New Pandemic Death Forecasts by State

H. D. Vinod and Katherine Theiss *

Aug. 6, 2020

Abstract

We offer state-wise forecasts of Covid-19 pandemic deaths for the week ending August 10, 2020, based on a sample selection model to correct for rationed and biased testing for the virus. We refer the reader to our earlier papers including Vinod and Theiss (2020b) and Vinod and Theiss (2020a), having earlier forecasts for the week ending August 3. We find that the overall trend in death forecasts is downward. We plot the ratio of new deaths per 100 new infections (ND/NI) measuring the success of the health-care establishment in reducing deaths. While there was a dramatic decline earlier, latest opening up decisions are preventing further declines. Our forecast of weekly new deaths during the week before November elections is below 5000.

Notes:
Funding: None to declare
Declaration of Interest: None to declare

1 Introductions to Our Predictions

Vinod and Theiss (2020b) describe how we correct for 'selection bias' in non-random testing for Covid-19 infections. We use a mild extension of Heckman (1979) two-equation model where the first so-called 'selection equation' uses the probit model to predict the probability that a randomly chosen person

*Vinod is Professor of Economics, vinod@fordham.edu and K. Theiss is a Ph.D. Candidate ktheiss@fordham.edu at Fordham University, Bronx, New York 10458.

from a state is selected for testing. The probit equation uses many state-level characteristics described in the paper. The second equation called 'observation equation' uses the generalized linear models (GLM) and inverse mills ratio (IMR) to provide relatively bias-free forecasts of cumulative Covid-19 deaths from the logarithm of one week lagged cumulative infections in a state. We use Poisson regression to ensure that predicted deaths are nonnegative.

2 Autoregressive distributed lag model adjusted forecasts

Our autoregressive distributed lag, ARDL$(1,1)$ model, has both autoregressive and distributed lag order 1. That is, the dependent variable y and the regressor z are both lagged precisely once.

We consider

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + \varepsilon_t. \tag{1}$$

Instead of using linear regression to fit (1), we use the generalized linear model (GLM) with a Poisson link function to make sure that the predicted new deaths are nonnegative. If the data contain zeros for (y_t, z_t), the GLM breaks down. Hence, we are unable to provide adjusted forecasts of new deaths for some states.

Model-based forecasts generally need a confidence interval, where the lower limit on deaths provides the best-case scenario, and the upper limit on deaths provides the worst-case scenario. We have here a very short and nonstationary time series. The 7-step maximum entropy bootstrap (meboot) algorithm described in Vinod and López-de-Lacalle (2009) is designed for developing confidence intervals in such situations. Fenga (2020) has used the meboot algorithm to study Covid-19 spread in Italian provinces.

In the present application, the meboot algorithm first creates a large number $J = 999$ of resamples representing an approximate ensemble (population) of analogous time series for both (y_t, z_t). It satisfies the ergodic theorem and has other desirable properties for inference. The observed time series is considered a random sample from the population of J time series, ready for statistical inference.

Next, we forecast y_{jT} for the last value y_T of the dependent variable (new deaths) repeatedly for $j = 1, \ldots J$ resamples. Sorting the J estimates

y_{jT} from the smallest to the largest give their order-statistics denoted by inserting parentheses in the subscript as $y_{(jT)}$. When $J = 999$, the 25-th and 975-th ordered prediction gives a 95% confidence interval. This method deletes the extreme 2.5% estimates at both ends to obtain our 95% confidence interval reported here.

3 Trend Assessment in US Death Forecasts

This section attempts to answer whether the United States is on a downward trend revealed by the parameters of our forecasting model from Vinod and Theiss (2020b). Denote cumulative death count by Y_{it}, lagged cumulative infection count by $X_{i,t-1}$, and let IMR_{it} denote the inverse Mills ratio correction for the data bias arising from the fact that Covid-19 testing is rationed and hence nonrandom. Now the conditional expectation is:

