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

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 $\mathcal{Y}_i = \{Y_{it} : 1 \leq t \leq T\}$ 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

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + u_{it} \qquad (1 \leq t \leq T) \tag{1.1}$$

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 $\{1, \ldots, n\}$ 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

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

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.

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.

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

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

Step 4. Let $S \subseteq \{1,\dots,n\}$ and $S' \subseteq \{1,\dots,n\}$ 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.

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 of the moments of the underlying densities, this would be much simpler and better. So it would be great if you could help with that.

3 Empirical analysis

In Section 3.1, we assess the finite sample performance of our clustering method by Monte-Carlo experiments. In Section 3.2, we apply the method to a sample of COVID-19 data from 104 different countries.

3.1 Simulation

Not done yet.

3.2 Analysis of COVID-19 data

We analyze data from 104 countries. We consider only those countries that have a total number of at least 20 000 cases of infection during the considered time period. 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 on the homepage of the European Center for Disease Prevention and Control (`https://www.ecdc.europa.eu`) and were downloaded on 25 February 2021.[1] As already mentioned in the Introduction, we take the first Monday 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 104 countries. The resulting dataset thus consists of $n = 104$ time series, each with $T = 192$ observations. Some of the time series contain negative values which we replaced by 0. Overall, this resulted in 14 replacements.

[1] ECDC switched to a weekly reporting schedule for the COVID-19 situation on 17 December 2020. Hence, all daily updates have been discontinued from 14 December. The downloaded daily data set thus presents historical data until 14 December 2020.

We assume that the data Y_{it} of each country i in our sample follow the nonparametric trend model

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + u_{it}$$

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 to compute the Nadaraya-Watson smoothers $\hat{m}_i$ and consider the bandwidths $h = 3.5/T$ and $h = 7/T$, which corresponds to effective sample sizes of 7 and 14 days of data respectively.

- We assume that the number of classes K is equal to 7. (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 $\{1, 1.01, 1.02, \ldots, \bar{b}\}$. 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 for the bandwidth choice $h = 3.5/T$ are presented in Figures 1–6:

- Figure 1 shows the dendrogram of the HAC algorithm for $h = 3.5/T$. The different colours of the countries correspond to the different classes they belong to.

- Figures 2a–2g 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 3 shows the estimated clusters in a world map. As in Figure 1, the different colours of the countries correspond to the different classes they belong to.

Figures 4–6 present the corresponding results for $h = 7/T$.

