Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Association between COVID-19 cases and international equity indices
Nick James a,∗, Max Menzies b,1

a School of Mathematics and Statistics, University of Sydney, NSW, Australia
b Yau Mathematical Sciences Center, Tsinghua University, Beijing, China

ABSTRACT
This paper analyzes the impact of COVID-19 on the populations and equity markets of 92 countries. We compare country-by-country equity market dynamics to cumulative COVID-19 case and death counts and new case trajectories. First, we examine the multivariate time series of cumulative cases and deaths, particularly regarding their changing structure over time. We reveal similarities between the case and death time series, and key dates that the structure of the time series changed. Next, we classify new case time series, demonstrate five characteristic classes of trajectories, and quantify discrepancy between them with respect to the behavior of waves of the disease. Finally, we show there is no relationship between countries' equity market performance and their success in managing COVID-19. Each country's equity index has been unresponsive to the domestic or global state of the pandemic. Instead, these indices have been highly uniform, with most movement in March.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction
COVID-19 has had an immense social and economic impact on countries around the world, claiming many lives, necessitating business closures, and sending financial markets into disarray. This paper addresses the following question: on a country-by-country basis, what has had the most impact on a country's stock market — its total cumulative cases, the growth in new daily cases, the return of second waves of the disease, or the worldwide state of the pandemic? The goal of this paper is to study the worldwide spread of COVID-19, analyze the various waves of the disease on a country-by-country basis, and show that the financial markets have been unresponsive to all developments in new or cumulative cases after March.

The pandemic has prompted a substantial amount of attention and research. Epidemiologists have analyzed the spread of COVID-19 and potential measures of containment [1–5], while clinical researchers have explored potential treatments for the disease [7–13]. In finance, many studies have observed the impact of COVID-19 on stock markets [14–16], particularly regarding financial contagion [17,18] and stability [19]. Within the nonlinear dynamics community, a majority of papers on COVID-19 have used new and traditional techniques to analyze and predict the spread of cases and deaths [20–23]. Of these, improvements to SIR models [24–30] and power-law models [31–34] have been the most popular. There is an absence of research that studies financial markets in conjunction with the spread of the virus.

For this goal, we use new and existing time series analysis techniques. Existing methods of time series analysis are diverse, including power-law models [31,35–38], and nonparametric methods such as distance analysis [39], distance correlation [40–42] and network models [43,44]. Time series analysis has been widely applied to both finance [45–50] and epidemiology [51,52], including COVID-19 [23,32,53].

We implement two methods of clustering time series, which have been previously used in various financial [54–56] and epidemiological applications, including inflammatory diseases [57], airborne diseases [58], Alzheimer’s disease [59], Ebola [60], SARS [61], and COVID-19 [53]. The two methods we use are hierarchical clustering [62,63] and the optimal one-dimensional implementation of K-means, Ckmeans.1d.dp [64].

In each of the proceeding three sections, we implement time series analysis and clustering for a different goal. In Section 2, we use a smoothed dynamic implementation of cluster analysis to track the worldwide spread of COVID-19, particularly the change in structure over time. In Section 3, we apply semi-metrics to sets of turning points to classify countries according to the disease’s first, second or third wave behavior. In Section 4, we use a new method to analyze case trajectories and equity markets in conjunction, and show the markets are highly concurrent with each other, not any country’s case counts. Section 5 summarizes our findings regarding the considerably different progression of COVID-19 and equity market trends of 2020.
2. Cumulative COVID-19 case and death spread

In this section, we use a dynamic and smoothed implementation of cluster analysis to study the worldwide spread of COVID-19, track the relationships between countries’ cumulative case and death counts, and detect changes in the structure of the two time series. Our data spans 12/31/2019 to 08/31/2020, a period of \( T = 245 \) days. We restrict attention to countries with more than 10,000 cumulative cases at the end of the data period, leaving \( n = 92 \) countries. We order these countries by alphabetical order and let \( x_{it} \) be the cumulative daily series of cumulative cases and deaths, respectively, for \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \).

2.1. Cluster-based methodology for multivariate time series

Following [23], this analysis proceeds in several steps, which are further explained in Appendix A. First, given the multivariate time series of cases or deaths, we generate a logarithmic distance matrix that captures the natural history of COVID-19. China was the first country to experience a severe number of cases, and was the unique country in the worst-affected cluster until late March. Then, Italy, Spain and the United States (US) join the worst-affected cluster, struggling to contain their case counts. From the beginning of April until the end of May, the US was the unique member of the worst-affected cluster, signifying how exceptionally it was impacted by COVID-19 cases relative to every other country. Brazil joins the worst-affected cluster at the start of June, and India at the start of August. By contrast, the United Kingdom (UK), Italy, Spain and Germany move to less affected clusters from the beginning of April, likely a result of strict lockdown procedures implemented in these countries.

Fig. 1(a) tracks the changing cluster membership of 15 select countries with respect to their case counts from February onwards, and captures the natural history of COVID-19. China was the first country to experience a severe number of cases, and was the unique country in the worst-affected cluster until late March. Then, Italy, Spain and the United States (US) join the worst-affected cluster, struggling to contain their case counts. From the beginning of April until the end of May, the US was the unique member of the worst-affected cluster, signifying how exceptionally it was impacted by COVID-19 cases relative to every other country. Brazil joins the worst-affected cluster at the start of June, and India at the start of August. By contrast, the United Kingdom (UK), Italy, Spain and Germany move to less affected clusters from the beginning of April, likely a result of strict lockdown procedures implemented in these countries.

