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 7, 2020

Contents

1 Acknowledgements

2 Management summary

3 Introduction

4 Problem description

The following specification is used by Adda 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:

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

In this thesis, we are interested in modelling the growth rate of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The growth rate is defined as follows:

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

Notice that Adda also models interaction between regions using the matrix $\widetilde{W}_{r,c}$. At first, we will neglect interactions between regions. The model becomes:

$$\begin{aligned} G_{r,t} = {} & G_{r,t-lag} S_{r,t-lag} \sum_{k=1}^{K} a^k_{within} W^k_{r,t-lag} \\ & + X_{r,t}\delta + \eta_{r,t} \end{aligned} \tag{2}$$

Note that the growth rate is a number between 0 and 1. As such, regular regression techniques are not applicable. A common censored model is the Tobit model. The following assumptions are made for the panel data Tobit model:

(I) The r-units ($r = 1, \ldots, R$) are a random sample. R should be large but T can be small.

(I) $G^*_{r,t} = x'_{r,t}\beta + \alpha_r + \epsilon_{r,t}$ and $G_{r,t} = min(max(0, G^*_{r,t}), 1)$.

(I) $\epsilon_{r,t} \sim N(0, \sigma^2)$, independent of $x_{r,t}$ for all t.

(I) $\alpha_r \sim N(0, \sigma^2_\alpha)$, independent of $x_{r,t}$ and $\epsilon_{r,t}$ for all t.

Limitations of the model are as follows:

- The size of our sample R is not large: $R = 21$, i.e. assumption (I) is violated.
- There is no reason to believe that all the regions are independent, i.e. assumption (I) is violated. For instance, if two regions are highly connected, it is likely that there is dependence. It may be that regions are independent, however, if they are not close to one another.

The spatial weighting matrix W_r has the following structure:

$$W_r = \begin{bmatrix} V_r & C_r \end{bmatrix},$$

where V_r consists of K_V time-varying regressors and C_r consists of K_C time-constant regressors, so $V_r \in \mathbb{R}^{T \times K_V}$ and $V_r \in \mathbb{R}^{T \times K_C}$. Taking an example:

$$W_r = \begin{bmatrix} V_r^{\text{schools closed}} & V_r^{\text{lockdown started}} & C_r^{\text{hospital beds}} & C_r^{\text{internet access}} \end{bmatrix}.$$

A description of the data that we use can be found in section 5. We note that the descriptive data (like demographics and economic data) that we use is assumed to be time-constant during the coronacrisis (due to lack of data). The time-varying information that we use consists binary indicators for whether certain policies (such as closing down schools or instigating a lockdown) were implemented. As such, W_r mostly contains time-constant information.

5 Materials

We will use the following specifications for the weights and regressors:

- $W_{r,t-lag}$ contains $K \coloneqq K_V + K_C$ 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 $\widetilde{K}$ 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.

6 Results

7 Conclusion

References

Adda, J. (2016). Economic activity and the spread of viral diseases: Evidence from high frequency data. *The Quarterly Journal of Economics*, *131*(2), 891–941.

A Tables