R: Spatial regression

Roger Bivand

Thursday, 5 September 2019, 15:12-15:36

Required current contributed CRAN packages:

I am running R 3.6.1, with recent `update.packages()`.

```
needed <- c("sf", "stars", "sp", "classInt", "raster", "colorspace", "RColorBrewer", "ggplot2", "cartog
```

Beijing data set

```
library(HSAR)
library(sp)
data(landSPDF)
data(landprice)
data(Beijingdistricts)
```

```
library(sf)
```

```
## Linking to GEOS 3.7.2, GDAL 3.0.1, PROJ 6.2.0
```

```
land_sf <- st_as_sf(landSPDF)
landprice_sf <- merge(land_sf, landprice, by="obs")
(landprice_sf <- landprice_sf[order(landprice_sf$district.id.x),])
```

```
## Simple feature collection with 1117 features and 12 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 428553.1 ymin: 4406815 xmax: 463693.2 ymax: 4440423
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=117 +k=1 +x_0=500000 +y_0=0 +ellps=krass +units=m +no_de
## First 10 features:
##      obs district.id.x lnprice   lnarea  lndcbd dsubway   dpark    dele
## 187  189             3 5.57430 10.27820 9.94866 6.83023 7.06579 6.81916
## 188  190             3 7.16382 11.58780 9.93534 7.14334 6.78243 6.67827
## 700  968             3 7.61282  8.94551 9.91779 7.64360 6.84364 4.60356
## 701  969             3 6.81564  5.81928 9.91940 7.64640 6.88254 4.10025
## 702  970             3 6.93528  7.71869 9.91752 7.65810 6.86760 4.53460
## 709  992             3 7.45757  9.20029 9.84785 7.78904 6.95662 7.05138
## 710  993             3 7.12569  7.97788 9.84388 7.81991 7.00792 7.11267
## 711  994             3 7.48522  7.78634 9.84203 7.83398 7.03089 7.13981
## 717 1001             3 5.87349 10.70910 9.95534 7.89121 7.11019 5.67984
## 181  183             5 6.79302  6.39403 9.92025 6.76006 6.23524 6.10494
##       popden crimerate district.id.y year                 geometry
## 187 0.548966  10.75110             3    1   POINT (430237 4422804)
## 188 0.548966  10.75110             3    1 POINT (430547.1 4423001)
## 700 0.548966  10.75110             3    0 POINT (431029.4 4423667)
## 701 0.548966  10.75110             3    0 POINT (431001.8 4423695)
## 702 0.548966  10.75110             3    0 POINT (431040.9 4423697)
## 709 0.548966  10.75110             3    0   POINT (432164 4422080)
## 710 0.548966  10.75110            3    0 POINT (432239.7 4422081)
## 711 0.548966  10.75110            3    0   POINT (432275 4422082)
## 717 0.548966  10.75110            3    0 POINT (430436.8 4424601)
## 181 1.407250   2.25832            5    1 POINT (430606.5 4420186)
```

