In search of w for the spatial lag model

Ghislain Geniaux

In search of W for the spatial lag model

[1] G. GENIAUX

[1]INRA SAD Ecodeveloppement Avignon (UR 767)

SEA 2012

Plan

Introduction

The spatial econometrics literature points out the importance of W choice but *"tells us little about adequate foundations for these choices"* (Harris et al., 2011)

The choice of W that corresponds to the true data generating process is crucial for the consistency of parameter estimates (Bhattacharjee et al., 2006) of Spatial Lag Model (SLM).

How econometricians can choose W ?

Harris et al. (2011) identify three main approaches.

- First a common practice that consists of comparing pre-specified versions of W using "goodness of fit statistics" like AIC to choose the best version of W. (LeSage and Fischer, 2008; Stakhovych and Bijmolt, 2009)

How econometricians can choose W ?

- A second approach starts with an unspecified spatial weight matrix and try to estimate a spatial weights matrices that are consistent with an observed pattern of spatial dependence. (Conley, 1999; Pinkse et al., 2002; Meen, 1996; Bhattacharjee et al., 2006)

How econometricians can choose W ?

- A third way is to use non parametric approaches for measuring spatial correlation of a single variable. López et al. (2010) uses the concept of symbolic entropy as a measure of spatial dependence. Other propositions use Moran or Ord and Getis (1995) local statistics (Aldstadt and Getis, 2006) to identify the most suitable W.

Aim of the paper

We propose in this article a parametric approach to estimate matrix W for SLM model at the frontier of the two first approaches discussed above.

The estimation procedure uses a (differentiable) flexible distance kernel with two parameters (the bandwidth h and the sharpness of weight decreasing around bandwidth k) to identify the matrix W.

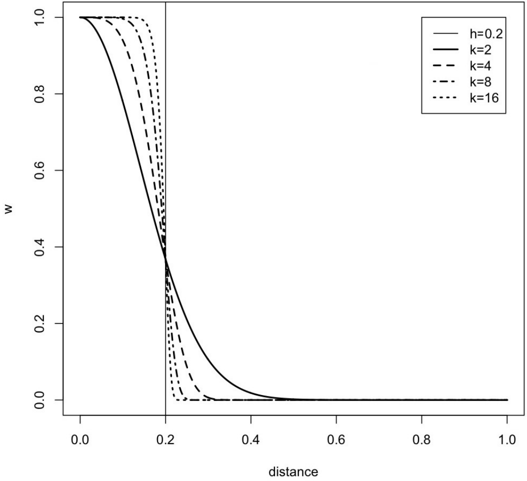
SLM with unknow W

We consider a spatial lag model (Anselin, 1980, 1988) with a weighting scheme based on distance that could be written in matrix form as:

$$Y = \lambda W(h, k)Y + X\beta + \epsilon \tag{1}$$

Gaussian kernel weighting scheme

$$w_{ij} = \frac{e^{(-d_{ij}/h)^k}}{\sum_j e^{(-d_{ij}/h)^k}}$$



IV estimates of SLM

Suppose first that h is known: $W(h) = W$.

Let us denote $\theta = [\lambda, \beta] \in \Theta$, where Θ is the parameter open space. IV estimation provides a consistent estimator of θ using $Z \in \Big[X, WX, W^2X, W^3X]$. Let note $Q = [WY, X]$, and $\tilde{Q} = P_Z Q = [P_Z X, P_Z WY]$, where $P_Z = Z(Z'Z)^{-1}Z'$ then we can rewrite the model (1) as:

$$Y = \tilde{Q}\theta + \epsilon \tag{2}$$

Assumptions

If we use a wrong $h_0 \neq h$ and/or $k_0 \neq k$ to estimate θ_{iv}, then θ_{iv} will be biased.

We could define $SSR_{iv}(h,k) = \epsilon(h,k)'\epsilon(h,k)$ the function that gives the sum of square of residuals of regression (2) for a given couple (h,k).

We do a conjecture that the minimization of $SSR_{iv}(h,k)$ will provide the true (h,k) and a consistent estimate of θ_{iv}.

A GNR for the SLM IV estimates

To simplify notation, we suppose first that k is known with $k = 2$.

