

# Modelling coronavirus with spatial proximity effects

Bernard Fingleton

Department of Land Economy, Silver Street, University of Cambridge, CB3 9EP, UK

Version 1

Very initial draft discussion paper. Comments and suggestions welcome

20/03/2020

## The model

The model used here (see Baltagi et al 2014, 2019) is summarised by the equation

$$\ln c_t = \gamma \ln c_{t-1} + \rho_1 \mathbf{W}_N \ln c_t + \beta \ln p_t + \theta \mathbf{W}_N \ln c_{t-1} + \varepsilon_t \quad (1)$$

Where  $c_t$  is the number of cases across  $N = 201$  countries at time  $t$  and  $p_t$  is the population across 201 countries at time  $t$ . In this case  $t$  represents months.  $\mathbf{W}_N$  is a (scaled)  $N$  by  $N$  connectivity matrix with zeros on the main diagonal, where

$$w_{ij}^* = d_{ij}^{-1}; \forall i, j, i \neq j \quad (2)$$

In which  $w_{ij}^*$  is the  $i, j$ th cell of matrix  $\mathbf{W}_N^*$  and  $d_{ij}^{-1}$  is the reciprocal of the great circle distance between countries  $i$  and  $j$ .  $\mathbf{W}_N$  is  $\mathbf{W}_N^*$  divided by its maximum absolute real eigenvalue. This means that  $\rho_1, \rho_2 < 1$  leads to a dynamically stable process as described below.

The model is saying that the number of cases in country  $i$  at time  $t$  depends on the number in the previous month, the population of the country, the number of cases in nearby countries, and the number of cases in nearby countries in the previous month. Obviously there are other factors, we call these unobservables, captured by  $\varepsilon_t$  which is a vector of random effects picking up unobservable time-invariant heterogeneity across countries (the net effect of many aspects of culture, climate etc etc), denoted by  $\mu_i, i = 1, \dots, N$ . In addition, it picks up idiosyncratic time and country varying shocks, denoted by  $v_{it}$ .

Also as is utilised in recent literature (Baltagi et al, 2019), we assume a spatial moving average process causing error dependence across countries, thus

$$\begin{aligned} \varepsilon_t &= u_t - \rho_2 \mathbf{W}_N u_t \\ u_{it} &= \mu_i + v_{it}, i = 1, \dots, N, t = 1, \dots, T \\ \mu_i &\sim iid(0, \sigma_\mu^2) \\ v_i &\sim iid(0, \sigma_v^2) \end{aligned} \quad (3)$$

The aim is to use the model to simulate realistic and plausible outcomes across countries and time. Normally one would estimate the parameters  $\gamma, \rho_1, \beta, \theta$  and  $\rho_2$  but in this case we simply assume values that make sense and correspond to what seems evident from data and other modelling exercises relating to coronavirus. Accordingly, we assume  $\gamma = 0.25, \rho_1 = 0.25, \beta = 0.5, \theta = -\rho_1 \gamma$

and  $\rho_2 = 0.75$ . The relationship between parameters as given by the assumed value of  $\theta$  is a property of this specification as explained in the cited literature. These assumptions are commensurate with a dynamically stable, stationary process. This means that in the long run the number of cases in each country tends to its own equilibrium level. Other choices for these parameters could lead to exponential growth in the number of cases with no long run equilibrium. Clearly this is an outcome to be avoided.

The equilibrium outcome at time  $T$  for this process is given by

$$\ln c_T = (\mathbf{B}_N - \mathbf{C}_N)^{-1} (\beta p_T + \mathbf{G}_N \mu) \quad (4)$$

In this,  $\mathbf{B}_N = (\mathbf{I}_N - \rho_1 \mathbf{W}_N)$ ,  $\mathbf{C}_N = (\gamma \mathbf{I}_N + \theta \mathbf{W}_N)$  and  $\mathbf{G}_N = (\mathbf{I}_N - \rho_2 \mathbf{W}_N)$ ,  $p_T$  is the population at  $T$  which is assumed to remain constant for each country over time, and  $\mu$  is an  $N$  by 1 vector of time-invariant heterogeneity across countries. This is generated by drawing at random from an  $N(0,0.01)$  distribution. In practice, the evolution towards equilibrium is given by the recursive iteration through to time  $T$  of

$$\ln c_t = \mathbf{B}_N^{-1} (\mathbf{C}_N \ln c_{t-1} + \beta p_t + \mathbf{G}_N \mu), t = 2, \dots, T \quad (5)$$

Where the process commences with  $\ln c_{i1} = 1$  where  $i$  denotes China, and

$\ln c_{j1} = 0, j = 1, \dots, N, j \neq i$  for all other countries. This gives exactly the outcome at time  $T$  as equation (4). The reason for the iteration is that we wish to study the evolution of the number of cases and the share of population infected through time until steady state has been reached. Also, the iteration involves intervention so that any country for which the number of cases exceeds 80% of the population has that number held rather than evolving further. This is based on the theory that previous cases have become immune to the virus, and avoids the outcome where a few countries have more than 100% of the population being infected. With this intervention, the outcomes of equations (4) and (5) are not exactly equal for this small number of countries.