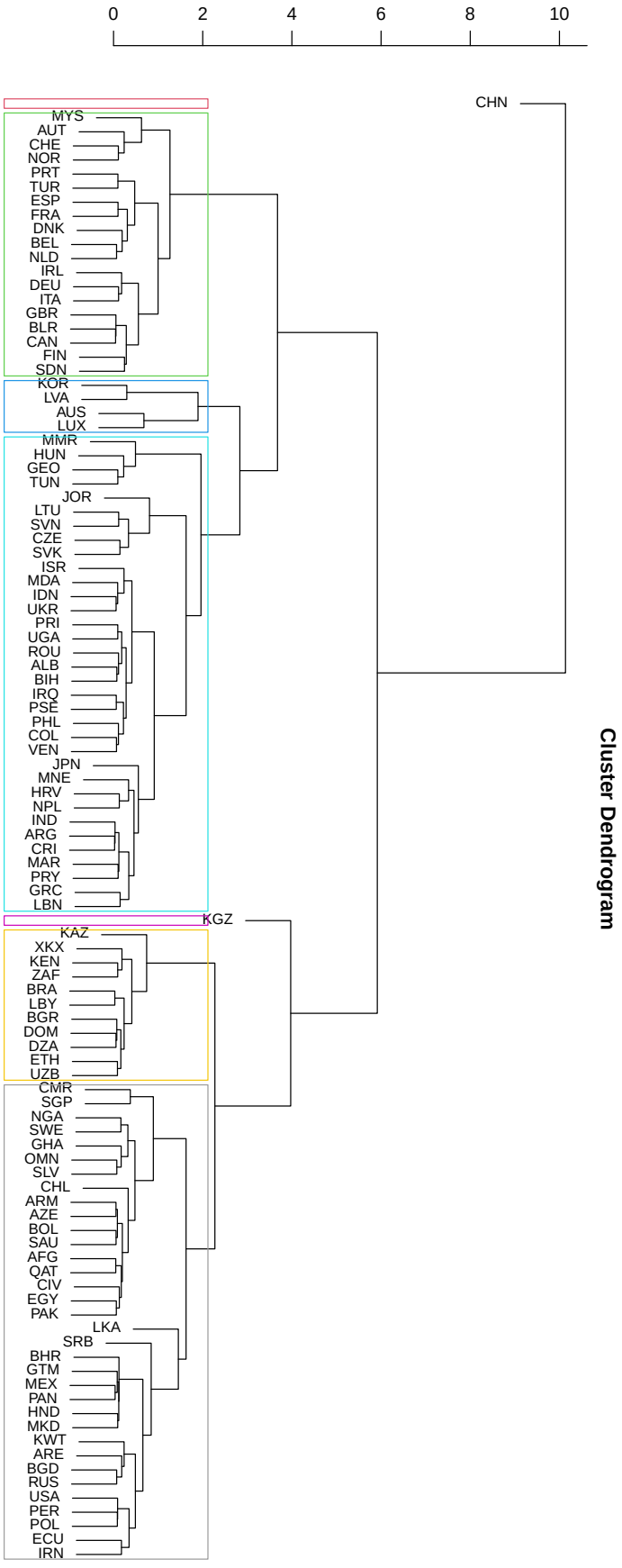
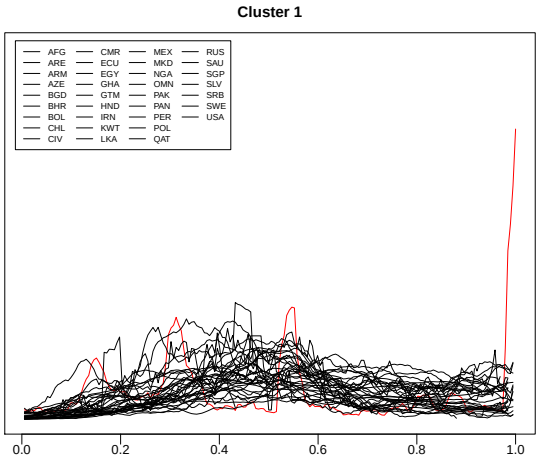
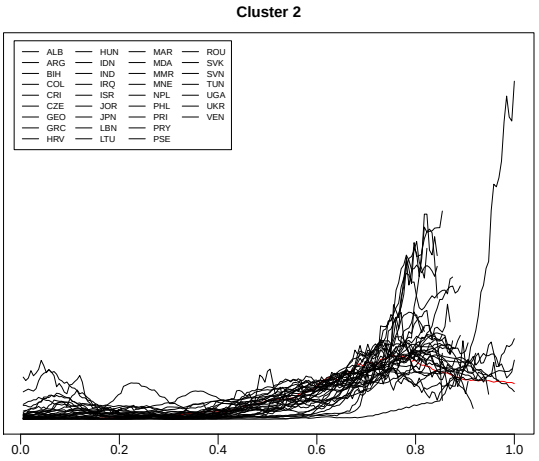


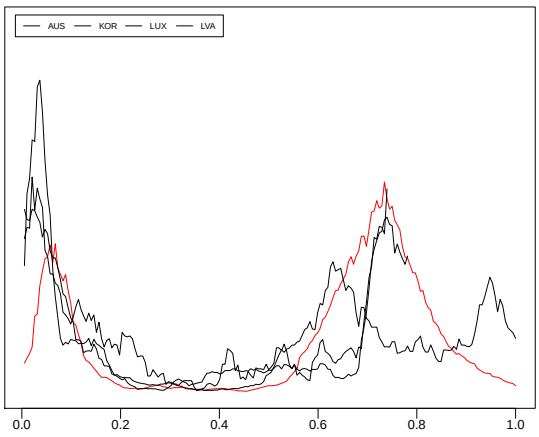
Figure 1: Results of HAC for $h = 3.5/T$.