Let note $\epsilon(h)$ the value of residual of model (2) using $\hat{\theta}_{iv}(h)$:

$$\epsilon(h) = Y - \hat{\lambda}_{iv}(h)\hat{P}_{Z(h)}W(h)Y - P_{Z(h)}X\hat{\beta}_{iv}(h) = Y - H(\hat{\theta}_{iv}(h)) \tag{3}$$

A GNR for the SLM IV estimates

Following Davidson et al. (1993), a *GNR* for such non linear regression can be defined as:

$$Y - H(\hat{\theta}_{iv}(\hat{h})) = \frac{\partial H}{\partial h}\Big|_{\hat{h}} b + \epsilon \tag{4}$$

Then:

$$\frac{\partial H}{\partial h}\Big|_{\hat{h}} = \hat{\theta}_{iv}^{h} P_{\hat{Z}} Q + \hat{\theta}_{iv} P_{\hat{Z}}^{h} Q + \hat{\theta}_{iv} P_{\hat{Z}} Q^{h}$$

Gradients definition

$$\hat{\theta}^h_{iv} = (\hat{Q}'P_{\hat{Z}}\hat{Q})^{-1}\Big[-(\hat{Q}'^h P_{\hat{Z}}\hat{Q} + \hat{Q}'P^h_{\hat{Z}}\hat{Q} + \hat{Q}'P_{\hat{Z}}\hat{Q}^h)(\hat{Q}'P_{\hat{Z}}\hat{Q})^{-1}\hat{Q}'P_{\hat{Z}}y + (\hat{Q}'^h P_{\hat{Z}} + \hat{Q}'P^h_{\hat{Z}})y\Big]$$

$$\hat{Q}^h = [0 \ \vdots \ \hat{W}^h Y]$$

$$\hat{W}^h_{ij} = \frac{2}{\hat{h}^3}\frac{e^{-(D_{ij}/\hat{h})^2}}{\sum_k e^{-(D_{ik}/\hat{h})^2}}\left[\frac{\sum_l (D^2_{ij} - D^2_{il})e^{-(D_{il}/\hat{h})^2}}{\sum_k e^{-(D_{il}/\hat{h})^2}}\right] = \frac{2W_{ij}}{h^3}(D^2_{ij} - \sum_j D^2_{ij}W_{ij})$$

$$P^h_{\hat{Z}} = \hat{Z}^h(\hat{Z}'\hat{Z})^{-1}\hat{Z}' + \hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'^h - \hat{Z}(\hat{Z}'\hat{Z})^{-1}(\hat{Z}'^h\hat{Z} + \hat{Z}'\hat{Z}^h)(\hat{Z}'\hat{Z})^{-1}\hat{Z}'$$

$$\hat{Z}^h = \Big[0 \ \vdots \ W^h X \ \vdots \ (\hat{W}^h\hat{W} + \hat{W}\hat{W}^h)X \ \vdots \ (\hat{W}^h\hat{W}\hat{W} + \hat{W}\hat{W}^h\hat{W} + \hat{W}\hat{W}\hat{W}^h)X\Big]$$

Gradients definition - general case

$$Y - H(\hat{\theta}_{iv}(\hat{h}, \hat{k})) = \frac{\partial H}{\partial h}\Big|_{\hat{h},\hat{k}} b_h + \frac{\partial H}{\partial k}\Big|_{\hat{h},\hat{k}} b_h + \epsilon$$

Where superscript h (resp. k) indicates a derivative with respect to h (resp. k):

