Clustering of epidemic time trends: the case of COVID-19

1 Approach 1

1.1 Model

Let Y_{it} be the number of new COVID-19 infections on day t in country i and suppose we observe a time series $\mathcal{Y}_i = \{Y_{it} : 1 \le t \le T\}$ for a large number n of different countries i. In order to align the data of different countries, we take the starting date $t = 1$ to be the first Monday after reaching 100 confirmed cases in each country. Considering the dates after reaching a certain level of confirmed cases is a common practice of "normalizing" the data (see e.g. Cohen and Kupferschmidt, 2020). Starting on a Monday additionally aligns the data across countries by the day of the week. This allows us to take care of possible weekly cycles in the data which are produced by delays in reporting new cases over the weekend.

We consider a simple nonparametric trend model for the time series $\mathcal{Y}_i$ in our sample. In particular, each time series $\mathcal{Y}_i$ is assumed to satisfy the nonparametric regression equation

$$Y_{it} = m_i\left(\frac{t}{T}\right) + u_{it} \qquad (1 \le t \le T) \qquad (1.1)$$

with $\mathbb{E}[u_{it}] = 0$, where m_i is an unknown smooth trend curve defined on $[0,1]$. As usual in nonparametric regression (see e.g. Robinson, 1989), we let the regression function m_i in model (1.1) depend on rescaled time t/T rather than on real time t. The assumptions on the error term u_{it} are discussed later.

Even though the trend functions m_i can be expected to be different across countries i, it is natural to assume that there are groups of countries i with similar trend curves m_i. We thus impose a group structure on the countries in our sample. Informally speaking, we suppose that the countries can be grouped into a small number of classes such that within each class, all countries have the same trend function up to certain transformations. Formally speaking, the class structure is defined as follows:

(G) The set of countries $\{1, \dots, n\}$ can be partitioned into K groups $\mathcal{G}_1, \dots, \mathcal{G}_K$ such that

$$m_i \in \mathcal{F}_k \quad \text{for all } i \in \mathcal{G}_k,$$

where $\mathcal{F}_k$ are function classes defined as

$$\mathcal{F}_k := \big\{ f : [0,1] \to \mathbb{R} \,\big|\, f = c \cdot g_k(b \cdot u) \text{ with } c > 0, b \in [1, \bar{b}] \text{ and } g_k \text{ a density} \big\}.$$

The parameter $\bar{b}$ is assumed to be known. For identification purposes, the classes are supposed to be distinct, i.e., $\mathcal{F}_k \cap \mathcal{F}_{k'} = \emptyset$ for any $k \neq k'$.

As can be seen, the elements of the class $\mathcal{F}_k$ are rescaled versions of a density function g_k. Hence, all countries i in the k-th group $\mathcal{G}_k$ have a trend curve m_i that is a rescaled version of g_k, in particular, $m_i(u) = c \cdot g_k(b \cdot u)$ for some constants c and b. We can regard c as a country-specific scaling parameter that accounts for the size of the country or population density. We introduce this additional parameter in order to be able to compare countries with vastly different population sizes, e.g., Luxembourg and Russia. The constant b can be interpreted as measuring the speed at which the epidemic develops in different countries. To see this, consider two countries i and j from the same group k whose trend functions are given by $m_i(u) = g_k(b_i \cdot u)$ and $m_j(u) = g_k(b_j \cdot u)$ with $b_j > b_i$. (For simplicity, we set $c_i = c_j = 1$.) Obviously, we can write $m_j(u) = m_i(\{b_j/b_i\}u)$ for $u \in [0, b_i/b_j]$. Hence, the trend m_j evolves qualitatively in the same way as m_i, but it evolves faster by the factor $b_j/b_i > 1$. In what follows, we call b the effective time parameter of the model.

Remark 1.1. *The functions m_i are not uniquely identified in terms of c, b and g_k. In particular, we can rewrite $m_i(u) = c \cdot g_k(b \cdot u)$ as $m_i(u) = \tilde{c} \cdot \tilde{g}_k(u)$, where $\tilde{g}_k(u) = g_k(b \cdot u)/\int_{-\infty}^{\infty} g_k(b \cdot v)dv$ and $\tilde{c} = c \int_{-\infty}^{\infty} g_k(b \cdot v)dv$. To uniquely determine the quantities c, b and g_k in the definition of the classes $\mathcal{F}_k$, we additionally impose the following constraint:*

(A) For some $i_k \in \mathcal{G}_k$, it holds that $m_{i_k}(u) = c \cdot g_k(b \cdot u)$ with $b = 1$ and for all other $i \in \mathcal{G}_k$, $m_i(u) = c \cdot g_k(b \cdot u)$ with $b \geq 1$.

Importantly, (A) is not an additional assumption but rather a harmless renormalization: If $m_i(u) = c \cdot g_k(b \cdot u)$ with $b > 1$ for all $i \in \mathcal{G}_k$, then we can replace c, b and g_k by renormalized versions $\tilde{c}$, $\tilde{b}$ and $\tilde{g}_k$ such that (A) holds. Also note that the classes $\mathcal{F}_k$ depend on the sample size n in general (i.e., $\mathcal{F}_k = \mathcal{F}_{k,n}$) if we impose the normalization (A). For simplicity, however, we suppress the dependence on n in the notation.

Remark 1.2. *We could replace the above definition of $\mathcal{F}_k$ by the more general version*

$$\mathcal{F}_k := \big\{ f : [0,1] \to \mathbb{R} \,\big|\, f = c \cdot g_k(b \cdot (u - u_0)) \ \text{with} \ c > 0, b \in [1, \bar{b}],$$
$$u_0 \geq 0 \ \text{and} \ g_k \ \text{a density}\big\}.$$

However, since we have aligned the data across countries i (by choosing the starting date $t = 1$ to be the first Monday after reaching the 100-th confirmed case in each country), we have implicitly normalized u_0 to be equal (to 0) in each country.

Obviously, the group structure in the data is not observed. In particular, the groups $\mathcal{G}_1, \ldots, \mathcal{G}_K$, the group-specific density functions $g_1, \ldots, g_K$ and the number of groups K are unknown in practice. In what follows, we construct a statistical procedure to estimate these quantities.

1.2 Clustering procedure

In this section, we construct a hierarchical clustering algorithm to estimate the unknown group structure from the data. We first design a suitable dissimilarity measure and then build a HAC (Hierarchical Agglomerative Clustering) algorithm based on this measure.

1.2.1 Construction of the dissimilarity measure

Step 1. For each i, estimate $m_i(u)$ by a Nadaraya-Watson estimator with a rectangular kernel and bandwidth h, where we choose h to be a multiple of 7 days, i.e., 1 week. This choice of bandwidth allows us to take care of possible weekly cycles in the data which are produced by delays in reporting new cases over the weekend. Formally, the estimator $\hat{m}_i(u)$ is defined as

$$\hat{m}_i(u) = \frac{\sum_{t=1}^{T} K_h(u - \frac{t}{T}) Y_{it}}{\sum_{t=1}^{T} K_h(u - \frac{t}{T})}$$

with K being a rectangular kernel and $K_h(x) = K(x/h)/h$.

Step 2. For a given value of $b \in [1, \bar{b}]$ and for a given pair of countries (i, j), define the statistic

$$\delta_{ij}(b) = \frac{1}{1/b} \int_0^{1/b} \left(\frac{\hat{m}_i(b \cdot u)}{\int_0^{1/b} \hat{m}_i(b \cdot v) dv/(1/b)} - \frac{\hat{m}_j(u)}{\int_0^{1/b} \hat{m}_j(v) dv/(1/b)} \right)^2 du.$$

This statistic measures a weighted L_2-distance between the functions $m_i(b \cdot u)$ and $m_j(u)$ on the interval $[0, 1/b]$. Note that $\delta_{ij}(b) \neq \delta_{ji}(b)$ for $i \neq j$ in general.

Step 3. Aggregate the statistics $\delta_{ij}(b)$ for different values of b. Specifically, take the infimum over all possible values of b to obtain the statistic

$$\Delta_{ij} = \min \left\{ \inf_{b \in [1, \bar{b}]} \delta_{ij}(b), \inf_{b \in [1, \bar{b}]} \delta_{ji}(b) \right\}.$$

Step 4. Let $S \subseteq \{1, \ldots, n\}$ and $S' \subseteq \{1, \ldots, n\}$ be two sets of time series from our sample. There are several ways to define a dissimilarity measure between S and S'. We work with the complete linkage measure of dissimilarity defined as

$$\mathcal{D}(S, S') = \max_{i \in S, j \in S'} \Delta_{ij}.$$

Alternatively, we may use a single or average linkage measure.

