

Analisi di regressione per dati spaziali

E.Chiola & P.Iezzi

15/7/2020

Analisi dataset: Boston

Nella seguente analisi il dataset in esame è denominato 'Boston' e raccoglie osservazioni inerenti le case e i vicinati nella città di Boston. Il documento mostrerà l'analisi del valore delle case (MEDV) in relazione alle altre variabili. Il dataset contiene 506 osservazioni per 23 variabili:

ID: Identità sequenziale

TOWN: Un fattore con i livelli suddivisi per città (in questo caso solo Boston, costante)

TOWNNO: Un vettore numerico corrispondente a TOWN

TRACT: Un vettore numerico di ID

LON: Longitudine

LAT: Latitudine

X: X Coordinate (UTM Zone 19)

Y: Y Coordinate (UTM Zone 19)

MEDV: Un vettore numerico del valore mediano delle case di proprietà in migliaia di dollari

CMEDV: Un vettore numerico del valore mediano corretto delle case di proprietà in migliaia di dollari

CRIM: Un vettore numerico dei crimini pro-capite

ZN: Un vettore numerico di proporzioni di terreni residenziali suddivisi in zone per lotti superiori a 25000 piedi quadrati per città (costante per tutti i tratti di Boston)

INDUS: Un vettore numerico di proporzioni di acri non al dettaglio per città (costante per tutti i tratti di Boston)

CHAS: Un fattoriale con livello 1 se l'area confina con il fiume Charles, 0 altrimenti

NOX: Un vettore numerico della concentrazione di ossidi nitrici (parti per 10 milioni)

RM: Un vettore numerico della media di stanze per abitazione

AGE: Un vettore numerico della proporzione di abitazioni di proprietà occupate costruire prima del 1940

DIS: Un vettore numerico delle distanze pesate dai 5 centri per l'impiego di Boston

RAD: Un vettore numerico indice di accessibilità alle autostrade

TAX: Un vettore numerico sull'entità delle tasse di proprietà per 10,000 dollari

PTRATIO: Un vettore numerico del rapporto alunno-insegnante

B: Un vettore numerico ottenuto dall'equazione $1000 \cdot (B_k - 0.63)^2$ con B_k indicante la popolazione afroamericana

LSTAT: Un vettore numerico del valore percentuale della popolazione in uno stato di povertà

Importo le librerie

```
#pulisco l'ambiente di lavoro
rm(list=ls())
#importo le librerie
library(spdep)
library(spatialreg)
library(tseries)
library(lmtest)
```

Importo il dataset

```
library(readr)
boston <- read_delim("C:/Users/Patrizio/Desktop/TUTTO/Ud'A/CLEBA/SUPERATE/DATA SCIENCE/3 progetto - Data Science/1. Importazione dati/boston.csv",
  ";", escape_double = FALSE, locale = locale(decimal_mark = ","),
  trim_ws = TRUE)
attach(boston)
options(scipen = 10) #annotazione numerica
```

```
#creazione matrice di contiguità
coord1<-as.matrix(boston[,7:8])
k10<-knearneigh(coord1, k=10) #nearest neighbour
k10_nb<-knn2nb(k10)
nn10w<-nb2listw(k10_nb,style="W") #matrice standardizzata per riga
nn10w_binary<-nb2listw(k10_nb, style = "B") #matrice dei pesi spaziali in formato listw di tipo binario
```

Test autocorrelazione

```
#indice di Moran
moran.test(MEDV, nn10w, zero.policy = TRUE, alternative = "greater")
```

```
##
##  Moran I test under randomisation
##
## data:  MEDV
## weights: nn10w
##
## Moran I statistic standard deviate = 28.872, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##      0.5307432582      -0.0019801980      0.0003404506
```

```
#indice di Geary
geary.test(MEDV, nn10w, zero.policy = TRUE, alternative = "greater")
```

```
##
```

```
## Geary C test under randomisation
##
## data: MEDV
## weights: nn10w
##
## Geary C statistic standard deviate = 25.491, p-value < 2.2e-16
## alternative hypothesis: Expectation greater than statistic
## sample estimates:
## Geary C statistic      Expectation      Variance
##      0.4325823459      1.0000000000      0.0004954871
```

