Analysing deaths and confirmed cases of COVID-19 pandemic by analytical approaches

Ray-Ming Chen

School of Mathematics and Statistics, Baise University, 21, Zhongshan No. 2 Road, Baise, Guangxi Province, China

1 Introduction

For the time being, there are over 180 million confirmed cases and around 4 million deaths of COVID-19 [1]. The variants of COVID-19 are still in full swing and spreading across the continents [2,3]. The spread of this pandemic is studied by many researchers [4–7]. There are many ways to look into the behaviours of the viruses or the pandemic itself [8,9], including for the efficacy of travel ban or lockdown [10]. Nowadays more and more papers are focusing on the efficacy and effectiveness of vaccine against all sorts of variants of SARS-CoV-2 [11,12]. Some researches even help shed light on resolving the pandemic via herd immunity [13] or map out the dynamical trajectory of the viruses [14,15]. Among those research, there are plenty of qualitative and quantitative research methods, in particular statistical methods: regression-related models [16], Mann-Whitney-U tests, Mann-Kendall tests, Spearman’s rho, etc [17,18]; factorial design [19]; artificial intelligence (AI) based methods [20,21] and some deep learning techniques in COVID-19 diagnosis [22,23]. Coupling with AI, an automatic reasoning for searching the hidden features of the trend of COVID-19 is also vital [24]. Some researches even have established the relations between cases and deaths of COVID-19 from demographic, economic and social perspectives [25].

An orthonormal basis [26] $\mathbb{B}_N$ (precisely $\mathbb{B}_{36}$), which is motivated by Fourier analysis [27], is employed. The underlying frequencies of data are taken into consideration. The COVID-19 database [28] which recorded the weekly COVID-19 cases and deaths from Week 15 to Week 51 (37 weeks in total) is utilised. By filtering out some non-essential data (countries), we obtain 43 countries as the research targets. By calculating the 36 (from Week 15 to Week 50) growth rates of the cases and deaths for each sampled country, we have a national case vector and a death vector. By transforming these two vectors into a set of coefficients, which is the result of the inner products via $\mathbb{B}_{36}$, we start to rank the coefficients by their face values and form ranking vectors for each country. The ranks indicate the strength (relation) between the growth rates and the underlying frequencies—a larger coefficient will be assigned a larger rank. Then the Manhattan metric [29] is applied to measure the distances between all the ranking vectors and yield a Manhattan distance matrix. Based on this matrix, we associate each country with its closer and further neighbours via the minimal pairing and maximal pairing. Then we re-analyse the collected growth rates of confirmed cases and deaths with another typical approach: cosine similarity. Finally, an approximation method based on spectral analysis to further predict the development of confirmed cases and deaths is devised.

In sum, the current work shows that:

- the patterned evolutionary correlation between countries not random, i.e., there are some fundamental factors that contribute to such relation;
• the correlated patterns for cases and deaths between countries bear no similarity at all;
• there is a strong discrepancy between evolution of cases and the one of deaths;
• the development of confirmed cases and deaths among countries are monotonic—some might increase the total cases, while some would decrease the total ones, if the detected features are preserved.

Based on the clustering technique in this study, it shall offer some knowledge for the policymakers in adopting the sensible measures for controlling the pandemic.

2 Methodology and procedures

By expanding the concept of the Fourier analysis, an orthogonal basis - which is served as the feature benchmarks to conduct the current study is introduced. Let \( \mathbb{N} \) denote the set of positive integers and \( \mathbb{R} \) denote the set of real numbers. For any \( \mathbf{e} \in \mathbb{R}^n \), \( \mathbf{e}_i \) is used to denote its \( i \)-th element, \( |\mathbf{e}| \) is used to denote its length and \( ||\mathbf{e}|| \) is used to denote the Euclidean norm. Assume \( |\mathbf{e}| = N+1 \), where \( N \) stands for a natural number. \( \Delta \mathbf{e} \) is used to denote its growth vector, i.e.,

\[
\Delta \mathbf{e} = \left( \frac{e_2 - e_1}{e_1}, \frac{e_3 - e_2}{e_2}, \ldots, \frac{e_{N+1} - e_N}{e_N} \right).
\]

Observe that \(|\Delta \mathbf{e}| = N\). This growth vector is the main research target, since we study the (weekly) grow rates of cases and deaths regarding COVID-19. For any two vectors \( \mathbf{e}, \mathbf{f} \in \mathbb{R}^n \), \( < \mathbf{e}, \mathbf{f} > \), or \( \mathbf{e} \cdot \mathbf{f} \) is used to denote their inner product.