Finally, we can use the constructions so far to compare the case and death time series in conjunction. Turning first to the number of clusters, let \( \hat{k}_c(t), \hat{k}_d(t) \) be the smoothed number for cases and deaths, respectively. Noticing a similarity between these functions, we can compute the most appropriate offset between them. Given a function \( f(t) \) and \( \delta > 0 \), we write \( f(t) = f(t + \delta) \). An appropriate offset can be computed by minimizing the \( L^1 \) norm between functions,

\[
\| \hat{k}_c(t) - \hat{k}_d(t) \|_{L^1} = \int |\hat{k}_c(t + \delta) - \hat{k}_d(t)| dt.
\]

2.2. Resultsof cluster-based analysis of cases and deaths

We now summarize the results of the three figures. Fig. 1(a) tracks the changing cluster membership of 15 select countries with respect to their case counts from February onwards, and captures the natural history of COVID-19. China was the first country to experience a severe number of cases, and was the unique country in the worst-affected cluster until late March. Then, Italy, Spain and the United States (US) join the worst-affected cluster, struggling to contain their case counts. From the beginning of April until the end of May, the US was the unique member of the worst-affected cluster, signifying how exceptionally it was impacted by COVID-19 cases relative to every other country. Brazil joins the worst-affected cluster at the start of June, and India at the start of August. By contrast, the United Kingdom (UK), Italy, Spain and Germany move to less affected clusters from the beginning of April, likely a result of strict lockdown procedures implemented in these countries.

Fig. 1(b) tracks the changing cluster memberships according to deaths for the same countries. Until mid-March, China was the only member of the worst-affected cluster. From then until mid-May, the US, UK, Spain, and Italy belong to the worst-affected cluster. Subsequently, the UK, Spain and Italy leave this cluster. As with cases, the US was the unique member of the worst cluster with respect to deaths for over a month, with Brazil joining at the end of June, and India joining just before the end of August. Given the similarity in case and death cluster behaviors, anomalous countries can be identified if they belong to a significantly different...
Fig. 2. (a) Smoothed number of clusters $\hat{k}(t)$ as a function of time for both cases and deaths. Similarity is observed up to an offset computed as $\delta = 27$. (b) Normalized difference between affinity matrices, with an optimal offset of $\tau = 15$.

Fig. 3. Hierarchical clustering on the distance $d(s, t)$ between adjacency matrices $\text{Adj}^{(t)}$ at different times $t$, for (a) cases and (b) deaths. Each cluster is an unbroken interval of dates. The three boundary dates 05/30, 07/02, 08/08 for cases and 06/13, 07/21, 08/18 for deaths herald significant changes in the structure of the multivariate time series on these dates.
case and death cluster at the same time. The most anomalous country is Singapore. On 08/31/2020, Singapore belonged to the fifth case cluster, but the ninth and least severe death cluster. Indeed, on this day, Singapore had 56771 cases and only 27 deaths, a lower death rate than any other country under consideration. In general, these heat maps have two advantages over other methodologies: first, they are an effective tool for visually comparing the severity of one country’s counts relative to the rest of the world. Next, they can highlight countries that are in significantly different clusters with respect to cases and deaths, such as Singapore.

Fig. 2(a) tracks the changing number of clusters for both cases and deaths. During February and March, the number of clusters rises substantially as the virus spreads to different countries at different rates. Subsequently, cluster numbers stabilize in April, May and June; then begin to decline as cumulative counts around the world begin to exhibit more homogeneity. Toward the end of the period, the greater number of case clusters than death clusters reflects the greater heterogeneity in death rates than of the period, the greater number of case clusters than death clusters. Subsequently, cluster numbers stabilize in April, with countries identified as contiguous intervals of dates: 05/01–05/29, 05/30–07/01, 07/02–08/07, and 08/08–31 for cases; 05/16–06/12, 06/13–07/20, 07/21–08/17, and 08/18–08/31 for deaths. Thus, Fig. 3(a) reveals marked transitions in the cluster structure on three dates: 03/30, 07/02, and 08/08, while Fig. 3(b) reveals transitions on 06/13, 07/21, and 08/18. The fact that all clusters are uncovered intervals means the transition dates between the adjacency matrices characterize significant changes in the cluster structure of the respective time series. In Fig. 3, we plot only the intervals of dates rather than all 123 individual dates, for sake of readability of the labels. This methodology reveals significant changes in a collection of time series without the use of change point detection algorithms [66]. Change point algorithms often require highly tuned parameters for good results, are usually applied in a univariate setting, and typically require the assumption of independence [67].

3. New case trajectories and wave behavior

In this section, we study the trajectories of new cases. Again, we restrict to the n = 92 countries with more than 10,000 total recorded cases as of 08/31/2020. Our goal is to algorithmically identify turning points in new case counts on a country-by-country basis and therefore determine which countries are in their first, second, or later waves of the disease. We also apply a measure of discrepancy between sets of turning points to compare this behavior between countries.

Following [68], we proceed in several steps, which are further detailed in Appendix B. Let \( z(t) \in \mathbb{R} \) be the time series of new daily cases, with countries ordered alphabetically. First, we apply a Savitzky–Golay filter to produce a collection of smoothed time series \( z_i(t) \), \( t = 1, \ldots, T \) and \( i = 1, \ldots, n \). Next, we apply a two-step algorithm where we select and then refine a set of turning points. We assign to each smoothed time series a non-empty set \( P_i \) and \( T_i \) of local maxima and local minima (troughs). Every sequence of turning points begins with a trough at \( t = 1 \), where there are zero cases, and alternates between trough and peak. These sequences determine if a given country is in a first or second wave of COVID-19. An assigned sequence of TP indicates the country is in its first wave, with case counts that have never materially decreased; a sequence of TPTPP indicates a country is in its second wave. A sequence of TPT indicates a country experienced one wave followed by a period of significant decline. If this trough occurs as the last day of the period, the cases are still in decline; if it occurs before the last day, the wave has reached a local minimum and is completely over.