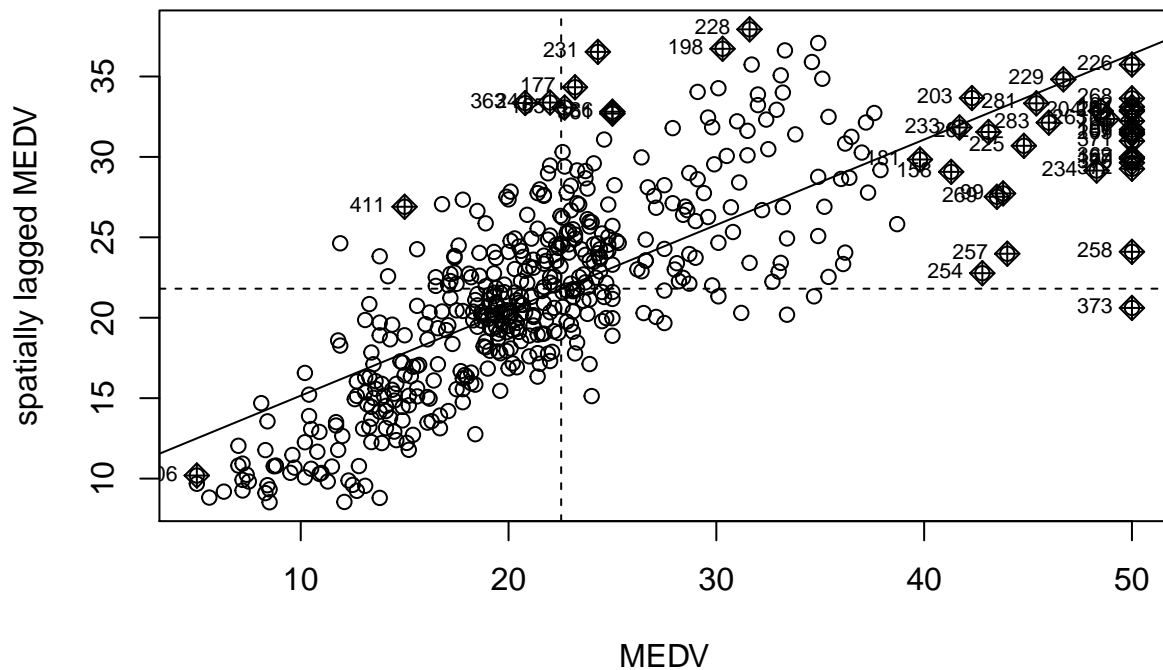
Autocorrelazione positiva perchè l'indice Moran è maggiore del valore atteso e, nel caso di Geary, il valore è inferiore rispetto alla suo valore atteso.

```
#indice di Getis-Ord
globalG.test(MEDV,nn10w_binary, zero.policy = TRUE,alternative = "greater") #i pesi sono binari
```

```
##
## Getis-Ord global G statistic
##
## data: MEDV
## weights: nn10w_binary
##
## standard deviate = 9.0069, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Global G statistic      Expectation      Variance
##      0.02091937745505      0.01980198019802      0.00000001539097
```

Possiamo notare che l'indice G è maggiore del suo valore atteso, quindi vi sono potenziali HOT SPOTS.

```
# Moran ScatterPlot
moran.plot(MEDV, nn10w, labels = NULL, xlab = NULL, ylab = NULL)
```



Utilizziamo il Moran scatterplot per raffigurare graficamente le osservazioni. Notiamo che i punti sono concentrati nei quadranti high-high e low-low, questo a conferma di una autocorrelazione positiva.

Applicazione modelli

```
#modello di regressione lineare multipla
reglin<-lm (log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT))
summary(reglin)
```

```
##
## Call:
## lm(formula = log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
##     AGE + TAX + PTRATIO + B + log(LSTAT))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.73967 -0.10856 -0.01359  0.10703  0.87777
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)  4.71511214  0.11208691  42.067    < 2e-16 ***
## CRIM        -0.00922400  0.00129098  -7.145 0.000000000000324 ***
## ZN          -0.00115613  0.00051187  -2.259   0.024340 *
## INDUS       -0.00040308  0.00236345  -0.171   0.864650
## CHAS         0.11192485  0.03571269   3.134   0.001827 **
```

```
## I(NOX^2)    -0.38774916  0.11416248  -3.396      0.000738 ***
## AGE         0.00203300  0.00049393   4.116 0.00004516155326 ***
## TAX         0.00007861  0.00009507   0.827      0.408677
## PTRATIO    -0.03619964  0.00534170  -6.777 0.00000000003502 ***
## B           0.00029330  0.00011053   2.654      0.008218 **
## log(LSTAT)  -0.47276782  0.02100059 -22.512      < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1973 on 495 degrees of freedom
## Multiple R-squared:  0.7716, Adjusted R-squared:  0.767
## F-statistic: 167.2 on 10 and 495 DF,  p-value: < 2.2e-16
```

```
AIC(reglin)
```

```
## [1] -193.5371
```

Le variabili sono tutte significative eccetto TAX e INDUS, un ottimo R-squared e un ANOVA eccellente.

```
#Jarque Bera test per analizzare la normalità degli errori
jarque.bera.test(reglin$residuals)
```

```
##
## Jarque Bera Test
##
## data: reglin$residuals
## X-squared = 119.58, df = 2, p-value < 2.2e-16
```

