Shackleton, Ryan; Das, Sonali; Gupta, Rangan

Book

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

University of Pretoria
Department of Economics Working Paper Series

Comparing Risk Profiles of International Stock Markets as Functional Data: COVID-19 versus the Global Financial Crisis

Ryan Shackleton
University of Pretoria
Sonali Das
University of Pretoria
Rangan Gupta
University of Pretoria

Working Paper: 2023-28
September 2023

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

Comparing Risk Profiles of International Stock Markets as Functional Data: COVID-19 versus the Global Financial Crisis

Ryan Shackleton[a], Sonali Das[b,*], Rangan Gupta[c]

[a]Department of Information Technology, University of Pretoria, Pretoria, South Africa

[b]Department of Business Management, University of Pretoria, Pretoria, South Africa

[c]Department of Economics, University of Pretoria, Pretoria, South Africa

[*] Corresponding author email: sonali.das@up.ac.za

Abstract

In this paper, we aim to provide a detailed econometric analysis of the realised 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 realised 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 realised 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 modelled by this infectious disease index, but only for periods experiencing an epidemic or pandemic.

JEL Codes: C22; C51; C52; C53; G15.

Keywords: Realised volatility; International stock markets; Functional data analysis.

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 to numerous countries worldwide, and was declared a global pandemic by March 2020. Due to the infectiousness and severity of COVID-19, governments enforced sudden and severe restriction measures on travel, social gatherings, and commercial and economic activities, resulting in unprecedented uncertainty within international financial markets [Díaz et al., 2022]. In the aftermath of the pandemic, it

has led to 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. It is a measure that 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 [Poon and Granger, 2003]. 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 [Zaremba et al., 2020].

There is a crucial connection between financial market uncertainty and investor confidence. Therefore, volatility acting as an indicator of financial risk [Poon and Granger, 2003] and as a result, it can aid governments, individual investors, portfolio fund managers, policymakers and financial institutions in making their respective decisions [Baek et al., 2020]. 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 [McAleer and Medeiros, 2008; Poon and Granger, 2003].

While volatility is a key indicator of the vulnerability of financial markets, developing models or forecasts using volatility presents an inherent problem [McAleer and Medeiros, 2008], because by its very nature, volatility is latent and, therefore, is not directly observable [Andersen et al., 2006; McAleer and Medeiros, 2008]. 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 [McAleer and Medeiros, 2008]. This presents an opportune to utilise the concept of *realised volatility*, which is the sum of squared intra-day returns at specific intervals [Bollerslev et al., 2003; Poon and Granger, 2003]. Typically, the sampling time for intra-day returns is five or ten minutes (sparse sampling), but it has been reported to be optimal in the range of thirty to sixty-five minutes [McAleer and Medeiros, 2008]. The smaller the sampling time (higher frequency sampling) results in higher accuracy and increased microstructure noise, presenting a useful trade-off [McAleer and Medeiros, 2008; Meddahi, 2002].

The realised 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 [Barndorff-Nielsen and Shephard, 2002]. Thus, the estimated realised volatility error, which is a function of the actual volatility, is zero. From the property mentioned above, it is evident that realised volatility is an unbiased indicator for actual volatility [Bollerslev et al., 2003; Barndorff-Nielsen and Shephard, 2002] and provides a consistent estimate of price fluctuations over a specific time [Andersen et al., 2006]. Generalised auto-regressive conditional heteroscedastic (GARCH) models are commonly utilised 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. Volatility is a latent variable that follows a stochastic process in this situation. In the literature, the GARCH and SV models are often utilised. However, these models suffer because the underlying volatility estimate is neither model free nor unconditional. The realised volatility has the advantage of being unconditional. Ultimately, by utilising realised volatility as a measure, the volatility becomes "observable" and can be modelled directly [McAleer and Medeiros, 2008]. Nonetheless, the realised volatility is still not free of the ever-present measurement errors [Bollerslev et al., 2003]. However, realised volatility resolves the latent variable issue and provides investors with an observable metric of uncertainty in the market. We use realised 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 realised volatility differ between the developed and emerging markets during the COVID-19 pandemic?
- Can the daily newspaper-based infectious disease index predict the realised volatility of international financial markets?

A statistical framework is required to analyse realised volatility, and we present a strong argument for using Functional data analysis (FDA) techniques in this endevour to analyse the role of uncertainty in international financial markets due to the COVID-19 pandemic, as well as predict the realised volatility using the infectious disease index through FDA techniques [Mangisa et al., 2019]. The realised volatility, the sum of squared intra-day returns, has been condensed into a single daily value. This daily value has been 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 analysing stock market data is time series, however, it suffers from many mathematical constraints. FDA offers an alternative approach to analysing time series data.

Past literature indicates that FDA has increased in popularity and is being used to better analyse time series data [Ullah and Finch, 2013]. 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 [Muller, 2011]. This applies to the realised volatility data set. Although a realised volatility value is computed daily, most major stock markets close trading on weekends and some public holidays. Multiple international financial markets were analysed 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 [Mangisa et al., 2019; Ullah and Finch, 2013].

Due to technological advances, a higher density of sampling can be done over a shorter time interval, generating discrete time series data sets of considerable size. Classical time series approaches often treat such discrete time series data as multivariate data [Ullah and Finch, 2013]. The classical time series multivariate approach fails to capture the functional behaviour that occurs when generating this type of discrete data [Ullah and Finch, 2013]. FDA is superior in this scenario because this approach handles the whole function as one single entity [Ullah and Finch, 2013; Mangisa et al., 2019], which results in no cause for concern about correlations between repeated measurements. In reality, it is common for measurement errors to be present in the data collection process. Wang et al. [2016], on the other hand, suggests that any measurement errors that occur throughout the data collection process, can be characterised 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 [Wang et al., 2016].

The functions that arise from FDA methods are assumed to be differentiable to some order [Mangisa et al., 2019]. 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 [Ullah and Finch, 2013]. Rate of change analysis, using phase-plane plots, can result in extremely useful insights relating to the financial markets. FDA allows the user the opportunity to conduct rate of change analysis in a more elegant fashion compared to classic time series methods. Furthermore, many traditional statistical tools rely on the property of stationarity being present within time series data. Stationarity is the concept whereby the process of generating a time series does not change over time. However, in most cases, time series data does violate this assumption [Mangisa et al., 2019; Das et al., 2019]. Financial data, such as the realised volatility of international markets, would require transformations to ensure mean-reversion if traditional statistical methods were applied [Das et al., 2019]. Conversely, Muller [2011] noted that, when using FDA, the underlying process does not have to be stationary.

Thus, it is clear FDA is a tool that overcomes many limitations associated with traditional time series statistical methods. It is moot to mention here that FDA is a complementary tool, and by no means replaces the traditional time series methods, but rather extends traditional statistical techniques by creating a function without extracting statistical indexes from the stock market data, which avoids critical information loss [Muller, 2011]. Due to the reasons mentioned above and because it is a tool not widely used, the FDA approach was utilised and the authors hope that this application will shed new light and lead to interesting insights contined in the results.

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. This claim was found to be accurate, however, only at low smoothing parameter values. For higher smoothing parameter value, less was the smoothing accuracy which was affected when utilising different B-spline orders. Second, a significant contribution of this research is in the extraction of the potential and kinetic energy from phase-plane plot analysis. To the authors' knowledge, a function to extract these quantities has not been attempted before and was not available at the time article was written. This extraction of energies from phase-plane plots can be applied to other applications as a visualisation tool, and thus is not limited to realised volatility as was applied in this work.

The application part of this research involved applying FDA techniques to realised 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 realised volatility of international financial markets. The review presented here contributes to the growing literature on the GFC, COVID-19 pandemic and realised volatility. The two data sets used provided an opportunity to demonstrate the FDA techniques mentioned, and to gain insights into the behaviour of realised 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 realised 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 paper is organised 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 utilised in this research. The first is a data set that contains realised 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.

2.1 Data

2.1.1 Realised Volatility Data Set

Realised volatility data were obtained from the Oxford-Man Institute of Quantitative Finance: Realised Library, created by Heber et al. [2009] 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 realised 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 [Heber et al., 2009]. The econometric measures available within the data set have been computed using high-frequency daily financial returns, with techniques summarised in Shephard and Sheppard [2010].

At the time of writing, this data set was last updated on 28 June 2022 and contained daily financial estimators for 31 stock market indices. It is in the comma-separated values (CSV) format with several econometric measures, and all are represented as columns in the file. Appendix A Table 6 contains a summary of all the econometric measures available. Table 6 has a row highlighted in bold, which is of particular importance. The *rv10* column contains the daily realised variance at a sampling frequency of 10 minutes. The daily realised 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 t-th day. In this case, the time-frequency is 10 minutes. Therefore, the *rv10* column contains values computed by summing the squared returns of a specific index every 10 minutes for each trading day. The high-frequency return data, used to compute the realised 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 [Heber et al., 2009]. The reason for choosing the realised variance with 10 minutes sampling frequency (*rv10*) instead of the 5-minute sampling frequency (*rv5*) is because the lower sampling frequency alleviates more noise [Andersen et al., 2006]. 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 [Andersen et al., 2006].

Referring back to Table 6, the *rv5_ss* and *rv10_ss* measures are identical to *rv5* and *rv10* measures, respectively. The data set contains the econometric measures summarised in Table 6 for 31 market indices over a specific range of time. These market indices and time intervals are summarised in Appendix A Table 7. 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). The rows in bold, in Table 7, are the leading market indices that represent those respective areas. 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 28 June 2022. However, the date range of interest for this research is between 04 January 2000 (ensuring the earliest available date for all respective markets to be the same) and 31 December 2021. A total time period of 22 years. The data set required moderate pre-processing, summarised in Section 2.2.1.

2.1.2 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. [2020b] with daily measures ranging from 01 January 1985 to date and is being updated daily. It is in CSV format with four columns as summarised in Appendix A Table 9. All four of these columns were utilised in this research. The date range of interest for this data set is between 04 January 2000 and 31 December 2021. This is 22 years worth of daily data, and the time period is consistent with the realised volatility data set.

The daily EMVID measure is computed using four sets of terms[1], as follows: E for {"economic, economy, financial"}; M for {"equity, equities, Standard and Poors, stock market"}; V for {"risk, risky, uncertain, uncertainty, volatility, volatile"}; and ID for {"coronavirus, disease, epidemic, ebola, flu, H5N1, H1N1, MERS, pandemic, SARS, virus"}.

Daily counts of newspaper articles that contain at least one of these terms within the four sets are obtained from roughly 3000 US newspapers [Baker et al., 2020b]. These daily counts are scaled by the total number of articles within the same day. Lastly, this series is multiplicatively re-scaled to match the level of the VIX (Chicago Board Options Exchange's CBOE Volatility Index) between 1990-2016. See Baker et al. [2019], for the creation of the original EMV index, without the infectious diseases term. This re-scaling is done by finding the ratio of EMVID to EMV articles and scaling the EMVID index accordingly [Baker et al., 2020b].

2.2 Transformations

Here, we outline the steps undertaken to prepare the data for the implementation of FDA on our dataset.

[1]These four sets of terms were taken verbatim from `http://policyuncertainty.com/infectious_EMV.html` and the author does not claim it as his own.

2.2.1 Data Pre-processing

The data pre-processing step involved aligning the realised volatility data. The underlying functions were over a year, and hence the smoothing was over a domain of a year length. Therefore, since the date range spans 22 years, the daily realised volatility data was grouped according to years. The average number of trading days per year was 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. This presents a problem when comparing different years within the same financial market. As the trading days shift, a realised volatility measure for a specific day may not exist within that specific year. Furthermore, to smooth discrete data and create multiple yearly curves or functions, all the years must have the same number of observations.

To solve this issue, grouping the data into years and aligning the dates was necessary. Algorithm 1 describes the function created to solve this problem. This algorithm groups the daily realised volatility values into their respective years and aligns the dates. When the algorithm encounters a day within a specific year without a realised volatility value, it assigns this day a missing (*NA*) value. This results in a data frame with a day-month column and yearly columns ranging from 2000 - 2021 containing the daily realised volatility. These yearly columns all contain an equal number of observations (days within the year). This data alignment was applied across all six market-specific data sets.

There were two pre-processing steps applied to the infectious disease data set, as follows: (i) **Extracting Date Range of Interest**: The raw EMVID data set contains data from 01 January 1985 to the present. However, the range of interest is between 04 January 2000 - 31 December 2021. For context, the first ten entries of the raw data set can be observed in Appendix A Table 10 and is an example of the structure of the raw data. The raw EMVID index data set was imported to R and saved as a data frame type. A new data frame was created, with the data between the specified date range. This new data frame also needed to undergo data alignment, as was the case with the realised volatility data, as the data needs to be in the same yearly column structure; (ii) **Data Alignment**: The final 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 realised volatility data set, a minor adaption of that alignment function was done. Algorithm 2 describes the function created to align the EMVID index data set. The result of this algorithm is a data frame, with a day-month column and yearly columns ranging from 2000 - 2021 containing the daily infectious EMVID index measure. This alignment was only applied once as the data set is utilised across all six markets.

2.2.2 The Smoothing Process

The next necessary step in FDA analysis 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 over the Fourier system for many reasons, however, the key reason was due to the periodicity of the underlying data sets. It was assumed that the realised volatility and EMVID index underlying data were not periodic. Considering the fact that B-splines accommodate non-periodic data, whereas the Fourier does not, it was a simple

choice. Additionally, B-splines are known to be superior computationally and have greater flexibility compared to other basis systems [Ramsay and Silverman, 2005]. 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 [Ramsay and Silverman, 2005] 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. According to Ramsay and Silverman [2005] the roughness penalty method is superior. This superiority is apparent since the least squares estimation method does not support derivative analysis and has little control over the amount of smoothing applied. The roughness penalty method avoids these limitations and often produces better results regarding smoothing and derivative estimation [Ramsay and Silverman, 2005]. Nonetheless, the most important parameter concerning the roughness penalty method is the smoothing parameter λ. It controls the trade-off between fitting the data and the variability of the function (fitting versus smoothing) [Ramsay and Silverman, 2002, 2005]. The smoothing parameter λ must be chosen such that penalised smoothing results in a good balance between overfitting and over-smoothing, which is a complex task [Kokoszka and Reimherr, 2018]. It is common to choose this parameter through visual inspection, however, the generalised 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 [Ramsay and Silverman, 2005].

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 optimised smoothing as well as providing a foundation for multiple types of FDA techniques. This flexibility and reusability are vital when multiple analyses, such as derivative analysis and functional regression, are carried out on multiple data sets. This point is an elegant segue to Sub-subsection 2.2.3, which deals with the practical implementation of the dynamic smoothing algorithm.

2.2.3 The Dynamic Smoothing Algorithm

The *fda* package in R was used extensively throughout this research Ramsay et al. [2022]. The goal of the DS algorithm was to ensure flexibility, reusability, repeatability, and the opportunity to optimise smoothing with regards to the smoothing parameter λ. Therefore, this algorithm was built to take three parameter arguments, namely, the data to be smoothed, the order of the B-spline polynomials m and the smoothing parameter, λ.

Some key functions used to develop the dynamic smoothing algorithm are summarised in Table 1. For further documentation on each function, such as function arguments and outputs, see Ramsay et al. [2022].

– Please include Table 1 about here. –

Three parameters determine the DS namely the knot sequence τ (this will infer the number of interior knots L), the order m of the polynomials, and the value for the smoothing parameter λ. Algorithm 3 describes the dynamic B-spline smoothing function developed. The algorithm takes three arguments, the data to be smoothed, the order m of the polynomials and the smoothing parameter λ. Thus, 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 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 [Ramsay et al., 2009]. 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 analysis is a key contribution of this research. Thus, the smoothing algorithm had to accommodate the calculation of the derivatives of the smooth functions. The roughness penalty method involves penalising the curvature of the highest-order derivative required in the analysis. 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. Algorithm 3 incorporates specific logic to deal with these missing values by excluding these arguments from the smoothing process.

We summarises the practical methods utilised to conduct derivative (rate of change) analyses on the realised volatility data set. Smooth curves of the realised volatility were generated using the dynamic smoothing algorithm developed in Sub-subsection 2.2.2. However, for the various analyses, the order m of the B-spline polynomials and the value of the smoothing parameter λ needed to be chosen. A smooth curve implies that derivatives exist and can be computed within the specified time interval of the observations [Mangisa et al., 2019]. The first derivative (velocity) and second derivative (acceleration) of these smooth curves are of particular interest as they are indicative of the rates of change. It has been mentioned before that conducting a derivative analysis requires smooth functions to have derivatives that exist and are smooth. This is important, as the choice of order of the polynomials m when smoothing the data will determine if the derivatives are smooth.

To ensure the highest order derivative n to be analysed is smooth, the derivative of order $n + 2$ must be penalised in the roughness penalty approach [Ramsay and Silverman, 2005]. In this case, the highest order of derivative of interest is the second derivative. Thus, the original function's fourth derivative ($2 + 2 = 4$) must be penalised 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 penalised. Therefore, for all derivative analyses conducted in this research,

an order six ($m = 6$) B-spline system was used in the smoothing process. In the case of the roughness penalty approach, the smoothing parameter λ controls the smoothness of the function. The trade-off between smoothness and fit is controlled by the smoothing parameter λ. The challenge lies in finding an optimal smoothing parameter that suits the data. For the case of the general estimation curves, the choice of smoothing parameter λ was made using the GCV criterion. The GCV criterion yields a smoothing parameter λ that is exceptionally close to the optimal value [Ramsay and Silverman, 2005].

– Please include Figure 1 about here. –

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 realised 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.

The realised volatility data was smoothed using the DS algorithm (Algorithm 3), which returns a smooth functional data object. The *fda* package contains the function `deriv.fd`, which takes a smooth functional data object as an argument as well as the order of derivative required, and returns the derivative of the functional data object.

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 [Ramsay and Silverman, 2005]. 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.

3.1.1.1 Choice of Smoothing Parameters

In theory, the smoothing parameters for the phase-plane plot analysis could be identical to the general derivative estimation parameters. This does hold true for the order of the B-spline basis but not for the value of the smoothing parameter. The B-spline basis order remains at six ($m = 6$) due to the reasons mentioned earlier. Conversely, a smoothing parameter of 10^4 produced indecipherable phase-plane plots due to under-smoothing. The data is not smooth "enough", so the derivatives vary in extremes, producing an indecipherable phase-plane plot. Theory suggests that, in the phase-plane plot case, instead, make the final choice of smoothing parameter λ by visual inspection [Ramsay and Silverman, 2005]. The most common technique employed when generating phase-plane plots is to start with a small value for the smoothing

parameter λ and increase it until the plots become stable, even at the cost of sacrificing a higher level of fit [Ramsay and Silverman, 2002]. Employing the technique mentioned above, it was found that a smoothing parameter of $\lambda = 10^8$ was sufficient to produce stable phase-plane plots. In summary, the smoothing parameters used to smooth the realised volatility data in the phase-plane plot case were a B-spline basis of order six ($m = 6$) and a smoothing parameter of $\lambda = 10^8$.