### Outcomes

For 201 countries for which we have data, we find one country (Germany) with 10million cases. Italy, the current hot spot, has nearly 9 million. China has 5.5 million (this is probably wrong !). As shown by Figure (1), most other countries have less than 2 million cases.

Figure 1

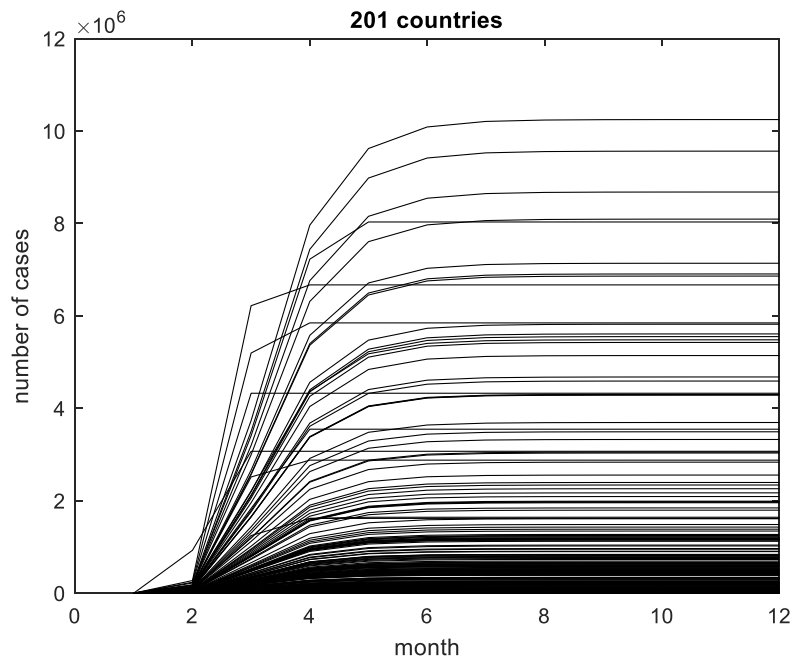


Figure 2 shows the geographical distribution of the number of cases at equilibrium (that is the levels at the right hand side of Figure 1). There are 5 circle sizes using quantiles ranging from 20%, 40%,60% and top 80% of number of cases. Circles are located at capital cities. The USA figure looks small compared with reality...it is only a simulation model! The fact that Europe is the epicentre is clearly evident.

Figure 2

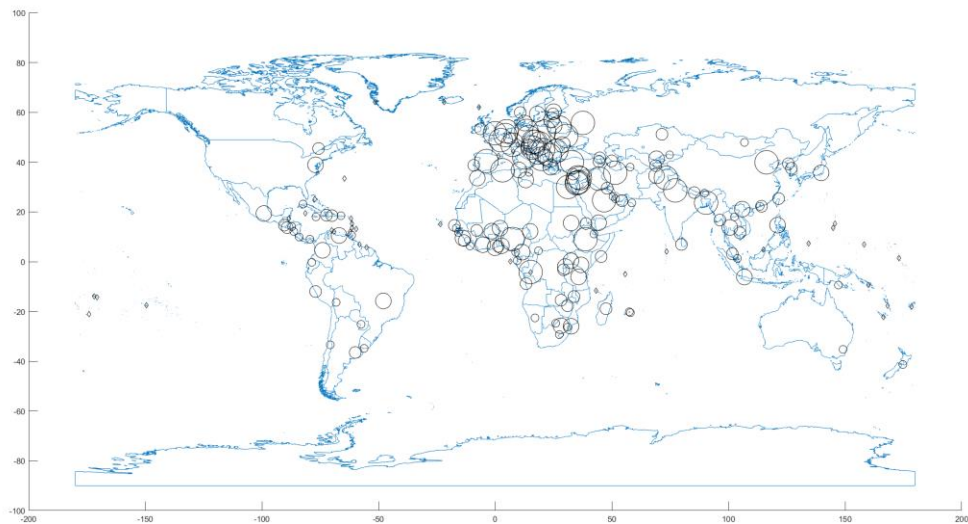


Figure 3 shows that the UK is clearly in the top 20%, alongside Italy. This doesn't do justice to the relatively greater number actually currently occurring in Italy.