To understand the idea behind the statistic Δ_{ij} and thus the dissimilarity measure $\mathcal{D}$, let us suppose for a moment that we could perfectly estimate m_i and m_j, that is, $\hat{m}_i = m_i$ and $\hat{m}_j = m_j$. For two countries i and j from the same group $\mathcal{G}_k$, it holds that $m_i(u) = c_i \cdot g_k(b_i \cdot u)$ and $m_j(u) = c_j \cdot g_k(b_j \cdot u)$ with some constants b_i, b_j, c_i, c_j (where $b_j \geq b_i$ w.l.o.g.). Hence, the two terms in the definition of $\delta_{ij}(b)$ can be written as

$$\frac{m_i(b \cdot u)}{\int_0^{1/b} m_i(b \cdot v) dv/(1/b)} = \frac{g_k(b \cdot b_i \cdot u)}{\int_0^{1/b} g_k(b \cdot b_i \cdot v) dv/(1/b)}$$

$$\frac{m_j(u)}{\int_0^{1/b} m_j(v) dv/(1/b)} = \frac{g_k(b_j \cdot u)}{\int_0^{1/b} g_k(b_j \cdot v) dv/(1/b)}.$$

The two right-hand sides of the above display become identical upon setting $b = b_j/b_i$, which implies that $\delta_{ij}(b) = 0$. This suggests the following: $\delta_{ij}(b)$ tends to be small for some $b \in [1, \bar{b}]$ if i and j belong to the same group $\mathcal{G}_k$. Similar considerations suggest that $\delta_{ij}(b)$ tends to be large for all $b \in [1, \bar{b}]$ if i and j belong to different groups. As a consequence, we expect the statistic Δ_{ij} to be small/large if i and j belong to the same group/different groups.

4

1.2.2 HAC algorithm

The HAC algorithm based on the dissimilarity measure $\mathcal{D}$ proceeds as follows:

Step 0 (Initialization). Let $\hat{\mathcal{G}}_i^{[0]} = \{i\}$ denote the ith singleton cluster for $1 \leq i \leq n$ and define $\{\hat{\mathcal{G}}_1^{[0]}, \ldots, \hat{\mathcal{G}}_n^{[0]}\}$ to be the initial partition of time series into clusters.

Step r (Iteration). Let $\hat{\mathcal{G}}_1^{[r-1]}, \ldots, \hat{\mathcal{G}}_{n-(r-1)}^{[r-1]}$ be the $n - (r-1)$ clusters from the previous step. Determine the pair of clusters $\hat{\mathcal{G}}_k^{[r-1]}$ and $\hat{\mathcal{G}}_{k'}^{[r-1]}$ for which

$$\mathcal{D}(\hat{\mathcal{G}}_k^{[r-1]}, \hat{\mathcal{G}}_{k'}^{[r-1]}) = \min_{1 \leq l < l' \leq n-(r-1)} \mathcal{D}(\hat{\mathcal{G}}_l^{[r-1]}, \hat{\mathcal{G}}_{l'}^{[r-1]})$$

and merge them into a new cluster.

Iterating this procedure for $r = 1, \ldots, n-1$ yields a tree of nested partitions $\{\hat{\mathcal{G}}_1^{[r]}, \ldots, \hat{\mathcal{G}}_{n-r}^{[r]}\}$, which can be graphically represented by a dendrogram. Roughly speaking, the HAC algorithm merges the n singleton clusters $\hat{\mathcal{G}}_i^{[0]} = \{i\}$ step by step until we end up with the cluster $\{1, \ldots, n\}$. In each step of the algorithm, the closest two clusters are merged, where the distance between clusters is measured in terms of the dissimilarity $\mathcal{D}$.

@Oliver:
- We're not sure whether the construction in Section 2.1 gives a particularly good dissimilarity measure. We in particular wonder whether the statistics $\delta_{ij}(b)$ can be replaced by something else. E.g., one may use a version of $\delta_{ij}(b)$ with a different normalization than the integrals $\int_0^{1/b} \hat{m}_i(b \cdot v)dv/(1/b)$ and $\int_0^{1/b} \hat{m}_j(v)dv/(1/b)$.
- The procedure described in Sections 2.1 and 2.2 is essentially the idea from your notes (which are attached to this pdf). The main difference is that the constants c and b are not estimated. The constant c "drops out" when considering the statistics $\delta_{ij}(b)$ and the constant b is taken care of by minimizing over it.
- Directly estimating the constants seems to be difficult. Is it really possible to identify the constants as moments of the underlying density g_k and thus to estimate them as mentioned in your notes? We ran into the following problem when trying to do so: Applying the substitution rule to the integrals $\int_0^1 u^\ell m_i(u)du$ produces integrals in terms of the density g_k. However, the integrals do in general not run over the whole support of g_k (but only over some part of it which depends on the specific values of the constants b and a in your notes) ... Maybe, we've misunderstood something here. If the constants could indeed be identified and estimated in terms

2 Approach 2

2.1 Model

We consider a variation of the model (1.1) which can be formally stated as follows. Each time series $\mathcal{Y}_i$ is assumed to satisfy the nonparametric regression equation

$$Y_{it} = m_i\left(\frac{t}{\sqrt{T}}\right) + u_{it} \qquad (1 \leq t \leq T) \tag{2.1}$$

with $\mathbb{E}[u_{it}] = 0$, where m_i is an unknown smooth trend curve defined on $(-\infty, \infty)$. The difference between the models lies in the fact that we work with a different rescaling device ($t/\sqrt{T}$ versus t/T). Furthermore, we impose a slightly different group structure than before:

(G′) The set of countries $\{1, \ldots, n\}$ can be partitioned into K groups $\mathcal{G}'_1, \ldots, \mathcal{G}'_K$ such that

$$m_i \in \mathcal{F}'_k \quad \text{for all } i \in \mathcal{G}'_k,$$

where $\mathcal{F}'_k$ are function classes defined as

$$\mathcal{F}'_k := \Big\{ f : \mathbb{R} \to \mathbb{R}_+ \,\big|\, f = c \cdot g_k\Big(\frac{u - a}{b}\Big) \text{ with } a, b, c > 0 \text{ and}$$
$$g_k \text{ a density with mean 0 and variance 1} \Big\}.$$

For identification purposes, the classes are supposed to be distinct, i.e., $\mathcal{F}'_k \cap \mathcal{F}'_{k'} = \emptyset$ for any $k \neq k'$.

As before, the elements of the class $\mathcal{F}'_k$ are rescaled versions of some density function g_k (which is unknown in practice). We still regard c as a country-specific scaling parameter that accounts for the size of the country or population density and the constant b can be interpreted as the time parameter of the model. We also introduce an additional parameter a that accounts for the differences in the start of the pandemic. We briefly commented on the possibility of introducing such a parameter in Remark 1.2.

2.2 Clustering procedure

For $i \in \mathcal{G}_k$, we can write $m_i(u) = c_i \cdot g_k(\frac{u-a_i}{b_i})$ and use the assumptions on the density function g_k to obtain that

$$
\begin{aligned}
\int_{-\infty}^{\infty} m_i(u)du &= \int_{-\infty}^{\infty} c_i \cdot g_k\Big(\frac{u-a_i}{b_i}\Big)du \\
&= c_i \int_{-\infty}^{\infty} b_i \cdot g_k(v)dv = c_i b_i, \\
\int_{-\infty}^{\infty} u m_i(u)du &= \int_{-\infty}^{\infty} c_i \cdot u g_k\Big(\frac{u-a_i}{b_i}\Big)du \\
&= c_i b_i \int_{-\infty}^{\infty} (a_i + b_i v) \cdot g_k(v)dv = c_i b_i a_i, \\
\int_{-\infty}^{\infty} u^2 m_i(u)du &= \int_{-\infty}^{\infty} c_i \cdot u^2 g_k\Big(\frac{u-a_i}{b_i}\Big)du \\
&= c_i b_i \int_{-\infty}^{\infty} (a_i + b_i v)^2 \cdot g_k(v)dv = c_i b_i a_i^2 + c_i b_i b_i^2.
\end{aligned}
$$

With

$$
m_i^*(u) = \frac{m_i(u)}{\int_{-\infty}^{\infty} m_i(v)dv},
$$

the parameters a_i, b_i and c_i can thus be expressed as

$$
a_i = \int_{-\infty}^{\infty} u\, m_i^*(u)du,
$$

$$
b_i = \left\{ \int_{-\infty}^{\infty} u^2 m_i^*(u)du - \Big(\int_{-\infty}^{\infty} u\, m_i^*(u)du \Big)^2 \right\}^{1/2},
$$

$$
c_i = \frac{\int_{-\infty}^{\infty} m_i(u)du}{\left\{ \int_{-\infty}^{\infty} u^2 m_i^*(u)du - \Big(\int_{-\infty}^{\infty} u\, m_i^*(u)du \Big)^2 \right\}^{1/2}}.
$$

We now compute a dissimilarity measure as follows:

