Co-funded by the
Erasmus+ Programme
of the European Union

Project funded by
European Commission Erasmus + Programme –Jean Monnet Action
Project number 553280-EPP-1-2015-1-IT-EPPJMO-MODULE

The spatial nonparametric (SNP) estimator: Monte Carlo experiment

M. Gerolimetto S. Magrini

Department of Economics - University Ca' Foscari of Venice

Pisa, October 23, 2017

Monte Carlo experiment

For the following Data Generating Process:

$$Y = M(X) + u \qquad \text{with } u = \rho Wu + \epsilon$$

we consider a set of nonlinear functions:

A $M(X) = sin(5\pi X)$

B $M(X) = 2 + sin(7.1(X - 3.2))$

C $M(X) = 1 - 48X + 218X^2 - 315X^3 + 145x^4$

D $M(X) = 10exp(-10X)$

E $M(X) = (-1 + 2X) + 0.95exp(-40(-1 + 2X)^2)$

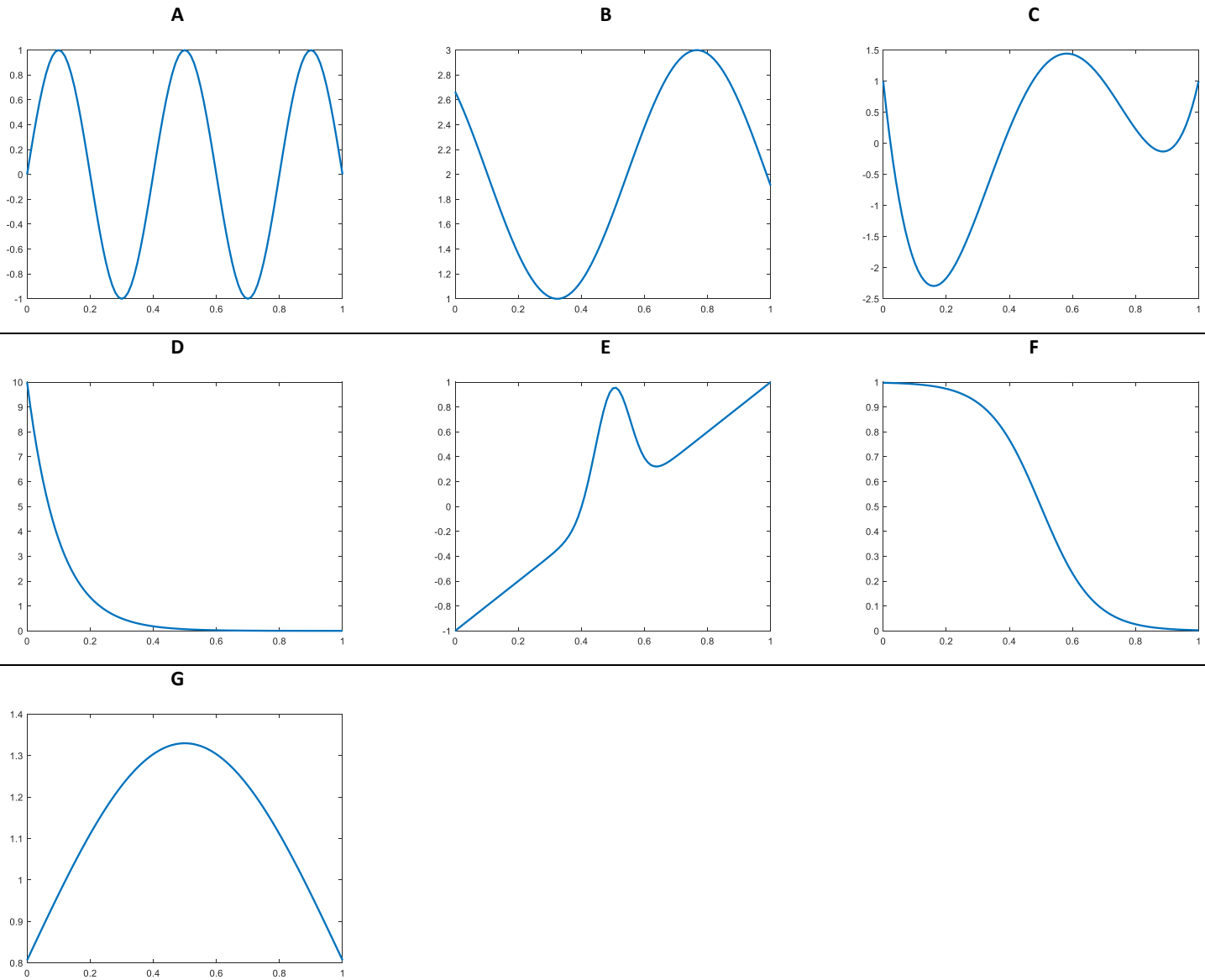
F $M(X) = 1/(1 + exp(-6 + 12X))$

G $M(x) = (0.3\sqrt{2\pi})^{-1} exp(-(X - 0.5)^2)$

where:

$X \sim U(0, 1)$
$\epsilon \sim N(0, \sigma^2)$, where σ is set to obtain pseudo-$R^2 = 0.3, 0.5, 0.7$
$\rho = 0.0, 0.2, 0.4, 0.6, 0.8$
two W matrices: 10% neighbors and contiguity from Voronoi tessellation
sample size (n) $= 50, 100, 200$
1000 Monte Carlo replications
two types of bandwidth: direct plug-in and cross-validation minimization

Monte Carlo functions



Monte Carlo results

Ratio (SNP over NP) of the median across replications of the MISE

W: 10% neighbors; bandwidth: cross-validation minimization; $\rho = 0, 0.4, 0.8$

		A			B			C		
pseudo-R^2	n	0	0.4	0.8	0	0.4	0.8	0	0.4	0.8