3.1.1.2 Generating a Phase-Plane Plot

Once the realised volatility data was smoothed using the DS algorithm, the result was a smooth functional data object which was passed as an argument to the `phaseplanePlot` function within the *fda* package. This generated phase-plane plot for each year pertaining to a specific market. Extracting the X and Y axes ranges from a phase-plane plot dynamically is no trivial task. The algorithm developed has to be dynamic to find the correct range for all plots between the years 2000 to 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 spiralling irregular shape. However, Figure 2 will suffice for an explanation.

– Please include Figure 2 about here. –

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*. These points mentioned above may be extracted by leveraging the knowledge that a phase-plane plot is simply the second derivative plotted against the first derivative. 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. Algorithm 4 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 utilised 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.

3.2 Functional Regression

We next summarises the practical methods implemented to conduct functional regression analyses on the realised volatility data set, using the daily newspaper-based infectious disease index as a predictor. The functional regression model implementation involves few steps: 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 realised volatility of the specific international financial markets. This data has been smoothed in the previous analyses, thus, it has the property of being functional. 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. The type of regression model most appropriate in this case is function-on-scalar regression. However, to conduct regression on the 22 realised 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. [2019]). However, all the daily EMVID index data points could be utilised 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.

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 realised volatility curves for years $i = 2000, 2001, \ldots, 2021$. The functional explanatory variable $x_{ij}(t)$ is the EMVID index curves over the same yearly i period.

3.2.1 Smoothing the Infectious Disease Index

The EMVID index data was smoothed using the dynamic smoothing 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.

– Please include Figure 3 about here. –

By observing Figure 3, it is clear that the smoothing parameter λ that minimises the GCV criterion is $\lambda = 10^6$. Thus, the EMVID index was smoothed using a sixth-order B-spline basis system ($m = 6$) with a smoothing parameter of $\lambda = 10^6$.

3.2.2 Choice of Functional Regression Smoothing Parameters

The choice of smoothing parameters for functional regression is twofold. It involves choosing smoothing parameters for the actual response variable $y_i(t)$ and the predicted response variable $\hat{y}_i(t)$. The general estimation analyses smoothed the realised volatility data using a sixth order B-spline ($m = 6$) and a smoothing parameter to the value of $\lambda = 10^4$ for all six markets. The EMVID index data was smoothed using a sixth-order B-spline ($m = 6$) and a smoothing parameter to the value of $\lambda = 10^6$ for all six markets. For this specific regression case, the derivative of the regression curves is not of interest. Thus, the order of the B-spline does not need to be as high as a sixth order. The question of accuracy between different B-spline basis system orders arises. The regression curves should exhibit the highest amount of accuracy possible. De Boor [2001] and Ramsay and Silverman [2005] both suggest that a fourth-order B-spline ($m = 4$) with knots at each argument smooths the data with the best accuracy. An experiment was created to test this claim. The realised volatility data for each market was smoothed using a range of B-spline orders ($m = 4, m = 5, m = 6$) over a range of smoothing parameter values ($\lambda = 10^{-3}$ to $\lambda = 10^4$ in increments of 10^1). A sample of 5% from each year was extracted, and the root mean square error (RMSE) metric was calculated. This was done over 25 trials. The results for each market can be seen in Figures 23 - 28 in Appendix E. These results reveal that the claim is partially correct. At low values of λ, the fourth order B-spline ($m = 4$) does have a lower RMSE value. However, as the smoothing parameter λ increases,

the claim does not hold, and the RMSE values are extremely similar across the three orders of B-splines. Due to this fact, and to be consistent across analyses, the B-spline order remains at six ($m = 6$) for the regression analysis.

The realised 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 smoothing parameters for the explanatory variable can be found in Sub-subsection 3.2.1. Both of these functional variables are used within the functional regression model. 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 minimises 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.

– Please include Figure 4 about here. –

The CVISE plot does not show a notable decrease in the CVISE score for a particular smoothing 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.

3.2.2.1 Important *fda* Functions

Some key functions used to develop the dynamic function-on-function regression algorithm are summarised in Table 2. For further documentation on each function, such as function arguments and outputs, see Ramsay et al. [2022].

– Please include Table 2 about here. –

3.2.2.2 Dynamic Function-On-Function Regression Function

A dynamic function-on-function regression algorithm was developed to carry out the regression analysis. Algorithm 5 produces a fitted regression model for the realised volatility of an international market using the EMVID index as the predictor variable. The algorithm is dynamic because it supports different international market volatility data and the smoothing parameters are variable. Algorithm 5 utilises the dynamic B-spline smoothing function developed in Algorithm 3 to smooth the realised volatility and EMVID index data. The EMVID index data contains values for dates that may not exist in the realised volatility data due to the aforementioned trading days issue. Thus, logic was incorporated to compare the market-specific realised volatility data set and the EMVID index data set to ensure both data sets have the same number of observations.

4 Empirical Findings

In this section, we start off with exploratory analysis of our data, before proceeding to our main results.

4.1 Exploratory Data Analysis

An initial exploratory data analysis was conducted on the realised volatility and infectious diseases index data sets. Figures 5 and 6 show the daily realised 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 realised 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 realised volatility in 2004 due to the UBS Securities scandal. China experienced high daily realised 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 realised volatility during the COVID-pandemic was less than the maximum daily realised volatility during the GFC.

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 realised 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 realised volatility, similar to the emerging markets.

– Please include Figures 5 and 6 about here. –

4.1.1 Realised Volatility Box Plots

Figures 7 and 8 show box plots of the daily realised 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 realised volatility, for each year. A blue cross depicts the mean for each year on the box plot. Across emerging and developed markets, the mean realised 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 realised 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 realised 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.

– Please include Figures 7 and 8 about here. –

4.2 EMVID Index Data

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. [2020a]'s claim that COVID-19 affected market volatilities on an unprecedented scale.

– Please include Figure 9 about here. –

4.3 Main Results

We present the results from the various functional data analysis techniques described earlier: general derivative curves, the phase-plane plot analysis, and the functional regression exercise results.

4.3.1 General Derivative Estimation

Derivatives were obtained after the realised volatility data had been smoothed using a B-spline basis system of order six ($m = 6$) and a smoothing parameter $\lambda = 10^4$ for all six markets. Important insights can be extracted from the derivative curves since they depict the rates at which the realised volatility (risk) changes. The results are centred around the COVID-19 period, as the focus of this research is COVID-19. However, the derivative results for the GFC period have also been included for comparison purposes. All curves were generated from the same realised volatility data set over the same period and are shown on the same scale. The derivative curves for the COVID-19 period can be found in Sub-subsubsection 4.3.1.1, the GFC period in Sub-subsection 4.3.1.2 and a comparison between the two in Sub-subsection 4.3.1.3.

4.3.1.1 Derivative Analysis: The COVID-19 Period

Figures 10 and 11 are the first and second derivative curves for the COVID-19 period, respectively. The year 2020 (the black curve) was when the COVID-19 pandemic was at its peak. These figures also include the derivative curves for the pre-COVID-19 years, 2018 as the green curve and 2019 as the blue curve, and the post-COVID-19 year 2021 as the red curve. COVID-19 was still a part of everyday life in 2021, however, it was not as significant and was therefore labelled as the post-COVID-19 year. The derivative curves for three emerging markets (Brazil, China and India) are on the left, and the three developed markets (the Eurozone, UK and US) are on the right of Figures 10 and 11.

From Figure 10 it is evident that the rate of change (velocity) of the realised volatility or risk did not fluctuate excessively in the years before the COVID-19 pandemic. This holds true across emerging and developed markets. However, when COVID-19 appeared, the realised volatility rate of change experienced a sharp increase. This fact is true for all the markets except for China. The sharp increase in the rate of change of risk occurred around March 2020,

the month when COVID-19 was declared a global pandemic. China did not experience as sharp a spike, however, the change in realised 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 incredibly similar. The sudden fluctuation in the rate of change of realised volatility was short-lived and did not persist beyond May 2020. Post COVID-19, during the 2021 year (the red curve in Figure 10), the change in realised 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.

Figure 11 are the second derivative (acceleration) curves of the realised volatility for the COVID-19 period. These curves depict the rate of change of the rate of change of the realised volatility or risk. The conclusions drawn from Figure 11 are identical to those in Figure 10. This figure does highlight Brazil as the market which experienced the most significant acceleration of realised volatility.

– Please include Figures 10 and 11 about here. –

4.3.1.2 Derivative Analysis: The Global Financial Crisis Period

Figures 12 and 13 are the first and second derivative curves for the GFC period, respectively. There was no specific date when the GFC started. However, it is synonymous with the collapse of the Lehman Brothers in 2008. The figures contain derivative curves for the two years prior, 2006 and 2007, and one year post 2009. The derivative curves for three emerging markets (Brazil, China and India) are on the left, and the three developed markets (the Eurozone, UK and US) are on the right of Figures 12 and 13.

Figure 12 indicates that during 2006 (the green curves), the rate of change of realised 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 realised volatility in May, 2006. During 2007, the emerging markets on the left of Figure 12 experienced significant movement in their respective realised volatility velocities. The developed markets on the right of Figure 12 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 realised 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 experienced an increase in the change in realised 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 realised volatility than the emerging markets. During 2009, all six markets experienced minor fluctuations throughout the year, however, not on the scale of 2008.

Figure 13 are the second derivative (acceleration) curves of the realised volatility for the COVID-19 period. These curves depict the rate of change of the rate of change of the realised volatility or risk. The conclusions drawn from Figure 13 are identical to those in Figure 12.

– Please include Figures 12 and 13 about here. –

4.3.1.3 Derivative Analysis: Crisis Comparison

Figures 14 and 15 are the first and second derivative curves for the COVID-19 (2020) and GFC (2008) years, respectively. Again, the derivative curves for the three emerging markets (Brazil, China and India) are on the left, and the three developed markets (the Eurozone, UK and US) are on the right of Figures 14 and 15.

Figures 14 and 15 allow for a direct comparison between the two crises. 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. The COVID-19 pandemic saw a sudden change in the realised volatility around March 2020 but proceeded to stabilise towards the end of May 2020.

– Please include Figures 14 and 15 about here. –

4.3.2 Phase-Plane Plots

The phase-plane plots used in this application were aimed to study the energy within the six international financial markets. The concept is derived from physics principles. Energy can exist within one of two states, kinetic energy and potential energy. Kinetic energy is associated with velocity (first derivative) and potential energy with acceleration (second derivative). Phase-plane plots are created by plotting the velocity on the X-axis and acceleration on the Y-axis. Kinetic energy is synonymous with work being done, in this application, it is movement within the market. Potential energy is synonymous with the outstanding capacity for work, in the market context, it is the availability of capital or resources to be invested in the financial market. The phase-plane plots in this section were generated using the process described in Sub-subsection 3.1.1. The phase-plane plots for three emerging markets (Brazil, China and India) are on the left, and the three developed markets (the Eurozone, UK and US) are on the right in this section's figures.

4.3.2.1 Phase-Plane Plots for the COVID-19 Period

Figure 16 contains the phase-plane plots of all six markets for 2020, the COVID-19 period. All plots have the same scale in order to compare them fairly [2]. Both the emerging and developed markets have remarkably similar phase-plane plot profiles, except for China. The five similar markets will be discussed first, followed by a separate discussion for China.

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 realised volatility. The velocity of the realised volatility reached its largest value just before March. This coincided with the peak of the COVID-19 pandemic when risk and uncertainty were at their highest. After May, the cycle capitulates and becomes exceptionally compressed. This indicates there was low potential and kinetic energy within the system. There was little capacity for trading (investors did not want to invest in the market), and the markets saw much less movement in the daily realised volatility from May to December. This phase represents the recovery period after the initial shock on the market

[2] This is the reason for some of the phase-plane plots being extremely small. Due to the scaling, it was impossible to elegantly represent the phase-plane plots for all six markets across multiple years on one figure.

volatilities due to COVID-19. All five market centres exist around zero, indicating neither a net positive nor negative velocity or acceleration. The emerging and developed markets exhibited very similar behaviour during the COVID-19 pandemic.

The phase-plane plot for China in 2020 exhibits a compact profile with reduced seasonal variability and looks similar to those during periods of war (see Ramsay and Silverman [2002] for further details). Starting from January, a cycle begins and terminates in June. This cycle is much smaller in comparison with the other five markets. It does indicate that the Chinese financial market saw a minor increase in the movement of daily realised volatility but not on a similar scale to the other markets. The compact profile of the phase-plane plot indicates little variability in the kinetic and potential energy within the system. Thus, the Chinese financial market has little capacity for trading but also saw less movement in the daily realised volatility. The fluctuations in uncertainty or risk were much lower for the Chinese market.

– Please include Figure 16 about here. –

4.3.2.2 Phase-Plane Plots for the Global Financial Crisis Period

Figure 17 contains the phase-plane plots of all six markets for 2008, the GFC period. Brazil, the Eurozone, the UK and the US have similar phase-plane plot profiles during this crisis. China and India differ slightly due to the shape of the dominant cycle. In January, all six markets start with a high velocity (large change in the daily realised volatility), but this velocity decreases rapidly as February approaches.

The phase-plane plots for Brazil, the Eurozone, the UK and the US depict a compressed cycle from February until August. This indicates that there was little capacity for trading in the markets. Investors did not have the capital or were not looking to invest in the market. This explains the low kinetic energy, as the market was experiencing minimal movement in the daily realised volatility. From September through to November, a large cycle appears. The kinetic energy increased, which implies a large shift in daily realised volatility occurred. The large shift in daily realised volatility or uncertainty was due to the collapse of the Lehman Brothers investment bank and the magnifying of the GFC. The collapse of the Lehman Brothers occurred in mid-September, which corresponds to the time at which the velocity of realised volatility was at its peak.

India also saw a notable cycle starting in September and continuing until November. This cycle was not as great in magnitude as the corresponding cycle for Brazil, the Eurozone, the UK and the US. China has a similar phase-plane plot profile to the one in COVID-19. The plot is exceptionally compressed and has little change in potential or kinetic energy. The low variation in the movement for the daily realised volatility indicates that the uncertainty in the Chinese market changes gradually and with a small magnitude. The most notable cycle occurs in March and April.

– Please include Figure 17 about here. –

4.3.2.3 A Comparison Between the Global Financial Crisis and COVID-19 Periods

Figures 18 and 19 contain the phase-plane plots, comparing the GFC (2008) and COVID-19 (2020) periods for the emerging and developed markets, respectively. The previous sections discuss the dynamics of each period individually, however, the comparison allows for insights into the difference in magnitudes of energy transfer within the system. Observing Figure 18, 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 19 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 realised 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.

– Please include Figures 18 and 19 about here. –

4.3.2.4 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 summarised 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.

– Please include Table 3 about here. –

In this application, the amount of kinetic energy within the system indicates the extent to which the velocity of daily realised 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 realised 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.

In this application, the amount of potential energy within the system indicates the extent to which the acceleration of daily realised 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 realised volatility

during the COVID-19 pandemic than the emerging markets. The realised 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.

4.3.3 Functional Regression

A regression analysis was conducted on the realised volatility data. The realised 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 realised volatility in 2020, but the aim is to quantify the relationship between infectious diseases and market volatility.

4.3.3.1 Correlation Analysis of The Realised Volatility and EMVID Index

Before proceeding with the regression analysis, the correlation coefficient was calculated for the realised 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 summarised in Figure 20.

– Please include Figure 20 about here. –

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 realised 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 realised 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 realised volatility for the year 2020.

4.3.3.2 Realised Volatility Regression Results

Figure 21 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 realised volatility, and the red curve depicts the predicted realised volatility. The regression results for the additional years can be found in Section G of Appendix F.

From Figure 21, it is clear that the EMVID index can be used as a predictor for the daily realised volatility of 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 realised 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 realised volatility as early as January.

For the years that do not contain any epidemics or pandemics, the regression model performs poorly, as seen by the results in Section G of Appendix F. This is to be expected, as the index is a newspaper-based equity market volatility due to infectious diseases index. Inherently,

only the years that contain infectious disease outbreaks monitored by this index will produce satisfactory results.

– Please include Figure 21 about here. –

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 realised volatility explained by the EMVID index. Table 4 summarises 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 realised 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 realised 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 realised volatility. By looking at the regression model output for 2008, in Figure 29 of Appendix F, it is evident that the regression model is performing poorly despite the R^2 value. Therefore, the root mean square error (RMSE) was calculated for each year across all six financial markets and is summarised in Table 5. This metric is a better measure of performance. The RMSE values for 2008 more accurately describe what is happening in Figure 29 of Appendix F. The RMSE values for the COVID-19 period are lower than all the other years, meaning that the regression model predicts the realised 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 realised volatility of international financial markets, provided that the period of interest experienced a major health crisis.

– Please include Tables 4 and 5 about here. –

Figure 22 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 22 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.

– Please include Figure 22 about here. –

5 Conclusions

Having analysed the realised 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.

The insights from the general derivative exercise resulted in the following findings: **First**, the realised volatility can successfully be modelled using functional data analysis (FDA) techniques. Specifically by using a B-spline basis system and the roughness penalty approach; **Second**, 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 realised 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 Organisation. Post-COVID-19, the change in realised volatility remained relatively stable across all six markets. Only Brazil, the Eurozone and the United Kingdom experienced minor fluctuations during 2021. Thus **fourth**, this 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). 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 stabilise. The fluctuations in realised volatility due to the GFC 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 offered additional findings due to the potential and kinetic energy extraction functions used. This is a significant contribution since, to the authors' knowledge, this type of analysis has not been done before. Additionally, no base function in R existed beforehand to extract these quantities, and in this research we developed a custom function to achieve this. This type of analysis can be extended to other applications, such as temperature or weather data, and is not limited to the data in this specific application. The potential and kinetic energies analysis produced the **seventh** finding in the form 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 **eighth** finding of this research, which is methodology specific. De Boor [2001] and Ramsay and Silverman [2005] 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 realised volatility of international financial markets using the EMVID index as a predictor. This analysis produced the **nineth** finding which was that while EMVID index is an accurate predictor of the realised 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.

