Received: 16 September 2023 | Revised: 19 April 2024 | Accepted: 29 May 2024
DOI: 10.1002/asmb.2879

RESEARCH ARTICLE

WILEY

Comparing risk profiles of international stock markets as functional data: COVID-19 versus the global financial crisis

Ryan Liam Shackleton[1] | **Sonali Das**[2] | **Rangan Gupta**[3]

[1]Department of Information Technology, University of Pretoria, Pretoria, South Africa
[2]Department of Business Management, University of Pretoria, Pretoria, South Africa
[3]Department of Economics, University of Pretoria, Pretoria, South Africa

Correspondence
Sonali Das, Department of Business Management, University of Pretoria, Pretoria, South Africa.
Email: sonali.das@up.ac.za

Abstract
In this article, we aim to provide a detailed econometric analysis of the realized volatility in international stock markets of Brazil, China, Europe, India, the United Kingdom, and the United States, which represent a mix of large developing, and developed markets. For our purpose, we use the functional data analysis (FDA) framework, whence discrete volatility data were first transformed into continuous functions, and thereafter, derivatives of the continuous functions were investigated, and kinetic and potential energy associated is the volatility system were extracted. Results revealed that COVID-19 indeed had a significant effect on international financial market volatility for all the countries, with the exception of China. The realized volatility of the international financial markets did return to their pre-COVID levels in May 2020, and this recovery time was significantly faster than the 2008 financial crisis recovery period. Within the FDA framework, we further investigated the role of uncertainty on the realized volatility, specifically from an outbreak of an infectious disease (such as COVID-19) and a daily newspaper-based infectious disease index as the predictor. The regression analysis showed that the volatility of financial markets can be accurately modeled by this infectious disease index, but only for periods experiencing an epidemic or pandemic.

KEYWORDS
COVID-19, functional data analysis, global financial crisis, infectious disease, international stock markets, realized volatility

1 | INTRODUCTION

In 2019 the novel severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) was identified in Wuhan city, the Capital of Hubei province, China. This highly contagious respiratory disease, commonly referred to as COVID-19, spread quickly and was declared a global pandemic by March 2020, and governments enforced sudden and severe restriction measures on travel, social gatherings, and commercial and economic activities, resulting in unprecedented uncertainty within international financial markets.[1] In the aftermath of the pandemic, it has presented a unique opportunity to gain valuable insights concerning the reaction of international financial markets to a global infectious health crisis, and to evaluate and compare the recovery trajectory across countries.

Volatility can be described as the dispersion of returns for a particular asset, and is often calculated using the standard deviation, or variance, of the specific asset returns, and can be interpreted as the amount of risk or uncertainty associated with a particular asset. Volatility is not the same as risk, but rather a measure that is interpreted as risk or uncertainty.[2] High uncertainty is often synonymous with high-risk assets, and can be used to interpret economic vulnerability, which is critical in the operation of financial markets.[3]

There is a crucial connection between financial market uncertainty and investor confidence, with volatility acting as an indicator of financial risk that can aid governments, individual investors, portfolio fund managers, policymakers, and financial institutions in making their respective decisions.[4] As per the Basel Accord established in 1996, it is compulsory for many financial institutions to forecast volatility, and use it as an input to risk-management.[2,5] While volatility is a key indicator of the vulnerability of financial markets, developing models or forecasts using volatility presents an inherent problem,[5] because by its very nature, volatility is latent and, therefore, is not directly observable.[5,6] Most models using latent volatility do not accurately capture all the facts (such as the declining autocorrelations within the squared returns) when dealing with financial time series.[5] This presents an opportune to utilize the concept of *realized volatility*, which is the sum of squared intra-day returns at specific intervals.[2,7] Typically, the sampling time for intra-day returns is 5 or 10 min (sparse sampling), but it has been reported to be optimal in the range of 30–65 min, with smaller sampling time (higher frequency sampling) resulting in higher accuracy and increased microstructure noise, presenting a useful trade-off.[5,8]

The realized volatility error, denoted by u_n, has the property of $E(u_n|\sigma_n^2) = 0$, where σ_n^2 is the actual volatility. Thus, the estimated realized volatility error, which is a function of the actual volatility, is zero. From the property mentioned above, it is evident that realized volatility is an unbiased indicator for actual volatility[7,9] and provides a consistent estimate of price fluctuations over a specific time.[6] Generalized auto-regressive conditional heteroscedastic (GARCH) models are commonly utilized to model standard time-varying volatility. The conditional variance is a deterministic function of model parameters and historical data. Stochastic volatility (SV) models are also popular among researchers. In the literature, while the GARCH and SV models are often utilized, these models suffer because the underlying volatility estimate is neither model free nor unconditional. The realized volatility has the advantage of being unconditional. Ultimately, by utilizing realized volatility as a measure, the volatility becomes "observable" and can be modeled directly.[5] Nonetheless, the realized volatility is still not free of the ever-present measurement errors,[7] but resolves the latent variable issue and provides investors with an observable metric of uncertainty in the market. We use realized volatility in this research which is based on intra-day data and is rich in information based on large number of observations being recorded daily, in turn, providing an observable and unconditional estimate of the process of volatility.

This research is important for a number of reasons, and ultimately aims to enable individual investors, policymakers, and governments to make better decisions in the future. Specifically, the research objectives in this study include, but are not limited to the following:

- How and to what extent did the COVID-19 pandemic affect the evolution of risk or uncertainty within international financial markets?
- How does this compare to one of the other recent Global Financial Crisis?
- Does the realized volatility differ between the developed and emerging markets during the COVID-19 pandemic?
- Can the daily newspaper-based infectious disease index predict the realized volatility of international financial markets?

A statistical framework is required to analyze realized volatility, and we present a strong argument for using Functional data analysis (FDA) techniques in this endeavor to analyze the role of uncertainty in international financial markets due to the COVID-19 pandemic, as well as to predict the realized volatility using the infectious disease index through FDA techniques.[10] The realized volatility, the sum of squared intra-day returns, has been condensed into a single daily value, and computed over several years, resulting in an extensive data set of discrete points over time, and therefore traditional time series methods would be a logical choice. The standard approach to analyzing stock market data is time series, however, it suffers from many mathematical constraints, with FDA offering an alternative approach to analyzing time series data.

Past literature indicates that FDA has increased in popularity and is being used to better analyze time series data.[11] One of the major drawbacks of classical time series approaches is that data observation intervals need to be equally spaced. The FDA approach is flexible as it does not require an equal time interval between observations.[12] This applies to the realized volatility data set. Although a realized volatility value is computed daily, most major stock markets close trading on weekends and some public holidays. Multiple international financial markets were analyzed and compared. However, the time interval between observations varies due to the time zones and public holidays. Thus, FDA is highly advantageous due to the flexible nature of the approach when it comes to timing intervals between observations. Additionally, the FDA approach considers that the values observed at different times for a single object may be dependent.[10,11] Further, it is common for measurement errors to be present in the data collection process. Wang et al.[13] suggests that any measurement errors that occur throughout the data collection process, can be characterized as random fluctuations around a smooth trajectory. FDA is well suited to combat the issue of measurement errors since repeated measurements are made for a single subject.

The functions that arise from FDA methods are assumed to be differentiable to some order. This is important, as additional information can be extracted from investigating the resulting functions' derivatives (rate of change), which is not easily obtained by using traditional statistical methods.[11] Rate of change analysis, using phase-plane plots, can result in extremely useful insights relating to the financial markets. Also, many traditional statistical tools rely on the property of stationarity being present within time series data, whereby the process of generating a time series does not change over time. However, in most cases, time series data does violate this assumption.[10,14] Financial data, such as the realized volatility of international markets, would require transformations to ensure mean-reversion if traditional statistical methods were applied.[14] Conversely, Muller[12] noted that, when using FDA, the underlying process does not have to be stationary.

Our paper makes multiple contributions. From a methodological standpoint, an extensive investigation into smoothing with B-splines was conducted with the aim to investigate the claim that a fourth-order B-spline with knots at each argument smooths the data with the best accuracy. Second, a significant contribution of this research is in the extraction of the potential and kinetic energy from phase-plane plot analysis. This extraction of energies from phase-plane plots can be applied to other applications as a visualization tool, and thus is not limited to realized volatility as was applied in this work.

The application part of this research involved applying FDA techniques to realized volatility data, and to daily newspaper-based equity market volatility due to the infectious diseases index data set. In the process, this research contains a review of the Global Financial Crisis (GFC) and the COVID-19 pandemic, and their effect on the realized volatility of international financial markets. The review presented here contributes to the growing literature on the GFC, COVID-19 pandemic, and realized volatility. The two data sets used provided an opportunity to demonstrate the FDA techniques mentioned, and to gain insights into the behavior of realized volatility of financial markets, with particular focus on COVID-19. The FDA techniques provided insightful findings and conclusions. In summary, this study contributes to the knowledge of how financial crises, specifically that induced by the COVID-19 pandemic, affected the realized volatility of both emerging and developed financial markets. Lastly, this research contributes to the area of literature that looks at daily newspaper-based equity market volatility due to the infectious diseases index.

The remainder of the article is organized as follows: In Section 2 we provide a detailed description of the data and its associated transformations required for the FDA approach. The Sections 3 and 4 are devoted to the methodology and empirical findings. Finally, Section 5 concludes the paper with a summary of empirical findings and its associated economic implications.

2 | DATA AND TRANSFORMATIONS

There were two different data sets utilized in this research. The first is a data set that contains realized metrics of 31 stock market indices. The second is a data set that contains a daily measure of equity market volatility (EMV) due to infectious diseases. The details of both data sets are provided below. Thereafter details of the transformation of the data as a pre-processing step for the FDA analysis is provided.

2.1 | Data: Realized volatility

Realized volatility data were obtained from the Oxford-Man Institute of Quantitative Finance: Realized Library, created by Heber et al.[15] and is available at https://realized.oxford-man.ox.ac.uk/data. It is freely available to the public without any restrictions, and hence there was no need for ethical clearances or non-disclosure agreements. This realized library is determined from underlying high-frequency data obtained from the Thomson Reuters Tick History database. This data is of high quality and has been cleaned to remove observations recorded outside the interval of when the exchange is open.[15] The econometric measures available within the data set have been computed using high-frequency daily financial returns, with techniques summarized in Shephard and Sheppard.[16] At the time of writing, this data set was last updated on June 28, 2022 and contained daily financial estimators for 31 stock market indices. The *rv10* column contains the daily realized variance at a sampling frequency of 10 min. The daily realized variance is computed using the sum of squared intra-day returns over a specific time-frequency. This can be described mathematically by the formula,

$$RV_t = \sum x_{j,t}^2, \tag{1}$$

where $x_{j,t}$ is a return value calculated by,

$$x_{j,t} = X_{t_{j,t}} - X_{t_{j-1,t}} \tag{2}$$

and $t_{j,t}$ is the time of a trade on the tth day. In this case, the time-frequency is 10 min. Therefore, the *rv10* column contains values computed by summing the squared returns of a specific index every 10 min for each trading day. The high-frequency return data, used to compute the realized variance, contains market microstructure noise. This is the reason for introducing a sampling frequency, as this averaging is beneficial in reducing the effect of microstructure noise.[15] The reason for choosing the realized variance with 10 min sampling frequency (*rv10*) instead of the 5-min sampling frequency (*rv5*) is because the lower sampling frequency alleviates more noise. Consequently, fewer intra-day returns are used in the calculation, and the measurement error may increase slightly, but the noise reduction is of a higher priority.[6]

The international financial markets this research focuses on are Brazil, China, the Eurozone, India, the United Kingdom (UK), and the United States (US). These six financial markets represent emerging (Brazil, China, and India) and developed markets (the Eurozone, UK and US). More specifically, Brazil is represented by the BVSP BOVESPA Index, China by the Shanghai Composite Index, the Eurozone by EURO STOXX 50 (50 stocks across 11 countries in Europe), India by NIFTY 50, the UK by FTSE 100 and the US by the S&P 500 Index. At the time of writing, the daily data is available up until June 28, 2022. However, the date range of interest for this research is between January 4, 2000 (ensuring the earliest available date for all respective markets to be the same) and December 31, 2021 (a total time period of 22 years).

2.2 | Data: Infectious disease data set

The newspaper-based equity market volatility due to infectious diseases (EMVID) tracker data was obtained from the Economic Policy Uncertainty website, available at http://policyuncertainty.com/infectious_EMV.html. It is freely available to the public without any restrictions, and for this too there was no need for ethical clearances or non-disclosure agreements. This data set was created by Baker et al.[17] with daily measures ranging from 01 January 1985 to date and is being updated daily. The date range of interest for this data set is between January 4, 2000 and December 31, 2021. This is 22 years worth of daily data, and the time period is consistent with the realized volatility data set.

To construct the EMVID,[17] specify four sets of terms, E: economic, economy, financial; M: “stock market”, equity, equities, “Standard and Poors”; V: volatility, volatile, uncertain, uncertainty, risk, risky; ID: epidemic, pandemic, virus, flu, disease, coronavirus, mers, sars, ebola, H5N1, H1N1, and then obtain daily counts of newspaper articles that contain at least one term in each of E, M, V, and ID across approximately 3000 US newspapers. The raw EMVID counts are scaled by the count of all articles in the same day and, finally, the authors multiplicatively rescale the resulting series to match the level of the Chicago Board Options Exchange (CBOE)'s Volatility Index VIX, by using the overall EMV index and scaling the EMVID index to reflect the ratio of EMVID articles to total EMV articles.

2.3 | Transformations

The data pre-processing step involved aligning the realized volatility data. Since the date range spans 22 years, the daily realized volatility data was grouped according to years, with average number of trading days per year being around 252.

However, this changed for some years and is also different across international financial markets. Additionally, trading days shift every year, and public holidays vary year on year, and leap years alter the number of days in a year. To solve this issue, grouping the data into years and aligning the dates was necessary. The next data pre-processing step involves aligning the EMVID index data. This data also needs to undergo alignment and have the daily EMVID index measure grouped by year. Since the structure of this data set varies from the realized volatility data set, a minor adaption of that alignment function was done.

The next necessary step in FDA analysis is the smoothing process to convert discrete observations to functions or curves. Two stages are required when going from discrete data to smooth curves or functions: First, a set of basis functions ϕ_k that are mathematically independent of each other must be created; and second, a smoothing method must be chosen to create a linear combination or weighted sum of these K basis functions, which best estimate a curve x from the discrete observations. The B-spline basis system was chosen for this analysis over the Fourier system for many reasons, however, the key reason was due to the aperiodicity of the underlying data sets.* Additionally, B-splines are known to be superior computationally and have greater flexibility compared to other basis systems.[18] As a reminder, a spline function $S(t)$ can be described mathematically by,

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau), \tag{3}$$

where the number of basis functions is $K = m + L - 1$ since all the supports for the B-spline functions start at the left boundary[18] with the assumption that all interior knots are discrete, the parameter $B_k(t, \tau)$ is the value of the B-spline basis function at time t with the knot sequence τ. Therefore, a B-spline function is determined by two parameters: the knot sequence τ (as this will infer the number of interior knots L) and the order m of the polynomials. The subsequent sections will expand on the choice of knot sequence τ and order m of the polynomials.

