

Modelling coronavirus with spatial proximity effects

Bernard Fingleton

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

Version 2

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

$$\ln c_t = k + \gamma \ln c_{t-1} + \rho_1 \mathbf{W}_N \ln c_t + \beta_1 \ln p_t + \beta_2 \ln v_t + \theta \mathbf{W}_N \ln c_{t-1} + \varepsilon_t \quad (1)$$

Where k is a constant, 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. Compared with the initial version of this WP, an additional variable has been introduced, namely the mean 2014-2018 international tourism receipts for travel items (current US\$) downloaded from the World bank databank website (<https://data.worldbank.org/indicator/ST.INT.TVLR.CD?view=map>). This is denoted by v in equation (1). One major limitation of the previous simulation was the low number of cases predicted for the USA, whereas we now know that the USA is a major epicentre of the pandemic. The new variable is designed to rectify this, by allowing the number of cases to also depend on the international tourism intensity of each country as indicated by v . It provides a supplementary measure of connectivity additional to the reciprocal of the great circle distance between countries. Note that data do not exist for some 34 of the 201 countries so a pragmatic solution has been to give them the nominal value of 1.0.

\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 $\rho_1 < 1$ 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 tourism intensity of the country (higher receipts approximate to more tourists), 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, as represented by ε_t . Generally ε_t 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} .

In recent literature (Baltagi et al, 2019), a spatial moving average process causing error dependence across countries has been adopted, thus

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

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 ρ_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

$$k = -16.5, \gamma = 0.25, \rho_1 = 0.25, \beta_1 = 0.4, \beta_2 = 0.75, \theta = -\rho_1 \gamma, \sigma_\mu^2 = 0.01, \sigma_v^2 = 0.001 \text{ and } \rho_2 = 0.1$$

. The relationship between parameters as given by the assumed value of θ 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. The rules governing dynamic stability are set out in detail in the literature cited. 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

$$\ln c_T = (\mathbf{B}_N - \mathbf{C}_N)^{-1} (\beta_1 p_T + \beta_2 v_T + \mathbf{G}_N \mu) \tag{4}$$

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, v_T is the tourism intensity, and μ is an N by 1 vector of time-invariant heterogeneity across countries. This is generated by drawing at random from an $N(0, \sigma_\mu^2)$ distribution. In practice, the evolution towards equilibrium is given by the recursive iteration through to time T of

$$\ln c_t = \mathbf{B}_N^{-1} (\mathbf{C}_N \ln c_{t-1} + \beta_1 p_t + \beta_2 v_t + \mathbf{G}_N \mu), t = 2, \dots, T \tag{5}$$

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, although not applicable in this case, the iteration could involve 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..

Outcomes

For 201 countries for which we have data, we find one country (USA) with nearly 6 million cases at steady state. China has more than 2 million. France, Germany, India, and Spain have more than 1 million cases. As shown by Figure (1), most other countries have less than 1 million cases.

Figure 1

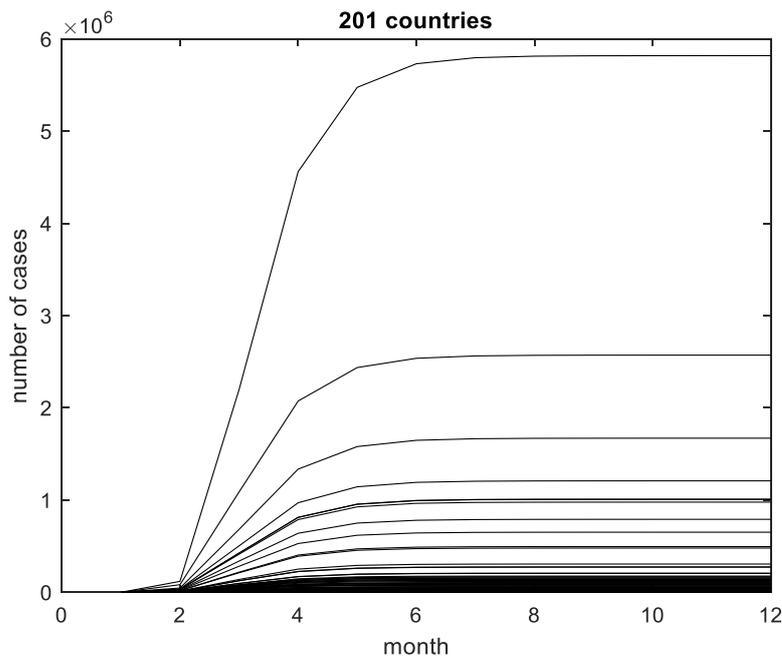


Figure 2 shows the geographical distribution of the number of cases at equilibrium (that is the levels at the right hand side of Figure 1). There are 6 circle sizes using quantiles ranging from 20%, 40%, 60%, 80% and top 99% of number of cases. Circles are located at capital cities. The USA and China stand out with the highest number of case. The fact that Europe became the epicentre is clearly evident.