Now, we turn our attention to the interpretation of the key findings. The COVID-19 pandemic significantly contributed to the financial markets' instability in 2020. The reason for this instability is that the COVID-19 pandemic introduced a large amount of uncertainty and fear among investors, with investors fleeing to low-risk safe haven assets [Corbet et al., 2021]. This uncertainty galvanised the freefall of most of the international financial markets. The government's policy responses to curb the spread of the virus also had huge implications on the markets [Zaremba et al., 2020; David et al., 2021]. During 2020, the risk started to evolve and increase uncertainty in the markets, which caused dramatic fluctuations in realised 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 realised 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 realised 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 the number of daily COVID-19 cases in China significantly affected the Chinese stock market. However, China implemented strict lockdown measures early on, and the daily infection rate was kept relatively low [Gao et al., 2021]. The low infection rate and proactive approach to handling the COVID-19 pandemic may have lessened the uncertainty among investors, which could be why China is not similarly affected.

The recovery time of the market volatilities was significantly quicker during the COVID-19 pandemic than that observed during the GFC. There are many reasons for the faster recovery of international financial markets during the COVID-19 pandemic such as, 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 [Claessens et al., 2010]. Due to this lack of adequate regulatory policy, 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) [The Reserve Bank of Australia, 2022] have a positive impact during a financial crisis. The response to a global crisis must be swift, systematic, and decisive [Lin, 2008]. 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 [Gao et al., 2021]. This was seen in the results, as the velocity of realised 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. The rich are primarily the people who invest in the financial market itself. Conversely, the origin of the GFC was the financial market itself. Thus, the investors and the rich were directly affected, which is why the recovery took much longer, along with the absent financial policies. The GFC affected the global financial markets and split over to many economies creating 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 [Gao et al., 2021]. 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 realised 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 realised volatility of international financial markets with a focus on COVID-19. Various FDA techniques were utilised throughout this research, and in this specific application, can be utilised 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 realised volatility. The application part of the study was completed using real-world volatility data. Since, volatility is an indicator of financial risk [Poon and Granger, 2003], the findings in this study can assist individual investors, policymakers and financial institutions in making informed decisions when confronted with a fiscal crises due to an infectious disease.

References

[Andersen et al., 2006] Torben Andersen, Tim Bollerslev, Peter Christoffersen, and Francis Diebold. Chapter 15 volatility and correlation forecasting. *Handbook of Economic Forecasting*, 1: 777–878, 12 2006. doi: 10.1016/S1574-0706(05)01015-3.

[Baek et al., 2020] Seungho Baek, Sunil K. Mohanty, and Mina Glambosky. COVID-19 and stock market volatility: An industry level analysis. *Finance Research Letters*, 37:101748, 2020. ISSN 1544-6123. doi: https://doi.org/10.1016/j.frl.2020.101748. URL https://www.sciencedirect.com/science/article/pii/S1544612320311843.

[Baker et al., 2019] Scott R Baker, Nicholas Bloom, Steven J Davis, and Kyle J Kost. Policy news and stock market volatility. Working Paper 25720, National Bureau of Economic Research, March 2019. URL http://www.nber.org/papers/w25720.

[Baker et al., 2020a] Scott R Baker, Nicholas Bloom, Steven J Davis, Kyle Kost, Marco Sammon, and Tasaneeya Viratyosin. The Unprecedented Stock Market Reaction to COVID-19. *The Review of Asset Pricing Studies*, 10(4):742–758, 07 2020a. ISSN 2045-9920. doi: 10.1093/rapstu/raaa008. URL https://doi.org/10.1093/rapstu/raaa008.

[Baker et al., 2020b] Scott R Baker, Nicholas Bloom, Steven J Davis, and Stephen J Terry. Covid-induced economic uncertainty. Working Paper 26983, National Bureau of Economic Research, April 2020b. URL http://www.nber.org/papers/w26983.

[Barndorff-Nielsen and Shephard, 2002] Ole E. Barndorff-Nielsen and Neil Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64(2):253–280, 2002. ISSN 13697412, 14679868. URL http://www.jstor.org/stable/3088799.

[Bollerslev et al., 2003] Tim Bollerslev, Francis Diebold, Torben Andersen, and Paul Labys. Modeling and forecasting realized volatility. *Econometrica*, 71:579–625, 03 2003. doi: 10.2139/ssrn.267792.

[Claessens et al., 2010] Stijn Claessens, Giovanni Dell'Ariccia, Deniz Igan, and Luc Laeven. Cross-country experiences and policy implications from the global financial crisis. *Economic Policy*, 25(62):267–293, 2010.

[Corbet et al., 2021] Shaen Corbet, Yang (Greg) Hou, Yang Hu, Les Oxley, and Danyang Xu. Pandemic-related financial market volatility spillovers: Evidence from the chinese COVID-19 epicentre. *International Review of Economics & Finance*, 71:55–81, 2021. ISSN 1059-0560. doi: https://doi.org/10.1016/j.iref.2020.06.022. URL https://www.sciencedirect.com/science/article/pii/S1059056020301350.

[Das et al., 2019] Sonali Das, Riza Demirer, Rangan Gupta, and Siphumlile Mangisa. The effect of global crises on stock market correlations: Evidence from scalar regressions via functional data analysis. *Structural Change and Economic Dynamics*, 50:132–147, 2019. ISSN 0954-349X. doi: https://doi.org/10.1016/j.strueco.2019.05.007.

[David et al., 2021] S.A. David, C.M.C. Inácio Jr., and José A. Tenreiro Machado. The recovery of global stock markets indices after impacts due to pandemics. *Research in International Business and Finance*, 55:101335, 2021. ISSN 0275-5319. doi: https://doi.org/10.

1016/j.ribaf.2020.101335. URL https://www.sciencedirect.com/science/article/pii/S0275531920309429.

[De Boor, 2001] Carl De Boor. *A practical guide to splines*. Springer New York, 27 edition, 2001. ISBN 978-0387953663.

[Díaz et al., 2022] Fernando Díaz, Pablo A. Henríquez, and Diego Winkelried. Stock market volatility and the COVID-19 reproductive number. *Research in International Business and Finance*, 59:101517, 2022. ISSN 0275-5319. doi: https://doi.org/10.1016/j.ribaf.2021.101517. URL https://www.sciencedirect.com/science/article/pii/S0275531921001380.

[Gao et al., 2021] Xue Gao, Yixin Ren, and Muhammad Umar. To what extent does covid-19 drive stock market volatility? A comparison between the U.S. and China. *Economic Research-Ekonomska Istraživanja*, 35:1–21, 04 2021. doi: 10.1080/1331677X.2021.1906730.

[Heber et al., 2009] Gerd Heber, Lunde Asger, Neil Shephard, and Kevin Sheppard. Oxford-man institute's realized library. Technical report, Oxford-Man Institute, University of Oxford, 2009. Library Version: 0.3.

[Kokoszka and Reimherr, 2018] Piotr Kokoszka and Matthew Reimherr. *Introduction to Functional Data Analysis*. Chapman and Hall/CRC, 1 edition, 2018. ISBN 9781315117416.

[Lin, 2008] Justin Yifu Lin. The impact of the financial crisis on developing countries. *Macroeconomics and Economic Growth: Economic Conditions and Volatility*, 2008. doi: 10.1596/26129.

[Mangisa et al., 2019] Siphumlile Mangisa, Sonali Das, Surajit Ray, and Gary Sharp. Functional regression models for South African economic indicators: A growth curve perspective. *OPEC Energy Review*, 43, 03 2019. doi: 10.1111/opec.12148.

[McAleer and Medeiros, 2008] Michael McAleer and Marcelo C. Medeiros. Realized volatility: A review. *Econometric Reviews*, 27(1-3):10–45, 2008. doi: 10.1080/07474930701853509. URL https://doi.org/10.1080/07474930701853509.

[Meddahi, 2002] Nour Meddahi. A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17(5):479–508, 2002. ISSN 08837252, 10991255. URL http://www.jstor.org/stable/4129268.

[Muller, 2011] Hans-Georg Muller. Functional data analysis. Technical report, Department of Statistics University of California, Davis, One Shields Ave., Davis, CA 95616, USA, 2011.

[Poon and Granger, 2003] Ser-Huang Poon and Clive W. J. Granger. Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2):478–539, 2003. ISSN 00220515. URL http://www.jstor.org/stable/3216966.

[Ramsay and Silverman, 2002] James Ramsay and Bernard Silverman. *Applied Functional Data Analysis: Methods and Case Studies*. Springer New York, NY, 1 edition, 2002. ISBN 978-0-387-22465-7.

[Ramsay and Silverman, 2005] James Ramsay and Bernard Silverman. *Functional Data Analysis*. Springer New York, NY, 2 edition, 2005. ISBN 978-0-387-22751-1.

[Ramsay et al., 2009] James Ramsay, Giles Hooker, and Spencer Graves. *Functional Data Analysis with R and MATLAB*. Springer New York, NY, 1 edition, 2009. ISBN 978-0-387-98185-7.

[Ramsay et al., 2022] J.O. Ramsay, Spencer Graves, and Giles Hooker. *Package 'fda' (Functional Data Analysis)*, 2022. URL https://cran.r-project.org/web/packages/fda/fda.pdf. Version 5.5.1.

[Shephard and Sheppard, 2010] Neil Shephard and Kevin Sheppard. Realising the future: Forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, 25(2):197–231, 2010. ISSN 08837252, 10991255. URL http://www.jstor.org/stable/40607017.

[The Reserve Bank of Australia, 2022] The Reserve Bank of Australia. The global financial crisis. https://www.rba.gov.au/education/resources/explainers/pdf/the-global-financial-crisis.pdf?v=2022-09-03-11-58-25, 2022. Accessed: 2022-09-03.

[Ullah and Finch, 2013] Shahid Ullah and Caroline Finch. Applications of functional data analysis: A systematic review. *BMC medical research methodology*, 13:43, 2013. doi: 10.1186/1471-2288-13-43.

[Wang et al., 2016] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295, 2016.

[Zaremba et al., 2020] Adam Zaremba, Renatas Kizys, David Y. Aharon, and Ender Demir. Infected markets: Novel coronavirus, government interventions, and stock return volatility around the globe. *Finance Research Letters*, 35:101597, 2020. ISSN 1544-6123. doi: https://doi.org/10.1016/j.frl.2020.101597. URL https://www.sciencedirect.com/science/article/pii/S1544612320306310.

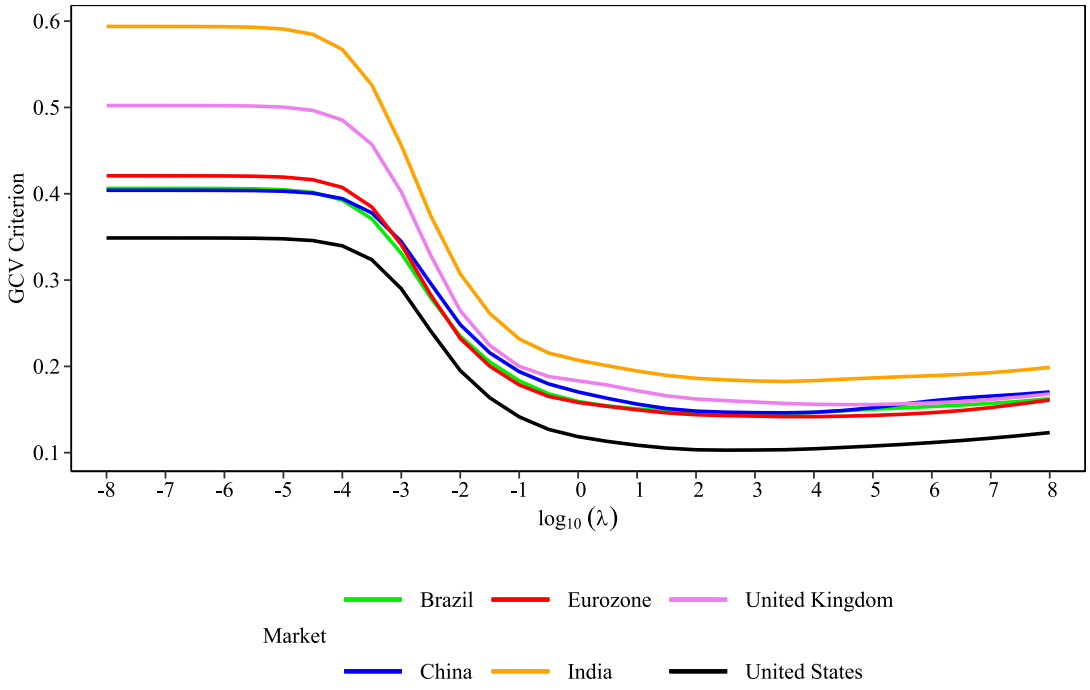


Figure 1: The GCV criterion for each financial market, with a B-spline system of order six ($m = 6$), over a range of smoothing parameter λ values.

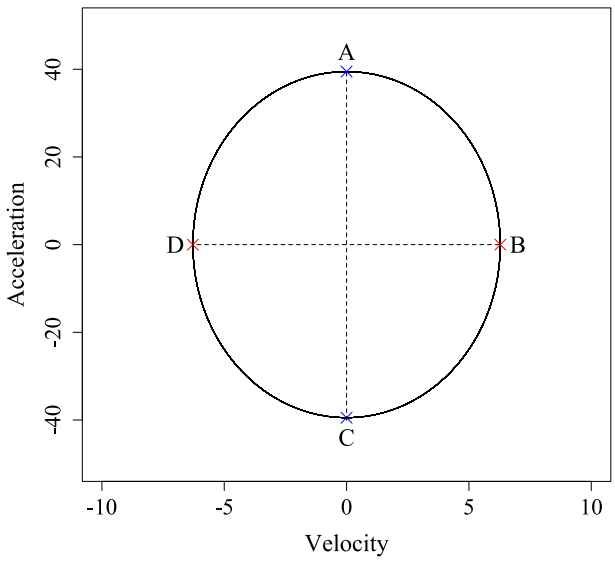


Figure 2: A phase-plane plot of the $\sin(2\pi t)$ function.

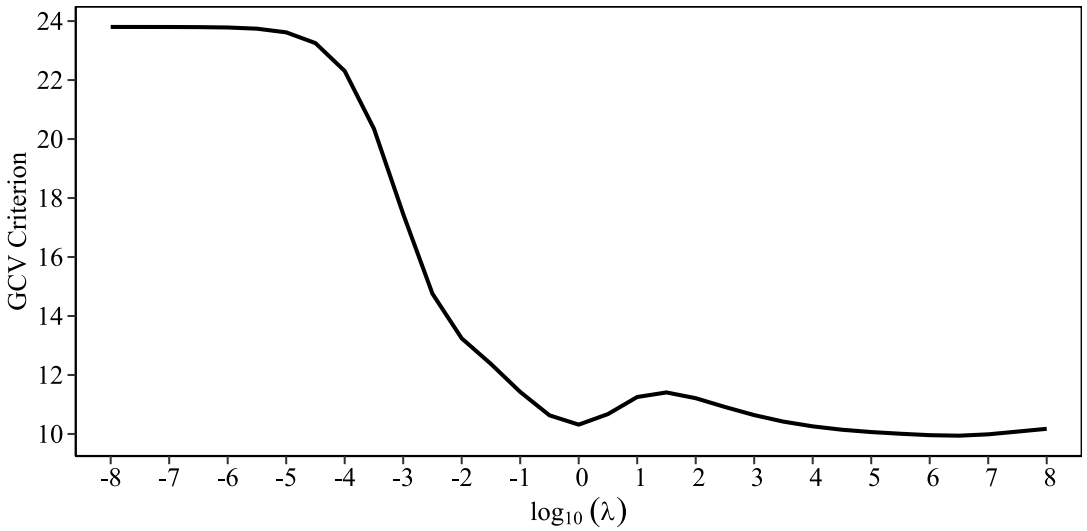


Figure 3: The EMVID index GCV criterion, with a B-spline system of order six ($m = 6$), over a range of smoothing parameter λ values.