Step 1. For each i, estimate $m_i(u)$ by a Nadaraya-Watson estimator with a rectangular kernel and bandwidth h:

$$
\hat{m}_i(u) = \frac{\sum_{t=1}^{T} K\Big(\frac{u-t/\sqrt{T}}{h}\Big) Y_{it}}{\sum_{t=1}^{T} K\Big(\frac{u-t\sqrt{T}}{h}\Big)} \tag{2.2}
$$

with K being a rectangular kernel and $K_h(x) = K(x/h)/h$. Here, we choose $h = 3.5/\sqrt{T}, 7/\sqrt{T}, \ldots$, and as before, this choice of bandwidth allows us to take care of possible weekly cycles in the data which are produced by delays

7

in reporting new cases over the weekend. We then estimate $m_i^*(u)$ and the parameters a_i, b_i and c_i by simply plugging (2.2) into the corresponding formulas:

$$\hat{m}_i^*(u) = \frac{\hat{m}_i(u)}{\int_{-\infty}^{\infty} \hat{m}_i(v) dv},$$

$$\hat{a}_i = \int_{-\infty}^{\infty} u\, \hat{m}_i^*(u) du,$$

$$\hat{b}_i = \left\{ \int_{-\infty}^{\infty} u^2 \hat{m}_i^*(u) du - \left(\int_{-\infty}^{\infty} u\, \hat{m}_i^*(u) du \right)^2 \right\}^{1/2},$$

$$\hat{c}_i = \frac{\int_{-\infty}^{\infty} \hat{m}_i(u) du}{\left\{ \int_{-\infty}^{\infty} u^2 \hat{m}_i^*(u) du - \left(\int_{-\infty}^{\infty} u\, \hat{m}_i^*(u) du \right)^2 \right\}^{1/2}}.$$

Step 2. By definition, for all $i \in \mathcal{G}_k$, we have

$$m_i(a_i + b_i\, v)/c_i = g_k(v).$$

Compute the estimator

$$\hat{p}_i(v) = \hat{m}_i(\hat{a}_i + \hat{b}_i\, v)/\hat{c}_i$$

and its normalized version

$$\hat{p}_i^*(v) = \frac{\hat{p}_i(v)}{\int_{-\infty}^{\infty} \hat{p}_i(w) dw}. \tag{2.3}$$

Note that $\hat{p}_i^*(v)$ is a proper density function since $\hat{p}_i^*(v) \geq 0$ by construction and

$$\int_{-\infty}^{\infty} \hat{p}_i^*(v) dv = \int_{-\infty}^{\infty} \frac{\hat{p}_i(v)}{\int_{-\infty}^{\infty} \hat{p}_i(w) dw} dv = 1.$$

Step 3. For a given pair of countries (i, j), define the statistic

$$\Delta'_{ij} = \int_{-\infty}^{\infty} \left(\sqrt{\hat{p}_i^*(v)} - \sqrt{\hat{p}_j^*(v)} \right)^2 dv,$$

which measures the (squared) Hellinger distance between the functions $m_i(a_i + b_i\, v)/c_i$ and $m_j(a_j + b_j\, v)/c_j$.

Step 4. Let $S \subseteq \{1, \ldots, n\}$ and $S' \subseteq \{1, \ldots, n\}$ be two sets of time series from our sample. As before, we work with the complete linkage measure of dissimi-

larity defined as

$$\mathcal{D}'(S, S') = \max_{i \in S, j \in S'} \Delta'_{ij}.$$

Alternatively, we may use a single or average linkage measure.

To obtain cluster estimates, we now run the HAC algorithm from the previous section based on $\mathcal{D}'$ rather than $\mathcal{D}$.

3   ToDos

3.1   Estimating the number of clusters

Approach 1. Not clear to me how to choose the number of clusters.

Approach 2. For $K = 1, 2, \ldots$, let $\{\hat{G}_1^{[K]}, \ldots, \hat{G}_K^{[K]}\}$ be the estimated groups. The group-specific densities g_k can be estimated by

$$\hat{g}_k^{[K]}(v) = \frac{1}{|\hat{G}_k^{[K]}|} \sum_{i \in \hat{G}_k^{[K]}} \hat{p}_i^*(v)$$

for $k = 1, \ldots, K$ with $\hat{p}_i^*(v)$ defined in (2.3). Use these to compute

$$\hat{\sigma}_i^2(K) = \frac{1}{T} \sum_{t=1}^{T} \left\{ Y_{it} - \hat{c}_i \hat{g}_k^{[K]} \left(\frac{t/\sqrt{T} - \hat{a}_i}{\hat{b}_i} \right) \right\}^2 \quad \text{for } i \in \hat{G}_k^{[K]}$$

and choose the number of clusters by maximizing the BIC criterion

$$\mathrm{BIC}(K) = T \sum_{i=1}^{n} \log(\hat{\sigma}_i^2(K)) - \log(nT) \left(K(n + T) + n \right).$$

3.2   Classification by deaths

In terms of the application, it is also of interest to do the classification by deaths to see the overlap between cases and deaths.

Figures 4 and 10 present the choropleth maps for the the results of the classification by deaths for the number of clusters $K = 12$. We can directly compare the results of classification by cases and classification by deaths: Figure 3 vs. Figure 4 and Figures 9 vs. Figure 10 for Approaches 1 and 2 respectively.

3.3 Determining common factors

We need to identify the common factors that place the countries in a particular grouping and determine their acceleration constants. One could do a logit or multinomial fit of the category (i.e., group membership of country i) on country specific predictors like demographics, income level, health care expenditure, lock down severity.

3.4 Other distance measures (in Approach 2)

Instead of the Hellinger distance, one may use

$$\Delta''_{ij} = \frac{\langle \hat{p}^*_i, \hat{p}^*_j \rangle}{\|\hat{p}^*_i\| \|\hat{p}^*_j\|}.$$

3.5 Clustering other features of the data

This section may be more about enhancing the application.

The trend is the dominant feature of the data but seasonality and comovement are also of interest. For example, suppose that

$$u_{it} = \sum_{j=1}^{J} \beta_{ij} D_{jt} + \varepsilon_{it},$$

where $J = 7$ for daily data, while $E(\varepsilon_{it}) = 0$ and

$$E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij}.$$

We may estimate $B = (\beta_{ij})$ and $\Sigma = (\sigma_{ij})$ using the time series data country by country. The question is whether there are distinct groups with regard to B, Σ. Specifically, by arranging the countries in groups determined by the trend, is Σ close to block diagonal? Can we also divide B into groups as in for group k there are coefficients $\gamma_{k1}, \ldots, \gamma_{kJ}$ and every member of group k has

$$\beta_{ij} = c_i \times \gamma_{kj}.$$

This reduces dimensionality from $n \times J$ to $n + K \times J$. This is a standard regression clustering.

4 Empirical analysis

In Section 4.1, we assess the finite sample performance of our clustering method by Monte-Carlo experiments. In Section 4.2, we apply the method to a sample of COVID-19 data from 104 different countries.

4.1 Simulation

Not done yet.

4.2 Analysis of COVID-19 data

4.2.1 Data

We analyze data from 160 countries. We consider only those countries that have a total number of at least $1\,000$ cases of infection during the considered time period. We also restrict our attention only to the countries for which the observed time period is around 300 days. For each country i, we observe a time series $\mathcal{Y}_i = \{Y_{it} : 1 \leq t \leq T\}$, where Y_{it} is the number of newly confirmed COVID-19 cases in country i on day t. The data are freely available in the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at John Hopkins University (`https://github.com/CSSEGISandData/COVID-19`) and were downloaded on 29 April 2021. For the application of Approach 1, we take the first Monday after reaching 100 confirmed cases in each country as the starting date $t = 1$, whereas for the application of Approach 2, we take the first day after reaching 100 confirmed cases in each country as the starting date $t = 1$. Beginning the time series of each country on the day when that country reached 100 confirmed cases is a common way of "normalizing" the data (see e.g. Cohen and Kupferschmidt, 2020). Additionally aligning the data by Monday allows to take care of possible weekly cycles in the data which are produced by delays in reporting new cases over the weekend. The time series length T is taken to be the longest interval for which we have observations for all 160 countries. The resulting dataset thus consists of $n = 160$ time series, each with $T = 296$ and $T = 301$ observations in case of Approach 1 or Approach 2 respectively. Some of the time series contain negative values which we replaced by 0. Overall, this resulted in 45 replacements.

4.2.2 Results for Approach 1

We assume that the data Y_{it} of each country i in our sample follow the nonparametric trend model from equation (1.1) and we impose the group structure (G) on the data. To implement the HAC algorithm, we make the following choices:

- We take $\bar{b} = 2$ in the definition of the classes $\mathcal{F}_k$, which is more than sufficient for our purposes: As discussed above, b is some sort of effective time parameter. Suppose that $m_j(u) = m_i(b \cdot u)$ for some $b > 1$ and all $u \in [0, 1/b]$. In this case, the trend m_j evolves more quickly than m_i by the factor b. By setting $\bar{b} = 2$, we allow the trend m_j of country j to evolve at most twice as quickly as the trend m_i of country i.