Finally, we measure distance between two sets of turning points using the semi-metrics between finite sets proposed in [69]. Given two non-empty finite sets \( A, B \), this is defined as

\[
D(A, B) = \frac{1}{2} \left( \sum_{a \in A} d(b, A) - \sum_{a \in B} d(a, B) \right).
\]

The semi-metric \( D(A, B) \) is symmetric, non-negative, and zero if and only if \( A = B \). Then, we define the \( n \times n \) turning point distance matrix \( D^{TP} \) by

\[
D_{ij}^{TP} = D(P_i, P_j) + D(T_i, T_j).
\]

Our algorithmic approach classifies the 92 countries under consideration into five characteristic classes. 15 countries, including Brazil, India and Argentina, displayed in Figs. 4(a), 4(b), 4(c), respectively, are assigned the sequence TP and determined to be in their first wave of the disease. 31 countries, including China 4(d), Sweden 4(e) and Russia 4(f) are assigned the sequence TPTP, indicating these countries experienced one wave of the disease—their counts have either reached a local minimum or are still in decline. 28 countries, including Spain 4(g), Italy 4(h), the UK 4(i) and Germany 4(j) are assigned TPTP and determined to be in the midst of their second wave. 14 countries, including the United States 4(k) and Singapore 4(l) are assigned TPTPT, indicating an ongoing or completed decline from a second wave. Notably, the United States was in a rapid decline in new cases as of 08/31/2020, but still a substantial number of ~40,000, while Singapore’s cases declined to nearly zero. Finally, 4 countries, that is, Portugal, Greece, Croatia and South Korea are determined to be in their third wave. Of the 92 countries analyzed, 24 exhibited their greatest case counts up to smoothing on the final day of the period. Of the 46 countries that experienced a second (or third) wave, their final wave was more severe for 28 of them. A complete classification of the 92 countries is included in Appendix C.

Fig. 5 displays hierarchical clustering on the 92 \( \times \) 92 matrix \( D^{TP} \). China is an outlier due to turning points that occurred much earlier than any other country, and relatively little activity in the disease after March. Excluding China, four primary clusters are revealed in this dendrogram, corresponding to differing behaviors of waves of the disease. The semi-metric in Eq. (5) prioritizes low minimal distances between sets, rather than the number of elements. Thus, the four countries in their third wave are assigned to the same cluster as the second wave countries due to low minimal distances between their turning point sets. Compared to other methodologies, the matrix \( D^{TP} \) is capable of quickly classifying countries into first, second or higher wave behaviors. This paper is the first time it has been used to examine different countries’ COVID-19 counts.
4. Equity market dynamics

In this section, we study the dynamics of 17 countries’ equity indices with respect to both pricing and 30-day rolling volatility. The data spans 01/01/2020 to 08/31/2020, a period of $T_1 = 175$ trading days. The countries analyzed are: Argentina, Australia, Brazil, Canada, China, France, Germany, India, Italy, Japan, Mexico, Russia, South Korea, Spain, Switzerland, the UK and the US. This list contains the top 15 economies in the world by nominal GDP [70], and at least one country from each of the five characteristic classes of behavior identified in Section 3.

Let $p_i(t)$ be the multivariate time series of each country’s daily closing equity prices, for $t = 1, \ldots, T_1$ and $i = 1, \ldots, 17$. Let $\sigma_i(t)$ be the 30-day rolling volatility, $t = 1, \ldots, T_1 - 30$. For each $t$, this is defined as the standard deviation of the previous 30 days of index data, normalized by $\sqrt{250}$, the number of trading days in a year. This measure of volatility is one of several that we could potentially consider. For example, [71] study the impact of COVID-19 on cryptocurrencies with a specific analysis of the max–min and absolute return volatility time series. We plot all 17 countries’ equity prices and rolling volatility in Fig. 6, respectively, with equity prices normalized to 1 as of the start of the year. Every index experiences a significant drop and a highly volatile period in March. At the end of the period, China’s index has risen the most relative to its value at the beginning of 2020.

Qualitatively, we make two striking observations: first, market dynamics have been highly uniform among the 17 countries, with China as the only exception. Secondly, market movement after March has been largely unaffected by the natural history of COVID-19 described in Sections 2 and 3, such as Brazil and India entering the worst-affected clusters for cases and deaths by the end of August, or the United States experiencing a large second wave in July, or many other developments.

We proceed to quantify and further elucidate both of these observations. First, we analyze all trajectories of equity prices and rolling volatilities in conjunction. Considering equity prices of a
Fig. 5. Turning point distance matrix $D_{TP}$, defined in Section 3, measures distance between sets of turning points in new case trajectories. Excluding the outlier China, four primary clusters of time series are identified with the following behaviors: 15 countries in their first wave, 31 countries declining from their first wave, 14 countries declining from their second wave, and a final cluster containing 28 countries currently in their second wave and 4 in their third.

Fig. 6. Equity market dynamics for 17 countries with respect to (a) adjusted closing equity prices, normalized to begin the year at 1, and (b) rolling volatility for the prior 30 days.

single country gives a function $p_i \in \mathbb{R}^T_1$. Let $\|p_i\| = \sum_{t=1}^{T_1} |p_i(t)|$ be its $L^1$ norm. We can define a normalized index price trajectory by $g_i = \frac{p_i}{\|p_i\|}$. Analogously, we define $\|\sigma_i\| = \sum_{t=1}^{T_1-30} |\sigma_i(t)|$ and the normalized volatility trajectory by $\sigma_i = \frac{\sigma_i}{\|\sigma_i\|}$. These vectors highlight the relative changes of price or volatility within the entire period. We then define two trajectory distance matrices, $D^p_{ij} = \|g_i - g_j\|$ and $D^\sigma_{ij} = \|\sigma_i - \sigma_j\|$. We analyze these distance matrices $D^p$ and $D^\sigma$, which are symmetric, real matrices with trace 0. As such, they can be diagonalized with real eigenvalues. To determine self-similarity within these indices with respect to prices and rolling volatility, we plot the absolute values of the eigenvalues $|\lambda_1| \leq \cdots \leq |\lambda_{17}|$ for each respective matrix in Fig. 7. Inspecting the collective similarity of Fig. 6, we expect that a large number $K$ of the 17 countries are highly similar with respect to equity price and rolling volatility, with a small number of outliers. Indeed, one would expect many volatility trajectories to behave similarly due to structural financial market factors such as volatility clustering. We examine the eigenvalues to estimate this $K$. If there is a large collection of highly similar elements in an $n \times n$ distance matrix $D$, the matrix would have the form

\[
\begin{pmatrix}
  c_1 & c_2 & c_3 & \cdots & c_K \\
  r_1 & 0 & * & \cdots & * \\
  r_2 & * & 0 & \cdots & * \\
  r_3 & * & * & 0 & \cdots & * \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  r_K & * & * & * & 0 \\
\end{pmatrix}
\]

where rows $r_1, \ldots, r_K$ are highly similar to one another and elements * are close to zero. Such a matrix is a small deformation away from a rank $n - (K - 1)$ matrix, and so $K - 1$ of the eigenvalues should be close to 0.

