

## Predicting mortality for Covid-19 in the US. Delayed Elasticity Method

### **Luis Ángel Hierro**

Senior Lecturer in Economics  
Department of Economics and Economic History (Former Head)  
Universidad de Sevilla  
Avda. Ramón y Cajal, 1  
41018 Sevilla (Spain)  
Email: [lhierro@us.es](mailto:lhierro@us.es)

### **Antonio José Garzón**

Postgraduate Teaching Assistant in Economics  
Department of Economics and Economic History  
Universidad de Sevilla  
Avda. Ramón y Cajal, 1  
41018 Sevilla (Spain)  
Email: [antoniojgarzon@gmail.com](mailto:antoniojgarzon@gmail.com)

### **Pedro Atienza-Montero (Corresponding Author)**

Lecturer in Economics  
Department of Economics and Economic History  
Universidad de Sevilla  
Avda. Ramón y Cajal, 1  
41018 Sevilla (Spain)  
Email: [atienza@us.es](mailto:atienza@us.es)

### **José Luis Márquez**

Medical Doctor (Ph. Dr.)  
Gastroenterology Unit (Former Head)  
Hospital Universitario Virgen del Rocio  
Avda. Manuel Siurot s/n  
41013-Sevilla (Spain)  
Email: [jlmarquezhuvr@gmail.com](mailto:jlmarquezhuvr@gmail.com)

**Abstract:** The evolution of the pandemic caused by COVID-19, its high reproductive number and the associated clinical needs, is overwhelming national health systems. We propose a method for predicting the number of deaths, and which will enable the health authorities of the countries involved to plan the resources needed to face the pandemic as many days in advance as possible. We employ OLS to perform the econometric estimation, and through RMSE, MSE, MAPE, and SMAPE forecast performance measures we select the best lagged predictor of both dependent variables. Our objective is rather to estimate a leading indicator of clinical needs. Having a forecast model available several days in advance can enable governments to more effectively face the gap between needs and resources triggered by the outbreak and thus reduce the deaths caused by COVID-19.

**Keywords:** COVID-19, predictions, clinical needs, mortality

Predictive models for Covid-19 focus on predicting the evolution of infections through exponential adjustment, without predefined models<sup>1,2</sup>, or with predefined infections<sup>3</sup> and number of deaths models<sup>4</sup>. These models can be highly complex<sup>3</sup>, based on very detailed epidemiological features of the disease studied and can allow clinical strategies to be analyzed, yet their predictive accuracy is low<sup>4</sup>. This makes them ineffective for prediction purposes and prevents health authorities from having reliable predictions in advance.

### Methodology

We propose the following method (Delayed Elasticity Method-DEM). Using officially published data from the the Johns Hopkins University CSSE<sup>5</sup>, we econometrically estimate the following equation:

$$\log(Deaths_t) = \alpha + \beta \log(Cases_{t-i}) + \epsilon_t \quad (1)$$

where  $i=1, 2 \dots, 10$  are the number of delays of the explanatory variable,  $Deaths_t$  is the total number of deaths up to day  $t$ ,  $Cases_t$  is the number of cases detected up to date  $t$ .

The coefficient  $\beta_{-i}$  is what in economics is called elasticity and it represents the relationship between the variation of the dependent and independent variables:

$$\beta_{-i} = \frac{\frac{\Delta \widehat{Deaths}_t}{\widehat{Deaths}_t}}{\frac{\Delta Cases_{t-i}}{Cases_{t-i}}} \quad (2)$$

After estimating equations using different lags, we select the one with the best forecast performance (which minimizes forecasting errors). We calculate RMSE<sup>3</sup> as an indicator of predictive accuracy. Other indicators can be used, such as MAE, MAPE or SMAPE. RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{t=T+1}^{T+N} (\hat{y} - y)^2}{N}} \quad (3)$$

where  $N$  is the number of out-of-sample observations, which we use to estimate the forecast performance of our estimate,  $\hat{y}$  is the estimated value of the dependent variable, and  $y$  is the actual value.