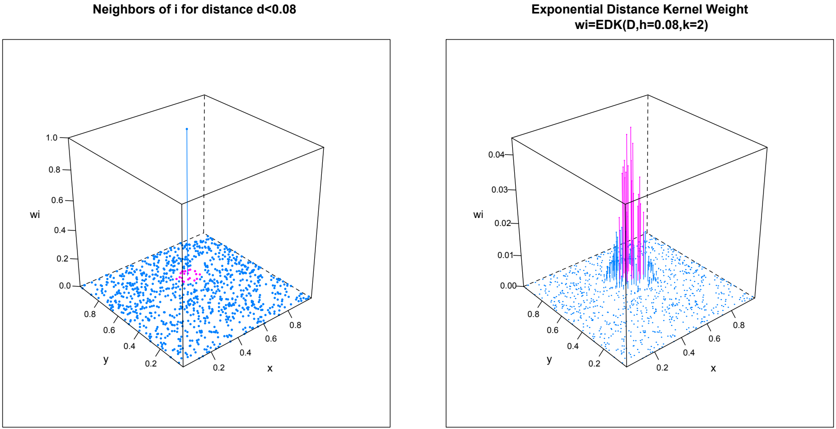
$$\begin{aligned} W_{ij}^h &= \frac{kW_{ij}}{h^{k+1}}\left(D_{ij}^k - \sum_j D_{ij}^k W_{ij}\right) \\ W_{ij}^k &= -W_{ij}\left(ln(D_{ij}/h)(D_{ij}/h)^k - \sum_j ln(D_{ij}/h)(D_{ij}/h)^k W_{ij}\right) \end{aligned}$$

MC Settings

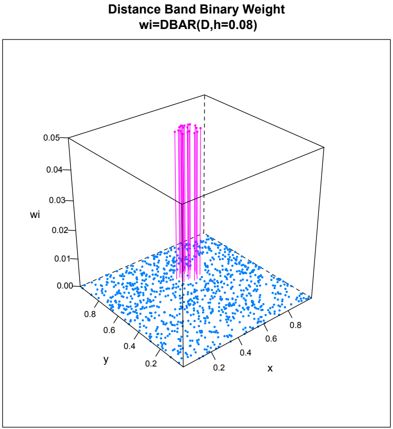
All experiments are based on the same value of $\beta = (1, 0.5)$, a single run of $X \sim N(0,1)$ and 1000 replications of $\epsilon \sim N(0,1)$. An experiment is defined by a quintuplet $(l, wt, lambda, h, k)$, where:

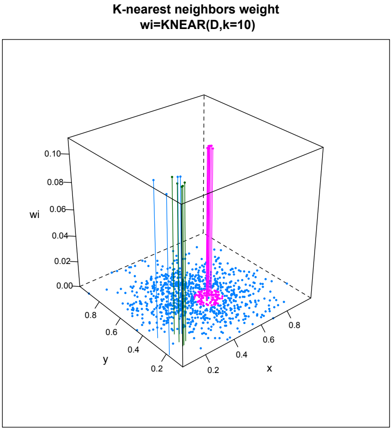
1. The true weighting scheme type, noted *wt* can be an exponential distance kernel (*EDK*), a distance band binary weight (*DBAR*) and a k-nearest neighbors (*KNEAR*).

MC Settings


Neighbors of i for distance d<0.08
wi
1.0
0.8
0.6
0.4
0.2
0.0
y
x
Exponential Distance Kernel Weight
wi=EDK(D,h=0.08,k=2)
wi
0.04
0.03
0.02
0.01
0.00
y
x

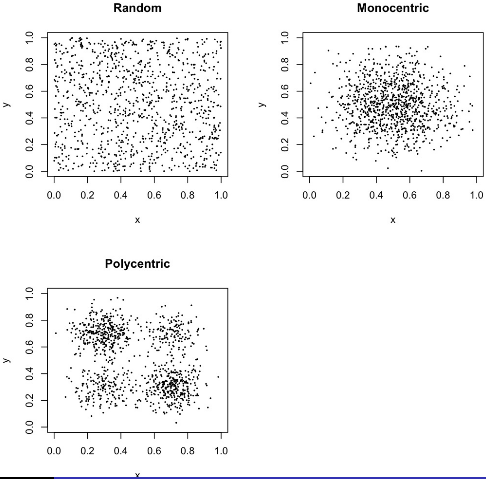
MC Settings


Distance Band Binary Weight
wi=DBAR(D,h=0.08)
wi
0.05
0.04
0.03
0.02
0.01
0.00
y
0.8
0.6
0.4
0.2
x
0.2
0.4
0.6
0.8


K-nearest neighbors weight
wi=KNEAR(D,k=10)
wi
0.10
0.08
0.06
0.04
0.02
0.00
y
0.8
0.6
0.4
0.2
x
0.2
0.4
0.6
0.8