For this study, the roughness penalty smoothing method was chosen as it often produces better results regarding smoothing and derivative estimation, with the smoothing parameter λ which controls the trade-off between fitting the data and the variability of the function (fitting versus smoothing).[18–20] It is common to choose this parameter through visual inspection, however, the generalized cross-validation (GCV) criterion is the preferable method. This automatic method, on average, yields a smoothing parameter λ that is exceptionally close to the optimal value.[18] In summary, it was vital to develop a dynamic smoothing (DS) algorithm with regards to the polynomial order m and smoothing parameter λ, as this allowed for the option of optimized smoothing as well as providing a foundation for multiple types of FDA techniques, allowing for flexibility and reusability when multiple analyses, such as derivative analysis and functional regression, are carried out on multiple data sets. This point is an elegant segue to Section 2.3.1, which deals with the practical implementation of the dynamic smoothing algorithm.

2.3.1 | The dynamic smoothing algorithm: Practical implementation

Some key functions used to develop the dynamic smoothing (DS) algorithm are summarized in Table 1. For further documentation on each function, such as function arguments and outputs, see Ramsay et al.[21]

Three parameters determine the DS namely the knot sequence τ, the order m of the polynomials, and the value for the smoothing parameter λ. Two of the three parameters, except for knot sequence τ, that determine the smoothing are dynamic and are chosen based on the specific purpose of the smoothing, such as derivative analysis or functional

TABLE 1 Key functions used to develop the dynamic smoothing algorithm.

Function	Description
`create.bspline.basis`	Create a B-spline basis
`int2Lfd`	Converts an integer to a linear differential operator
`fdPar`	Defines a functional parameter object
`smooth.basis`	Constructs a functional data object by smoothing data using a roughness penalty
`fd`	Defines a functional data object

regression, for example. The choice of knot sequence τ will infer the number of interior knots L and the number of basis functions K within the B-spline basis system since $K = m + L - 1$ when the B-spline functions start at the left boundary. However, in most applications, a knot will exist at every breakpoint except for the beginning and end of the interval range.[22] This ensures that the adjoining polynomials have derivatives that match at the point where they join. Thus, the number of basis functions becomes $K = m + L - 2$ when implemented practically. For this research, it was decided to place a breakpoint at each argument value. This knot sequence, τ, ensures that the smoothing of regions with high and low curvature will be catered for across all markets and years of interest.

3 | EMPIRICAL METHODS

In this section, we describe in great detail the econometric approaches undertaken for our analyses.

3.1 | Derivative analysis

A derivative (rate of change) analysis is a key contribution of this research, with the first derivative (velocity) and second derivative (acceleration) of these smooth curves being of particular interest as they are indicative of the rates of change. A final, but important note on the dynamic smoothing algorithm, is the handling of missing values within the data. The `smooth.basis` function in the *fda* package does not support the smoothing of data containing missing values[†].

To ensure the highest order derivative n to be analyzed is smooth, the derivative of order $n + 2$ must be penalized in the roughness penalty approach,[18] in our case it is the second derivative. Thus, the original function's fourth derivative $(2 + 2 = 4)$ must be penalized to ensure the second derivative is smooth. Furthermore, the order of a B-spline basis must be at least two higher than the highest derivative to be penalized. Therefore, for all derivative analyses conducted in this research, an order six $(m = 6)$ B-spline system was used in the smoothing process.

Figure 1 is a plot of the GCV criterion using a six-order B-spline basis system $(m = 6)$ over a range of smoothing parameter λ values. The GCV criterion plot indicates that the optimal smoothing parameter λ value, across all six markets, is when $\lambda = 10^4$, as this is where the GCV is at a minimum. Therefore, the realized volatility data for the general derivative estimation case was smoothed using the dynamic smoothing algorithm with a sixth-order B-spline basis system $(m = 6)$ and a smoothing parameter $\lambda = 10^4$ for all six markets.

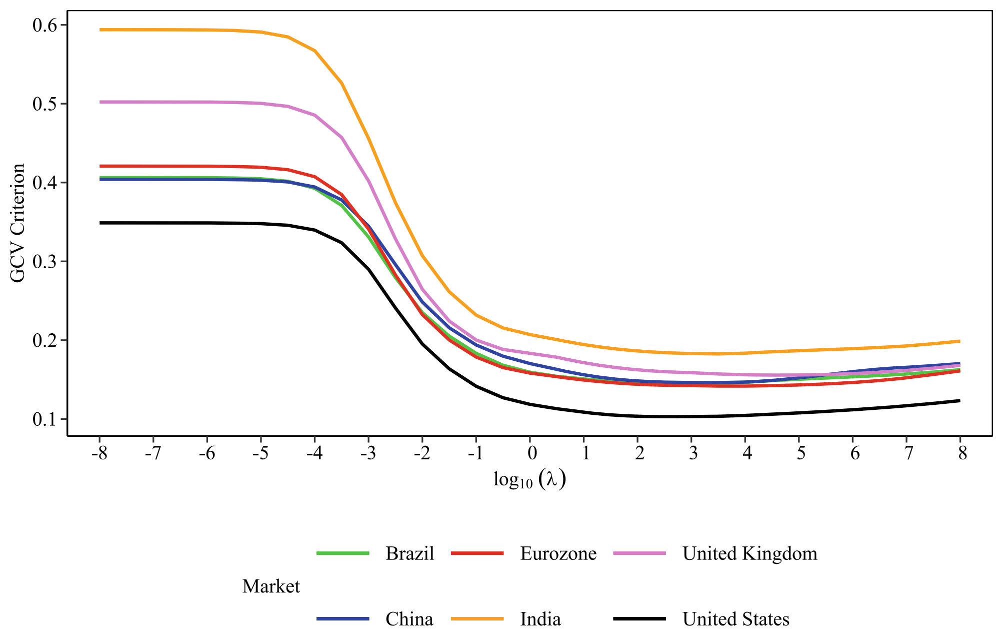


FIGURE 1 The GCV criterion for each financial market, with a B-spline system of order six $(m = 6)$, over a range of smoothing parameter λ values.

3.1.1 | Phase-plane plot analysis

A phase-plane plot is a plot of the first derivative (velocity) against the second derivative (acceleration) of a particular smooth function. Phase-plane plot analysis provides completely different insights due to the nature of plotting one order of derivative versus the other, and are useful as they illustrate the exchange of energy (potential versus kinetic) within a system, and many insights into a particular system can be derived from such plots. A phase-plane plot's X-axis range (width) and Y-axis range (height) relate to the amount of kinetic and potential energy, respectively. From an economics perspective, kinetic energy corresponds to the manufacturing and shipping of commodities (economic movement), and potential energy corresponds to the amount of capital, resources, and material available to bring about economical movement.[18] This concept can be extended to international financial market volatilities. Therefore, concerning the market volatility phase-plane plots, the extraction of the amount of kinetic energy (X-axis range) and potential energy (Y-axis range) has the potential to expound valuable market insights.

3.1.2 | Generating a phase-plane plot

Once the realized volatility data was smoothed using the DS algorithm, the resultant smooth functional data object was passed as an argument to the `phaseplanePlot` function within the *fda* package. Extracting the X and Y axes ranges from a phase-plane plot dynamically is no trivial task, as the algorithm developed needs to be dynamic to find the correct range for all plots between the years 2000 and 2021, across all six international financial markets. Figure 2 is a straightforward example of a phase-plane plot since most phase-plane plots have a spiraling irregular shape. However, Figure 2 will suffice for an explanation.

By observing Figure 2, the X-axis range is defined as the difference between point B and point D. Similarly, the Y-axis range is defined as the difference between point A and point C. It is clear that point A and point C exist where the first derivative (velocity) is zero. Likewise, point B and point D exist where the second derivative (acceleration) is zero. The flaw in this approach is that the matrix returned by this plot contains a set of discrete values at which the functional data object was evaluated. Therefore one cannot test for values at precisely zero. To avoid this issue, testing for a change in the sign (positive to negative, or negative to positive) of the respective derivative provides the index at which the points exist. An algorithm was developed to execute the process of extracting the X and Y spans over the years and for all the countries of interest.[‡] This functional data object is identical to the one utilized to create the phase-plane plots. Therefore, the smoothing parameters are identical, and the B-spline order remains at $m = 6$ and the smoothing parameter at $\lambda = 10^8$ for this specific analysis.

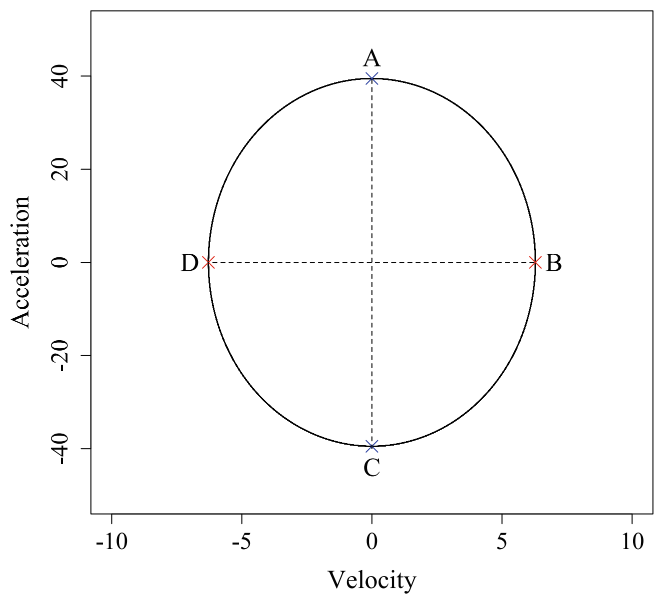


FIGURE 2 A phase-plane plot of the $\sin(2\pi t)$ function.

3.2 | Functional regression

We next summarizes the practical methods implemented to conduct functional regression analyses on the realized volatility data set, using the daily newspaper-based infectious disease index as a predictor. The functional regression model implementation involves few steps namely: The choice of functional regression model, smoothing of the relevant data sets, regression smoothing parameters, essential *fda* functions, and the functional regression algorithm are all necessary to produce a functional regression model. The variable of interest (response variable) in this study is the realized volatility of the specific international financial markets, as a functional data object. Therefore, the type of model to be implemented is a functional response regression model. The EMVID index is the explanatory (predictor) variable here and had the property of being scalar as it contains daily EMVID index values over the 22 years of interest. The response variable is functional, and the explanatory variable is scalar. As such, the type of regression model most appropriate in this case is function-on-scalar regression. However, to conduct regression on the 22 realized volatility curves, only 22 EMVID index data points could be used within the regression model. Aggregating data, such as taking the mean for each year, is not desirable as it could result in information loss (see Das et al.[14]). However, all the daily EMVID index data points could be utilized in the regression if the EMVID index data is also smoothed to produce yearly curves. In this case, both variables are functional, and the type of regression model most appropriate is function-on-function regression. The function-on-function regression has the following mathematical notation,

$$y_i(t) = \beta_0(t) + \sum_{j=1}^{q-1} x_{ij}(t)\beta_j(t) + \epsilon_i(t), \tag{4}$$

where $y_i(t)$ are realized volatility curves for years $i = 2000, 2001, \dots, 2021$. The functional explanatory variable $x_{ij}(t)$ is the EMVID index curves over the same yearly i period. The EMVID index data was smoothed using the DS algorithm developed. A sixth-order B-spline basis system ($m = 6$) was chosen to be consistent across analyses. The smoothing parameter was once again chosen by using the GCV criterion. The GCV criterion for the EMVID index over a range of smoothing parameter values λ is given in Figure 3.

The realized volatility response curves y_i, across each market, were smoothed using a sixth order B-spline ($m = 6$) and a smoothing parameter of $\lambda = 10^4$ as per Figure 1 and consistent with the general derivative estimation analysis case. The output of the functional regression model must also be smooth curves. The smoothing parameters for the predicted curves $\hat{y}_i$ (the result of the regression analysis) need to be chosen. The B-spline order was kept at six ($m = 6$), however, the smoothing parameter value was chosen based on the cross-validated integrated squared error (CVISE) score. The smoothing parameter λ, which minimizes the CVISE score, is the optimal choice of smoothing parameter for regression analysis. Figure 4 is a plot of the CVISE score using a six-order B-spline basis system ($m = 6$) over a range of smoothing parameter λ values. The CVISE plot does not show a notable decrease in the CVISE score for a particular smoothing

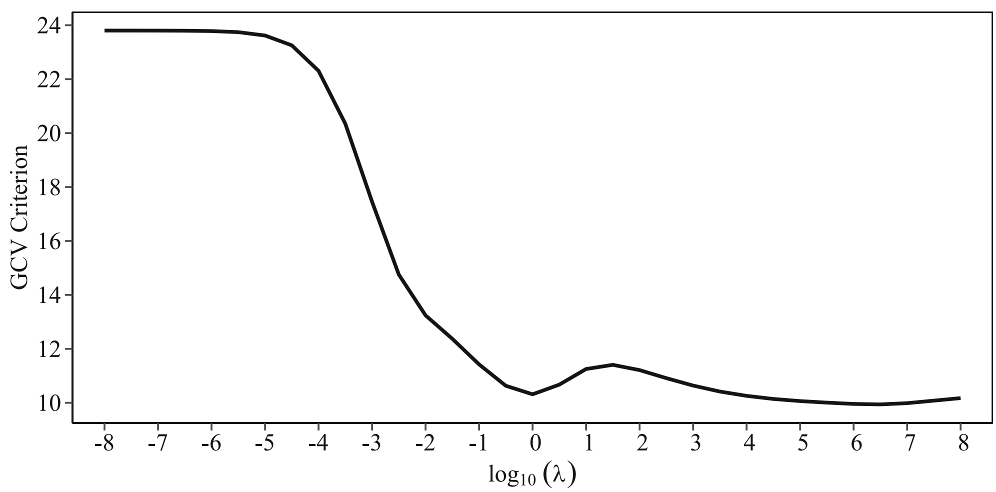


FIGURE 3 The EMVID index GCV criterion, with a B-spline system of order six ($m = 6$), over a range of smoothing parameter λ values.