Finally, we select the estimate with the lagged explanatory variable that shows the lowest value in this indicator, and we make the corresponding prediction in total values.

## Results

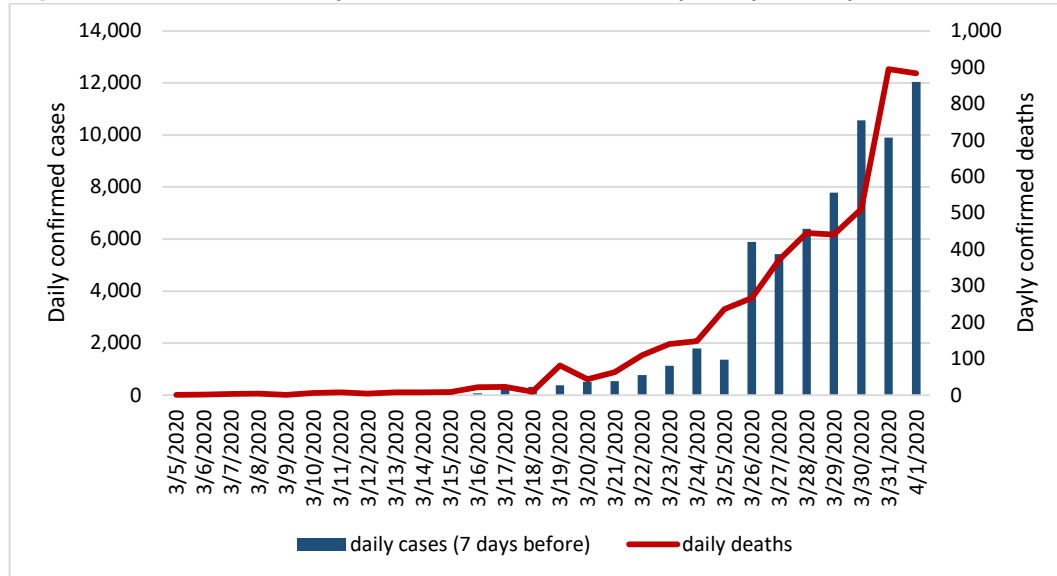
The estimation sample spans from 3/4/2020 to 3/29/2020. We left March 30, 31 and April 1 as out-of-sample observations in order to measure the forecast performance of the estimated model. Estimation is performed through the OLS estimator. We select the model including seven delays, since it shows the lowest RMSE value. The equation of the model that evidences the best forecast performance is the following:

$$\log(\widehat{Deaths}_t) = -0.7543 + 0.8227 * \log(Cases_{t-7}) + \epsilon_t \quad (4)$$

The delayed elasticity,  $\beta = 0.8227$ , implies that a 1% increase in the number of infected cases predicts a 0.82% increase in the number of deaths seven days later. The estimate presents a high goodness of fit, with an R-square of 0.99.

Figure 1 displays the evolution of daily new confirmed deaths and new confirmed cases 7 days before.

**Figure 1.** Evolution of daily confirmed deaths and 7-day delayed daily confirmed cases.



Source: authors' own compilation and Johns Hopkins University CSSE

Table 1 shows the number of actual and estimated deaths, as well as the errors for each time period. To obtain the total number of deaths, we carry out the following transformation