Cluster 1
AFG ARE ARM AZE BGD BHR BOL CHL CIV
CMR ECU EGY GHA GTM HND IRN KWT LKA
MEX MKD NGA OMN PAK PAN PER POL QAT
RUS SAU SGP SLV SRB SWE USA
0.0 0.2 0.4 0.6 0.8 1.0

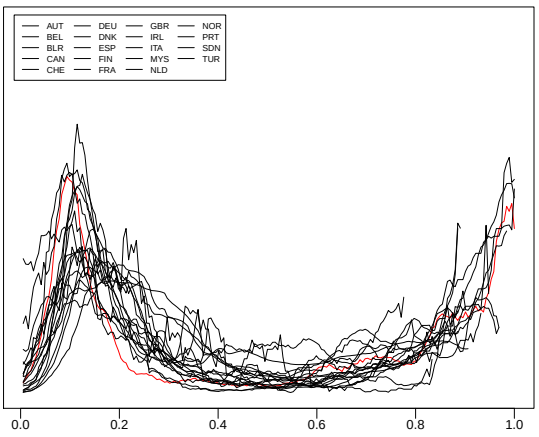
(a)


Cluster 2
ALB ARG BIH COL CRI CZE GEO GRC HRV
HUN IDN IND IRQ ISR JOR JPN LBN LTU
MAR MDA MMR MNE NPL PHL PRI PRY PSE
ROU SVK SVN TUN UGA UKR VEN
0.0 0.2 0.4 0.6 0.8 1.0

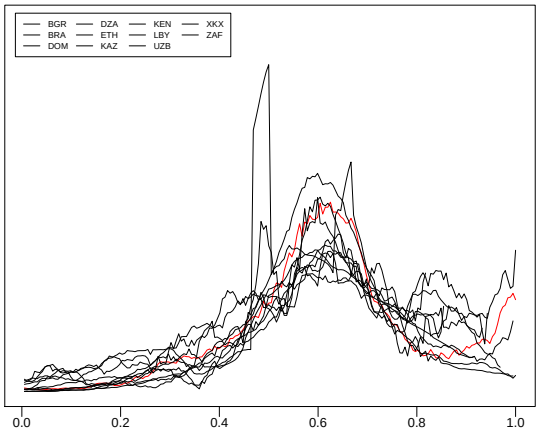
(b)


Cluster 3
AUS KOR LUX LVA
0.0 0.2 0.4 0.6 0.8 1.0

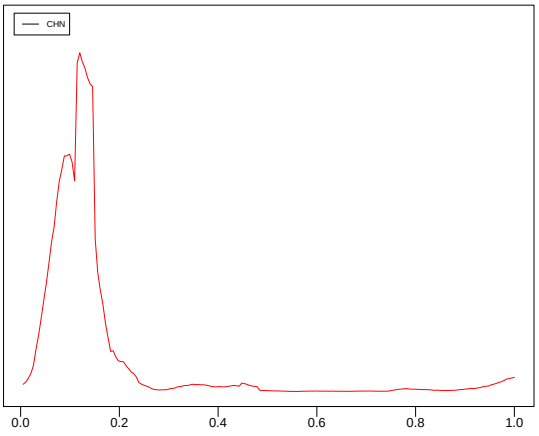
(c)


Cluster 4
AUT BEL BLR CAN CHE
DEU DNK ESP FIN FRA
GBR IRL ITA MYS NLD
NOR PRT SDN TUR
0.0 0.2 0.4 0.6 0.8 1.0

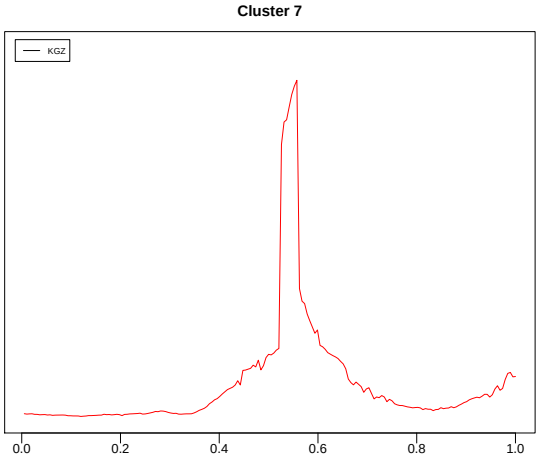
(d)


Cluster 5
BGR BRA DOM
DZA ETH KAZ
KEN LBY UZB
XKX ZAF
0.0 0.2 0.4 0.6 0.8 1.0

(e)


Cluster 6
CHN
0.0 0.2 0.4 0.6 0.8 1.0

(f)



(g)

Figure 2: Clusters produced by the algorithm. Each panel presents appropriately scaled curve estimates $\hat{m}_i$ that belong to a particular cluster. The bandwidth h is taken to be $3.5/T$.