Figure 2

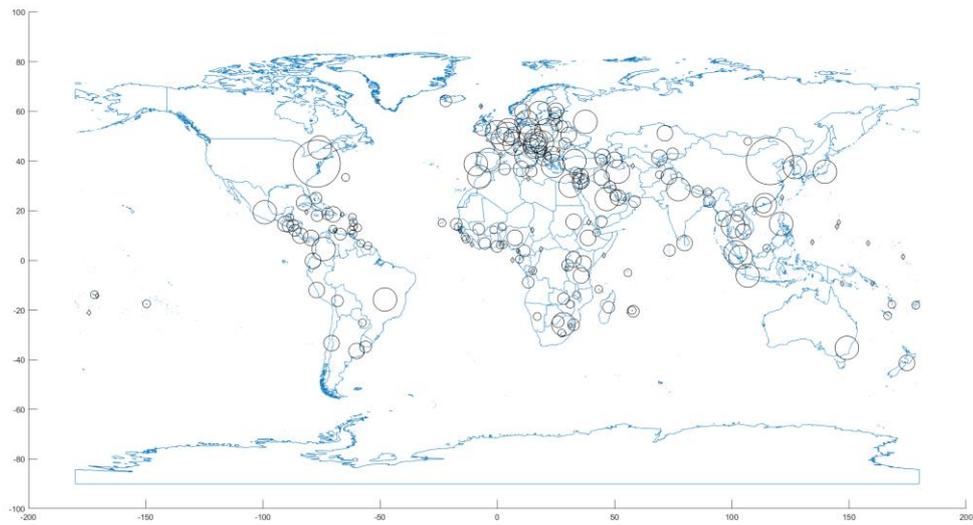


Figure 3 shows that the number of cases is relatively large in countries with large populations, modified by the tourism intensity.

Figure 3



Figure 4 shows that the share of infected is mainly below about 2% of the total population. The outlier is Macao, with more than 7% of the population infected. Also in the top group with more than 1.5% infected are Austria, France, Hong Kong, Luxembourg, Spain, Great Britain and the USA.

Figure 4

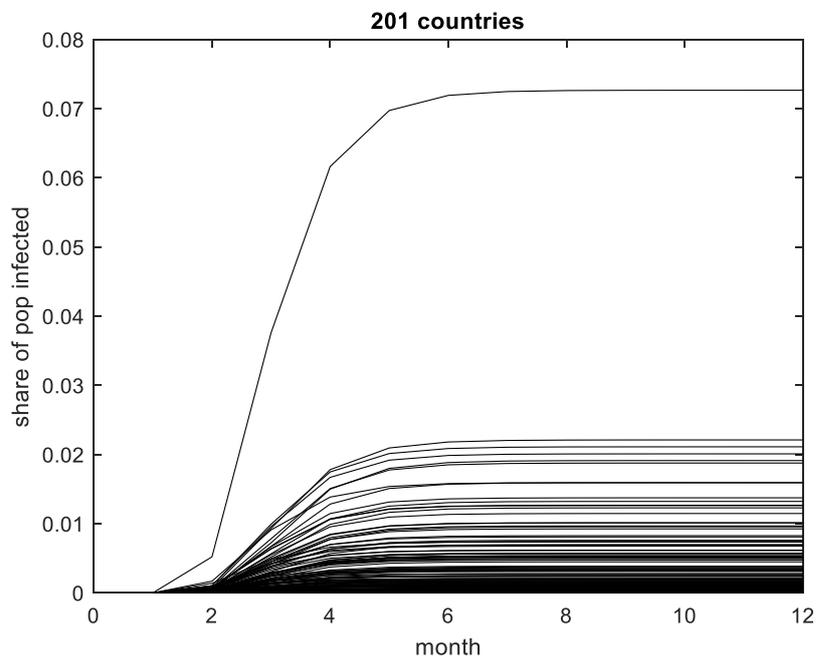


Figure 5 shows the global distribution of the infected proportions, bigger circles mean bigger proportions.

