Model 3 clean_script

Jun Qian, Lucas Schroyer, Ryan Mitchell, Oliver Chang

4/8/2021

```
load("pwd/model_2_workspace.RData")
```

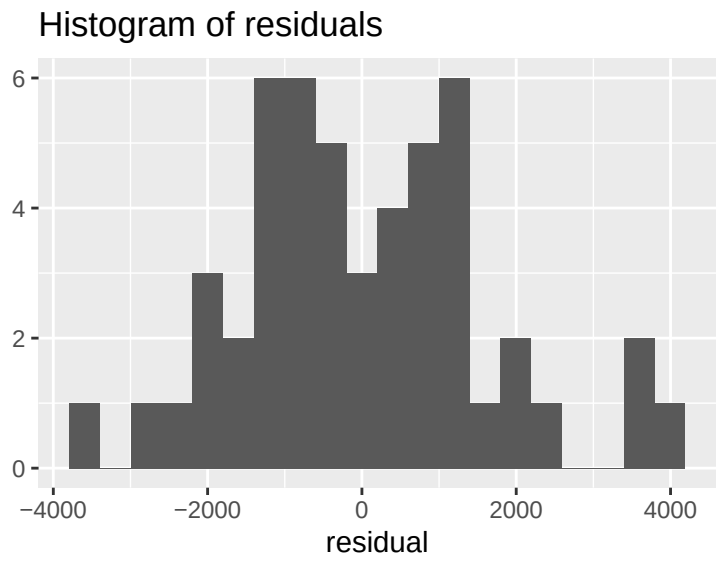
Third Model

Model 3:

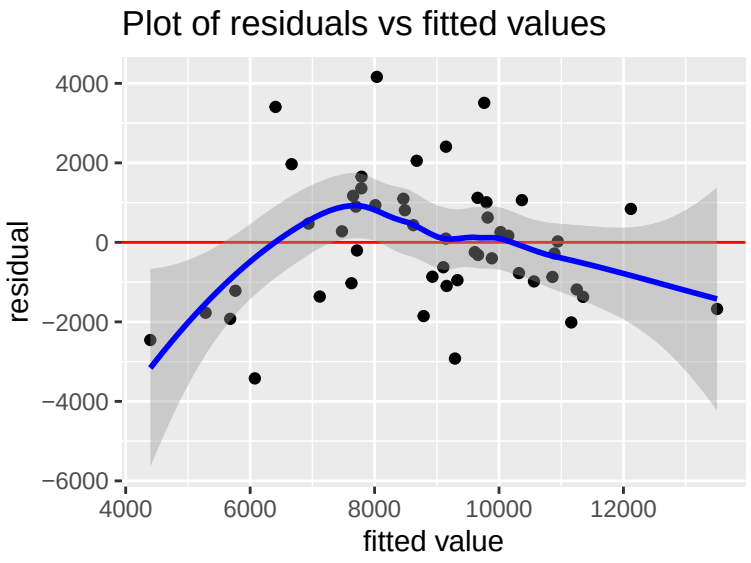
```
##
## Call:
## lm(formula = cases_per_100k_at_365d ~ median_transit_change +
##     pop_pct_age_0_24 + population_density + mask_mandate_days +
##     unemployment_benefits_days + increased_weekly_unemployment_insurance_amt_thru_jul31 +
##     business_closed_days_round1 + travel_quarantine_mandate_days +
##     stay_at_home_days, data = final_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3421.8 -1079.3   -90.2   990.0  4162.1
##
## Coefficients:
##                                                  Estimate Std. Error
## (Intercept)                                     -4556.676   4720.588
## median_transit_change                              54.348     28.765
## pop_pct_age_0_24                                  569.302    129.214
## population_density                                  5.790      1.575
## mask_mandate_days                                  -1.188      2.766
## unemployment_benefits_days                         -3.873      4.991
## increased_weekly_unemployment_insurance_amt_thru_jul31   -2.339      2.031
## business_closed_days_round1                       -16.598     23.816
## travel_quarantine_mandate_days                     -1.779      2.379
## stay_at_home_days                                   1.841      4.816
##                                                 t value Pr(>|t|)
## (Intercept)                                      -0.965 0.340206
## median_transit_change                             1.889 0.066103 .
## pop_pct_age_0_24                                  4.406 7.68e-05 ***
## population_density                                3.677 0.000694 ***
## mask_mandate_days                                -0.429 0.669918
## unemployment_benefits_days                       -0.776 0.442269
## increased_weekly_unemployment_insurance_amt_thru_jul31  -1.151 0.256369
## business_closed_days_round1                      -0.697 0.489895
## travel_quarantine_mandate_days                   -0.748 0.458904
```

```
## stay_at_home_days                                          0.382 0.704251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1781 on 40 degrees of freedom
## Multiple R-squared:  0.5695, Adjusted R-squared:  0.4727
## F-statistic:  5.88 on 9 and 40 DF,  p-value: 3.369e-05
```

```
## Analysis of Variance Table
##
## Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
##     population_density + mask_mandate_days + unemployment_benefits_days +
##     increased_weekly_unemployment_insurance_amt_thru_jul31 +
##     business_closed_days_round1 + travel_quarantine_mandate_days +
##     stay_at_home_days
## Model 2: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
##     population_density
##   Res.Df       RSS Df Sum of Sq      F Pr(>F)
## 1     40 126808331
## 2     46 144077125 -6 -17268794 0.9079 0.4991
```




```
## `geom_smooth()` using formula 'y ~ x'
```