Figure 3

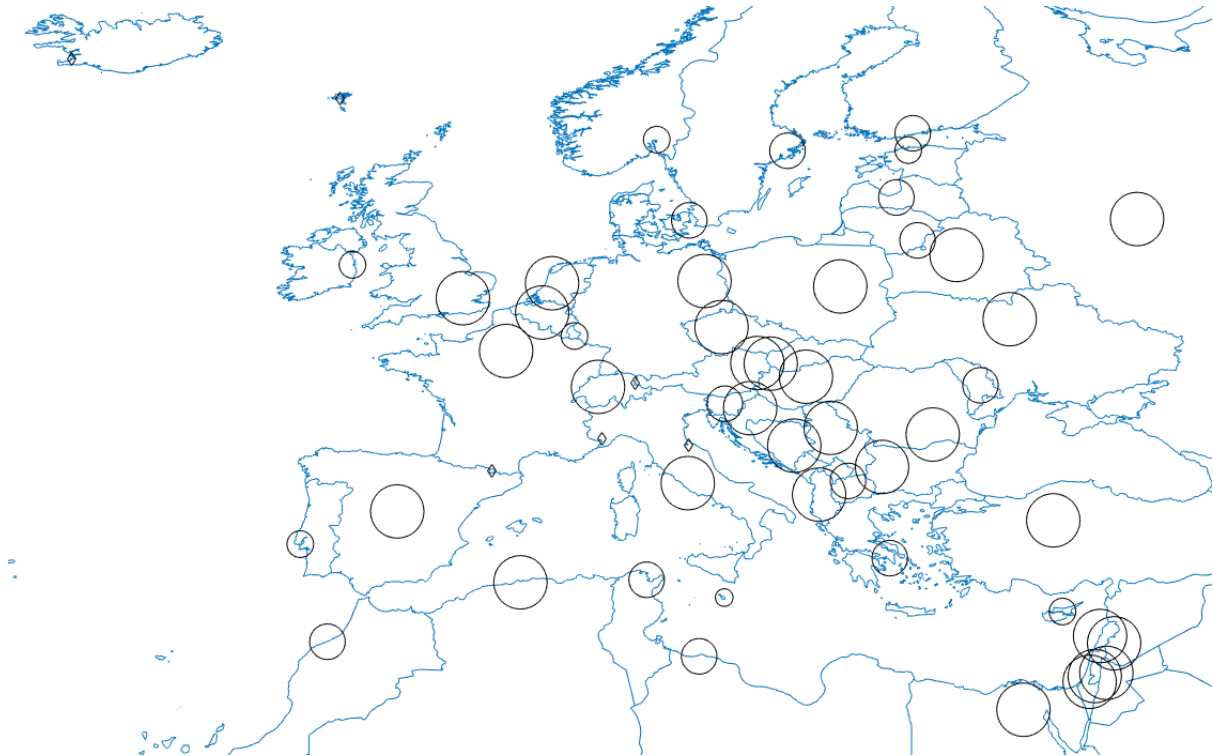


Figure 4 shows that the share of infected is mainly about 5-10% of the total population. Note the cut off at 80% assuming inherent immunity from previously infected people.

Figure 4

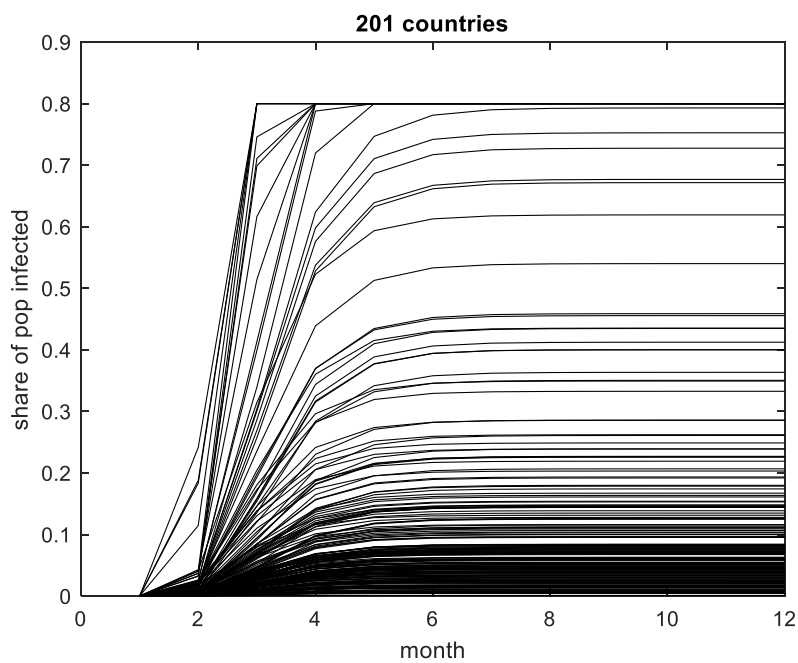


Figure 5 shows the global distribution of the infected proportions, bigger circles mean bigger proportions.

Figure 5

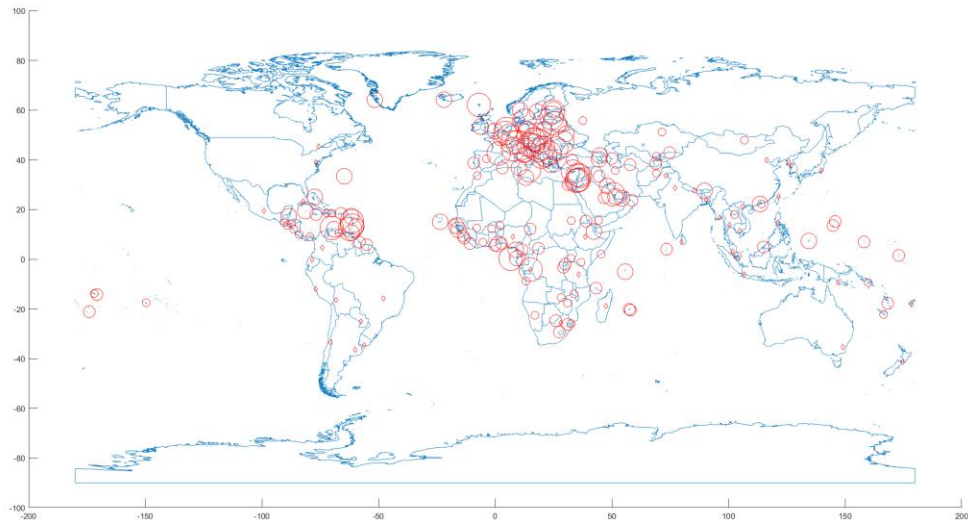


Figure 6 shows the proportions of the population of each country infected. Note in this simulation, Italy (Rome) is on a par (roughly) with the UK, and less than San Marino.