MC Settings

2. l is the type of spatial distribution of coordinates. $l \in$ *(random, monocentric, polycentric)*



MC Settings

3. $\lambda \in (0.2, 0.4, 0.6)$.
4. $h \in (0.008, 0.015, 0.03)$ for all wt.
5. $k \in (2, 3, 12)$ for $wt = EDK$

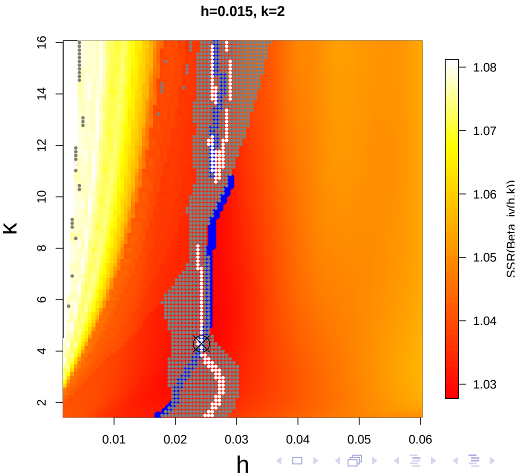
MC Settings

We draw three types of experiments :

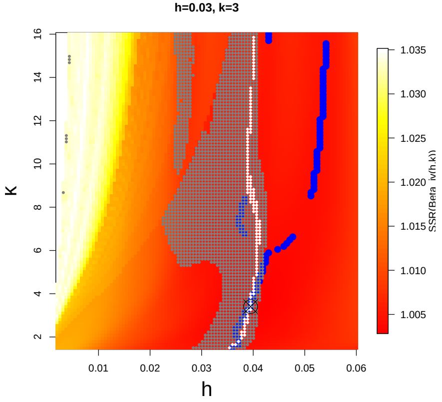
1. In first type of experiments, we try to minimize $SSR_{iv}(h)$ with respect to h with $wt = EDK$ and considering a known $k = 2$.
2. In second type of experiments, we try to minimize $SSR_{iv}(h, k)$ with respect to (h, k) with $wt = EDK$.
3. In the last type of experiments, we use non differentiable weighting schemes ($wt \in \ (DBAR, KNEAR)$) for generating data but use differentiable exponential distance kernel for identifying W.

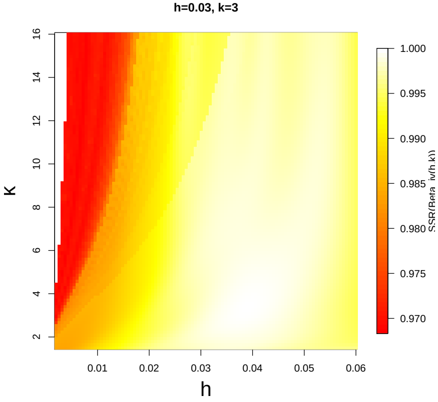
Exploring $SSR_{iv}(h, k)$

Surface plot of $SSR_{iv}(h, k)$ for $\lambda = 0.2,\ k = 2$ and $h = 0.015$. We use a 100 x 100 grid with $h \in [0.002, 0.1]$ and $k \in [1, 16]$.