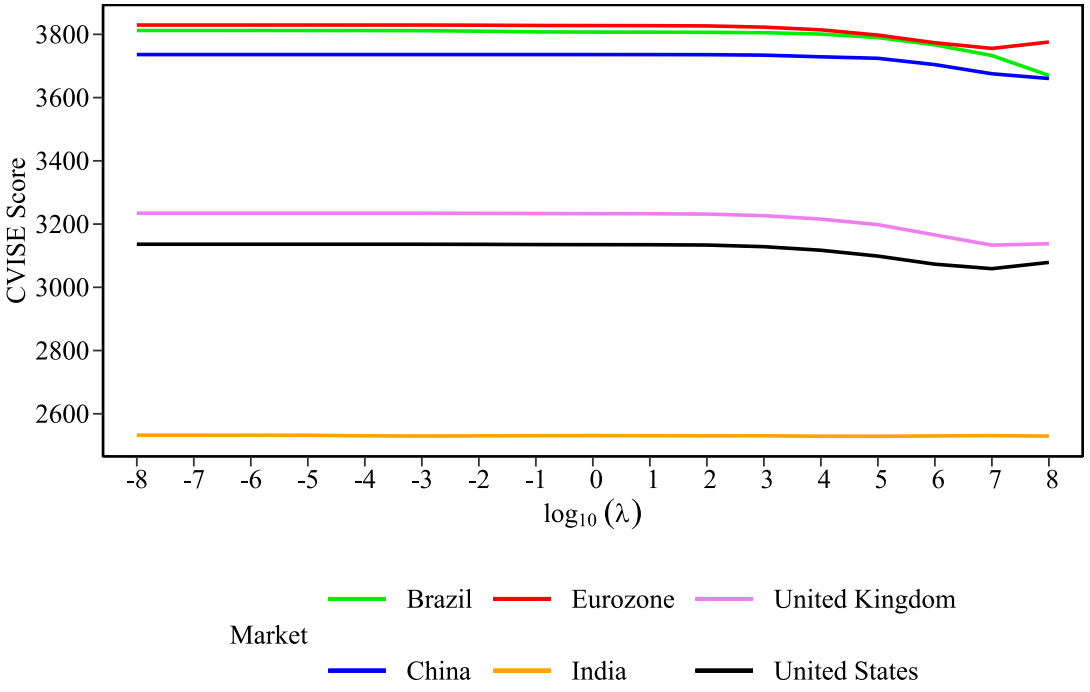
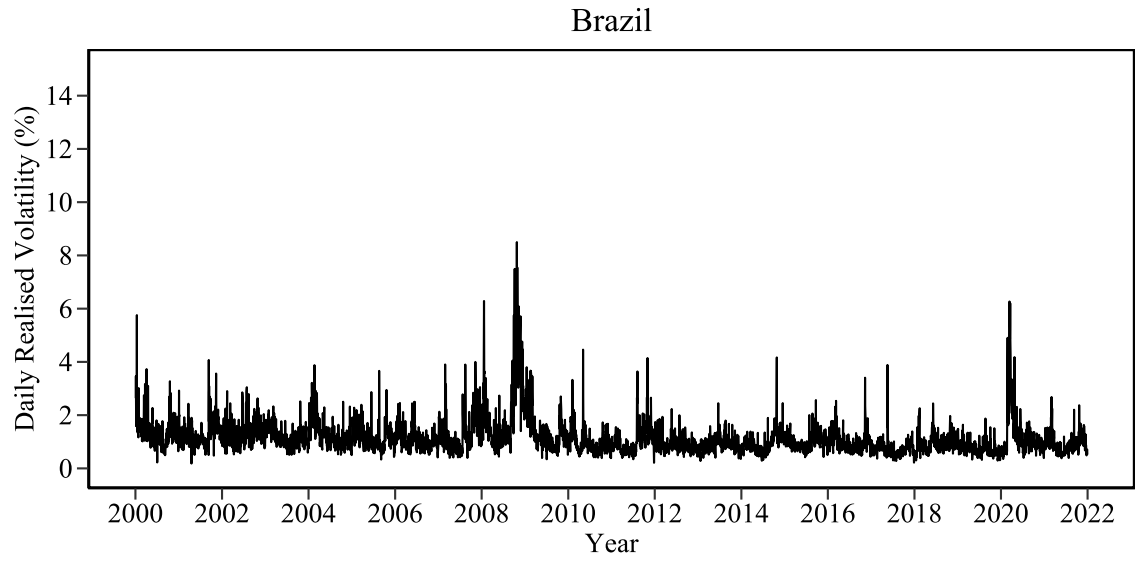
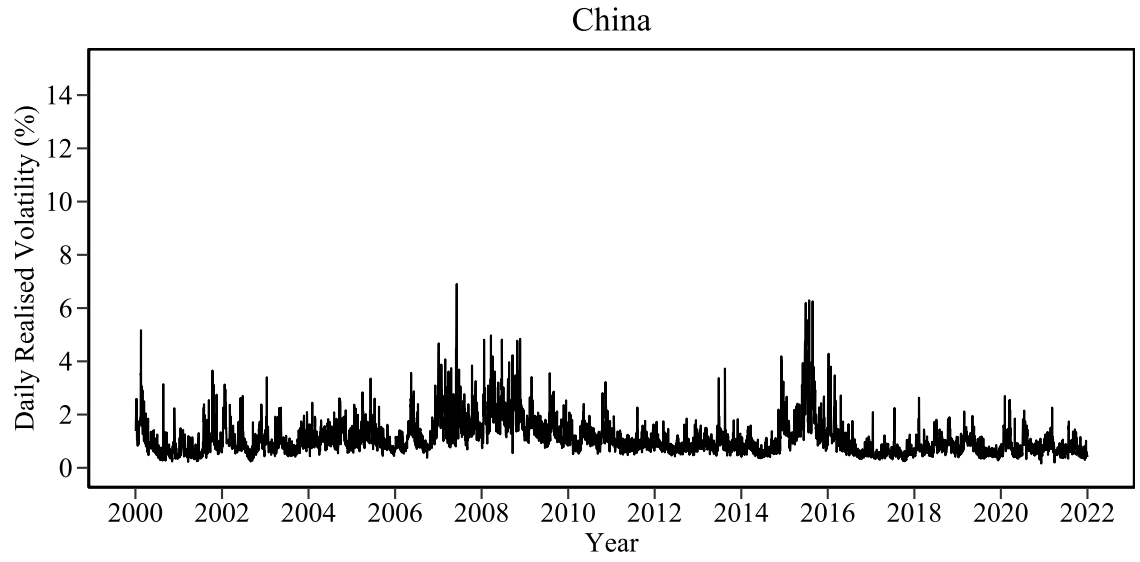


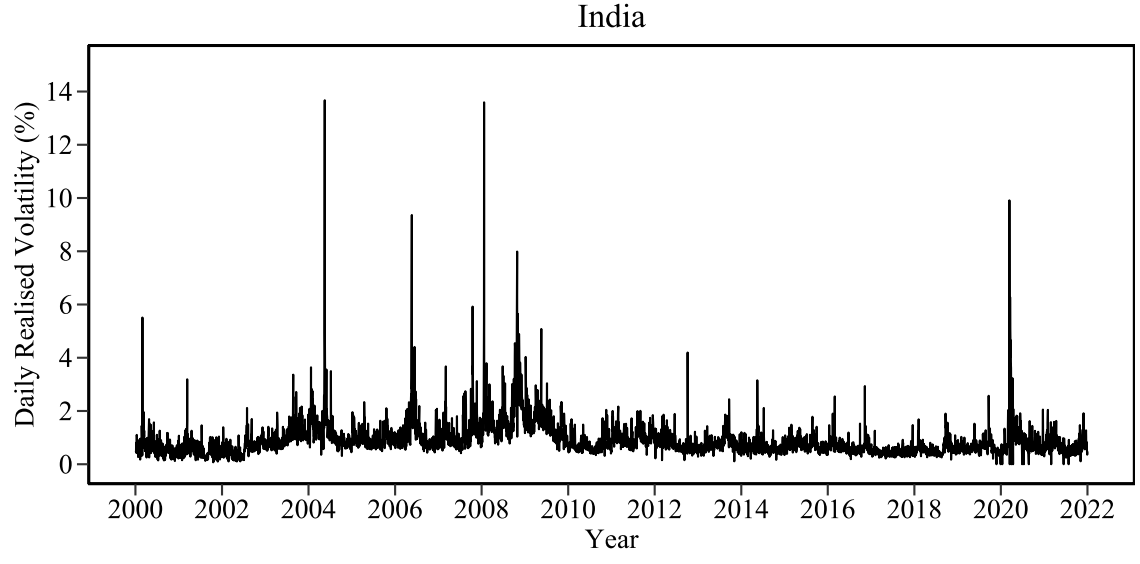
Figure 4: The CVISE score for each financial market, with a B-spline system of order six ($m = 6$), over a range of smoothing parameter λ values.



(a) Brazil

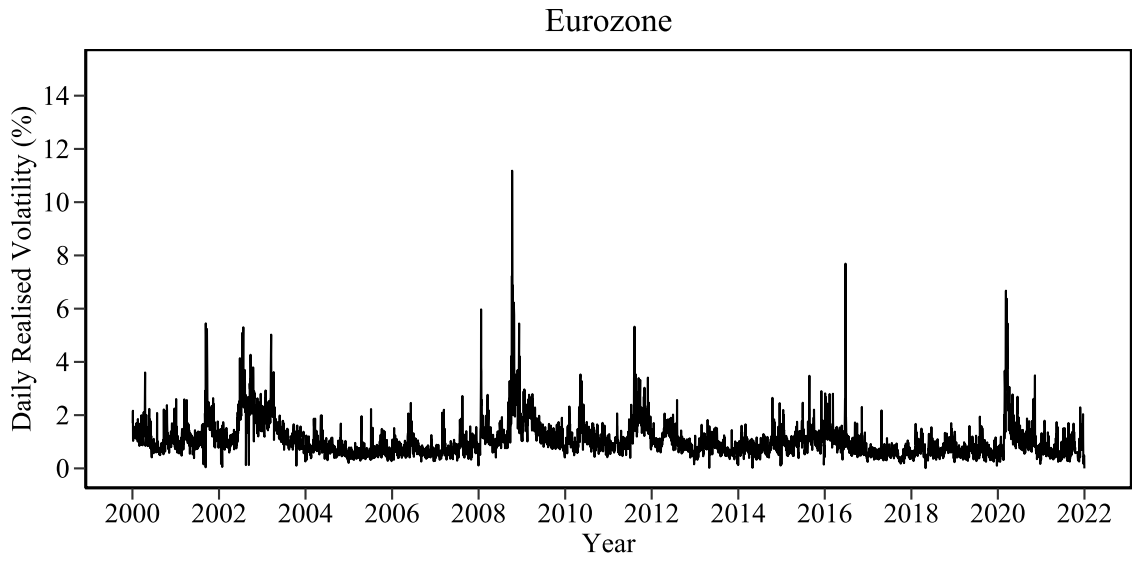


(b) China

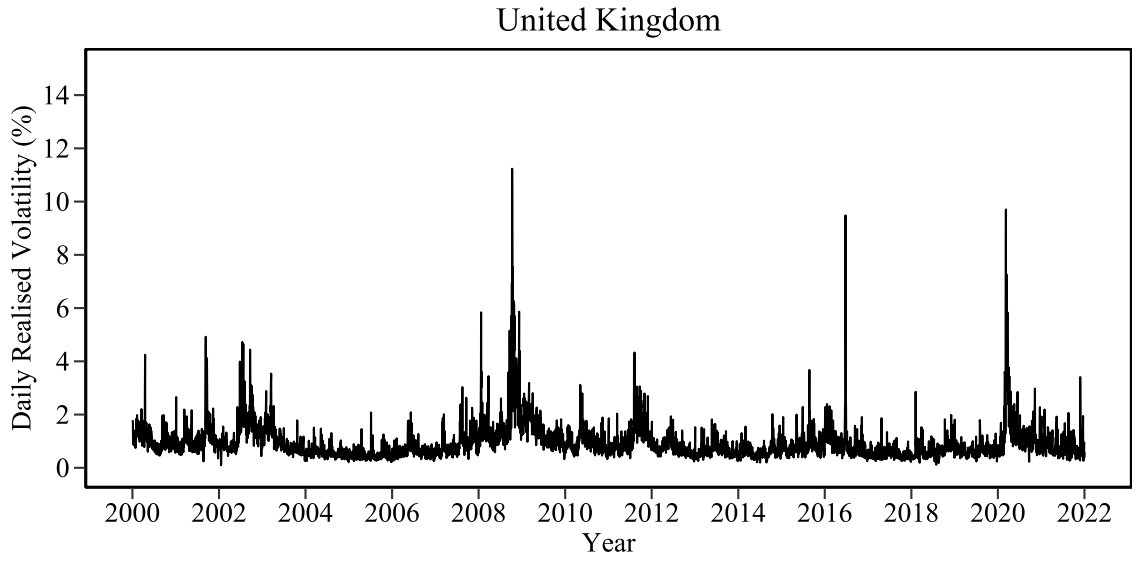


(c) India

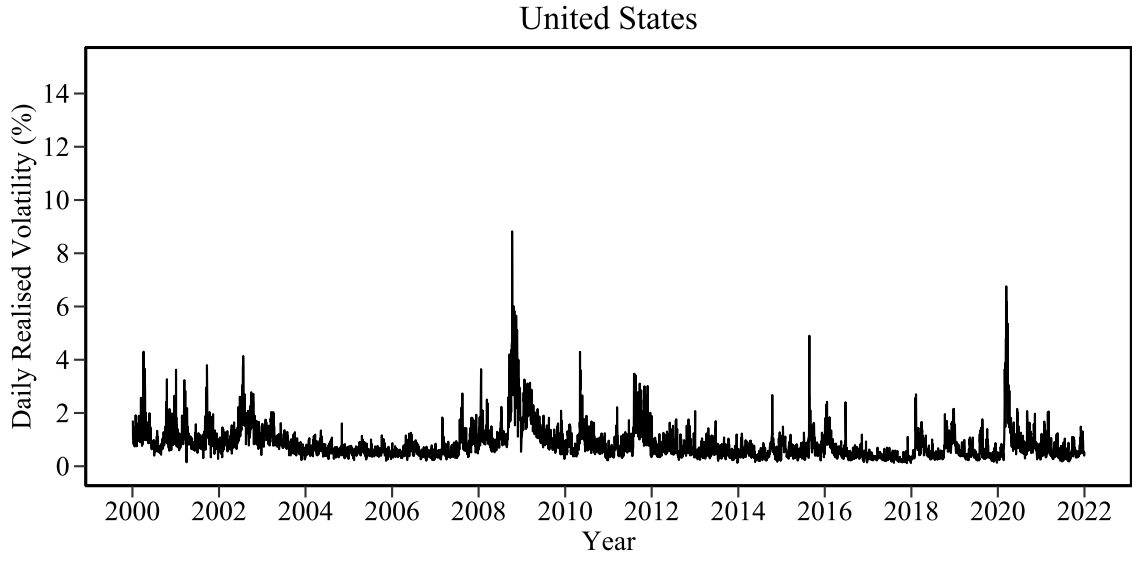
Figure 5: Time series plots of the daily realised volatility, from $2000 - 2021$, for the emerging markets.



(a) The Eurozone

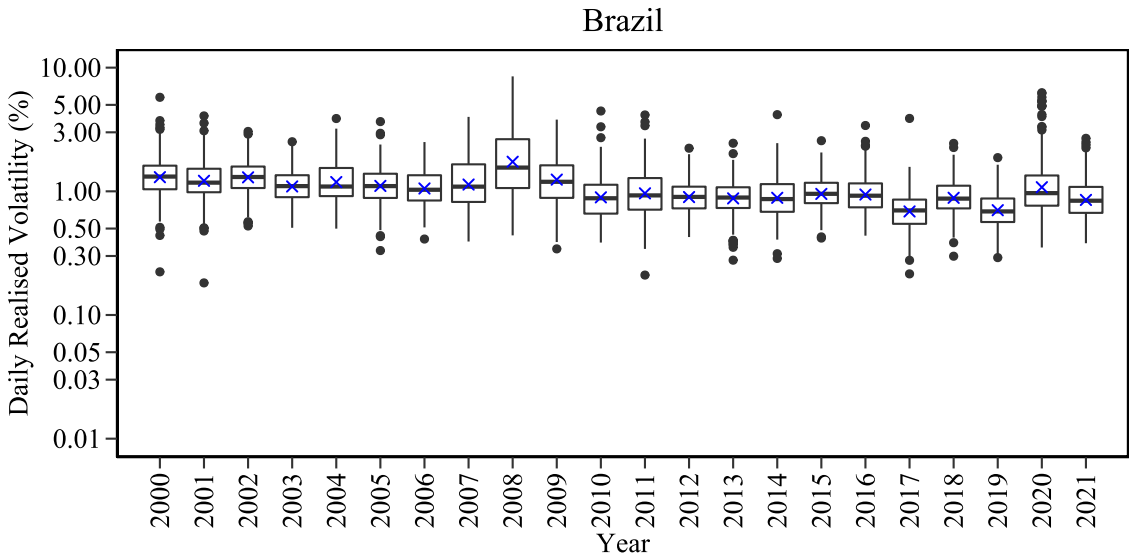


(b) United Kingdom

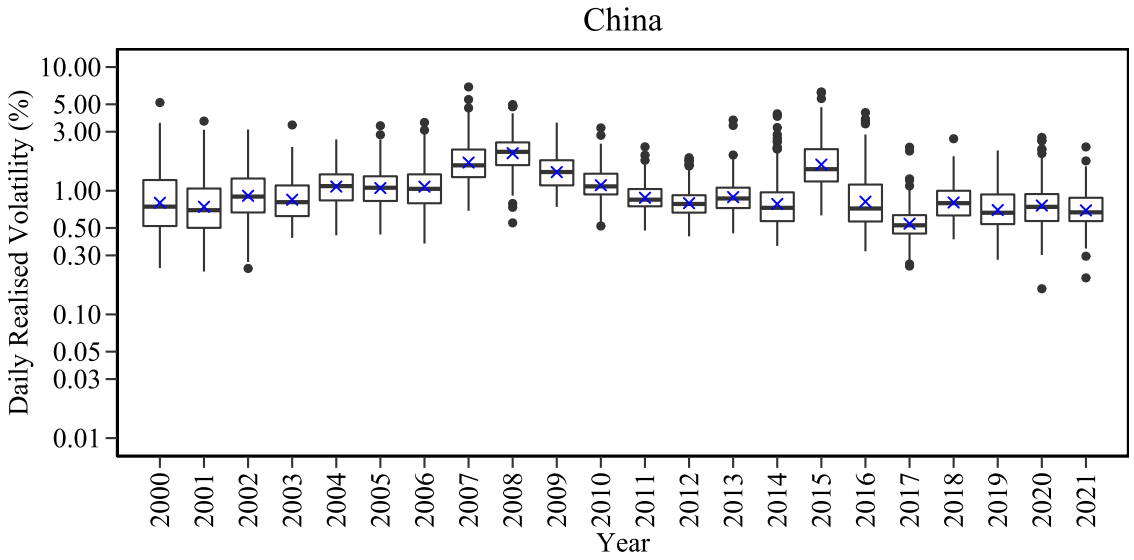


(c) United States

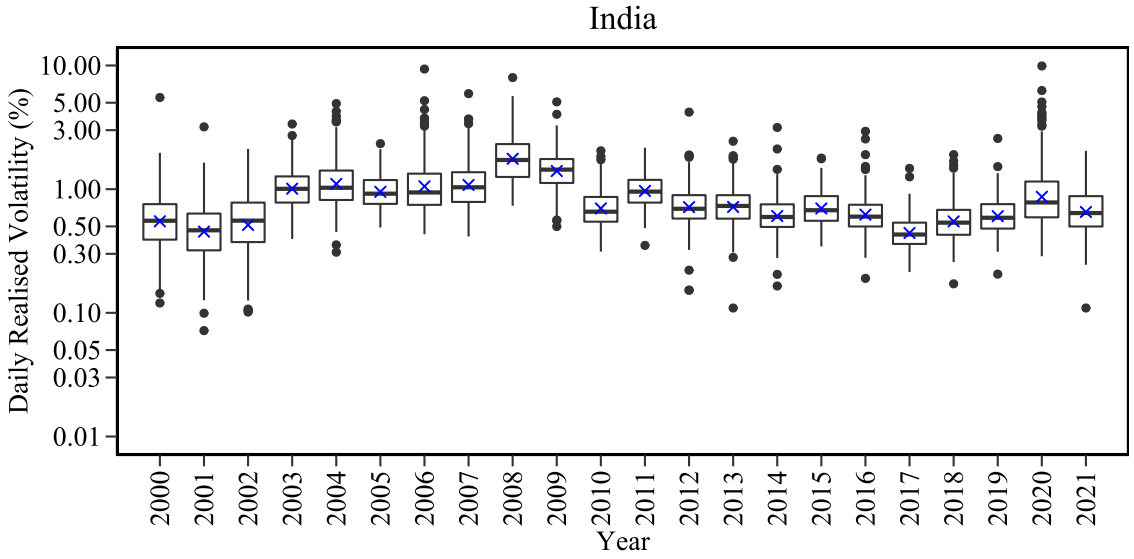
Figure 6: Time series plots of the daily realised volatility, from $2000 - 2021$, for the developed markets.