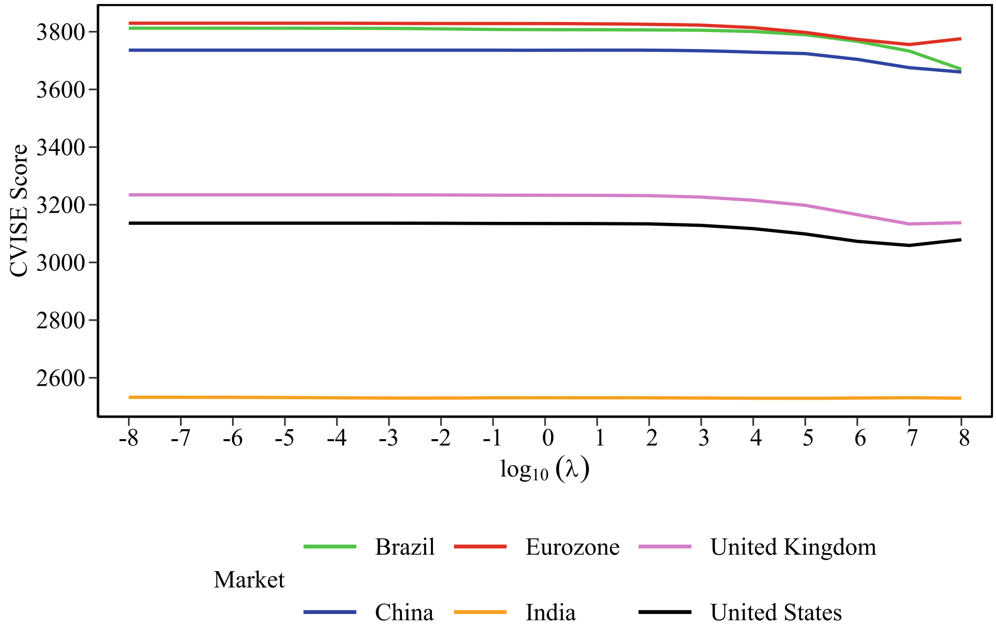


FIGURE 4 The CVISE score for each financial market, with a B-spline system of order six ($m = 6$), over a range of smoothing parameter λ values.

TABLE 2 Key functions used to develop the dynamic function-on-function regression algorithm.

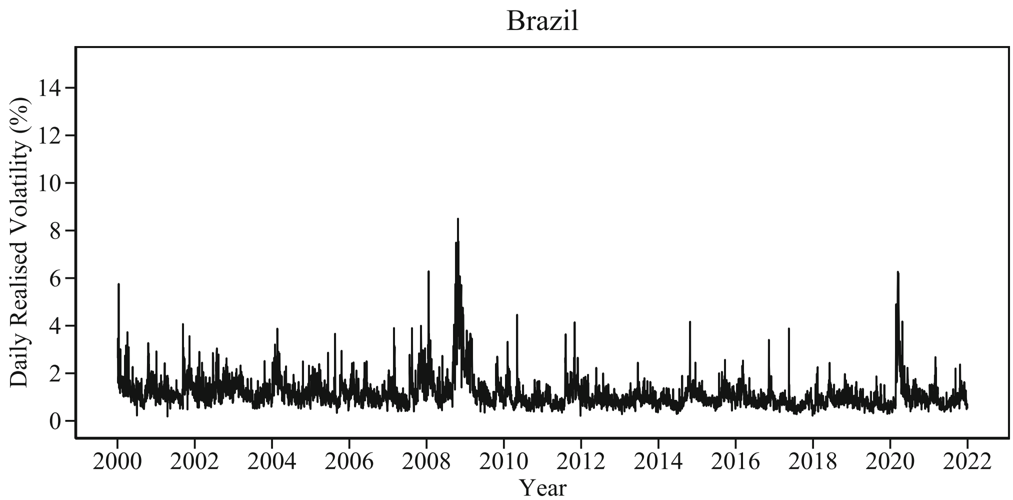
Function	Description
`fd`	Defines a functional data object
`fdPar`	Defines a functional parameter object
`fRegress`	Carry out various types of functional regression analysis

parameter value. Thus, the smoothing parameter was chosen to be $\lambda = 10^4$ consistent with the response variable. In summary, the smoothing parameters for both the actual y_i and predicted $\hat{y}_i$ response curves were chosen to be $m = 6$ and $\lambda = 10^4$ for the regression analysis.

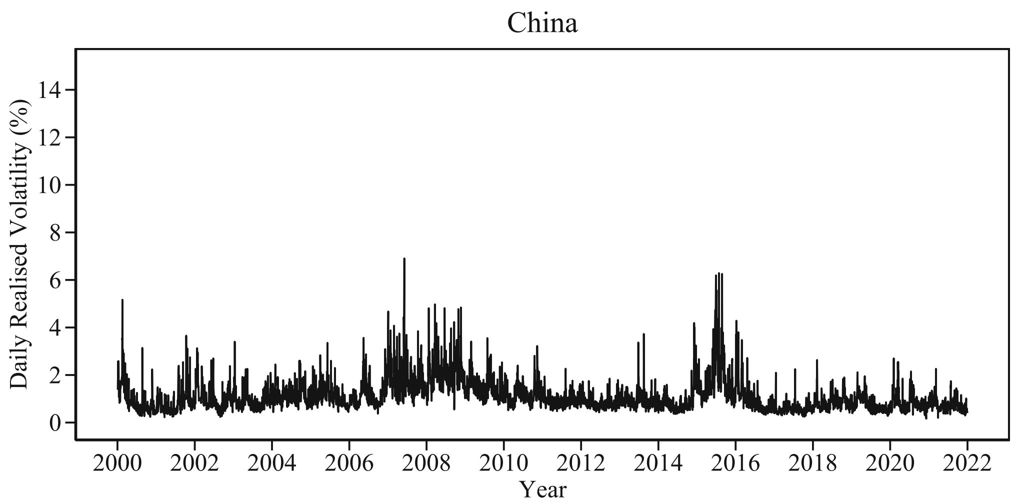
Some key functions used to develop the dynamic function-on-function regression algorithm are summarized in Table 2. For further documentation on each function, such as function arguments and outputs, see Ramsay et al.[21] The EMVID index data contains values for dates that may not exist in the realized volatility data due to the aforementioned trading days issue. Thus, logic was incorporated to compare the market-specific realized volatility data set and the EMVID index data set to ensure both data sets have the same number of observations.

4 | EMPIRICAL FINDINGS

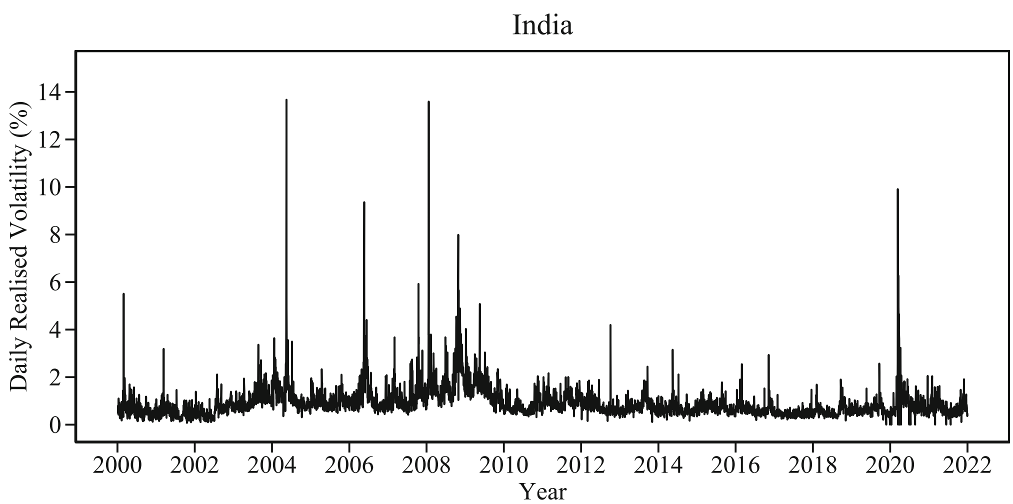
An initial exploratory data analysis was conducted on the realized volatility and infectious diseases index data sets. Figures 5 and 6 show the daily realized volatility (%) during the years 2000–2021, for the emerging and developed markets, respectively. From Figure 5, it is clear that two or three notable periods experienced excessively high realized volatility values across the three emerging markets. For Brazil and India, these notable periods correspond to the years 2008/9 (Global Financial Crisis), and 2020 (COVID-19 pandemic). India also experienced a notable increase in daily realized volatility in 2004 due to the UBS Securities scandal. China experienced high daily realized volatility values during the Global Financial Crisis (GFC). However, it did not show this during the COVID-19 pandemic year. The second notable period increase in volatility for the Chinese market was during the 2015/16 period when the Chinese stock market crashed. The maximum daily realized volatility during the COVID-pandemic was less than the maximum daily realized volatility during the GFC.



(A) Brazil

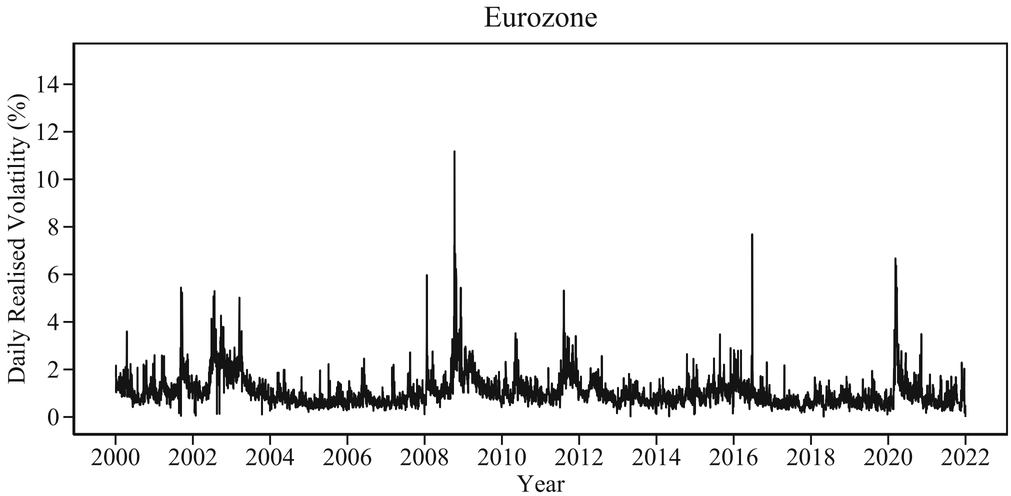


(B) China

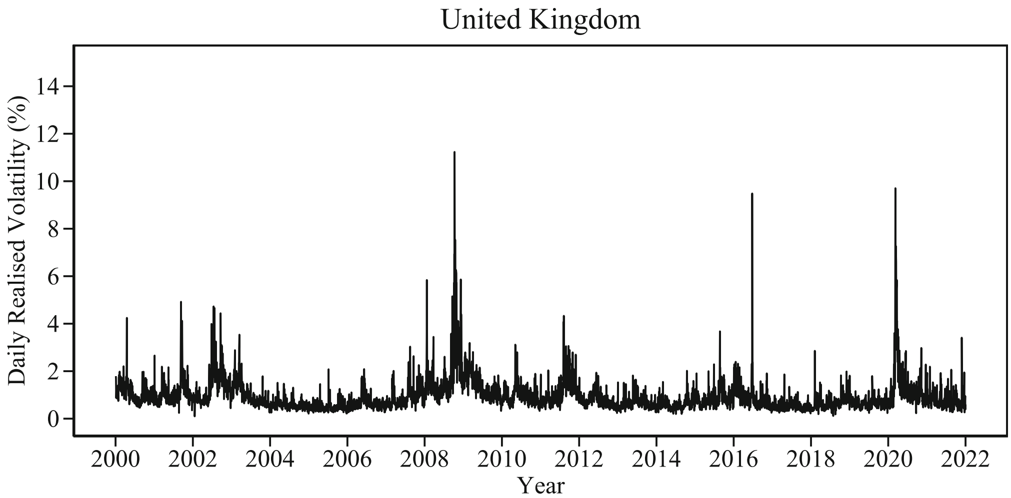


(C) India

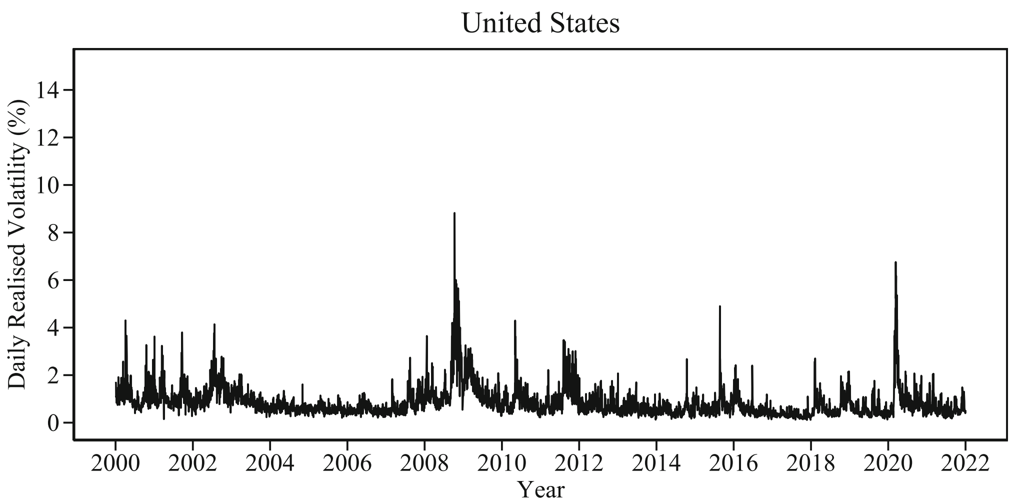
FIGURE 5 Time series plots of the daily realized volatility, from 2000 to 2021, for the emerging markets. (A) Brazil, (B) China, (C) India.



(A) The Eurozone



(B) United Kingdom



(C) United States

FIGURE 6 Time series plots of the daily realized volatility, from 2000 to 2021, for the developed markets. (A) The Eurozone, (B) United Kingdom, (C) United States.

Figure 6 shows similar findings for the developed markets. The Eurozone area, the United Kingdom (UK) and the United States (US) show three notable increases in daily realized volatility over the 2000–2021 time period. All three markets experienced increases during the years 2008/9 (Global Financial Crisis), 2015/16 (the Chinese stock market turbulence, the UK European Union membership referendum, and the H1N1 and Ebola epidemics) and 2020 (COVID-19 pandemic). The developed markets all indicate that the GFC period had the highest daily realized volatility, similar to the emerging markets.