Figure 5

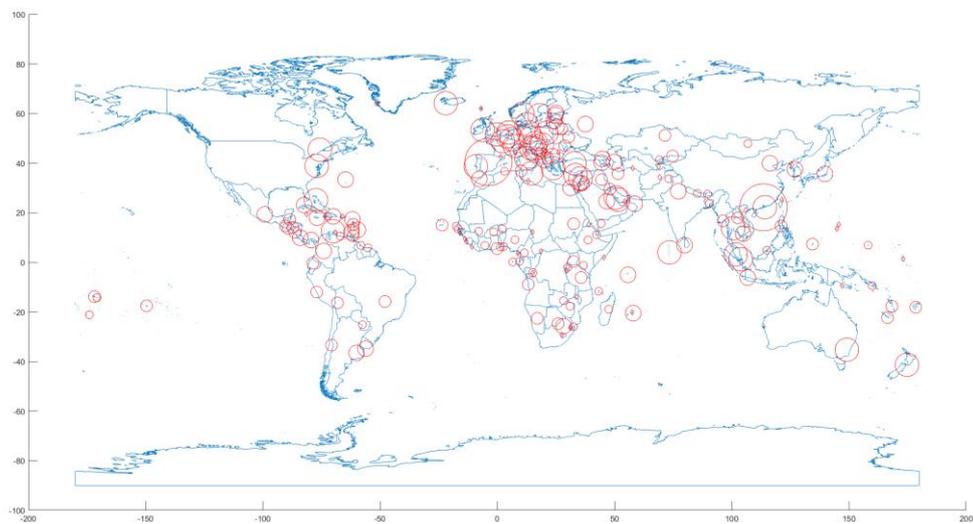


Figure 6 shows the proportions of the population of each European country infected. Note that Spain stands out as second only to Macau.