- We use a rectangular kernel $K(\cdot)$ to compute the Nadaraya-Watson smoothers $\hat{m}_i$ and consider the bandwidth $h = 7/T$, which corresponds to effective sample sizes of 14 days of data.

- We assume that the number of classes K is equal to 12. We also produce the choropleth maps for $K = 10$ and $K = 15$ clusters. (We haven't considered the problem of estimating K yet. So far, K is handpicked. @Oliver: Any ideas how to estimate K???)

- The values $\delta_{ij}(b)$ and $\delta_{ji}(b)$ are calculated not for all $b \in [1, \bar{b}]$, but on a grid that consists of the values $\{1, 1.05, 1.1, \ldots, \bar{b}\}$. Since this grid is finite, we write $\min_{b \in [1, \bar{b}]} \delta_{ij}(b)$ and $\min_{b \in [1, \bar{b}]} \delta_{ji}(b)$ instead of $\inf_{b \in [1, \bar{b}]} \delta_{ij}(b)$ and $\inf_{b \in [1, \bar{b}]} \delta_{ji}(b)$ in what follows.

The estimation results are presented in Figures 1–6:

- Figures 1–3 show the estimated clusters in a world map for different numbers of clusters $K = 10, 15, 12$, where each country is coloured according to the class it belongs to. Figure 4 presents the results of classification by deaths for $K = 12$.

- Figures 5 shows the dendrogram of the HAC algorithm with a border drawn around each of $K = 12$ clusters.

- Figures 6a–6l depict the trend functions of the countries in the different estimated classes, each figure corresponding to a specific class. Since the functions m_i in a given class $\mathcal{G}_k$ are only identical up to the scaling by the constants c and b, we do not plot the (estimated) functions m_i themselves. We rather proceed as

