Assignment 2

Spatial Regression Models

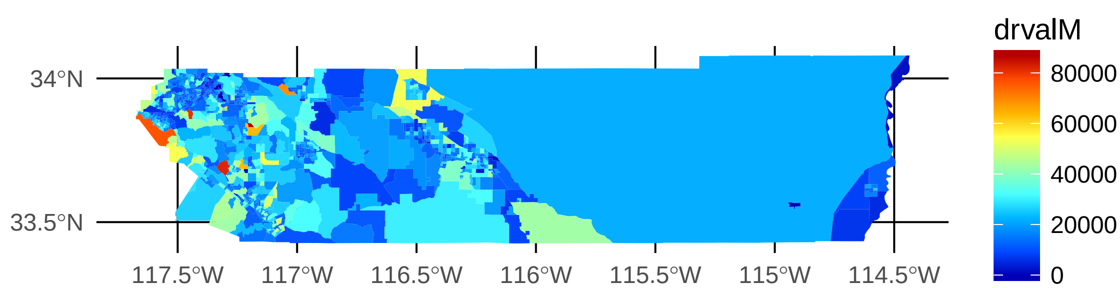
Juan Carlos Villaseñor-Derbez

2018-03-05

```
riverside <- read_sf(dsn = "./shp", layer = "LPA_Pop_Char_bg") %>%
  filter(countyname == "Riverside") %>%
  mutate(location = case_when(LPAgrp == 0 ~ "None",
                              LPAgrp == 1 ~ "Rural",
                              LPAgrp == 2 ~ "Exurb",
                              LPAgrp == 3 ~ "Suburb",
                              TRUE ~ "Center"),
         LPAgrp = fct_relevel(location, "Center", "Suburb", "Exurb", "Rural"))
```

Visualize the data

```
ggplot(riverside) +
  geom_sf(aes(fill = drvalM), color = "transparent") +
  theme_minimal() +
  scale_y_continuous(breaks = c(33.5, 34)) +
  scale_fill_gradientn(colors = colorRamps::matlab.like(20))
```



Fit the models

Non spatial OLS

The model is specified by:

$$drvalM = \beta_0 + \sum_{s=1}^{7} \sigma_s HHSIZE_s + \sum_{v=1}^{8} \gamma_v HHVEH_v + \sum_{l=1}^{4} \lambda_l LPAgrp_l + \epsilon$$

Which assumes that outcome variables are a function of household size, household vehicles, and location of the block group

```
frmla <- formula("drvalM ~ HHSIZE1 + HHSIZE2 + HHSIZE3 + HHSIZE4 +
                  HHSIZE5 + HHSIZE6 + HHSIZE7 + HHVEH1 + HHVEH2 +
                  HHVEH3 + HHVEH4 + HHVEH5 + HHVEH6 + HHVEH7 + HHVEH8 + LPAgrp")

NonSpOLS <- lm(formula = frmla, data = riverside)
```

Define 10 nearest neighbors for spatial models

```
ids <- rownames(riverside)

knn <- riverside %>%
  as("Spatial") %>%
  coordinates() %>%
  knearneigh(k=10) %>%
  knn2nb(knn = ., row.names = ids) %>%
  nb2listw(style = "W")
```

Spatial Lag

The spatial lag model assumes that, in addition to variables in the ordinary least squares regression, the Y values of a blockgroup are also influenced by the Y values of their neighboring block groups.

$$drvalM = \beta_0 + \sum_{s=1}^{7} \sigma_s HHSIZE_s + \sum_{v=1}^{8} \gamma_v HHVEH_v + \sum_{l=1}^{4} \lambda_l LPAgrp_l + \rho WY + \epsilon$$

```
SpLag <- lagsarlm(formula = frmla, data = riverside, listw = knn)
```

Spatial Error

The spatial error model assumes that the observations of Y are correlated with the neighboring block groups' errors. That is, the ϵ term in the non-spatial regression model is correlated with $]epslion$ of the neighbors. Thismodel is specified as

$$drvalM = \beta_0 + \sum_{s=1}^{7} \sigma_s HHSIZE_s + \sum_{v=1}^{8} \gamma_v HHVEH_v + \sum_{l=1}^{4} \lambda_l LPAgrp_l + \upsilon$$