Regression Table

```
robust_se_1 <- coeftest(model_1_final,
                        vcovHC(model_1_final, type = 'HC3'))[ , "Std. Error"]

robust_se_2 <- coeftest(model_2_final,
                        vcovHC(model_2_final, type = 'HC3'))[ , "Std. Error"]

robust_se_3 <- coeftest(model_3_final,
                        vcovHC(model_3_final, type = 'HC3'))[ , "Std. Error"]
```

```
# Print results
# stargazer(model_1_final, model_2_final, model_3_final,
#           type = "text",
#           se = list(robust_se_1, robust_se_2, robust_se_3),
#           title = "Table 1: OLS models for COVID-19 Spread",
#           column.sep.width = "0.5pt",
#           align=TRUE,
#           header = FALSE,
#           font.size = "small",
#           single.row = TRUE,
#           out = "pwd/fit_lm.txt") # to put coefficients and standard errors on same line


# jp
# stargazer(model_1_final, model_2_final, model_3_final,
#           type = 'text',
#           title = "Table 1: OLS models for COVID-19 Spread",
#           # float.env = "sidewaystable",
#           covariate.labels=c("median_transit_change",
#                              "pop_pct_age_0_24",
#                              "population_density",
#                              "mask_mandate_days",
```

```
#                                 "unemployment_benefits",
#                                 "increased_weekly_unemp",
#                                 "business_closed_days",
#                                 "travel_quarantine_man",
#                                 "stay_at_home_days"),
#           # rownames = FALSE,
#           se = list(robust_se_1, robust_se_2, robust_se_3)
#           # font.size=6,
#           # header=FALSE, # to get rid of r package output text
#           # single.row = TRUE, # to put coefficients and standard errors on same line
#           # no.space = TRUE, # to remove the spaces after each line of coefficients
#           # column.sep.width = "1pt", # to reduce column width
#           # font.size = "small" # to make font size smaller
#
# )
```

Table 1: OLS models for COVID-19 Spread

	Dependent variable:		
	cases_per_100k_at_365d		
	(1)	(2)	(3)
median_transit_change	79.250** (32.943)	82.223*** (31.481)	54.348 (45.057)
pop_pct_age_0_24		628.592*** (154.756)	569.302*** (160.445)
population_density		4.665** (2.213)	5.790** (2.789)
mask_mandate_days			-1.188 (3.116)
unemployment_benefits_days			-3.873 (5.252)
increased_weekly_unemployment_insurance_amt_thru_jul31			-2.339 (2.110)
business_closed_days_round1			-16.598 (23.841)
travel_quarantine_mandate_days			-1.779 (3.121)
stay_at_home_days			1.841 (5.388)
Constant	10,370.170*** (610.184)	-10,160.590** (5,070.544)	-4,556.676 (5,891.585)
Observations	50	50	50
R2	0.168	0.511	0.570
Adjusted R2	0.151	0.479	0.473
Residual Std. Error	2,259.262 (df = 48)	1,769.777 (df = 46)	1,780.508 (df = 40)
F Statistic	9.714*** (df = 1; 48)	16.018*** (df = 3; 46)	5.880*** (df = 9; 40)
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 1: stargazer output

Plots, Figures, and Tables

Do the plots, figures and tables that the team has chosen to include successfully move forward the argument that they are making? Has the team chosen the most effective method (a table or a chart) to display their evidence? Is that table or chart the most communicative it could be? Is every plot, figure, and table that is included in the report referenced in the narrative argument?

Assessment of the CLM.

Has the team presented a sober assessment of the CLM assumptions that might be problematic for their model? Have they presented their analysis about the consequences of these problems (including random sampling) for the models they estimate? Did they use visual tools or statistical tests, as appropriate? Did they respond appropriately to any violations?

An Omitted Variables Discussion.

Did the report miss any important sources of omitted variable bias? Are the estimated directions of bias correct? Was their explanation clear? Is the discussion connected to whether the key effects are real or whether they may be solely an artifact of omitted variable bias?

Conclusion.

Does the conclusion address the research question? Does it raise interesting points beyond numerical estimates? Does it place relevant context around the results?

Are there any other errors, faulty logic, unclear or unpersuasive writing, or other elements that leave you less convinced by the conclusions?

General Notes:

"In principle the SE reflects the degree of uncertainty or the lack of information for getting a 'good' (that is reliable) estimate of a parameter. Therefore if you keep everything else the same (eg the same variation in the response, the same number of observations) but you increase the number of separate parameters to be estimated there will be less information per parameter to get the estimate, and hence larger standard error. Precisely what happens will depend on the the degree of variation in the additional X variable that is included and how colinear it is with already included variables."

Known IID violations: Geo-spatial dependence (states near each other are not independent... physical proximity) Policy coordination dependence (states near each-other have coordinated policies (like NY/NJ quarantine policies, etc))

Other limitations: Mobility data is based on Google-Maps cell phone users. Not everyone has access to a smart phone or uses Google Maps and allows their location to be traced, so this data may not be representative of the population. Additionally, we do not have absolute numbers, only relative change data.