(a) Brazil

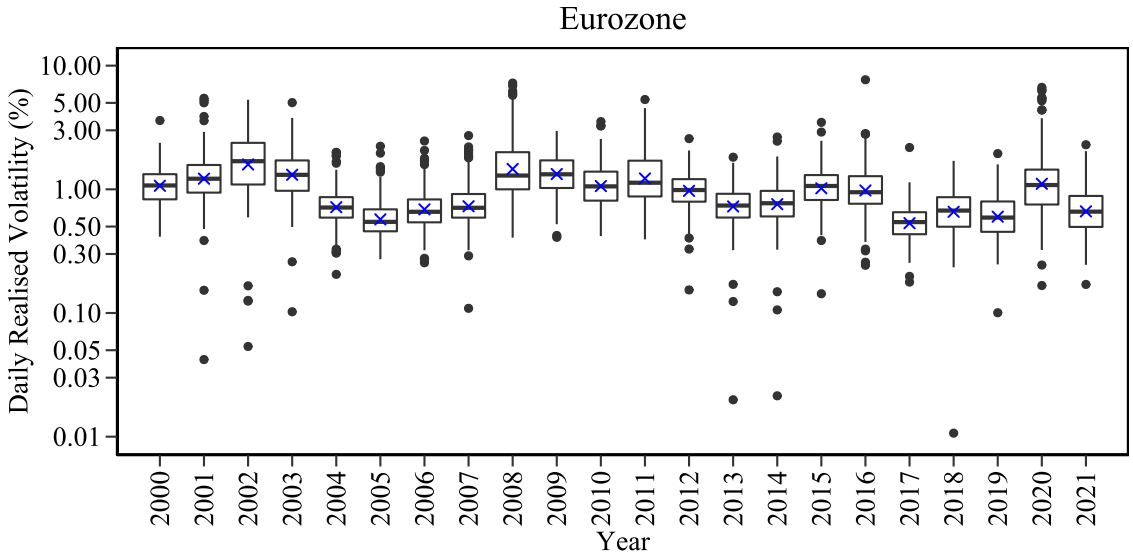


(b) China

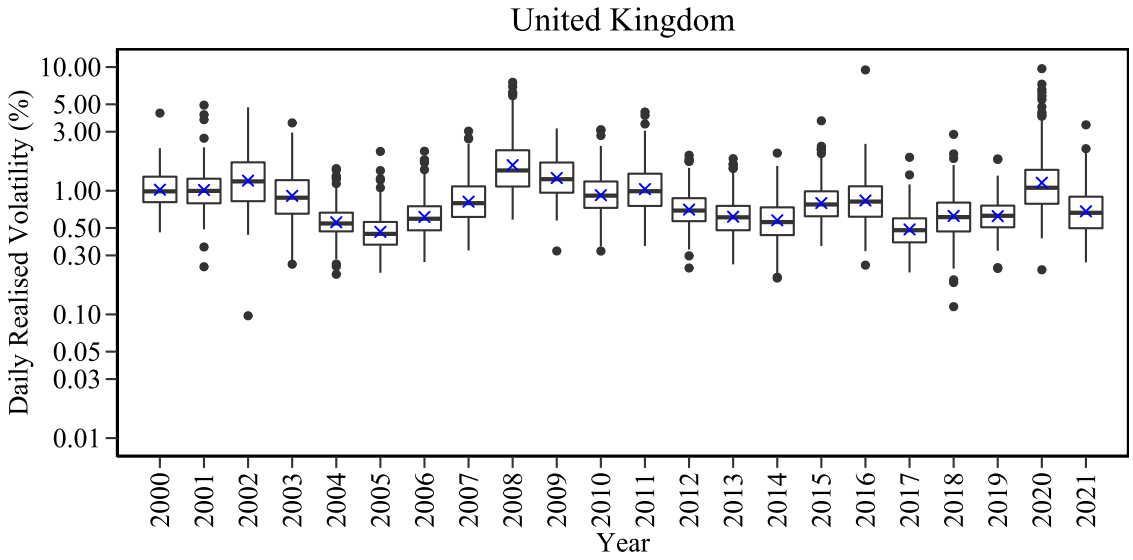


(c) India

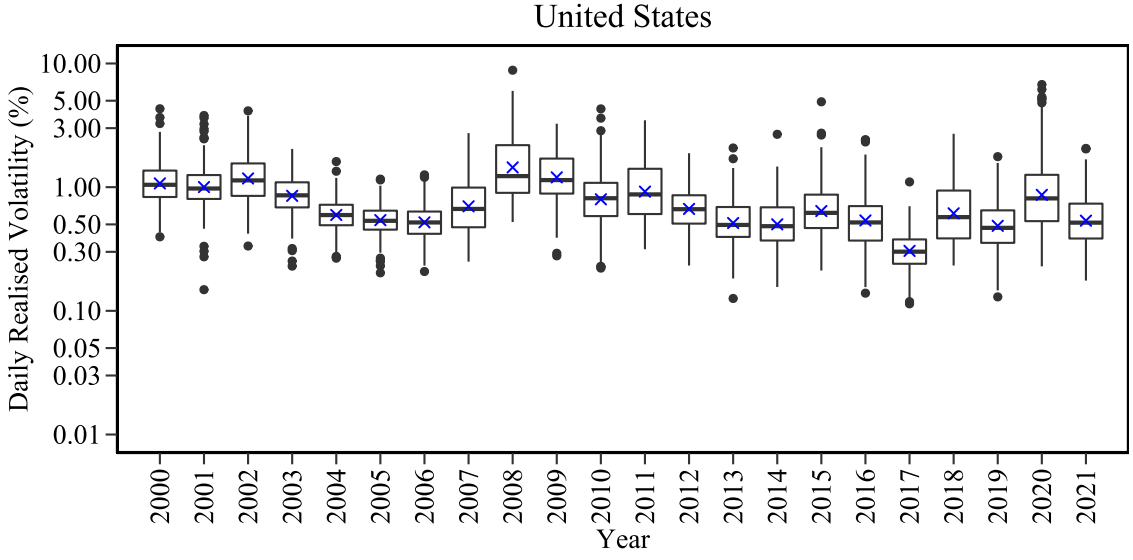
Figure 7: Box plots of the daily realised volatility, from 2000 – 2021, for the emerging markets.



(a) The Eurozone



(b) United Kingdom



(c) United States

Figure 8: Box plots of the daily realised volatility, from 2000 − 2021, for the developed markets.

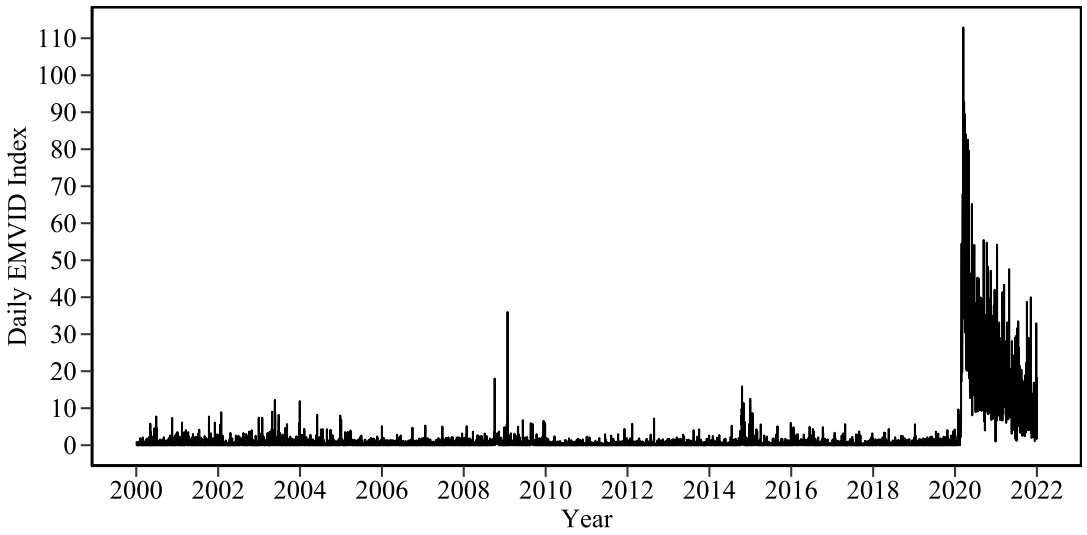


Figure 9: The daily EMVID index from $2000 - 2021$.

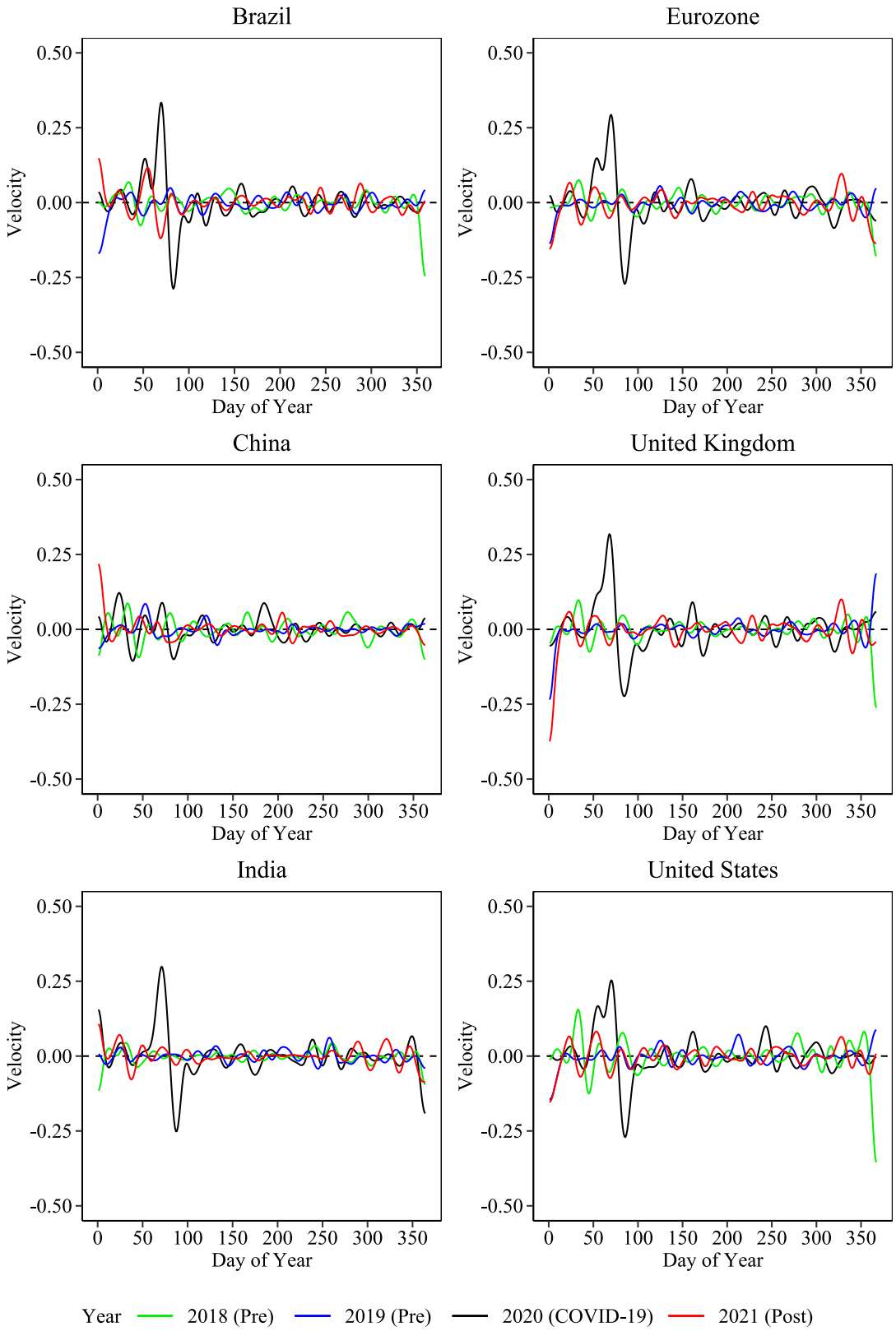


Figure 10: The first derivative of the realised volatility over the COVID-19 period.

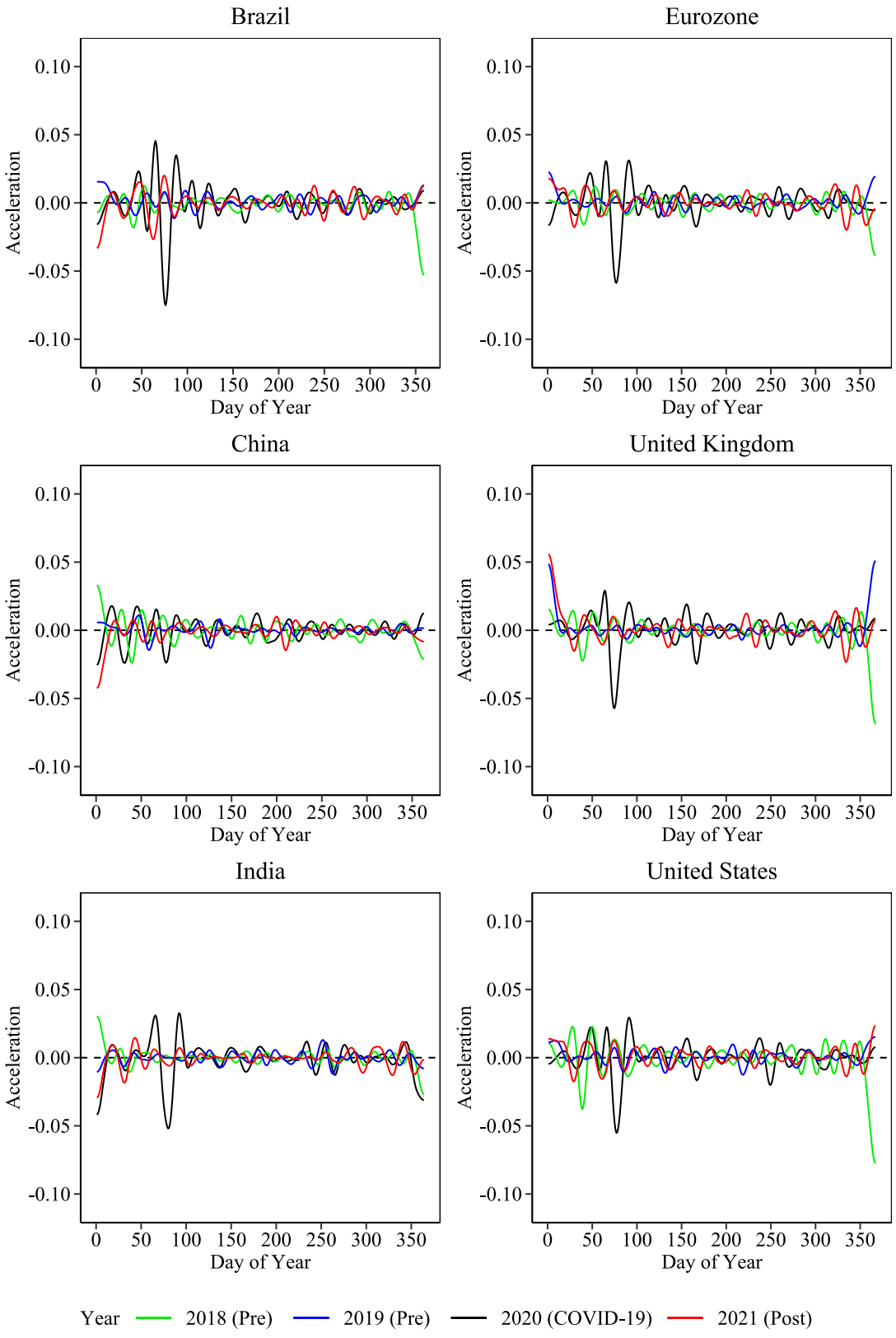


Figure 11: The second derivative of the realised volatility over the COVID-19 period.

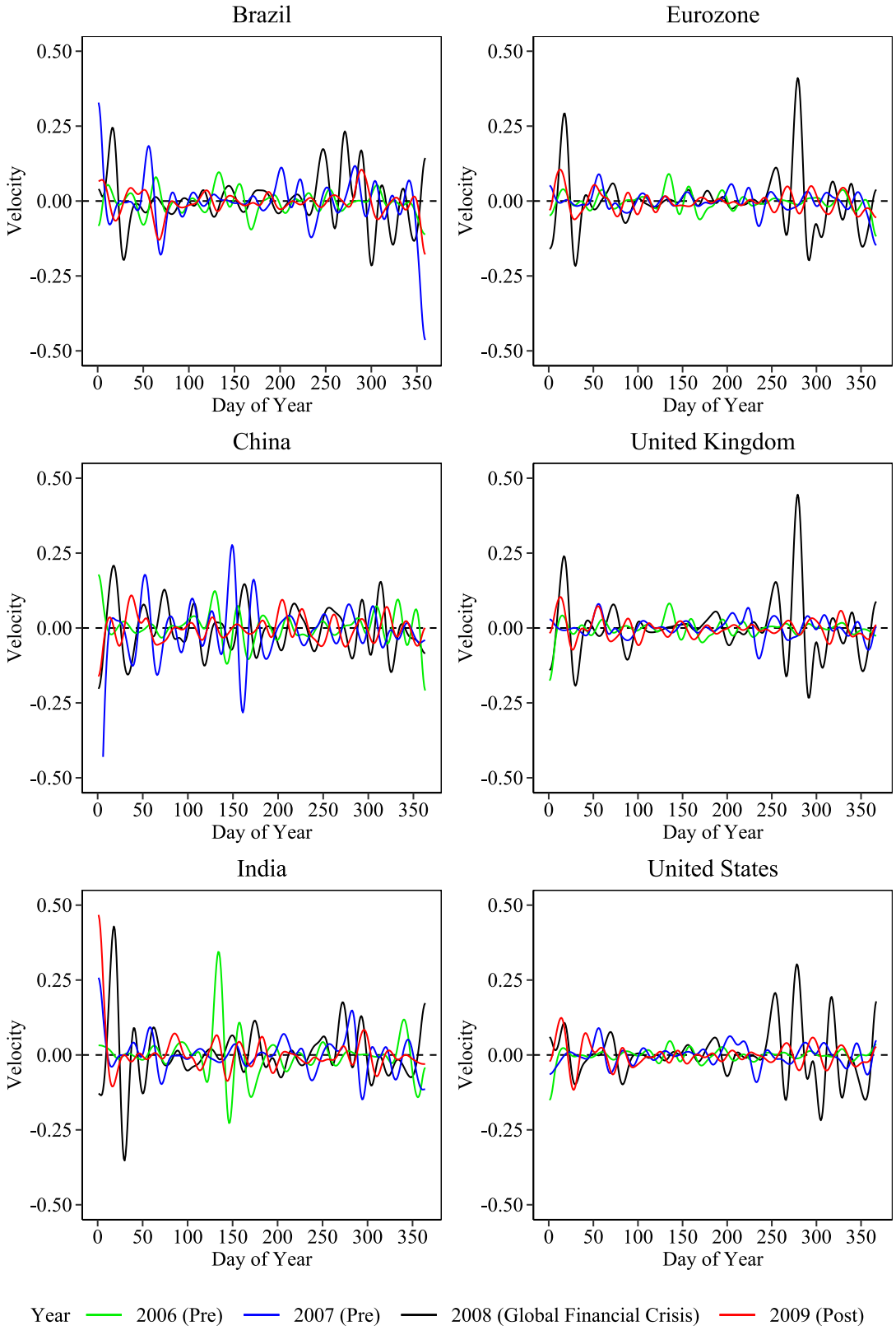


Figure 12: The first derivative of the realised volatility over the GFC period.

