Lab 5

Antonio Jurlina

10/17/2020

According to all the gathered statistics, model analyses and tests, the model choice can be narrowed down to two, the simultaneous autoregressive error model (SARerr) and the simultaneous autoregressive lag model (SARlag). The conditionally autoregressive model (CAR) is only slightly less efficient, while the ordinary least squares (OLS) approach provides the least satisfactory statistics. Both the SARerr and the SARlag have highest Log Likelihood, and lowest AIC and BIC criteria values, relative to the OLS and the CAR models (Table 1).

First, the Lagrange Multiplier test (Table 2) identifies the spatial error and spatial lag models (besides the SARMA approach which is not under consideration here) as the most viable ones. The low p-values suggest that adopting the OLS approach within a more expansive framework (SARerr and SARlag) would add statistically significant improvements to model fit.

Next, focusing on the normality of the residuals, we can use the results from Table 4 and Figure 1 to understand the distributions. Shapiro-Wilk normality test outcomes indicate a strong likelihood that residuals across all four models are not normally distributed. Furthermore, the Jarque Bera test (Table 3) confirms this as well.

Unfortunately, it seems that the heteroskedasticity of the residuals is present across all 4 models (Figure 2). This is is further confirmed with a Breusch-Pagan test (Table 5) rejecting the homoskedasticity hypothesis on OLS, SARerr and SARlag models. Unfortunately, a valid Breusch-Pagan test for a spautolm object in R has not been sufficiently developed yet and the available approaches might confuse spatial autocorrelation with heteroskedasticity.

Moran's I test (Table 6) strongly indicates the presence of spatial autocorrelation in the OLS approach, as well as the possibility of some in the CAR approach. However, it fails to reject the absence of spatial autocorrelation for both the SARerr and SARlag models.

Finally, the estimates across all four models are fairly similar (Tables 7-10). The X value is not estimated with significance, while Y and the Jarman Index are consistently significant, even at the 99% level. This is to be expected, as our main worry lies with the accuracy of the standard errors, in the presence of aforementioned issues.

Table 1: Diagnostic result comparison

model	Log Likelihood	AIC	BIC
OLS	-787.2561	1584.512	1600.747
SARerr	-765.5657	1543.131	1562.614
SARlag	-766.3806	1544.761	1564.243
CAR	-778.1354	1568.271	1587.753

Table 2: Lagrange multiplier diagnostics for spatial dependence

test	statistic	p.value	parameter
LMerr	42.19598	0.00000	1
LMlag	46.19572	0.00000	1
RLMerr	3.42183	0.06434	1
RLMlag	7.42157	0.00644	1
SARMA	49.61755	0.00000	2

Table 3: Jarque Bera test

model	statistic	p.value	parameter
OLS	17.63913	0.00015	2
SARerr	31.83172	0.00000	2
SARlag	26.24660	0.00000	2
CAR	36.17169	0.00000	2

Table 4: Shapiro-Wilk normality test

model	statistic	p.value
OLS	0.97506	0.00179
SARerr	0.96666	0.00017
SARlag	0.96192	0.00005
CAR	0.96137	0.00004

Table 5: Breusch-Pagan test

model	statistic	p.value	parameter
OLS	45.41050	0	3
SARerr	35.32473	0	3
SARlag	33.83480	0	3

Table 6: Moran I test under randomisation

model	estimate1	estimate2	estimate3	statistic	p.value
OLS	0.30306	-0.00532	0.00212	6.69165	0.00000
SARerr	0.01766	-0.00532	0.00212	0.49912	0.30885
SARlag	0.05060	-0.00532	0.00212	1.21381	0.11241
CAR	0.08124	-0.00532	0.00212	1.87874	0.03014