Thus, given an appropriate threshold $\epsilon$, if $|\lambda_1| \leq \cdots \leq |\lambda_{K-1}| \leq \epsilon$, then we can deduce $K$ indices are similar with respect to price or volatility. This is a concise measure of the number of
indices that are similar within the collection studied. In Fig. 7(a), we can set $\epsilon = 0.2$ and observe that 15 eigenvalues are less than this threshold, suggesting high similarity among 16 of the index prices, with China as the clear outlier. In Fig. 7(b), 14 eigenvalues are under a threshold of $\epsilon = 0.3$, suggesting broad similarity among 15 of the indices with respect to volatility. Returning to Fig. 6, this broad similarity is even more striking when examining the significant changes over time in market behavior.

Further, the eigenvalue analysis provides a measure of the scale of the matrix. Since the distance matrices $D$ are symmetric, they can be conjugated by an orthogonal matrix to yield a diagonal matrix of eigenvalues. As a consequence, the operator norm [72] of $D$ coincides with the largest eigenvalue $|\lambda_n|$. That is,

$$\max_{x \in \mathbb{R}^{17} - \{0\}} \|Dx\|_2 = \|D\|_{op} = |\lambda_n|. \quad (7)$$

We can see from Fig. 7 that the operator norm for $D^{vol}$ is approximately four times that of $D^p$. Both distance trajectory matrices are normalized, so a direct comparison is appropriate. Similarly, when comparing $L^2$ matrix norms, $\|D^{vol}\|_2 = 4.14$, while $\|D^p\|_2 = 1.09$. That is, there is a higher degree of collective similarity among indices with respect to price trajectories than volatility. This is a surprising result, given the expected similarity in collective volatility behavior due to volatility clustering. This result may differ if we were to compare price and volatility trajectories of assets in different financial sectors.

Finally, we quantify the extent to which large changes in index price coincide with high case counts. For this purpose, we take all 17 equity price time series $p_i(t)$ and the corresponding countries’ new case time series $z_i(t) \in \mathbb{R}_{>0}$. Large values of $z_i(t)$ mean that the disease is spreading rapidly in that country. On the other hand, large values of the absolute value log returns $|R_i(t)| = \log \left( \frac{p_i(t)}{p_i(t-1)} \right)$ indicate significant changes in the value of the market. Since equity data is only applicable on weekdays, we restrict the new case time series to the weekdays to yield a time series $w_i(t), t = 1, \ldots, T_i$. As new cases are lower on the weekends, this provides a good representation of the trajectory of new cases in each country.

We define a symmetric $34 \times 34$ matrix $M$ that compares the concurrence of these changes. The entries of $M$ are normalized inner products between time series to measure the extent of overlap between market movement and new cases. We define

$$\|R_i\|_2 = \left( \sum_{t=1}^{T_i} |R_i(t)|^2 \right)^{1/2}, \quad (8)$$

$$\|w_i\|_2 = \left( \sum_{t=1}^{T_i} w_i(t)^2 \right)^{1/2}, \quad (9)$$

$$\langle R_i, w_j \rangle_n = \frac{1}{\|R_i\|_2 \|w_j\|_2} \sum_{t=1}^{T_i} R_i(t)w_j(t). \quad (10)$$

The pairing $\langle \cdot, \cdot \rangle_n$ is a normalized inner product that measures the concurrence of large changes in the time series more accurately than the correlation between price and new case time series. The matrix $M$ is defined as follows:

$$M_{ij} = \begin{cases} \langle |R_i(t)|, |R_j(t)| \rangle_n & \text{if } 1 \leq i, j \leq 17, \\ \langle w_{i-17}(t), w_{j-17}(t) \rangle_n & \text{if } 18 \leq i, j \leq 34, \\ \langle |R_i(t)|, |R_j(t)| \rangle_n & \text{if } 1 \leq i \leq 17, 18 \leq j \leq 34. \end{cases} \quad (11)$$

As all the sequences $|R_i(t)|, w_j(t)$ are non-negative, all entries of $M$ are non-negative. In general, given non-negative functions $f, g$, $\langle f, g \rangle_n = 1$ if and only if $f = ag$ for some $\alpha > 0$, while $\langle f, g \rangle_n = 0$ if and only if the non-zero values of $f$ and $g$ are disjoint sets.

In Fig. 8 we perform hierarchical clustering on the matrix $M$ and reveal several insights. First, China is highly anomalous with respect to both case counts and its index. Indeed, China recorded a large number of cases only during January and February, with few cases since and no subsequent wave. Second, China is also relatively anomalous with respect to its index. We can see two particular periods in Fig. 6 where China did not undergo similar large changes as other countries. In March, China’s index experienced a less severe drawdown than every other country; in July, China experienced a period of significant positive growth, unlike any other country.

Third, the dendrogram reveals a high level of similarity among equity indices, excluding China’s, visible in the clear subcluster in the center of the dendrogram. These 16 equity indices form a submatrix in which the mean of all the entries is 0.78, indicating high concurrence of large price movements. Turning to the remaining indices, the dendrogram reveals more heterogeneity, yet some similarity, between the same 16 countries’ case counts. While the 16 equity indices form one prominent subcluster, the same countries’ case counts split into two subclusters. The normalized inner product produces high association between countries whose peaks in cases occurred at similar times. Indeed, the first cluster generally experienced much earlier peaks, as can be seen for Italy 4(b) and Germany 4(j), while the second cluster experienced large case counts much later, such as Brazil 4(a), India 4(b) and Argentina 4(c). Even within the two subclusters, there is less similarity between case counts than there is for indices. This is reflected in the tree of Fig. 8, where branches belonging to countries’ equity indices are split much lower in the tree’s structure.
Most significantly, the figure reveals that there is no concurrence at all between large changes in countries’ case counts and their equity indices. Excluding China’s index, all other equity indices have moved together closely — even China itself exhibited some similarity with other indices.