follows: Consider a specific class $\mathcal{G}_k$, write the trends in this class as $m_i(u) = c_i \cdot g_k(b_i \cdot u)$ with some constants c_i and b_i, and fix some country $i_k \in \mathcal{G}_k$. It holds that $m_i(u) = \{c_{i_k}/c_i\} \cdot m_i(\{b_{i_k}/b_i\}u)$ for all $i \in \mathcal{G}_k$. Hence, the functions $\{c_{i_k}/c_i\} \cdot m_i(\{b_{i_k}/b_i\}u)$ are identical for all $i \in \mathcal{G}_k$. Rather than plotting the estimates $\hat{m}_i(u)$, we thus pick some $i_k \in \hat{\mathcal{G}}_k$ and plot the renormalized functions $\{\hat{c}_{i_k}/\hat{c}_i\} \cdot \hat{m}_i(\{\hat{b}_{i_k}/\hat{b}_i\}u)$, where $\hat{c}_{i_k}/\hat{c}_i$ and $\hat{b}_{i_k}/\hat{b}_i$ are estimates of the ratios c_{i_k}/c_i and b_{i_k}/b_i. (The index i_k is chosen such that $\hat{b}_{i_k}/\hat{b}_i < 1$ for all $i \in \hat{\mathcal{G}}_k$.)
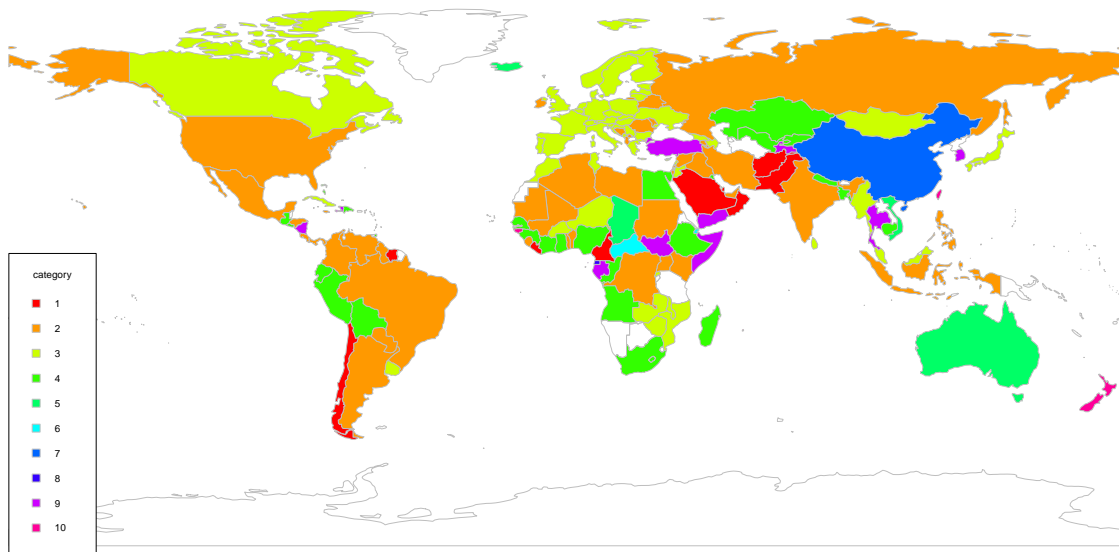
Figure 1: Results of HAC for 10 clusters (Approach 1) on a map: each country is coloured according to the group it belongs to.
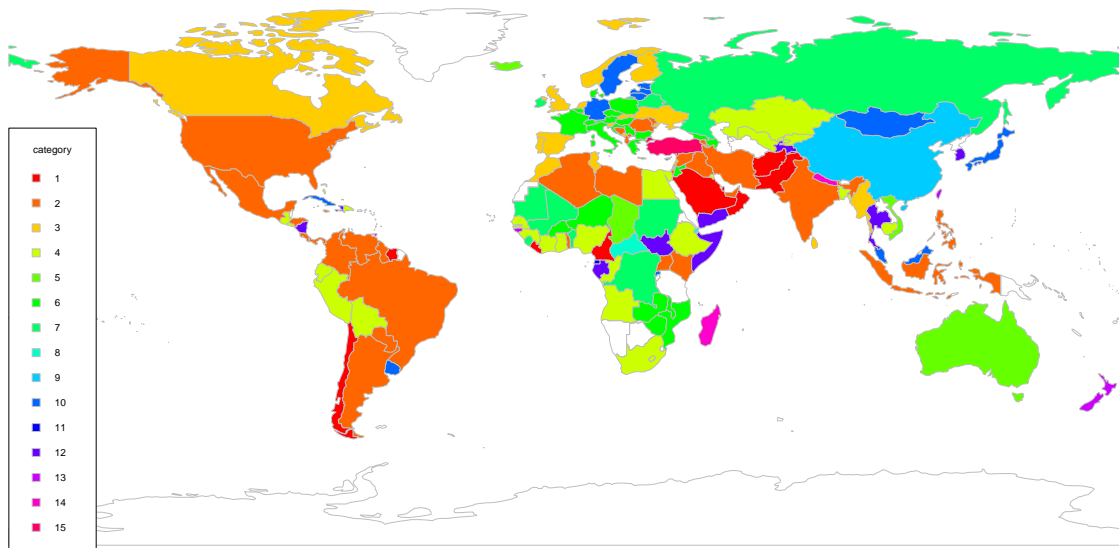
Figure 2: Results of HAC for 15 clusters (Approach 1) on a map: each country is coloured according to the group it belongs to.
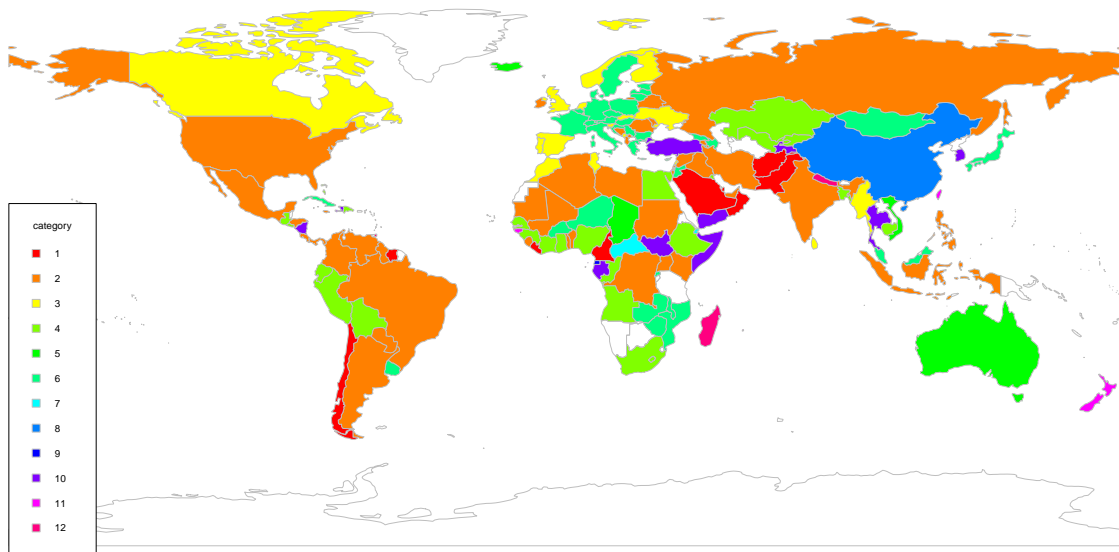
Figure 3: Results of HAC for the number of cases for 12 clusters (Approach 1) on a map: each country is coloured according to the group it belongs to.
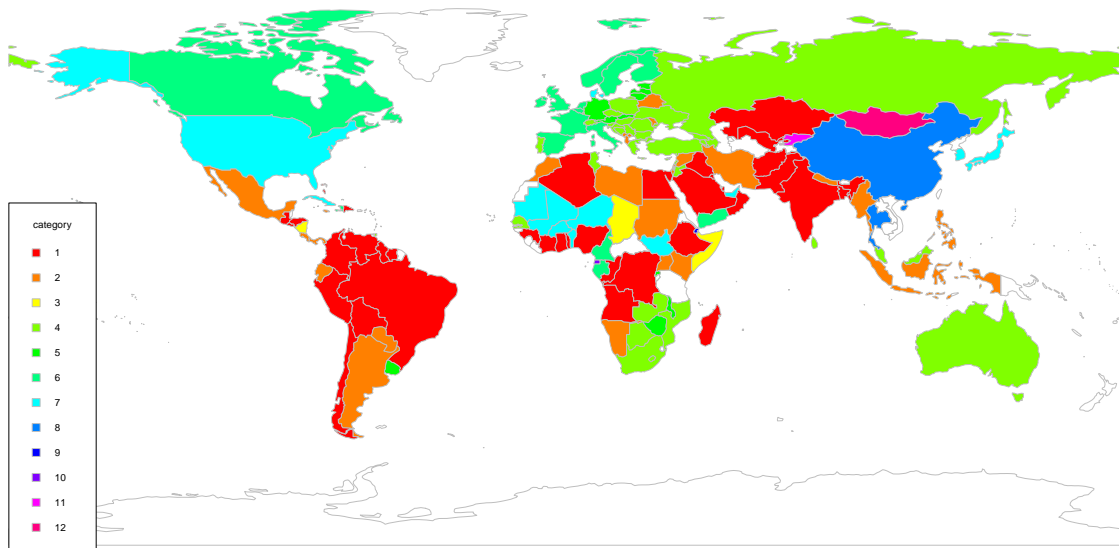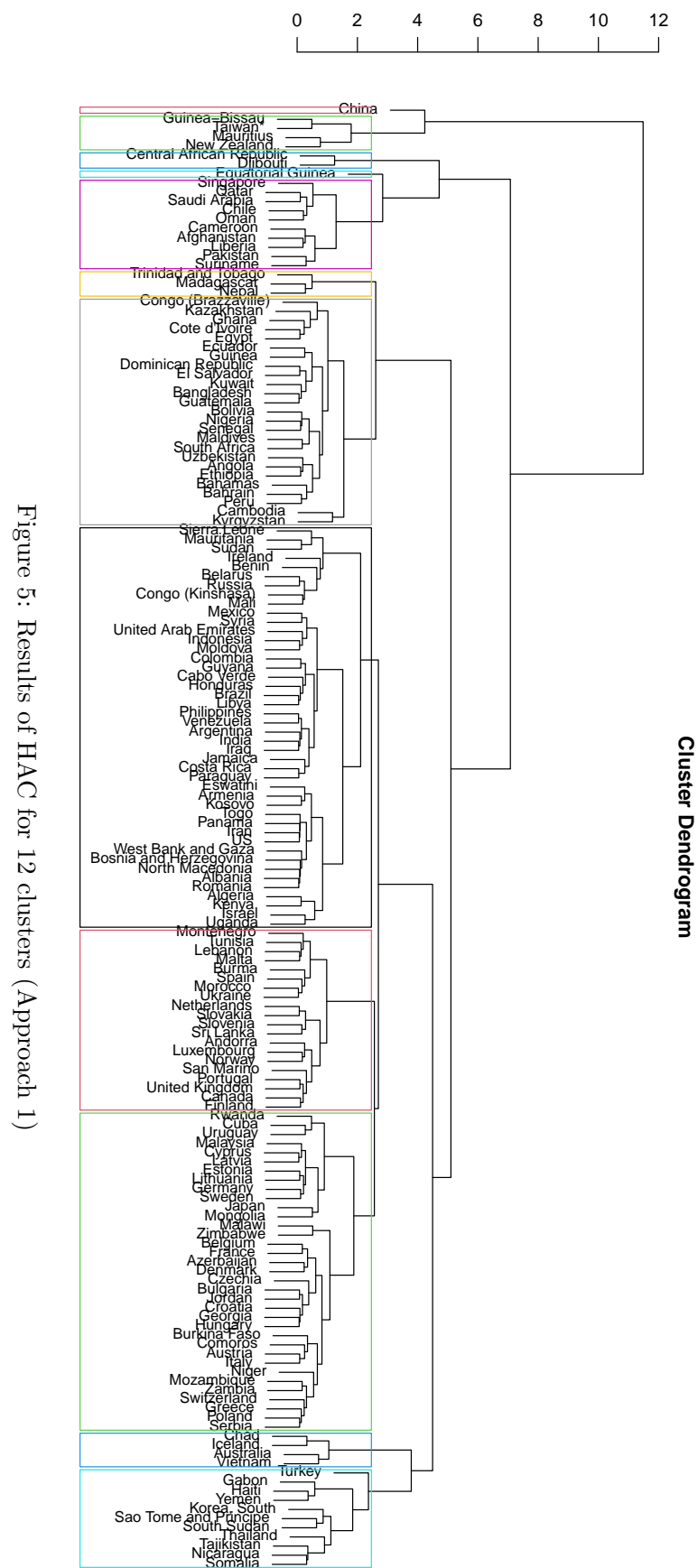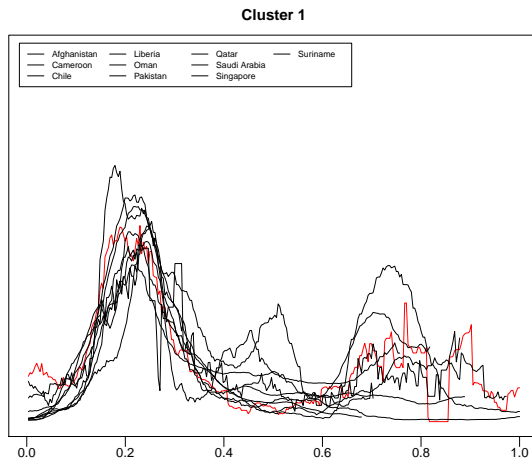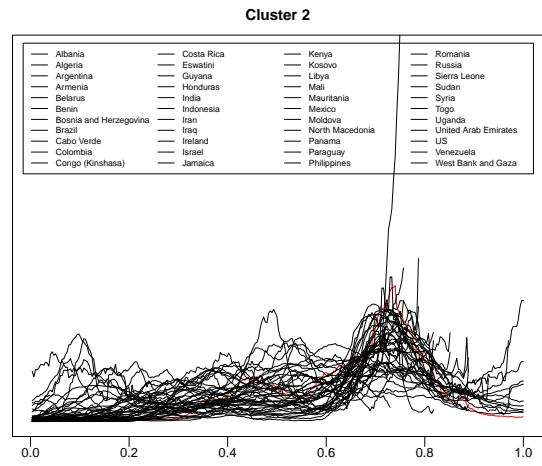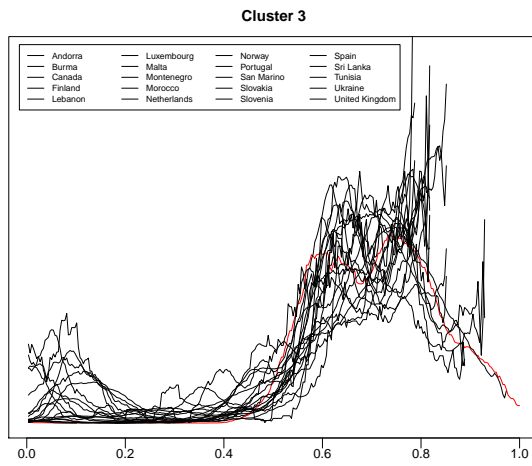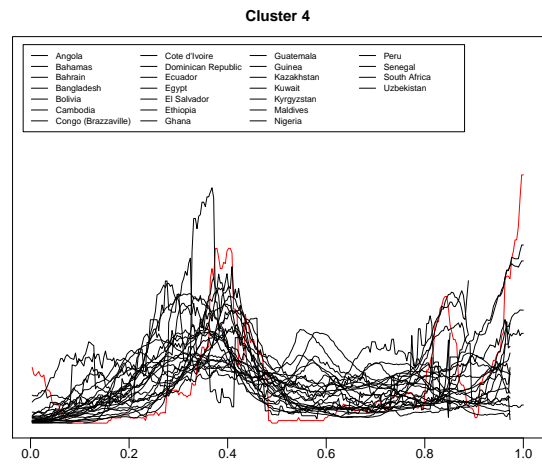

Figure 4: Results of HAC for the number of deaths for 12 clusters (Approach 1) on a map: each country is coloured according to the group it belongs to.