0.7	50	0.98	0.98	0.98	0.93	0.91	0.88	0.93	0.90	0.88
0.7	100	0.97	0.96	0.96	0.94	0.90	0.84	0.90	0.88	0.87
0.7	200	0.96	0.95	0.95	0.94	0.88	0.78	0.88	0.83	0.82
0.5	50	0.98	0.98	0.98	0.94	0.91	0.90	0.94	0.91	0.91
0.5	100	0.97	0.96	0.97	0.93	0.87	0.86	0.92	0.89	0.89
0.5	200	0.95	0.95	0.96	0.92	0.84	0.82	0.87	0.82	0.82
0.3	50	0.99	0.98	0.98	0.96	0.95	0.93	0.97	0.95	0.94
0.3	100	0.98	0.97	0.97	0.91	0.88	0.89	0.91	0.90	0.90
0.3	200	0.96	0.96	0.96	0.92	0.82	0.83	0.87	0.83	0.83

		D			E			F		
pseudo-R^2	n	0	0.4	0.8	0	0.4	0.8	0	0.4	0.8
0.7	50	0.89	0.85	0.90	0.96	0.94	0.94	0.98	0.98	0.96
0.7	100	0.81	0.78	0.86	0.94	0.93	0.95	0.97	0.94	0.92
0.7	200	0.80	0.72	0.78	0.92	0.89	0.92	0.94	0.89	0.85
0.5	50	0.92	0.88	0.95	0.99	0.99	0.97	1.00	0.98	0.96
0.5	100	0.84	0.79	0.88	0.95	0.94	0.95	0.98	0.98	0.98
0.5	200	0.79	0.72	0.81	0.91	0.90	0.93	0.97	0.95	0.93
0.3	50	0.95	0.96	0.97	1.00	1.01	0.96	1.01	1.00	0.91
0.3	100	0.88	0.84	0.91	0.99	0.98	0.98	0.99	1.00	1.00
0.3	200	0.83	0.75	0.83	0.95	0.93	0.95	0.99	0.99	0.99

		G		
pseudo-R^2	n	0	0.4	0.8
0.7	50	0.82	0.77	0.81
0.7	100	0.76	0.72	0.77
0.7	200	0.76	0.60	0.69
0.5	50	0.82	0.76	0.82
0.5	100	0.72	0.70	0.79
0.5	200	0.80	0.59	0.73
0.3	50	0.86	0.82	0.87
0.3	100	0.76	0.70	0.82
0.3	200	0.79	0.62	0.75

Monte Carlo results

Ratio (SNP over NP) of the median across replications of the MISE

W: contiguity from Voronoi tessellation; bandwidth: cross-validation minimization; $\rho = 0, 0.4, 0.8$