The insights provided by Fig. 8 are enhanced by the specific methodology presented above. For example, we trialed the commonly used Pearson correlation and distance correlation \[40\] measures between time series. The correlation does not perform as well due to the considerable flatness of the China cases time series, and does not identify it as an outlier. In addition, using correlation does not split the 16 countries (other than China) into two subclusters as described above. The distance correlation is not suitable in this setting: two time series \(x_t\) and \(y_t\) with any linear relationship, such as \(x_t = ay_t + b\), are determined to have maximal distance correlation 1. This is not desired if \(a\) is negative, meaning that \(x_t\) increases or decreases in the opposite direction of \(y_t\). As such, using distance correlation performs even more poorly, for instance allocating Canada and the UK to their own cluster and several other poor results.

5. Conclusion

In this paper, we analyze the natural history of COVID-19 across the world in conjunction with the stock market activity of 17 countries. Qualitatively and quantitatively we demonstrate that market movements have been highly uniform between these 17 countries, with China as the only exception.

In Section 2, we analyze the structure of the multivariate time series of COVID-19 cases and deaths. Our analysis isolates the US as the unique member of the worst-affected cluster with respect to both cumulative cases and deaths for over a month, reflecting its exceptional impact by COVID-19. Subsequently, Brazil and India join that cluster, as their counts rose rapidly. The dendrograms in Fig. 3 each exhibit four contiguous intervals of dates, allowing us to observe key dates when the structure of the world’s case counts changed substantially. With respect to cases, these dates are 05/30, 07/02, and 08/08. Indeed, all these dates herald significant shifts in the status of the disease around the world. On 05/30, Russia and Brazil enter the worst-affected cluster, on the same day as the latter reported a record number of cases \[73\]. On 07/02, several countries that had been heavily impacted earlier, such as China, South Korea, Singapore, and the Netherlands, enter less-affected clusters. This follows from June, a period of steady decline in Europe \[74\]. On 08/08, both Singapore and the Netherlands move back into more severely affected clusters. Also around this time, India, Brazil, much of Africa and South America experience significantly more cases \[75\], while cases in Europe demonstrate a slower increase.

In Section 3, we identify five characteristic behaviors of new case trajectories between countries. 24 countries exhibit their greatest counts up to smoothing on the final day of the period. 46 countries experience a second wave, with 28 experiencing a more severe second wave than the first. Singapore and Australia responded quickly to the virus \[76\], and South Korea was hailed for its early contact tracing success \[77\], yet all three of these countries experience second waves, and South Korea exhibits its greatest case counts at the end of the period. Italy and Spain were acknowledged to have imposed lockdowns too late in March \[78\], with case counts eventually declining in May. Nonetheless, both of these countries experience second waves, with Spain’s more severe than its first. Overall, long first waves and the return of second waves contribute to high case counts toward the end of the data period.

Despite the substantial activity in COVID-19 cases after March, the heterogeneity of subsequent waves and the number of countries with peaks in new cases, no discernible impact on financial indices was observed from March. In Section 4, we apply a new method to analyze collective equity market dynamics across 17 countries in conjunction with their new case counts. Eigenvalue analysis indicates high similarity between 16 countries’ equity prices, with China as the only outlier. We introduce an inner product pairing that demonstrates little concurrence between the profound market movements observed in March and development in COVID-19 cases.

Overall, we have chronicled the natural history of COVID–19 together with the market movements during 2020. Despite substantial heterogeneity in the new case trajectories on a country-by-country basis and frequent changes in the order and structure of most affected countries in cumulative cases, we have observed high homogeneity in the markets. All have moved together with substantial drawdown in March, followed by steady recovery, and no qualitative or quantitative relationship to any developments in COVID–19.

Our methods and results could be used by both financial practitioners and policymakers. The identification of a single country’s index behaving differently from the rest of the collection, in this case China’s, could provide opportunities for diversification of
holdings and portfolio gains. More broadly, studying the sensitivity of financial securities to external factors is fundamental to market analysts and strategists. The finding that the equity markets have been highly uniform, despite considerably differing trends in COVID-19 cases and deaths, may encourage policymakers to be more aggressive in pursuing containment measures. That is, the finding that the equity indices of countries that responded more aggressively to the pandemic have not been negatively impacted could provide confidence in prioritizing public health with reduced concern about harming their equity index.

Diverse future research possibilities exist with these methods. Market dynamics could be analyzed across different sectors, rather than countries. It is conceivable that during the pandemic, equity movements may be highly homogeneous within certain sectors such as technology, but heterogeneous when comparing retail and health care stocks. One could analyze fixed income assets such as government and corporate bonds, and determine if there is a bifurcation between equities and fixed income assets during market crises. Alternative asset classes such as cryptocurrencies could be studied, building on the work of [79], who analyze the co-movement between COVID-19 and Bitcoin. One could also apply our analysis to alternative time series, such as max–min and absolute return volatility [71,80].

In this section, we provide more details for the identification of turning points of a new case time series \( z(t) \). First, some smoothing of the counts is necessary due to irregularities in the data set, and discrepancies between different data sources. There are consistently lower counts on the weekends, and some negative counts due to retroactive adjustments. The Savitzky–Golay filter ameliorates these issues by combining polynomial smoothing with a moving average computation – this moving average eliminates all but a few small negative counts; we simply replace these negative counts with zero. This yields a smoothed time series \( \tilde{z}(t) \in \mathbb{R}_{>0} \). Subsequently, we perform a two-step process to select and then refine a non-empty set \( P \) of local maxima (peaks) and \( T \) of local minima (troughs).