Rifiutiamo l'ipotesi H0 quindi gli errori non si distribuiscono secondo la normale.

```
#Breusch-Pagan
bptest(reglin,studentize = TRUE)
```

```
##
## studentized Breusch-Pagan test
##
## data: reglin
## BP = 35.333, df = 10, p-value = 0.0001095
```

```
bptest(reglin,studentize = FALSE)
```

```
##
## Breusch-Pagan test
##
## data: reglin
## BP = 76.724, df = 10, p-value = 0.00000000002195
```

Osserviamo che siamo in presenza di eteroschedasticità

```
#verifichiamo l'ipotesi di autocorrelazione dei residui con il test di Moran e con il Lagrange Multiplier
lm.morantest(reglin,nn10w,alternative = "two.sided")
```

```
##
## Global Moran I for regression residuals
##
## data:
## model: lm(formula = log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
## AGE + TAX + PTRATIO + B + log(LSTAT))
## weights: nn10w
##
## Moran I statistic standard deviate = 20.725, p-value < 2.2e-16
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##      0.3505668868      -0.0116744583      0.0003054883
```

```
lm.LMtests(reglin, nn10w, test=c("LMerr", "LMlag", "RLMerr","RLMlag"))
```

```
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
## AGE + TAX + PTRATIO + B + log(LSTAT))
## weights: nn10w
##
## LMerr = 351.5, df = 1, p-value < 2.2e-16
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
## AGE + TAX + PTRATIO + B + log(LSTAT))
## weights: nn10w
##
## LMlag = 188.44, df = 1, p-value < 2.2e-16
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
## AGE + TAX + PTRATIO + B + log(LSTAT))
## weights: nn10w
##
## RLMerr = 183.4, df = 1, p-value < 2.2e-16
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
```

```
## data:
## model: lm(formula = log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
## AGE + TAX + PTRATIO + B + log(LSTAT))
## weights: nn10w
##
## RLMlag = 20.343, df = 1, p-value = 0.000006473
```

Controllando i due test LMerr e LMLag non è possibile scegliere quello più adatto, per cui si procede con l'analisi delle versioni robuste di questi due test e si conclude che RLMerr ha un p-value più basso per cui il modello Spatial Error (SEM) è più efficiente.

Stima dei modelli

```
#Spatial Autoregressive Model (SAM)
Pure<-spautolm(log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT), da
summary(Pure)
```

```
##
## Call: spautolm(formula = log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
## AGE + TAX + PTRATIO + B + log(LSTAT), data = boston, listw = nn10w)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6482970 -0.0848290 -0.0038748  0.0910343  0.7990140
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.214294821  0.129746485  32.4810 < 2.2e-16
## CRIM        -0.006678113  0.001082217  -6.1708 0.0000000006796
## ZN          -0.000145861  0.000498399  -0.2927  0.769782
## INDUS       -0.002622075  0.002808574  -0.9336  0.350512
## CHAS         0.022860345  0.033514865   0.6821  0.495178
## I(NOX^2)    -0.203211071  0.157727720  -1.2884  0.197619
## AGE          0.000061464  0.000515788   0.1192  0.905144
## TAX         -0.000160226  0.000110592  -1.4488  0.147394
## PTRATIO     -0.015426266  0.005719737  -2.6970  0.006996
## B           0.000561163  0.000118710   4.7272 0.0000022767570
## log(LSTAT)  -0.387160535  0.019612855 -19.7401 < 2.2e-16
##
## Lambda: 0.8307 LR test value: 216.61 p-value: < 2.22e-16
## Numerical Hessian standard error of lambda: 0.035154
##
## Log likelihood: 217.0725
## ML residual variance (sigma squared): 0.022573, (sigma: 0.15024)
## Number of observations: 506
## Number of parameters estimated: 13
## AIC: -408.14
```

Il modello autoregressivo spaziale è di difficile interpretazione. Notiamo che il valore di lambda (rho sulle slide) è elevato, questo significa che c'è un'elevata autocorrelazione.