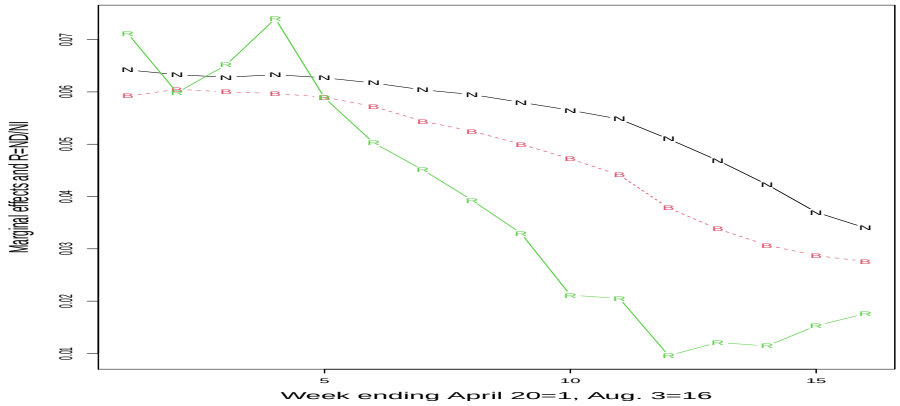
$$E(Y_{it} \mid X_{i,t-1}\,, IMR_{it}) = \beta_0 + \beta_1 log(X_{i,t-1}) + \beta_2(IMR_{it}). \tag{2}$$

We have fitted these models for many weeks, starting with the week ending April 20, 2020. One version retains the bias correction term IMR, and the other version (without IMR) assumes $\beta_2 = 0$. The estimation of parameters uses a Generalized Linear Model (GLM) with the Poisson distribution providing a so-called 'link function' to predict Covid-19 deaths based on the (logs of) cumulative infections from the previous week. Instead of ordinary least squares (OLS), the R software has 'glm' function, which uses maximum likelihood (ML) for parameter estimation to ensure that predicted deaths are nonnegative. Leeper (2018) provides tools for computing 'marginal effect' measured by new deaths from new infections.

Figure 1 reports the path of marginal effects over the entire range of our model estimates. We have two paths marked 'B' for models with the IMR bias correction and 'N' for no correction. Both paths have a downward direction suggesting that the US is on the right path. The third path is the ratio of new deaths per 100 new infections (ND/NI) and marked 'R' for ratio.

New deaths per 100 new infections have come down from a high of 7 deaths in April to less than 2 by August 3, implying that the virus is becoming less lethal over the 16 weeks as doctors and hospitals are learning to manage and treat the novel virus. As the lock-downs are ending in most states, the ratio has inched up in recent weeks.

Figure 1: US-wide Downward Trend in Marginal Effect



4 Forecasts for Aug. 10 made on Aug. 3

This section reports three tables analogous to the Tables in Vinod and Theiss (2020a). The following three tables provide detailed forecasts for each state, with identical names for the reported columns. The column entitled 'Party' reports state governor's party affiliation (democrat or republican).

Our estimates of new deaths are derived from the national model designed for predicting cumulative deaths from the logarithm of lagged cumulative infections. They are obtained by subtraction. Let $\hat{C}uD_T$ denote the latest week's cumulative death prediction for a state, and let $\hat{C}uD_{T-1}$ denote the cumulative deaths predicted for that state for the previous week. Thus, our preliminary prediction of new deaths for each state is computed as: $max(0, (\hat{C}uD_T - \hat{C}uD_{T-1}))$, where we prevent negative estimates by construction.

The column entitled 'Party' reports state governor's party affiliation (democrat or republican).

The third to fifth columns report confidence intervals using the R package for maximum entropy bootstrap (meboot) to construct 95% confidence intervals around forecasts. Column 'LowLim' gives the lower limit of the confidence interval, representing a best-case scenario for new deaths. Column 'UpLim' gives the upper limit of the confidence interval, representing a worst-case scenario for new Covid-19 deaths.

The column entitled 'Mid' provides our adjusted forecasts for each state for the week ending Aug. 10, using information up to Aug. 3. If the bootstrap does not report any adjusted forecast in the 'Mid' column for technical reasons, we insert the midpoint of the two confidence limits.

The column entitled 'ND/NI' reports the ratio of new deaths per 100 new infections. These are data numbers for the previous week without any involvement of our models.

The columns entitled 'Err803' in Tables 1 to 3 report the difference forecast error defined by subtracting actual new deaths during the week ending Aug. 3 from those forecast by our model.