Exploring $SSR_{iv}(h, k)$





MC Results for k known and $wt = EDK$

$n = 1000$, $r = 1000$, $\beta = (1, 0.5)$, $wt = EDK$, $l = random$

Exp. parameters			stat.	$\hat{\beta}_{ols}$		h estimates			$\hat{\theta}_{iv}(h^\star)$		
h	k	λ		$\beta 0_{ols}$	$\beta 1_{ols}$	$h^\star$	h_{gnr}	h_{ignr}	$\beta 0_{iv}$	$\beta 1_{iv}$	λ_{iv}
0.008	2	0.2	MEAN	1.2422	0.5118	0.0085	0.0088	0.0087	0.9952	0.4994	0.2033
			RMSE	*0.2456*	*0.0162*	*0.0035*	*0.0031*	*0.004*	*0.042*	*0.0108*	*0.022*
0.015	2	0.2	MEAN	1.2454	0.5088	0.0158	0.0153	0.0168	0.9927	0.4993	0.2054
			RMSE	*0.2488*	*0.014*	*0.0056*	*0.0036*	*0.0094*	*0.0466*	*0.0109*	*0.0271*
0.03	2	0.2	MEAN	1.2464	0.5024	0.0354	0.0238	0.0446	0.9783	0.4993	0.2171
			RMSE	*0.2497*	*0.011*	*0.0175*	*0.0081*	*0.0776*	*0.0665*	*0.0108*	*0.0472*
0.008	2	0.4	MEAN	1.6507	0.5546	0.0081	0.0083	0.0081	0.9973	0.4994	0.4013
			RMSE	*0.6531*	*0.0561*	*0.0016*	*0.0015*	*0.0016*	*0.0452*	*0.0111*	*0.019*
0.015	2	0.4	MEAN	1.6579	0.542	0.0152	0.0152	0.0154	0.9961	0.4993	0.402
			RMSE	*0.6602*	*0.0437*	*0.002*	*0.0018*	*0.0024*	*0.0503*	*0.0111*	*0.0232*
0.03	2	0.4	MEAN	1.6591	0.5149	0.0314	0.0244	0.0329	0.9881	0.4993	0.4068
			RMSE	*0.6613*	*0.0186*	*0.0065*	*0.0061*	*0.013*	*0.0703*	*0.0109*	*0.0377*
0.008	2	0.6	MEAN	2.4717	0.6595	0.0082	0.0083	0.0082	0.9979	0.4994	0.6006
			RMSE	*1.4743*	*0.1604*	*0.0011*	*0.001*	*0.0011*	*0.0479*	*0.0115*	*0.0141*
0.015	2	0.6	MEAN	2.4869	0.6217	0.0154	0.0153	0.0155	0.9968	0.4993	0.601
			RMSE	*1.4893*	*0.1227*	*0.0014*	*0.0012*	*0.0015*	*0.0548*	*0.0115*	*0.0177*
0.03	2	0.6	MEAN	2.4872	0.5451	0.0312	0.025	0.0316	0.9909	0.4993	0.6034
			RMSE	*1.4894*	*0.0468*	*0.004*	*0.0051*	*0.0058*	*0.0797*	*0.0111*	*0.0293*

Table 1: Results of MC simulations for k known, random location and gaussian distance kernel.

Results for k unknown and $wt = EDK$

$n = 1000$, $r = 300$, $\beta = (1, 0.5)$, $wt = EDK$, $l = random$, $k^\star \in [1.5, 16]$										
Exp. parameters			stat.	$\hat{\beta}_{ols}$		h, k estimates		$\hat{\theta}_{iv}$		
h	k	λ		$\beta 0_{ols}$	$\beta 1_{ols}$	$h^\star$	$k^\star$	$\beta 0_{iv}$	$\beta 1_{iv}$	λ_{iv}
0.015	2	0.2	MEAN	1.2455	0.5085	0.0176	5.7751	0.9836	0.4988	0.2131
			RMSE	*0.2484*	*0.0141*	*0.0068*	*5.4635*	*0.0497*	*0.0111*	*0.0316*
0.03	2	0.2	MEAN	1.2465	0.5022	0.0398	7.9627	0.9561	0.4989	0.2352
			RMSE	*0.2494*	*0.0113*	*0.021*	*8.2962*	*0.0759*	*0.0111*	*0.0563*
0.015	3	0.2	MEAN	1.244	0.5106	0.0155	5.3964	0.9892	0.4989	0.2085
			RMSE	*0.247*	*0.0155*	*0.0054*	*4.1402*	*0.0438*	*0.0112*	*0.0255*
0.03	3	0.2	MEAN	1.2467	0.5042	0.0328	7.4014	0.9698	0.4988	0.2241
			RMSE	*0.2495*	*0.0119*	*0.0123*	*6.7558*	*0.0615*	*0.0111*	*0.0444*
0.015	12	0.2	MEAN	1.2205	0.5126	0.0139	9.9576	0.9902	0.4988	0.2085
			RMSE	*0.2237*	*0.017*	*0.0036*	*3.877*	*0.0414*	*0.0112*	*0.0245*
0.03	12	0.2	MEAN	1.246	0.5068	0.0302	11.2826	0.9851	0.499	0.2118
			RMSE	*0.2489*	*0.0131*	*0.0062*	*4.2003*	*0.0502*	*0.0111*	*0.0321*