Following [68], we apply a two-step algorithm to the smoothed time series \( \tilde{z}(t) \). The first step produces an alternating sequence of troughs and peaks, beginning with a trough at \( t = 1 \), where there are zero cases. The second step refines this sequence according to chosen conditions and parameters. The most important conditions to identify a peak or trough, respectively, in the first step, are the following:

\[
\tilde{z}(t_0) = \max [\tilde{z}(t) : \max [1, t_0 - l] \leq t \leq \min (t_0 + l, T)], \quad (B.1)
\]

\[
\tilde{z}(t_0) = \min [\tilde{z}(t) : \max [1, t_0 - l] \leq t \leq \min (t_0 + l, T)], \quad (B.2)
\]

where \( l \) is a parameter to be chosen. Following [68], we select \( l = 17 \), which accounts for the 14-day incubation period of the virus [88] and less testing on weekends. Defining peaks and troughs according to this definition alone has several flaws, such as the potential for two consecutive peaks.

Instead, we implement an inductive procedure to choose an alternating sequence of peaks and troughs. Suppose \( t_0 \) is the last determined peak. We search in the period \( t > t_0 \) for the first of two cases: if we find a time \( t_1 > t_0 \) that satisfies (B.2) as well as a non-triviality condition \( \tilde{z}(t_1) < \tilde{z}(t_0) \), we add \( t_1 \) to the set of troughs and proceed from there. If we find a time \( t_1 > t_0 \) that satisfies (B.1) and \( \tilde{z}(t_0) \geq \tilde{z}(t_1) \), we ignore this lower peak as redundant; if we find a time \( t_1 > t_0 \) that satisfies (B.1) and \( \tilde{z}(t_1) > \tilde{z}(t_0) \), we remove the peak \( t_0 \) replace it with \( t_1 \) and continue from \( t_1 \). A similar process applies from a trough at \( t_0 \).

At this point, the time series is assigned an alternating sequence of troughs and peaks. However, some turning points are immaterial and should be excluded. The second step is a flexible approach introduced in [68] for this purpose. In this paper, we

To choose the number of clusters at each \( t \), we average six methods described in [65]. These methods are as follows: Pbisection index [82], Silhouette score [83], KL index [84], C index [85], McClain–Rao index [86] and Dunn index [87], but other methods could be used alternatively. Let the cluster numbers computed by these methods be \( k_1(t), \ldots, k_6(t) \), respectively. We define \( k_\text{av}(t) = \frac{1}{6} \sum_{j=1}^{6} k_j(t) \). This value is not necessarily an integer, so we cannot cluster with it directly. The function \( k_\text{av}(t) \) is approximately locally stationary. So we may apply exponential smoothing to produce a smoothed integer value \( \tilde{k}(t) \). It is with this number that we cluster. Doing so at each \( t \) produces a time-varying partition of the 92 countries into clusters, and defines an adjacency matrix \( A(t) \) for every time \( t \).

Appendix B. Turning point methodology

In this section, we provide more details for the identification of turning points of a new case time series \( z(t) \). First, some smoothing of the counts is necessary due to irregularities in the data set, and discrepancies between different data sources. There are consistently lower counts on the weekends, and some negative counts due to retroactive adjustments. The Savitzky–Golay filter ameliorates these issues by combining polynomial smoothing with a moving average computation – this moving average eliminates all but a few small negative counts; we simply replace these negative counts with zero. This yields a smoothed time series \( \tilde{z}(t) \in \mathbb{R}_{>0} \). Subsequently, we perform a two-step process to select and then refine a non-empty set \( P \) of local maxima (peaks) and \( T \) of local minima (troughs).

Following [68], we apply a two-step algorithm to the smoothed time series \( \tilde{z}(t) \). The first step produces an alternating sequence of troughs and peaks, beginning with a trough at \( t = 1 \), where there are zero cases. The second step refines this sequence according to chosen conditions and parameters. The most important conditions to identify a peak or trough, respectively, in the first step, are the following:

\[
\tilde{z}(t_0) = \max [\tilde{z}(t) : \max [1, t_0 - l] \leq t \leq \min (t_0 + l, T)], \quad (B.1)
\]

\[
\tilde{z}(t_0) = \min [\tilde{z}(t) : \max [1, t_0 - l] \leq t \leq \min (t_0 + l, T)], \quad (B.2)
\]

where \( l \) is a parameter to be chosen. Following [68], we select \( l = 17 \), which accounts for the 14-day incubation period of the virus [88] and less testing on weekends. Defining peaks and troughs according to this definition alone has several flaws, such as the potential for two consecutive peaks.

Instead, we implement an inductive procedure to choose an alternating sequence of peaks and troughs. Suppose \( t_0 \) is the last determined peak. We search in the period \( t > t_0 \) for the first of two cases: if we find a time \( t_1 > t_0 \) that satisfies (B.2) as well as a non-triviality condition \( \tilde{z}(t_1) < \tilde{z}(t_0) \), we add \( t_1 \) to the set of troughs and proceed from there. If we find a time \( t_1 > t_0 \) that satisfies (B.1) and \( \tilde{z}(t_0) \geq \tilde{z}(t_1) \), we ignore this lower peak as redundant; if we find a time \( t_1 > t_0 \) that satisfies (B.1) and \( \tilde{z}(t_1) > \tilde{z}(t_0) \), we remove the peak \( t_0 \) replace it with \( t_1 \) and continue from \( t_1 \). A similar process applies from a trough at \( t_0 \).