$$\widehat{Deaths}_t = \exp(\log(\widehat{Deaths}_t)) \quad (5)$$

**Table 1.** Actual vs estimated deaths and estimate error.

| Date             | Cases          | Deaths       | Deaths (Est) | Error          | Error %     |
|------------------|----------------|--------------|--------------|----------------|-------------|
| 3/4/2020         | 149            | 11           | 13           | 2.09           | 19.03       |
| 3/5/2020         | 217            | 12           | 13           | 1.28           | 10.68       |
| 3/6/2020         | 262            | 14           | 14           | -0.34          | 2.45        |
| 3/7/2020         | 402            | 17           | 15           | -1.86          | 10.95       |
| 3/8/2020         | 518            | 21           | 16           | -4.77          | 22.72       |
| 3/9/2020         | 583            | 22           | 20           | -1.55          | 7.05        |
| 3/10/2020        | 959            | 28           | 24           | -4.17          | 14.91       |
| 3/11/2020        | 1,281          | 36           | 29           | -7.13          | 19.82       |
| 3/12/2020        | 1,663          | 40           | 39           | -0.67          | 1.68        |
| 3/13/2020        | 2,179          | 47           | 46           | -1.08          | 2.29        |
| 3/14/2020        | 2,727          | 54           | 65           | 11.32          | 20.95       |
| 3/15/2020        | 3,499          | 63           | 80           | 17.46          | 27.72       |
| 3/16/2020        | 4,632          | 85           | 89           | 3.68           | 4.33        |
| 3/17/2020        | 6,421          | 108          | 134          | 25.56          | 23.67       |
| 3/18/2020        | 7,783          | 118          | 169          | 51.48          | 43.63       |
| 3/19/2020        | 13,677         | 200          | 210          | 10.07          | 5.04        |
| 3/20/2020        | 19,100         | 244          | 262          | 18.38          | 7.53        |
| 3/21/2020        | 25,489         | 307          | 316          | 8.56           | 2.79        |
| 3/22/2020        | 33,276         | 417          | 387          | -29.61         | 7.10        |
| 3/23/2020        | 43,847         | 557          | 488          | -69.04         | 12.40       |
| 3/24/2020        | 53,740         | 706          | 638          | -67.63         | 9.58        |
| 3/25/2020        | 65,778         | 942          | 748          | -194.16        | 20.61       |
| 3/26/2020        | 83,836         | 1,209        | 1,189        | -19.81         | 1.64        |
| 3/27/2020        | 101,657        | 1,581        | 1,565        | -15.75         | 1.00        |
| 3/28/2020        | 121,478        | 2,026        | 1,985        | -41.33         | 2.04        |
| 3/29/2020        | 140,886        | 2,467        | 2,471        | 4.40           | 0.18        |
| <b>3/30/2020</b> | <b>161,807</b> | <b>2,978</b> | <b>3,101</b> | <b>123.09</b>  | <b>4.13</b> |
| <b>3/31/2020</b> | <b>188,172</b> | <b>3,873</b> | <b>3,666</b> | <b>-206.86</b> | <b>5.34</b> |
| <b>4/1/2020</b>  | <b>213,372</b> | <b>4,757</b> | <b>4,329</b> | <b>-427.56</b> | <b>8.99</b> |
| <b>4/2/2020</b>  |                |              | <b>5,286</b> |                |             |
| <b>4/3/2020</b>  |                |              | <b>6,194</b> |                |             |
| <b>4/4/2020</b>  |                |              | <b>7,172</b> |                |             |
| <b>4/5/2020</b>  |                |              | <b>8,102</b> |                |             |
| <b>4/6/2020</b>  |                |              | <b>9,079</b> |                |             |

Source: authors' own compilation and Johns Hopkins University CSSE. Deaths(est) values are rounded to integer values. Out-of-sample dates in bold.

### Discussion

The result of the DEM is a model with a very high predictive accuracy, offering a seven-day advance for deaths in the US. This advanced forecast is what can allow the authorities to improve resources planning to face the clinical needs-health resources gap caused by the epidemic.

The DEM is replicable to any country, state, region, city or hospital area, and is applicable to other clinical situations apart from deaths and can also be applied to anticipate health needs (ICUs, hospital beds...).

The DEM does, nevertheless, evidence certain shortcomings that must be taken into account and that basically affect available data. It only uses one independent variable; detected cases. If the number of tests for possible Covid-19 infected cases were to be increased, the model would tend to overestimate future deaths and would need to be recalibrated. The same effect would be caused by applying new clinical strategies that reduce the fatality ratio. These effects can be corrected by performing a re-estimation when the model loses its predictive accuracy.