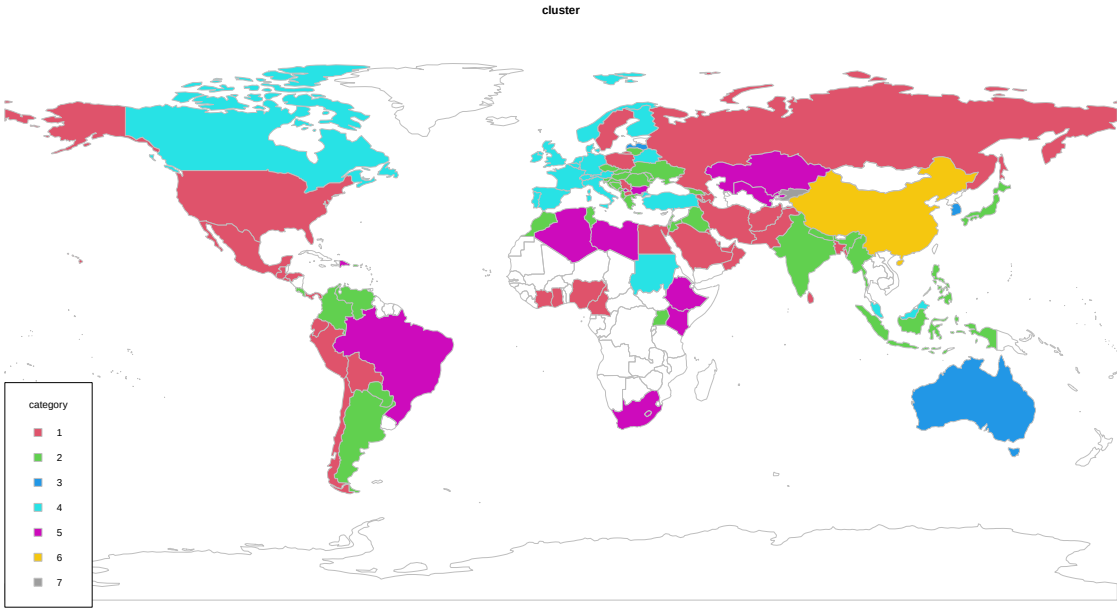


Figure 3: Results of HAC for $h = 3.5/T$ on a map: each country is coloured according to the group it belongs to.

