	costs / inventory	Net profit margin = (net profit / sales revenues)	short term credit / sales revenues	
	costs / sales revenues		current liability / sales revenues	
	costs / accounts payable			

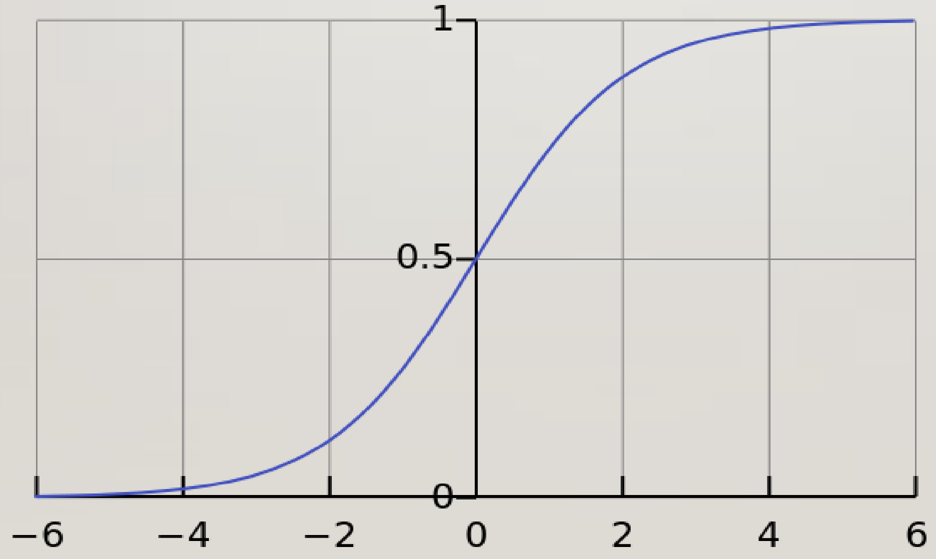
MODEL INPUTS - MULTICOLINEARITY



Probability of default (PD) = current ratio + sales revenues to accounts receivable + sales revenues to assets + Net profit margin + ROA+ ROE + leverage + liabilities / assets + collateral to outstanding credit