		A			B			C		
pseudo-R^2	n	0	0.4	0.8	0	0.4	0.8	0	0.4	0.8
0.7	50	0.98	0.97	0.97	0.93	0.92	0.88	0.93	0.91	0.88
0.7	100	0.97	0.96	0.96	0.94	0.89	0.82	0.90	0.87	0.87
0.7	200	0.96	0.95	0.95	0.94	0.78	0.77	0.88	0.81	0.81
0.5	50	0.98	0.98	0.98	0.94	0.91	0.90	0.94	0.91	0.90
0.5	100	0.97	0.96	0.96	0.93	0.85	0.84	0.92	0.89	0.90
0.5	200	0.95	0.95	0.95	0.92	0.76	0.79	0.87	0.81	0.82
0.3	50	0.99	0.98	0.98	0.96	0.95	0.92	0.97	0.95	0.94
0.3	100	0.98	0.97	0.97	0.91	0.86	0.89	0.91	0.88	0.90
0.3	200	0.96	0.95	0.96	0.92	0.76	0.81	0.87	0.81	0.83

		D			E			F		
pseudo-R^2	n	0	0.4	0.8	0	0.4	0.8	0	0.4	0.8
0.7	50	0.89	0.85	0.87	0.96	0.94	0.95	0.98	0.98	0.96
0.7	100	0.81	0.77	0.83	0.94	0.92	0.94	0.97	0.92	0.91
0.7	200	0.80	0.70	0.78	0.92	0.89	0.92	0.94	0.83	0.83
0.5	50	0.92	0.88	0.91	0.99	0.98	0.97	1.00	0.99	0.97
0.5	100	0.84	0.77	0.85	0.95	0.93	0.95	0.98	0.99	0.97
0.5	200	0.79	0.70	0.80	0.91	0.89	0.93	0.97	0.93	0.92
0.3	50	0.95	0.93	0.97	1.00	1.01	0.99	1.01	1.01	0.95
0.3	100	0.88	0.85	0.89	0.99	0.98	0.97	0.99	1.01	0.97
0.3	200	0.83	0.73	0.83	0.95	0.93	0.94	0.99	0.98	0.98

		G		
pseudo-R^2	n	0	0.4	0.8
0.7	50	0.82	0.78	0.79
0.7	100	0.76	0.69	0.75
0.7	200	0.76	0.55	0.68
0.5	50	0.82	0.76	0.80
0.5	100	0.72	0.68	0.79
0.5	200	0.80	0.55	0.72
0.3	50	0.86	0.82	0.83
0.3	100	0.76	0.70	0.80
0.3	200	0.79	0.56	0.73

Monte Carlo results

Ratio (SNP over NP) of the median across replications of the MISE

W: 10% neighbors; bandwidth: direct plug-in; $\rho = 0, 0.4, 0.8$

		A			B			C		
pseudo-R^2	n	0	0.4	0.8	0	0.4	0.8	0	0.4	0.8
0.7	50	0.74	0.68	0.64	0.59	0.50	0.47	0.68	0.58	0.56
0.7	100	0.46	0.39	0.36	0.51	0.42	0.39	0.58	0.51	0.49
0.7	200	0.34	0.31	0.29	0.41	0.35	0.33	0.47	0.42	0.42
0.5	50	0.77	0.70	0.66	0.71	0.59	0.55	0.74	0.66	0.61
0.5	100	0.53	0.46	0.41	0.58	0.50	0.46	0.64	0.58	0.54
0.5	200	0.41	0.36	0.33	0.48	0.41	0.38	0.55	0.50	0.49
0.3	50	0.85	0.76	0.71	0.83	0.70	0.64	0.85	0.77	0.68
0.3	100	0.63	0.54	0.50	0.66	0.59	0.56	0.74	0.68	0.63
0.3	200	0.48	0.43	0.39	0.57	0.50	0.46	0.64	0.59	0.57
		D			E			F		
pseudo-R^2	n	0	0.4	0.8	0	0.4	0.8	0	0.4	0.8
0.7	50	0.78	0.73	0.81	0.86	0.79	0.73	0.86	0.77	0.66
0.7	100	0.71	0.66	0.70	0.73	0.65	0.60	0.76	0.69	0.62
0.7	200	0.61	0.56	0.62	0.57	0.50	0.47	0.63	0.56	0.53
0.5	50	0.87	0.82	0.87	0.93	0.89	0.83	0.98	0.93	0.79
0.5	100	0.79	0.75	0.79	0.83	0.78	0.71	0.90	0.86	0.77
0.5	200	0.70	0.65	0.70	0.69	0.63	0.58	0.75	0.70	0.65
0.3	50	0.95	0.93	0.93	1.00	1.00	0.91	1.11	1.10	0.87
0.3	100	0.88	0.86	0.89	0.95	0.92	0.87	1.04	1.05	0.98
0.3	200	0.81	0.76	0.82	0.86	0.79	0.73	0.91	0.89	0.84
		G								
pseudo-R^2	n	0	0.4	0.8						
0.7	50	0.59	0.56	0.70						
0.7	100	0.51	0.48	0.61						
0.7	200	0.43	0.40	0.48						
0.5	50	0.69	0.65	0.77						
0.5	100	0.62	0.58	0.70						
0.5	200	0.51	0.47	0.54						
0.3	50	0.81	0.79	0.84						
0.3	100	0.75	0.72	0.82						
0.3	200	0.63	0.56	0.64						

