Causal Effect of Population with Masks on Early Covid19 Expansion

Group 6: Tian Xie, Ruoheng Sheng, Wenqian Xing, Shijie Mao

April 2022

1 Summary

Motivated by one recent paper and Covid19 datasets posted by its authors[1], we decide to analyze the causal inference on whether wearing masks can influence the spread of Covid19 in United States. With the comprehensive data-set[2][3] and necessary preprocessing, we have constructed appropriate DAG and use different techniques to estimate the ATE including **IP Weighting, Standardization G-estimation**. Moreover, we also find an instrument variable **Party** to estimate ATE. With different techniques, we compare the results and discover they give similar but slightly different estimations.

2 Acquire Data for causal Inference

2.1 Raw Data

The Covid19 data-set[2] has different statistics for each state in America. It has 151 column variables for over 50 states. The columns include social economic statistics, census statistics, and health statistics (e.g. population wearing mask). To estimate the influence of wearing mask on Covid cases, we use the total number of cases posted by CDC on August 1, 2021[3]. The reason for using cases one year after the statistics for wearing mask is we assume the influence of wearing mask was quite profound.

2.2 Data Preprocessing

After ensuring no missing values or problematic outliers, we have done following preprocessing: **Firstly**, We construct the column **case** (which is our y) by dividing the total cases in one state by its total population posted in 2020 Census, and then multiplying the number by 1000. Therefore, the correct understanding for y would by case number in 1000 people for a typical state. **Secondly**, we use z-score of percentage of people wearing mask as treatment. If z-score is larger than 1, we encode treatment as 1. Otherwise, we encode it as 0. **Finally**, we use label encoder to encode categorical data such as **Party** and finally get the usable data[4].

2.3 EDA

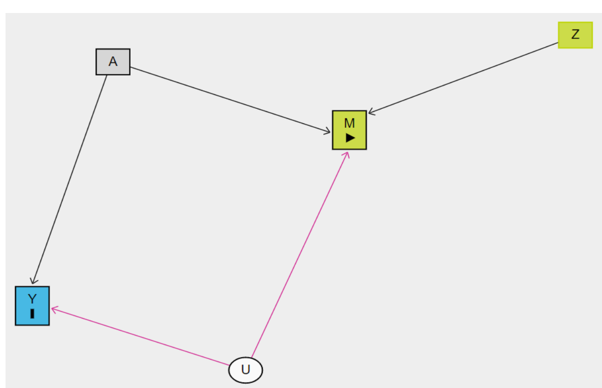
For exploratory data analysis, we inspect correlations between different factors and wearing mask. We also inspect correlations between different factors and Covid Cases. With the results, we have found several possible common causes as confounders. Here are their names and their correlations with treatment and effect:

factor	corr with treatment	corr with effect
residential percent change from baseline	0.79	-0.24
Number Homeless	0.36	0.31
Percent unemployed in 2018	0.44	-0.22
Population density per square miles	0.36	-0.12

Above 4 factors are our confounders. Also, we assume **Party** to be a possible instrument variable because they have direct influence on masking policy but may not directly influence getting infected by Covid. However, we also know there are possible backdoor paths open. But it is worthwhile having an experiment to see the results (section 3.4).

2.4 DAG

With above analysis, we come up with the DAG:



M is treatment (z-score encoding for masking percent), Y is result (Cases in 1000), A is known confounders, U is possible unobserved confounders, Z is party, which is a possible instrument variable.

3 Treatment Effect Computation

3.1 IP Weighting

According to above DAG, we finish IP weighting by fitting a Logistic Regression between M and A. Then we use the weights to fit weighted least square regression between Y and M. Details of code can be seen in our Jupyter notebook[5]. Here is the result:

```
                            WLS Regression Results
==============================================================================
Dep. Variable:                      Y   R-squared:                       0.017
Model:                            WLS   Adj. R-squared:                 -0.003
Method:                 Least Squares   F-statistic:                    0.8484
Date:                Sat, 23 Apr 2022   Prob (F-statistic):              0.362
Time:                        14:39:02   Log-Likelihood:                -294.74
No. Observations:                  51   AIC:                             593.5
Df Residuals:                      49   BIC:                             597.3
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    118.5515     14.211      8.342      0.000      89.994     147.109
M            -18.8432     20.458     -0.921      0.362     -59.954      22.268
==============================================================================
Omnibus:                       93.798   Durbin-Watson:                   2.304
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2266.114
Skew:                           5.134   Prob(JB):                         0.00
Kurtosis:                      33.999   Cond. No.                         2.58
==============================================================================
```

Therefore, by using IP weighting method, the treatment effect is -18.843. But this result is NOT significant due to the relative high p-value.

3.2 Standardization

According to above DAG, we finish standardization by fitting an ordinary least square regression between Y and M plus A (we do not use interaction terms). We also use robust covariance estimator. Then, we use the model to predict Y when M = 0 or 1. Finally, we get the estimation.

However, the OLS result suggests strong multicollinearity may exist. We look back to our correlation part (2.3) and decide to remove **residential percent change from baseline** and **Percent unemployed in 2018** because they have very high correlation with treatment.