Figure 6

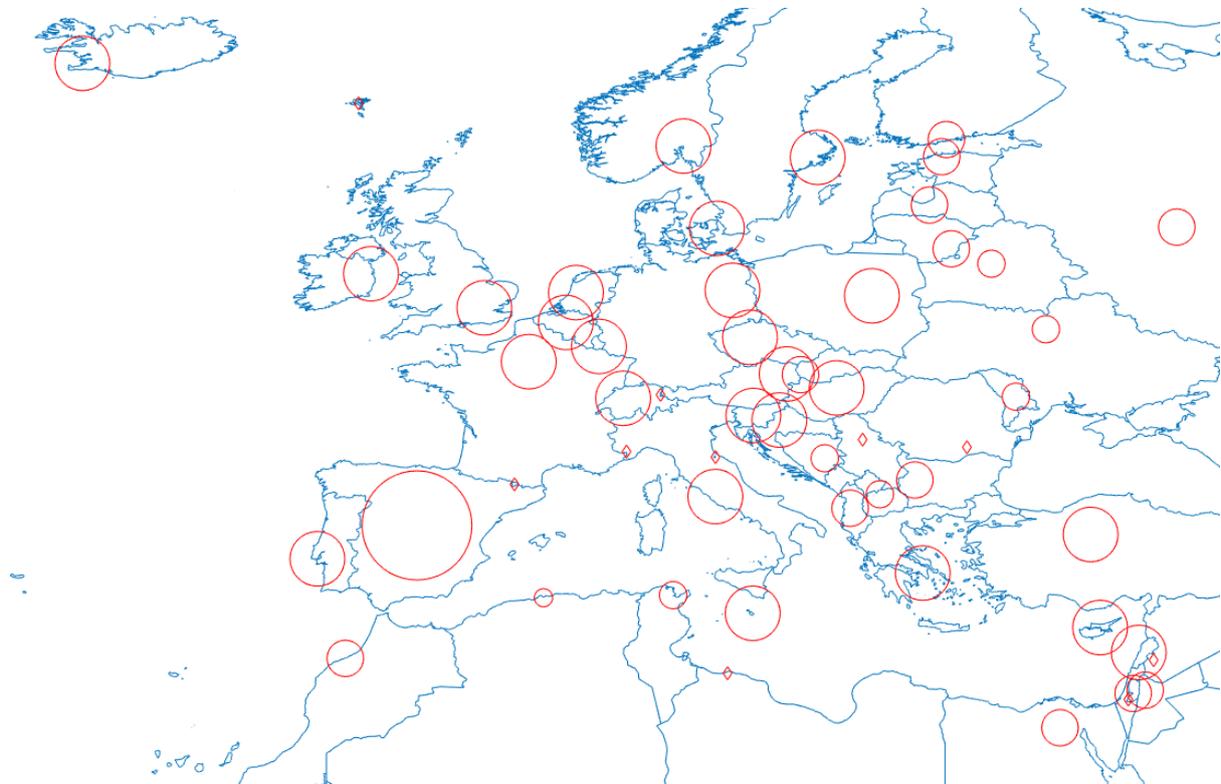


Figure 7 is the difference between the number of cases per month, indicating the new cases. Because the cases level off at equilibrium (figure 1) new cases fall to zero. Although there is in theory an upper limit of 80% infection rate, which would cause the number of new cases to fall to zero, no country hits this barrier.

Figure 7

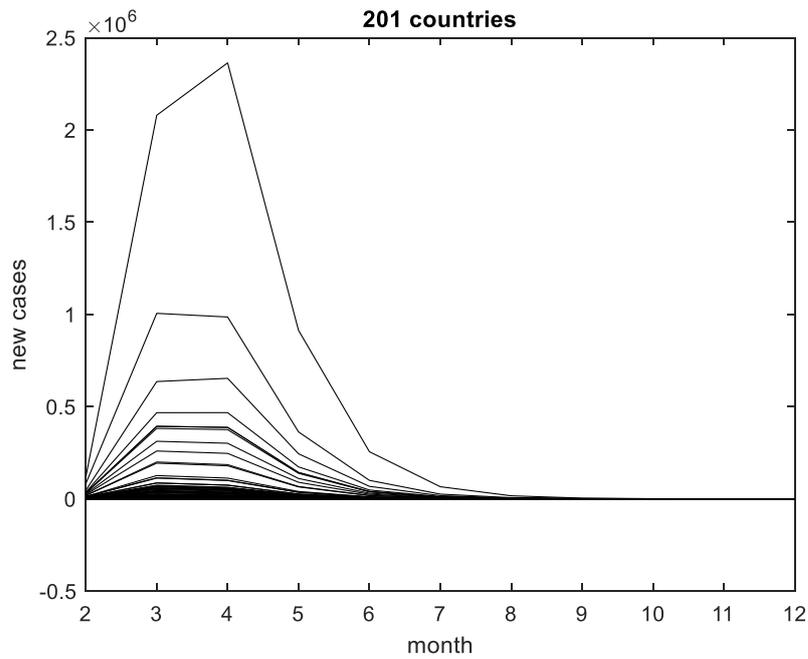


Figure 8 shows the paths for number of cases for Great Britain, Italy which was the older epicentre, and the USA which is the new epicentre. For GB and Italy cases reach nearly 1 million, the number rise to well over 5 million for the USA.

Figure 8

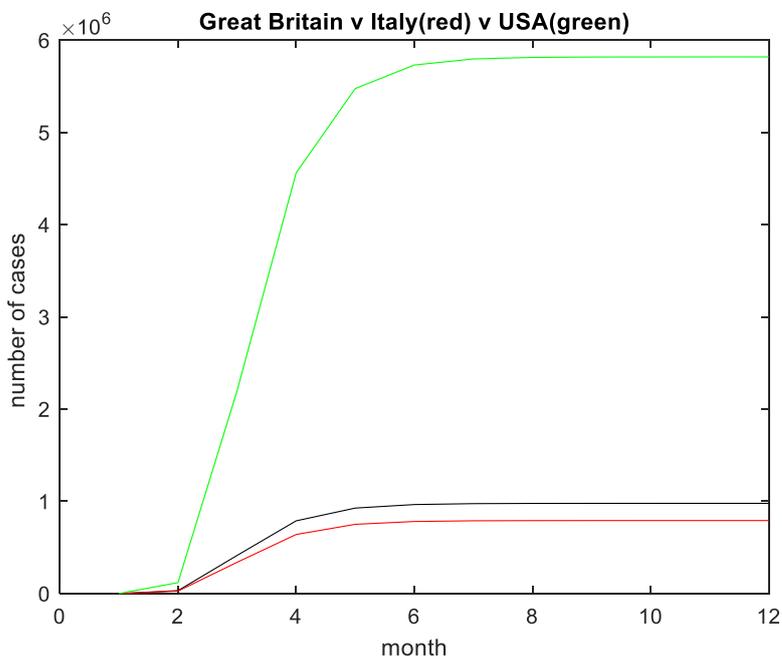


Figure 9 shows new cases peaking at about month 4, falling to zero by month 9.