```
all.equal(landprice_sf$district.id.x, landprice_sf$district.id.y)
```

```
## [1] TRUE
```

```
library(spatialreg)
```

```
## Loading required package: spData
```

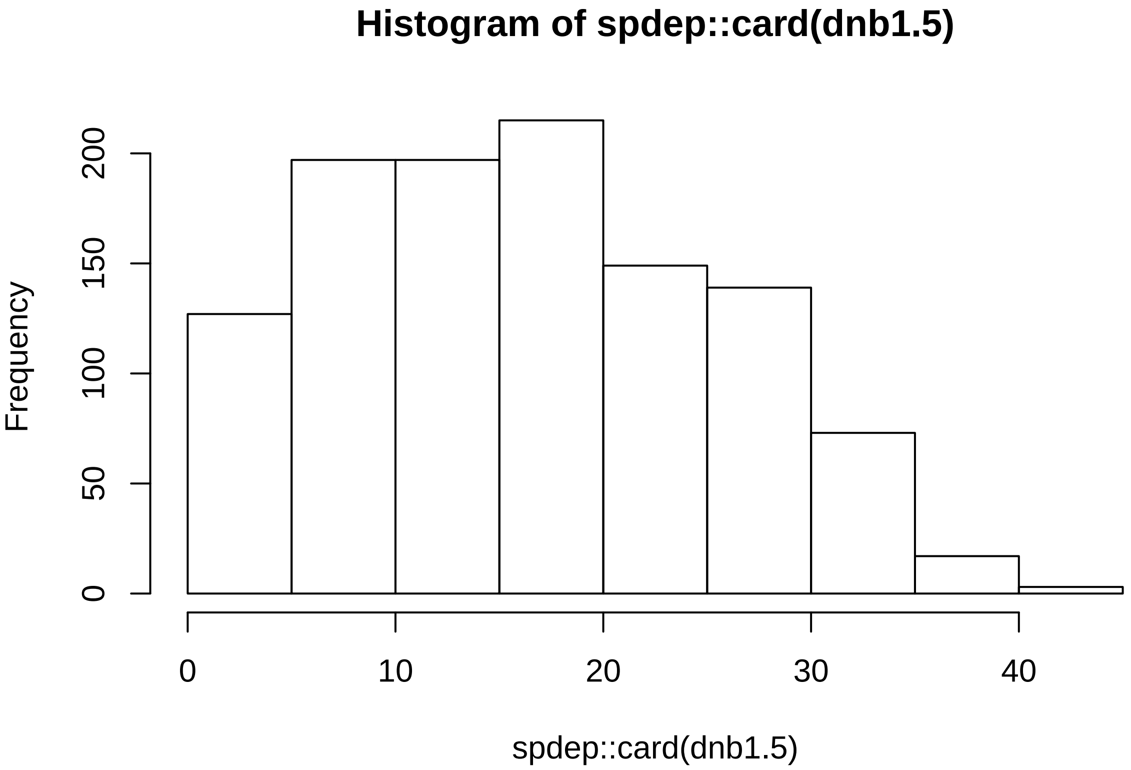
```
## Loading required package: Matrix
```

```
## Registered S3 methods overwritten by 'spatialreg':
##   method                   from
##   residuals.stsls          spdep
##   deviance.stsls           spdep
##   coef.stsls               spdep
##   print.stsls              spdep
##   summary.stsls            spdep
##   print.summary.stsls      spdep
##   residuals.gmsar          spdep
##   deviance.gmsar           spdep
##   coef.gmsar               spdep
##   fitted.gmsar             spdep
##   print.gmsar              spdep
##   summary.gmsar            spdep
##   print.summary.gmsar      spdep
##   print.lagmess            spdep
##   summary.lagmess          spdep
##   print.summary.lagmess    spdep
##   residuals.lagmess        spdep
##   deviance.lagmess         spdep
##   coef.lagmess             spdep
##   fitted.lagmess           spdep
##   logLik.lagmess           spdep
##   fitted.SFResult          spdep
##   print.SFResult           spdep
##   fitted.ME_res            spdep
##   print.ME_res             spdep
##   print.lagImpact          spdep
##   plot.lagImpact           spdep
##   summary.lagImpact        spdep
##   HPDinterval.lagImpact    spdep
##   print.summary.lagImpact  spdep
##   print.sarlm              spdep
##   summary.sarlm            spdep
##   residuals.sarlm          spdep
##   deviance.sarlm           spdep
##   coef.sarlm               spdep
##   vcov.sarlm               spdep
##   fitted.sarlm             spdep
##   logLik.sarlm             spdep
##   anova.sarlm              spdep
##    predict.sarlm            spdep
##    print.summary.sarlm      spdep
##    print.sarlm.pred         spdep
##    as.data.frame.sarlm.pred spdep
##    residuals.spautolm       spdep
##    deviance.spautolm        spdep
##    coef.spautolm            spdep
##    fitted.spautolm          spdep
##    print.spautolm           spdep
##    summary.spautolm         spdep
##    logLik.spautolm          spdep
##    print.summary.spautolm   spdep
##    print.WXImpact           spdep
##    summary.WXImpact         spdep
##    print.summary.WXImpact   spdep
##    predict.SLX              spdep
```