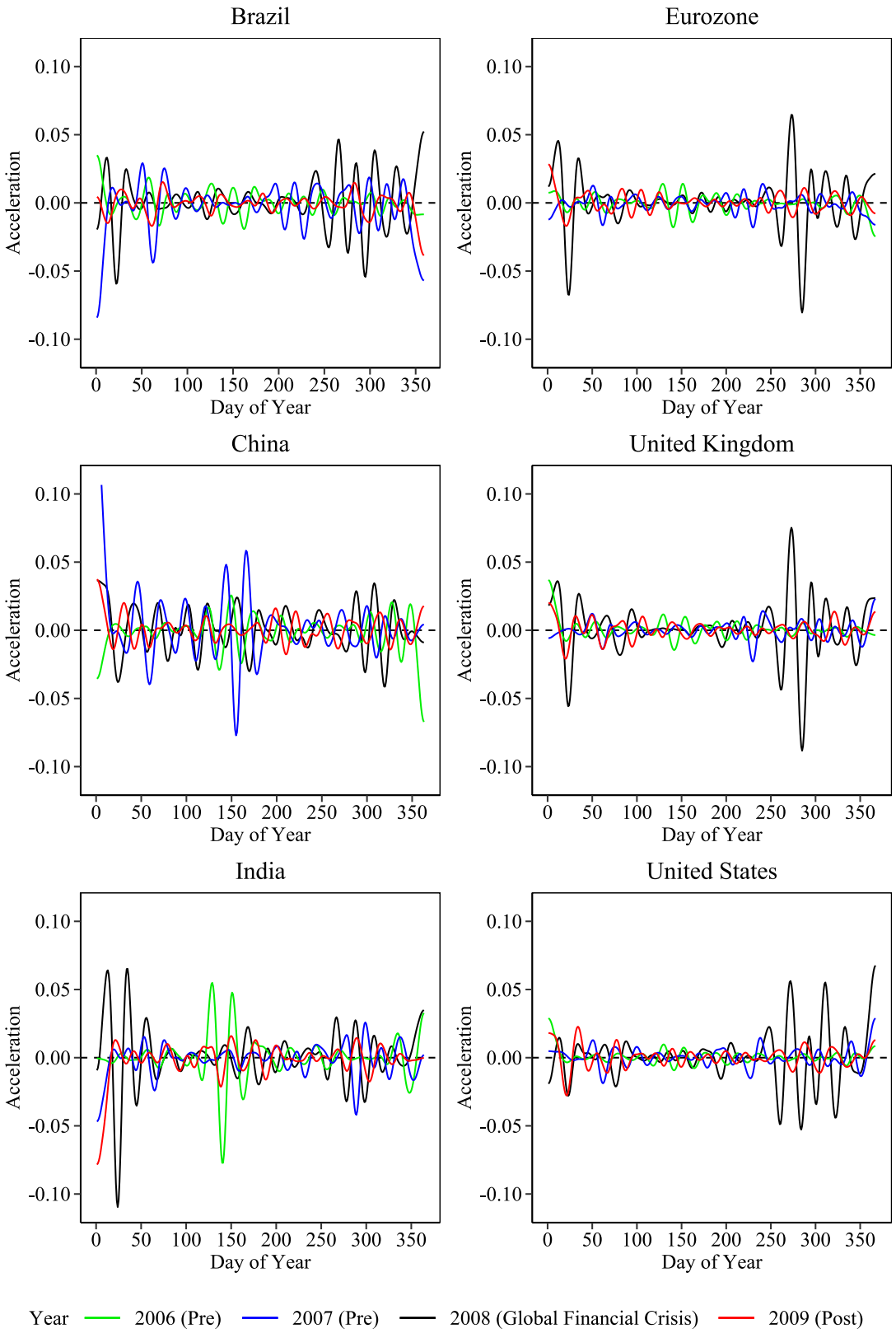


Figure 13: The second derivative of the realised volatility over the GFC period

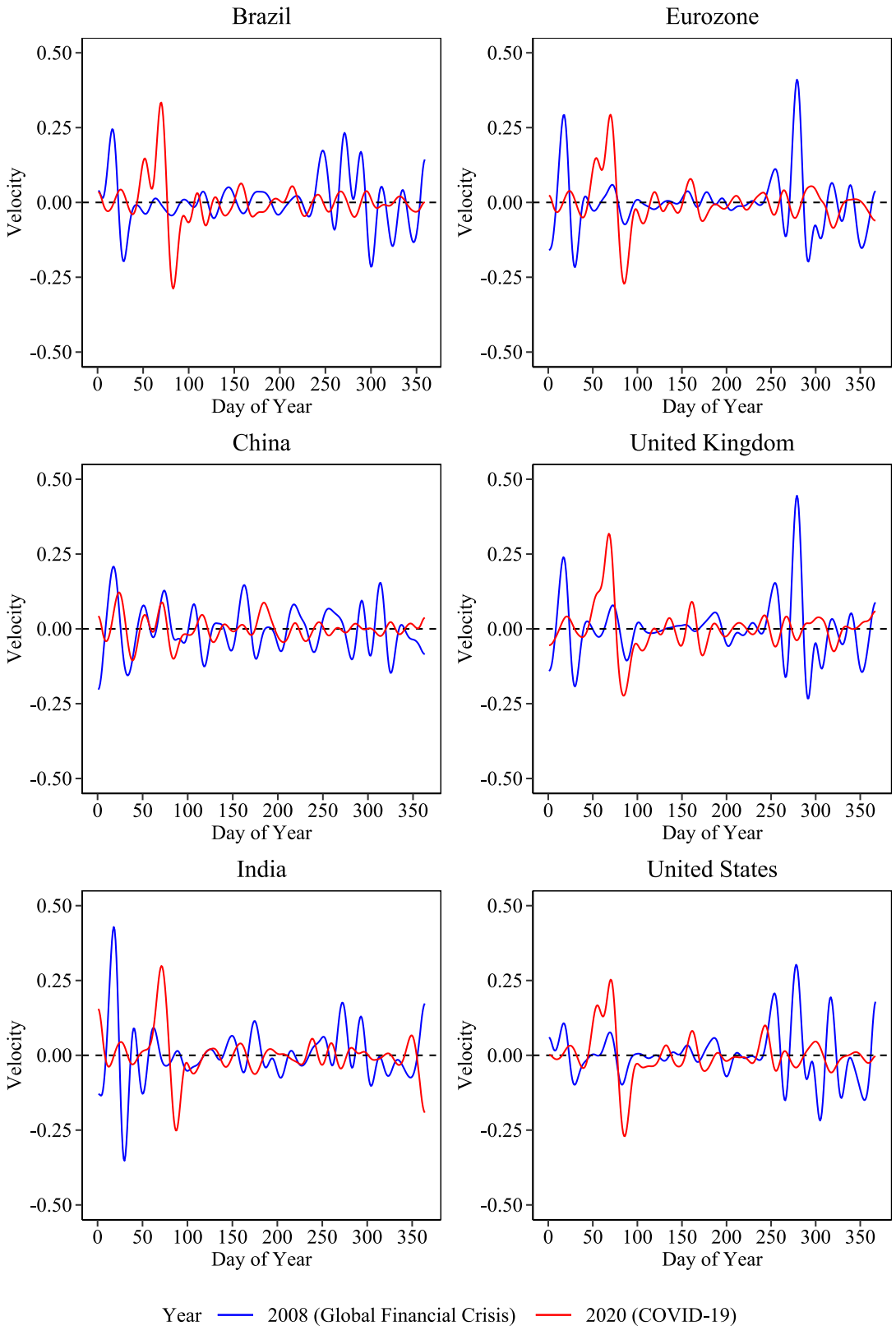


Figure 14: The first derivative of the realised volatility over the COVID-19 and GFC periods.

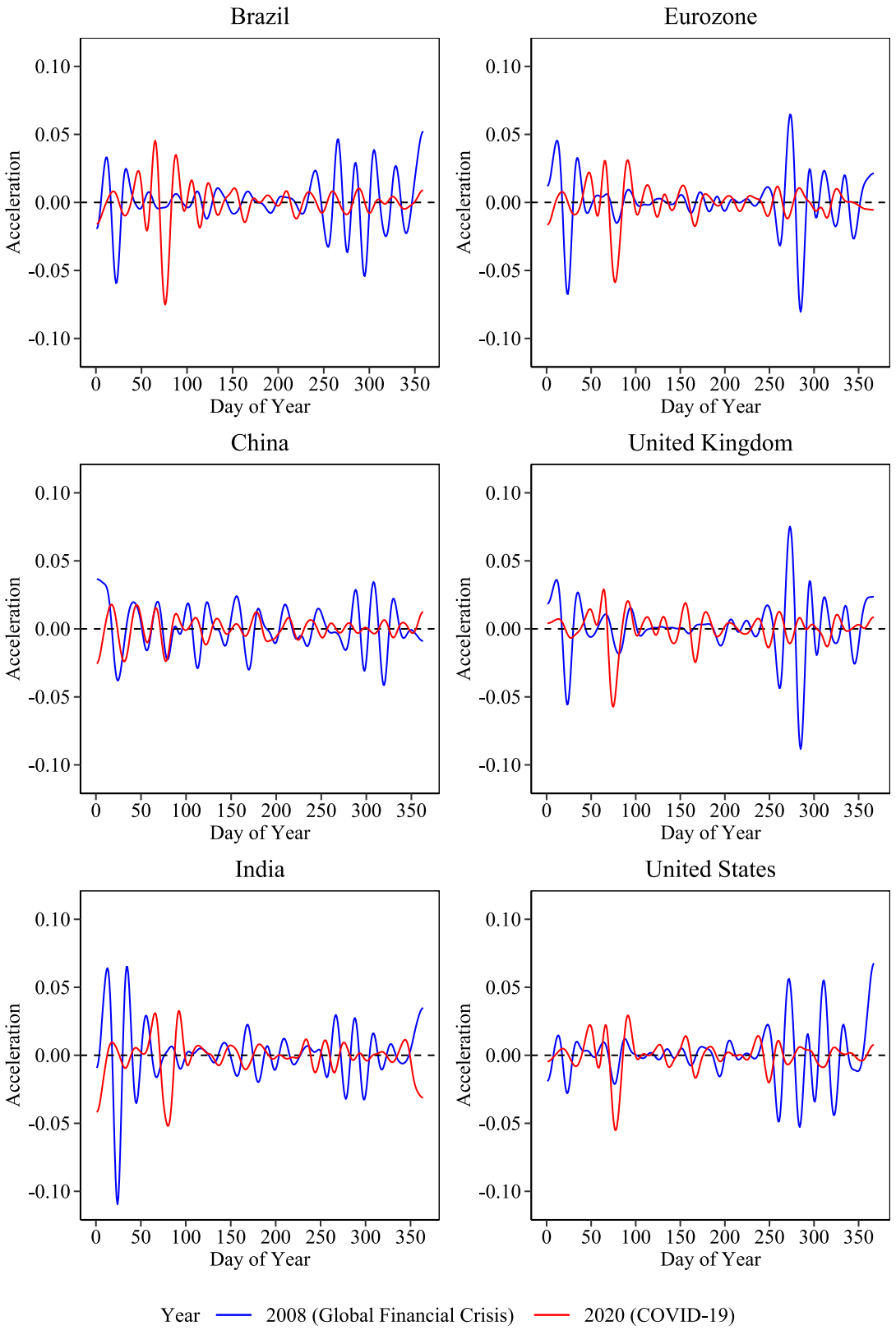


Figure 15: The second derivative of the realised volatility over the COVID-19 and GFC periods.

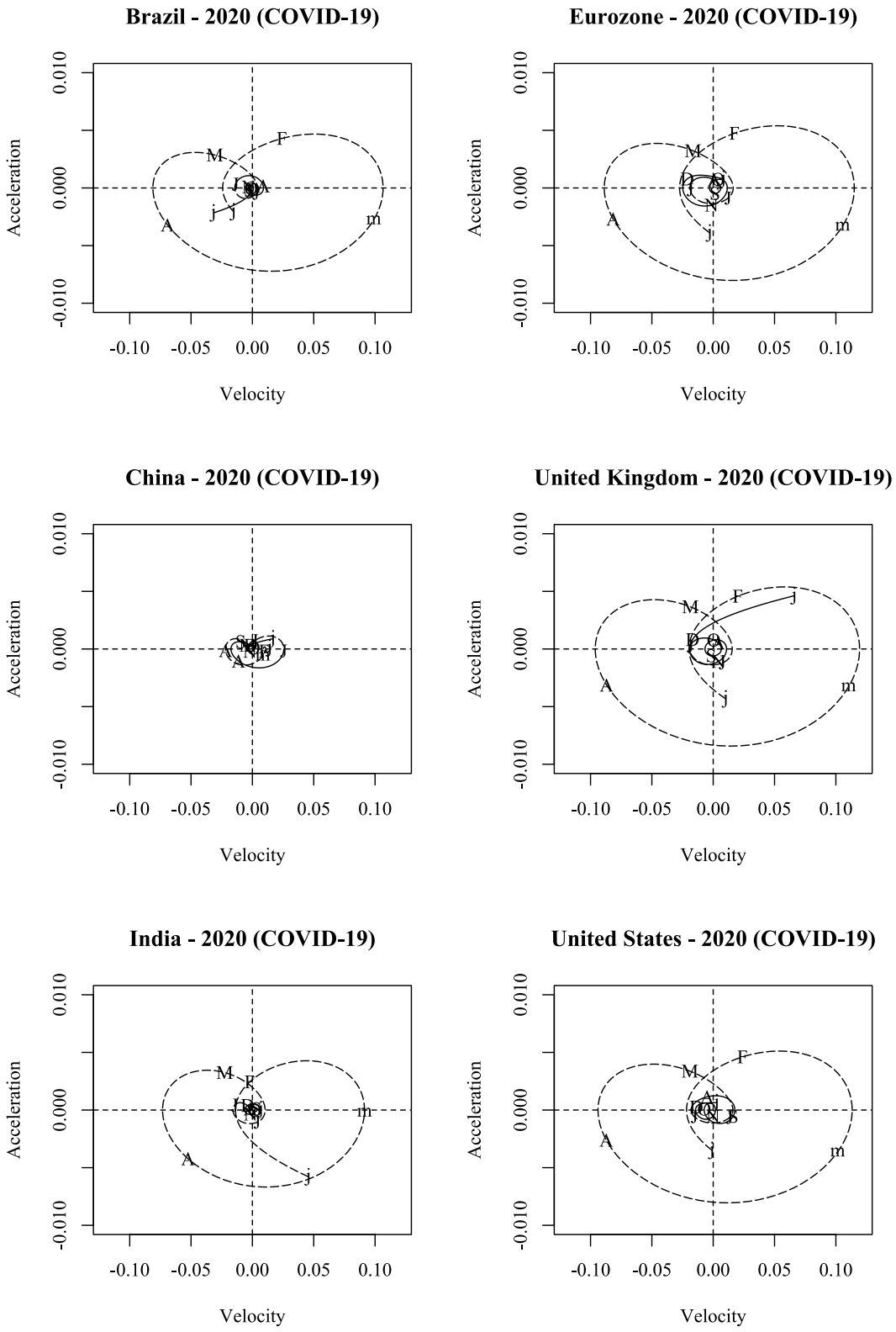


Figure 16: Phase-plane plots of all six markets for the COVID-19 (2020) period.