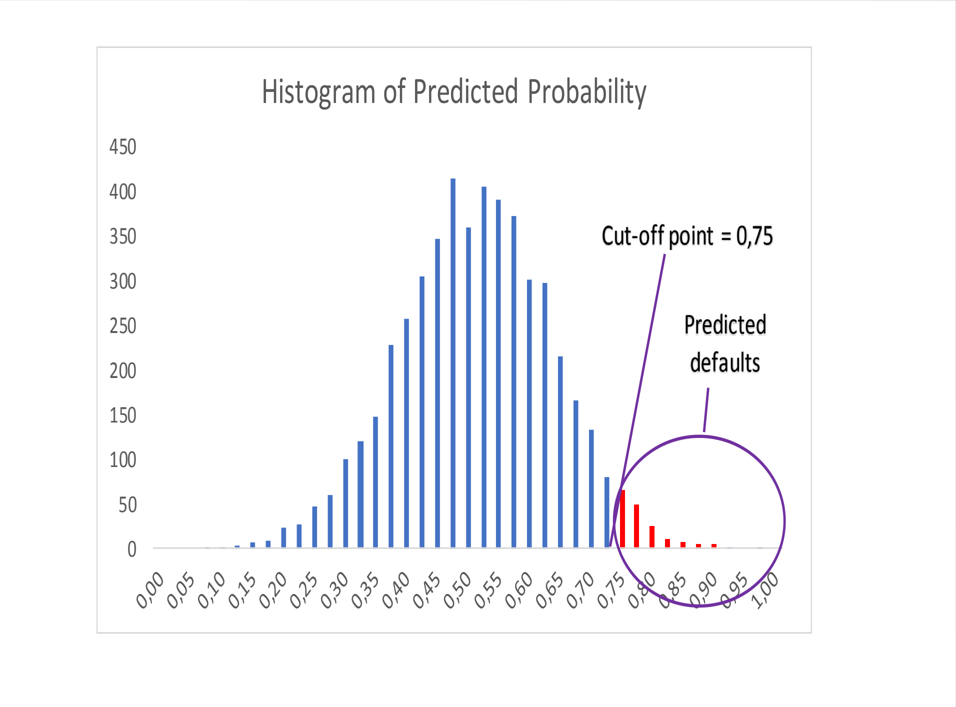
BINARY CHOICE MODEL - LOGIT

$$\text{Prob}(Y = 1 \mid \mathbf{x}) = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'\boldsymbol{\beta})} = \Lambda(\mathbf{x}'\boldsymbol{\beta}).$$



FROM PROBABILITY TO DEFAULT

- Once the model is estimated on the **training data**, we use the model on the **test data** to get **predicted probabilities**.
- Once the probabilities were calculated, the companies were **predicted/classified** as Non-default / Default according to appropriate **cut-off** point
- **Predicted** Non-default / Default are compared with **Historical** Non-default/Default



PREDICTION ACCURACY MEASURES

- Confusion matrix

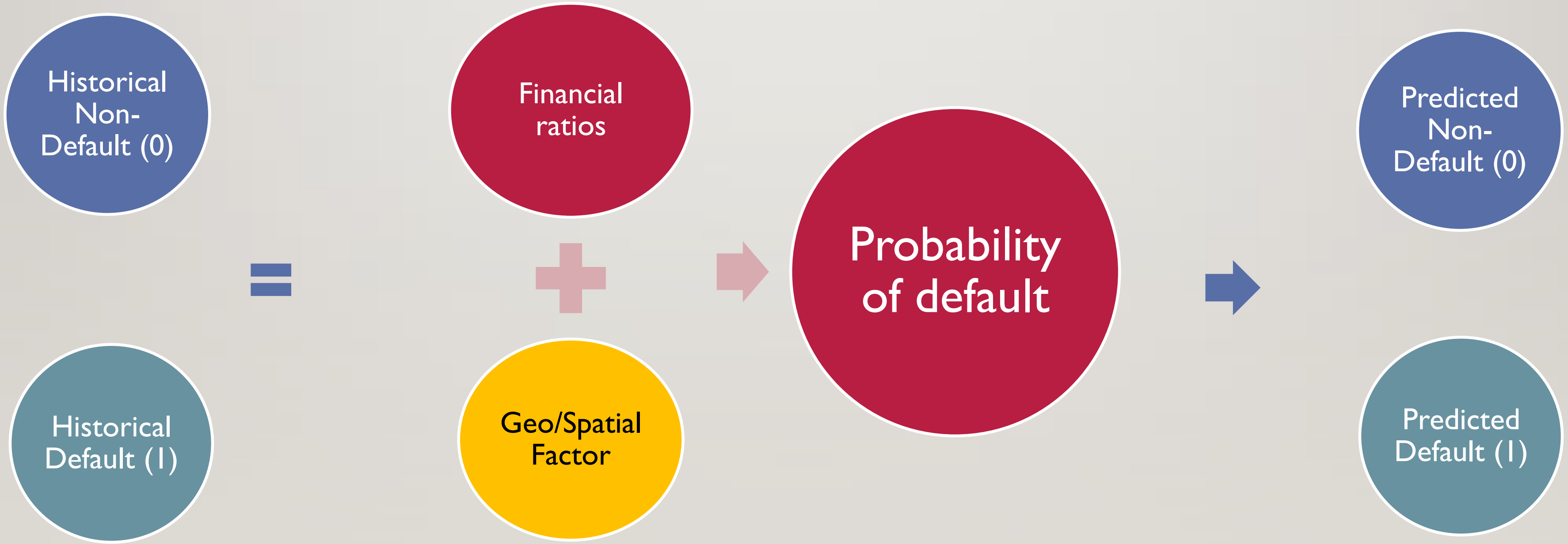
		actual	
		0	1
predicted	0	TP – True Positive	FP - False Positive
	1	FN – False Negative	TN – True Negative

