hw2models

Gabriel Gilling

12/1/2020

In this lab I am interested in assessing the determinants of preferences for environmental regulations in the 116th Congressional Districts. The response variable is an index I built using CCES questionnaires (see previous lab) to measure respondent preferences at the CD level. For the sake of this lab, I am using a simple aggregated measure, where responses are tallied and then summarised at the CD level.

I have 4 dependent variables: the number of GHG emitting facilities (GHGEF) in a CD (using count by location in QGIS), the average level of C02 emissions from those GHGEFs within a 10km buffer (computed in QGIS), a Party Idenfication variable constructed from average PID responses in the CCES (again tallied and aggregated at the CD level) and finally an index variable representing House member roll call votes accessible at [govtrack](https://www.govtrack.us/congress/bills/subjects/environmental_protection/6038?fbclid=IwAR1lLZp4jvnExWOOBPN8xC1j6lTNkuX4y_fYt5HvfeTy9mzXKrciKdlh1Io#terms2=6040&bill_type%5B%5D=3,1,5) for bills passed around environmental protection in the 116th congress.

The idea is to understand what plays the most important role in determining environmental preferences: do voters pick up on queues from their representatives? or are their preferences amongst party lines? or finally, are they a matter of proximity to GHGEFs, a proxy for employment and economic welfare?

Because of the highly polarized nature of the 116th congress, I am testing the hypothesis that the determinants of support for regulation is a matter of political ideology above all else.

Let’s start by loading in our packages.

library(spdep)

## Loading required package: sp

## Loading required package: spData

## To access larger datasets in this package, install the spDataLarge  
## package with: `install.packages('spDataLarge',  
## repos='https://nowosad.github.io/drat/', type='source')`

## Loading required package: sf

## Linking to GEOS 3.8.1, GDAL 3.1.1, PROJ 6.3.1

library(rgdal)

## rgdal: version: 1.5-18, (SVN revision 1082)  
## Geospatial Data Abstraction Library extensions to R successfully loaded  
## Loaded GDAL runtime: GDAL 3.1.1, released 2020/06/22  
## Path to GDAL shared files: /Library/Frameworks/R.framework/Versions/4.0/Resources/library/rgdal/gdal  
## GDAL binary built with GEOS: TRUE   
## Loaded PROJ runtime: Rel. 6.3.1, February 10th, 2020, [PJ\_VERSION: 631]  
## Path to PROJ shared files: /Library/Frameworks/R.framework/Versions/4.0/Resources/library/rgdal/proj  
## Linking to sp version:1.4-4  
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,  
## use options("rgdal\_show\_exportToProj4\_warnings"="none") before loading rgdal.

library(spatialreg)

## 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

##   
## Attaching package: 'spatialreg'

## The following objects are masked from 'package:spdep':  
##   
## anova.sarlm, as\_dgRMatrix\_listw, as\_dsCMatrix\_I, as\_dsCMatrix\_IrW,  
## as\_dsTMatrix\_listw, as.spam.listw, bptest.sarlm, can.be.simmed,  
## cheb\_setup, coef.gmsar, coef.sarlm, coef.spautolm, coef.stsls,  
## create\_WX, deviance.gmsar, deviance.sarlm, deviance.spautolm,  
## deviance.stsls, do\_ldet, eigen\_pre\_setup, eigen\_setup, eigenw,  
## errorsarlm, fitted.gmsar, fitted.ME\_res, fitted.sarlm,  
## fitted.SFResult, fitted.spautolm, get.ClusterOption,  
## get.coresOption, get.mcOption, get.VerboseOption,  
## get.ZeroPolicyOption, GMargminImage, GMerrorsar, griffith\_sone,  
## gstsls, Hausman.test, HPDinterval.lagImpact, impacts, intImpacts,  
## Jacobian\_W, jacobianSetup, l\_max, lagmess, lagsarlm, lextrB,  
## lextrS, lextrW, lmSLX, logLik.sarlm, logLik.spautolm, LR.sarlm,  
## LR1.sarlm, LR1.spautolm, LU\_prepermutate\_setup, LU\_setup,  
## Matrix\_J\_setup, Matrix\_setup, mcdet\_setup, MCMCsamp, ME, mom\_calc,  
## mom\_calc\_int2, moments\_setup, powerWeights, predict.sarlm,  
## predict.SLX, print.gmsar, print.ME\_res, print.sarlm,  
## print.sarlm.pred, print.SFResult, print.spautolm, print.stsls,  
## print.summary.gmsar, print.summary.sarlm, print.summary.spautolm,  
## print.summary.stsls, residuals.gmsar, residuals.sarlm,  
## residuals.spautolm, residuals.stsls, sacsarlm, SE\_classic\_setup,  
## SE\_interp\_setup, SE\_whichMin\_setup, set.ClusterOption,  
## set.coresOption, set.mcOption, set.VerboseOption,  
## set.ZeroPolicyOption, similar.listw, spam\_setup, spam\_update\_setup,  
## SpatialFiltering, spautolm, spBreg\_err, spBreg\_lag, spBreg\_sac,  
## stsls, subgraph\_eigenw, summary.gmsar, summary.sarlm,  
## summary.spautolm, summary.stsls, trW, vcov.sarlm, Wald1.sarlm