#Spatial Lag Model (SLM)

```
Spatial_lag<-lagsarlm(log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT), data = boston, listw = nn10w, summary(Spatial_lag)
```

```
##
## Call:lagsarlm(formula = log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
##     AGE + TAX + PTRATIO + B + log(LSTAT), data = boston, listw = nn10w,
##     Durbin = FALSE, tol.solve = 1e-14)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.61410 -0.10647 -0.01531  0.10099  0.86018
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.446137962  0.189159233  12.9316 < 2.2e-16
## CRIM        -0.005939408  0.001114221  -5.3305 0.00000009792
## ZN          -0.000592151  0.000432934  -1.3678  0.171386
## INDUS       -0.000612468  0.001997768  -0.3066  0.759166
## CHAS         0.021611773  0.030659180   0.7049  0.480870
## I(NOX^2)    -0.074182206  0.097673884  -0.7595  0.447560
## AGE         0.002027660  0.000418473   4.8454 0.00000126369
## TAX         0.000094229  0.000080841   1.1656  0.243775
## PTRATIO     -0.013678662  0.004733990  -2.8895  0.003859
## B           0.000227754  0.000093937   2.4246  0.015327
## log(LSTAT)  -0.375324010  0.019255814 -19.4915 < 2.2e-16
##
## Rho: 0.50898, LR test value: 146.01, p-value: < 2.22e-16
## Asymptotic standard error: 0.03739
##      z-value: 13.613, p-value: < 2.22e-16
## Wald statistic: 185.3, p-value: < 2.22e-16
##
## Log likelihood: 181.7716 for lag model
## ML residual variance (sigma squared): 0.027812, (sigma: 0.16677)
## Number of observations: 506
## Number of parameters estimated: 13
## AIC: -337.54, (AIC for lm: -193.54)
## LM test for residual autocorrelation
## test value: 67.119, p-value: 2.2204e-16
```

Il modello spazialmente ritardato prevede un'influenza sulla variabile dipendente per merito delle regioni vicini. Osserviamo un rho maggiore di zero, quindi vi è dipendenza spaziale positiva.

#Spatial Error Model (SEM)

```
Spatial_Error<-errorsarlm((log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT)), data = boston, listw = nn10w, Durbin = FALSE, zero.policy = TRUE, tol.solve = 1e-14) summary(Spatial_Error)
```

```
##
## Call:errorsarlm(formula = (log(MEDV) ~ CRIM + ZN + INDUS + CHAS +
##     I(NOX^2) + AGE + TAX + PTRATIO + B + log(LSTAT)), data = boston,
##     listw = nn10w, Durbin = FALSE, zero.policy = TRUE, tol.solve = 1e-14)
```



```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6482970 -0.0848290 -0.0038748  0.0910343  0.7990140
##
## Type: error
## Coefficients: (asymptotic standard errors)
##              Estimate Std. Error z value      Pr(>|z|)
## (Intercept)  4.214294826  0.129746484  32.4810    < 2.2e-16
## CRIM        -0.006678113  0.001082217  -6.1708 0.0000000006796
## ZN          -0.000145861  0.000498399  -0.2927    0.769782
## INDUS       -0.002622075  0.002808574  -0.9336    0.350512
## CHAS         0.022860346  0.033514865   0.6821    0.495178
## I(NOX^2)     -0.203211069  0.157727719  -1.2884    0.197619
## AGE          0.000061464  0.000515788   0.1192    0.905144
## TAX         -0.000160226  0.000110592  -1.4488    0.147394
## PTRATIO     -0.015426266  0.005719737  -2.6970    0.006996
## B            0.000561163  0.000118710   4.7272 0.0000022767573
## log(LSTAT)  -0.387160536  0.019612855 -19.7401    < 2.2e-16
##
## Lambda: 0.8307, LR test value: 216.61, p-value: < 2.22e-16
## Asymptotic standard error: 0.032152
##      z-value: 25.837, p-value: < 2.22e-16
## Wald statistic: 667.54, p-value: < 2.22e-16
##
## Log likelihood: 217.0725 for error model
## ML residual variance (sigma squared): 0.022573, (sigma: 0.15024)
## Number of observations: 506
## Number of parameters estimated: 13
## AIC: -408.14, (AIC for lm: -193.54)
```

In questo modello l'influenza tra le regioni si manifesta come termine di errore poichè è un parametro di disturbo.