Figure 6

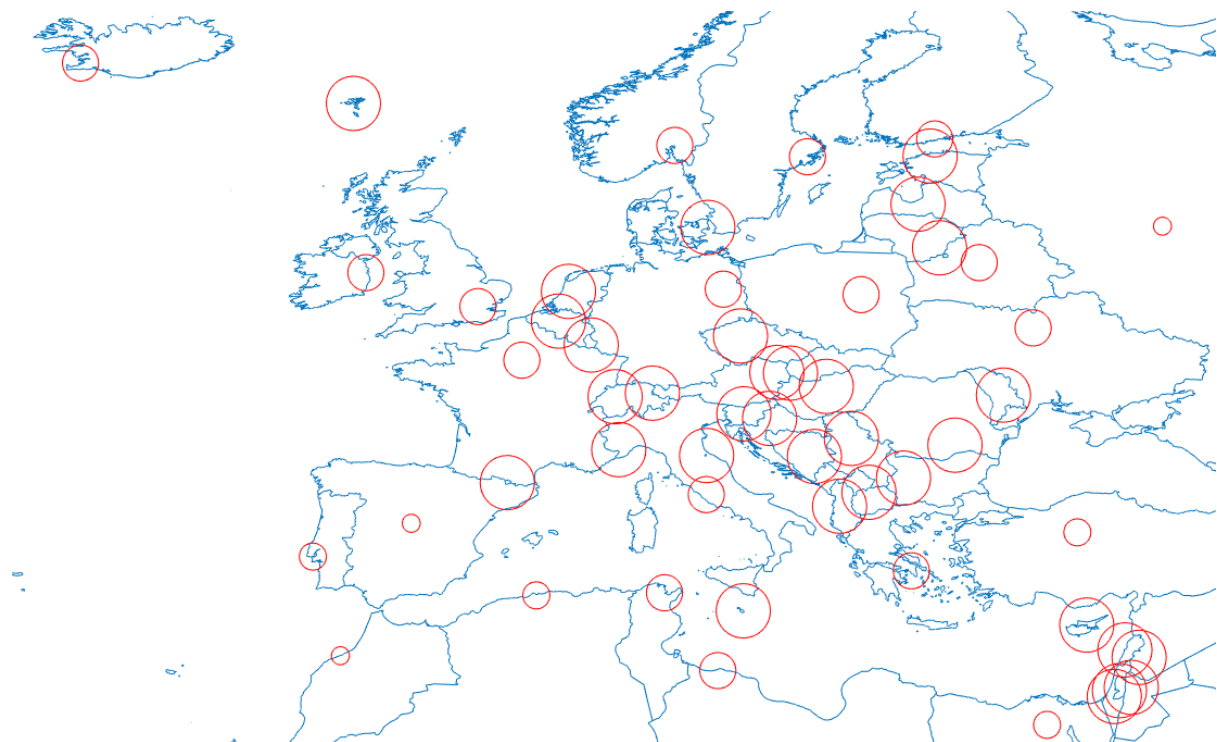


Figure 7 is the difference between the number of cases per month, indicating the new cases. Because the cases level off at equilibrium (figure 1) new cases fall to zero. Those countries that peak early are though that quickly hit the 80% barrier and so there were no new cases.

Figure 7

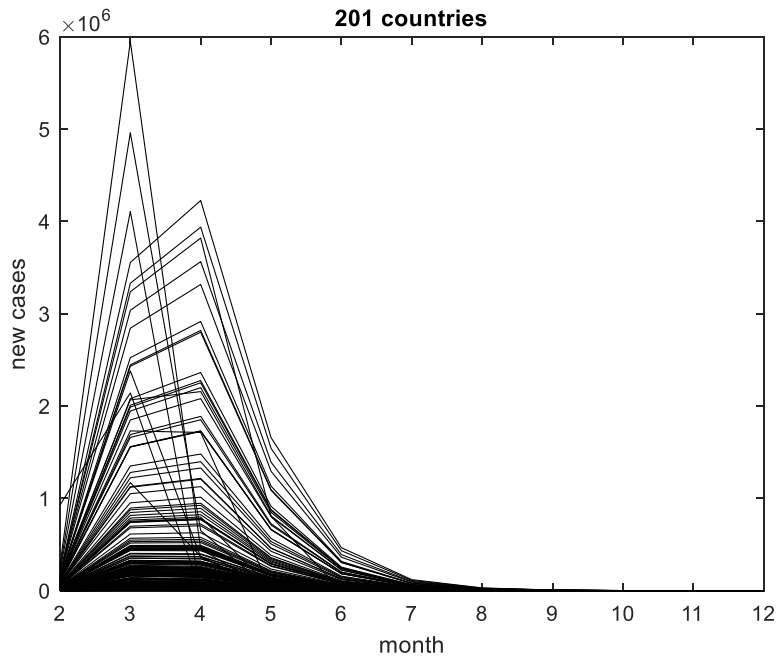


Figure 8 shows the paths for number of cases for Great Britain and Italy. For GB, cases reach nearly 7 million. For Italy, the number is nearly 9 million.