Table 1: Adjusted forecasts of new deaths for the week ending Aug. 10, 2020. Col. 'Mid' has forecasts within a 95% interval, ND/NI is 100 times ratio of new deaths to new infections during the week ending Aug. 3

No.	State	Party	LowLim	Mid	UpLim	ND/NI	Err803
1	AK	Rep	4	4	5	0.5	-1
2	AL	Rep	125	167	201	1.2	63
3	AR	Rep	39	50	69	1.3	-24
4	AZ	Rep	260	490	540	3	55
5	CA	Dem	708	822	978	1.7	-296
6	CO	Dem	42	52	62	0.9	21
7	CT	Dem	28	68	107	1.8	46
8	DC	Dem	7	10	18	0.9	7
9	DE	Dem	22	33	37	0.9	31
10	FL	Rep	982	1132	1411	2.1	-284
11	GA	Rep	232	282	337	1.4	-12
12	HI	Dem	0	0	0	0	2
13	IA	Rep	23	29	36	1.5	-20
14	ID	Rep	27	51	54	1.6	-23
15	IL	Dem	96	190	284	1.1	50
16	IN	Rep	58	92	125	1.3	32
17	KS	Dem	24	28	34	1.1	-8

Table 2: Adjusted forecasts of new deaths for the week ending Aug. 10, 2020. Col. 'Mid' has forecasts within a 95% interval, ND/NI is 100 times ratio of new deaths to new infections during the week ending Aug. 3

No.	State	Party	LowLim	Mid	UpLim	ND/NI	Err803
18	KY	Dem	35	37	41	0.9	3
19	LA	Dem	178	276	326	2.2	-2
20	MA	Rep	61	155	249	4.1	68
21	MD	Rep	44	87	129	1.2	3
22	ME	Dem	3	3	4	3.6	-2
23	MI	Dem	41	88	136	1.1	44
24	MN	Dem	24	41	58	0.8	7
25	MO	Rep	40	61	69	0.5	1
26	MS	Rep	133	158	208	2.6	-65
27	MT	Dem	4	11	18	2	-11
28	NC	Dem	122	143	186	1.6	-74
29	ND	Rep	5	7	8	0.5	1
30	NE	Rep	7	14	15	0.8	2
31	NH	Rep	7	10	13	4	1
32	NJ	Dem	446	580	714	1.5	601
33	NM	Dem	37	41	45	2	4
34	NV	Dem	66	107	111	1.5	-19

Table 3: Adjusted forecasts of new deaths for the week ending Aug. 10, 2020. Col. 'Mid' has forecasts within a 95% interval, ND/NI is 100 times ratio of new deaths to new infections during the week ending Aug. 3

No.	State	Party	LowLim	Mid	UpLim	ND/NI	Err803
35	NY	Dem	70	160	250	1.2	98
36	OH	Rep	145	187	235	2.2	-30
37	OK	Rep	50	54	67	0.9	-16
38	OR	Dem	23	29	41	1.7	-11
39	PA	Dem	80	149	218	1.5	85
40	RI	Dem	10	18	26	0.8	9
41	SC	Rep	174	287	353	2.7	32
42	SD	Rep	9	10	12	2.1	-3
43	TN	Rep	103	123	149	0.8	15
44	TX	Rep	550	1338	1968	2.3	383
45	UT	Rep	19	27	34	1.1	3
46	VA	Dem	77	140	162	1.9	-44
47	VT	Rep	0	0	0	4	-1
48	WA	Dem	50	58	62	1.7	-47
49	WI	Dem	40	48	53	0.9	-25
50	WV	Rep	5	12	13	1.2	-6
51	WY	Rep	1	1	1	0.6	0
sum			5336	7960	10272	80.80	643
mean			104.63	156.08	201.41	1.58	12.61

The last two rows of Table 3 suggest the following. Our model predicts 7960 deaths during the week ending Aug. 10, 2020, with a 95% range of [5336, 10272]. The average ND/NI ratio is 1.58 for the week ending Aug. 3, close to the 1.60 of the previous week's average for July 27.

Our forecasting has been only one-week ahead. Next section reports tentative long term forecasts of Covid-19 deaths till the week ending Monday, November 2, 2020

5 Weekly US New Death Forecasts Till Election

Long term forecasts are generally unreliable. The best-fitting autoregressive integrated moving average (ARIMA) for forecasting new deaths in the US is the random walk or ARIMA(0,1,0). A random walk suggests that the best forecast for new deaths for the week ending Aug. 10 equals the latest value on August 3. Even though the ratio of new deaths to 100 new infections is trending down in Figure 1, the best fitting model for forecasting the ratio is also a random walk.

Fortunately, the best fitting model to forecast new infections is ARIMA(2, 1,0), offering meaningful h-step ahead forecasts. We use these forecasts and multiply by the random walk forecast of the ND/NI ratio to forecast new deaths till November 2, the day before elections. This will require 13-week ahead forecasting based on data till the week ending Aug. 3, to do all forecasts. Figure 2 shows the path followed by the ARIMA model in the portion to the left of the vertical line referring to the week ending Aug. 3. Note that the curve to the right of the vertical line represent long-term model projections. The path of new infections during the future weeks appears to be visually plausible.