**Orthogonal basis** Define real functions \( I(x) = \sqrt{\frac{N}{2}} \) and \( b_m(x) := \sqrt{\frac{2}{N}} \cos(\frac{2\pi}{N} \cdot m \cdot x) \) and \( \hat{b}_m(x) := \sqrt{\frac{2}{N}} \cdot \sin(\frac{2\pi}{N} \cdot m \cdot x) \), where \( x \in \{1, 2, \ldots, N\} \). Define \( \mathbb{B}_N \) as:

\[
\mathbb{B}_N := \left\{ \{i, b_1, \hat{b}_1, b_2, \hat{b}_2, \ldots, b_m, \hat{b}_m, \ldots, b_{\frac{N}{2}+1}, \hat{b}_{\frac{N}{2}+1}, \hat{b}_N \}, \{i, \hat{b}_1, b_1, \hat{b}_2, b_2, \ldots, \hat{b}_m, b_m, \ldots, \hat{b}_{\frac{N}{2}+1}, b_{\frac{N}{2}+1}, \hat{b}_N \} \right\} , \quad \text{if } N \text{ is even};
\]

\[
\mathbb{B}_N := \left\{ \{i, b_1, \hat{b}_1, b_2, \hat{b}_2, \ldots, b_m, \hat{b}_m, \ldots, b_{\frac{N-1}{2}+1}, \hat{b}_{\frac{N-1}{2}+1}, \hat{b}_N \}, \{i, \hat{b}_1, b_1, \hat{b}_2, b_2, \ldots, \hat{b}_m, b_m, \ldots, \hat{b}_{\frac{N-1}{2}+1}, b_{\frac{N-1}{2}+1}, \hat{b}_N \} \right\} , \quad \text{if } N \text{ is odd};
\]

where \( \hat{b}_N := \sqrt{\frac{1}{2}} \cdot b_N \). By some manipulation of mathematical operations, \( \mathbb{B}_N \) is proved to be an orthogonal basis for all natural number \( N \).

**Manhattan metric and cosine similarity** Let \( \mathbf{e}, \mathbf{f} \in \mathbb{R}^n \) be arbitrary. A Manhattan metric \( d \) over \( \mathbb{R}^n \) is defined by [29]

\[
d(\mathbf{e}, \mathbf{f}) := \sum_{i=1}^{n} |\mathbf{e}_i - \mathbf{f}_i| .
\]

It faithfully adds up all the projected distances with respect to every individual dimensions. Cosine similarity between them is defined by

\[
\cos(\mathbf{e}, \mathbf{f}) := \frac{\mathbf{e} \cdot \mathbf{f}}{|\mathbf{e}| \cdot |\mathbf{f}|} .
\]

This is a classical approach measuring the relation or similarity between two vectors.

**Data** The weekly data (up to Week 51, 2020) of the reported COVID-19 total confirmed cases and deaths worldwide [28] are download. The samples are the countries. To reduce biased sampling which derives from the insufficient population, inadequate healthcare support and missing data, only these countries satisfying the following three criteria simultaneously would be qualified for the sampled countries:

1. countries whose population are more than ten millions;
2. countries whose healthcare systems are ranked \([30]\) among the top 100;
3. countries whose data from Week 15 to Week 51, Year 2020 are available.

After filtering out the non-essential samples by the above criteria, we obtain 43 countries as shown in Table 1—each of which contains 37 weekly data (from Week 15 to Week 51) of total confirmed cases and deaths.

**Procedures** The current work is conducted result spectral analysis by the following procedures:

1. Prepare and compile the weekly accumulated confirmed cases and deaths from Week 15 to Week 51 in Year 2020 with respect to the sampled 43 countries. The sampled data are shown in Table 6.

2. Calculate the \( t \)-th week’s growth rates for confirmed cases of COVID-19 by the formula

\[
g_t^i = \frac{Week_t^i + 1 - Week_t^{i-1}}{1 + Week_t^i} ,
\]

where \( Week_t^i \) denotes the total number of confirmed cases at Week \( t \) for country \( i \). For my analytical purpose, the denominator is deliberately added by 1 to avoid divisor being 0. Similarly, we could calculate the \( t \)-th week’s growth rates of deaths for country \( i \) by \( h_t^i \). Then for each country \( i \), a growth vector \( g_t^i = (g_t^i)_{i=15}^{50} \) and \( h_t^i = (h_t^i)_{i=15}^{50} \) are formed. The calculated results are tabulated in Table 7.