**Author Contributions:** Luis Ángel Hierro: study design, literature search (Google Scholar) and writing of the Summary, Research in Context, Introduction, and Discussion. Antonio José Garzón: econometric treatment, data analysis, and writing of the Methods and Results. Pedro Atienza: data collection, data analysis, Figures, Tables, References and corresponding author. José Luis Márquez: literature search (PubMed), evaluation of the impact of the results, revision of the text and structure for adaptation to J. Clin. Med., and adaptation of the language to medical scientific terminology. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- 1 Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *The Lancet* 2020; published online Mar 13. DOI: 10.1016/S0140-6736(20)30627-9
- 2 Jung SM, Akhmetzhanov AR, Hayashi K, et al. Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: Inference using exported cases. *Journal of Clinical medicine* 2020; 9(2):523. Published online Feb 14. DOI: 10.3390/jcm9020523
- 3 Al-qaness MA, Ewees AA, Fan H, Abd El Aziz M. Optimization Method for Forecasting Confirmed Cases of COVID-19 in China. *Journal of Clinical Medicine* 2020; 9(3):674. Published online Mar 2. DOI:10.3390/jcm9030674.
- 4 Flaxman S, Mishra S Gandy S et al. Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. 30 March 2020 - Imperial College London <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/COVID-19/> (accessed March 31, 2020)
- 5 Dong, Ensheng, Hongru Du, and Lauren Gardner. "An interactive web-based dashboard to track COVID-19 in real time." *The Lancet Infectious Diseases* (2020).

### Additional information

#### Other forecast performance indicator:

Mean Absolute Error (MAE):

$$MAE = \frac{\sum_{t=T+1}^{T+N} |\hat{y}_t - y_t|}{N} \quad (2A)$$

Mean Absolute Percentage Error (MAPE):

$$MSE = \frac{\sum_{t=T+1}^{T+N} (\hat{y}_t - y_t)^2}{N} \quad (3A)$$

Symmetric Mean Absolute Percentage Error (SMAPE):

$$SMAPE = \frac{\sum_{t=T+1}^{T+N} |\hat{y}_t - y_t| / y_t}{(|y_t| + |\hat{y}_t|) / 2} \times \frac{100}{N} \quad (4A)$$

**Table A.1.** Out-of-sample forecast performance for deaths

| Lags     | RMSE          | MAE           | MAPE        | SMAPE       |
|----------|---------------|---------------|-------------|-------------|
| 1        | 1462.45       | 1382.47       | 41.25       | 52.37       |
| 2        | 1284.91       | 1206.2        | 35.84       | 44.03       |
| 3        | 1075.51       | 1006.51       | 29.85       | 35.34       |
| 4        | 885.1         | 823.97        | 24.36       | 27.92       |
| 5        | 656.56        | 601.78        | 17.62       | 19.45       |
| 6        | 357.85        | 280.45        | 7.59        | 8.04        |
| <b>7</b> | <b>216.25</b> | <b>197.38</b> | <b>6.68</b> | <b>6.42</b> |
| 8        | 646.64        | 635.86        | 19.59       | 17.84       |
| 9        | 1117.68       | 1080.37       | 32.75       | 28.11       |
| 10       | 1545.14       | 1465.83       | 43.82       | 35.83       |

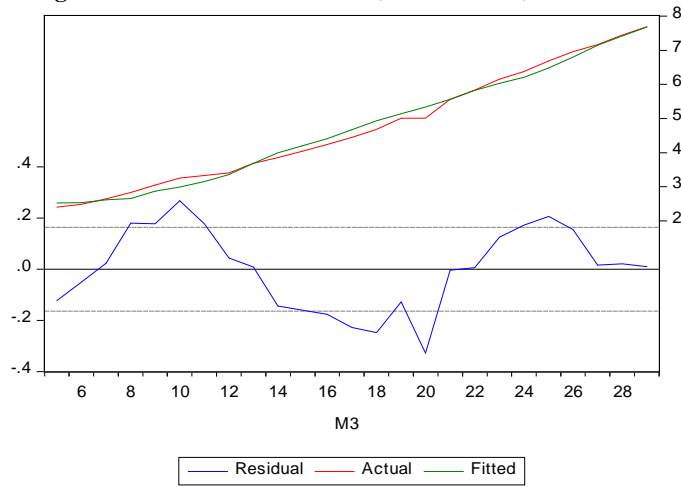
Sources: authors' own compilation. Selected lag in bold.