```
dnb1.5 <- spdep::dnearneigh(landprice_sf, 0, 1500, row.names=as.character(landprice_sf$obs))
dists <- spdep::nbdists(dnb1.5, st_geometry(landprice_sf))
edists <- lapply(dists, function(x) exp((-((x/1000)^2))/(1.5^2)))
ozpo <- spdep::set.ZeroPolicyOption(TRUE)
set.ZeroPolicyOption(TRUE)
```

```
## [1] FALSE
```

```
lw <- spdep::nb2listw(dnb1.5, glist=edists, style="W")
hist(spdep::card(dnb1.5))
```



```
landprice_sf$fyear <- factor(landprice_sf$year + 2003)
landprice_sf$price <- exp(landprice_sf$lnprice)
landprice_sf$area <- exp(landprice_sf$lnarea)
```

```r
landprice_sf$Dcbd <- exp(landprice_sf$lndcbd)
landprice_sf$Dsubway <- exp(landprice_sf$dsubway)
landprice_sf$Dpark <- exp(landprice_sf$dpark)
landprice_sf$Dele <- exp(landprice_sf$dele)
landprice_sf$f_district.id <- factor(landprice_sf$district.id.x)
(t1 <- table(table(landprice_sf$f_district.id)))
```

```
## 
##  1  2  3  4  5  6  7  8  9 10 11 12 14 15 16 17 18 19 20 21 23 25 26 27 28 
##  7  8 10  4  5 10  8 10  8  4  6  2  4  3  1  2  5  2  1  1  1  1  1  1  2 
## 31 32 33 52 
##  1  1  1  1
```

```r
Beijingdistricts$id1 <- Beijingdistricts$id+1
all.equal(unique(landprice_sf$district.id.x), Beijingdistricts$id1)
```

```
## [1] TRUE
```

```r
(Beijingdistricts_sf <- st_as_sf(Beijingdistricts))
```

```
## Simple feature collection with 111 features and 2 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: 426987.3 ymin: 4403559 xmax: 467920.9 ymax: 4443287
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=117 +k=1 +x_0=500000 +y_0=0 +ellps=krass +units=m +no_de
## First 10 features:
##   id id1                       geometry
## 0  2   3 MULTIPOLYGON (((428183.1 44...
## 1  4   5 MULTIPOLYGON (((432472.2 44...
## 2  6   7 MULTIPOLYGON (((432446.1 44...
## 3  7   8 MULTIPOLYGON (((433534.5 44...
## 4  8   9 MULTIPOLYGON (((443807.9 44...
## 5  9  10 MULTIPOLYGON (((444461 4420...
## 6 10  11 MULTIPOLYGON (((447530.6 44...
## 7 11  12 MULTIPOLYGON (((443849.6 44...
## 8 12  13 MULTIPOLYGON (((446810 4417...
## 9 13  14 MULTIPOLYGON (((445954.8 44...
```

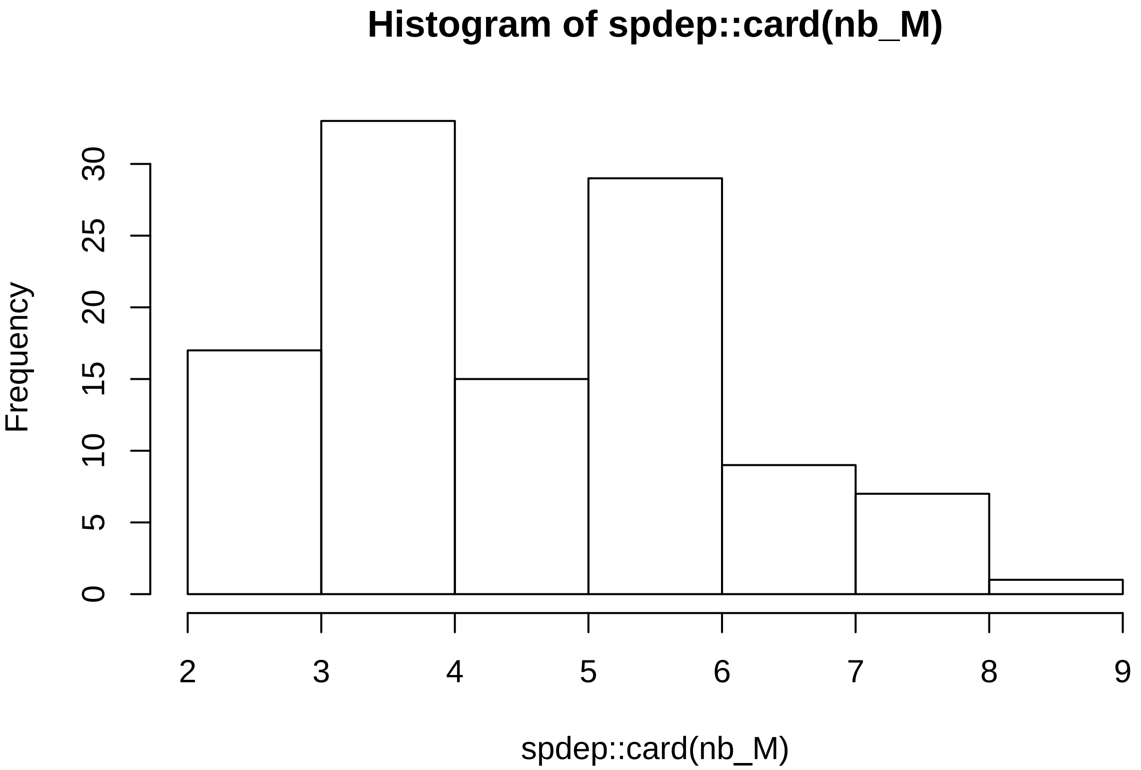
```r
t2 <- table(sapply(st_contains(Beijingdistricts_sf, landprice_sf), length))
all.equal(t1, t2)
```

```
## [1] TRUE
```

```r
nb_M <- spdep::poly2nb(Beijingdistricts, queen=FALSE, row.names=as.character(Beijingdistricts$id1))