3. Choose an orthonormal basis \( \mathbb{B}_{36} \) and rename them by \( \beta_1 \equiv I, \beta_2 \equiv b_1, \beta_3 \equiv b_1, \beta_4 \equiv b_2, \beta_5 \equiv b_2, \ldots, \beta_{35} \equiv \hat{b}_{17}, \beta_{36} \equiv \sqrt{\frac{1}{2}} \cdot b_{18} \).
Table 1  Sampled countries with labels

| No. | 1   | 2   | 3   | 4   |
|-----|-----|-----|-----|-----|
| Country | Algeria | Argentina | Australia | Bangladesh |
| 5   |    |     |     |     |
| Country | Belgium | Benin | Canada | Chile |
| 9   |    |     |     |     |
| Country | Cuba | Czechia | Congo | Egypt |
| 13  |    |     |     |     |
| Country | France | Germany | Greece | Guatemala |
| 17  |    |     |     |     |
| Country | Indonesia | Iran | Italy | Japan |
| 21  |    |     |     |     |
| Country | Jordan | Kazakhstan | Malaysia | Mexico |
| 25  |    |     |     |     |
| Country | Morocco | Netherlands | Philippines | Poland |
| 29  |    |     |     |     |
| Country | Portugal | Romania | Saudi Arabia | Senegal |
| 33  |    |     |     |     |
| Country | South Korea | Spain | Sri Lanka | Sweden |
| 37  |    |     |     |     |
| Country | Thailand | Tunisia | Turkey | Ukraine |
| 41  |    |     |     |     |
| Country | United Kingdom | United States | Venezuela |

These 43 labelled countries are sampled based on three criteria: over 10 million population, top 100 healthcare system, and available data for the set periods.

Fig. 1  Trend of weekly total cases of COVID-19. These plots correspond directly to Table 6. It depicts the weekly total cases of COVID-19 for the labelled countries 1 to 4 and 40 to 43.
3 Results

In correspondence to the procedures described in Sect. 2, we embark on data analysis and produce the results in this section with the help of some R 4.1.0 programs [31]. The raw data for weekly total confirmed cases and deaths are further processed and graphed in Figs. 1 and 6. Due to limited space, we choose labelled countries 1 to 4 and 40 to 43 throughout this study for demonstrative purpose. The visualised growth rates for confirmed cases and deaths are graphed in Fig. 2. As expected, the trends for confirmed cases and deaths are very similar, but different in scale. However, this synchrony falls apart as more features and factors are taken into consideration.

To further reveal the characteristics of the trends, we delve into the growth rates of the total numbers. The visualised results are graphed in Fig. 2.

This visualisation is not sufficient enough to look into the underlying features of the changes. To investigate the hidden features of the trend, we avail of spectral analysis to decompose the national trends into different magnitude of frequencies. The reason that the growth rates are adopted for decomposition rather than the total cases is for the further comparison and analysis between countries, since the growth rate in essence is already scaling the data and that makes the decompositions comparable. The visualised results are graphed in Fig. 3.

Since the data collected might still contain some noise, we process the resulting Fourier coefficients by ranking. An alternative is to compare the coefficients directly by some similarity measures, but this approach might contain some noise and distort the results. The ranks are assigned directly by comparing the values of the coefficients. The higher values will be assigned higher ranks. After this ranking, we start to measure the distances for the ranking vectors of the countries via Manhattan metric, which is a straightforward metric reflecting the difference between two vectors. The calculated results are shown in Table 2.

Clustering analysis Based on Table 2, we analyse the distance matrix by categorising the countries with similar features by two methods for contrasts: minimal pairing and maximal pairing. The minimal one reveals the closer neighbours who share the similar features, whereas the maximal one discloses the further neighbours who bear the most dissimilarities. The results are collectively presented in Fig. 4.

Cosine similarity For further comparison, we contrast the spectral analysis with typical cosine similarity analysis. The targets for cosine similarity are the growth rates of confirmed cases and deaths. In other words, it calculates the similarities between the graphs in Fig. 3. Cosine similarity, which is defined in Equation 2, is explicitly or implicitly used in many fields, since it is a composite index which yields a much intuitive interpretation via geometrical notions. Since it is a static indicator, the analysis per se does not really
reveal the internal difference between the structures or trends. The results are presented in Table 3.