#Spatial Durbin Model (SDM)

```
Spatial_Durbin<- lagsarlm((log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT)), data = boston, listw = nn10w,
summary(Spatial_Durbin)
```

```
##
## Call:lagsarlm(formula = (log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
##      AGE + TAX + PTRATIO + B + log(LSTAT)), data = boston, listw = nn10w,
##      Durbin = TRUE, zero.policy = TRUE, tol.solve = 1e-14)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6369456 -0.0838804 -0.0046662  0.0878169  0.8484354
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
##              Estimate Std. Error z value      Pr(>|z|)
## (Intercept)  1.597645181  0.268751637   5.9447 0.0000000027698
## CRIM        -0.006578673  0.001074536  -6.1223 0.0000000009221
## ZN           0.000096557  0.000525740   0.1837    0.8542807
```

```
## INDUS          -0.003384253  0.002920499  -1.1588      0.2465408
## CHAS           0.006316430  0.033650365   0.1877      0.8511058
## I(NOX^2)       -0.221132342  0.170299149  -1.2985      0.1941178
## AGE           -0.000526203  0.000535805  -0.9821      0.3260603
## TAX           -0.000186146  0.000115204  -1.6158      0.1061377
## PTRATIO       -0.011434510  0.005870899  -1.9477      0.0514558
## B             0.000619839  0.000119468   5.1883 0.0000002122020
## log(LSTAT)     -0.366916057  0.019852347 -18.4823      < 2.2e-16
## lag.CRIM       -0.000055264  0.002472417  -0.0224      0.9821669
## lag.ZN         -0.000113601  0.000995942  -0.1141      0.9091875
## lag.INDUS      0.002455182  0.004130507   0.5944      0.5522432
## lag.CHAS       0.064981737  0.060933629   1.0664      0.2862272
## lag.I(NOX^2)   -0.136841125  0.217961005  -0.6278      0.5301193
## lag.AGE        0.003073073  0.000892588   3.4429      0.0005756
## lag.TAX        0.000438199  0.000183850   2.3835      0.0171509
## lag.PTRATIO    -0.008896632  0.009904610  -0.8982      0.3690621
## lag.B          -0.000701724  0.000174814  -4.0141 0.0000596710871
## lag.log(LSTAT) 0.182430765  0.044916732   4.0615 0.0000487516362
##
## Rho: 0.70646, LR test value: 141.44, p-value: < 2.22e-16
## Asymptotic standard error: 0.044945
## z-value: 15.718, p-value: < 2.22e-16
## Wald statistic: 247.07, p-value: < 2.22e-16
##
## Log likelihood: 233.5391 for mixed model
## ML residual variance (sigma squared): 0.021927, (sigma: 0.14808)
## Number of observations: 506
## Number of parameters estimated: 23
## AIC: -421.08, (AIC for lm: -281.64)
## LM test for residual autocorrelation
## test value: 0.32278, p-value: 0.56994
```

```
#SARAR
```

```
SARAR<- saccsarlm((log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT))
summary(SARAR)
```

```
##
## Call:saccsarlm(formula = (log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
## AGE + TAX + PTRATIO + B + log(LSTAT)), data = boston, listw = nn10w,
## Durbin = FALSE, zero.policy = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6549744 -0.0890314 -0.0045025  0.0897357  0.8138451
##
## Type: sac
## Coefficients: (asymptotic standard errors)
##              Estimate Std. Error z value      Pr(>|z|)
## (Intercept)  3.74487685  0.30026716  12.4718      < 2.2e-16
## CRIM        -0.00655239  0.00108794  -6.0227 0.000000001715
## ZN          -0.00020941  0.00049634  -0.4219    0.673099
## INDUS       -0.00213869  0.00277520  -0.7706    0.440918
## CHAS         0.02234975  0.03352240   0.6667    0.504957
## I(NOX^2)    -0.13493066  0.15673774  -0.8609    0.389310
```

```
## AGE          0.00024226  0.00051343  0.4719      0.637034
## TAX          -0.00015293  0.00010889 -1.4045      0.160178
## PTRATIO      -0.01500651  0.00569956 -2.6329      0.008465
## B            0.00053747  0.00011811  4.5504 0.000005354751
## log(LSTAT)   -0.38843172  0.01957104 -19.8473     < 2.2e-16
##
## Rho: 0.14234
## Asymptotic standard error: 0.079723
##      z-value: 1.7855, p-value: 0.074187
## Lambda: 0.78149
## Asymptotic standard error: 0.048331
##      z-value: 16.169, p-value: < 2.22e-16
##
## LR test value: 219.49, p-value: < 2.22e-16
##
## Log likelihood: 218.5149 for sac model
## ML residual variance (sigma squared): 0.022779, (sigma: 0.15093)
## Number of observations: 506
## Number of parameters estimated: 14
## AIC: -409.03, (AIC for lm: -193.54)
```

Il modello contempla gli effetti spaziali propri del modello spazialmente ritardato (SLM), dove la var. dip. è influenzata dalle unità vicine, ed anche effetti previsti dal modello con errori spazialmente autocorrelati (SEM), dove l'influenza è nei residui. Da un confronto dell'AIC con quello del Durbin si evince che la spiegazione dell'autocorrelazione dei residui non è così rilevante da migliorare la regressione.