- Positive (P) - the number of non-default cases ("0"s) in the data
- Negatives (N) - the number of default cases ("1"s) in the data
- Sensitivity (recall or True Positive rate) = TP / (TP + FN)
- Specificity (True negative rate) = TN / (TN + FP)
- Precision = TP/(TP + FP)
- Negative predictive value = TN / (TN + FN)
- Accuracy = (TP + TN)/ (P +N)
- F1 score = 2 * precision * sensitivity / (precision + sensitivity)
- (F1 score is the harmonic mean of precision and sensitivity)

NEW PARADIGM
– INCLUDE GEO & SPATIAL FACTOR

- Fernandes and Artes (2016) argue that adding kriging outcome variable in the logistic model improves its accuracy
- Albuquerque, Medina and Silva (2016) construct credit scoring models using Geographically Weighted Logistic Regression (GWLR) techniques

BINARY CHOICE MODEL- SPATIAL DIMENSION



NEW PARADIGM – INCLUDE GEO FACTOR

- Introducing a geo-component
- PD = current ratio + sales revenues to accounts receivable + sales revenues to assets + Net profit margin + ROA+ ROE + leverage + liabilities / assets + collateral to outstanding credit + **distance to capital**
- PD = current ratio + sales revenues to accounts receivable + sales revenues to assets + Net profit margin + ROA+ ROE + leverage + liabilities / assets + collateral to outstanding credit + **geographical dummy (rural / urban / cosmopolitan)**

INTRODUCING A GEO-COMPONENT

- **<u>Distance to capital</u>** does not have significant influence on the estimated probability of default

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3698  -0.5730  -0.4653  -0.3402   2.9367
Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)        -2.155e+00  4.065e-01  -5.301 1.15e-07 ***
Size3               5.142e-01  2.278e-01   2.257 0.024024 *
SECTORN             1.692e+00  5.251e-01   3.222 0.001272 **
revenues to assets -5.223e-01  1.006e-01  -5.192 2.08e-07 ***
ROA                -5.298e+00  2.901e+00  -1.826 0.067828 .
oblig. to assets    1.416e+00  3.745e-01   3.782 0.000156 ***
dist_from_centre    8.271e-06  1.120e-03   0.007 0.994107
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 2936.9  on 3856  degrees of freedom
Residual deviance: 2779.4  on 3821  degrees of freedom
AIC: 2851.4
```

INTRODUCING A GEO-COMPONENT

- <u>**Companies in rural municipalities**</u> have to some extent significantly lower estimated probability of default than the companies in the urban or cosmopolitan municipality

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3731  -0.5791  -0.4624  -0.3354   2.9572

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)        -1.9327884  0.4011482  -4.818 1.45e-06 ***
Size3               0.4574680  0.2200163   2.079 0.037595 *
SECTORN             1.5740263  0.5243641   3.002 0.002684 **
revenues to assets -0.5444658  0.0992840  -5.484 4.16e-08 ***
ROA                -5.9083472  2.8491434  -2.074 0.038105 *
obliga.to assets    1.3700925  0.3637948   3.766 0.000166 ***
rural□             -0.3355379  0.2033449  -1.650 0.098924 .
urban              -0.0349231  0.1031850  -0.338 0.735023
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 3096.1  on 4031  degrees of freedom
Residual deviance: 2918.2  on 3995  degrees of freedom
AIC: 2992.2
```