Figure 8

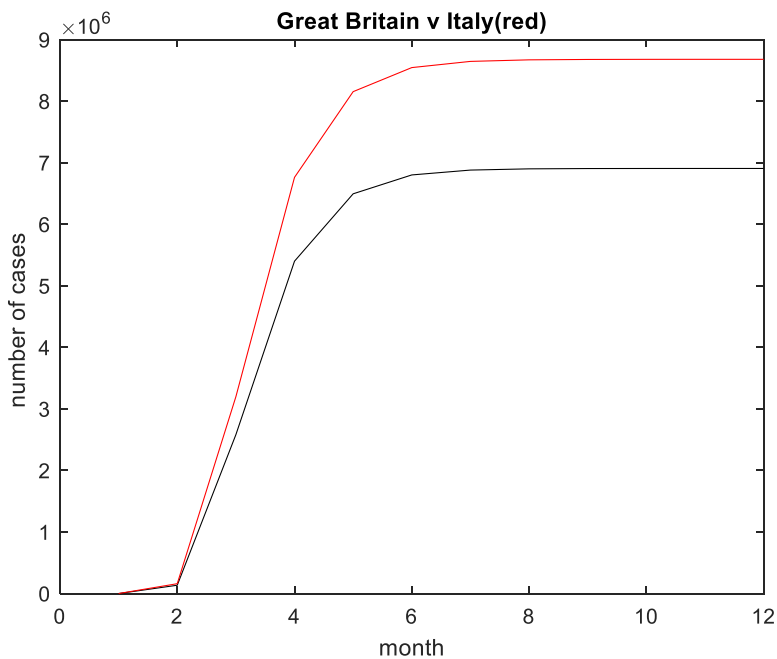


Figure 9 shows new cases peaking at about month 4, falling to zero by month 9.

Figure 9

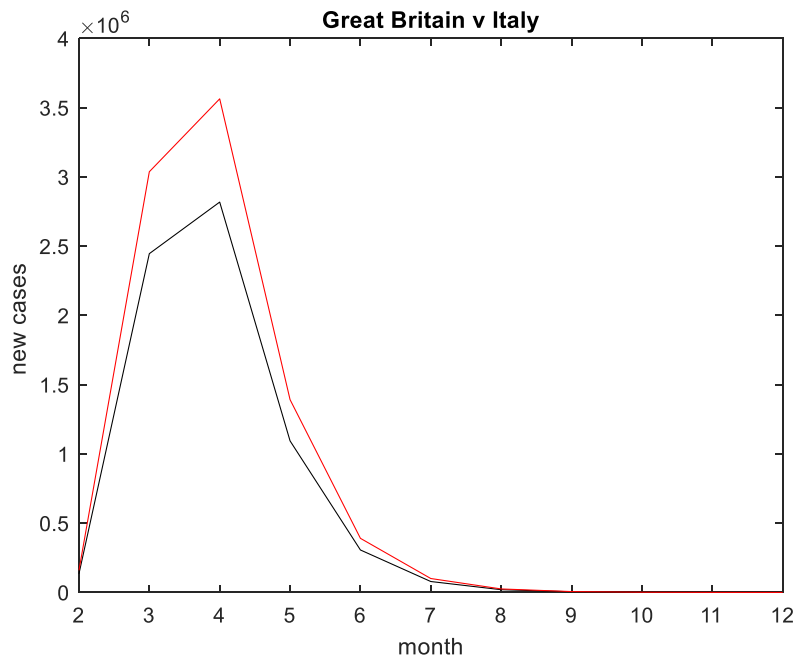
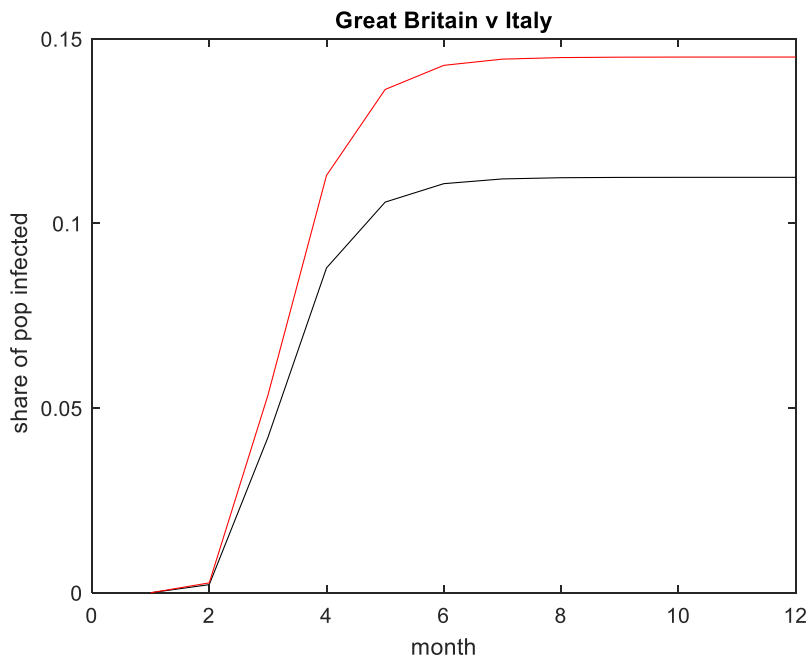