Summary

1. From Figs. 1 and 6, we observe that the trends between confirmed cases and deaths are pretty much the same.
2. From Fig. 2, the trends of growth rates for confirmed cases and deaths are not similar to each other. Furthermore, the growth rates for deaths are much stabler than the ones for the confirmed cases;
3. From Fig. 3, one could observe that the Fourier coefficients for confirmed cases and deaths are in line with each other; there is no obvious leading frequencies that decide the trend of growth rates;
4. From Fig. 4, we observe that the graphs are disconnected from the spectral features’ point of view. The sub-graphs indicate the closeness between their features. Each group shall share some hidden properties which are not necessary linking to the geographical locations;
5. From Fig. 4, we find the maximal clusters do not form a connected graph—this indicates the characteristic of the grouping is distinct and clear-cut, i.e., there are some separable features among the countries;
6. Comparing Fig. 5 with Fig. 4, we find cosine similarities lead to few representing countries than the spectral-featured counterparts; in addition, the confirmed cases have even fewer representing countries than the death—this shall indicate the deaths might be due to the circumstances of each individual countries—for example, the healthcare systems;
7. From Table 4, the closest pair is the minimal pairing for confirmed case via spectral analysis and the maximal pairing for deaths via cosine similarity. This indicates there two methods indeed produce substantially different results (if the two methods are similar to each other, we shall expect a closer relation between $f_{q \_mc}$ and $c_{s \_mc}$ than the one between $f_{q \_mc}$ and $c_{s \_Mc}$).

Prediction Now we show how to further approximate the weekly COVID-19 confirmed cases and deaths based on obtained features. In this section, for any vector $z_i$, we use $z_{i,j}$ to denote its $j$-th element in $z_i$. Let $v_i^t$ and $w_i^t$ denote the total confirmed cases and total deaths by Week $t$ for Country $i$, respectively. Define the growth
Fig. 4 Minimal and maximal pairings for cases and deaths. These figures are based on Table 2. They reveal the distances between ranked Fourier coefficients of the 43 countries. The lower the distance is, the closer the features are.
Fig. 5 Minimal and maximal pairings for cases and deaths by typical cosine values. These figures are based on Table 3. They reveal the similarities between weekly growth rates of COVID-19 for the 43 countries.
rates $g^t_i := \frac{v^{i+1}_t - v^t_i}{v^t_i}$ and $h^t_i := \frac{w^{i+1}_t - w^t_i}{w^t_i}$. For each $n \in \mathbb{N}$, let vector

$$g_i[15 + n : 50 + n] = (g_i[14+n+j])_{j=1}^{36},$$
$$h_i[15 + n : 50 + n] = (h_i[14+n+j])_{j=1}^{36},$$

and let vector

$$p_i[15 + n, 50 + n] = (<g_i[15 + n : 50 + n], \beta_k>)_{k=1}^{36},$$
$$q_i[15 + n, 50 + n] = (<h_i[15 + n : 50 + n], \beta_k>)_{k=1}^{36}.$$

Now we want to estimate the values, which extend the values in Table 6, of $\{v^{51+n}_i\}_{n \in \mathbb{N}}$ and $\{w^{51+n}_i\}_{n \in \mathbb{N}}$ based on the derived features. To preserve the features as much as possible, the decision is made by choosing the $g_i^{51}$ such that $||p_i[16,51] - p_i[15,50]||$, or for computational purpose $||p_i[16,51] - p_i[15,50]||^2$, is minimized. Then by the definition, $v_i^{52} = v_i^{51} \cdot (1 + g_i^{51})$ is computed. Hence the general problem could be formulated by finding the optimal $g_i^{50+n}$ such that

$$||p_i[15 + n : 50 + n] - p_i[15 : 50]||^2 = ||(<g_i[15 + n : 50 + n], \beta_k> - <g_i[15 : 50], \beta_k>)_{k=1}^{36}||^2$$

is minimized. The implementation is described in the following claim:

### Claim 1

$g_i^{50+n} = g_i^{50} - \frac{\alpha^n_i}{36} \sum_{k=1}^{36} (\beta_{k,36})^2$, and $v_i^{50+n+1} = v_i^{50+n} \cdot (1 + g_i^{50+n})$ for any arbitrary $n \in \mathbb{N}$.