Where υ is given by $\lambda W \upsilon + \epsilon$

These first three models can be examined in table 1.

```r
SpaErr <- errorsarlm(formula = frmla, data = riverside, listw = knn)
```

SARAR

Finally, the SARAR model assumes that Y is explained by a set of explanatory variables, neighboring Ys and neighboring errors:

$$drvalM = \beta_0 + \sum_{s=1}^{7} \sigma_s HHSIZE_s + \sum_{v=1}^{8} \gamma_v HHVEH_v + \sum_{l=1}^{4} \lambda_l LPAgrp_l + \rho WY + \upsilon$$

Where υ is given by $\lambda W \upsilon + \epsilon$

```r
sarar <- sacsarlm(formula = frmla, data = riverside, listw = knn)
```

Spatial Lag with Lagged explanatory variables

```r
SpLagLag <- lagsarlm(formula = frmla, data = riverside, listw = knn, type= "mixed")
```

SARAR with Lagged explanatory variables

```r
sararlag <- sacsarlm(formula = frmla, data = riverside, listw = knn, type= "sacmixed")
```

Model interpretations and findings

Overall model interpretation is very similar. While the `impacts` function provides a more clear interpretation of the effects by separating the direct from the indirect interactions, we can visualize the effects with a termplot (Figure 2). Full model coefficients and goodness of fit are provided in table 1 and the printed `impacts`.

We observe that groupblock where households are located further away, have higher values of `drvalM` (Fig 2A). The SARAR and SARAR lagged model have higer estimates for these coefficients. Household size has different effects on `drvalM` (Fig 2B). We observe that households made up from 1 and 2 people drive less. Then, households made of 3 and 4 people drive more, and then the estimate is again reduced. Increasing the number of vehicles in a household block group also resulted in increases in `drvalM`, with exception of households with 7 vehicles (Fig 2C).