Figure 10 shows that about 10% of the UK population will be infected, compared with nearly 15% for Italy.

Figure 10



What can be done? We can simulate the impact of reduced connectivity. I don't want to take any of this too seriously, but when connectivity is reduced so that  $\rho_1 = \rho_2 = 0.15$  we see fewer cases and lower infection rates. Compare figures 1 and 11, we see the number of cases diminish from a maximum of 10 million to about 3.5 million. India and China now top the charts.

Figure 11

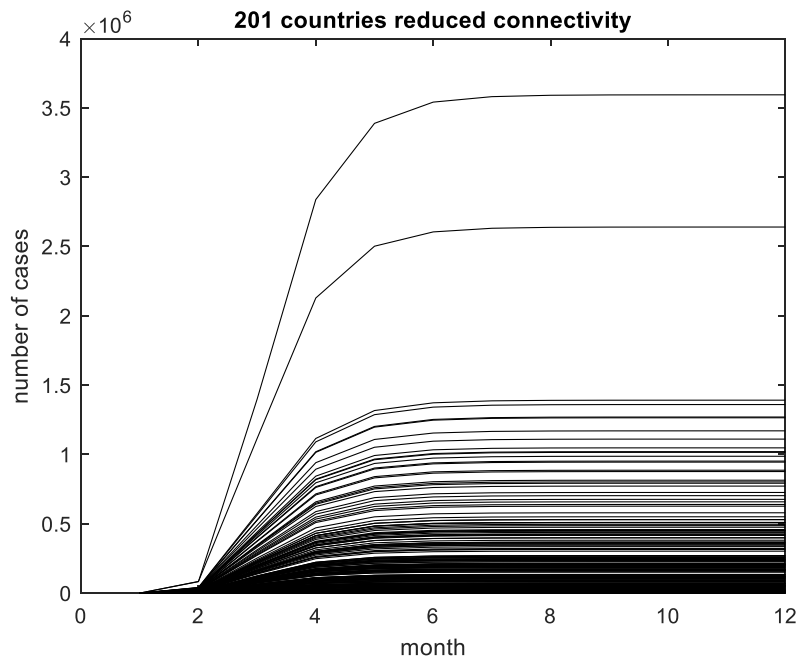
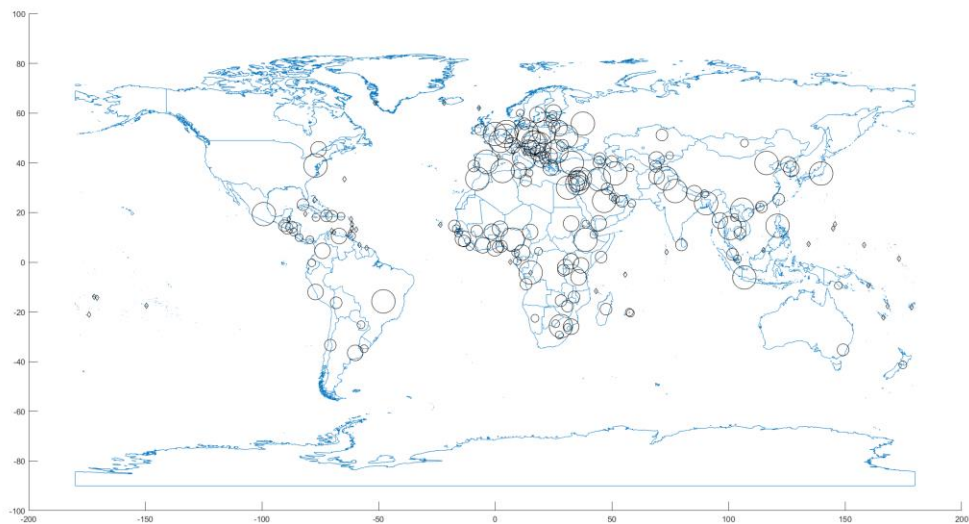


Figure 12 gives the global distribution of number of cases with 5 groups based on the quantiles ranging from 20%, 40%, 60% and top 80% of number of cases.

Figure 12





In Figure 13 we zoom in to show number of cases across Europe, from Figure 12.