Let’s read in our generate shapefile with all of the data.

shp\_df <- readOGR("/Users/gabgilling/Documents/Documents - Gabriel’s MacBook Pro/GitHub/GIS-Project/shapefiles")

## OGR data source with driver: ESRI Shapefile   
## Source: "/Users/gabgilling/Documents/Documents - Gabriel’s MacBook Pro/GitHub/GIS-Project/shapefiles", layer: "final"  
## with 444 features  
## It has 22 fields  
## Integer64 fields read as strings: ALAND AWATER f\_field\_1 f\_GEOID GEOID2 votes\_byGE

For model one simply run an ordinary linear regression. Include at least one control variable.

m1 <- lm(f\_pref\_ind ~ NUMPOINTS + GHG.QUANTI + votes\_by\_1 + as.numeric(f\_pid), data = shp\_df@data)  
summary(m1)

##   
## Call:  
## lm(formula = f\_pref\_ind ~ NUMPOINTS + GHG.QUANTI + votes\_by\_1 +   
## as.numeric(f\_pid), data = shp\_df@data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.8933 -0.5234 -0.0016 0.4726 2.7755   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.072e-01 3.838e-01 1.582 0.11451   
## NUMPOINTS -5.934e-03 1.829e-03 -3.244 0.00129 \*\*   
## GHG.QUANTI 9.931e-08 2.003e-07 0.496 0.62031   
## votes\_by\_1 1.338e+00 1.313e-01 10.193 < 2e-16 \*\*\*  
## as.numeric(f\_pid) -5.573e-01 4.131e-01 -1.349 0.17821   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.795 on 344 degrees of freedom  
## (95 observations deleted due to missingness)  
## Multiple R-squared: 0.3707, Adjusted R-squared: 0.3634   
## F-statistic: 50.67 on 4 and 344 DF, p-value: < 2.2e-16

This model’s output is surprising: while the congressional roll call votes are highly significant, PID isn’t. On the other hand, the number of GHGEFs in a CD is quite predictive of preferences: the higher the number of GHGEFs in a CD, the less a respodent is likely to support regulations. The sign of the coefficient for *votes\_by\_1* makes sense: the more a congressperson votes in favor of regulating the environment, the more likely his electorate is as well.

I am using a queen’s contiguity weights, which offer a less stringent definition of what constitutes a CD’s neighbor: any CD that shares a vertex, not just an edge, will be included. This makes intuitive sense as we do not think that spatial dependency in the variables we are assessing is just a matter of sharing a border, but rather whether there is connectivity at all between the CDs.

Let’s add the continguity and create the weights, and then run the spatial dependence tests:

First, Moran’s I:

list.queen <-poly2nb(shp\_df, queen = T)  
  
W <- nb2listw(list.queen, style = "W", zero.policy = T)  
  
moran.lm <- lm.morantest(m1, W, alternative = "two.sided", zero.policy = T)  
  
print(moran.lm)

##   
## Global Moran I for regression residuals  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ NUMPOINTS + GHG.QUANTI + votes\_by\_1 +  
## as.numeric(f\_pid), data = shp\_df@data)  
## weights: W  
##   
## Moran I statistic standard deviate = 7.3332, p-value = 2.247e-13  
## alternative hypothesis: two.sided  
## sample estimates:  
## Observed Moran I Expectation Variance   
## 0.245579775 -0.006869700 0.001185116

The p-value is extremely low and we can conclude spatial dependence is an issue.

Now let’s run more tests:

print(lm.LMtests(m1, W, test = "all", zero.policy = T))

##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ NUMPOINTS + GHG.QUANTI + votes\_by\_1 +  
## as.numeric(f\_pid), data = shp\_df@data)  
## weights: W  
##   
## LMerr = 49.257, df = 1, p-value = 2.246e-12  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ NUMPOINTS + GHG.QUANTI + votes\_by\_1 +  
## as.numeric(f\_pid), data = shp\_df@data)  
## weights: W  
##   
## LMlag = 78.475, df = 1, p-value < 2.2e-16  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ NUMPOINTS + GHG.QUANTI + votes\_by\_1 +  
## as.numeric(f\_pid), data = shp\_df@data)  
## weights: W  
##   
## RLMerr = 0.63026, df = 1, p-value = 0.4273  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ NUMPOINTS + GHG.QUANTI + votes\_by\_1 +  
## as.numeric(f\_pid), data = shp\_df@data)  
## weights: W  
##   
## RLMlag = 29.849, df = 1, p-value = 4.672e-08  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ NUMPOINTS + GHG.QUANTI + votes\_by\_1 +  
## as.numeric(f\_pid), data = shp\_df@data)  
## weights: W  
##   
## SARMA = 79.105, df = 2, p-value < 2.2e-16

The output above tells us that our model suffers from both spatial autocorrelation for both the error and the lag, but that using Robust Errors allievates the issue.