#Spatial Durbin Error Model (SDEM)

```
SDEM<- errorsarlm((log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT)),
summary(SDEM)
```

```
##
## Call:errorsarlm(formula = (log(MEDV) ~ CRIM + ZN + INDUS + CHAS +
##      I(NOX^2) + AGE + TAX + PTRATIO + B + log(LSTAT)), data = boston,
##      listw = nn10w, Durbin = TRUE, zero.policy = TRUE, tol.solve = 1e-14)
##
## Residuals:
##      Min        1Q      Median        3Q       Max
## -0.6668435 -0.0853004 -0.0072786  0.0841756  0.8195714
##
## Type: error
## Coefficients: (asymptotic standard errors)
##      Estimate      Std. Error  z value      Pr(>|z|)
## (Intercept)   4.8250534498  0.3069130138  15.7212     < 2.2e-16
## CRIM          -0.0069401029  0.0010978317  -6.3216 0.0000000002588
## ZN            -0.0000032426  0.0005091528  -0.0064   0.9949185
## INDUS         -0.0035981011  0.0028987496  -1.2413   0.2145098
## CHAS          0.0160805795  0.0332428878   0.4837   0.6285776
## I(NOX^2)      -0.1603414528  0.1676131785  -0.9566   0.3387611
## AGE           -0.0001814994  0.0005251267  -0.3456   0.7296211
## TAX           -0.0001590160  0.0001166612  -1.3631   0.1728643
## PTRATIO       -0.0127144978  0.0058134508  -2.1871   0.0287365
## B             0.0005706147  0.0001179158   4.8392 0.0000013038332
## log(LSTAT)    -0.3755592923  0.0196579099 -19.1047     < 2.2e-16
```

```
## lag.CRIM      -0.0083373875  0.0042485270  -1.9624      0.0497138
## lag.ZN        -0.0003901015  0.0013846663  -0.2817      0.7781508
## lag.INDUS     0.0046498729  0.0064877896   0.7167      0.4735522
## lag.CHAS      0.0169747594  0.0999682226   0.1698      0.8651662
## lag.I(NOX^2)  -0.4900572971  0.3279720329  -1.4942      0.1351222
## lag.AGE       0.0048852752  0.0014093004   3.4665      0.0005274
## lag.TAX       0.0003668200  0.0003003446   1.2213      0.2219610
## lag.PTRATIO   -0.0272404521  0.0151347897  -1.7999      0.0718833
## lag.B         -0.0002985301  0.0003231393  -0.9238      0.3555678
## lag.log(LSTAT) -0.1738796183  0.0634356350  -2.7410      0.0061245
##
## Lambda: 0.75573, LR test value: 131.26, p-value: < 2.22e-16
## Asymptotic standard error: 0.041909
##      z-value: 18.033, p-value: < 2.22e-16
## Wald statistic: 325.18, p-value: < 2.22e-16
##
## Log likelihood: 228.4514 for error model
## ML residual variance (sigma squared): 0.0221, (sigma: 0.14866)
## Number of observations: 506
## Number of parameters estimated: 23
## AIC: -410.9, (AIC for lm: -281.64)
```

#Spatial Lag X (SLX)

```
SLX<-lmSLX(log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT), data=boston, weights=weights)
summary(SLX)
```

```
##
## Call:
## lm(formula = formula(paste("y ~ ", paste(colnames(x)[-1], collapse = "+"))),
##     data = as.data.frame(x), weights = weights)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.65568 -0.10116 -0.00742  0.09663  0.96199
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.200532352  0.164430450  31.628 < 2e-16 ***
## CRIM          -0.007241708  0.001299749  -5.572 0.0000000419 ***
## ZN             0.000009096  0.000636058   0.014  0.988597
## INDUS        -0.004500541  0.003532933  -1.274  0.203315
## CHAS          -0.011193748  0.040709129  -0.275  0.783457
## I.NOX.2.      -0.302938235  0.206033225  -1.470  0.142119
## AGE           -0.000729771  0.000648185  -1.126  0.260778
## TAX           -0.000212442  0.000139329  -1.525  0.127973
## PTRATIO       -0.010663120  0.007101717  -1.501  0.133881
## B              0.000727854  0.000144524   5.036 0.0000006708 ***
## log.LSTAT.    -0.360935703  0.024008949 -15.033 < 2e-16 ***
## lag.CRIM      -0.010181435  0.002863087  -3.556  0.000413 ***
## lag.ZN        -0.001561351  0.001201585  -1.299  0.194421
## lag.INDUS     0.002569794  0.004997105   0.514  0.607307
## lag.CHAS      0.281350966  0.072860270   3.862  0.000128 ***
## lag.I.NOX.2.  -0.584597952  0.259439460  -2.253  0.024685 *
## lag.AGE       0.005839638  0.001046463   5.580 0.0000000400 ***
```

```
## lag.TAX          0.000844650  0.000220531  3.830      0.000145 ***
## lag.PTRATIO      -0.046692925  0.011596419 -4.026 0.0000656874 ***
## lag.B            -0.000794293  0.000211413 -3.757      0.000193 ***
## lag.log.LSTAT.   -0.199849908  0.045373895 -4.405 0.0000130564 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1792 on 485 degrees of freedom
## Multiple R-squared:  0.8155, Adjusted R-squared:  0.8079
## F-statistic: 107.2 on 20 and 485 DF,  p-value: < 2.2e-16
```