Figure 9

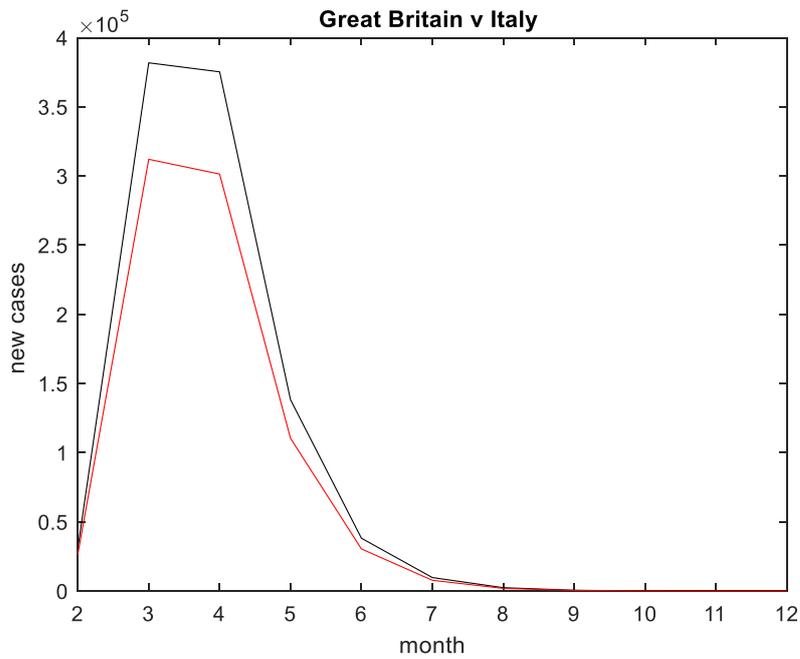
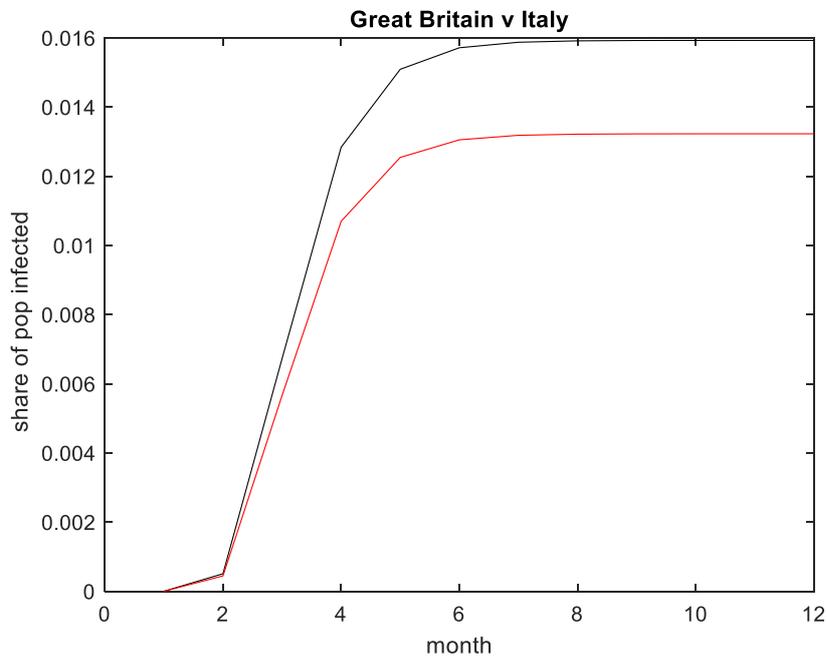


Figure 10 shows that between about 1.3% and 1.6% of the UK and Italian population will be infected.

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 $\rho_1 = 0.15$ we see fewer cases and lower infection rates. Compare figures 1 and 11, we see the number of cases diminish from a maximum of nearly 6 million to just below 3.5 million.