Comparing the models

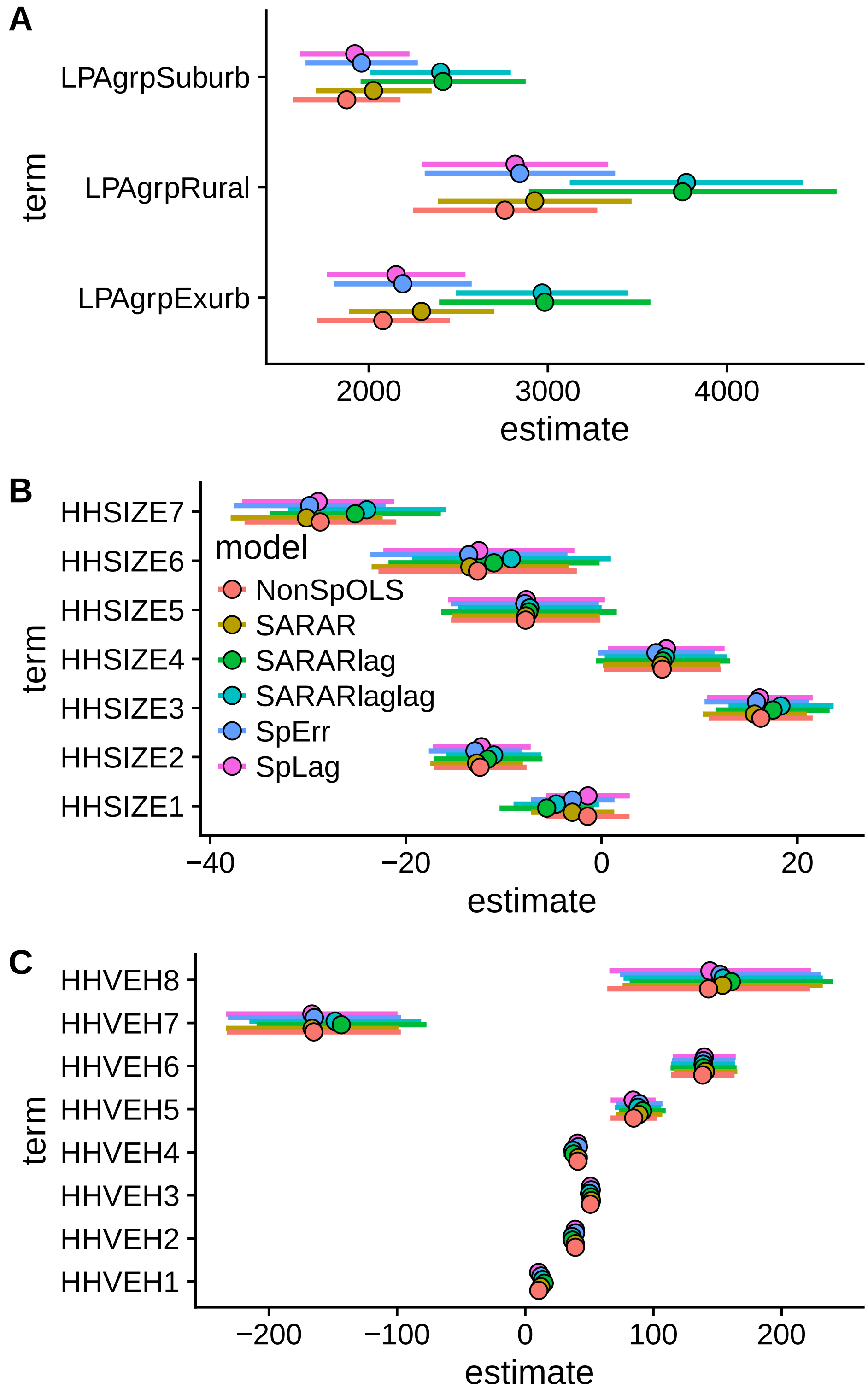


Table 1: Coefficient estimates for a non-spatial OLS regression, a regression with spatial lags, and a regression with spatial errors. Type of model is indicated above each column.