Figure 13

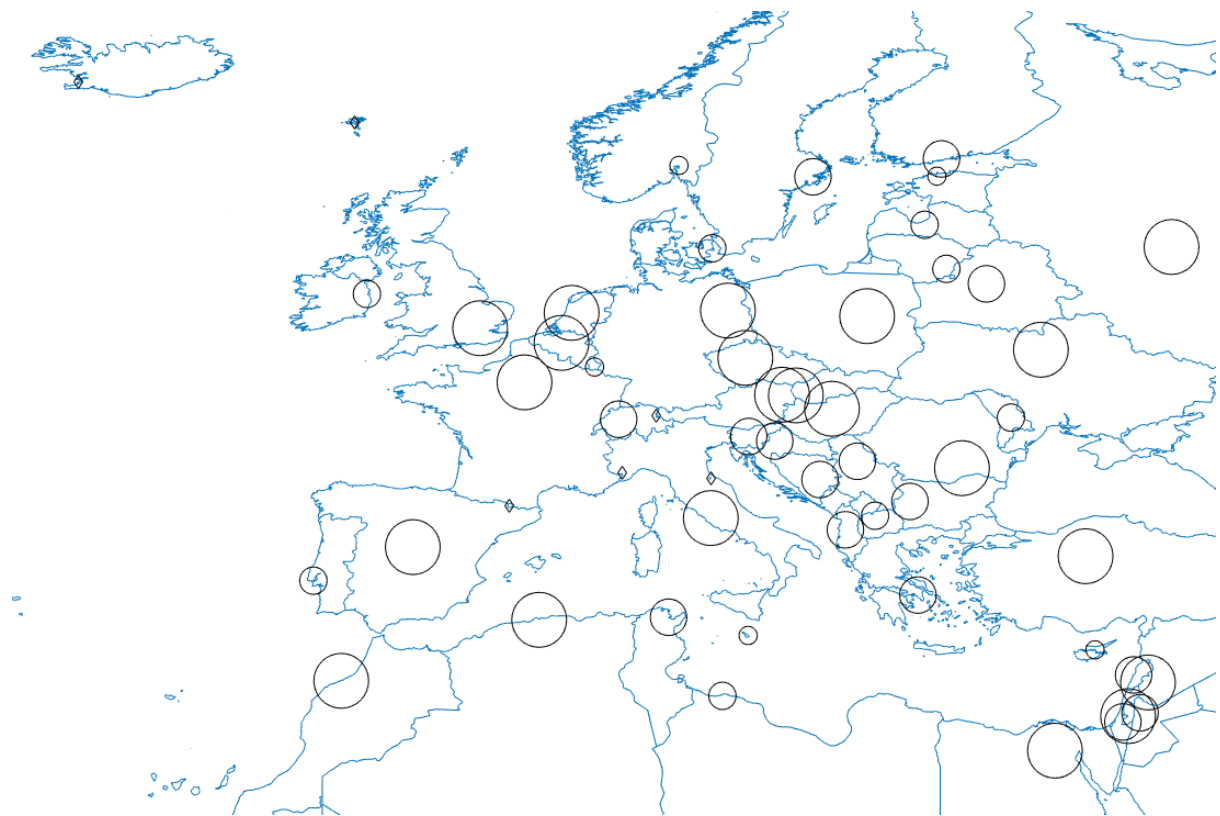


Figure 14 gives share of population infected. Comparing Figure 4 and 14, there appears to be a toning down of infectivity.

Figure 14

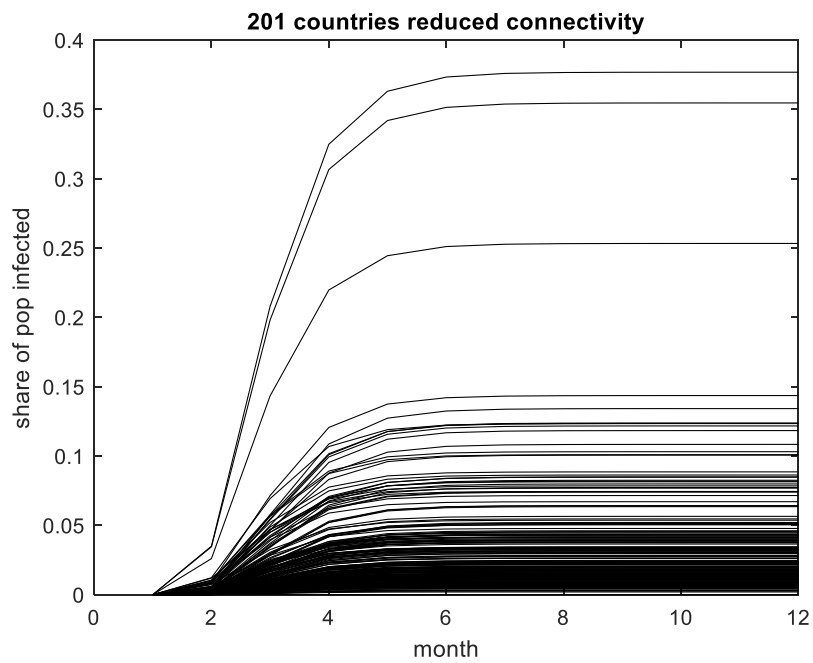


Figure 15 gives the global distribution of infected population shares.