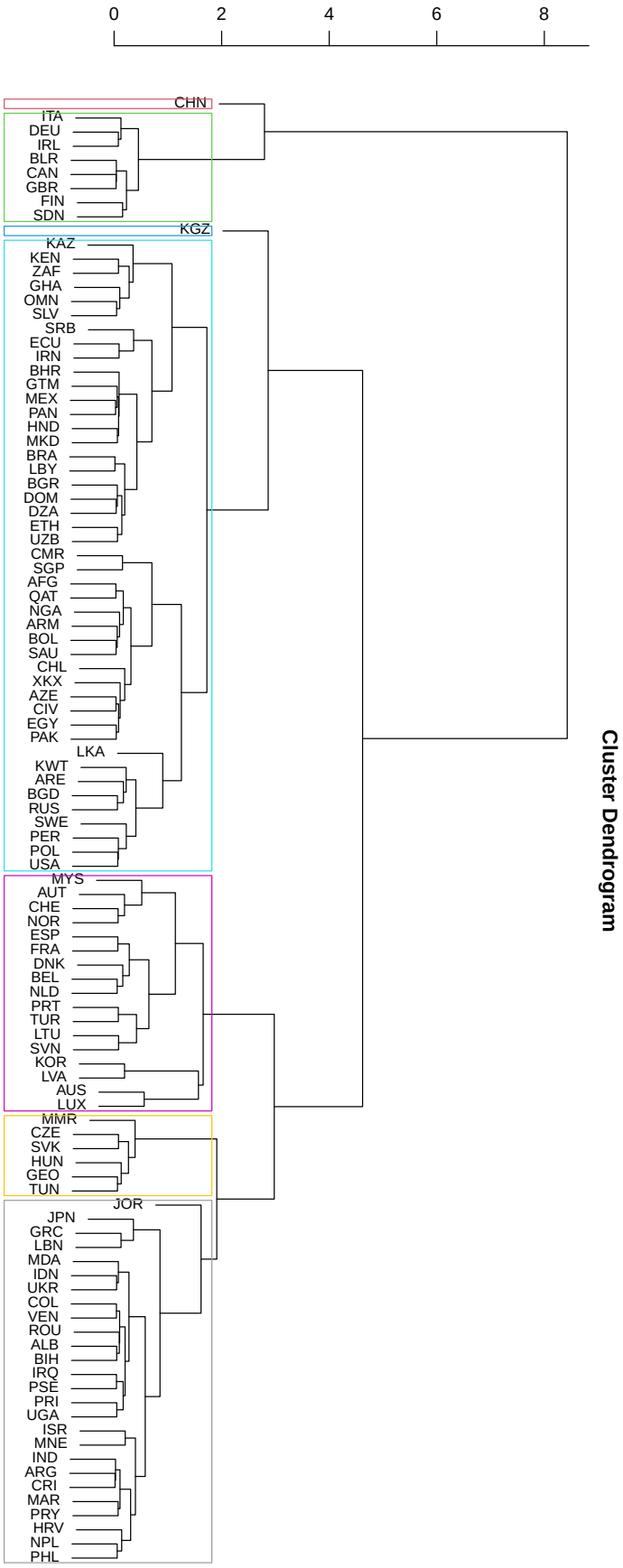
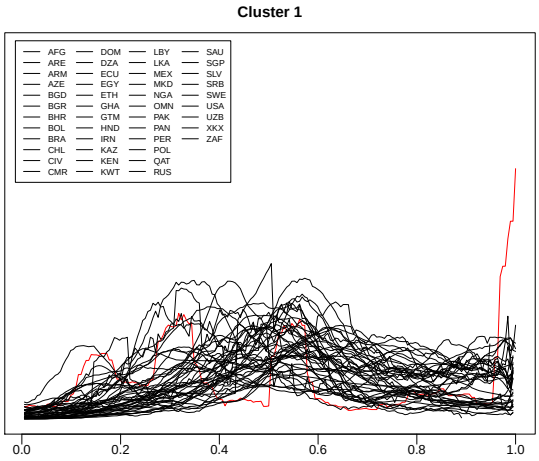
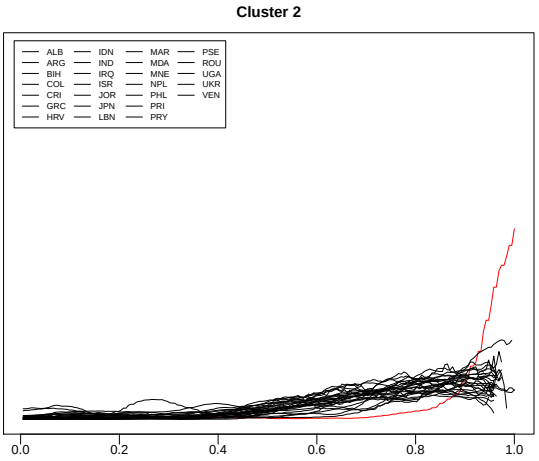


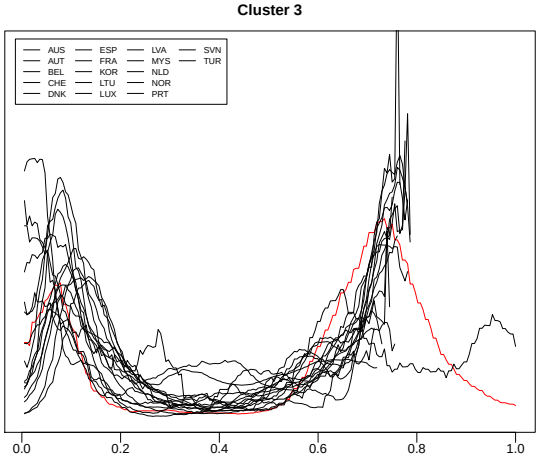
Figure 4: Results of HAC for $h = 7/T$.


Cluster 1
AFG DOM LBY SAU
ARE DZA LKA SGP
ARM ECU MEX SLV
AZE EGY MKD SRB
BGD ETH NGA SWE
BGR GHA OMN USA
BHR GTM PAK UZB
BOL HND PAN XXX
BRA IRN PER ZAF
CHL KAZ POL
CIV KEN QAT
CMR KWT RUS
0.0 0.2 0.4 0.6 0.8 1.0