Figure 5: Results of HAC for 12 clusters (Approach 1)

16

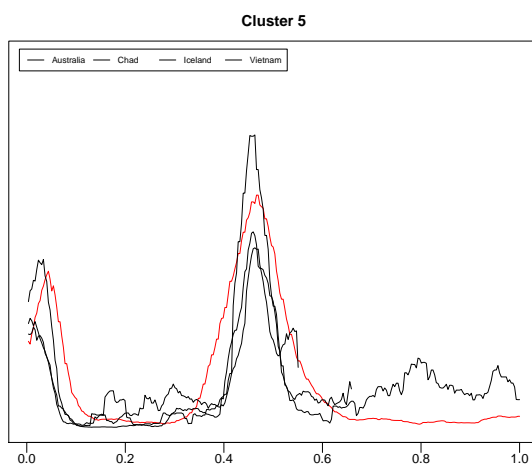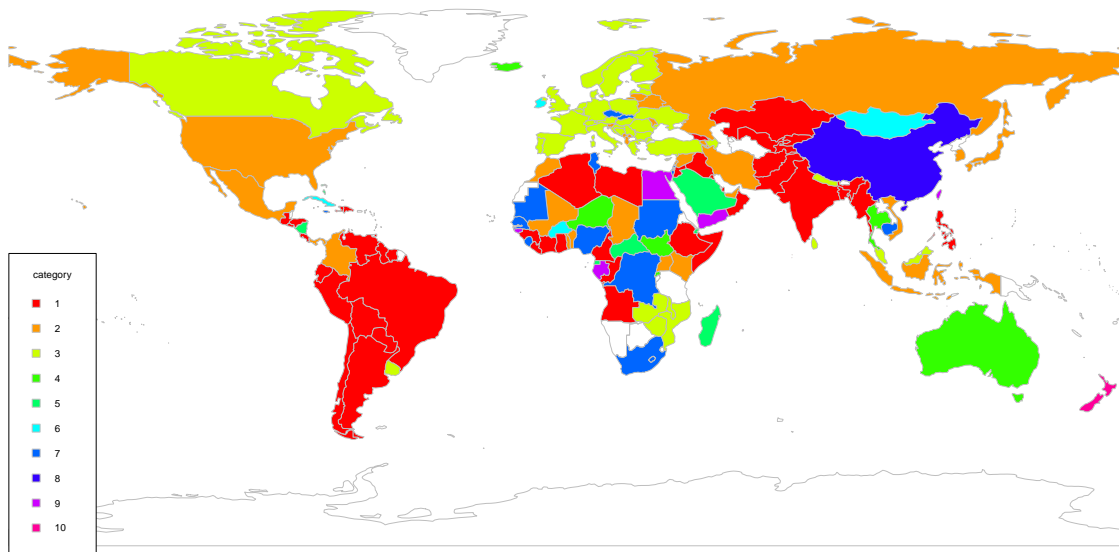## Cluster 1

Afghanistan — Liberia — Qatar — Suriname
Cameroon — Oman — Saudi Arabia
Chile — Pakistan — Singapore

Cluster 2

Albania — Costa Rica — Kenya — Romania
Algeria — Eswatini — Kosovo — Russia
Argentina — Guyana — Libya — Sierra Leone
Armenia — Honduras — Mali — Sudan
Belarus — India — Mauritania — Syria
Benin — Indonesia — Mexico — Togo
Bosnia and Herzegovina — Iran — Moldova — Uganda
Brazil — Iraq — North Macedonia — United Arab Emirates
Cabo Verde — Ireland — Panama — US
Colombia — Israel — Paraguay — Venezuela
Congo (Kinshasa) — Jamaica — Philippines — West Bank and Gaza

(a)   (b)

Cluster 3

Andorra — Luxembourg — Norway — Spain
Burma — Malta — Portugal — Sri Lanka
Canada — Montenegro — San Marino — Tunisia
Finland — Morocco — Slovakia — Ukraine
Lebanon — Netherlands — Slovenia — United Kingdom

Cluster 4

Angola — Cote d'Ivoire — Guatemala — Peru
Bahamas — Dominican Republic — Guinea — Senegal
Bahrain — Ecuador — Kazakhstan — South Africa
Bangladesh — Egypt — Kuwait — Uzbekistan
Bolivia — El Salvador — Kyrgyzstan
Cambodia — Ethiopia — Maldives
Congo (Brazzaville) — Ghana — Nigeria

(c)   (d)

Cluster 5

Australia — Chad — Iceland — Vietnam

Cluster 6

Austria — Czechia — Japan — Poland
Azerbaijan — Denmark — Jordan — Rwanda
Belgium — Estonia — Latvia — Serbia
Bulgaria — France — Lithuania — Sweden
Burkina Faso — Georgia — Malawi — Switzerland
Comoros — Germany — Malaysia — Uruguay
Croatia — Greece — Mongolia — Zambia
Cuba — Hungary — Mozambique — Zimbabwe
Cyprus — Italy — Niger

(e)   (f)

17

Figure 6: Clusters produced by the algorithm (Approach 1). Each panel presents appropriately scaled curve estimates $\hat{m}_i$ that belong to a particular cluster. The bandwidth h is taken to be $7/T$.

4.2.3   Results for Approach 2

We assume that the data Y_{it} of each country i in our sample follow the nonparametric trend model (2.1) and we impose the group structure (G$'$) on the data. To implement the HAC algorithm, we make the following choices:

- We use a rectangular kernel $K(\cdot)$ to compute the Nadaraya-Watson smoothers $\hat{m}_i$ from (2.2) and consider the bandwidth $h = 7/\sqrt{T}$, which corresponds to an effective sample size of 14 days of data.

- We assume that the number of classes K is equal to 12. We also produce the choropleth maps for $K = 10$ and $K = 15$ clusters.

The estimation results are presented in Figures 7–12:

- Figures 7–9 show the estimated clusters in a world map for different numbers of clusters $K = 10, 15, 12$, where each country is coloured according to the class it belongs to. Figure 10 presents the results of classification by deaths for $K = 12$.

- Figures 11 shows the dendrogram of the HAC algorithm with a border drawn around each of $K = 12$ clusters.

- Figures 12a–12l depict the trend functions of the countries in the different estimated classes, each figure corresponding to a specific class. Since the functions m_i in a given class $\mathcal{G}_k$ are only identical up to the scaling by the constants a, b and c, we do not plot the (estimated) functions m_i themselves. We rather plot the estimates $\hat{p}_i(v) = \hat{m}_i(\hat{a}_i + \hat{b}_i \, v)/\hat{c}_i$. Moreover, even though m_i is defined on $(-\infty, \infty)$, we restrict our attention to $[-5, 5]$, which seems to be enough for all of the countries.

Figure 7: Results of HAC for 10 clusters (Approach 2) on a map: each country is coloured according to the group it belongs to.
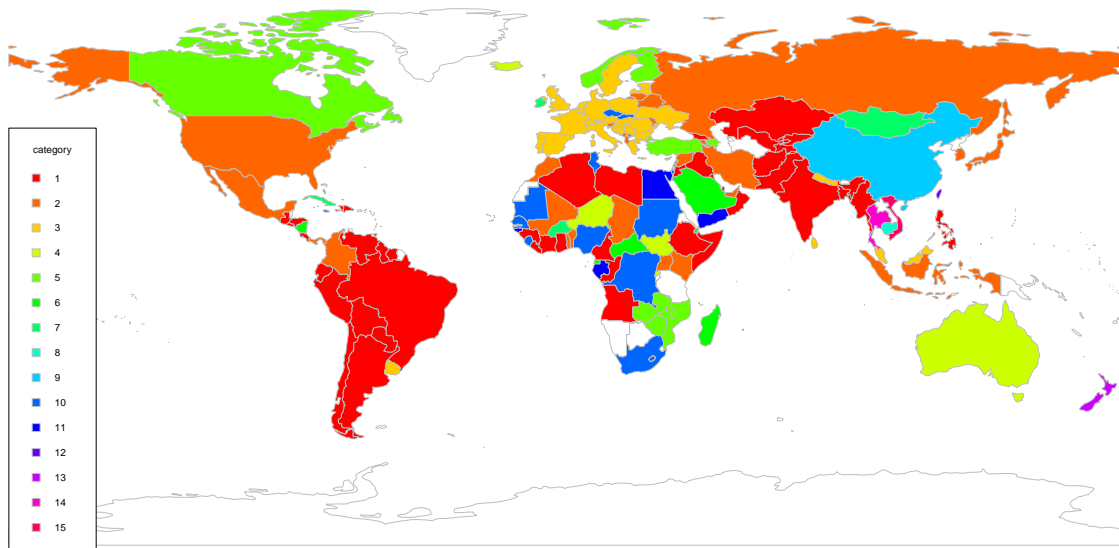