M <- as(spdep::nb2listw(nb_M, style="B"), "CsparseMatrix")
dim(M)
```

```
## [1] 111 111
```

```r
hist(spdep::card(nb_M))
```



```r
form <- log(price) ~ log(area) + log(Dcbd) + log(Dele) + log(Dpark) + log(Dsubway) +
  crimerate + popden + fyear
OLS <- lm(form, data=landprice_sf)
summary(OLS)
```

```
##
## Call:
## lm(formula = form, data = landprice_sf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5915 -0.5752 -0.0496  0.5206  3.4042
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.153917   0.370103  38.243  < 2e-16 ***
## log(area)   -0.008253   0.018675  -0.442  0.65863
## log(Dcbd)   -0.250601   0.047752  -5.248 1.84e-07 ***
## log(Dele)   -0.085528   0.032308  -2.647  0.00823 **
## log(Dpark)  -0.284372   0.046046  -6.176 9.24e-10 ***
## log(Dsubway) -0.245748   0.034755  -7.071 2.73e-12 ***
## crimerate    0.007668   0.004458   1.720  0.08575 .
## popden       0.032827   0.010304   3.186  0.00148 **
## fyear2004   -0.164503   0.058380  -2.818  0.00492 **
## fyear2005    0.017635   0.124986   0.141  0.88782
## fyear2006   -0.120314   0.107209  -1.122  0.26201
## fyear2007    0.551384   0.117431   4.695 3.00e-06 ***
## fyear2008    0.396172   0.129571   3.058  0.00229 **
## fyear2009    2.113691   0.228688   9.243  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8335 on 1103 degrees of freedom
## Multiple R-squared:  0.3524, Adjusted R-squared:  0.3448
## F-statistic: 46.17 on 13 and 1103 DF,  p-value: < 2.2e-16
```

```
spdep::lm.morantest(OLS, listw=lw)
```

```
##
##  Global Moran I for regression residuals
##
## data:
## model: lm(formula = form, data = landprice_sf)
## weights: lw
##
## Moran I statistic standard deviate = 15.1, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Observed Moran I      Expectation         Variance
##     0.1944768641    -0.0054494313     0.0001753028
```

```
spdep::lm.LMtests(OLS, listw=lw, test="all")
```

```
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = form, data = landprice_sf)
## weights: lw
##
## LMerr = 207.15, df = 1, p-value < 2.2e-16
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = form, data = landprice_sf)
## weights: lw
##
## LMlag = 92.712, df = 1, p-value < 2.2e-16
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = form, data = landprice_sf)
## weights: lw
##
## RLMerr = 118.22, df = 1, p-value < 2.2e-16
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = form, data = landprice_sf)
