The paper solves the bias in data arising from non-random testing of entire population for Covid-19. A two-equation model overcomes the bias by using the inverse Mills ratio and improves forecasts of new deaths in all nine out of nine weekly out-of-sample comparisons. We identify the political party of the governors of states with desirable negative and undesirable positive trends.

```
library(sampleSelection)
library(margins)
library(dplyr)
##Pull Data
Cumulative_Deaths=read.csv("Cumulative Deaths.csv")
Cumulative_Infections=read.csv("Cumulative Infections.csv")
Covid=read.csv("Covid_0615.csv")
##clean data
Covid$State.Abbreviation<-Covid$STA
Final_Dataset<-merge(Covid,Cumulative_Deaths,by="State.Abbreviation")
Data<-merge(Final_Dataset,Cumulative_Infections,by="State.Abbreviation")
2020/06/15: First Step Probit Model

myprobit<-glm(TI~HES+CPT+UI+HR+HI, family=binomial(link="probit"),data=Data)
##TI is Testing Indicator (i.e. 1 for tested and 0 for not tested)
##HES is Hospital Employee Share
##CPT is portion of population that Commutes on Public Transit
##UI in Uninsured Population
##HR is Hypertension Rate (one of the leading comorbidities)
#HI is Household Income

summary(myprobit)
##
## Call:
## glm(formula = TI ~ HES + CPT + UI + HR + HI, family = binomial(link =
"probit"),
##     data = Data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7296  -0.3849  -0.3553  -0.3376   2.5180
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.4319390  0.0636110 -38.231  < 2e-16 ***
## HES          0.0663127  0.0081659   8.121 4.64e-16 ***
## CPT          0.0193450  0.0004345  44.523  < 2e-16 ***
## UI          -0.0070335  0.0009966  -7.058 1.69e-12 ***
## HR           0.0198518  0.0011021  18.012  < 2e-16 ***
## HI           0.0021614  0.0004608   4.690 2.73e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 254187  on 499999  degrees of freedom
## Residual deviance: 250590  on 499994  degrees of freedom
## AIC: 250602
##
## Number of Fisher Scoring iterations: 5
2020/06/15: Probit Model Marginal Effects
```

```
#Marginal Effects of probit model
effects_pro=margins(myprobit) #print(effects_pro)
summary(effects_pro)
```

```
1   CPT 0.0025685341    5.801388e-05    44.274475   0.000000e+00    0.0024548290
0.0026822392
2   HES 0.0088046672    1.084429e-03    8.119170    4.693813e-16    0.0066792246
0.0109301098
3   HI  0.0002869787    6.119577e-05    4.689519    2.738481e-06    0.0001670372
0.0004069203
4   HR  0.0026358250    1.465264e-04    17.988742   2.387156e-72    0.0023486386
0.0029230114
5   UI  -0.0009338730   1.323531e-04    -7.055923   1.714593e-12
-0.0011932802   -0.0006744657
5 rows
```

2020/06/15: Extract inverse Mills ratio from probit model

```
##Compute inverse Mills ratio from probit model usine 'invMillsRatio' command
from the 'sampleSelection' package.
IMR=invMillsRatio(myprobit,all=FALSE)
Data$IMR<-invMillsRatio(myprobit)$IMR1
##create table of IMR values by state
IMR.Index=cbind(as.character(Data$State.Abbreviation),Data$IMR)
IMR_table=distinct(data.frame(IMR.Index))
```

2020/06/15: Second Step Heckman Equation (i.e. with IMR)

```
##GLM with Poisson distribution to estimate cumulative deaths based on the log
of cumulative infections from the previous week, as well as the IMR.
##Use 'glm' command with 'family=poisson(link="log")'
##Only ran the model on individuals who were tested (i.e. TI = 1).
##CD is "Cumulative Deaths"
##CI is "Cumulative Infections"

glm_IMR_0615_CD<-glm(X20200615_CD~log(X20200608_CI)
+IMR,family=poisson(link="log"),
                     data=subset(Data,TI==1))
```

fig-out-of-sample-all

```
#Compute fitted values for each state
Tested_Individuals<-subset(Data,TI==1)
Fitted_CD_IMR=fitted(glm_IMR_0615_CD)
Fitted_CD=fitted(glm_0615_CD)

Results_Table=cbind(as.character(Tested_Individuals$State.Abbreviation),Tested_
Individuals$X20200622_CI,
               Tested_Individuals$X20200622_CD,Tested_Individuals$
X20200615_CD,Fitted_CD_IMR,Fitted_CD,
               Tested_Individuals$IMR)

#Pull fitted value for each state
Final_Results=distinct(data.frame(Results_Table))

View(Final_Results)
```

Assessing Nation-wide Forecast

```
#Compare fitted values with in-sample (2020/06/15) and out-of-sample
(2020/06/22) actual deaths

Actual_Infections_Out_Sample=sum(as.numeric(Final_Results[,2]))

Actual_Deaths_Out_Sample=sum(as.numeric(Final_Results[,3]))
Predict_Deaths_IMR=sum(as.numeric(Final_Results$Fitted_CD_IMR))
Predict_Deaths_no_IMR=sum(as.numeric(Final_Results$Fitted_CD))
Out_of_sample_per_diff_CD_IMR=(Predict_Deaths_IMR-Actual_
Deaths_Out_Sample)/Actual_Deaths_Out_Sample
Out_of_sample_per_diff__CD_no_IMR=(Predict_Deaths_no_IMR-
Actual_Deaths_Out_Sample)/Actual_Deaths_Out_Sample

Actual_Deaths_In_Sample=sum(as.numeric(Final_Results[,4]))
In_sample_per_diff_CD_IMR=(Predict_Deaths_IMR-Actual_Deaths_In_Sample)/Actual_
Deaths_In_Sample
In_sample_per_diff_CD_no_IMR=(Predict_Deaths_no_IMR-Actual_
Deaths_In_Sample)/Actual_Deaths_In_Sample

print(rbind(Actual_Deaths_In_Sample,In_sample_per_diff_CD_
IMR,In_sample_per_diff_CD_no_IMR))
##                                       [,1]
## Actual_Deaths_In_Sample       1.098220e+05
## In_sample_per_diff_CD_IMR     4.064085e-03
## In_sample_per_diff_CD_no_IMR -4.705036e-02
print(rbind(Actual_Infections_Out_Sample,Actual_Deaths_Out_
Sample,Predict_Deaths_no_IMR, Out_of_sample_per_diff_CD_IMR,
Out_of_sample_per_diff__CD_no_IMR))
##                                            [,1]
## Actual_Infections_Out_Sample       2.290489e+06
## Actual_Deaths_Out_Sample           1.139440e+05
## Predict_Deaths_no_IMR              1.046548e+05
## Out_of_sample_per_diff_CD_IMR     -3.225860e-02
## Out_of_sample_per_diff__CD_no_IMR -8.152394e-02
```

In all nine cases using IMR bias correction improves out-of-sample forecasts of Covid-19 deaths in US as a whole

```
cbind(Week,IMR_Perc_Diff,Without_IMR_Perc_Diff)
##        Week IMR_Perc_Diff Without_IMR_Perc_Diff
##   [1,]    1        -25.27                -26.96
##   [2,]    2        -21.61                -22.50
##   [3,]    3        -17.66                -18.73
##   [4,]    4        -12.14                -13.78
##   [5,]    5         -9.38                -11.28
##   [6,]    6         -7.25                 -9.70
##   [7,]    7         -5.53                 -9.01
##   [8,]    8         -3.94                 -8.03
```