Figure 8: Results of HAC for 15 clusters (Approach 2) on a map: each country is coloured according to the group it belongs to.
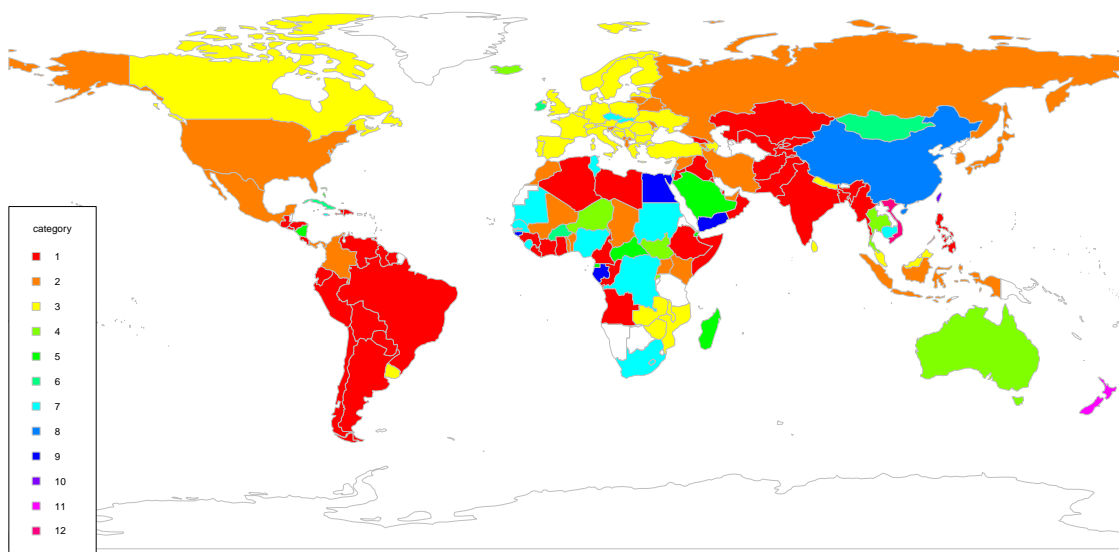
Figure 9: Results of HAC for the number of cases for 12 clusters (Approach 2) on a map: each country is coloured according to the group it belongs to.
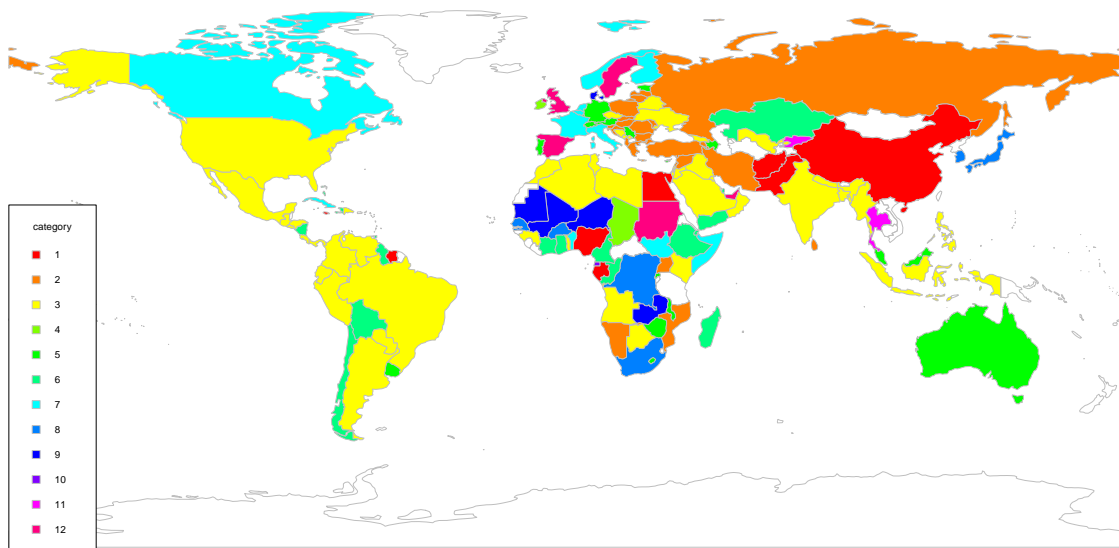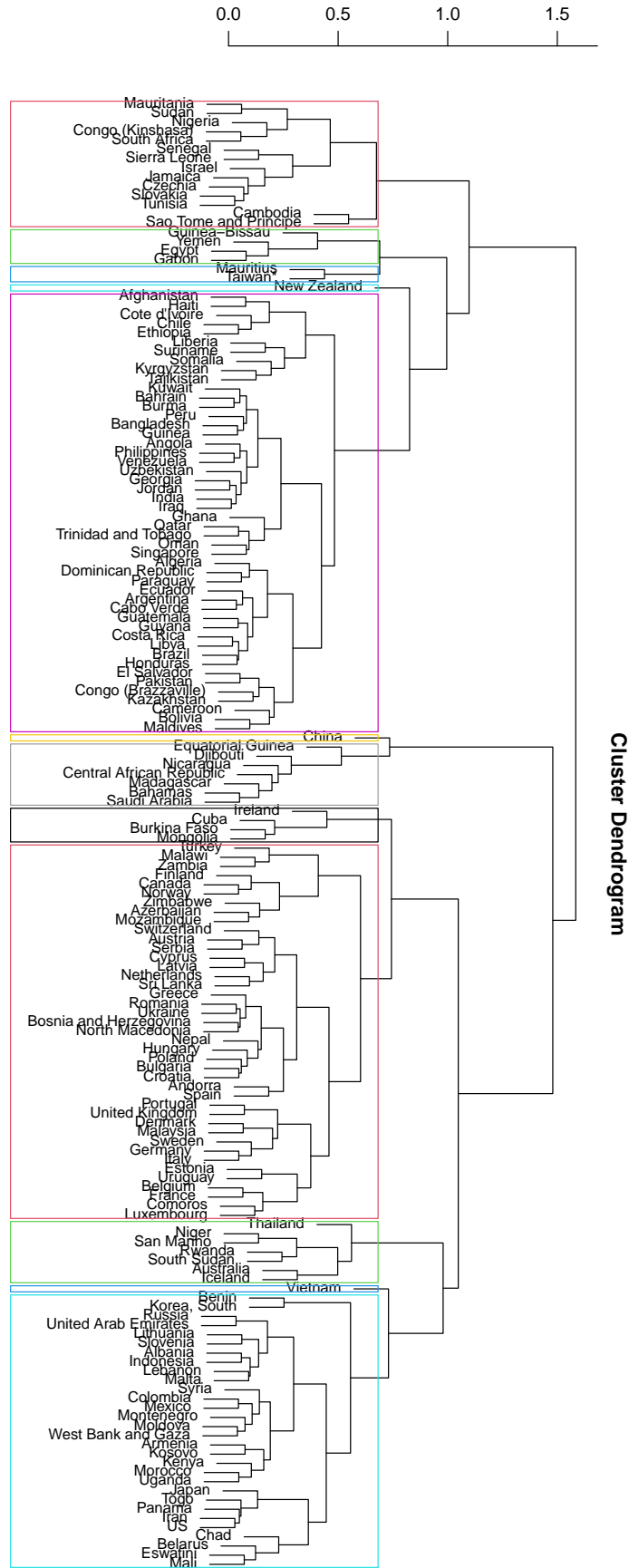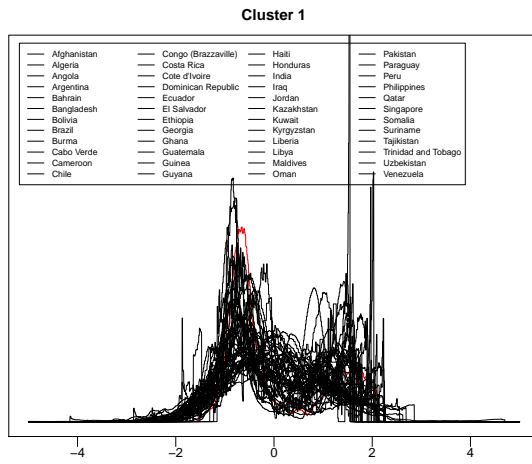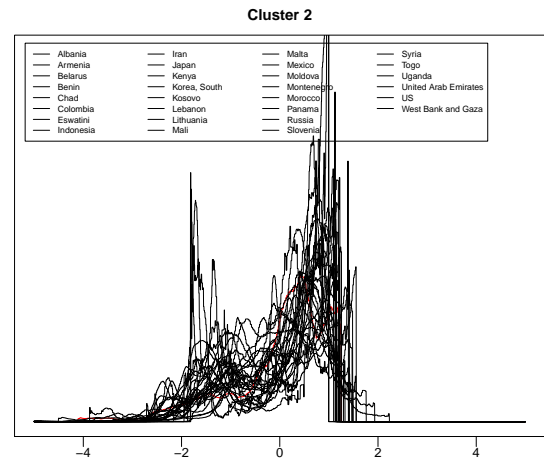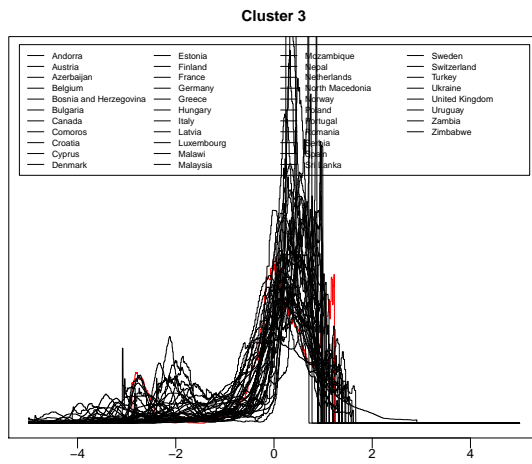