#General Nesting Model (GNS)

```
GNS<-sacsarlm(log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT), data=boston)
summary(GNS)
```

```
##
## Call:sacsarlm(formula = log(MEDV) ~ CRIM + ZN + INDUS + CHAS + I(NOX^2) +
##      AGE + TAX + PTRATIO + B + log(LSTAT), data = boston, listw = nn10w,
##      Durbin = TRUE)
##
## Residuals:
##      Min        1Q      Median        3Q       Max
## -0.6333673 -0.0824864 -0.0043353  0.0867222  0.8455880
##
## Type: sacmixed
## Coefficients: (asymptotic standard errors)
##      Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.382267841  0.394595247  3.5030  0.0004600
## CRIM         -0.006631017  0.001077587 -6.1536 0.0000000007575
## ZN           0.000111872  0.000526134  0.2126  0.8316153
## INDUS        -0.003301055  0.002900499 -1.1381  0.2550792
## CHAS          0.004258021  0.033628666  0.1266  0.8992421
## I(NOX^2)     -0.228671938  0.169584329 -1.3484  0.1775214
## AGE          -0.000537845  0.000535415 -1.0045  0.3151187
## TAX          -0.000187497  0.000114045 -1.6441  0.1001640
## PTRATIO      -0.011633421  0.005847485 -1.9895  0.0466489
## B             0.000629109  0.000119196  5.2779 0.0000001306435
## log(LSTAT)   -0.367428581  0.019846156 -18.5138 < 2.2e-16
## lag.CRIM      0.000648042  0.002532155  0.2559  0.7980086
## lag.ZN        -0.000076911  0.000968422 -0.0794  0.9366994
## lag.INDUS     0.002366105  0.003983107  0.5940  0.5524887
## lag.CHAS      0.060414707  0.059631842  1.0131  0.3109989
## lag.I(NOX^2) -0.096585150  0.219109136 -0.4408  0.6593517
## lag.AGE       0.002878215  0.000912848  3.1530  0.0016160
## lag.TAX       0.000422392  0.000179146  2.3578  0.0183833
## lag.PTRATIO   -0.006598915  0.010140429 -0.6508  0.5152059
## lag.B         -0.000715222  0.000167434 -4.2717 0.0000194029808
## lag.log(LSTAT) 0.207208529  0.053742853  3.8556  0.0001155
##
## Rho: 0.74888
## Asymptotic standard error: 0.071918
##      z-value: 10.413, p-value: < 2.22e-16
## Lambda: -0.12525
## Asymptotic standard error: 0.20215
```

```
##      z-value: -0.61959, p-value: 0.53553
##
## LR test value: 249.88, p-value: < 2.22e-16
##
## Log likelihood: 233.7093 for sacmixed model
## ML residual variance (sigma squared): 0.02166, (sigma: 0.14717)
## Number of observations: 506
## Number of parameters estimated: 24
## AIC: -419.42, (AIC for lm: -193.54)
```

Include tutti i tipi di effetti di interazione.

Confronto modelli con AIC

```
comp<-cbind(AIC(reglin),AIC(Pure),AIC(Spatial_lag),AIC(Spatial_Error),AIC(Spatial_Durbin),AIC(SARAR),AIC(SDEM),AIC(SLX),AIC(GNS))
colnames(comp)<-c("reglin","Pure","Spatial Lag","Spatial Error","Spatial Durbin","SARAR","SDEM","SLX","GNS")
round(comp,7)
```

```
##      reglin      Pure Spatial Lag Spatial Error Spatial Durbin      SARAR
## [1,] -193.5371 -408.1449  -337.5431    -408.1449    -421.0782 -409.0297
##      SDEM      SLX      GNS
## [1,] -410.9029 -281.6421 -419.4185
```

Il modello con AIC più basso è lo Spatial Durbin (SDM)