Results $k = 2$ unknown and $wt = EDK$

$n = 1000$, $r = 300$, $\beta = (1, 0.5)$, $wt = EDK$, $l = random$, $k^\star \in [1.5, 16]$										
Exp. parameters			stat.	$\hat{\beta}_{ols}$		h, k estimates		$\hat{\theta}_{iv}$		
h	k	λ		$\beta 0_{ols}$	$\beta 1_{ols}$	$h^\star$	$k^\star$	$\beta 0_{iv}$	$\beta 1_{iv}$	λ_{iv}
0.015	2	0.4	MEAN	1.6576	0.5417	0.0154	2.9559	0.993	0.499	0.404
			RMSE	*0.6597*	*0.0435*	*0.004*	*2.5002*	*0.0481*	*0.0113*	*0.0225*
0.03	2	0.4	MEAN	1.6592	0.5147	0.0323	3.5379	0.9777	0.4989	0.4133
			RMSE	*0.6612*	*0.0187*	*0.0093*	*3.8923*	*0.072*	*0.0112*	*0.0387*
0.015	3	0.4	MEAN	1.6541	0.5501	0.0149	3.5756	0.9957	0.4991	0.4024
			RMSE	*0.6562*	*0.0518*	*0.0031*	*1.7849*	*0.0446*	*0.0114*	*0.0195*
0.03	3	0.4	MEAN	1.6594	0.5228	0.0299	4.1771	0.9848	0.4988	0.409
			RMSE	*0.6613*	*0.0257*	*0.0054*	*3.342*	*0.061*	*0.0112*	*0.0315*
0.015	12	0.4	MEAN	1.5926	0.5565	0.0147	11.2667	0.9932	0.4988	0.4044
			RMSE	*0.5947*	*0.058*	*0.0018*	*2.3694*	*0.0431*	*0.0114*	*0.0199*
0.03	12	0.4	MEAN	1.6579	0.5341	0.0299	11.6918	0.9921	0.499	0.4045
			RMSE	*0.6599*	*0.0362*	*0.0017*	*3.3903*	*0.0514*	*0.0113*	*0.0254*

Results $k = 2$ unknown and $wt = EDK$

$n = 1000$, $r = 300$, $\beta = (1, 0.5)$, $wt = EDK$, $l = random$, $k^\star \in [1.5, 16]$										
Exp. parameters			stat.	$\hat{\beta}_{ols}$		h, k estimates		$\hat{\theta}_{iv}$		
h	k	λ		$\beta 0_{ols}$	$\beta 1_{ols}$	$h^\star$	$k^\star$	$\beta 0_{iv}$	$\beta 1_{iv}$	λ_{iv}
0.015	2	0.6	MEAN	2.486	0.6212	0.015	2.0626	0.9962	0.4991	0.6014
			RMSE	1.4882	0.1222	0.003	0.7786	0.0521	0.0117	0.0168
0.03	2	0.6	MEAN	2.4872	0.5449	0.0307	2.3044	0.9854	0.4989	0.6058
			RMSE	1.4892	0.0467	0.006	1.6008	0.0793	0.0113	0.0284
0.015	3	0.6	MEAN	2.4785	0.6466	0.015	3.1016	0.998	0.4992	0.6007
			RMSE	1.4807	0.1476	0.0021	0.8826	0.0477	0.0117	0.0146
0.03	3	0.6	MEAN	2.4868	0.5673	0.0296	3.2238	0.9887	0.4989	0.6044
			RMSE	1.4888	0.0686	0.0044	1.7018	0.0681	0.0115	0.0238
0.015	12	0.6	MEAN	2.3416	0.6645	0.015	11.8847	0.9962	0.499	0.6015
			RMSE	1.3439	0.1654	9e-04	1.4611	0.0445	0.0119	0.0142
0.03	12	0.6	MEAN	2.4834	0.6005	0.0301	11.7019	0.9942	0.499	0.6022
			RMSE	1.4855	0.1015	0.0011	2.8155	0.0565	0.0116	0.0192