PREDICTION WITH GEO-MODEL - DISTANCE TO CAPITAL

Base model

quantile	TP	FP	FN	TN	sensitivity	accuracy	recall	F1score	precision	true negative rate
0,25	75	8	856	149	0,1483	0,2059	0,0806	0,1479	0,9036	0,9490
0,50	116	10	815	147	0,1528	0,2417	0,1246	0,2195	0,9206	0,9363
0,75	161	16	770	141	0,1548	0,2776	0,1729	0,2906	0,9096	0,8981
0,90	219	18	712	139	0,1633	0,3290	0,2352	0,3750	0,9241	0,8854
0,95	243	20	688	137	0,1661	0,3493	0,2610	0,4070	0,9240	0,8726
0,99	397	34	534	123	0,1872	0,4779	0,4264	0,5830	0,9211	0,7834

Model with geo-variable (distance to the capital)

quantile	TP	FP	FN	TN	sensitivity	accuracy	recall	F1score	precision	true negative rate
0,25	75	8	856	149	0,1483	0,2059	0,0806	0,1479	0,9036	0,9490
0,50	116	10	815	147	0,1528	0,2417	0,1246	0,2195	0,9206	0,9363
0,75	161	16	770	141	0,1548	0,2776	0,1729	0,2906	0,9096	0,8981
0,90	219	18	712	139	0,1633	0,3290	0,2352	0,3750	0,9241	0,8854
0,95	243	20	688	137	0,1661	0,3493	0,2610	0,4070	0,9240	0,8726
0,99	398	33	533	124	0,1887	0,4798	0,4275	0,5844	0,9234	0,7898

PREDICTION WITH GEO-MODEL - GEOGRAPHICAL DUMMY

Base model

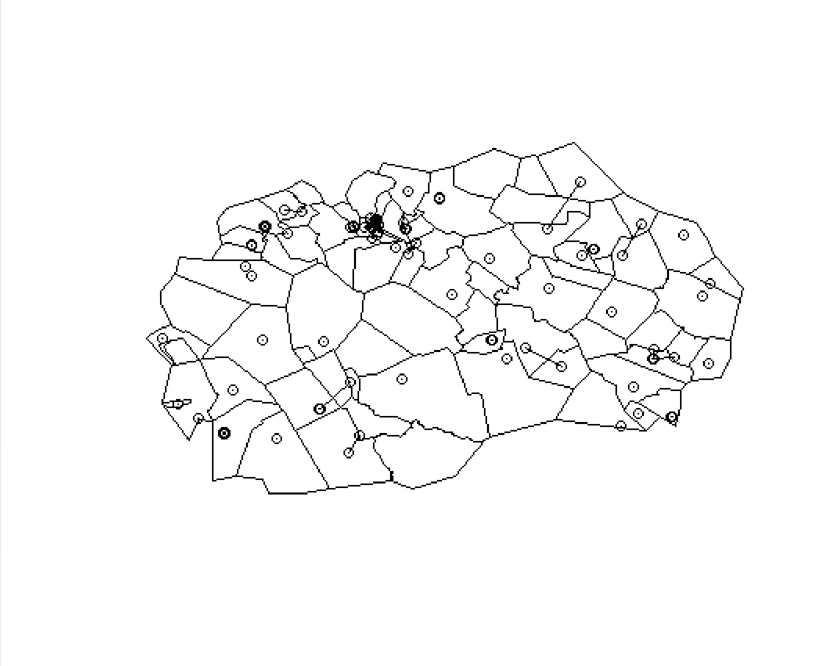
quantile	TP	FP	FN	TN	sensitivity	accuracy	F1 score	precision	recall	true negative rate
0,25	77	8	887	158	0,1512	0,2080	0,0799	0,1468	0,9059	0,9518
0,50	119	10	845	156	0,1558	0,2434	0,1234	0,2177	0,9225	0,9398
0,75	173	17	791	149	0,1585	0,2850	0,1795	0,2998	0,9105	0,8976
0,90	235	18	729	148	0,1688	0,3389	0,2438	0,3862	0,9289	0,8916
0,95	274	21	690	145	0,1737	0,3708	0,2842	0,4353	0,9288	0,8735
0,99	445	35	519	131	0,2015	0,5097	0,4616	0,6163	0,9271	0,7892