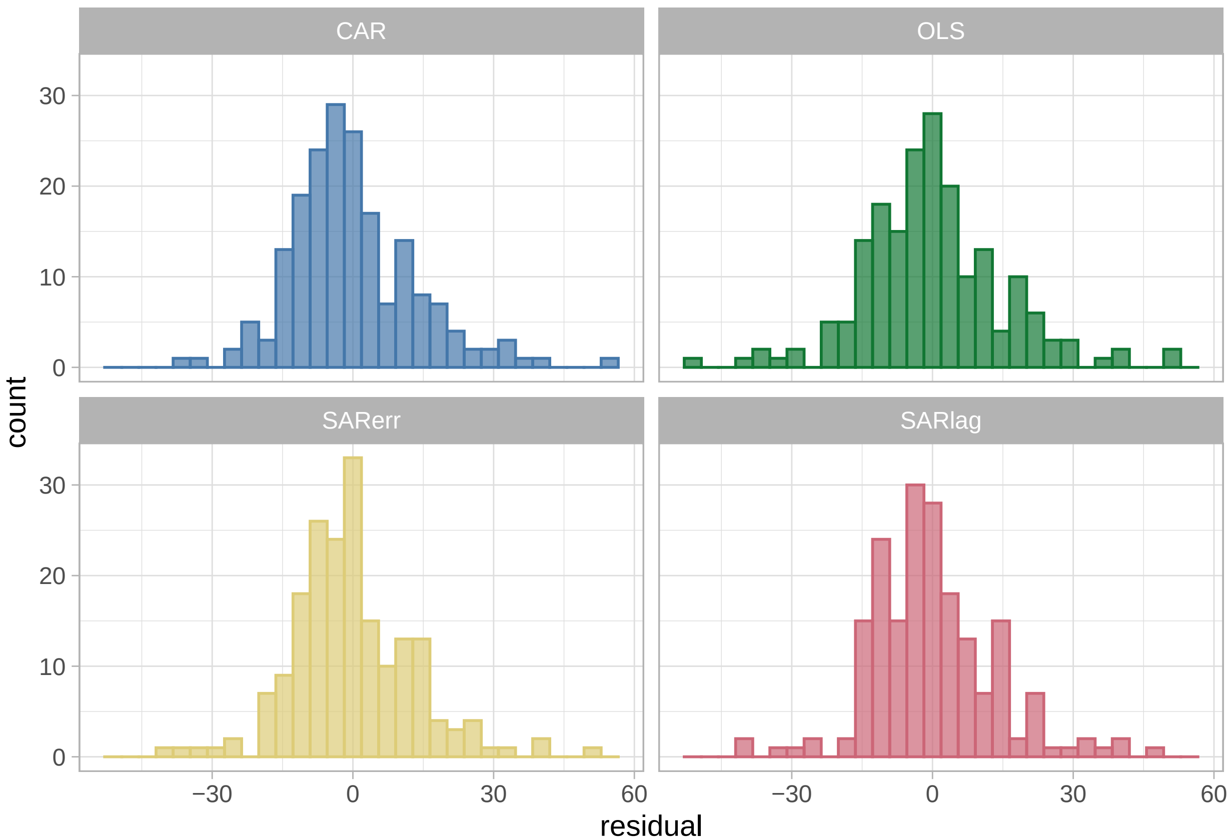

Figure 1: Distributions of model residuals

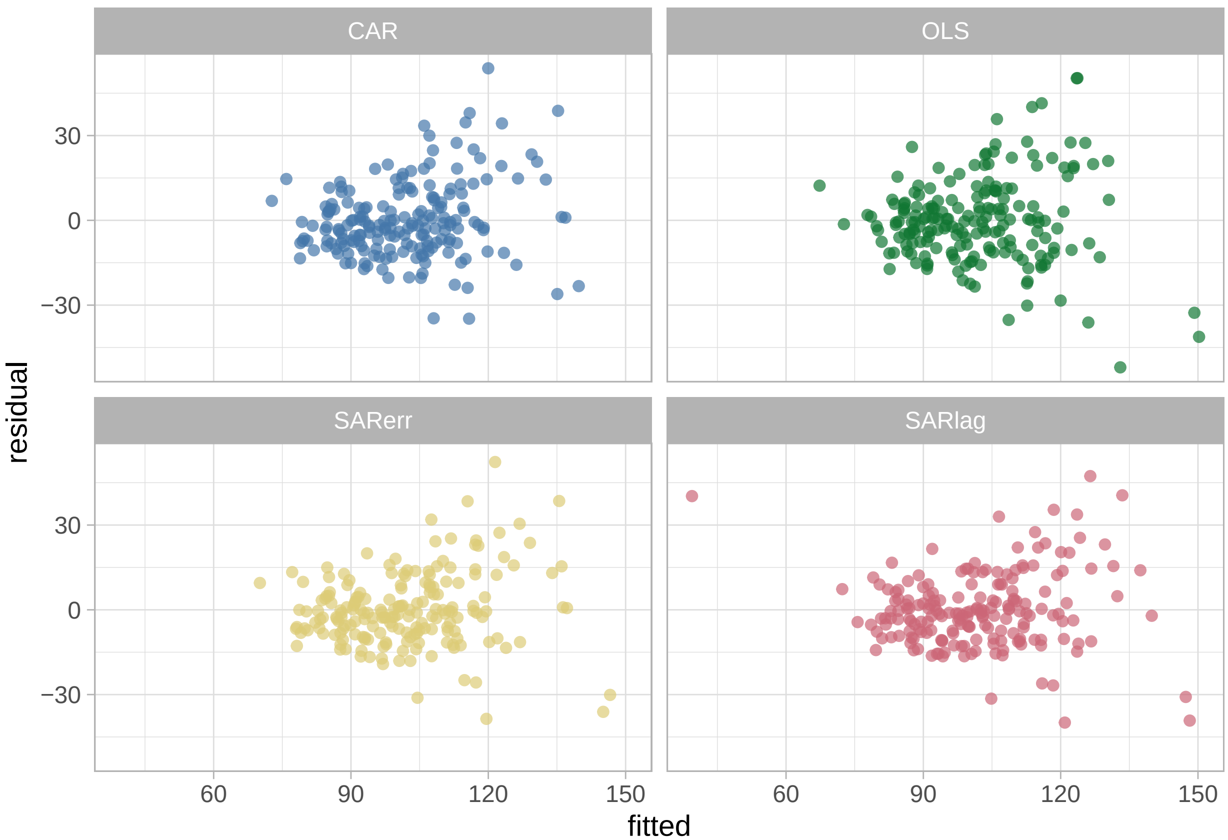

Figure 2: Residuals vs. fitted values across all models

Table 7: Ordinary least squares model (OLS)

term	estimate	std.error	statistic	p.value
(Intercept)	29.41390	8.47002	3.47271	0.00064
x	0.00030	0.00016	1.81224	0.07156
y	0.00066	0.00010	6.69506	0.00000
jarman_sc	0.45650	0.05768	7.91402	0.00000