**Table A2.** Estimation output for equation (1)

| Variable           | Coefficient | Std. Error            | t-Statistic | Prob.  |
|--------------------|-------------|-----------------------|-------------|--------|
| LOG(CASES(-7))     | 0.843878    | 0.016864              | 50.03987    | 0.0000 |
| C                  | -0.920055   | 0.117526              | -7.828543   | 0.0000 |
| R-squared          | 0.990898    | Mean dependent var    | 4.728880    |        |
| Adjusted R-squared | 0.990503    | S.D. dependent var    | 1.677035    |        |
| S.E. of regression | 0.163435    | Akaike info criterion | -0.708180   |        |
| Sum squared resid  | 0.614356    | Schwarz criterion     | -0.610670   |        |
| Log likelihood     | 10.85225    | Hannan-Quinn criter.  | -0.681134   |        |
| F-statistic        | 2503.989    | Durbin-Watson stat    | 0.482811    |        |
| Prob(F-statistic)  | 0.000000    |                       |             |        |

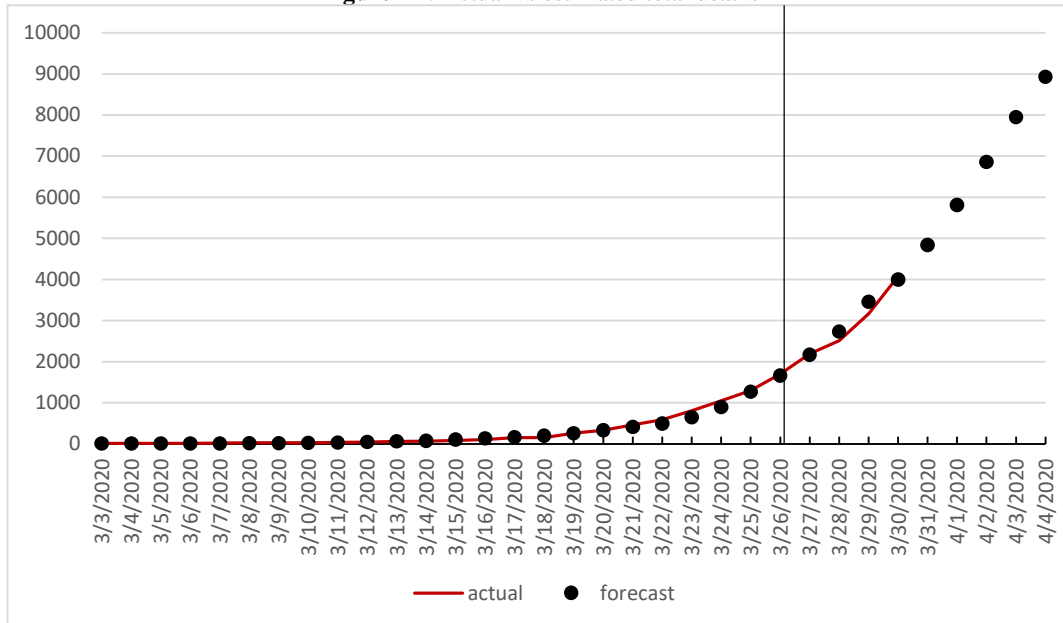
Source: Authors' own compilation

**Figure A2.** Model estimate values, actual values, and residuals



Source: authors' own compilation

**Figure A1.** Actual vs estimated total deaths



Source: authors' own compilation