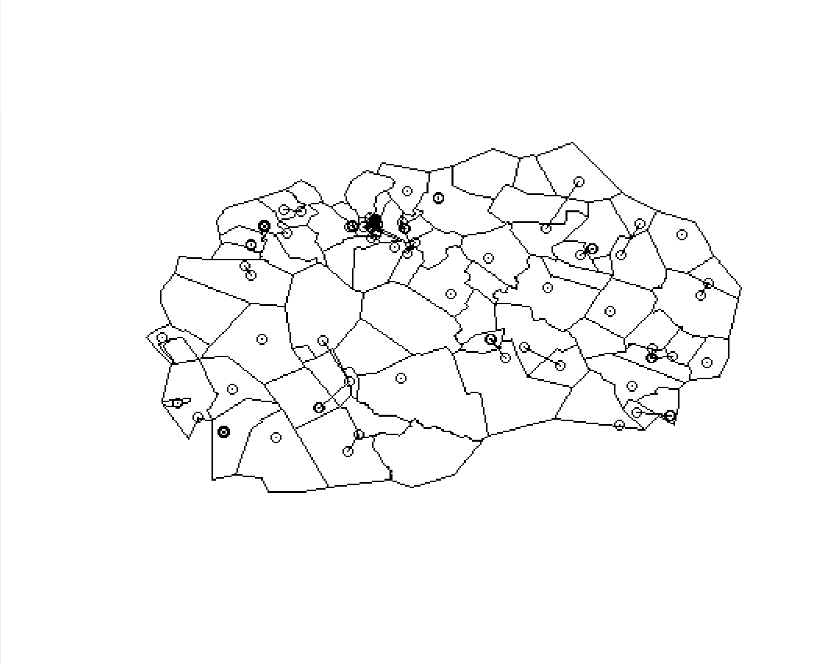
Model with dummy geo-variable (cosmopolitan/urban/rural)

quantile	TP	FP	FN	TN	sensitivity	accuracy	F1 score	precision	recall	true negative rate
0,25	77	9	887	157	0,1504	0,2071	0,0799	0,1467	0,8953	0,9458
0,50	121	11	843	155	0,1553	0,2442	0,1255	0,2208	0,9167	0,9337
0,75	172	17	792	149	0,1583	0,2841	0,1784	0,2984	0,9101	0,8976
0,90	232	19	732	147	0,1672	0,3354	0,2407	0,3819	0,9243	0,8855
0,95	270	21	694	145	0,1728	0,3673	0,2801	0,4303	0,9278	0,8735
0,99	446	38	518	128	0,1981	0,5080	0,4627	0,6160	0,9215	0,7711

NEW PARADIGM
– INCLUDE SPATIAL FACTOR

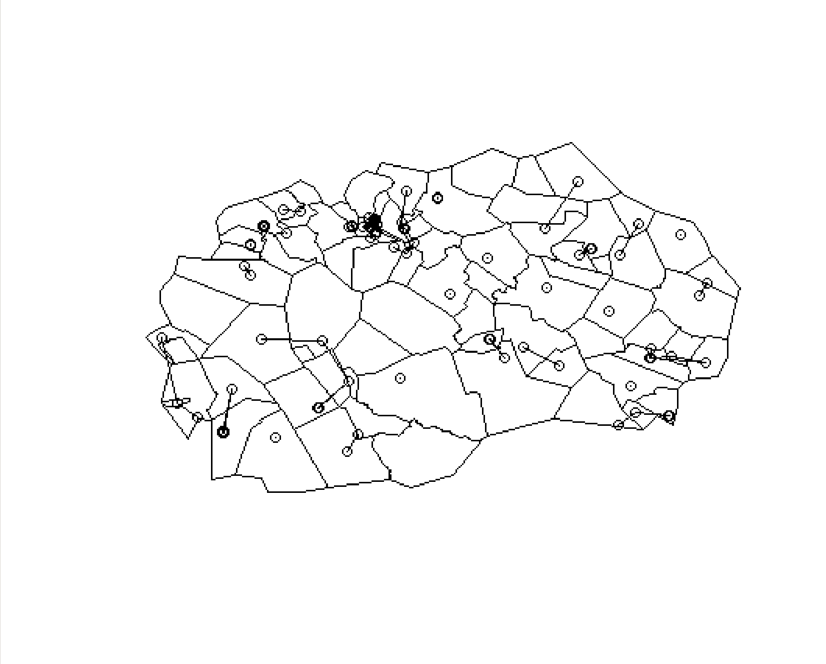
Build spatial component (1) : Coordinates -> Knn nearest neighbor object -> Neighbour list object