Raffinamento modello

```
#Spatial Durbin sena variabile ZN
Spatial_Durbin_NO_ZN<-lagsarlm((log(MEDV) ~ CRIM + INDUS + CHAS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT)), data = boston, listw = nn10w,
summary(Spatial_Durbin_NO_ZN)
```

```
##
## Call:lagsarlm(formula = (log(MEDV) ~ CRIM + INDUS + CHAS + I(NOX^2) +
##      AGE + TAX + PTRATIO + B + log(LSTAT)), data = boston, listw = nn10w,
##      Durbin = TRUE, zero.policy = TRUE, tol.solve = 1e-14)
##
## Residuals:
##      Min      1Q    Median      3Q      Max
## -0.6369697 -0.0829474 -0.0044358  0.0889404  0.8479568
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
##      Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.597791846  0.251291379  6.3583 0.0000000002040
## CRIM        -0.006581074  0.001074471 -6.1249 0.0000000009072
## INDUS       -0.003486394  0.002866654 -1.2162  0.2239129
## CHAS         0.006350815  0.033613361  0.1889  0.8501420
## I(NOX^2)    -0.223785018  0.169701570 -1.3187  0.1872703
## AGE        -0.000530757  0.000535066 -0.9919  0.3212228
```

```
## TAX -0.000183787 0.000114455 -1.6058 0.1083259
## PTRATIO -0.011625094 0.005778127 -2.0119 0.0442290
## B 0.000619431 0.000119450 5.1857 0.0000002152074
## log(LSTAT) -0.367438106 0.019646499 -18.7025 < 2.2e-16
## lag.CRIM -0.000042641 0.002470936 -0.0173 0.9862316
## lag.INDUS 0.002569596 0.004061230 0.6327 0.5269206
## lag.CHAS 0.065400974 0.060238362 1.0857 0.2776104
## lag.I(NOX^2) -0.133016717 0.216260325 -0.6151 0.5385040
## lag.AGE 0.003065461 0.000858103 3.5724 0.0003538
## lag.TAX 0.000434825 0.000180446 2.4097 0.0159645
## lag.PTRATIO -0.008655771 0.009318964 -0.9288 0.3529751
## lag.B -0.000701710 0.000174736 -4.0158 0.0000592381402
## lag.log(LSTAT) 0.183012203 0.044628101 4.1008 0.0000411673005
##
## Rho: 0.70632, LR test value: 144.38, p-value: < 2.22e-16
## Asymptotic standard error: 0.044762
## z-value: 15.78, p-value: < 2.22e-16
## Wald statistic: 248.99, p-value: < 2.22e-16
##
## Log likelihood: 233.5222 for mixed model
## ML residual variance (sigma squared): 0.021929, (sigma: 0.14809)
## Number of observations: 506
## Number of parameters estimated: 21
## AIC: -425.04, (AIC for lm: -282.66)
## LM test for residual autocorrelation
## test value: 0.30334, p-value: 0.5818
```

I p-value sono divenuti in generale più significativi ma la variabile CHAS non spiega quasi nulla, procediamo con la rimozione.

#Spatial Durbin senza variabili ZN e CHAS

```
Spatial_Durbin_NO_ZN_CHAS<-lagsarlm((log(MEDV) ~ CRIM + INDUS + I(NOX^2)+AGE + TAX + PTRATIO + B + log(LSTAT)), data = boston, listw = nn10w,
summary(Spatial_Durbin_NO_ZN_CHAS)
```

```
##
## Call:lagsarlm(formula = (log(MEDV) ~ CRIM + INDUS + I(NOX^2) + AGE +
## TAX + PTRATIO + B + log(LSTAT)), data = boston, listw = nn10w,
## Durbin = TRUE, zero.policy = TRUE, tol.solve = 1e-14)
##
## Residuals:
## Min 1Q Median 3Q Max
## -0.6404921 -0.0831352 -0.0073632 0.0909888 0.8570458
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.54083742 0.24449179 6.3022 0.0000000002934
## CRIM -0.00653191 0.00107329 -6.0859 0.0000000011584
## INDUS -0.00328513 0.00286143 -1.1481 0.2509383
## I(NOX^2) -0.21000697 0.16946819 -1.2392 0.2152671
## AGE -0.00050176 0.00053448 -0.9388 0.3478471
## TAX -0.00018350 0.00011433 -1.6050 0.1084844
## PTRATIO -0.01115137 0.00576584 -1.9340 0.0531080
```

```

## B          0.00061537  0.00011932   5.1571 0.0000002507743
## log(LSTAT) -0.36854133  0.01963068 -18.7737    < 2.2e-16
## lag.CRIM   0.00042529  0.00245691   0.1731    0.8625739
## lag.INDUS  0.00250945  0.00406098   0.6179    0.5366135
## lag.I(NOX^2) -0.12773821  0.21581639  -0.5919    0.5539284
## lag.AGE    0.00312442  0.00085351   3.6607    0.0002516
## lag.TAX    0.00041387  0.00017976   2.3023    0.0213169
## lag.PTRATIO -0.00951225  0.00922991  -1.0306    0.3027333
## lag.B      -0.00068620  0.00017424  -3.9383 0.0000820558349
## lag.log(LSTAT) 0.18510639  0.04434205   4.1745 0.0000298626103
##
## Rho: 0.72471, LR test value: 166.85, p-value: < 2.22e-16
## Asymptotic standard error: 0.042913
##      z-value: 16.888, p-value: < 2.22e-16
## Wald statistic: 285.21, p-value: < 2.22e-16
##
## Log likelihood: 232.4765 for mixed model
## ML residual variance (sigma squared): 0.021925, (sigma: 0.14807)
## Number of observations: 506
## Number of parameters estimated: 19
## AIC: -426.95, (AIC for lm: -262.1)
## LM test for residual autocorrelation
## test value: 0.32316, p-value: 0.56971

```

L'AIC è diminuito ulteriormente.