Figure 11

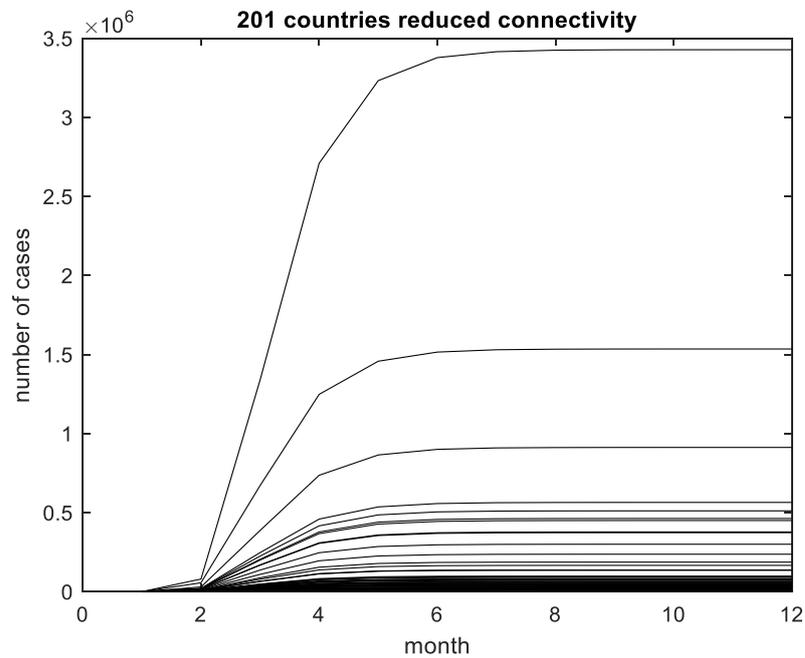
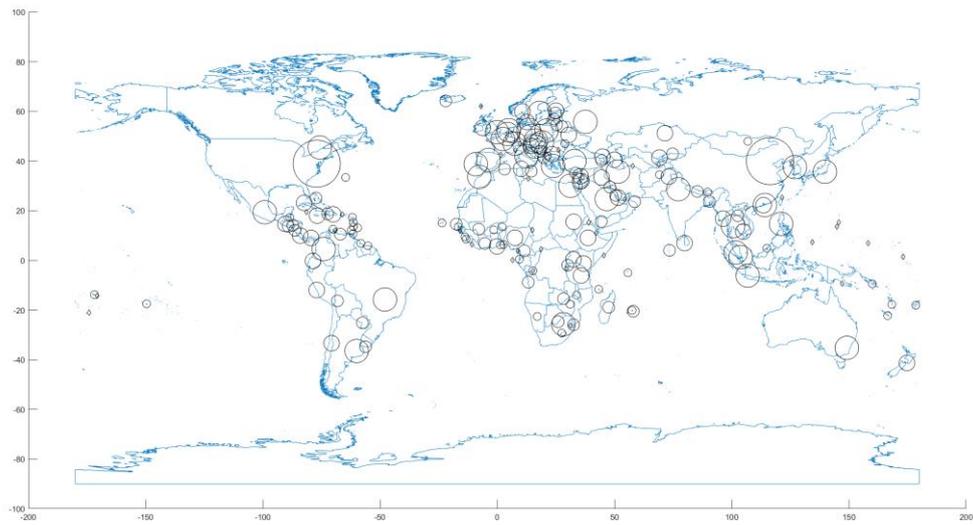


Figure 12 gives the global distribution of number of cases with 6 groups based on the quantiles ranging from 20%, 40%, 60%, 80% and 99% of number of cases.