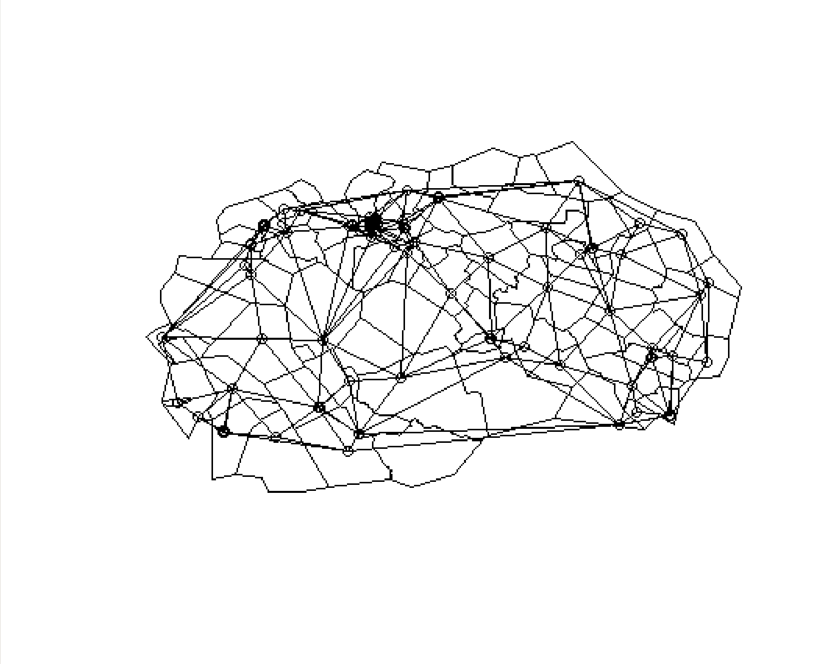




NEW PARADIGM
– INCLUDE SPATIAL FACTOR

Build spatial component (2) : Coordinates -> Knn nearest neighbor object -> Neighbour list object





NEW PARADIGM – INCLUDE SPATIAL FACTOR

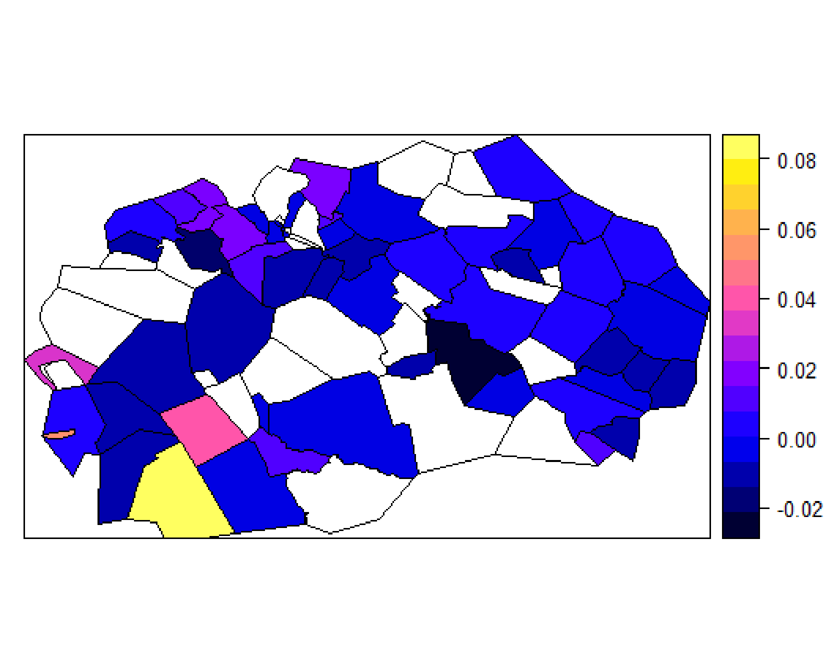
- Use Neighbour list object in the Spatial Filtering.
- When fitting the model with Spatial Filtering, Moran eigenvector GLM filtering is used as in Bivand (2008), which uses brute force to search the set of eigenvectors of the matrix ***MWM***:

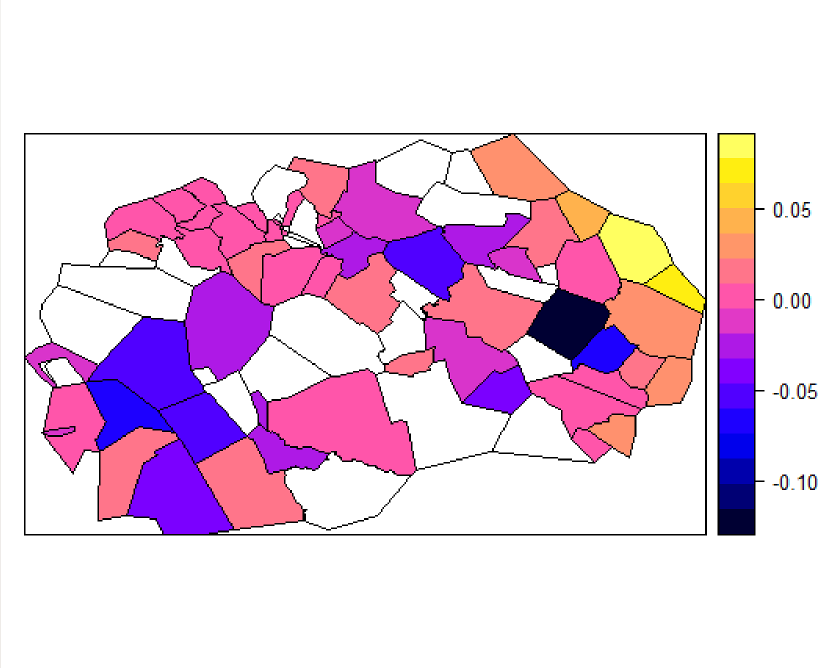
$$M = I - X(X^t X)^{-1} X^t$$

- ***M*** is a symmetric and idempotent projection matrix and ***W*** are the spatial weights.
- Once the spatial filter is applied, spatial eigenvectors are used in the main equation.
- PD = current ratio + sales revenues to accounts receivable + sales revenues to assets + Net profit margin + ROA+ ROE + leverage + liabilities / assets + collateral to outstanding credit + fitted spatial eigenvector