Table 8: Simultaneous autoregressive error model (SARerr)

term	estimate	std.error	statistic	p.value
(Intercept)	40.60861	13.93295	2.91457	0.00356
x	0.00035	0.00032	1.11051	0.26678
y	0.00078	0.00019	4.03645	0.00005
jarman_sc	0.30696	0.05397	5.68718	0.00000
lambda	0.61578	0.07072	8.70736	0.00000

Table 9: Simultaneous autoregressive lag model (SARlag)

term	estimate	std.error	statistic	p.value
rho	0.48125	0.06941	6.93355	0.00000
(Intercept)	6.75851	8.13908	0.83038	0.40632
x	0.00017	0.00014	1.20115	0.22969
y	0.00030	0.00010	2.91715	0.00353
jarman_sc	0.33243	0.05153	6.45122	0.00000

Table 10: Conditionally autoregressive model (CAR)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	50.71421	10.28520	4.93079	0.00000
x	0.00035	0.00022	1.60628	0.10821
y	0.00087	0.00013	6.72046	0.00000
jarman_sc	0.18701	0.05777	3.23696	0.00121

```
## R version 3.6.2 (2019-12-12)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.29       broom_0.7.0      spatialreg_1.1-5 Matrix_1.2-18
##  [5] tseries_0.10-47  lmtest_0.9-38    zoo_1.8-8        ggthemes_4.2.0
##  [9] rgdal_1.5-16     classInt_0.4-3   spdep_1.1-5      spData_0.3.8
## [13] rebus_0.1-3      lubridate_1.7.9  tmap_3.1         sf_0.9-5
## [17] sp_1.4-2         forcats_0.5.0    stringr_1.4.0    dplyr_1.0.2
## [21] purrr_0.3.4      readr_1.3.1      tidyr_1.1.2      tibble_3.0.3
## [25] ggplot2_3.3.2    tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] leafem_0.1.3          colorspace_1.4-1      deldir_0.1-29
##  [4] ellipsis_0.3.1        class_7.3-17          rprojroot_1.3-2
##  [7] leaflet_2.0.3         base64enc_0.1-3       fs_1.5.0
## [10] dichromat_2.0-0       rstudioapi_0.11       farver_2.0.3
## [13] fansi_0.4.1           xml2_1.3.2            codetools_0.2-16
## [16] splines_3.6.2         jsonlite_1.7.0        tmaptools_3.1
## [19] dbplyr_1.4.4          png_0.1-7             compiler_3.6.2
## [22] httr_1.4.2            backports_1.1.9       assertthat_0.2.1
## [25] cli_2.0.2             htmltools_0.5.0       tools_3.6.2
## [28] coda_0.19-3           gtable_0.3.0          glue_1.4.2
## [31] rebus.base_0.0-3      gmodels_2.18.1        Rcpp_1.0.5
## [34] cellranger_1.1.0      raster_3.3-13         vctrs_0.3.4
## [37] gdata_2.18.0          nlme_3.1-149          leafsync_0.1.0
## [40] crosstalk_1.1.0.1     lwgeom_0.2-5          xfun_0.16
## [43] rebus.datetimes_0.0-1 rvest_0.3.6           lifecycle_0.2.0
## [46] rebus.numbers_0.0-1   gtools_3.8.2          XML_3.99-0.3
## [49] LearnBayes_2.15.1     MASS_7.3-52           scales_1.1.1
## [52] hms_0.5.3             parallel_3.6.2        expm_0.999-5
## [55] RColorBrewer_1.1-2    curl_4.3              quantmod_0.4.17
## [58] yaml_2.2.1            stringi_1.4.6         highr_0.8
## [61] e1071_1.7-3           TTR_0.24.2            boot_1.3-25
## [64] rlang_0.4.7           pkgconfig_2.0.3       evaluate_0.14
## [67] lattice_0.20-41       labeling_0.3          htmlwidgets_1.5.1
## [70] tidyselect_1.1.0      here_0.1              magrittr_1.5
## [73] R6_2.4.1              generics_0.0.2        DBI_1.1.0
## [76] pillar_1.4.6          haven_2.3.1           withr_2.2.0
## [79] xts_0.12-0            units_0.6-7           stars_0.4-3
## [82] abind_1.4-5           rebus.unicode_0.0-2   modelr_0.1.8
## [85] crayon_1.3.4          KernSmooth_2.23-17    rmarkdown_2.3
## [88] grid_3.6.2            readxl_1.3.1          blob_1.2.1
```

```
## [91] reprex_0.3.0          digest_0.6.25          munsell_0.5.0
## [94] viridisLite_0.3.0     quadprog_1.5-8
```