Figure 15

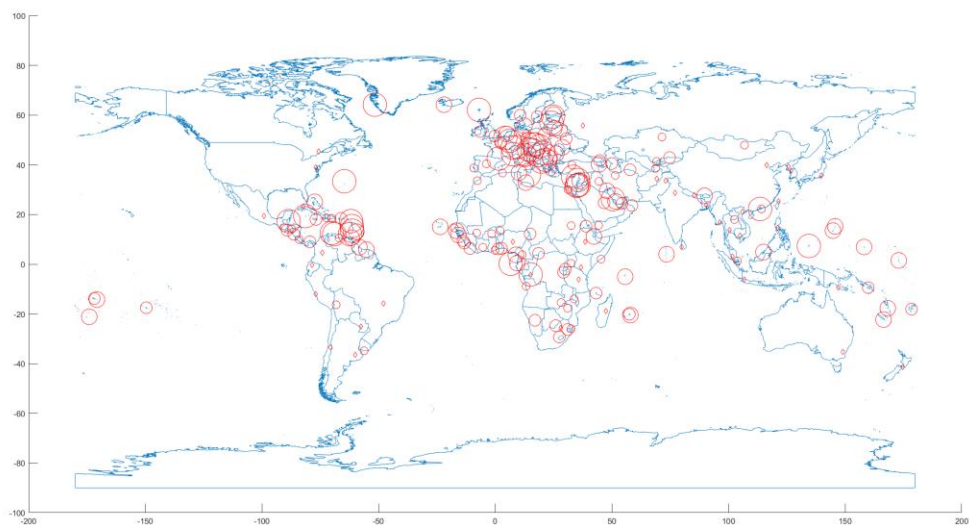
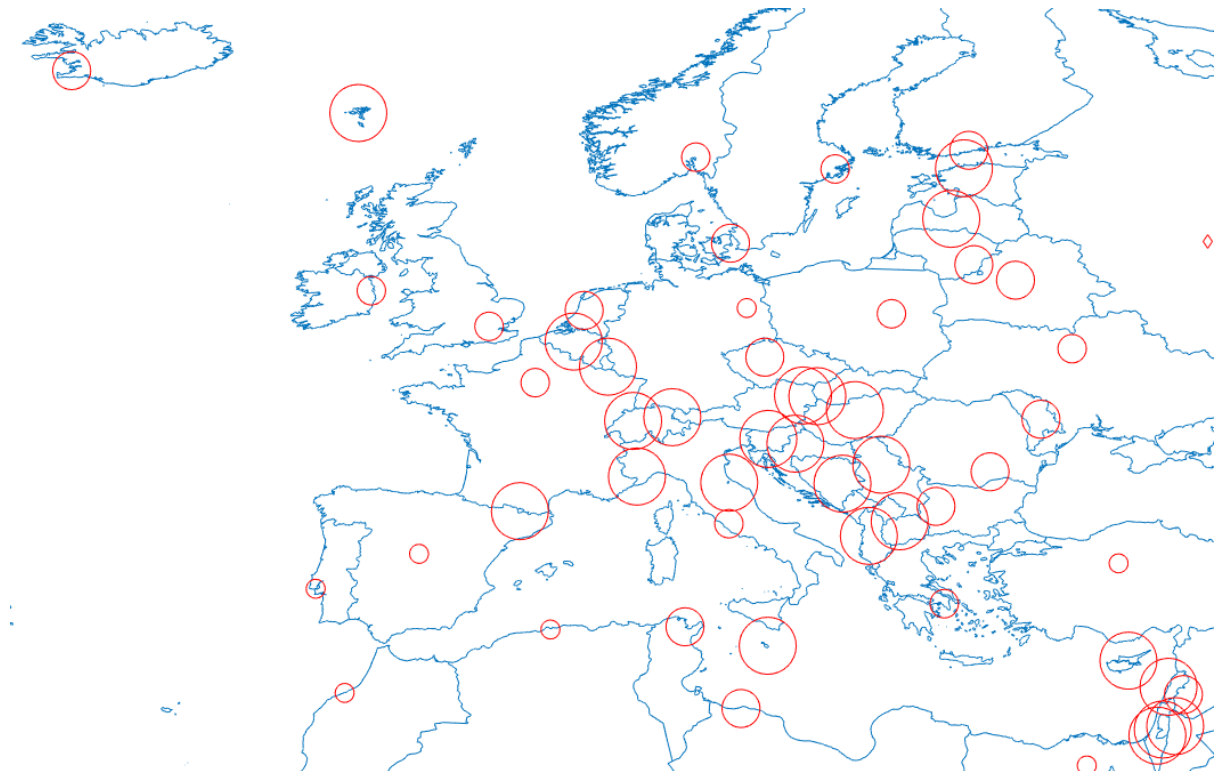


Figure 16 gives the European distribution of infected population shares.

Figure 16



### Conclusion

The message is, connectivity matters, and reducing it either internationally or on a personal level can have a big impact on the transmission of the virus.

### References

Baltagi BH, Fingleton B, Pirotte A (2014) Estimating and forecasting with a dynamic spatial panel model. *Oxford Bulletin of Economics and Statistics* 76:112–138

Baltagi B, Fingleton B, Pirotte A (2019) 'A Time-Space Dynamic Panel Data Model with Spatial Moving Average Errors' *Regional Science and Urban Economics* 76 13-31