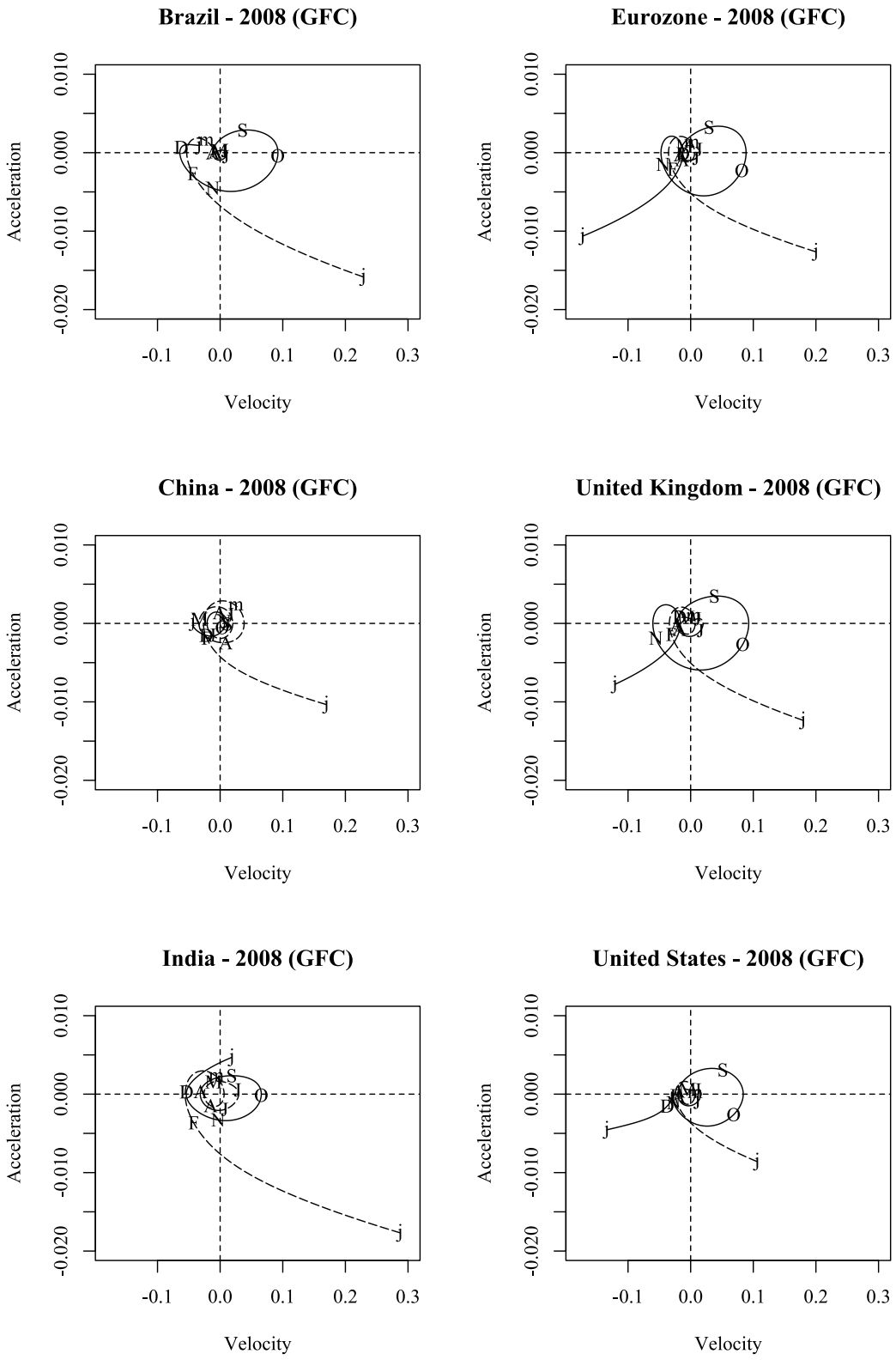
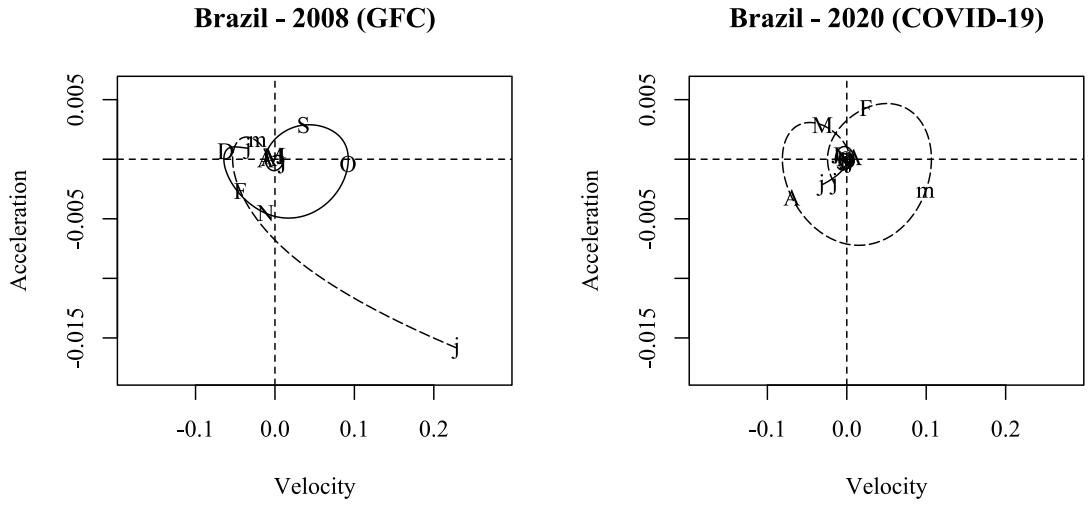
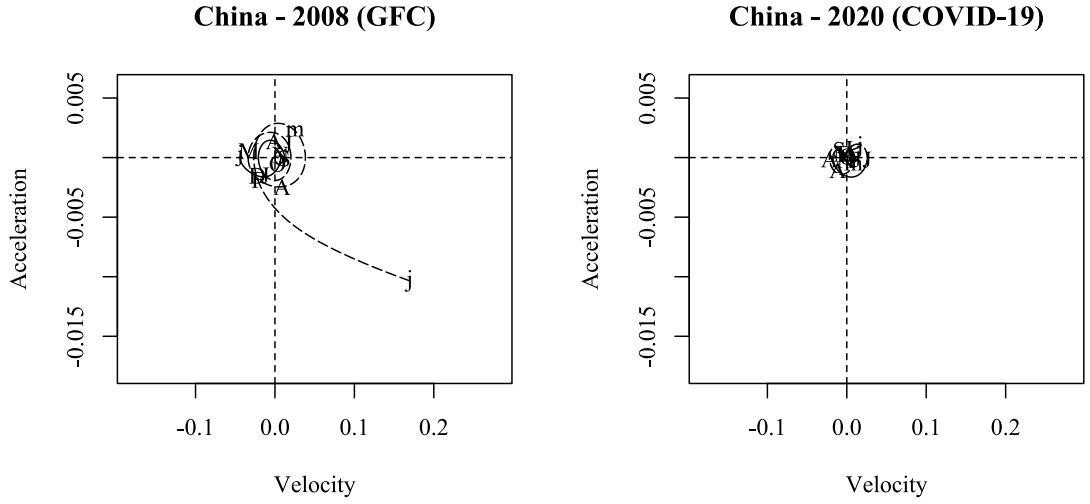


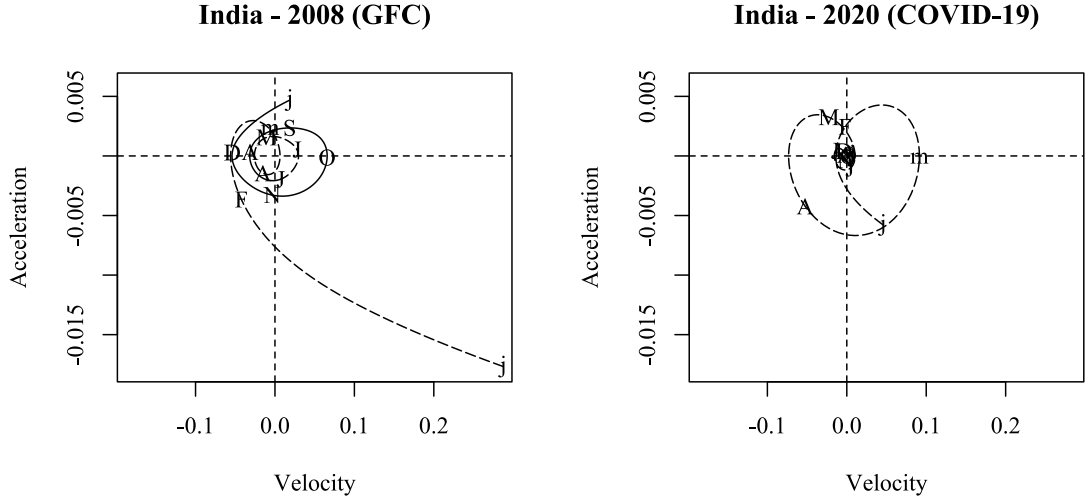
Figure 17: Phase-plane plots of all six markets for the GFC (2008) period.



(a) Brazil

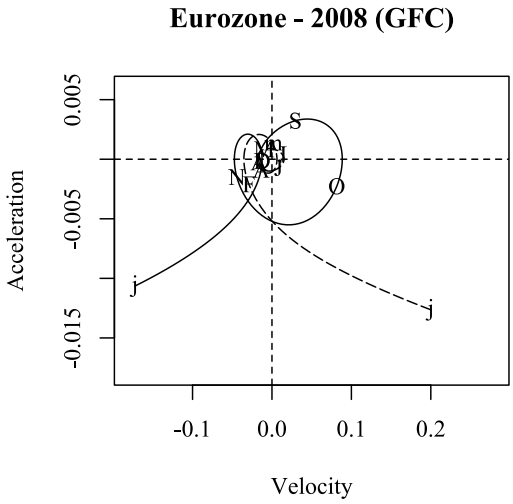


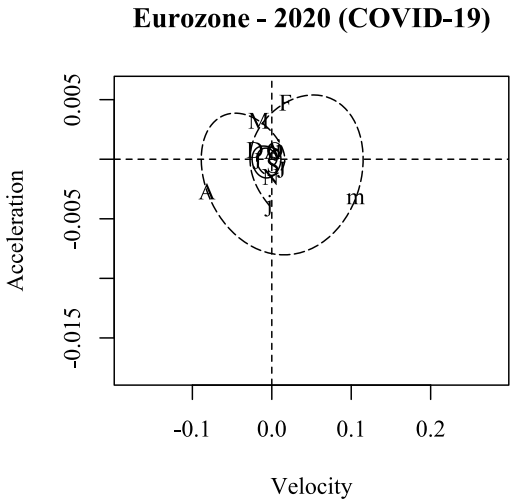
(b) China



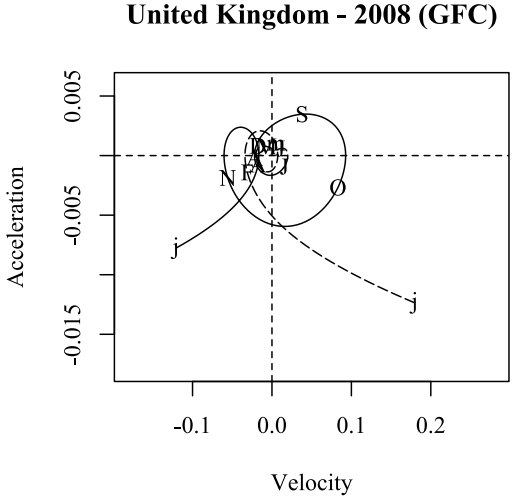
(c) India

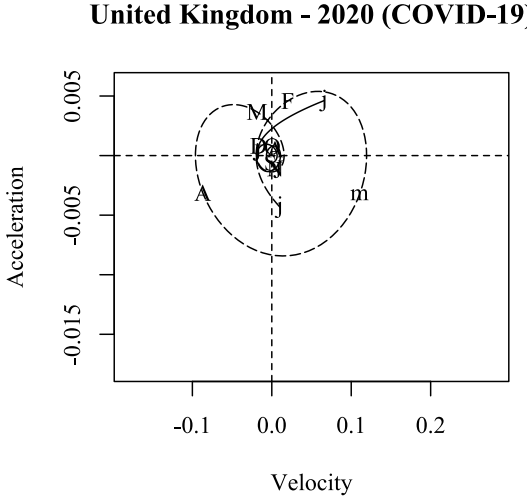
Figure 18: Phase-plane plots showing a comparison of the GFC (2008) and COVID-19 (2020) period for the emerging markets.



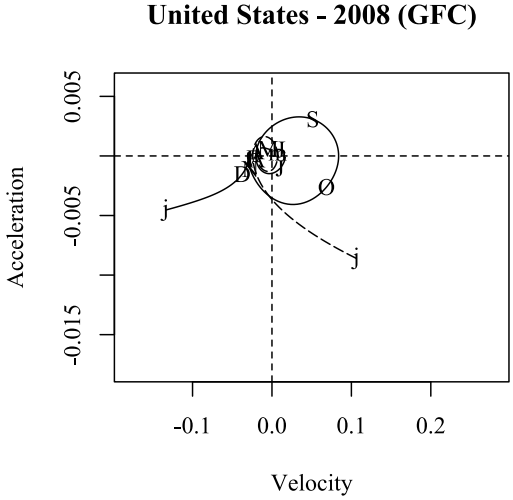


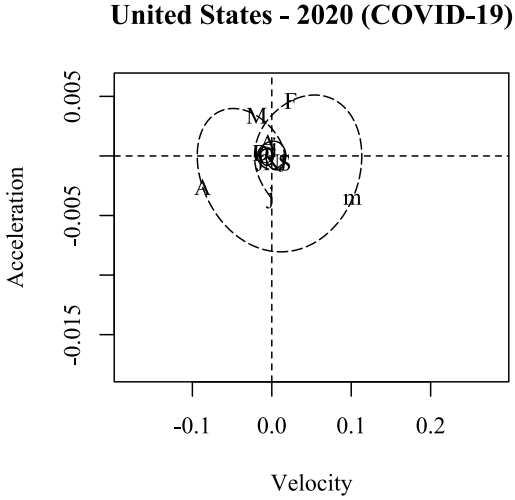
(a) The Eurozone





(b) United Kingdom





(c) United States

Figure 19: Phase-plane plots showing a comparison of the GFC (2008) and COVID-19 (2020) period for the developed markets.

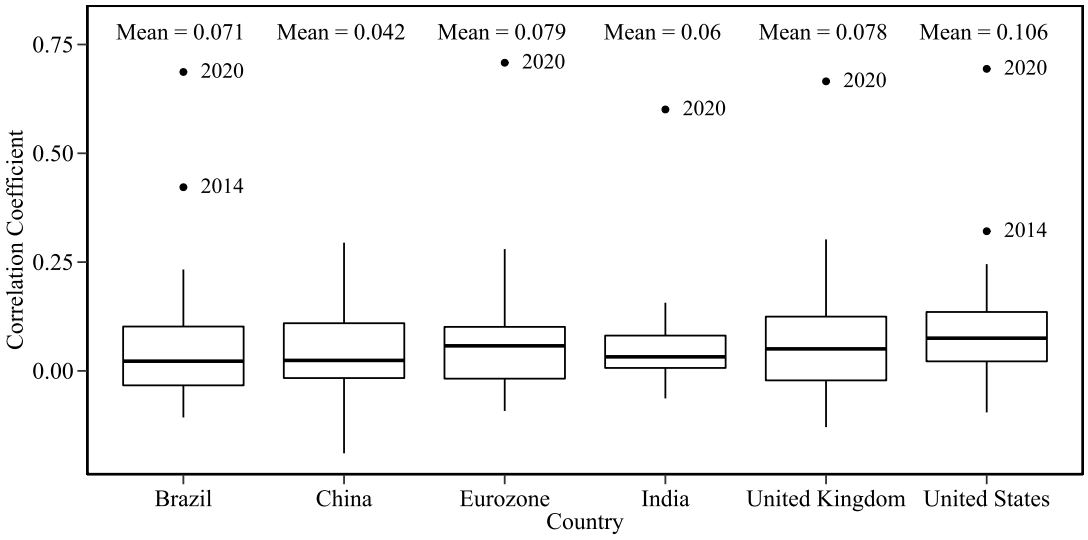


Figure 20: Box plots of the yearly correlation between each market's realised volatility and the EMVID index.

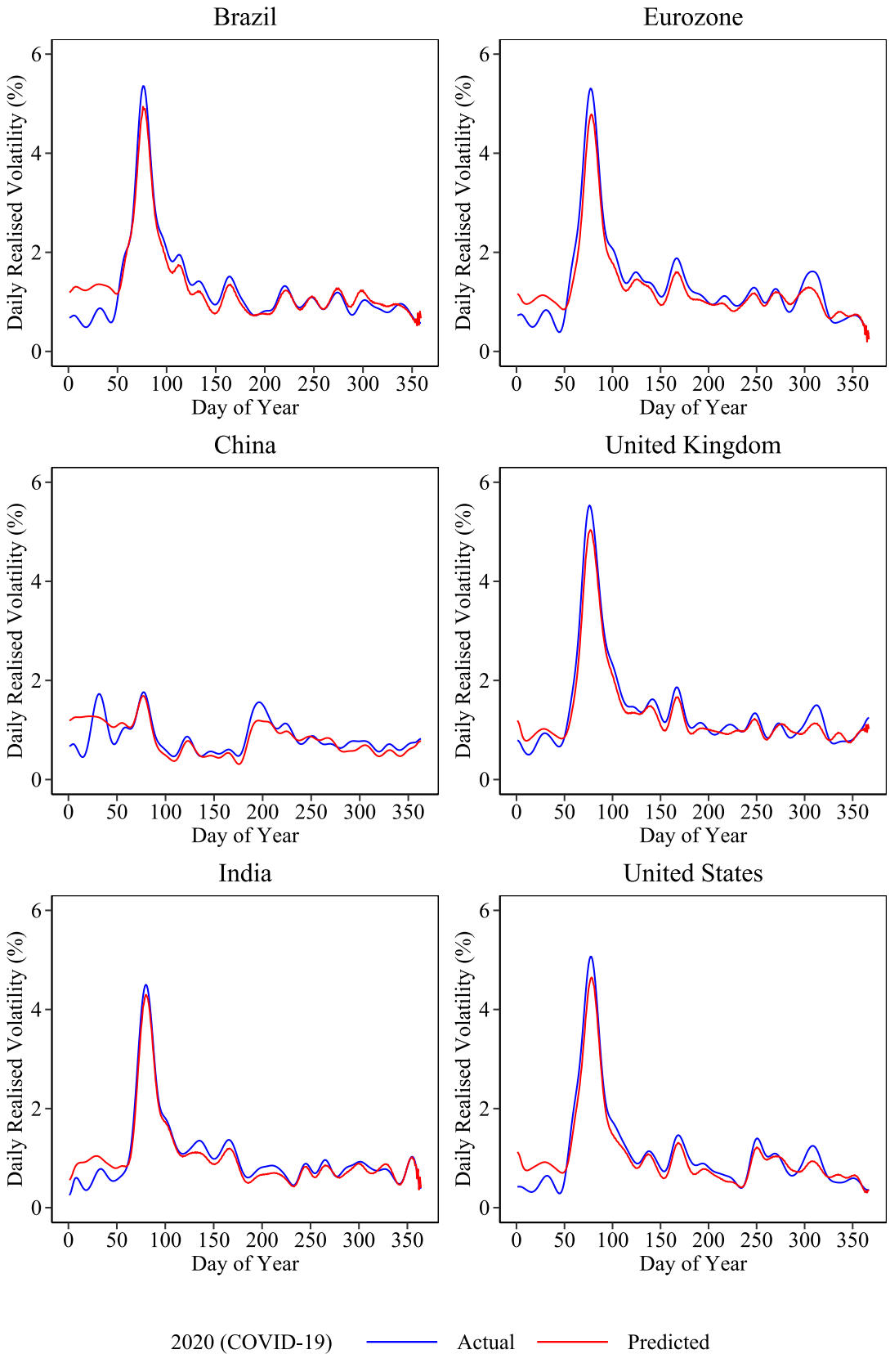


Figure 21: The actual (blue) and predicted (red) daily realised volatility for all six markets during the COVID-19 period.