Figures 7 and 8 show box plots of the daily realized volatility (%), for each year during the 2000–2021 period, for the emerging and developed markets, respectively. These figures illustrate more comprehensive insights, into some basic statistics of the daily realized volatility, for each year. A blue cross depicts the mean for each year on the box plot. Across emerging and developed markets, the mean realized volatility is often the highest during the 2009/9 years, corresponding to the GFC. This is in accordance with the trend seen in the time series plots. This suggests that, during the GFC, the increase in realized volatility was sustained for an extended period. Most notably, the mean for the COVID-19 pandemic (2020) across the markets was not significantly higher than in other years. This is not in accordance with the trend seen in the time series plots. In summary, during the COVID-19 pandemic, there were days with extremely high realized volatilities but the mean for the entire year was not substantial. This indicates that the increase in market volatility during the pandemic was relatively short-term.

Figure 9 is a time series plot of the daily newspaper-based equity market volatility due to the infectious diseases (EMVID) index during the 2000–2021 time period. There was a moderate spike in the EMVID index in 2009, which corresponds to the H1N1 (Swine Flu) pandemic. Furthermore, the EMVID index saw a minor increase in 2015 because of the Ebola or Zika virus epidemics. The EMVID index reached an all-time high in 2020 due to the COVID-19 pandemic. The EMVID index values during this period are unparalleled and, comparatively, reduce the previous spikes to minor status. Since this index is based on market volatility due to infectious diseases, it is clear that COVID-19 substantially affected market volatility. This supports Baker et al.[23]'s claim that COVID-19 affected market volatilities on an unprecedented scale.

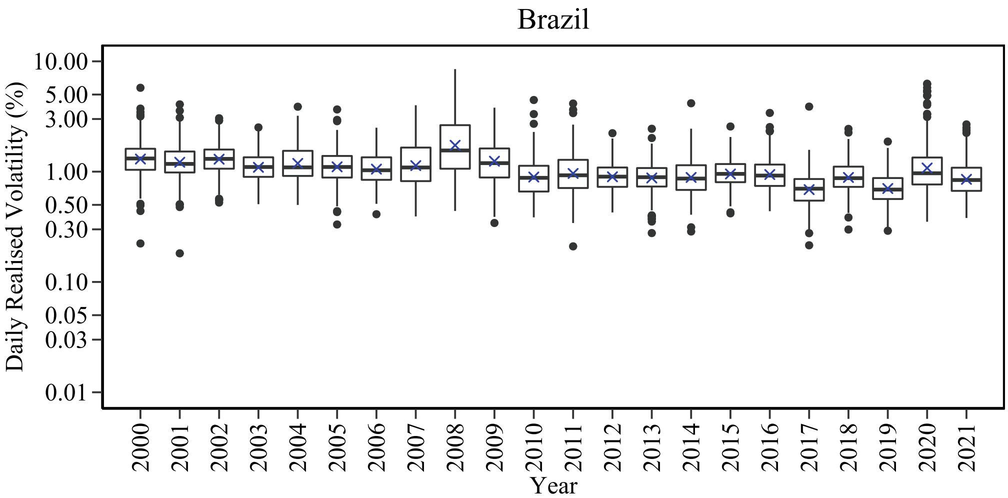
4.1 | Results from the derivative analyses

4.1.1 | Derivative analysis: The COVID-19 period

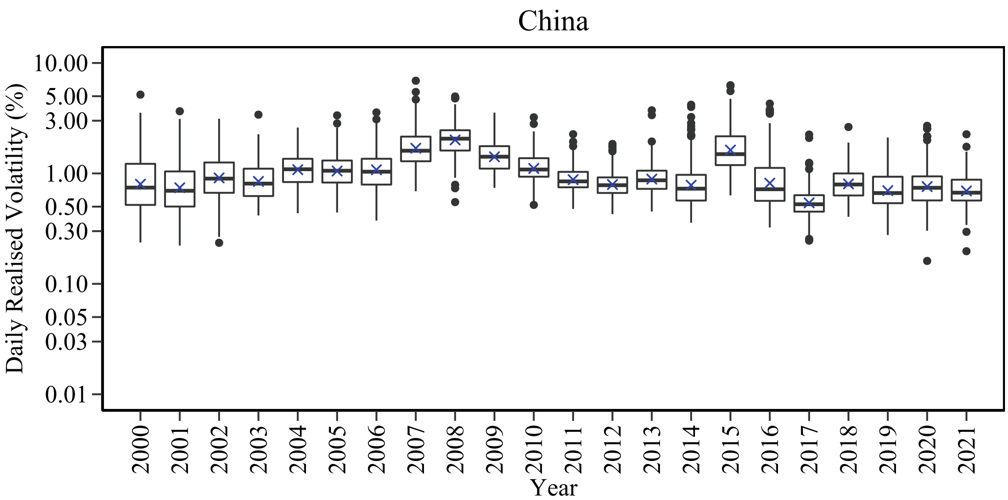
The rate of change (velocity) of the realized volatility, or risk, did not fluctuate excessively in the years before the COVID-19 pandemic across both the emerging and developed markets. However, when COVID-19 appeared, the realized volatility rate of change experienced a sharp increase, except for China. China did not experience as sharp a spike, however, the change in realized volatility did increase from January to March of 2020, just not on the same scale as the other markets. The change in volatility profiles (magnitude and shape) for Brazil, India, the Eurozone, the United Kingdom and the US during the COVID-19 period are very similar, with the sudden fluctuation in the rate of change of realized volatility being short-lived and did not persist beyond May 2020. Post COVID-19, during the 2021 year, the change in realized volatility remained relatively stable across all markets. Only Brazil, the Eurozone and the UK showed minor fluctuations. Thus, the extreme changes in volatility were limited to March, April and May of 2020 and did not persist for an extended period. The second derivative (acceleration) curves of the realized volatility for the COVID-19 period were similar to the first derivative.[§]

4.1.2 | Derivative analysis: The global financial crisis period

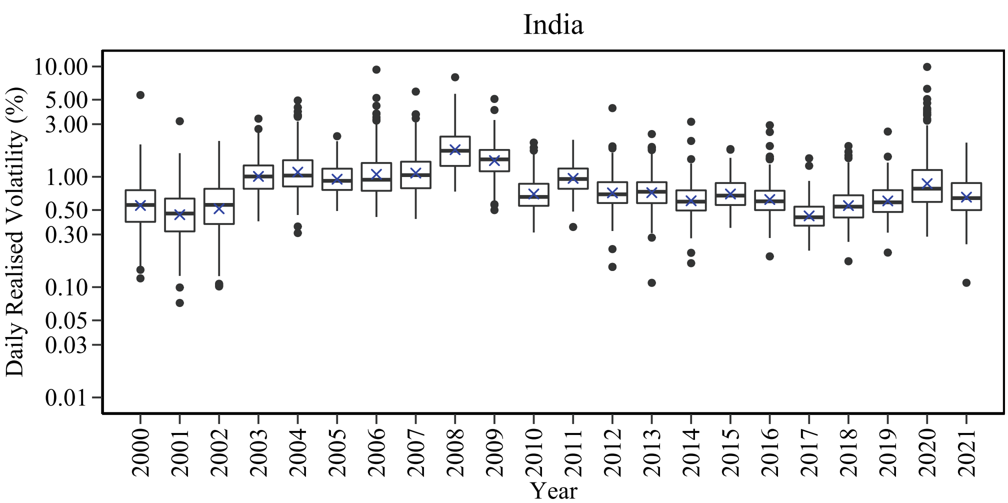
The first and second derivative curves for the GFC period indicated that there was no specific date when the GFC started, it is however synonymous with the collapse of the Lehman Brothers in 2008. During 2006 the rate of change of realized volatility was relatively stable throughout the year for most of the markets. China and India were the exceptions, both experienced fluctuations in the velocity of realized volatility in May, 2006. During 2007, the emerging markets experienced significant movement in their respective realized volatility velocities. The developed markets did not follow this trend. When the GFC started to gain momentum in 2008, none of the six markets escaped its effect. Both the emerging and developed markets saw an increase in the rate of change of realized volatility. These fluctuations in risk persisted for the majority of 2008, unlike the short-lived effect COVID-19 had on the markets. At the beginning of 2008, all markets



(A) Brazil

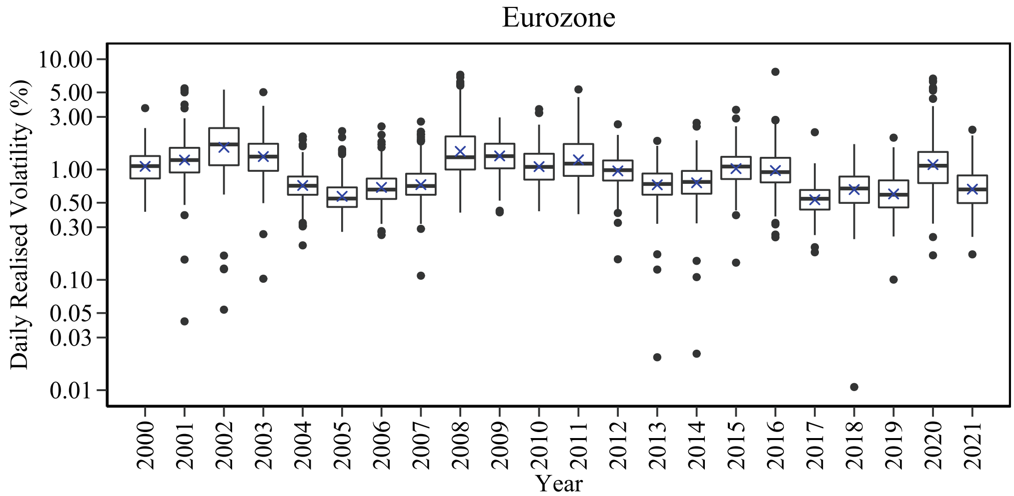


(B) China

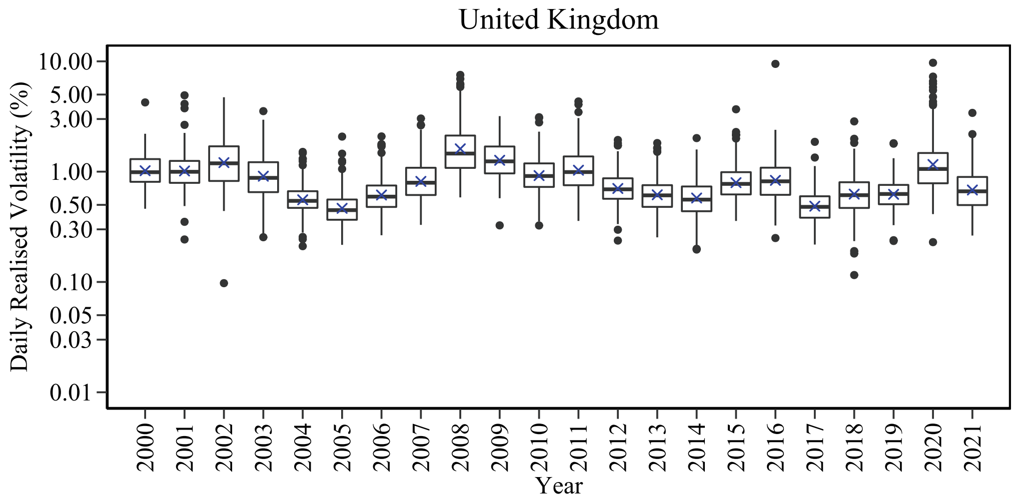


(C) India

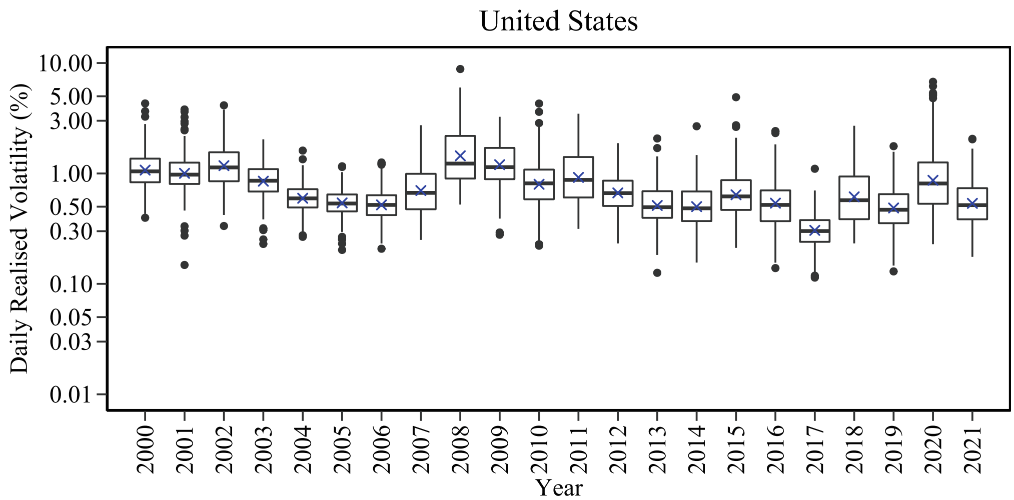
FIGURE 7 Box plots of the daily realized volatility, from 2000 to 2021, for the emerging markets. (A) Brazil, (B) China, (C) India.



(A) The Eurozone



(B) United Kingdom



(C) United States

FIGURE 8 Box plots of the daily realized volatility, from 2000 to 2021, for the developed markets. (A) The Eurozone, (B) United Kingdom, (C) United States.

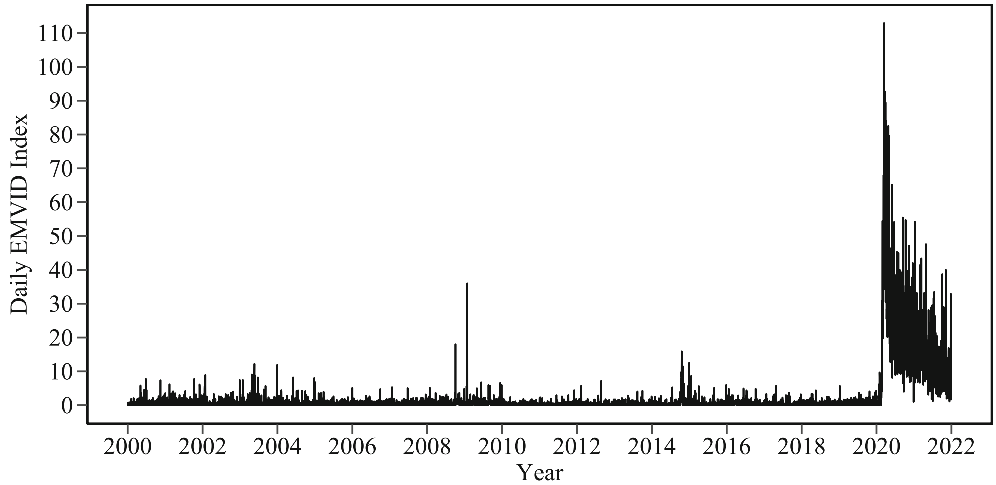


FIGURE 9 The daily EMVID index from 2000 to 2021.

experienced an increase in the change in realized volatility, with India having the most notable shift. The three emerging markets have similar profiles, with a small spike at the beginning of the year and large spikes in velocities starting in September of 2008, which coincides with the collapse of the Lehman Brothers. The developed markets seemed to experience greater shifts in the change of realized volatility than the emerging markets. During 2009, all six markets experienced minor fluctuations throughout the year, however, not on the scale of 2008.¶

4.1.3 | Derivative analysis: Crisis comparison

Figures 10 and 11 are the first and second derivative curves specifically for the COVID-19 (2020) and GFC (2008) years, respectively, with the derivative curves for the three emerging markets (Brazil, China and India) on the left, and the three developed markets (the Eurozone, UK and US) on the right. Figures 10 and 11 thus allow for a direct comparison between the two crises years. The conclusion taken away from these two figures is that the risk instability caused by the GFC persisted for much longer than the instability caused by the COVID-19 pandemic, while for the COVID-19 pandemic year it saw a sudden change in the realized volatility around March 2020, but proceeded to stabilize toward the end of May 2020.