Results k and wt unknown

$n = 1000$, $r = 300$, $\beta = (1, 0.5)$
wt=randomly $(EDK, DBAR, KNEAR)$
$l = random$, k =randomly $(2, 3, 12)$

h	λ	TRUE wt	freq	% good pred.
0.015	0.2	DBAR	101	29.7
0.015	0.2	EDK	96	98.96
0.015	0.2	KNEAR	103	100
0.015	0.4	DBAR	92	40.22
0.015	0.4	EDK	113	100
0.015	0.4	KNEAR	95	100
0.015	0.6	DBAR	96	60.42
0.015	0.6	EDK	93	100
0.015	0.6	KNEAR	111	100

$n = 1000$, $r = 300$, $\beta = (1, 0.5)$
wt=randomly $(EDK, DBAR, KNEAR)$
$l = random$, k =randomly $(2, 3, 12)$

h	λ	TRUE wt	freq	% good pred.
0.03	0.2	DBAR	107	70.09
0.03	0.2	EDK	94	70.21
0.03	0.2	KNEAR	99	100
0.03	0.4	DBAR	112	95.54
0.03	0.4	EDK	87	95.4
0.03	0.4	KNEAR	101	100
0.03	0.6	DBAR	99	96.97
0.03	0.6	EDK	92	100
0.03	0.6	KNEAR	109	100

Conclusion

The properties of the estimator in finite sample estimator appear satisfactory. When the type of weighting scheme is known (gaussian, tri-cube, distance band or others distance based weight matrices), the minimizing algorithm allows a good approximation of the bandwidth, and easy to use for moderate sample size.

Conclusion

When the type of weighting scheme is unknown, the procedure proposed in our last MC experiments highlights very promising results.

For large datasets, the use of gradients (GNR) must be limited to cases where the spatial autocorrelation is high ($\lambda >= 0.3$) and with a convenient number of mean neighbors (h not to large).

Conclusion

Demonstrate the consistency of the estimator will be a priority of our future works.

References I

Aldstadt, J., Getis, A., 2006. Using amoeba to create a spatial weights matrix and identify spatial clusters. Geographical Analysis 38 (4), 327–343.

Anselin, L., 1980. Estimation methods for spatial autoregressive structures: A study in spatial econometrics. Program in Urban and Regional Studies, Cornell University.

Anselin, L., 1988. Spatial econometrics : Methods and Models. Wiley Online Library.

Bhattacharjee, A., Jensen-Butler, C., of St Andrews. Department of Economics, U., University of St Andrews. Centre for Research into Industry, Enterprise, F., the Firm (CRIEFF), 2006. Estimation of spatial weights matrix in a spatial error model, with an application to diffusion in housing demand. University of St Andrews. Department of Economics.

Conley, T., 1999. Gmm estimation with cross sectional dependence. Journal of econometrics 92 (1), 1–45.

Davidson, R., MacKinnon, J., Davidson, J., 1993. Estimation and inference in econometrics. Oxford University Press New York.

Harris, R., Moffat, J., Kravtsova, V., 2011. In search of ‘w’. Spatial Economic Analysis 6 (3), 249–270.

References II

LeSage, J., Fischer, M., 2008. Spatial growth regressions: Model specification, estimation and interpretation. Spatial Economic Analysis 3 (3), 275–304.

López, F., Matilla-García, M., Mur, J., Marín, M., 2010. A non-parametric spatial independence test using symbolic entropy. Regional Science and Urban Economics 40 (2-3), 106–115.

Meen, G., 1996. Spatial aggregation, spatial dependence and predictability in the uk housing market. Housing Studies 11 (3), 345–372.

Ord, J., Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geographical analysis 27 (4), 286–306.

Pinkse, J., Slade, M., Brett, C., 2002. Spatial price competition: a semiparametric approach. Econometrica 70 (3), 1111–1153.

Stakhovych, S., Bijmolt, T., 2009. Specification of spatial models: A simulation study on weights matrices. Papers in Regional Science 88 (2), 389–408.