

# UBrisk method to predict COVID-19 Spain

Montserrat Guillen\*

Dept. Econometría, Estadística y Economía Aplicada,  
Riskcenter-IREA,  
Universitat de Barcelona

April 2, 2020

## Resumen

Summary of the method used by UBriskcenter to predict five indicators by region and in Spain. The results are forwarded daily to the cooperative predictor of CEMAT (Comité Español de Matemáticas) at MODES.

## 1. Introduction

This is the first draft version to explain the method. Details will follow.

Data provided by: `covid19.isciii.es/resources/serie_historica_acumulados.csv`.

We aim to predict cumulated total cases of emergency room (*uci*), hospitalizations (*hospitalizados*), deaths (*fallecidos*). We also aim to predict new cases (*nuevos*) and total cumulated confirmed cases (*confirmados*). Predictions are for the forthcoming days 1 to 7. These predictions are updated daily. The results are sent to CEMAT, `prediccion.covid19@udc.gal`, via file `UBRisk_mm_dd.xlsx`. Where `mm_dd` is the month and day of the available data. Predictions apply to the next seven days.

## 2. Method

Notation  $t$  is time in days, starting 12/03/2020,  $i$  is the index for Autonomous Community denoted a CCAA (there are 19 CCAA),  $s$  subindex refers to Spain. Jointpoint regression was used in [1]. It seems that a quadratic shape emerges for cases. This principle is applied to new cases and daily deaths.

Indicators are defined in Table 1. For region  $i$  and time  $t$ ,  $N_{jit}$  is the observed cumulated count indicator  $j = 1, 2, 3, 5$ . Note that we want to predict four cumulative counts and also an estimate of new cases,  $n_{4t}$  *nuevos casos*, which is not cumulated. We observed up to day  $T$ , we will predict  $T + 1, T + 2, \dots, T + 7$ .

---

\*

Cuadro 1: Indicators

variable	j	notation	cumulated
uci	1	$N_{1it}$	YES
hospitalizados	2	$N_{2it}$	YES
fallecidos	3	$N_{3it}$	YES
nuevos	4	$n_{4it}$	NO
confirmados	5	$N_{5it}$	YES

We introduce the data on population in each CCAA to analyse model results later in a comparable scale. Let  $pop_i$  be the population of each CCAA,  $pop_s$  is the population in Spain. These data were taken from the INE [https://www.ine.es/prensa/cp\\_j2019\\_p.pdf](https://www.ine.es/prensa/cp_j2019_p.pdf).

## 2.1. Fitting curve to new cases and deaths by CCAA and to Spain

Consider a linear model with the dependent variable  $\log((n_{4it} + 1)/pop_i)$ , and the linear predictor as:

$$\beta_0^{(i)} + \beta_1^{(i)}t + \beta_2^{(i)}t^2,$$

for  $i = 1, \dots, 19$ . Similarly for Spain.

For the series of log-transformed relative population adjusted new cases ( $\log((n_{4it} + 1)/pop_i)$ ), we fit a quadratic shape for each CCAA. We use OLS and obtain parameter estimates and their variance-covariance matrix.

From the series of log-transformed relative population adjusted cumulated deaths ( $N_{3it}$ ), we first find daily counts as:

$$n_{3it} = N_{3it} - N_{3i(t-1)}.$$

Similarly, for the series of log-transformed relative daily deaths ( $\log((n_{3it} + 1)/pop_i)$ ), not cumulated deaths, we fit a quadratic curve. We use OLS and obtain parameter estimates and their corresponding variance-covariance matrix.

Parameter estimates of the curves have an interpretation in terms of relative speed in the advance of the epidemic, turn point and time to certain events. Comparisons can be made between CCAA, due to the fact that all series are relative to the region population.

## 2.2. Prediction of cumulated cases and new cases by CCAA and for Spain

Point prediction  $\hat{n}_{4i(T+h)}$  are obtained for  $h = 1, 2, \dots, 7$  from OLS log-linear regression as:

$$\exp(\hat{\beta}_0^{(i)} + \hat{\beta}_1^{(i)}(T+h) + \hat{\beta}_2^{(i)}(T+h)^2) \times pop_i - 1,$$

for  $i = 1, \dots, 19$  and  $h = 1, \dots, 7$ , where  $\hat{\beta}_k^{(i)}$ ,  $k = 0, 1, 2$  are obtained via OLS. Similarly for Spain. A corrector may be applied latter to sum to Spain predicted totals.

Prediction of cumulated cases is obtained as follows:

$$\hat{N}_{5i(T+h)} = N_{5iT} + \sum_{m=1}^h \hat{n}_{4i(T+m)},$$

for  $h = 1, \dots, 7$ .

## 2.3. Prediction of cumulated deaths by CCAA and to Spain

Point predictions  $\hat{n}_{3it}$  are obtained for  $t = T+1, T+2, \dots, T+7$  from the corresponding OLS log-linear regression as before. Corrections may be applied latter to sum to Spain predicted totals.

Prediction of cumulated deaths is obtained as follows for each  $i$ -th region and s:

$$\hat{N}_{3i(T+h)} = n_{3iT} + \sum_{m=1}^h \hat{n}_{3i(T+m)}, \quad h = 1, \dots, 7.$$

## 2.4. Prediction of other indicators

Simple linear regressions are fitted for  $N_{1it}$  as a function of  $N_{2it}$ ,  $F_1^{(i)}$ ,  $N_{2it}$  as a function of  $N_{5it}$ ,  $F_2^{(i)}$ , for each  $i$ -th region and for Spain, subscript  $s$ . These are currently highly linearly correlated.

Predictions are updated sequentially as:

$$\hat{N}_{2i(T+h)} = F_2^{(i)}(\hat{N}_{5i(T+h)}), \quad h = 1, \dots, 7.$$

Then,

$$\hat{N}_{1i(T+h)} = F_1^{(i)}(\hat{N}_{2i(T+h)}), \quad h = 1, \dots, 7.$$

## 3. Basis of the epidemiological curve

Cumulated cases are the integral of the so called “epidemiological curve”. We estimated daily the basis of the curve and monitor its trends. We do not

estimated the shape of the curve, but rather the time to reach a certain % of the population. The larger the length of the basis, the lower the kurtosis.

Estimates, currently not forwarded to the cooperative group are calculated daily for one region. Interestingly, the estimated length of the basis is slightly increasing each day. This denotes that the pace of contagions has substantially reduced, and it is consistent with current  $R0$  estimates obtained elsewhere.

## Referencias

- [1] Al Hasan SM, Saulam J, Kanda K and Hirao T. The novel coronavirus disease (COVID-19) outbreak trends in mainland China: a joinpoint regression analysis of the outbreak data from January 10 to February 11, 2020. [Submitted]. Bull World Health Organ. E-pub: 17 February 2020. doi: <http://dx.doi.org/10.2471/BLT.20.253153>