4.1.4 | Phase-plane plots: A comparison between the global financial crisis and COVID-19 periods

The phase-plane plots used in this application were aimed to study the energy within the six international financial markets and were created by plotting the velocity on the X-axis and acceleration on the Y-axis. Starting from January, the small "j", a large cycle begins and progresses until May. This cycle has a large radius, indicating a large amount of energy transfer within the event. A large amount of kinetic energy from January to May of 2020 indicates that these markets experienced significant movement in the daily realized volatility.

Figures 12 and 13 comparing the GFC (2008) and COVID-19 (2020) periods for the emerging and developed markets, respectively. Observing Figure 12, it is clear to see that for Brazil and India, the potential and kinetic energy exchange in 2020 is greater than in 2008. For China, the opposite of this holds true. The large cycles also do not occur in the same months, which is expected, as each crisis occurs in its own time frame. Figure 13 depicts a similar trend for the developed markets. The potential and kinetic energy exchange during 2020 was greater than 2008. This indicates, for both the emerging and developed markets, that the rate of change (not the magnitude) in realized volatility during COVID-19 was greater than during the GFC. Therefore, the uncertainty surrounding the financial markets during the COVID-19 period increased and subsided rapidly. In comparison, the uncertainty surrounding the GFC came about at a slower rate.

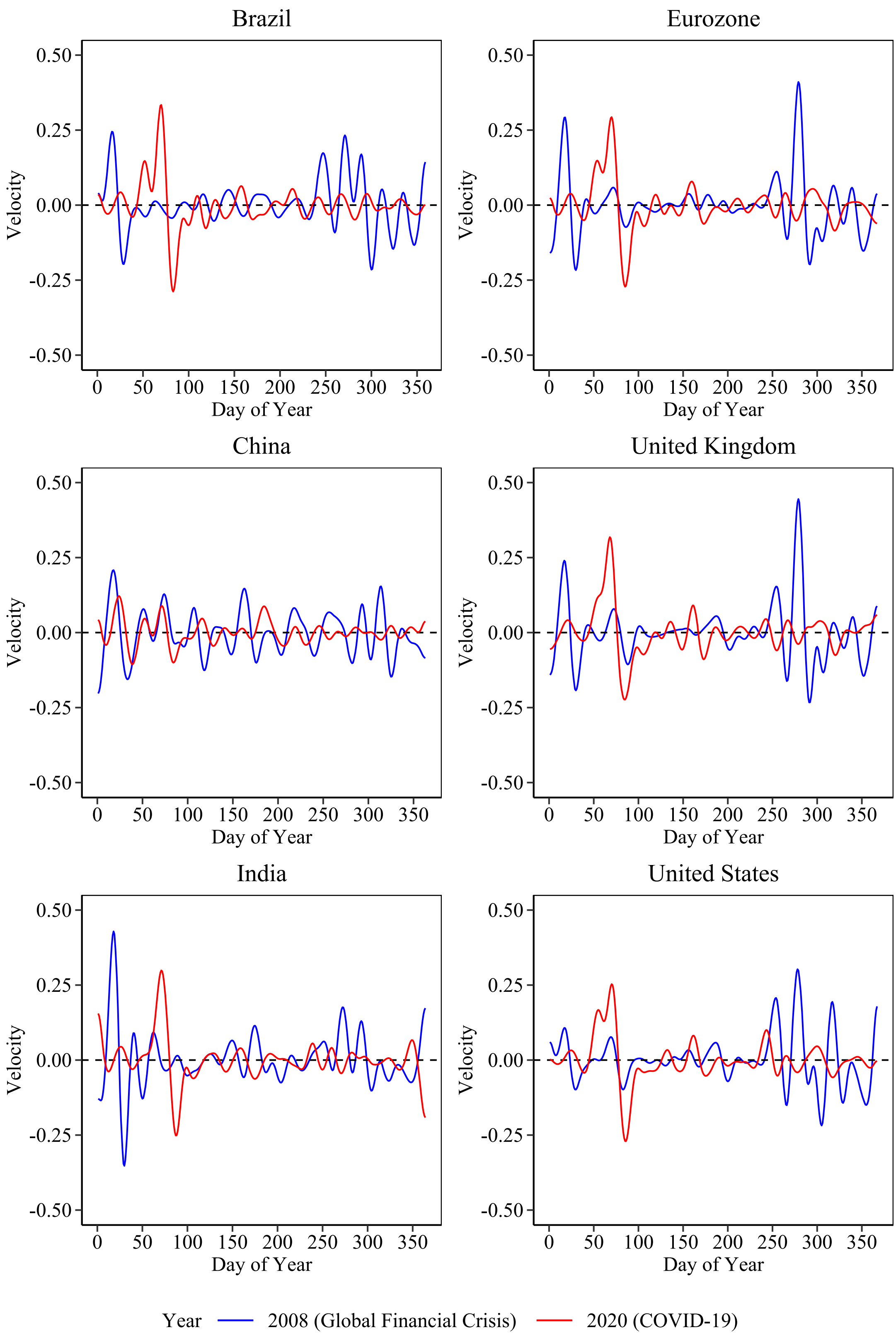


FIGURE 10 The first derivative of the realized volatility over the COVID-19 and GFC periods.

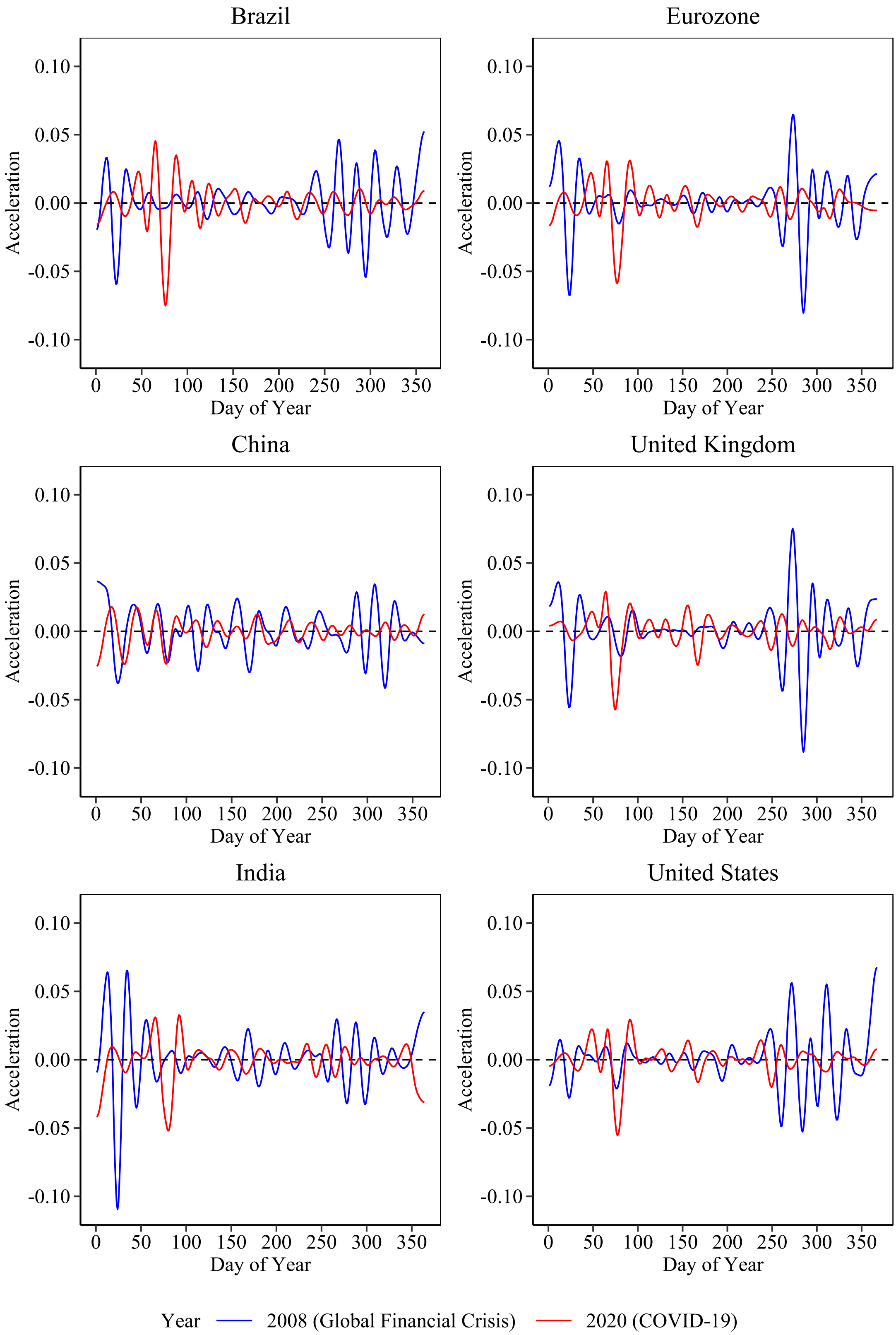
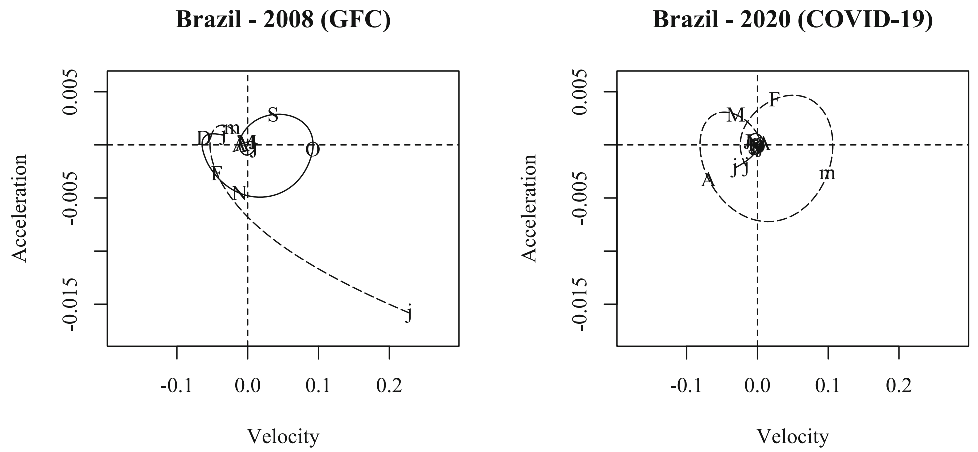
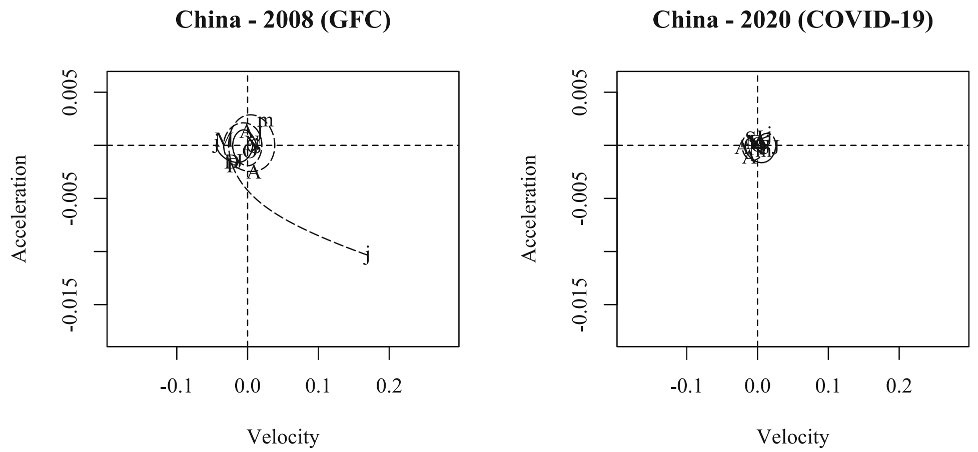


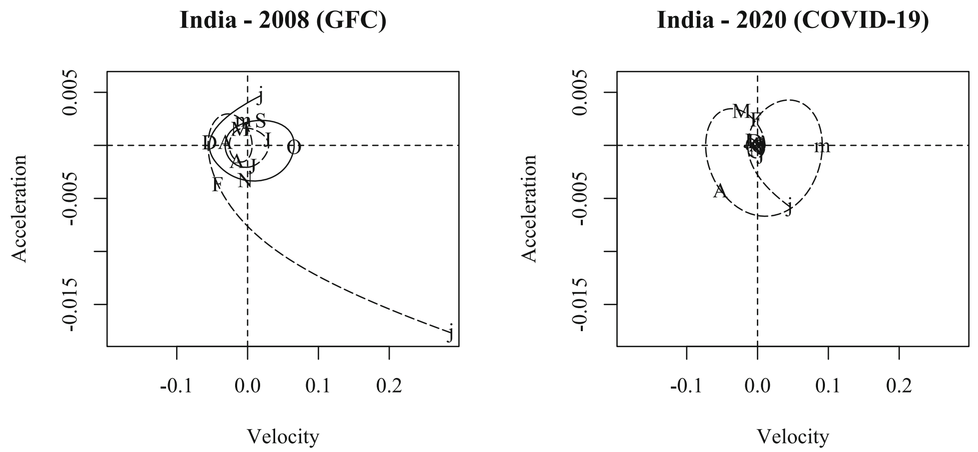
FIGURE 11 The second derivative of the realized volatility over the COVID-19 and GFC periods.