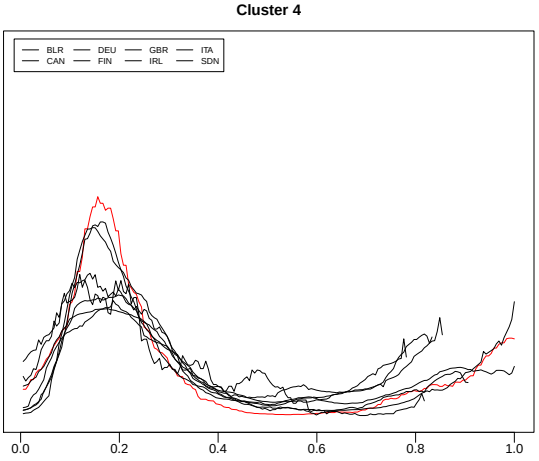
(a)


Cluster 2
ALB IDN MAR PSE
ARG IND MDA ROU
BIH IRQ MNE UGA
COL ISR NPL UKR
CRI JOR PHL VEN
GRC JPN PRI
HRV LBN PRY
0.0 0.2 0.4 0.6 0.8 1.0

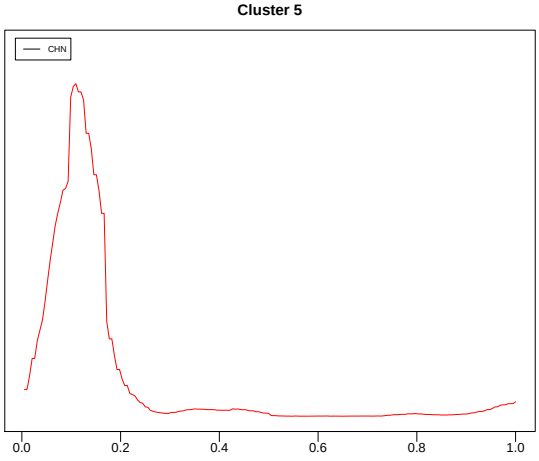
(b)


Cluster 3
AUS ESP LVA SVN
AUT FRA MYS TUR
BEL KOR NLD
CHE LTU NOR
DNK LUX PRT
0.0 0.2 0.4 0.6 0.8 1.0

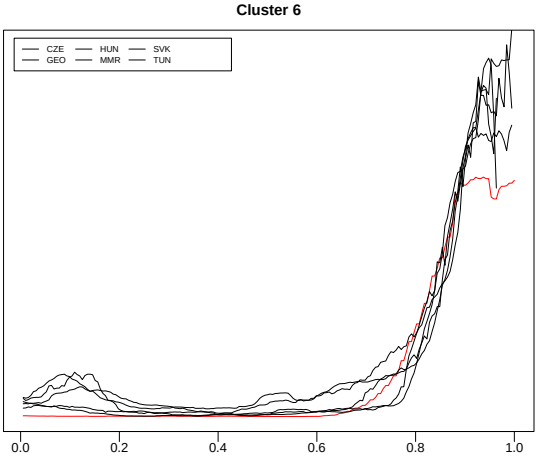
(c)


Cluster 4
BLR DEU GBR ITA
CAN FIN IRL SDN
0.0 0.2 0.4 0.6 0.8 1.0

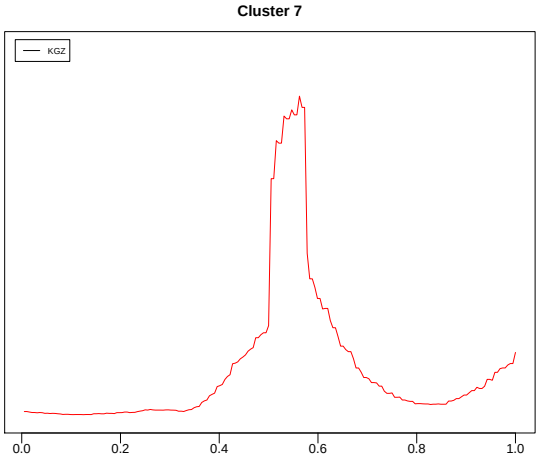
(d)


Cluster 5
CHN
0.0 0.2 0.4 0.6 0.8 1.0

(e)


Cluster 6
CZE HUN SVK
GEO MMR TUN
0.0 0.2 0.4 0.6 0.8 1.0

(f)



(g)

Figure 5: Clusters produced by the algorithm. Each panel presents appropriately scaled curve estimates $\hat{m}_i$ that belong to a particular cluster. The bandwidth h is taken to be $7/T$.