where $\alpha_i^n = \sum_{k=1}^{36} \beta_{k,36} \cdot (\Delta^n(g_i) \cdot \beta^{(35)}_i)$; where $\Delta^n(g_i) = (g_i[14+n+j] - g_i[15+n+35])_{j=1}^{36}$; and where $\beta^{(35)}_k = [\beta_{k,1}, \beta_{k,2}, \ldots, \beta_{k,35}]$

### Proof

Define $\psi^n_i(g_i^{50+n}) = ||p_i[15 + n : 50 + n] - p_i[15 : 50]||^2$. Then

$$\psi^n_i(g_i^{50+n}) = \sum_{k=1}^{36} \sum_{j=1}^{36} (g_i^{14+n+j} \cdot \beta_{k,j} - g_i^{14+n} \cdot \beta_{k,j})^2$$

$$= \sum_{k=1}^{36} \sum_{j=1}^{36} (g_i^{50+n} \cdot \beta_{k,36} - g_i^{50} \cdot \beta_{k,36} - \sum_{j=1}^{35} (g_i^{14+n+j} - g_i^{14+n}) \cdot \beta_{k,j})^2.$$
and second order derivative $2 \cdot \sum_{k=1}^{36} (\beta_{k,36})^2 > 0$ the result follows immediately. \hfill $\square$

Corollary 1 $h_i^{50+n} = h_i^{50} - \frac{\gamma^n_{i}}{\sum_{k=1}^{36} (\beta_{k,36})^2}$, and $w_i^{50+n+1}$

$= w_i^{50+n} \cdot (1 + h_i^{50+n})$ for any arbitrary $n \in \mathbb{N}$,

where $\gamma^n_i = \sum_{k=1}^{36} \beta_{k,36} \cdot (\Delta^n(h_i) \cdot \beta_{k,36}^{(35)})$; where

$\Delta^n(h_i) = [h_i^{14+n+35} - h_i^{14+n+35}]_{j=1}^{35}$.

Proof The result could be deduced by the same inference above. \hfill $\square$

Based on this alternating inductive steps, we predict next four weeks’ confirmed cases and deaths for the 43
Table 5 Prediction of confirmed cases and deaths

| Country | Week | 52   | 53   | 54   | 55   |
|---------|------|------|------|------|------|
| 1       | 2498 | 2012 | 1621 | 1305 |
| 2       | 68343| 95408| 133193| 185942|
| 3       | 419  | 1050 | 2633 | 6602 |
| 4       | 7979 | 6254 | 4902 | 3842 |
| ...     | ...  | ...  | ...  | ...  |
| 40      | 56615| 45577| 36691| 29537|
| 41      | 288387| 436014| 659211| 996665|
| 42      | 1681616| 1780656| 1885529| 1996578|
| 43      | 3071 | 3254 | 3447 | 3653 |

This table shows the prediction of weekly accumulated confirmed cases (upper block) and deaths (lower block) from Week 52 to Week 55 for the sampled 43 countries. The prediction is based on the spectral analysis.

Based on these characteristics, there are a couple of points I like to address:

- Relatively speaking, the main advantage for statistical approach, in particular regression-related methods, over spectral analysis lies in its interpretability of the causal variables. But this might also be a disadvantage since one needs to specify the independent variables, which is not a case for spectral analysis. Henceforth, for automatic reasoning, the spectral analysis shall turn out to be more effective, but for meaningful interpretation, statistical approach is preferable. The choice between them would depend on the one’s purpose.

As for the study, there are some points worth noticing and enhancing.

- Some of the results about causal relations in this study might not comply with other researches [25]. This is reasonable, since the approach we adopt focus more on feature detection, not solely on causal relation finding.

4 Discussion and conclusion

The main purpose for this study is to extract the patterns of evolution of COVID-19 regarding confirmed cases and deaths across the globe and to predict the future trend via spectral analysis. The results are presented in Sect. 3. The characteristics of this approach in this study go as follows:

- Spectral analysis is applied on detecting and tracking the hidden features of growth rates of confirmed cases and deaths of COVID-19. This method offers some advantages over statistical approaches: there is no predetermined independent variable association and no need to consider or interpret the interactions between chosen factors. This characteristic provides a much efficient way for automatic reasoning, though the features extracted are somehow mechanic.