(A) Brazil

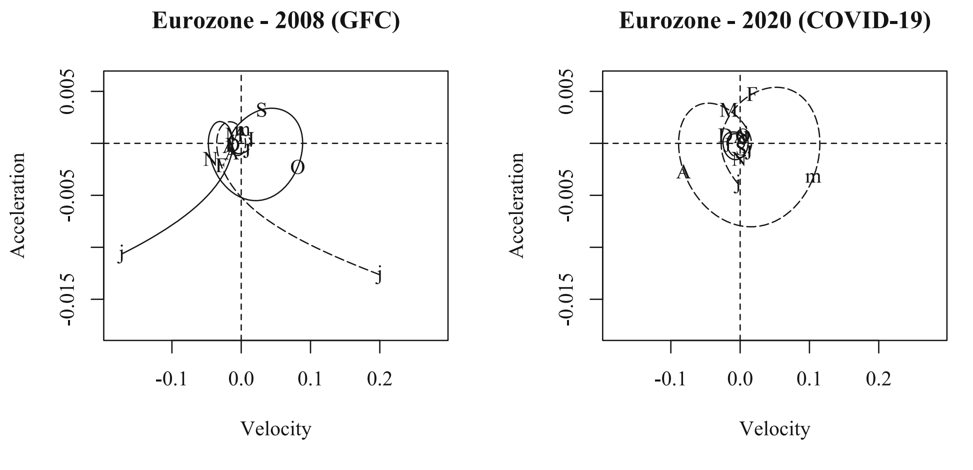


(B) China

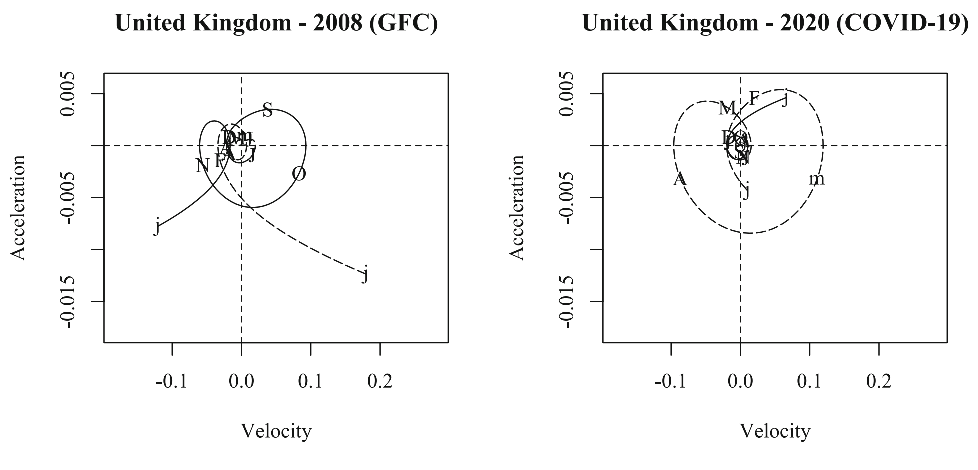


(C) India

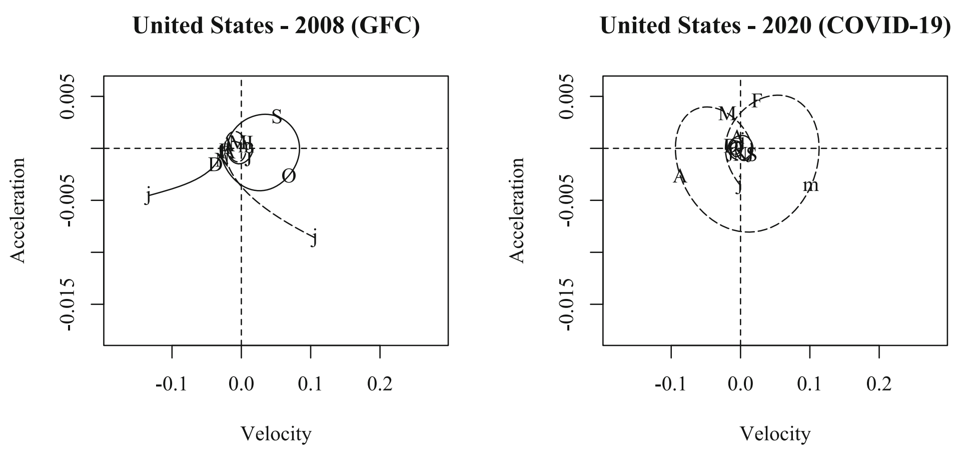
FIGURE 12 Phase-plane plots showing a comparison of the GFC (2008) and COVID-19 (2020) period for the emerging markets. (A) Brazil, (B) China, (C) India.



(A) The Eurozone



(B) United Kingdom



(C) United States

FIGURE 13 Phase-plane plots showing a comparison of the GFC (2008) and COVID-19 (2020) period for the developed markets. (A) The Eurozone, (B) United Kingdom, (C) United States.

4.1.5 | Kinetic and potential energy extraction

The difference between the maximum and minimum velocity value (X-axis range) on a phase-plane plot is a quantitative measure representing the amount of kinetic energy within the system. Likewise, the difference between the maximum and minimum acceleration value (Y-axis range), on a phase-plane plot is a quantitative measure representing the amount of potential energy within the system. The total kinetic and potential energy values were extracted from each phase-plane plot and are summarized in Table 3. The values were extracted for the entire range of interest, namely from 2000 to 2021, to get a sense of what the values would be both with and without a financial crisis. The results of the Global Financial Crisis and COVID-19 pandemic years are highlighted in red.

The amount of kinetic energy within the system indicates the extent to which the velocity of daily realized volatility fluctuated throughout the year. The results in Table 3 suggest that the amount of kinetic energy experienced during 2008 and 2020 was much higher than all the other years, indicating some form of financial instability in the markets. It can also be seen that during the COVID-19 pandemic, the amount of kinetic energy experienced was greater than the GFC across all six markets. This means that the change in daily volatility was much faster during COVID-19. The initial shock of COVID-19 had a significant and swift effect on the realized volatility of both emerging and developed markets. Again, the exception to this trend is China. The amount of kinetic energy experienced by the Chinese market is quite similar across the years, crisis or no crisis. The GFC had higher kinetic energy than the COVID-19 pandemic.

TABLE 3 A summary of the total kinetic energy KE_t (X-axis range) and total potential energy PE_t (Y-axis range), extracted from the phase-plane plots, for each market.

	Country											
	Emerging markets						Developed markets					
	Brazil		China		India		The Eurozone		United Kingdom		United States	
Year	KE_t	PE_t	KE_t	PE_t	KE_t	PE_t	KE_t	PE_t	KE_t	PE_t	KE_t	PE_t
2000	0.051	0.004	0.084	0.010	0.043	0.003	0.031	0.003	0.030	0.004	0.037	0.002
2001	0.054	0.003	0.057	0.004	0.032	0.003	0.083	0.005	0.063	0.004	0.051	0.005
2002	0.043	0.003	0.051	0.007	0.031	0.002	0.103	0.008	0.085	0.006	0.075	0.005
2003	0.023	0.001	0.049	0.004	0.023	0.002	0.061	0.003	0.037	0.002	0.022	0.001
2004	0.038	0.003	0.050	0.003	0.125	0.008	0.028	0.002	0.017	0.001	0.020	0.001
2005	0.044	0.004	0.024	0.002	0.035	0.004	0.016	0.001	0.016	0.001	0.016	0.001
2006	0.055	0.004	0.071	0.004	0.112	0.005	0.039	0.002	0.041	0.002	0.015	0.001
2007	0.073	0.005	0.050	0.003	0.063	0.005	0.035	0.003	0.043	0.003	0.055	0.004
2008	0.157	0.009	0.082	0.007	0.121	0.012	0.136	0.007	0.153	0.008	0.115	0.006
2009	0.074	0.004	0.058	0.006	0.051	0.005	0.021	0.003	0.023	0.003	0.030	0.001
2010	0.059	0.004	0.045	0.004	0.027	0.002	0.066	0.004	0.057	0.003	0.063	0.004
2011	0.051	0.004	0.018	0.002	0.028	0.003	0.064	0.003	0.055	0.003	0.055	0.003
2012	0.030	0.002	0.028	0.003	0.019	0.002	0.027	0.001	0.025	0.002	0.020	0.002
2013	0.029	0.002	0.031	0.002	0.036	0.002	0.027	0.003	0.025	0.002	0.025	0.003
2014	0.031	0.002	0.059	0.006	0.019	0.002	0.040	0.004	0.034	0.004	0.045	0.005
2015	0.021	0.002	0.108	0.007	0.027	0.002	0.045	0.005	0.045	0.003	0.055	0.003
2016	0.051	0.002	0.024	0.002	0.027	0.003	0.056	0.004	0.061	0.004	0.028	0.002
2017	0.020	0.002	0.017	0.002	0.011	0.002	0.019	0.001	0.011	0.001	0.010	0.001
2018	0.043	0.002	0.035	0.003	0.035	0.002	0.029	0.003	0.026	0.003	0.046	0.003
2019	0.018	0.003	0.037	0.003	0.021	0.001	0.030	0.003	0.023	0.002	0.030	0.002
2020	0.188	0.010	0.048	0.002	0.165	0.009	0.204	0.012	0.215	0.012	0.207	0.011
2021	0.035	0.001	0.032	0.002	0.019	0.002	0.043	0.005	0.033	0.004	0.036	0.004

The amount of potential energy within the system indicates the extent to which the acceleration of daily realized volatility fluctuated throughout the year. According to Table 3, the potential energy across the six markets is higher during years that experienced a financial crisis. The developed markets experienced a higher acceleration in realized volatility during the COVID-19 pandemic than the emerging markets. The realized volatility or uncertainty had a higher acceleration during the COVID-19 pandemic than during the GFC. This supports the aforementioned theory that market uncertainty arose much more quickly during COVID-19 but did not persist for an extended period.

4.2 | Functional regression

The realized volatility, for each market from 2000 to 2021, was predicted using a function-on-function regression model. The daily newspaper-based equity market volatility due to infectious diseases (EMVID) index was used as the only predictor variable. The regression analysis was carried out for all 22 years, but the focus is on 2020, the COVID-19 period. It is known that the COVID-19 pandemic affected the realized volatility in 2020, but the aim is to quantify the relationship between infectious diseases and market volatility.

4.2.1 | Correlation analysis of the realized volatility and EMVID index

Before proceeding with the regression analysis, the correlation coefficient was calculated for the realized volatility and EMVID index to investigate the relationship between the two variables. The correlation coefficient was calculated for each year across the six financial markets and is summarized in Figure 14.

The mean correlation coefficient for each market is shown above each box plot. The mean correlation coefficient for each market is near zero, indicating that almost no linear relationship exists between the realized volatility data and the EMVID index data. However, the year 2020 was an outlier for all markets. The 2014 year was also an outlier for Brazil and the US. This higher correlation could be due to the Ebola epidemic in 2014. In 2020, the correlation between the realized volatility and the EMVID index is greater than 0.60 for most of the markets. This indicates a moderately positive linear relationship. The EMVID index is only suitable as a predictor for realized volatility for the year 2020.[#]

4.2.2 | Realized volatility regression results

Figure 15 contains the results of the function-on-function regression model, for the COVID-19 pandemic period (2020), across the six markets. The blue curve depicts the actual realized volatility, and the red curve depicts the predicted realized volatility[||]. From Figure 15, it is clear that the EMVID index can be used as a predictor for the daily realized volatility of

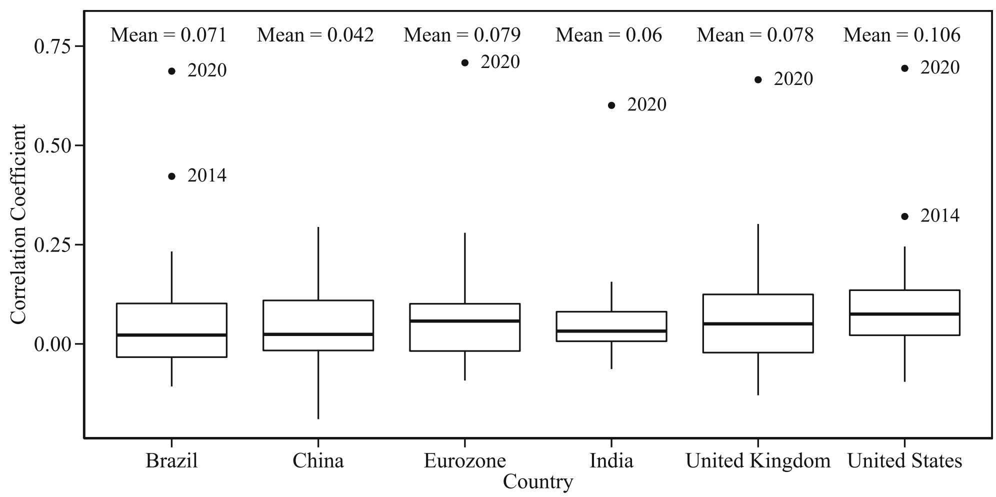


FIGURE 14 Box plots of the yearly correlation between each market's realised volatility and the EMVID index.

