UBrisk method to predict COVID-19 Spain

Montserrat Guillen[*]
Dept. Econometra, Estadstica y Economa Aplicada,
Riskcenter-IREA,
Universitat de Barcelona

April 1, 2020

Abstract

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 (Comit Espaol de Matemticas) at MODES.

1 Introduction

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

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

We aim to predict cumulated cases (forthcomming days 1 to 7) daily and provide these predictions to CEMAT, `prediccion.covid19@udc.gal`.

2 Method

Notation t is time in days, starting 20/02/2020, i is the index for Autonomous Community denoted a CCAA (there are 17 CCAA), s subindex refers to Spain. Jointpoint regression was used in [1].

Indicators in Table 1. n_{jit} is the observed number of cumulated cases of each indicator $j = 1, ..., 5$ (note that n_{4t} *nuevos casos* is not cumulated), for region i and time t. We observed up to day T, we will predict $T + 1, T + 2, ...T + 7$.

We introduce the data on population in each CCAA to make results comparable. pop_i is the population of each CCAA, pop_s is the population in Spain.

---

[*]

Table 1: Indicators

uci
hospitalizados
fallecidos
nuevos
confirmados

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

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

$$\beta_0 + \beta_1 t + \beta_2 + t^2$$

Similarly for Spain.

For the series of new cases ($j = 4$), we fit only a quadratic shape. We fit using OLS and obtain parameter estimates.

For the series of deaths ($j = 3$), not cumulated deaths, we fit only a quadratic shape. We fit using OLS and obtain parameter estimates.

2.2 Prediction of cumulated and new cases by CCAA and to Spain

Point prediction $\hat{n}_{4it})$ is obtained as $t = T + 1, T + 2, ...T + 7$ from OLS log-linear regression. Corrector will be applied latter to sum to Spain predicted total.

Prediction of cumulated cases is obtained as follows:

$$\hat{n}_{5i(T+k)} = n_{5iT} + \sum_{k=1}^{7} \hat{n}_{4i(T+k)}.$$

2.3 Prediction of deaths by CCAA and to Spain

Point prediction $\hat{n}_{3it})$ is obtained as $t = T + 1, T + 2, ...T + 7$ from OLS log-linear regression. Corrector will be applied latter to sum to Spain predicted total.

Prediction of cumulated deaths is obtained as follows:

$$\hat{N}_{3i(T+k)} = n_{3iT} + \sum_{k=1}^{7} \hat{n}_{3i(T+k)}.$$

2.4 Prediction of other indicators

Linear regressions are fitted for n_{1it} as a function of n_{2it}, $\hat{F}_1$, n_{2it} as a function of n_{5it}, $\hat{F}_2$.

Predictions are updated sequentially as:

$$\hat{n}_{2i(T+k)} = \hat{F}_2(\hat{n}_{5i(T+k)}).$$

Then,

$$\hat{n}_{1i(T+k)} = \hat{F}_1(\hat{n}_{2i(T+k)}).$$

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.

References

[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