Figure 12



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

Figure 13

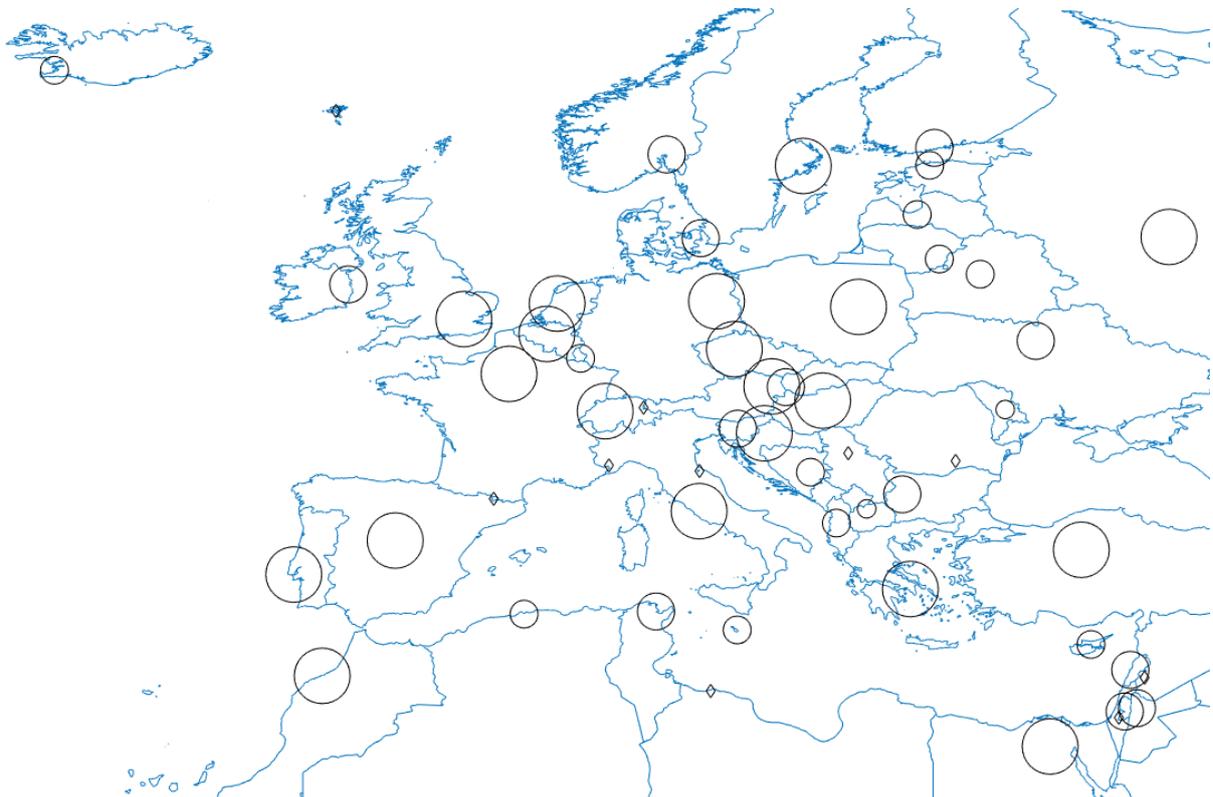


Figure 14 gives share of population infected. Comparing Figure 4 and 14, there appears to be a toning down of infectivity. Macau still leads but with just above 4%, followed by Hong Kong (1.15%), Spain(1.12%) and the USA(1.13%). All other countries have an infectivity rate of less than 1%.

Figure 14

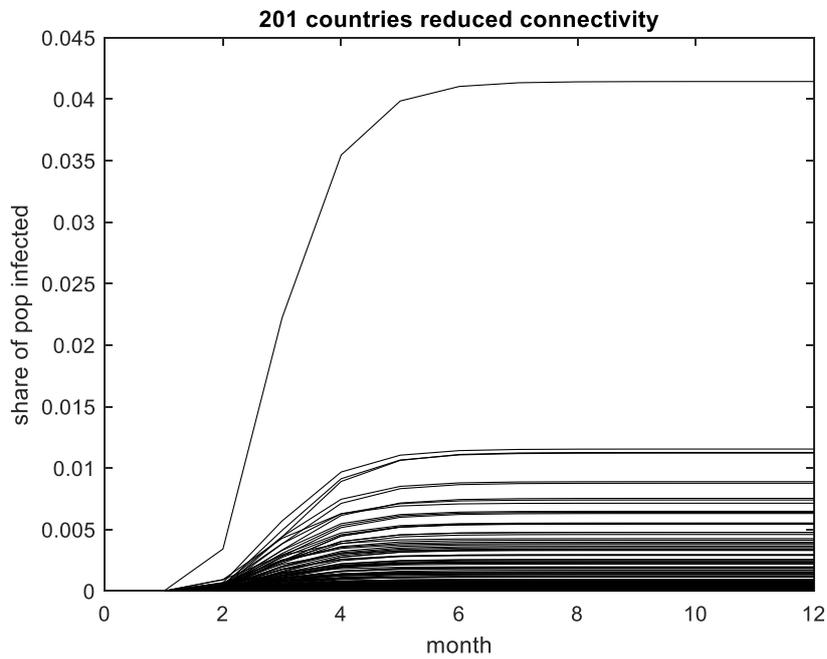


Figure 15 gives the global distribution of infected population shares.