- By Manhattan metric and ranking techniques, we could then perform spectral clustering which groups the countries with similar features. Since this is a feature-based clustering, the groups could be easily identified by their representing frequencies. Unlike the classical cosine similarity analysis which is a composite and descriptive index and hard to identify their representing properties of the clusters, spectral analysis delves into the internal structures or dynamical properties of the trend.

- Spectral analysis could also be applied in approximation problem and this is also conducted in the prediction of confirmed cases and deaths of COVID-19 in this study.

- Based on these characteristics, there are a couple of points I like to address:

- When one is interested in finding out the fundamental features, or the benchmarks, of time series, then spectral analysis is the candidate, i.e., the internal structures of data are revealed via the magnitudes of the frequencies; cosine similarity, on the other hand, is a composite and static indicator for relation between vectors—it is much intuitive in geometrical interpretation, but ambiguous in revealing or comparing the internal structures.

- As for the study, there are some points worth noticing and enhancing.

- Some of the results about causal relations in this study might not comply with other researches [25]. This is reasonable, since the approach we adopt focus more on feature detection, not solely on causal relation finding.

- One could also delve into the shift of phrases of the frequencies by lifting the constraint on weekly growth rates. This might yield an even more dynamical pictures of the evolutions.

- The samples filtered are based on some criteria. One could loosen or strengthen the criteria to compare the results generated.

- During the reviewing process of this manuscript, there is a new variant Omicron [32,33] whose dynamical behaviour is worth further investigating [34].
Acknowledgements This work is supported by the Humanities and Social Science Research Planning Fund Project under the Ministry of Education of China (No. 20XJA-GAT001).

Data Availability Statement All data used in this article are included in the manuscript.

Appendix

Refined raw data

See Table 6.

Table 6 Raw weekly total cases and total deaths of COVID-19

| Country | Week | 15 | 16 | 17 | ... | 49 | 50 | 51 |
|---------|------|----|----|----|------|----|----|----|
| 1       | 594  | 715 | 753 | ... | 6031 | 3850 | 3101 |
| 2       | 649  | 727 | 837 | ... | 44299 | 35067 | 48955 |
| 3       | 578  | 290 | 101 | ... | 72   | 66   | 167 |
| 4       | 533  | 1835| 2960| ... | 15138 | 12988 | 10180 |
| ...     | ...  | ... | ... | ... | ...   | ...   | ... |
| 40      | 1469 | 2672| 3168| ... | 90627 | 87360 | 70327 |
| 41      | 32205| 32027| 32830| ... | 105915 | 126161 | 190744 |
| 42      | 219936| 202116| 206223| ... | 1373677| 1499756| 1588085 |
| 43      | 33   | 75  | 69  | ... | 2402 | 2735 | 2898 |

This is a partial raw data for weekly confirmed cases (upper block) and deaths (lower block) of COVID-19 for 15th week 2020 to 51st week 2020 for 43 sampled countries. The sampling is based on the size of population in a country, the availability of COVID-19 data and the healthcare system in a country.

Deaths from Week 15 to Week 51

See Fig. 6.

Fig. 6 Trend of weekly total deaths of COVID1-19. These plots correspond directly to Table 6.
Weekly growth rates for cases and deaths

See Table 7.

Table 7 Weekly growth rates of cases and deaths of COVID-19

| Country | Week | 15  | 16  | 17  | ... | 48  | 49  | 50  |
|---------|------|-----|-----|-----|-----|-----|-----|-----|
| 1       | 0.2  | 0.05| 0.45| ... | −0.18| −0.36| −0.19|
| 2       | 0.12 | 0.15| 0.2 | ... | −0.09| −0.21| 0.4  |
| 3       | −0.50| −0.65| −0.13| ... | 0.00 | −0.08| 1.51 |
| 4       | 2.44 | 0.61| 0.36| ... | 0.00 | −0.14| −0.22 |
| ...     |     |     |     |     |     |     |     |     |
| 40      | 0.82 | 0.19| 0.04| ... | −0.18| −0.04| −0.19|
| 41      | −0.01| 0.03| −0.05| ... | 0.01 | 0.19 | 0.51 |
| 42      | −0.08| 0.02| −0.07| ... | 0.21 | 0.09 | 0.06 |
| 43      | 1.24 | −0.08| −0.53| ... | 0.09 | 0.14 | 0.06 |

This table partially presents the results of the growth rates of weekly cases and deaths from Week 15 to Week 50 for 43 sampled countries. Each $i$ row vector $g_i$ is $(g_i^t)_{t=15}^{50}$ whose elements are defined in Equation 3.