## weights: lw
##
## RLMlag = 3.7725, df = 1, p-value = 0.0521
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = form, data = landprice_sf)
## weights: lw
##
## SARMA = 210.93, df = 2, p-value < 2.2e-16
```

```
SLX <- lmSLX(form, data=landprice_sf, listw=lw)
summary(impacts(SLX))
```

```
## Impact measures (SLX, estimable, n-k):
##                    Direct     Indirect        Total
## log(area)    -0.027393865  0.090812258  0.063418393
## log(Dcbd)    -0.586485128  0.362485133 -0.223999995
## log(Dele)     0.017324132 -0.159256358 -0.141932226
## log(Dpark)   -0.091259953 -0.212539298 -0.303799251
## log(Dsubway) -0.176849400 -0.146595779 -0.323445180
## crimerate     0.003700684  0.006267203  0.009967887
## popden        0.003018936  0.038261481  0.041280417
## fyear2004    -0.203603540  0.448210479  0.244606939
## fyear2005    -0.037845735  0.874443554  0.836597820
## fyear2006    -0.117303538  0.212143967  0.094840429
## fyear2007     0.689975245 -0.246715186  0.443260059
## fyear2008     0.465854623 -0.266940282  0.198914342
## fyear2009     2.203749162  0.406971450  2.610720612
## ========================================================
## Standard errors:
##                   Direct   Indirect       Total
## log(area)    0.019571325 0.04274094 0.041806133
## log(Dcbd)    0.123554643 0.13231893 0.053287909
## log(Dele)    0.051056912 0.06662771 0.043850829
## log(Dpark)   0.107566463 0.13104491 0.061869504
## log(Dsubway) 0.059404607 0.07922483 0.049145364
## crimerate    0.009139954 0.01132696 0.005692364
## popden       0.017518547 0.02466541 0.014932683
## fyear2004    0.059967444 0.14620810 0.144181383
## fyear2005    0.125200756 0.34549622 0.348471521
## fyear2006    0.108193628 0.31019156 0.314906166
## fyear2007    0.126343166 0.24786610 0.234084184
## fyear2008    0.131769313 0.36565659 0.364477241
## fyear2009    0.226910003 0.50284041 0.549874172
## ========================================================
## Z-values:
##                  Direct   Indirect      Total
## log(area)    -1.3996939  2.1247136  1.5169639
## log(Dcbd)    -4.7467672  2.7394806 -4.2035802
## log(Dele)     0.3393102 -2.3902422 -3.2367056
## log(Dpark)   -0.8484053 -1.6218813 -4.9103231
## log(Dsubway) -2.9770318 -1.8503767 -6.5813976
## crimerate      0.4048909  0.5532996  1.7510981
## popden         