Figure 10: Results of HAC for the number of deaths for 12 clusters (Approach 2) on a map: each country is coloured according to the group it belongs to.
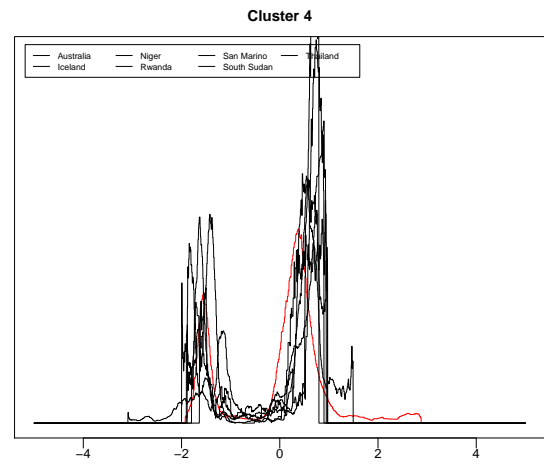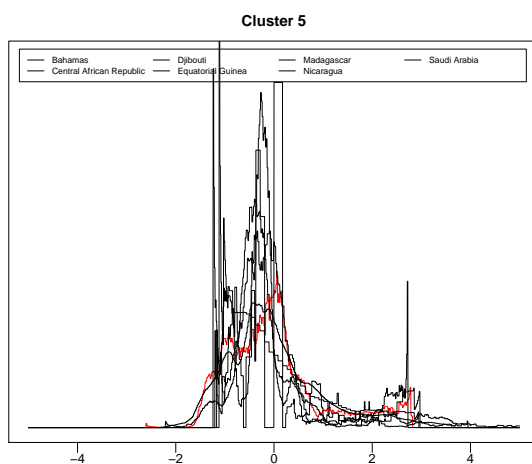
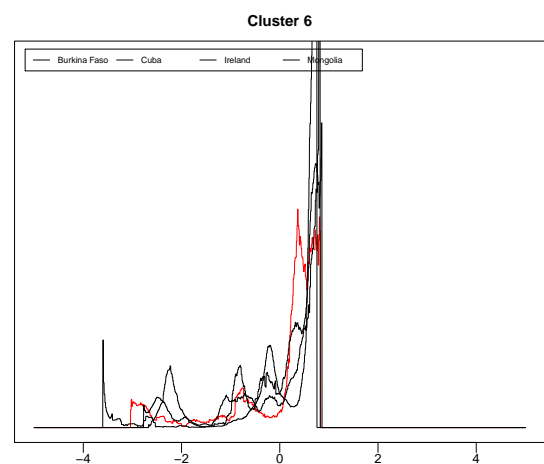Figure 11: Results of HAC for 12 clusters (Approach 2)
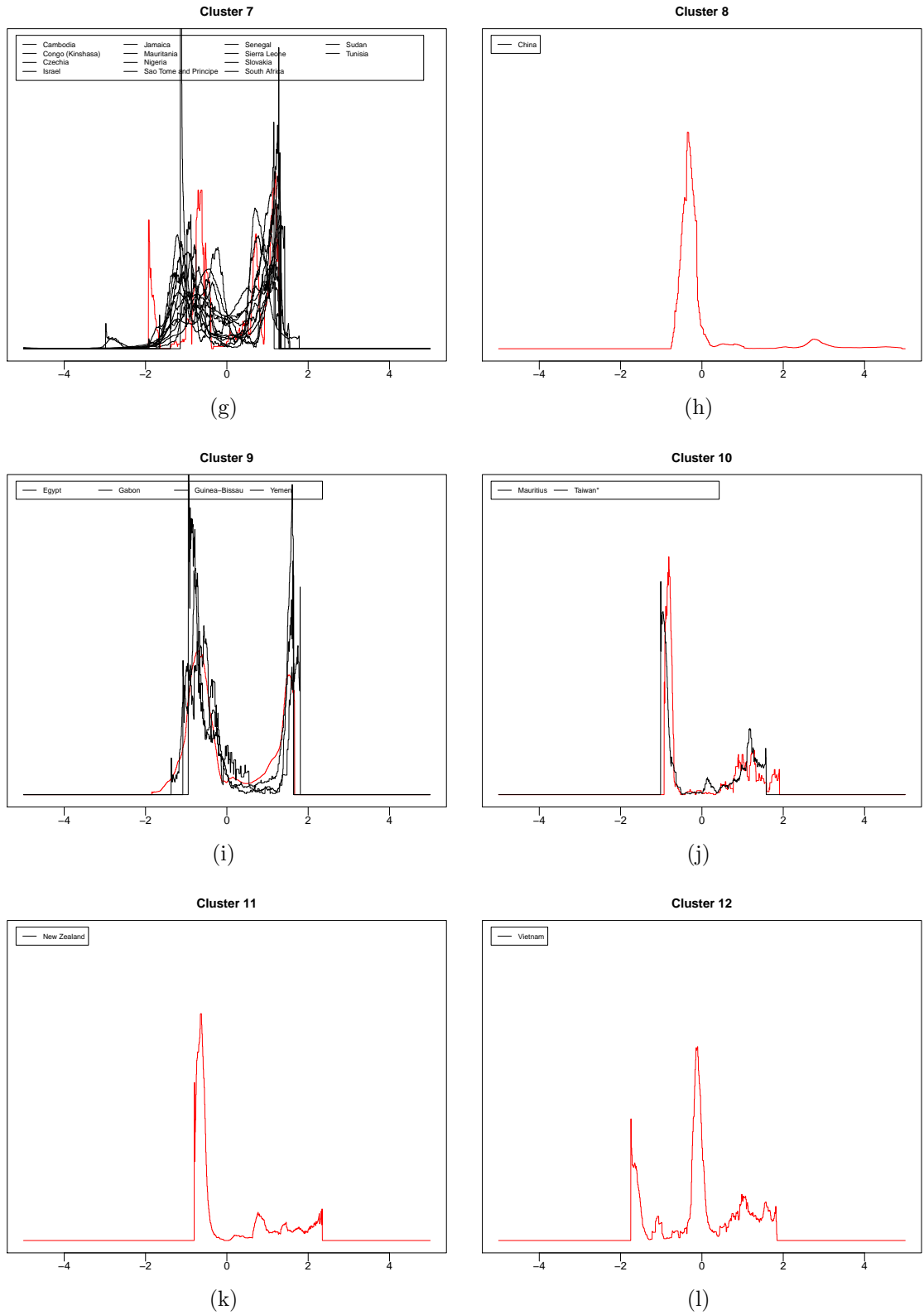
(a)

(b)

(c)

(d)

(e)

(f)

Figure 12: Clusters produced by the algorithm (Approach 2). Each panel presents appropriately scaled curve estimates $\hat{m}_i$ that belong to a particular cluster. The bandwidth h is taken to be $7/T$.

References

COHEN, J. and KUPFERSCHMIDT, K. (2020). Countries test tactics in 'war' against COVID-19. *Science*, **367** 1287–1288.

ROBINSON, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change* (P. Hackl, ed.). Springer, 253–264.