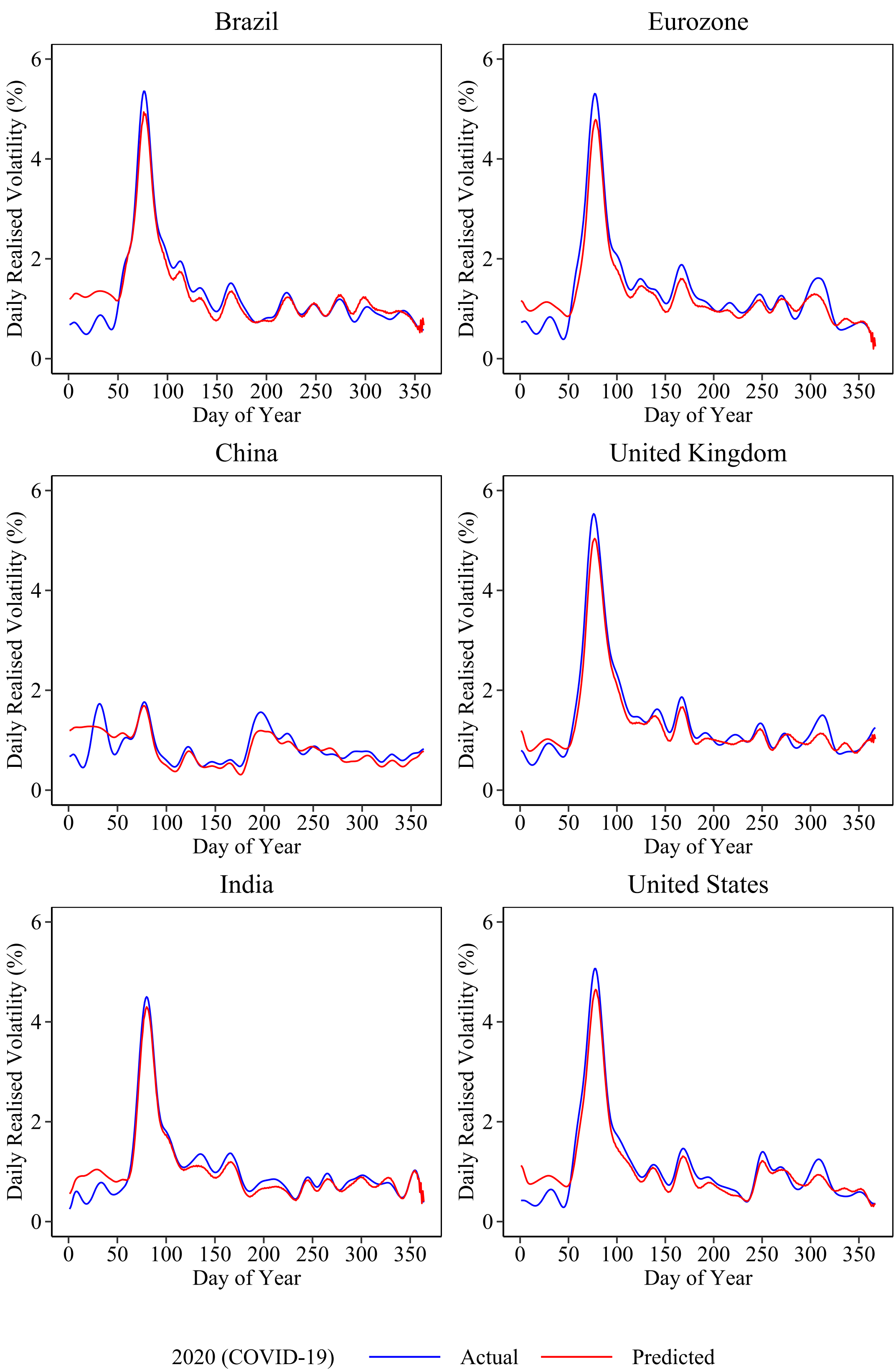


FIGURE 15 The actual (blue) and predicted (red) daily realised volatility for all six markets during the COVID-19 period.

international financial markets. The predicted curves (in red) are very similar to the actual curves (in blue). The results show a substantial spike in daily realized volatility, due to COVID-19, during March 2020. The Chinese market does not have a similar profile to the other markets but does show an increase in daily realized volatility as early as January.

To measure the fit of the regression model, the R-squared (R^2) metric was calculated for each year across the six markets. This quantitative measure represents the proportion of variance for the daily realized volatility explained by the EMVID index. Table 4 summarizes the R^2 metric for the six markets, with 2020 highlighted in red, as the focus is COVID-19. The regression model performs exceptionally well for the 2020 year since the R^2 value is close to one for all markets except China. In 2020, Brazil, India, the Eurozone, the UK and the US have R^2 values over 0.92. The EMVID index is a good predictor of realized volatility for this year. This is expected because of the COVID-19 pandemic. The predicted volatility for the Chinese market has an R^2 value of 0.600, which is still acceptable. All years, apart from 2020, have a much lower R^2 value, which indicates the EMVID index is a poor predictor of realized volatility for years with no severe health crises.

Although the R^2 value can quantify the variation explained by the predictor variable, it can be misleading when it comes to regression performance. This is evident by observing the moderate R^2 values for the 2008 year. There was no significant disease epidemic during this year, but the EMVID index can explain some variation in the realized volatility. By looking at the regression model output for 2008, in Figure 16, it is evident that the regression model is performing poorly

TABLE 4 A summary of the (R^2) measure, across all six markets, over the period of interest.

	Country					
	Emerging markets			**Developed markets**		
	Brazil	**China**	**India**	**The Eurozone**	**United Kingdom**	**United States**
Year	R^2					
2000	0.350	0.207	0.120	0.053	0.001	0.135
2001	0.193	0.008	0.015	0.254	0.298	0.237
2002	0.127	0.329	0.002	0.186	0.229	0.078
2003	0.478	0.066	0.001	0.013	0.001	0.013
2004	0.157	0.031	0.280	0.086	0.040	0.030
2005	0.396	0.028	0.159	0.197	0.103	0.142
2006	0.041	0.008	0.116	0.001	0.001	0.056
2007	0.490	0.485	0.172	0.065	0.086	0.053
2008	0.489	0.001	0.121	0.309	0.437	0.471
2009	0.389	0.001	0.033	0.017	0.002	0.008
2010	0.179	0.103	0.298	0.001	0.000	0.012
2011	0.271	0.026	0.017	0.166	0.268	0.183
2012	0.028	0.025	0.076	0.000	0.010	0.006
2013	0.085	0.130	0.002	0.002	0.002	0.010
2014	0.328	0.057	0.172	0.204	0.108	0.342
2015	0.208	0.006	0.002	0.052	0.015	0.004
2016	0.272	0.541	0.055	0.024	0.044	0.009
2017	0.231	0.037	0.013	0.003	0.026	0.011
2018	0.086	0.005	0.115	0.052	0.007	0.004
2019	0.002	0.111	0.064	0.052	0.119	0.075
2020	0.921	0.600	0.950	0.947	0.976	0.954
2021	0.365	0.474	0.431	0.005	0.158	0.229

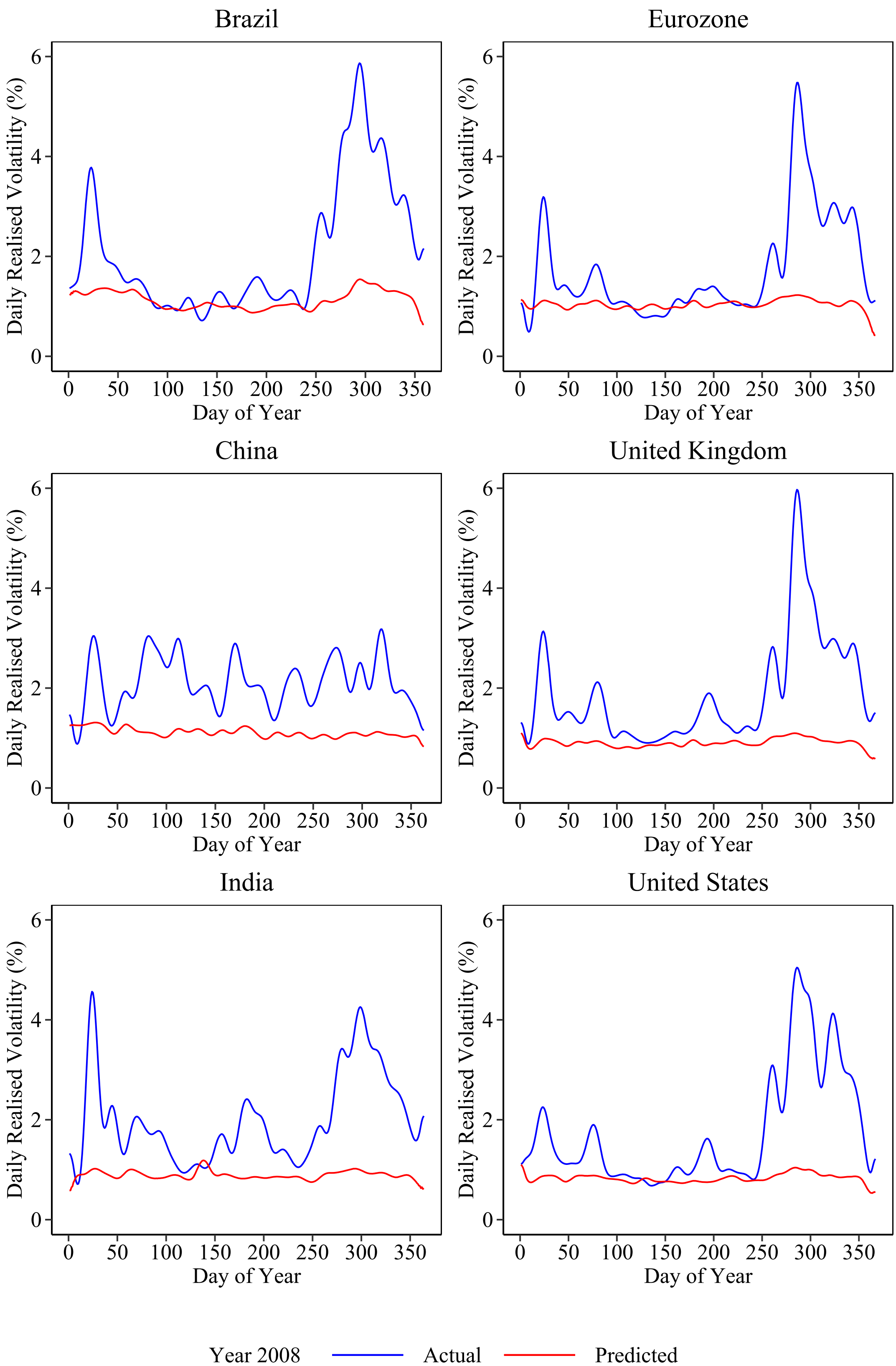


FIGURE 16 The actual (blue) and predicted (red) daily realised volatility for all six markets during the year 2008.

TABLE 5 A summary of the RMSE measure across all six markets over the period of interest.

	Country					
	Emerging markets			Developed Markets		
	Brazil	China	India	The Eurozone	United Kingdom	United States
Year	RMSE					
2000	0.462	0.606	0.358	0.338	0.380	0.466
2001	0.400	0.477	0.423	0.585	0.412	0.428
2002	0.371	0.396	0.395	1.078	0.745	0.677
2003	0.187	0.362	0.355	0.683	0.430	0.278
2004	0.433	0.268	0.771	0.336	0.347	0.249
2005	0.240	0.322	0.225	0.445	0.427	0.291
2006	0.335	0.444	0.733	0.395	0.340	0.331
2007	0.458	0.921	0.488	0.342	0.303	0.331
2008	1.514	1.133	1.403	1.190	1.412	1.415
2009	0.594	0.529	0.740	0.589	0.658	0.763
2010	0.337	0.282	0.245	0.395	0.340	0.392
2011	0.367	0.256	0.277	0.686	0.534	0.563
2012	0.307	0.326	0.225	0.295	0.278	0.229
2013	0.382	0.262	0.272	0.340	0.306	0.328
2014	0.303	0.547	0.282	0.301	0.334	0.328
2015	0.253	1.074	0.243	0.307	0.298	0.356
2016	0.287	0.531	0.287	0.453	0.426	0.398
2017	0.446	0.572	0.442	0.500	0.407	0.517
2018	0.318	0.361	0.347	0.379	0.310	0.406
2019	0.465	0.431	0.305	0.430	0.284	0.352
2020	0.270	0.216	0.177	0.260	0.206	0.228
2021	0.376	0.234	0.288	0.541	0.468	0.435

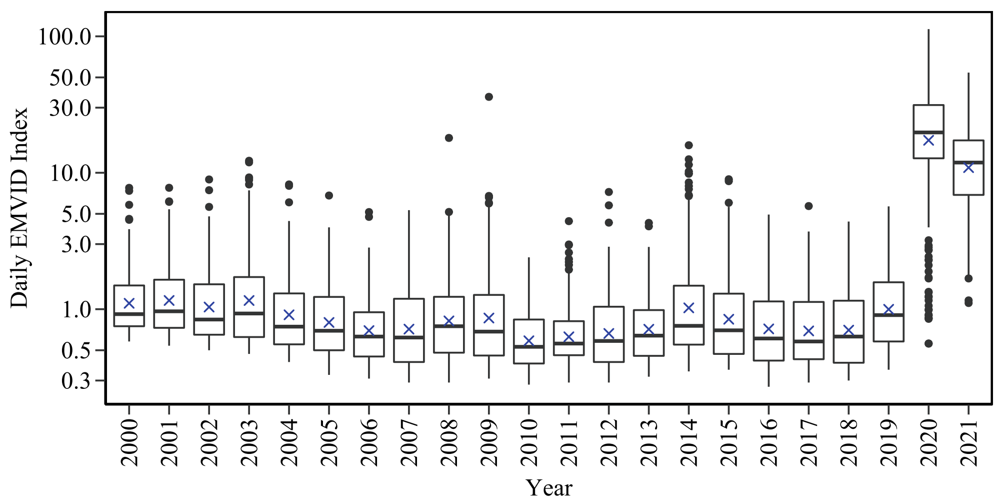


FIGURE 17 Box plots of the daily EMVID index from 2000 to 2021.

despite the R^2 value. Therefore, the root mean square error (RMSE) was calculated for each year across all six financial markets and is summarized in Table 5. This metric is a better measure of performance. The RMSE values for 2008 more accurately describe what is happening in Figure 16. The RMSE values for the COVID-19 period are lower than all the other years, meaning that the regression model predicts the realized volatility more accurately for this specific year. This is expected due to the nature of the predictor variable. In summary, the R^2 and RMSE metrics indicate that the EMVID index can be used as a predictor for the realized volatility of international financial markets, provided that the period of interest experienced a major health crisis.

Figure 17 is a series of box plots of the daily newspaper-based equity market volatility due to infectious diseases (EMVID) index for each year during the 2000–2021 time period. The 2020 and 2021 years had significantly higher mean EMVID index values than the other years in this period. The trend seen in Figure 17 corroborates the trend seen in Figure 9. Additionally, it suggests that the effects of the COVID-19 pandemic potentially spilled over into the following year.

5 | KEY FINDINGS AND CONCLUDING REMARKS

Having analyzed the realized volatility of developed and emerging stock markets using FDA, with comparisons across the global financial crisis and the COVID-19 episodes, first, we aim to provide a detailed summary of our econometric findings before discussing the implications of our results, followed by some concluding remarks of this research.

Insights from the general derivative exercise resulted in the following findings: *First* being that realized volatility can successfully be modeled using functional data analysis (FDA) techniques, specifically by using a B-spline basis system and the roughness penalty approach; *Second*, that the COVID-19 pandemic was a significant contributor to the instability of the financial markets in 2020; *Third*, it is clear from the derivative curves that the velocity of the realized volatility, or risk, did not fluctuate excessively in the years before the COVID-19 pandemic. However, the rate of risk grew rapidly at the beginning of March 2020, when COVID-19 was officially declared a global pandemic by the World Health Organization. Post-COVID-19, the change in realized volatility remained relatively stable across all six markets. Only Brazil, the Eurozone, and the United Kingdom experienced minor fluctuations during 2021. *Fourth*, significant effect on the market volatilities was seen for Brazil, India, the Eurozone, the United Kingdom, and the United States. Thus, the similar market volatility profiles across most of the markets, with China being the exception, indicate that the emerging and developed markets were affected by the COVID-19 pandemic on a similar scale. *Fifth*, the recovery time of the market volatilities was significantly quicker during the COVID-19 pandemic than that observed during the Global Financial Crisis (GFC), and the rate of change of uncertainty returned to pre-COVID levels in May 2020. Thus, the uncertainty due to COVID-19 only persisted from March to May 2020 and then proceeded to stabilize. This is different from the GFC, whence the fluctuations in realized volatility began as early as 2007 and persisted until early 2009. This finding held across emerging and developed markets. The *sixth* finding of this research is that the phase-plane plot exercise revealed that the initial shock on the markets due to COVID-19 was substantial and greater than the initial shock observed due to the 2008 GFC, which while spread more quickly across the financial markets, did not persist for a prolonged period. The smoothing trial exercise produced the *seventh* finding of this research, which is methodology specific. De Boor[25] and Ramsay and Silverman[18] both suggest that a fourth-order B-spline ($m = 4$) with knots at each argument smooths the data with the best accuracy. The smoothing exercise found this claim accurate but only at smoothing parameter values below $\lambda = 10^{-1}$. The difference in accuracy is marginal when using higher B-spline orders if a smoothing parameter greater than $\lambda = 10^{-1}$ is applied.