Figure 15

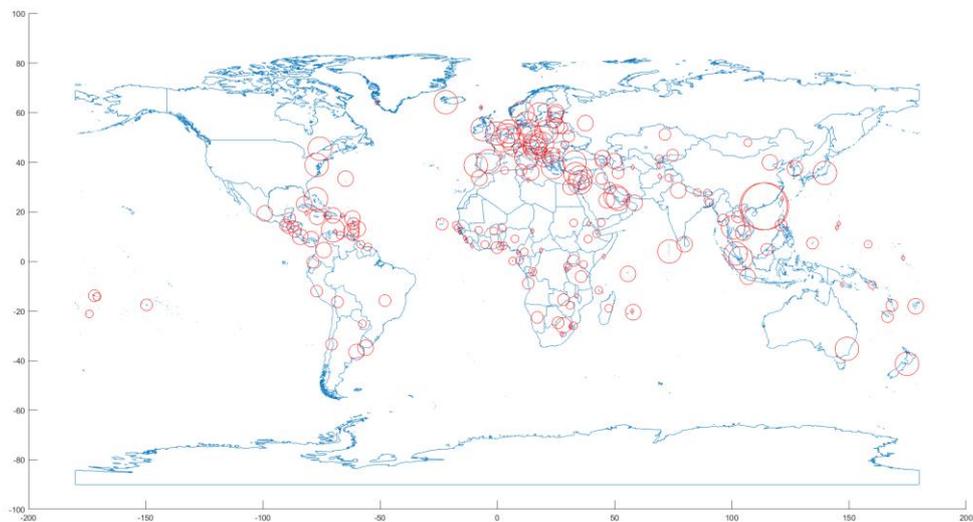
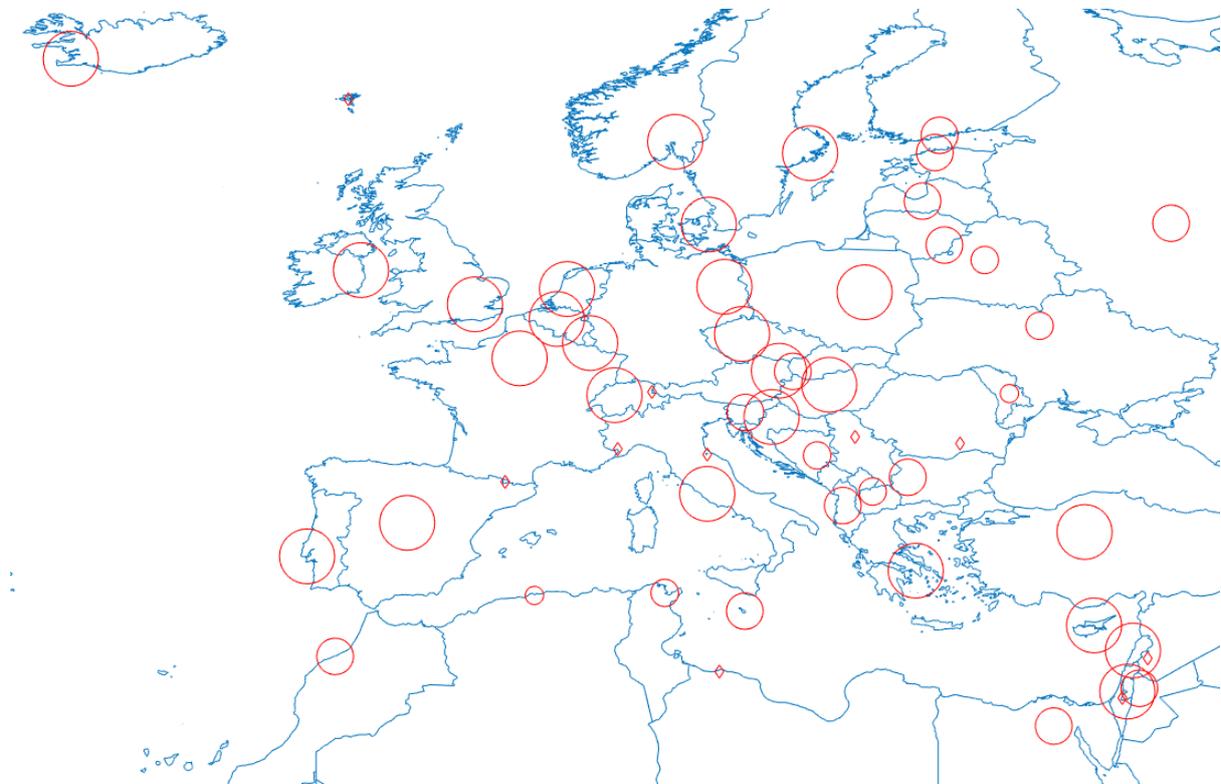


Figure 16 gives the European distribution of infected population shares, these are mainly below 1% although Spain is in the top four globally with 1.12%.

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