With this modification, we get the new result:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      Y   R-squared:                       0.029
Model:                            OLS   Adj. R-squared:                 -0.033
Method:                 Least Squares   F-statistic:                    0.4685
Date:                Sat, 23 Apr 2022   Prob (F-statistic):              0.706
Time:                        14:39:03   Log-Likelihood:                -300.43
No. Observations:                  51   AIC:                             608.9
Df Residuals:                      47   BIC:                             616.6
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    123.1542     16.370      7.523      0.000      90.221     156.087
M            -18.4344     29.145     -0.633      0.530     -77.066      40.197
A2            -0.0002      0.001     -0.401      0.690      -0.001       0.001
A4            -0.0046      0.008     -0.555      0.581      -0.021       0.012
==============================================================================
Omnibus:                       99.591   Durbin-Watson:                   2.311
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2860.582
Skew:                           5.598   Prob(JB):                         0.00
Kurtosis:                      37.940   Cond. No.                     6.28e+04
==============================================================================
```

By using standardization method, the treatment effect is -18.434 which is quite similar to the IP weighting result. And it is also NOT significant. Code details are also in Jupyter Notebook[5].

3.3 G-estimation

Based on standardization, we also use the remaining two factors and treatment to try G-estimation. We fit a logistic regression model with these factors, and use the root finding tool to find appropriate ATE to make the coefficient of Y_0 0. Here is the result:

```
scipy.optimize.root_scalar(g_est_solver, bracket = [-100,100],
                           method = 'bisect', maxiter=50)
```

```
     converged: True
          flag: 'converged'
function_calls: 49
    iterations: 47
          root: -11.250867943572018
```

(a) Root finding

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                      M   No. Observations:                   51
Model:                            GLM   Df Residuals:                       47
Model Family:                Binomial   Df Model:                            3
Link Function:                  logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -20.861
Date:                Sat, 23 Apr 2022   Deviance:                       41.722
Time:                        14:39:06   Pearson chi2:                     42.3
No. Iterations:                     8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -2.7882      1.042     -2.676      0.007      -4.830      -0.746
A2             0.0001   6.66e-05      1.864      0.062   -6.36e-06       0.000
A4             0.0074      0.003      2.536      0.011       0.002       0.013
H_phi      -1.989e-07      0.006  -3.41e-05      1.000      -0.011       0.011
==============================================================================
```

(b) G-estimation Result

By using g-estimation method, the treatment effect is -11.25 which is still reasonably similar to the result of standardization method and IP-weighting. Code details are also in Jupyter Notebook[5].

3.4 Construct Instrument Variable

We would try to find instrument variables to estimate a more unbiased ATE. The potential instrument variables of our treatment variable could be the political party governing the state. We fit a two stage least square using the instrument variable: In stage one, we fit an ordinary least square between treatment and instrument. In stage two, we use predictions for treatment produced by the model in stage one to estimate ATE. Here is the result:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      Y   R-squared:                       0.098
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                    0.9767
Date:                Sat, 23 Apr 2022   Prob (F-statistic):              0.442
Time:                        14:44:33   Log-Likelihood:                -298.55
No. Observations:                  51   AIC:                             609.1
Df Residuals:                      45   BIC:                             620.7
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        284.2217     85.988      3.305      0.002     111.033     457.410
pred_M       -83.2198    130.913     -0.636      0.528    -346.891     180.452
A1            -7.4533      7.137     -1.044      0.302     -21.829       6.922
A2             0.0001      0.001      0.192      0.848      -0.001       0.001
A3           -15.6636     13.196     -1.187      0.241     -42.241      10.914
A4             0.0047      0.010      0.471      0.640      -0.015       0.025
==============================================================================
Omnibus:                       92.045   Durbin-Watson:                   2.354
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2099.828
Skew:                           5.003   Prob(JB):                         0.00
Kurtosis:                      32.800   Cond. No.                     2.82e+05
==============================================================================
```

Here we observe a wider difference between the ATE estimations produced by 3.1 to 3.3. This may happen because of the following reasons: **Firstly**, F statistics is very low, suggesting the IV is very weak and the estimation would be very volatile. **Moreover**, it is hard for us to make sure no backdoor paths exist between instrument and result. However, we should emphasize our result is normalized by 1000, which means the estimation -83 and -18 do not have that much difference as you may perceive.
Therefore, by having a weak instrument Variable, the treatment effect is -83.25 which is still within a comparatively reasonable range of results of standardization method,IP-weighting and G-estimation.

3.5 Evaluations for Different Techniques and Conclusion

Finally, we would compare the results of different techniques and give comments. By combining all those analyses, we would also try to come up with a reasonable estimation of the ATE. Code details are also in Jupyter Notebook[5].

4 Model Evaluation & Conclusion

According to above sections, we use four methods to estimate ATE of treatment (wearing mask) on effect (Covid cases). The result suggests wearing mask would make Covid cases less at the magnitude of **0.18%**. The methods are quite consistent, and the result aligns with general understanding. However, several weaknesses exist for our study:

1. The results are not significant. This may occur mainly because of **lack of data**. we do not gain more detailed datasets. We use data aggregated by each state, but if we have access for each county or even city, our estimations for ATE would have a much narrower confidence interval. Thus, a meaningful future research could be elaborate our study with city-level data.

2. The instrument variable may not be satisfying enough. Just as we mentioned in section 3.4, **Party** is only a weak instrument, and the ATE estimate is a bit different from other methods. So another possible direction for future study would be finding better instrument variables.

5 Reference

[1] Chernozhukov, Victor, Hiroyuki Kasahara, and Paul Schrimpf. "Causal impact of masks, policies, behavior on early covid-19 pandemic in the US." Journal of econometrics 220.1 (2021): 23-62.

[2] **Predictor Data (causes)**: https://raw.githubusercontent.com/ubcecon/covid-impact/master/cases_and_policies/data/covidstates.csv

[3] **Result Data**: https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36/data

[4] **Final Data for Causal Inference**:https://github.com/TianXie1999/4578-Final-Proj/blob/main/data_for_causal_inference.csv

[5] **Project Code**: https://github.com/TianXie1999/4578-Final-Proj/blob/main/project_code.ipynb