Finally, to have an explicit understanding of the relationship between the uncertainty in financial markets and the COVID-19 pandemic, the daily newspaper-based infectious disease (EMVID) index was used in a regression exercise. A functional regression model was developed to predict the realized volatility of international financial markets using the EMVID index as a predictor. This analysis produced the *eighth* finding which was that while EMVID index is an accurate predictor of the realized volatility of financial markets, it is only in the years that markets experience a severe health crisis, as 2020 did with the COVID-19 pandemic. An interpretation of these key findings is discussed in the section below.

The COVID-19 pandemic significantly contributed to the financial markets' instability in 2020 and introduced a large amount of uncertainty and fear among investors, with investors fleeing to low-risk safe haven assets.[26] This uncertainty galvanized the freefall of most of the international financial markets. Governments' policy responses to curb the spread of the virus also had huge implications on the markets.[3,27] During 2020, the risk started to evolve and increase uncertainty in the markets, which caused dramatic fluctuations in realized volatility. In 2021, the post-COVID-19 year, the market volatilities were relatively stable because the financial markets had recovered from the shock of COVID-19. The minor fluctuations in realized volatility could be due to several factors. However, a possible explanation is that small outbreaks of COVID-19 occurred. Once again, increasing uncertainty among investors.

The emerging and developed markets, except for China, were affected by the COVID-19 pandemic on a similar scale. Emerging markets tend to follow developed markets, and as the developed markets were experiencing large increases in realized volatility it spilt over to the emerging markets. The Chinese financial market was the exception, it did experience an increase in uncertainty over this period, but it was not on the same scale as the rest. This is surprising since the COVID-19 pandemic began in China. This could be attributed to the fact that while the number of daily COVID-19 cases in China significantly affected the Chinese stock market, China implemented strict lockdown measures early on, and the daily infection rate was kept relatively low,[28] which may have lessened the uncertainty among investors, and could be why China was not similarly affected.

The recovery time of the market volatilities was significantly quicker during the COVID-19 pandemic than that observed during the GFC, and may be attributed, but not limited to, financial policies, technology stocks, the nature of the crisis and the dynamic adaptability of the stock market. The GFC was an event that exposed the lack of adequate regulatory policy,[29] due to which many vital lessons were learnt during the GFC that could be applied to the COVID-19 financial crisis. The central banks discovered that financial policies such as lowering interest rates, quantitative easing, capital injections into financial institutions and increased government spending (fiscal policy)[30] have a positive impact during a financial crisis, and that response to a global crisis must be swift, systematic, and decisive.[31] Therefore, when the COVID-19 pandemic appeared, the central banks were better prepared and knew what to do. Once the COVID-19 crisis was declared a pandemic, many countries acted swiftly and introduced economic rescue measures. For example, in May 2020, the US Federal Reserve reduced interest rates to zero and brought back quantitative easing, which positively affected the stock market.[28] This was seen in the results, as the velocity of realized volatility started to return to pre-COVID levels at the end of May 2020. Due to the lockdown measures put in place by governments, society relied heavily on technology for remote work and entertainment. The technology stocks saw great returns during the COVID-19 period and may have contributed to this fast recovery of the markets.

When discussing the recovery of financial markets, a fact that cannot be ignored is the nature of the crisis that caused the instability. The COVID-19 pandemic was a health and humanitarian crisis that brought about a financial crisis due to the reduced investment in risky assets. The entire global population was affected by the pandemic. However, the rich had better resources to survive the shock than the poor. Such rich are primarily the people who invest in the financial market itself. Conversely, the origin of the GFC was the financial market. Thus, during the GFC the investors and the rich were directly affected, which is why the recovery took much longer, along with the lack of financial policies, and created what is known as the "Great Recession."

Another potential reason for the relatively short-term impact of COVID-19 on the volatilities of international stock markets is the "dynamic adaptability" of the stock markets.[28] This is a situation whereby a new epidemic arises, and the market reacts strongly. However, as time passes, the market adapts to the epidemic. The phase-plane analysis supports this since the potential and kinetic energies during COVID-19 were higher than during the GFC, but the duration of the crisis was shorter. This suggests that the markets during COVID-19 experienced an initial but substantial shock, which spread quickly across markets, but the shock did not last, and the recovery was swift.

We found that the EMVID index can be used as a predictor for the realized volatility only if the markets experience a severe health crisis, as 2020 did with the COVID-19 pandemic. It is obvious that only the years that contain infectious disease outbreaks monitored by this index will produce satisfactory results. Naturally, the regression model performed poorly for the additional years in the period of interest, which did not contain any epidemics or pandemics. However, this was expected due to the inherent nature of the predictor variable. The R^2 and RMSE metrics act as a quantitative measure for the relationship between infectious diseases and financial market volatilities. Additionally, the regression analysis acts as a robustness check and confirms that the COVID-19 virus increased uncertainty, significantly affecting international market volatility.

This research aimed to investigate the realized volatility of international financial markets with a focus on COVID-19. Various FDA techniques were utilized throughout this research, and in this specific application, can be utilized as an alternative methodology to the conventional time series approach. On an academic level, the research contributes to the knowledge of basis systems, smoothing methods, derivative analysis and functional regression contributed to the growing theory on FDA techniques and literature. Furthermore, a significant contribution of this research is the extraction of the potential and kinetic energy from the phase-plane plot analysis, which can be applied to other applications and not limited to realized volatility. The application part of the study was completed using real-world volatility data. Since, volatility is an indicator of financial risk,[2] the findings in this study can assist individual investors, policymakers and financial institutions in making informed decisions when confronted with a financial crises due to an infectious disease. As part of future work, given that stock market volatility is known to be affected by large number of macroeconomic and financial variables, as part of future research, it would interesting to extend the financial regression analysis presented here with one predictor to multiple predictors using machine learning techniques, such as random forests, as detailed in Reference 32.

ACKNOWLEDGMENTS

We would like to thank an anonymous referee for many helpful comments. However, any remaining errors are solely ours.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Oxford-Man Institute of Quantitative Finance: Realized Libra at https://realized.oxford-man.ox.ac.uk/data. These data were derived from the following resources available in the public domain: (EMVID) tracker data was obtained from the Economic Policy U, http://policyuncertainty.com/infectious_EMV.html.

ENDNOTES

*Based on the suggestions of an anonymous referee, a Fourier-based smoothing was also performed, however, band results did not change. Details of which can be provided on reasonable request from the authors.

†An algorithm was developed that incorporates specific logic to deal with these missing values by excluding these arguments from the smoothing process. This algorithm can be made available upon reasonable request to the authors.

‡This algorithm can be made available upon reasonable request to the authors.

§The first and second derivative plots for realized volatility around the COVID-19 period (2018-2021), for all the six countries, are available as supplementary material to this article.

¶The first and second derivative plots for realized volatility, around the GFC period (2006–2009), for all the six countries are available as Supplementary Material to this article.

#Based on the suggestions of an anonymous referee, when we controlled for new COVID-19 cases, following,[24] as a possible alternative metric of capturing "real" uncertainty, we found that, in general, the predictive content in the EMVID index on its own remained robust, to the extent that both these measures capture complementary information. Complete details of the functional regression analyses conducted in this regard is available upon request from the authors. But it must be realized that, the EMVID index is able to capture uncertainty in financial markets beyond the COVID-19, by considering news regarding other infectious diseases as well, such as, MERS, SARS, Ebola, H5N1, or H1N1.

‖The regression results for the additional years are available as supplementary material corresponding to this article.

ORCID

Sonali Das https://orcid.org/0000-0002-6350-4279

REFERENCES

1. Díaz F, Henríquez PA, Winkelried D. Stock market volatility and the COVID-19 reproductive number. *Research in International Business and Finance*. Vol 59. Elsevier; 2022:101517. https://www.sciencedirect.com/science/article/pii/S0275531921001380
2. Poon S-H, Granger CWJ. Forecasting volatility in financial markets: A review. *J Econ Lit*. 2003;41(2):478-539. http://www.jstor.org/stable/3216966
3. Zaremba A, Kizys R, Aharon DY, Demir E. Infected markets: Novel coronavirus, government interventions, and stock return volatility around the globe. *Financ Res Lett*. 2020;35:101597. https://www.sciencedirect.com/science/article/pii/S1544612320306310
4. Baek S, Mohanty SK, Glambosky M. COVID-19 and stock market volatility: An industry level analysis. *Financ Res Lett*. 2020;37:101748. doi:10.1016/j.frl.2020.101748
5. McAleer M, Medeiros MC. Realized volatility: A review. *Econ Rev*. 2008;27(1-3):10-45. doi:10.1080/07474930701853509
6. Andersen T, Bollerslev T, Christoffersen P, Diebold F. Chapter 15 volatility and correlation forecasting. *Handbook of Economic Forecasting*. Vol 1. Elsevier; 2006:777-878.
7. Bollerslev T, Diebold F, Andersen T, Labys P. Modeling and forecasting realized volatility. *Econometrica*. 2003;71:579-625. doi:10.2139/ssrn.267792
8. Meddahi N. A theoretical comparison between integrated and realized volatility. *J Appl Econ*. 2002;17(5):479-508. http://www.jstor.org/stable/4129268
9. Barndorff-Nielsen OE, Shephard N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J R Stat Soc B (Stat Methodol)*. 2002;64(2):253-280. http://www.jstor.org/stable/3088799
10. Mangisa S, Das S, Ray S, Sharp G. Functional regression models for South African economic indicators: A growth curve perspective. *OPEC Energy Rev*. 2019;43:3. doi:10.1111/opec.12148
11. Ullah S, Finch C. Applications of functional data analysis: A systematic review. *BMC Med Res Methodol*. 2013;13:43. doi:10.1186/1471-2288-13-43
12. Muller H-G. *Functional Data Analysis*. Technical report. Department of Statistics University of California, Davis; 2011.
13. Wang J-L, Chiou J-M, Müller H-G. Functional data analysis. *Annu Rev Stat Appl*. 2016;3(1):257-295.

14. Das S, Demirer R, Gupta R, Mangisa S. The effect of global crises on stock market correlations: Evidence from scalar regressions via functional data analysis. *Struct Chang Econ Dyn*. 2019;50:132-147. doi:10.1016/j.strueco.2019.05.007
15. Heber G, Asger L, Shephard N, Sheppard K. *Oxford-Man Institute's Realized Library*. Technical report. Oxford-Man Institute, University of Oxford; 2009 Library Version: 0.3.
16. Shephard N, Sheppard K. Realising the future: Forecasting with high-frequency-based volatility (heavy) models. *J Appl Econ*. 2010;25(2):197-231. http://www.jstor.org/stable/40607017
17. Baker SR, Bloom N, Davis SJ, Terry SJ. *Covid-Induced Economic Uncertainty*. Working Paper 26983, National Bureau of Economic Research; 2020 http://www.nber.org/papers/w26983
18. Ramsay J, Silverman B. *Functional Data Analysis*. 2nd ed. Springer; 2005.
19. Kokoszka P, Reimherr M. *Introduction to Functional Data Analysis*. 1st ed. Chapman and Hall/CRC; 2018.
20. Ramsay J, Silverman B. *Applied Functional Data Analysis: Methods and Case Studies*. 1st ed. Springer; 2002.
21. Ramsay JO. Spencer Graves, and Giles Hooker. Package 'fda' (Functional Data Analysis). 2022. Version 5.5.1 https://cran.r-project.org/web/packages/fda/fda.pdf
22. Ramsay J, Hooker G, Graves S. *Functional Data Analysis with R and MATLAB*. 1st ed. Springer; 2009.
23. Baker SR, Bloom N, Davis SJ, Kost K, Sammon M, Viratyosin T. The Unprecedented Stock Market Reaction to COVID-19. *Rev Asset Pricing Stud*. 2020;10(4):742-758. doi:10.1093/rapstu/raaa008
24. Agosto A, Giudici P. A Poisson autoregressive model to understand Covid-19 contagion dynamics. *Risks*. 2020;8(3):77-85. doi:10.3390/risks8030077
25. De Boor C. *A Practical Guide to Splines*. 27th ed. Springer; 2001.
26. Corbet S, Hou Y(G), Hu Y, Oxley L, Xu D. Pandemic-related financial market volatility spillovers: Evidence from the Chinese COVID-19 epicentre. *Int Rev Econ Financ*. 2021;71:55-81. https://www.sciencedirect.com/science/article/pii/S1059056020301350
27. David SA, Inácio CMC Jr, Machado JAT. The recovery of global stock markets indices after impacts due to pandemics. *Res Int Bus Financ*. 2021;55:101335. https://www.sciencedirect.com/science/article/pii/S0275531920309429
28. Gao X, Ren Y, Umar M. To what extent does Covid-19 drive stock market volatility? A comparison between the U.S. and China. *Econom Res-Ekonomska Istraživanja*. 2021;35:1-21. doi:10.1080/1331677X.2021.1906730
29. Claessens S, Dell'Ariccia G, Igan D, Laeven L. Cross-country experiences and policy implications from the global financial crisis. *Econ Policy*. 2010;25(62):267-293.
30. The Reserve Bank of Australia. The global financial crisis. 2022 https://www.rba.gov.au/education/resources/explainers/pdf/the-global-financial-crisis.pdf?v=2022-09-03-11-58-25
31. Lin JY. The impact of the financial crisis on developing countries. *Macroeconomics and Economic Growth: Economic Conditions and Volatility*. World Bank; 2008. doi:10.1596/26129
32. Gertheiss J, Rügamer D, Liew B, Greven S. Functional data analysis: An introduction and recent developments. 2024.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Shackleton RL, Das S, Gupta R. Comparing risk profiles of international stock markets as functional data: COVID-19 versus the global financial crisis. *Appl Stochastic Models Bus Ind*. 2024;40(4):1153-1181. doi: 10.1002/asmb.2879