0.1723280  1.5512203  2.7644339
## fyear2004     -3.3952346  3.0655654  1.6965224
## fyear2005     -0.3022804  2.5309786  2.4007638
## fyear2006     -1.0842001  0.6839128  0.3011704
## fyear2007      5.4611204 -0.9953567  1.8935925
## fyear2008      3.5353802 -0.7300300  0.5457524
## fyear2009      9.7119965  0.8093452  4.7478509
##
## p-values:
##               Direct     Indirect  Total
## log(area)     0.16160499 0.0336105 0.1292758
## log(Dcbd)     2.0669e-06 0.0061536 2.6273e-05
## log(Dele)     0.73437605 0.0168373 0.0012092
## log(Dpark)    0.39621231 0.1048288 9.0926e-07
## log(Dsubway)  0.00291054 0.0642593 4.6605e-11
## crimerate     0.68555771 0.5800583 0.0799290
## popden        0.86317968 0.1208489 0.0057022
## fyear2004     0.00068570 0.0021726 0.0897870
## fyear2005     0.76243832 0.0113745 0.0163609
## fyear2006     0.27827611 0.4940303 0.7632845
## fyear2007     4.7314e-08 0.3195628 0.0582791
## fyear2008     0.00040719 0.4653718 0.5852361
## fyear2009     < 2.22e-16 0.4183166 2.0559e-06
```

```r
spdep::lm.morantest(SLX, listw=lw)
```

```
##
##  Global Moran I for regression residuals
##
## data:
## model: lm(formula = formula(paste("y ~ ", paste(colnames(x)[-1],
## collapse = "+"))), data = as.data.frame(x), weights = weights)
## weights: lw
##
## Moran I statistic standard deviate = 14.294, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Observed Moran I      Expectation         Variance
##     0.1768367391    -0.0095750057     0.0001700666
```

```r
spdep::lm.LMtests(SLX, listw=lw, test="all")
```

```
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = formula(paste("y ~ ", paste(colnames(x)[-1],
## collapse = "+"))), data = as.data.frame(x), weights = weights)
## weights: lw
##
## LMerr = 171.28, df = 1, p-value < 2.2e-16
##
##
##  Lagrange multiplier diagnostics for spatial dependence
```

```
##
## data:
## model: lm(formula = formula(paste("y ~ ", paste(colnames(x)[-1],
## collapse = "+"))), data = as.data.frame(x), weights = weights)
## weights: lw
##
## LMlag = 166.26, df = 1, p-value < 2.2e-16
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = formula(paste("y ~ ", paste(colnames(x)[-1],
## collapse = "+"))), data = as.data.frame(x), weights = weights)
## weights: lw
##
## RLMerr = 7.3124, df = 1, p-value = 0.006848
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = formula(paste("y ~ ", paste(colnames(x)[-1],