Fourier coefficients

See Table 8.

Table 8 Fourier coefficients, or inner products, for 36 frequencies (or features) with respect to weekly growth rates of cases (upper block) and deaths (lower block)

| Country | Feature | $I$  | $b_1$ | $b_2$ | ... | $b_{17}$ | $b_{17}$ | $b_{18}$ |
|---------|---------|-----|-------|-------|-----|----------|----------|----------|
| 1       |         | 0.2 | 0.2   | 0.2   | ... | 0.03     | −0.2     | 0.35     |
| 2       |         | 0.82| 0.69  | 0.13  | ... | −0.57    | −0.12    | 0.64     |
| 3       |         | −0.01| 0.51  | −1.46 | ... | −0.7     | 0.09     | 0.86     |
| 4       |         | 0.8  | 0.88  | 0.76  | ... | 0.59     | 0.78     | 0.8      |
| ...     |         |     |       |       |     |          |          |          |
| 40      |         | 0.66| −0.2  | 0.35  | ... | 0.16     | −0.23    | 0.45     |
| 41      |         | −0.15| −0.69 | 0.12  | ... | 0.09     | −0.77    | 0.4      |
| 42      |         | −0.02| 0.01  | 0.04  | ... | 0.14     | −0.32    | 0.34     |
| 43      |         | 1.62| 0.9   | −0.2  | ... | 0.21     | 1.35     | 1.84     |

Ranking the Fourier coefficients

See Table 9.
Table 9  Ranking the Fourier coefficients for cases (upper block) and deaths (lower block) calculated in Table 8

| Country | Feature | $I$ | $b_1$ | $b_1$ | $b_2$ | $b_2$ | $b_3$ | $b_3$ | $b_{15}$ | $b_{16}$ | $b_{16}$ | $b_{17}$ | $b_{17}$ | $b_{18}$ |
|---------|---------|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1       |         | 27  | 25    | 26    | 4     | 30    | 23    |       | 36     | 15     | 2     | 16     | 8     | 33     |
| 2       |         | 36  | 35    | 22    | 20    | 6     | 26    |       | 2      | 5      | 14    | 1      | 16    | 34     |
| 3       |         | 19  | 30    | 2     | 5     | 15    | 34    |       | 32     | 16     | 1     | 8      | 22    | 35     |
| 4       |         | 34  | 36    | 32    | 30    | 29    | 17    |       | 6      | 2      | 3     | 31     | 33    | 35     |
|         |         |     |       |       |       |       |       |       |       |       |       |       |       |       |
| 40      |         | 36  | 2     | 33    | 7     | 26    | 10    |       | 23     | 17     | 18    | 21     | 1     | 35     |
| 41      |         | 10  | 2     | 24    | 28    | 4     | 34    |       | 12     | 16     | 21    | 36     | 21    | 35     |
| 42      |         | 20  | 23    | 26    | 2     | 7     | 14    |       | 36     | 17     | 1     | 30     | 3     | 35     |
| 43      |         | 35  | 30    | 9     | 6     | 4     | 5     |       | 2      | 8      | 11    | 20     | 33    | 36     |

The higher the coefficients are, the higher the ranks are. The higher ranks indicate the main features of weekly growth rates of COVID-19 in terms of the chosen 36 frequencies.
### Prediction from Week 52 to Week 54

See Figs. 7 and 8.

#### Prediction from Week 52 to Week 54

| Week 52 | Week 53 | Week 54 |
|---------|---------|---------|
| USA | China | India |
| 1000 | 2000 | 3000 |

#### References

1. Worldometer, COVID-19 PANDEMIC (2021). [https://www.worldometers.info/coronavirus/](https://www.worldometers.info/coronavirus/). Accessed 28 June 2021

2. E. Mahase, Covid-19: What have we learnt about the new variant in the UK? BMJ 2020, 371 (2020). [https://doi.org/10.1136/bmj.m4944](https://doi.org/10.1136/bmj.m4944)

3. A.M. Pollock, Asymptomatic transmission of covid-19. BMJ 2020, 371 (2020). [https://doi.org/10.1136/bmj.m4851](https://doi.org/10.1136/bmj.m4851)