At this point, the time series is assigned an alternating sequence of troughs and peaks. However, some turning points are immaterial and should be excluded. The second step is a flexible approach introduced in [68] for this purpose. In this paper, we
introduce new conditions within this framework. First, let $t_m$ be the global maximum of $\tilde{z}(t)$. If this is not unique, we declare $t_m$ to be the first global maximum. This time $t_m$ is always declared a peak during the first step detailed above. Given any other peak $t_1$, we compute the peak ratio $\frac{\tilde{z}(t_1)}{\tilde{z}(t_m)}$. We select a parameter $\delta$, and if $\frac{\tilde{z}(t_1)}{\tilde{z}(t_m)} < \delta$, we remove the peak $t_1$. If two consecutive troughs $t_0$, $t_2$ remain, we remove $t_0$ if $\tilde{z}(t_0) > \tilde{z}(t_2)$, and remove $t_2$ if $\tilde{z}(t_0) \leq \tilde{z}(t_2)$. That is, we ensure the sequence of peaks and troughs remains alternating. In our implementation, we choose $\delta = 0.05$. Unlike [68], we remove earlier peaks, not just subsequent peaks, according to this condition.

Finally, we use the same log-gradient function between times $t_1 < t_2$, defined as

$$\text{log-grad}(t_1, t_2) = \frac{\log \tilde{z}(t_2) - \log \tilde{z}(t_1)}{t_2 - t_1}. \quad (B.3)$$

The numerator equals log(\frac{\tilde{z}(t_2)}{\tilde{z}(t_1)}), a “logarithmic rate of change.” Unlike the standard rate of change given by $\frac{\tilde{z}(t_2) - \tilde{z}(t_1)}{t_2 - t_1}$, the logarithmic change is symmetrically between $(-\infty, \infty)$. Let $t_1$, $t_2$ be adjacent turning points (one a trough, one a peak). We choose a parameter $\epsilon = 0.007$; if

$|\text{log-grad}(t_1, t_2)| < \epsilon, \quad (B.4)$

that is, the average logarithmic change is less than 0.7%, we remove $t_2$ from our sets of peaks and troughs. If $t_2$ is not the final turning point, we also remove $t_1$.

Appendix C. Classification of countries by wave behavior

In Table C.1, we classify all $n = 92$ countries into 5 different characteristic classes according to the methodology of Section 3.

| Country behaviors | First wave (TP) | Over first wave (TPT) | Second wave (TPTP) | Over second wave (TPTPT) | Third wave (TPTPTP) |
|-------------------|-----------------|-----------------------|--------------------|--------------------------|---------------------|
| Argentina         | Afghanistan     | Austria                | Algeria            |                          |                     |
| Brazil            | Armenia         | Belarus                | Australia          |                          |                     |
| Colombia          | Azerbaijan      | Bosnia & Herzegovina   | Belgium            |                          |                     |
| Costa Rica        | Bahrain         | Canada                 | Bulgaria           |                          |                     |
| India             | Bangladesh      | Czech Republic         | Israel             |                          |                     |
| Indonesia         | Bolivia         | Denmark                | Japan              |                          |                     |
| Iraq              | Cameroon        | Ecuador                | Kosovo             |                          |                     |
| Lebanon           | Chile           | Ethiopia               | North Macedonia    |                          |                     |
| Libya             | China           | France                 | Panama             |                          |                     |
| Palestine         | Côte d’Ivoire   | Germany                | Serbia             |                          |                     |
| Paraguay          | Dominican Republic | Honduras            | Singapore          |                          |                     |
| Philippines       | DR Congo        | Iran                   | United States      |                          |                     |
| Senegal           | Egypt           | Ireland                | Uzbekistan         |                          |                     |
| Ukraine           | El Salvador     | Italy                  | Zambia             |                          |                     |
| Venezuela         | Ghana           | Moldova                |                    |                          |                     |
|                   | Guatemala       | Morocco                |                    |                          |                     |
|                   | Kazakhstan      | Nepal                  |                    |                          |                     |
|                   | Kenya           | Netherlands            |                    |                          |                     |
|                   | Kuwait          | Norway                 |                    |                          |                     |
|                   | Kyrgyzstan      | Peru                   |                    |                          |                     |
|                   | Madagascar      | Poland                 |                    |                          |                     |
|                   | Mexico          | Romania                |                    |                          |                     |
|                   | Nigeria         | Spain                  |                    |                          |                     |
|                   | Oman            | Sudan                  |                    |                          |                     |
|                   | Pakistan        | Switzerland            |                    |                          |                     |
|                   | Puerto Rico     | Turkey                 |                    |                          |                     |
|                   | Qatar           | United Arab Emirates   |                    |                          |                     |
|                   | Russia          | United Kingdom         |                    |                          |                     |
|                   | Saudi Arabia    |                        |                    |                          |                     |
|                   | South Africa    |                        |                    |                          |                     |
|                   | Sweden          |                        |                    |                          |                     |

References

[1] G. Wang, Y. Zhang, J. Zhao, J. Zhang, F. Jiang. Mitigate the effects of home confinement on children during the COVID-19 outbreak, Lancet 395 (10228) (2020) 945–947, http://dx.doi.org/10.1016/s0140-6736(20)30547-x.
[2] M. Chinazzi et al., The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, Science 368 (2020) 395–400, http://dx.doi.org/10.1126/science.abax9577.
[3] Y. Liu, A.A. Gayle, A. Wilder-Smith, J. Rocklov, The reproductive number of COVID-19 is higher compared to SARS coronavirus, J. Travel Med. 27 (2) (2020) http://dx.doi.org/10.1093/jtm/ttaa021.
[4] Y. Fang, V. Nie, M. Penny, Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis, J. Med. Virol. 92 (6) (2020) 645–659, http://dx.doi.org/10.1002/jmv.25750.
[5] P. Zhou et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (7798) (2020) 270–273, http://dx.doi.org/10.1038/s41586-020-2012-7.
[6] J. Dehning, J. Zierenberg, F.P. Spitzner, M. Wilbrad, J.P. Neto, M. Wilczek, V. Priesemann, Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions, Science 369 (6500) (2020) eabb9789, http://dx.doi.org/10.1126/science.abb9789.
[7] F. Jiang, L. Deng, L. Zhang, Y. Cai, C.W. Cheung, Z. Xia, Review of the clinical characteristics of coronavirus disease 2019 (COVID-19), J. Gen. Intern. Med. 35 (5) (2020) 1545–1549, http://dx.doi.org/10.1007/s11606-020-05762-w.
[8] Z.Y. Zu, M.D. Jiang, P.P. Xu, W. Chen, Q.Q. Ni, G.M. Lu, L.J. Zhang, Coronavirus disease 2019 (COVID-19): A perspective from China, Radiology (2020) 200490, http://dx.doi.org/10.1148/radiol.2020200490.
[9] G. Li, E.D. Clercq, Therapeutic options for the 2019 novel coronavirus (2019-ncov), Nat. Rev. Drug Discov. 19 (3) (2020) 149–150, http://dx.doi.org/10.1038/s41573-020-00016-0.
[10] L. Zhang, Y. Liu, Potential interventions for novel coronavirus in China: A systematic review, J. Med. Virol. 92 (5) (2020) 479–490, http://dx.doi.org/10.1002/jmv.25707.
[11] M. Wang, R. Cao, L. Zhang, X. Yang, J. Liu, M. Xu, Z. Shi, Z. Hu, W. Zhong, G. Xiao, Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-ncov) in vitro, Cell Res. 30 (3) (2020) 269–271, http://dx.doi.org/10.1038/s41422-020-0282-z.
[12] B. Cao et al., A trial of Lopinavir-Ritonavir in Covid-19, N. Engl. J. Med. 382 (21) (2020) e68, http://dx.doi.org/10.1056/nejmc2008043.
[64] H. Wang, M. Song, Ckmeans.1d.dp: Optimal k-means clustering in one dimension by dynamic programming, R J. 3 (2) (2011) 29–33, http://dx.doi.org/10.32614/rj-2011-015.