Monte Carlo results

Ratio (SNP over NP) of the median across replications of the MISE

W: contiguity from Voronoi tessellation; bandwidth: direct plug-in; $\rho = 0, 0.4, 0.8$

		A			B			C		
pseudo-R^2	n	0	0.4	0.8	0	0.4	0.8	0	0.4	0.8
0.7	50	0.74	0.69	0.64	0.59	0.51	0.47	0.68	0.59	0.56
0.7	100	0.46	0.38	0.36	0.51	0.39	0.37	0.58	0.48	0.47
0.7	200	0.34	0.30	0.28	0.41	0.33	0.33	0.47	0.40	0.41
0.5	50	0.77	0.70	0.66	0.71	0.59	0.53	0.74	0.68	0.63
0.5	100	0.53	0.44	0.41	0.58	0.47	0.45	0.64	0.55	0.53
0.5	200	0.41	0.35	0.32	0.48	0.38	0.36	0.55	0.46	0.45
0.3	50	0.85	0.76	0.73	0.83	0.72	0.64	0.85	0.76	0.71
0.3	100	0.63	0.52	0.49	0.66	0.58	0.54	0.74	0.67	0.61
0.3	200	0.48	0.41	0.36	0.57	0.47	0.44	0.64	0.55	0.53
		D			E			F		
pseudo-R^2	n	0	0.4	0.8	0	0.4	0.8	0	0.4	0.8
0.7	50	0.78	0.73	0.77	0.86	0.79	0.74	0.86	0.75	0.68
0.7	100	0.71	0.63	0.67	0.73	0.63	0.59	0.76	0.65	0.61
0.7	200	0.61	0.52	0.58	0.57	0.49	0.45	0.63	0.54	0.48
0.5	50	0.87	0.83	0.85	0.93	0.90	0.84	0.98	0.92	0.80
0.5	100	0.79	0.73	0.76	0.83	0.75	0.71	0.90	0.81	0.73
0.5	200	0.70	0.61	0.67	0.69	0.60	0.55	0.75	0.67	0.61
0.3	50	0.95	0.94	0.92	1.00	1.01	0.91	1.11	1.15	0.95
0.3	100	0.88	0.84	0.86	0.95	0.91	0.86	1.04	1.04	0.88
0.3	200	0.81	0.73	0.76	0.86	0.75	0.68	0.91	0.86	0.77
		G								
pseudo-R^2	n	0	0.4	0.8						
0.7	50	0.59	0.56	0.64						
0.7	100	0.51	0.47	0.57						
0.7	200	0.43	0.38	0.47						
0.5	50	0.69	0.64	0.72						
0.5	100	0.62	0.56	0.67						
0.5	200	0.51	0.45	0.53						
0.3	50	0.81	0.78	0.82						
0.3	100	0.75	0.69	0.78						
0.3	200	0.63	0.54	0.60						

Conclusions on the Monte Carlo experiment

Results show

- SNP **outperforms** polynomial regression (NP)
- this is confirmed:
  - for various functional forms
  - for all considered ρ values
  - for all considered sample sizes
  - for all considered **pseudo**-R^2 values
  - for both spatial weight matrices in the DGP

Hence

- SNP is a valuable tool for nonparametric regression when data are spatially dependent
- SNP can be used to estimate the mean function within Hyndman's *mean-bias* adjustment thus improving the properties of the conditional density estimator