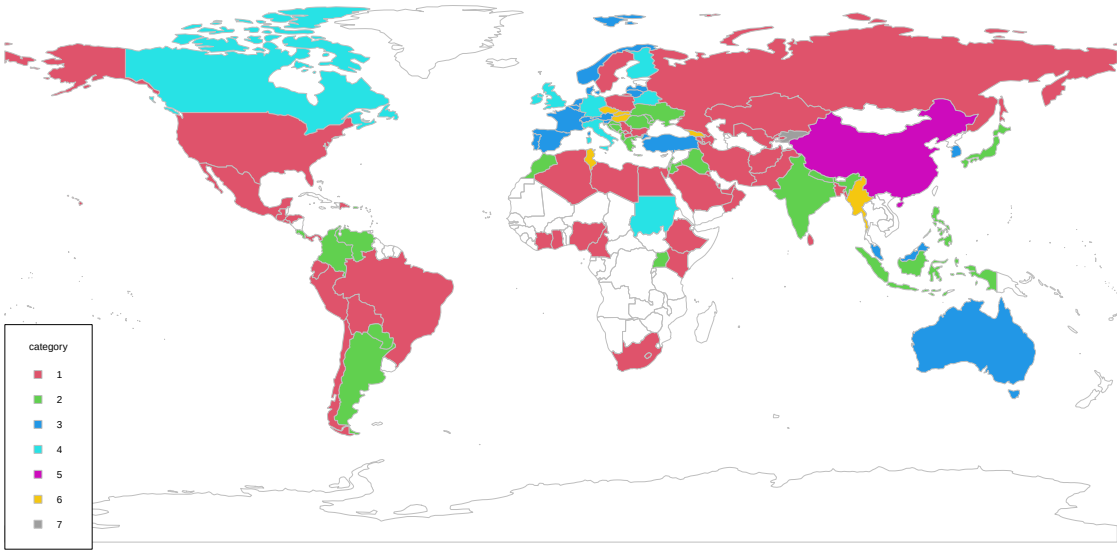


Figure 6: Results of HAC for $h = 7/T$ on a map: each country is coloured according to the group it belongs to.

References

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

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

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.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, **58** 236–244.

As we know, There are known knowns. There are things we know we know. We also know There are known unknowns. That is to say, We know there are some things We do not know. But there are also unknown unknowns, The ones we don't know We don't know. Donald Rumsfeld, U.S. Secretary of Defence

Suppose that

$$y_{it} = m_i(t/T) + u_{it},$$

where $i = 1, \ldots, n$ and $t = 1, \ldots, T$; both n and T are large. We suppose that there exists J classes of functions

$$\mathcal{G}_j = \{f : f(u) = cg_j((u-a)/b),\ a, b, c \in \mathbb{R}_+,\ g_j \text{ a density}\}.$$

We suppose that for any i, $m_i(\cdot) \in \mathcal{G}_j$ for some j, that is, for some j there exists a_i, b_i, c_i with

$$m_i(u) = c_i g_j((u-a_i)/b_i)/b_i$$

for all $u \in [0,1]$.

Estimation. Suppose that there is only one group with unknown $g(.)$. Given unrestricted estimates $\widehat{m}_i(\cdot)$, we can estimate c_i by $\int_0^1 \widehat{m}_i(u)du$ and work with the ratio $\widehat{m}_i^*(u) = \widehat{m}_i(u)/\int_0^1 \widehat{m}_i(u)du$. We may make different assumptions here about a, b. For example, a is the mean and b is the standard deviation of the density g. In that case we can estimate a_i by $\int_0^1 u\widehat{m}_i^*(u)du$ and b_i by the square root of $\int_0^1 u^2\widehat{m}_i^*(u)du - (\int_0^1 u\widehat{m}_i^*(u)du)^2$. Alternatively, median and interquartile range work. Then one can estimate $g(u)$ by

$$\frac{1}{n}\sum_{i=1}^{n} \widehat{b}_i \widehat{m}_i^* \left(u\widehat{b}_i + \frac{\widehat{a}_i}{\widehat{b}_i}\right).$$

Now suppose that there are multiple $g's$. The recovery of the constants a, b, c only uses the individual regression function. But now the uncertainty is around which i to average over. This can be addressed by the clustering algorithms.

The parameter estimates $\widehat{a}_i, \widehat{b}_i, \widehat{c}_i$ are $\sqrt{T}$ consistent, whereas the estimates of g_j will be $\sqrt{nTh}$ consistent.