Figure 2: US-wide New Infection from ARIMA(2,1,0) model. Model uses data to the left of the vertical line.

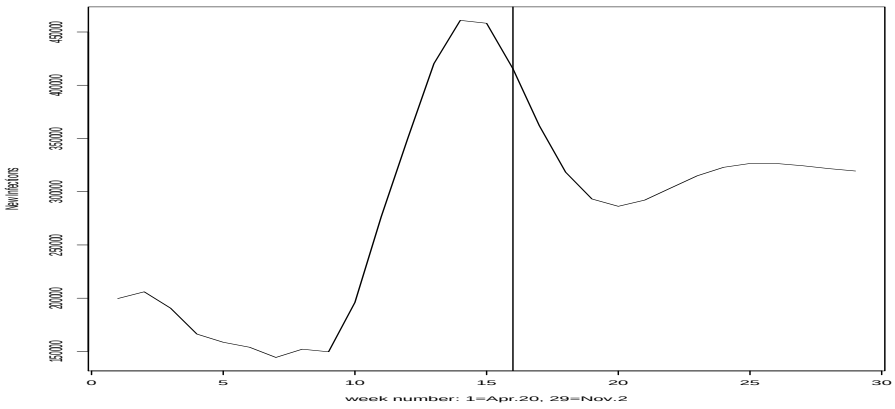


Table 4 has our long-term forecasts till the US elections. The first column

has dates in 'mdd' format. That is, the first row '810' refers to the forecast for the week ending August 10, 2020. The 13-th row marked '1102' refers to the week ending November 2, 2020. The columns entitled 'NI' and 'ND' have new infections and new deaths during that week.

Recall the last line of Table 3, where we report the averge of ND/NI, or new deaths per 100 new infections to be 1.58. This measures the success of American doctors in treating Covid-19. It is fair to assume that with newer treatments coming on line the average ratio will be 1.5 or smaller in the future. Hence our conservative estimate of new deaths from new infections is given by ND=0.015(NI).

Similar to Section 2 we use the 7-step maximum entropy bootstrap to construct 95% confidence intervals for NI and ND with the column heading 'Lo-' for the lower limit and 'Up-' for the upper limit.

Table 4: US-wide new infection (NI) and new death (ND) forecasts with 95% confidence intervals till November elections

	Date	NI-Lo	NI	NI-Up	ND-Lo	ND	ND-Up
1	810	253370	362091	438963	3801	5431	6584
2	817	250189	318335	438963	3753	4775	6584
3	824	252306	293162	438963	3785	4397	6584
4	831	252306	286281	438963	3785	4294	6584
5	907	250189	292097	438963	3753	4381	6584
6	914	250096	303585	438963	3751	4554	6584
7	921	250096	314972	438963	3751	4725	6584
8	928	250096	322923	438963	3751	4844	6584
9	1005	250096	326515	438963	3751	4898	6584
10	1012	250096	326501	438963	3751	4898	6584
11	1019	250096	324392	438963	3751	4866	6584
12	1026	250096	321687	438963	3751	4825	6584
13	1102	250096	319436	438963	3751	4792	6584

The pandemic must be dealt with by using similar projections at all local levels. The key question for policy makers is whether the weekly additions of new infections (NI) and new deaths (ND) can be managed with the available medical personnel, supplies and health-care facilities.

References

Fenga, L., 2020. Forecasting the CoViD-19 diffusion in Italy and the related occupancy of Intensive Care Units. Technical Report. Italian National Institute of Statistics ISTAT, Rome, Italy 00184. URL: https://www.medrxiv.org/content/10.1101/2020.03.30.20047894v1.full.pdf.

Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.

Leeper, T.J., 2018. margins: Marginal Effects for Model Objects. R package version 0.3.23.

Vinod, H.D., López-de-Lacalle, J., 2009. Maximum entropy bootstrap for time series: The meboot R package. Journal of Statistical Software 29, 1–19. URL: http://www.jstatsoft.org/v29/i05/.

Vinod, H.D., Theiss, K., 2020a. Bias-free Forecast-driven Guidelines for Opening Pandemic-ravaged Economies. Discussion Paper No. 2020-07-27. Department of Economics, Fordham University. URL: http://ssrn.com/abstract=3659766.

Vinod, H.D., Theiss, K., 2020b. A Novel Solution to Biased Data in Covid-19 Incidence Studies. Discussion Paper No. 2020-06-29. Department of Economics, Fordham University. URL: http://ssrn.com/abstract=3637682.