NEW PARADIGM
– INCLUDE SPATIAL FACTOR

Build spatial component (3) : Spatial eigenvectors



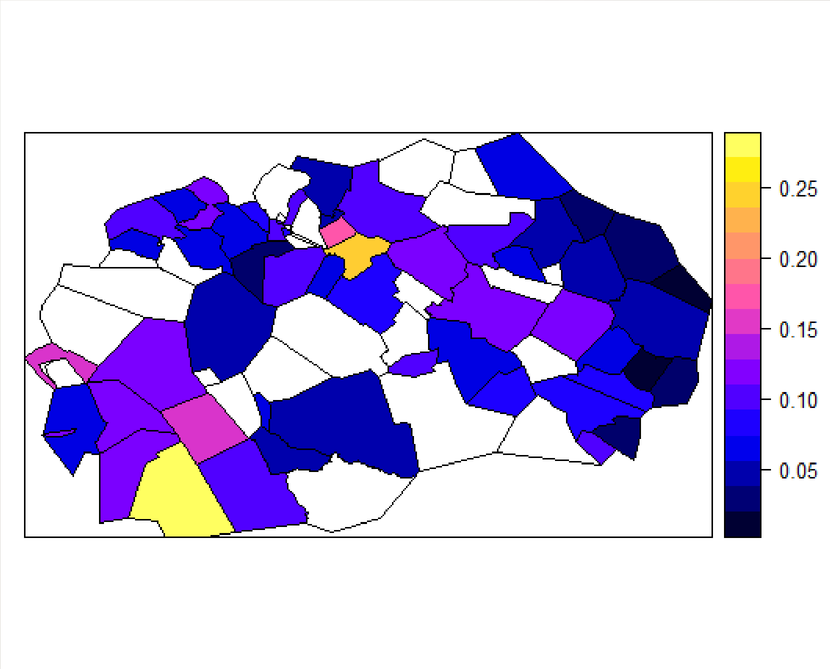


NEW PARADIGM – INCLUDE SPATIAL FACTOR

BASE MODEL		**SPATIAL MODEL**	
Coefficients:		Coefficients:	
(Intercept)	-2.028309	(Intercept)	-1.998662
current ratio	-0.105135	current ratio	-0.112115
sales revenues to accounts receivable	0.002093	sales revenues to accounts receivable	0.002318
sales revenues to assets	-0.191145	sales revenues to assets	-0.190830
Net profit margin	4.772017	Net profit margin	4.816988
ROA	-14.642937	ROA	-15.011321
ROE	3.232302	ROE	3.292886
leverage	0.022535	leverage	0.020073
liabilities / assets	0.246656	liabilities / assets	0.218358
collateral to outstanding credit	0.010920	collateral to outstanding credit	0.003887
		fitted(spatial)vec1	-7.262897
		fitted(spatial)vec2	-2.609019
Degrees of Freedom: 1105 Total (i.e. Null); 1096 Residual		Degrees of Freedom: 1105 Total (i.e. Null); 1094 Residual	
Null Deviance: 689.6		Null Deviance: 689.6	
Residual Deviance: 657.5 AIC: 677.5		Residual Deviance: 650.8 AIC: 668.9	

INTRODUCING SPATIAL COMPONENT

- Estimated probability of default using spatial model



PREDICTION WITH SPATIAL MODEL

- Prediction results depend on the created weight matrix
- With the increase of the neighbor links, prediction by the spatial model increases and slightly outperforms the base model.
- The form of the graph of neighbor relationship determines the significance of the spatial autocorrelation tests

	base model	1 link	2 links	4 links	triangulation links
AIC	677,5	671,5	675,9	682	668,9
number of links in the neighbor object		1	2	4	5,96
used eigenvectors		9	8	5	2
ANOVA Pr(>Chi)		0,00426	0,02413	0,09667	0,001826
average prediction accuracy (from all accuracy measures)	0,4360	0,4319	0,4346	0,4362	0,4400
Moran I statistic p value		0,1510	0,1603	0,0043	0,0217
Observed Moran I		0,0382	0,0541	0,0540	0,0347

CONCLUSION

- In general, spatial filter enhance the fit and can slightly improve the prediction of the credit risk model. Effect of adding eigenvectors to the base model is higher then the effect of adding some geographical dummy or geographical variable like distance to the capital.
- Still, geo-dummies help us to detect that companies in rural municipalities have lower probability of default, compared with rural/urban areas, although not very significant.
- It should be noted however that the fit and prediction results depend on the created weight matrix when using spatial filtering. With the increase of the neighbor links, the prediction by the spatial model increase and slightly outperform the base model.
- It was confirmed that the form of the neighbor relationship determines the significance of the spatial autocorrelation tests.
- Positive autocorrelation indicate existence of clusters of defaults within geographical area, which could confirm the need for use of spatial techniques.