## collapse = "+"))), data = as.data.frame(x), weights = weights)
## weights: lw
##
## RLMlag = 2.2935, df = 1, p-value = 0.1299
##
##
##  Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = formula(paste("y ~ ", paste(colnames(x)[-1],
## collapse = "+"))), data = as.data.frame(x), weights = weights)
## weights: lw
##
## SARMA = 173.57, df = 2, p-value < 2.2e-16
```

```r
e <- eigenw(lw)
SDEM <- errorsarlm(form, data=landprice_sf, listw=lw, Durbin=TRUE, control=list(pre_eig=e))
summary(impacts(SDEM))
```

```
## Impact measures (SDEM, estimable, n):
##                    Direct      Indirect        Total
## log(area)    -0.025312192  0.0899318870  0.064619695
## log(Dcbd)    -0.588066020  0.3297094158 -0.258356604
## log(Dele)    -0.002311498 -0.1215960956 -0.123907593
## log(Dpark)   -0.081889383 -0.2187482895 -0.300637673
## log(Dsubway) -0.168100990 -0.1136152278 -0.281716218
## crimerate     0.006368132 -0.0007289184  0.005639214
## popden        0.004766271  0.0394667820  0.044233053
## fyear2004    -0.206852047  0.2244675077  0.017615461
## fyear2005    -0.047901527  0.5532037864  0.505302260
## fyear2006    -0.132401052  0.0290395422 -0.103361510
## fyear2007     0.693289805 -0.4672464525  0.226043352
## fyear2008     0.485618731 -0.1354646418  0.350154090
## fyear2009     2.299169487  0.7608233880  3.059992875
## ========================================================
## Standard errors:
##                    Direct   Indirect       Total
## log(area)    0.018213195 0.05439058 0.059111166
## log(Dcbd)    0.125909677 0.14337589 0.091299942
## log(Dele)    0.045766222 0.08001448 0.071996590
## log(Dpark)   0.101522411 0.14359787 0.096960618
## log(Dsubway) 0.054624512 0.09528292 0.079291144
## crimerate    0.008310533 0.01266441 0.009369951
## popden       0.015882383 0.02871415 0.023371534
## fyear2004    0.055951753 0.19532568 0.209065378
## fyear2005    0.118378556 0.44432426 0.480862483
## fyear2006    0.102489167 0.35908167 0.392488897
## fyear2007    0.116560989 0.32719106 0.350864647
## fyear2008    0.123098193 0.45396729 0.483946755
## fyear2009    0.222961274 0.56093975 0.677756394
## ========================================================
## Z-values:
##                    Direct    Indirect       Total
## log(area)    -1.38977220  1.65344597  1.09318930
## log(Dcbd)    -4.67053869  2.29961551 -2.82975649
## log(Dele)    -0.05050663 -1.51967605 -1.72102031
## log(Dpark)   -0.80661386 -1.52333933 -3.10061630
## log(Dsubway) -3.07739115 -1.19239863 -3.55293420
## crimerate     0.76627244 -0.05755644  0.60184029
## popden        0.30009801  1.37447145  1.89260378
## fyear2004    -3.69697167  1.14919608  0.08425815
## fyear2005    -0.40464699  1.24504519  1.05082488
## fyear2006    -1.29185411  0.08087169 -0.26334888
## fyear2007     5.94787165 -1.42805385  0.64424659
## fyear2008     3.94497044 -0.29840177  0.72353846
## fyear2009    10.31196782  1.35633709  4.51488603
##
## p-values:
##              Direct     Indirect Total
## log(area)    0.16459806 0.09824  0.27431069
## log(Dcbd)    3.0041e-06 0.02147  0.00465834
## log(Dele)    0.95971867 0.12859  0.08524714
## log(Dpark)   0.41988899 0.12767  0.00193118
## log(Dsubway) 0.00208821 0.23310  0.00038096
## crimerate    0.44351421 0.95410  0.54728045
## popden       0.76410240 0.16930  0.05841058
## fyear2004    0.00021819 0.25048  0.93285119
## fyear2005    0.68573701 0.21312  0.29333903
## fyear2006    0.19640767 0.93554  0.79228169
## fyear2007    2.7165e-09 0.15328  0.51941554
## fyear2008    7.9810e-05 0.76540  0.46934913
## fyear2009    < 2.22e-16 0.17499  6.3351e-06
```

```
LR1.sarlm(SDEM)
```

```
##
##  Likelihood Ratio diagnostics for spatial dependence
```

```
##
## data:
## Likelihood ratio = 106.85, df = 1, p-value < 2.2e-16
## sample estimates:
## Log likelihood of spatial error model
##                             -1296.305
##           Log likelihood of OLS fit y
##                             -1349.730
```

```
Hausman.test(SDEM)
```

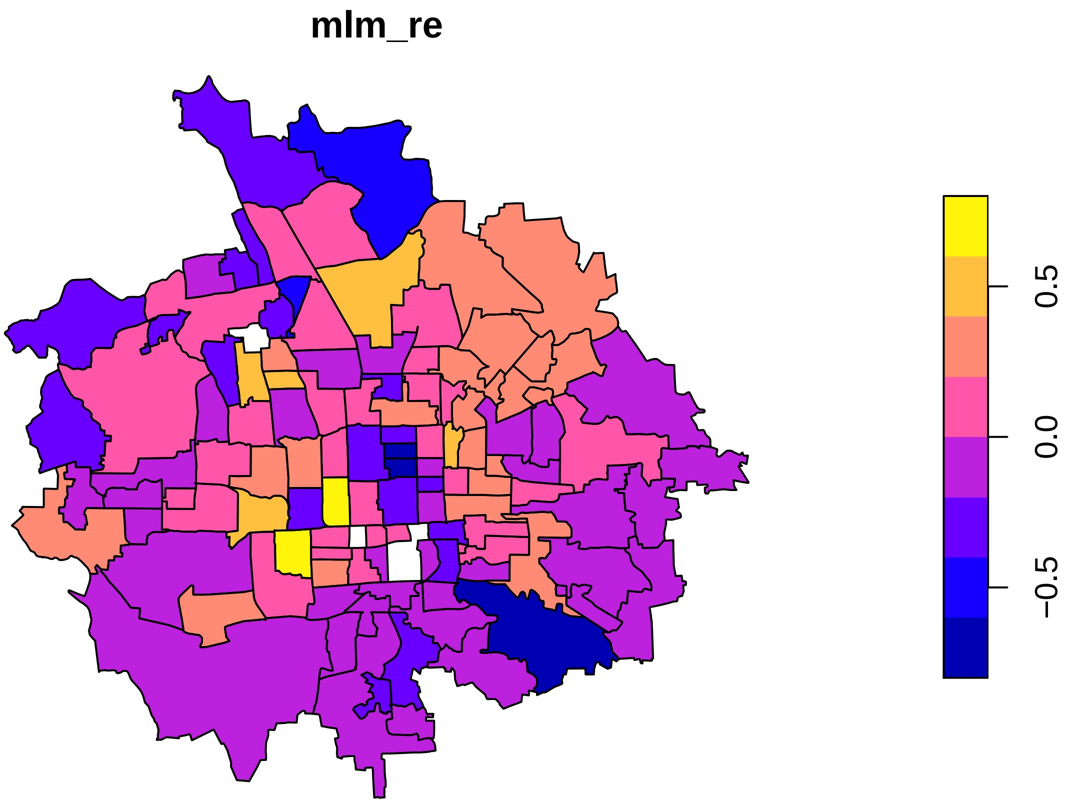
```
##
##  Spatial Hausman test (asymptotic)
##
## data:  NULL
## Hausman test = 12.872, df = 27, p-value = 0.99
```

```
library(lme4)
mlm_1 <- lmer(update(form, . ~ . + (1 | f_district.id)), data=landprice_sf, REML=FALSE)
Beijingdistricts_sf$mlm_re <- ranef(mlm_1)[[1]][,1]
```

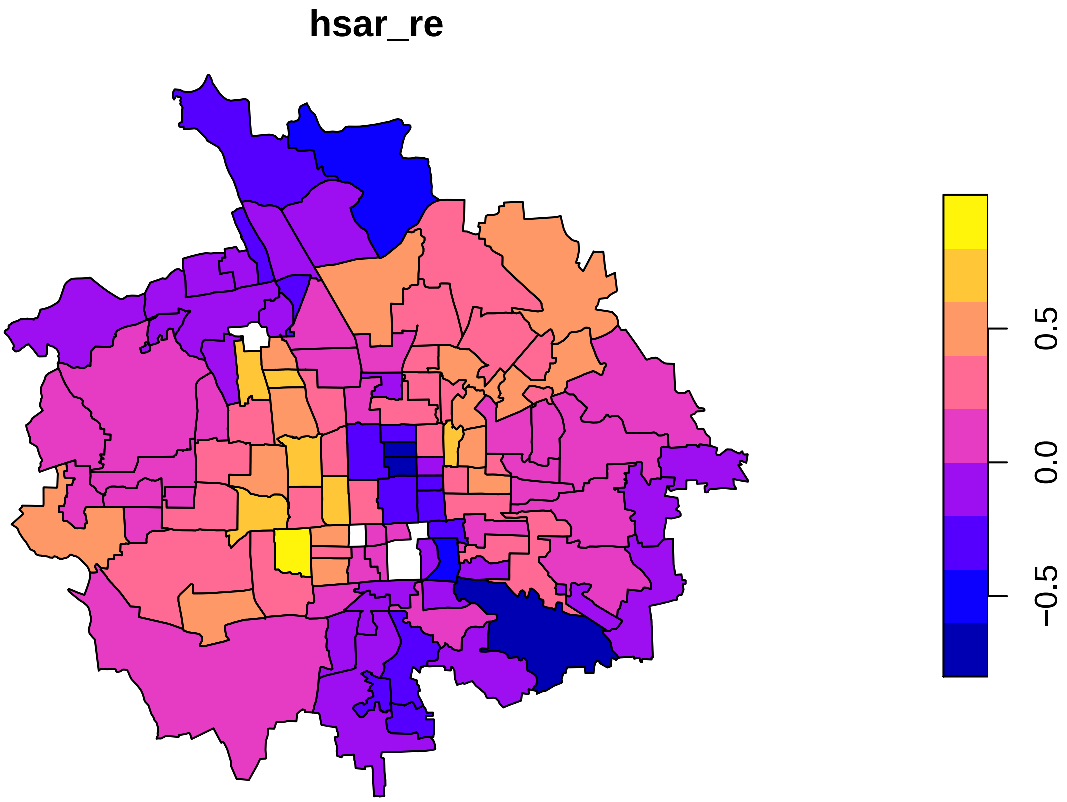
```
library(Matrix)
suppressMessages(library(MatrixModels))
Delta <- as(model.Matrix(~ -1 + f_district.id, data=landprice_sf, sparse=TRUE), "dgCMatrix")
```

```
m_hsar <- hsar(form, data=landprice_sf, W=NULL, M=M, Delta=Delta, burnin=500, Nsim=5000, thinning=1)
Beijingdistricts_sf$hsar_re <- m_hsar$Mus[1,]
```

```
plot(Beijingdistricts_sf[,"mlm_re"])
```



```r
plot(Beijingdistricts_sf[,"hsar_re"])
```


hsar_re
0.5
0.0
−0.5