[65] P. Radchenko, G. Mukherjee, Convex clustering via ℓ1 fusion penalization, J. R. Stat. Soc. Ser. B Stat. Methodol. 79 (5) (2017) 1527–1546, http://dx.doi.org/10.1111/rssb.12226.

[66] S. Aminikhahemiani, D.J. Cook, A survey of methods for time series change point detection, Knowl. Inf. Syst. 51 (2) (2016) 339–367, http://dx.doi.org/10.1007/s10115-016-0987-1.

[67] M. Lavielle, G. Teyssiére, Detection of multiple change-points in multivariate time series, Lith. Math. J. 46 (3) (2006) 287–306, http://dx.doi.org/10.1007/s10986-006-0028-9.

[68] N. James, M. Menzies, COVID-19 in the United States: Trajectories and second surge behavior, Chaos 30 (2020) 091102, http://dx.doi.org/10.1063/5.0024204.

[69] N. James, M. Menzies, L. Azizi, J. Chan, Novel semi-metrics for multivariate change point analysis and anomaly detection, Physica D 412 (2020) 132636, http://dx.doi.org/10.1016/j.physd.2020.132636.

[70] GDP (current US$), The World Bank, September 21, 2020, 2020, https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?year_high_desc=true.

[71] M.B. Arouxet, A.F. Bariviera, V.E. Pastor, V. Vampa, Covid-19 impact on cryptocurrencies: evidence from a wavelet-based Hurst exponent, 2020, arXiv:2009.05652.

[72] W. Rudin, Functional Analysis, McGraw-Hill Science, 1991.

[73] A. Boadle, Brazil has record new coronavirus cases, surpasses France in deaths, 2020, Reuters, May 31, 2020, https://www.reuters.com/article/us-health-coronavirus-brazil/brazils-coronavirus-outbreak-worsens-as-total-cases-near-500000-idUSKBN2360U8.

[74] S. Neuman, France announces further reopening amid declining number of coronavirus cases, 2020, NPR, June 15, 2020, https://www.npr.org/sections/health-shots/2020/06/15/876953360/france-announces-further-reopening-amid-declining-number-of-coronavirus-cases.

[75] S. Neuman, Global coronavirus cases hit 20 million as pandemic accelerates, 2020, Sydney Morning Herald, August 11, 2020, https://www.smh.com.au/world/europe/who-only-10-per-cent-of-the-way-to-funding-coronavirus-fight-20200811-p55kzf.html.

[76] J. McCurry, Test, trace, contain: how South Korea flattened its coronavirus curve, 2020, The Guardian, April 23, 2020, https://www.theguardian.com/world/2020/apr/23/test-trace-contain-how-south-korea-flattened-its-coronavirus-curve.

[77] A. McCann, N. Popovich, J. Wu, Italy's virus shutdown came too late, What happens now? 2020, The New York Times, April 5, 2020, https://www.nytimes.com/interactive/2020/04/05/world/europe/italy-coronavirus-lockdown-reopen.html.

[78] J.W. Goodell, S. Goutte, Co-movement of COVID-19 and bitcoin: Evidence from wavelet coherence analysis, Finance Res. Lett. (2020) 101625, http://dx.doi.org/10.1016/j.frl.2020.101625.

[79] S. McDonell, Coronavirus: us and australia close borders to Chinese ar- rival, 2020, BBC. https://www.bbc.com/news/world-51338899. (Accessed 1 February 2020).

[80] N. James, M. Menzies, J. Chan, Changes to the extreme and erratic behaviour of cryptocurrencies during COVID-19, Physica A 565 (2021) 125581, http://dx.doi.org/10.1016/j.physa.2020.125581.

[81] Our World in Data, 2020. https://ourworldindata.org/coronavirus-source-data. (Accessed 6 September 2020).

[82] G.W. Milligan, An examination of the effect of six types of error perturbation on fifteen clustering algorithms, Psychometrika 45 (3) (1980) 325–342, http://dx.doi.org/10.1007/bf02293907.

[83] J. McClain, V.R. Rao, CLUSTISZ: A program to test for the quality of clustering of a set of objects, J. Mark. Res. 12 (1975) 456–460.

[84] J. Dunn, Well-separated clusters and optimal fuzzy partitions, J. Cybern. 4 (1) (1974) 95–104, http://dx.doi.org/10.1080/01969772408546059.

[85] S.A. Lauer, et al., The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application, Ann. Intern. Med. 172 (9) (2020) 577–582, http://dx.doi.org/10.7326/m20-0504.