Let’s run a 2nd model with robust standard errors:

err <- errorsarlm(m1, data = shp\_df@data, W, zero.policy = T)  
  
summary(err)

##   
## Call:errorsarlm(formula = m1, data = shp\_df@data, listw = W, zero.policy = T)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.955974 -0.454522 -0.011158 0.423073 2.207302   
##   
## Type: error   
## Coefficients: (asymptotic standard errors)   
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 9.8224e-01 3.8852e-01 2.5282 0.01147  
## NUMPOINTS -1.3474e-03 1.8630e-03 -0.7232 0.46953  
## GHG.QUANTI -3.3163e-07 2.0470e-07 -1.6201 0.10521  
## votes\_by\_1 9.2665e-01 1.3105e-01 7.0708 1.541e-12  
## as.numeric(f\_pid) -9.0106e-01 4.0845e-01 -2.2061 0.02738  
##   
## Lambda: 0.53156, LR test value: 52.3, p-value: 4.7651e-13  
## Asymptotic standard error: 0.056798  
## z-value: 9.3588, p-value: < 2.22e-16  
## Wald statistic: 87.587, p-value: < 2.22e-16  
##   
## Log likelihood: -386.4962 for error model  
## ML residual variance (sigma squared): 0.50113, (sigma: 0.70791)  
## Number of observations: 349   
## Number of parameters estimated: 7   
## AIC: 786.99, (AIC for lm: 837.29)

This is extremely interesting: when using robust standard errors the number of GHGEFs is no longer significant, however, a respondent’s PID is now statistically significant. This goes in hand with the original hypothesis: in this polarized climate, ideological attitudes (your party ID and how your congressperson votes) trumps economic determinants.

Let’s run a third model, taking out economic factors:

m2 <- lm(f\_pref\_ind ~votes\_by\_1 + as.numeric(f\_pid), data = shp\_df@data)  
  
W <- nb2listw(list.queen, style = "W", zero.policy = T)  
  
moran.lm2 <- lm.morantest(m2, W, alternative = "two.sided", zero.policy = T)  
  
print(moran.lm2)

##   
## Global Moran I for regression residuals  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ votes\_by\_1 + as.numeric(f\_pid), data =  
## shp\_df@data)  
## weights: W  
##   
## Moran I statistic standard deviate = 8.4366, p-value < 2.2e-16  
## alternative hypothesis: two.sided  
## sample estimates:  
## Observed Moran I Expectation Variance   
## 0.286683038 -0.004972128 0.001195092

print(lm.LMtests(m2, W, test = "all", zero.policy = T))

##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ votes\_by\_1 + as.numeric(f\_pid), data =  
## shp\_df@data)  
## weights: W  
##   
## LMerr = 67.125, df = 1, p-value = 2.22e-16  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ votes\_by\_1 + as.numeric(f\_pid), data =  
## shp\_df@data)  
## weights: W  
##   
## LMlag = 93.208, df = 1, p-value < 2.2e-16  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ votes\_by\_1 + as.numeric(f\_pid), data =  
## shp\_df@data)  
## weights: W  
##   
## RLMerr = 0.0029351, df = 1, p-value = 0.9568  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ votes\_by\_1 + as.numeric(f\_pid), data =  
## shp\_df@data)  
## weights: W  
##   
## RLMlag = 26.087, df = 1, p-value = 3.265e-07  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = f\_pref\_ind ~ votes\_by\_1 + as.numeric(f\_pid), data =  
## shp\_df@data)  
## weights: W  
##   
## SARMA = 93.211, df = 2, p-value < 2.2e-16

We run into the same issues as before, so we can simply run a robust error model.

err2 <- errorsarlm(m2, data = shp\_df@data, W, zero.policy = T)  
  
summary(err2)

##   
## Call:errorsarlm(formula = m2, data = shp\_df@data, listw = W, zero.policy = T)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.9342713 -0.4627456 -0.0098688 0.4554420 1.9807789   
##   
## Type: error   
## Coefficients: (asymptotic standard errors)   
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.93569 0.38840 2.4091 0.01599  
## votes\_by\_1 0.97026 0.13008 7.4588 8.726e-14  
## as.numeric(f\_pid) -0.99100 0.40504 -2.4467 0.01442  
##   
## Lambda: 0.52152, LR test value: 60.079, p-value: 9.1038e-15  
## Asymptotic standard error: 0.057511  
## z-value: 9.0681, p-value: < 2.22e-16  
## Wald statistic: 82.23, p-value: < 2.22e-16  
##   
## Log likelihood: -388.1018 for error model  
## ML residual variance (sigma squared): 0.50724, (sigma: 0.71221)  
## Number of observations: 349   
## Number of parameters estimated: 5   
## AIC: 786.2, (AIC for lm: 844.28)

We can interpret the model output as such: for *votes\_by\_1* comparing two respondents that share the same ideological score, an increase of 1 unit in the roll call vote index is associated by an increase of 0.97 units in the environmental regulation preference index. This isn’t very intuitively understandable but the gist is the same: when an congressperson favors regulation, his constituents have a tendency to favor them too (the opposite is true). The standard error is tiny compared to the coefficient, so the variable is very significant.