	Dependent variable:		
	drvalM		
	OLS	*spatial autoregressive*	*spatial error*
	(1)	(2)	(3)
HHSIZE1	−1.430 (4.232)	−1.398 (4.263)	−2.974 (4.242)
HHSIZE2	−12.424*** (4.717)	−12.281** (4.982)	−12.946*** (4.710)
HHSIZE3	16.269*** (5.293)	16.139*** (5.387)	15.813*** (5.284)
HHSIZE4	6.191 (5.963)	6.613 (5.939)	5.551 (5.952)
HHSIZE5	−7.771 (7.600)	−7.698 (7.994)	−7.846 (7.548)
HHSIZE6	−12.668 (10.126)	−12.545 (9.744)	−13.577 (10.039)
HHSIZE7	−28.757*** (7.727)	−28.964*** (7.738)	−29.835*** (7.727)
HHVEH1	10.581** (5.291)	10.503* (5.413)	12.272** (5.282)
HHVEH2	39.140*** (4.395)	39.030*** (4.641)	39.299*** (4.389)
HHVEH3	50.908*** (5.170)	51.000*** (5.384)	51.528*** (5.155)
HHVEH4	41.028*** (7.328)	40.905*** (7.428)	41.437*** (7.303)
HHVEH5	84.636*** (17.947)	84.241*** (17.500)	89.092*** (17.810)
HHVEH6	138.506*** (24.509)	139.759*** (24.471)	139.073*** (24.498)
HHVEH7	−164.990** (67.636)	−166.465** (66.784)	−164.605** (67.187)
HHVEH8	143.056* (78.890)	144.147* (78.404)	152.125* (78.015)
LPAgrpSuburb	1,876.430*** (297.789)	1,922.062*** (304.721)	1,958.931*** (311.764)
LPAgrpExurb	2,078.742*** (370.112)	2,152.185*** (384.865)	2,189.119*** (384.948)
LPAgrpRural	2,759.469*** (513.867)	2,816.420*** (517.750)	2,842.660*** (530.481)
Constant	−1,188.886*** (323.243)	−1,086.913*** (372.804)	−1,280.219*** (339.001)
Observations	1,030	1,030	1,030
R^2	0.959		
Adjusted R^2	0.958		
Log Likelihood		−9,419.357	−9,417.929
σ^2		5,137,769.000	5,116,317.000
Akaike Inf. Crit.		18,880.710	18,877.860
Residual Std. Error	2,288.306 (df = 1011)		
F Statistic	1,314.339*** (df = 18; 1011)		
Wald Test (df = 1)		0.379	3.169*
LR Test (df = 1)		0.390	3.245*
Note:			*p<0.1; **p<0.05; ***p<0.01

```r
impacts(sarar, listw = knn)
```

```
## Impact measures (sac, exact):
##                    Direct     Indirect       Total
## HHSIZE1         -2.997239   0.04051842   -2.956721
## HHSIZE2        -12.774161   0.17268852  -12.601472
## HHSIZE3         15.621827  -0.21118493   15.410642
## HHSIZE4          6.087507  -0.08229446    6.005213
## HHSIZE5         -7.745026   0.10470176   -7.640324
## HHSIZE6        -13.463435   0.18200653  -13.281428
## HHSIZE7        -30.163122   0.40776260  -29.755359
## HHVEH1          12.236555  -0.16542085   12.071134
## HHVEH2          39.138571  -0.52909793   38.609473
## HHVEH3          51.651990  -0.69826160   50.953729
## HHVEH4          41.328755  -0.55870611   40.770049
## HHVEH5          88.819080  -1.20070791   87.618372
## HHVEH6         140.634558  -1.90117963  138.733378
## HHVEH7        -166.297846   2.24811086 -164.049735
## HHVEH8         154.092816  -2.08311617  152.009700
## LPAgrpSuburb 2025.880511 -27.38702929 1998.493482
## LPAgrpExurb  2293.792078 -31.00881345 2262.783265
## LPAgrpRural  2926.550222 -39.56280552 2886.987417
```

```r
impacts(SpLagLag, listw = knn)
```

```
## Impact measures (mixed, exact):
##                    Direct     Indirect       Total
## HHSIZE1         -4.644181    34.834975   30.190794
## HHSIZE2        -11.029135    18.326192    7.297057
## HHSIZE3         18.315757    18.512828   36.828585
## HHSIZE4          6.513253    28.119349   34.632602
## HHSIZE5         -7.340759   -19.829420  -27.170179
## HHSIZE6         -9.234656    49.798709   40.564053
## HHSIZE7        -24.001434    33.802088    9.800654
## HHVEH1          13.548385   -43.307200  -29.758815
## HHVEH2          36.598702    -6.740896   29.857805
## HHVEH3          50.269179   -18.929675   31.339505
## HHVEH4          37.255709   -25.060329   12.195380
## HHVEH5          88.014856  -148.448107  -60.433252
## HHVEH6         138.834306     1.632463  140.466769
## HHVEH7        -148.352502     9.009130 -139.343373
## HHVEH8         154.540602  -374.901609 -220.361007
## LPAgrpSuburb 2401.082186 -1201.939724 1199.142463
## LPAgrpExurb  2967.737816 -1022.013810 1945.724006
## LPAgrpRural  3773.414666   292.444813 4065.859479
```

```r
impacts(sararlag, listw = knn)
```

```
## Impact measures (sacmixed, exact):
##                    Direct     Indirect       Total
## HHSIZE1         -4.915105    38.941904   34.026799
## HHSIZE2        -11.251686    21.466435   10.214749
## HHSIZE3         17.941866    23.883388   41.825254
## HHSIZE4          6.782305    29.470232   36.252537
## HHSIZE5         -7.785577   -18.822431  -26.608008
```

```
## HHSIZE6        -9.974738    57.796703   47.821964
## HHSIZE7       -24.487880    38.405469   13.917589
## HHVEH1         13.912896   -49.147232  -35.234336
## HHVEH2         36.715293    -8.343657   28.371636
## HHVEH3         50.774249   -22.976123   27.798127
## HHVEH4         37.295976   -29.369562    7.926414
## HHVEH5         88.451380  -169.902463  -81.451083
## HHVEH6        139.350823     8.574885  147.925708
## HHVEH7       -144.220655   -38.567390 -182.788045
## HHVEH8        152.076819  -488.883067 -336.806248
## LPAgrpSuburb 2390.957324 -1267.838956 1123.118368
## LPAgrpExurb  2964.801147  -968.159452 1996.641694
## LPAgrpRural  3761.569971   547.575724 4309.145695
```