Predicting The Incidence Rate And Case Fatality Rate Of The Novel Coronavirus SARS-CoV-2

by
Mike Weltevrede (SNR 1257560)

A thesis submitted in partial fulfillment of the requirements for the degree of Master in Econometrics and Mathematical Economics.

Tilburg School of Economics and Management
Tilburg University

Supervised by:
dr. Otilia Boldea (Tilburg University)
dr. George Knox (Tilburg University)

Date:
April 3, 2020

Contents

1 Acknowledgements

2 Management summary

3 Introduction

4 Problem description

We will use open source data on Italian regions. Adda uses the following specification to model the incidence rate $Inc_{r,t}$ for several viruses, being the percentage of the population in a region r who have the virus at a time t:

$$\begin{aligned} Inc_{r,t} = {} & Inc_{r,t-lag} S_{r,t-lag} \sum_{k=1}^{K} a^k_{within} W^k_{r,t-lag} \\ & + \sum_{c \neq r} Inc_{c,t-lag} S_{r,t-lag} \sum_{k=1}^{\tilde{K}} a^k_{between} \widetilde{W}^k_{r,c,t-lag} \\ & + X_{r,t}\delta + \eta_{r,t} \end{aligned} \tag{1}$$

Adda models the susceptible population as the total population who currently do not have the virus and who are not immune. That is, let S denote the fraction of individuals who are susceptible to contracting the disease, I the fraction of individuals who are infected, and R the fraction of individuals who have recovered but are still immune. Then:

$$\begin{cases} \frac{dI(t)}{dt} = \alpha S(t) I(t) - \beta I(t) \\ \frac{dR(t)}{dt} = \beta I(t) - \lambda R(t) \\ \frac{dS(t)}{dt} = -\alpha S(t) I(t) + \lambda R(t) \end{cases}$$

For simplification purposes, we will neglect that people lose their immunity, i.e. $\lambda = 0$. That is, we define

$$\begin{cases} \frac{dI(t)}{dt} & = \alpha S(t) I(t) - \beta I(t) \\ \frac{dR(t)}{dt} & = \beta I(t) \end{cases}$$

In this thesis, we are interested in modelling the growth rate of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). As such, the dependent variable changes compared to Adda. We define the growth rate as:

$$G_{r,t} = \frac{\text{The number of new cases}}{\text{The number of active cases}}.$$

Firstly, we will also neglect interactions between regions. The model becomes:

$$G_{r,t} = G_{r,t-lag} S_{r,t-lag} \sum_{k=1}^{K} a_{within}^{k} W_{r,t-lag}^{k} \\ + X_{r,t}\delta + \eta_{r,t} \tag{2}$$

For us, we will use the following specifications for the regressors:

- $W_{r,t-lag}$ contains K region-specific variables that potentially influence the transmission rate of SARS-CoV-2 within a region r. We split these in several categories:

  Economic

  - The amount of freight being transported by plane from and to the region (not available interregionally).
  - The amount of freight being transported by ship from and to the region (not available interregionally).
  - The amount of arrivals at tourist accommodations.
  - The GDP at current market prices per inhabitant.
  - The disposable income per inhabitant.
  - The amount of journeys made for transport of freight by road by loading and unloading region.

  Demographics, social, etcetera

  - The area size.
  - The median age and median age squared.
  - The population number.
  - The percentage of people at risk of poverty or social exclusion.
  - The percentage of people with broadband access.
  - The percentage of people who used internet to contact the public authorities in the last year.
  - The percentage of people that attained a certain education level.

  Medical

  - The average length-of-stay in a hospital.
  - The crude death rate for several different diseases.
  - The number of health personnel (doctors and nurses).
  - The number of hospital beds.

  Travelling

  - The number of passengers travelling by plane from and to the region (not available interregionally).
  - The number of passengers travelling by ship from and to the region (not available interregionally).
  - The length of railroads, motorways, navigable rivers, etcetera.

- $X_{r,t}$ contains certain fixed effects to control for, such as a binary indicator whether the day was on a weekend.

When we will also consider interactions between regions, we will define $\widetilde{W}_{r,t-lag}$ to contain $\tilde{K}$ variables that potentially influence the transmission rate of SARS-CoV-2 across regions:

- Amount of passengers that travelled from region c to region r via railroad.
- Amount of freight that travelled from region c to region r via railroad.
- A binary indicator indicating whether the regions border each other.
- The distance between the largest (most populous) cities in the regions.
- The population ratios.
- The log regional GDP ratios.

5 Materials

6 Results

7 Conclusion

References

A Tables