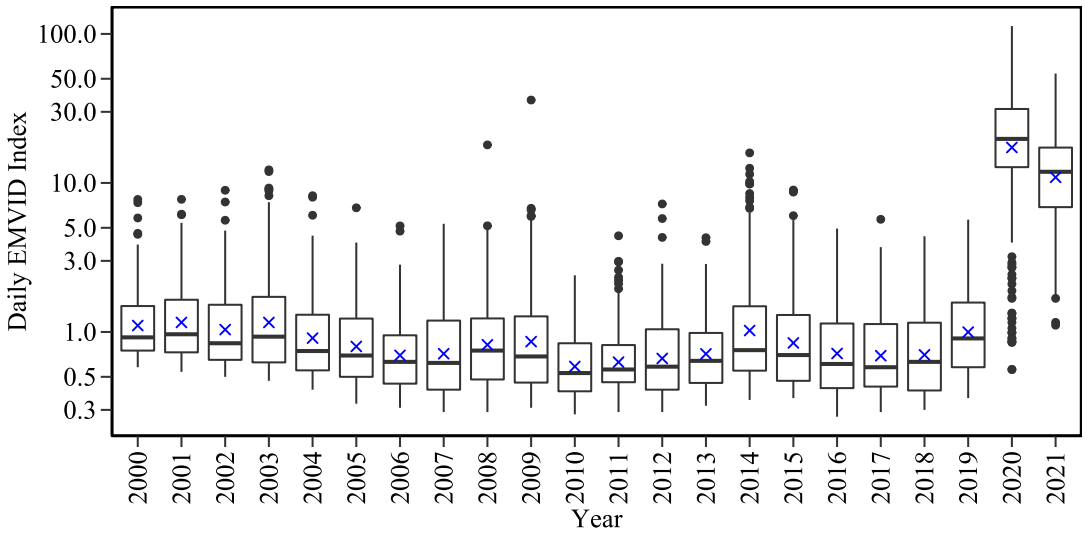


Figure 22: Box plots of the daily EMVID index from $2000 - 2021$.

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

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

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.

Year	**Country**											
	Emerging Markets						**Developed Markets**					
	Brazil		**China**		**India**		**The Eurozone**		**United Kingdom**		**United States**	
	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

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

Year	Country					
	Emerging Markets			Developed Markets		
	Brazil	China	India	The Eurozone	United Kingdom	United States
	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

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

Year	Country					
	Emerging Markets			Developed Markets		
	Brazil	China	India	The Eurozone	United Kingdom	United States
	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

A Algorithms

Algorithm 1: Realised Volatility Data Alignment Function

```
/* RVdata: A market-specific data frame containing the
   misaligned daily realised volatility data                    */
Input: RVdata
/* RVdataAligned: A data frame containing the yearly
   columns of the daily realised volatility, aligned, so
   each year has the same number of days                        */
Output: RVdataAligned
Function RVdataAlignment(RVdata):
    /* Add a year column                                        */
    year ⟵ format(RVdata$Date, "%Y")
    RVdata$year ⟵ year
    /* Add a month-day column                                   */
    monthDay ⟵ format(RVdata$Date, "%m-%d")
    RVdata$monthDay ⟵ monthDay
    /* Create an empty list to store the yearly data            */
    YearList ⟵ list()
    /* Iterate from year 2000 to 2021                           */
    for (i = 2000; i < 2022; i = i + 1) do
        /* Store monthDay and yearly realised volatility
           data                                                 */
        year_i ⟵ subset(RVdata, Year == i)
        year_i ⟵ subset(year_i, select(monthDay, "realised_volatility")
        /* Append this to the list                              */
        YearList ⟵ list(year_i)
    end
    /* Perform outer join to align days of year, insert NA
       where rows have no data                                  */
    RVdataAligned ⟵ OuterJoin(YearList, by = monthDay)
    /* Rename columns from 2000 - 2021, except for the
       monthDay column                                          */
    YearNames ⟵ vector(2000:2021)
    for (j = 2; j < ncol(RVdataAligned); j = j + 1) do
        names(RVdataAligned)[j] ⟵ YearNames[j − 1]
    end
    return RVdataAligned
```

Algorithm 2: EMVID Data Alignment Function

```
/* evmidData: The EMVID data frame for interest                    */
Input: evmidData
/* evmidDataAligned: A data frame containing the yearly
   columns of the daily EMVID index measure, aligned, so
   each year has the same number of days                           */
Output: evmidDataAligned
Function EMVIDdataAlignment(evmidData):
    /* Add a month-day column                                      */
    monthDay ⟵ format(evmidData$Date, "%m-%d")
    evmidData$monthDay ⟵ monthDay
    /* Create an empty list to store the yearly data               */
    YearList ⟵ list()
    /* Iterate from year 2000 to 2021                              */
    for (i = 2000; i < 2022; i = i + 1) do
        /* Store monthDay and EMVID index yearly data              */
        year_i ⟵ subset(evmidData, Year == i)
        year_i ⟵ subset(year_i, select(monthDay, "daily_infect_emv_index")
        /* Append this to the list                                 */
        YearList ⟵ list(year_i)
    end
    /* Perform outer join to align days of year, insert NA
       where rows have no data                                     */
    evmidDataAligned ⟵ OuterJoin(YearList, by = monthDay)
    /* Rename columns from 2000 - 2021, except for the
       monthDay column                                             */
    YearNames ⟵ vector(2000:2021)
    for (j = 2; j < ncol(evmidDataAligned); j = j + 1) do
        names(evmidDataAligned)[j] ⟵ YearNames[j − 1]
    end
    return evmidDataAligned
```

Algorithm 3: Dynamic B-spline Smoothing Function

```
/* alignedDataFrame: The aligned realised volatility or EMVID data
   frame (data to be smoothed)                                        */
/* norder: The order of the B-spline                                  */
/* loglambda: The smoothing parameter value                           */
Input: alignedDataFrame, norder, loglambda
/* fdSmooth: A smooth functional data object                          */
Output: fdSmooth
Function smooth_data_bspline(alignedDataFrame, norder, loglambda):

    /* Drop the MonthDay column                                       */
    dropColumnAlignedData ⟵ subset(alignedDataFrame, select = −c(MonthDay))
    /* Convert to a matrix                                            */
    MyDataMatrix ⟵ as.matrix(dropColumnAlignedData)
    /* Define variables to pass into create.bspline.basis             */
    dataRange ⟵ c(1, dim(MyDataMatrix)[1])
    /* Interior knots placed at every argument value (each day in the
       year)                                                          */
    KnotValues ⟵ c(1: dim(MyDataMatrix)[1])
    norder ⟵ norder
    nbasis ⟵ length(KnotValues) + norder - 2
    fdnames ⟵ list("Days of year", colnames(dropColumnAlignedData))

    /* Create the basis object                                        */
    basisObj ⟵ create.bspline.basis(rangeval = dataRange, nbasis = nbasis, norder = norder,
      breaks = KnotValues)
    /* create the linear differential operator object                 */
    Lfdobj ⟵ int2Lfd(max(0, norder-2))
    /* Create the functional parameter object                         */
    fdParobj ⟵ fdPar(fdobj = basisObj, Lfdobj = Lfdobj, lambda = 10^(loglambda))
    /* Create the coefficients                                        */
    numberCol ⟵ dim(MyDataMatrix)[2]
    coefs ⟵ matrix(0, nrow = nbasis, ncol = numberCol)

    /* Smooth using a roughness penalty and dealing with NA values   */
    for (i in 1:numberCol) do
        curve ⟵ MyDataMatrix[,i]
        SmoothCurve ⟵ smooth.basis(argvals = KnotValues[!is.na(curve)],y =
          curve[!is.na(curve)], fdParobj = fdParobj, wtvec = NULL, fdnames = fdnames,
          covariates= NULL, method = "chol")
        coefs[,i] ⟵ SmoothCurve$fd$coefs
    fdSmooth ⟵ fd(coef = coefs, basisobj = basisObj , fdnames = fdnames)
    return fdSmooth
```

Algorithm 4: The Kinetic and Potential Energy Extraction Function

```
/* fdSmooth: A smooth functional data object                                  */
Input: fdSmooth
/* XYrange: A data frame containing the yearly X-Axis and Y-Axis range        */
Output: XYrange
Function x_y_range_extraction(fdSmooth):

    /* Create an empty x and y range vector to store the values               */
    xRangeVec ⟵ vector()
    yRangeVec ⟵ vector()
    /* Define the number of columns (equal to the number of years)            */
    numberCol ⟵ dim(fdSmooth$coefs)[2]
    /* Vector of argument values (days of the year) to evaluate the fd object at */
    evalArg ⟵ vector(fdSmooth$basis$rangeval[1]:fdSmooth$basis$rangeval[2])

    /* Extract the yearly X & Y range from the phaseplanePlot                 */
    for (i in 1:numberCol) do
        /* Create the phaseplanePlot for a specific year                      */
        tempPhasePlanePlot ⟵ phaseplanePlot(evalarg = evalArg, fdobj = fdSmooth[j], returnMatrix = TRUE)
        /* Detect sign change of the first derivative                         */
        signChangeXaxis ⟵ diff(sign(tempPhasePlanePlot[, 1]))
        /* Find the indices of these sign changes                             */
        XaxisIndices ⟵ which(signChangeXaxis != 0)
        /* Extract the second derivative value at those indices               */
        yVals ⟵ tempPhasePlanePlot[, 2][XaxisIndices]
        /* Find the minimum and maximum value                                 */
        minYVal ⟵ min(yVals)
        maxYVal ⟵ max(yVals)
        /* Calculate the Y-Axis range                                         */
        yRange ⟵ maxYVal − minYVal
        /* Append yRange of that year to the storage vector                   */
        yRangeVec ⟵ vector(yRangeVec, yRange)

        /* Detect sign change of the second derivative                        */
        signChangeYaxis ⟵ diff(sign(tempPhasePlanePlot[, 2]))
        /* Find the indices of these sign changes                             */
        YaxisIndices ⟵ which(signChangeYaxis != 0)
        /* Extract the first derivative value at those indices                */
        xVals ⟵ tempPhasePlanePlot[, 1][YaxisIndices]
        /* Find the minimum and maximum value                                 */
        minXVal ⟵ min(xVals)
        maxXVal ⟵ max(xVals)
        /* Calculate the X-Axis range                                         */
        xRange ⟵ maxXVal − minXVal
        /* Append xRange of that year to the storage vector                   */
        xRangeVec ⟵ vector(xRangeVec, xRange)
    /* Create the years column                                                */
    YearCol ⟵ fdSmooth$fdnames[[2]]
    /* Create a data frame with the results                                   */
    XYrange ⟵ data.frame(YearCol, xRangeVec, yRangeVec)
    return XYrange
```

Algorithm 5: Dynamic Function-On-Function Regression Function

```
/* RValignedDataFrame: The aligned realised volatility data frame
   containing the market-specific data                              */
/* EMVIDalignedDataFrame: The aligned EMVID data frame              */
/* norder: The order of the B-spline                                */
/* loglambda_RV: The realised volatility smoothing parameter value  */
/* loglambda_EMVID: The EMVID smoothing parameter value             */
Input: RValignedDataFrame, EMVIDalignedDataFrame, norder, loglambda_RV,
       loglambda_EMVID
/* fRegressFit: List containing the functional regression fitted
   objects                                                          */
Output: fRegressFit
Function function_on_function_regression(RValignedDataFrame,
 EMVIDalignedDataFrame, norder, loglambda_RV, loglambda_EMVID):

    /* Smooth the market-specific realised volatility data         */
    smoothMarketfd ⟵ smooth_data_bspline(RValignedDataFrame, norder, loglambda_RV)
    /* Remove dates from the EMVID data frame to match RV data frame
       */
    equalDateEMVIDdf <- remove_emvid_dates(EMVIDalignedDataFrame,
     RValignedDataFrame)
    /* Smooth the EMVID index data                                 */
    smoothEMVIDfd ⟵ smooth_data_bspline(EMVIDalignedDataFrame, norder,
     loglambda_EMVID)

    /* Create the xfdlist object                                   */
    constMap ⟵ rep(1, dim(smoothMarketfd$coef)[2])
    emvidList ⟵ list(constMap=constMap, emvidfd=smoothEMVIDfd)

    /* Create regression model objects                             */
    beta1 ⟵ with(smoothMarketfd, fd(basisobj=basis, fdnames=fdnames))
    beta2 <- with(smoothEMVIDfd, fd(basisobj=basis, fdnames=fdnames))
    betalist <- list(constMap=fdPar(beta1), emvidfd=fdPar(beta2))

    /* Fit the regression model                                    */
    fRegressFit ⟵ fRegress(y = smoothMarketfd, xfdlist = emvidList , betalist = betalist)
    return fRegressFit
```

B Realised Volatility Data Set Information

This section contains the metadata for the realised volatility data set used in this research. Table 6 contains a summary of the econometric measures available within the realised volatility data set [3]. Table 7 contains a summary of the 31 market indices in the realised volatility data set and their respective date ranges for which data was available at the time of writing [4]. These tables are used as a reference to explain the data-processing steps required in this research.

[3] These econometric measures were taken verbatim from https://realized.oxford-man.ox.ac.uk/data and the author does not claim it as his own.

[4] These market indices and time intervals were taken verbatim from https://realized.oxford-man.ox.ac.uk/data and the author does not claim it as his own.

Table 6: A summary of the econometric measures within the realised volatility data set [Heber et al., 2009].

Column	Measure
bv	Bipower Variation (5-min)
bv_ss	Bipower Variation (5-min Sub-sampled)
close_price	Closing (Last) Price
close_time	Closing Time
medrv	Median Realised Variance (5-min)
nobs	Number of Observations
open_price	Opening (First) Price
open_time	Opening Time
open_to_close	Open to Close Return
rk_parzen	Realized Kernel Variance (Non-Flat Parzen)
rk_th2	Realized Kernel Variance (Tukey-Hanning(2))
rk_twoscale	Realized Kernel Variance (Two-Scale/Bartlett)
rsv	Realized Semi-variance (5-min)
rsv_ss	Realized Semi-variance (5-min Sub-sampled)
rv10	**Realized Variance (10-min)**
rv10_ss	Realized Variance (10-min Sub-sampled)
rv5	Realised Variance (5-min)
rv5_ss	Realized Variance (5-min Sub-sampled)

Table 7: A summary of the 31 market indices in the realised volatility data set and their respective date ranges for which data is available [Heber et al., 2009].

Symbol	Name	Earliest Available	Latest Available
.AEX	AEX index	January 03, 2000	June 28, 2022
.AORD	All Ordinaries	January 04, 2000	June 28, 2022
.BFX	Bell 20 Index	January 03, 2000	June 28, 2022
.BSESN	S&P BSE Sensex	January 03, 2000	June 28, 2022
.BVLG	PSI All-Share Index	October 15, 2012	March 18, 2022
.BVSP	**BVSP BOVESPA Index**	**January 03, 2000**	**June 28, 2022**
.DJI	Dow Jones Industrial Average	January 03, 2000	June 28, 2022
.FCHI	CAC 40	January 03, 2000	June 28, 2022
.FTMIB	FTSE MIB	June 01, 2009	June 28, 2022
.FTSE	**FTSE 100**	**January 04, 2000**	**June 28, 2022**
.GDAXI	DAX	January 03, 2000	June 28, 2022
.GSPTSE	S&P/TSX Composite index	May 02, 2002	June 28, 2022
.HSI	HANG SENG Index	January 03, 2000	June 24, 2022
.IBEX	IBEX 35 Index	January 03, 2000	June 28, 2022
.IXIC	Nasdaq 100	January 03, 2000	June 28, 2022
.KS11	Korea Composite Stock Price Index (KOSPI)	January 04, 2000	June 28, 2022
.KSE	Karachi SE 100 Index	January 03, 2000	June 28, 2022
.MXX	IPC Mexico	January 03, 2000	June 28, 2022
.N225	Nikkei 225	February 02, 2000	June 28, 2022
.NSEI	**NIFTY 50**	**January 03, 2000**	**June 28, 2022**
.OMXC20	OMX Copenhagen 20 Index	October 03, 2005	June 28, 2022
.OMXHPI	OMX Helsinki All Share Index	October 03, 2005	June 28, 2022
.OMXSPI	OMX Stockholm All Share Index	October 03, 2005	June 28, 2022
.OSEAX	Oslo Exchange All-share Index	September 03, 2001	June 28, 2022
.RUT	Russel 2000	January 03, 2000	June 28, 2022
.SMSI	Madrid General Index	July 04, 2005	June 28, 2022
.SPX	**S&P 500 Index**	**January 03, 2000**	**June 28, 2022**
.SSEC	**Shanghai Composite Index**	**January 04, 2000**	**June 28, 2022**
.SSMI	Swiss Stock Market Index	January 04, 2000	June 28, 2022
.STI	Straits Times Index	January 03, 2000	June 28, 2022
.STOXX50E	**EURO STOXX 50**	**January 03, 2000**	**June 28, 2022**

C Realised Volatility Data Set Structure

The first 10 records from the realised volatility data set are illustrated in Table 8. This table is used as a reference to explain the data-processing steps required in this research.

Table 8: The first 10 records of the raw realised volatility data set.

	Symbol	close_time	open_to_close	open_time	rv10	open_price	rk_twoscale	bv_ss	bv	rv5	medrv	rk_th2	rv10_ss	rsv	rv5_ss	rk_parzen	rsv_ss	nobs	close_price
2000-01-03 00:00:00+00:00	.AEX	163015	-0.000340	90101	0.000178	675.67	0.000103	0.000100	0.000100	0.000130	0.000050	0.000102	0.000178	0.000046	0.000130	0.000179	0.000046	1795	675.44
2000-01-04 00:00:00+00:00	.AEX	163016	-0.033606	90416	0.000261	664.2	0.000199	0.000207	0.000207	0.000201	0.000075	0.000201	0.000261	0.000147	0.000201	0.000423	0.000147	1785	642.25
2000-01-05 00:00:00+00:00	.AEX	163016	-0.001675	90016	0.000714	633.37	0.000325	0.000361	0.000361	0.000491	0.000166	0.000345	0.000714	0.000328	0.000491	0.000324	0.000328	1801	632.31
2000-01-06 00:00:00+00:00	.AEX	163002	-0.013130	90016	0.000182	632.46	0.000218	0.000258	0.000258	0.000225	0.000152	0.000221	0.000182	0.000116	0.000225	0.000219	0.000116	1799	624.21
2000-01-07 00:00:00+00:00	.AEX	163016	0.025013	90046	0.000157	628.93	0.000126	0.000130	0.000130	0.000138	0.000040	0.000123	0.000157	0.000048	0.000138	0.000155	0.000048	1798	644.86
2000-01-10 00:00:00+00:00	.AEX	163017	0.005081	90146	0.000126	651.82	0.000074	0.000089	0.000089	0.000109	0.000032	0.000073	0.000126	0.000062	0.000109	0.000086	0.000062	1794	655.14
2000-01-11 00:00:00+00:00	.AEX	170656	-0.012511	90033	0.000130	657.11	0.000100	0.000124	0.000124	0.000127	0.000068	0.000103	0.000130	0.000079	0.000127	0.000118	0.000079	1795	648.94
2000-01-12 00:00:00+00:00	.AEX	163015	-0.007366	90101	0.000150	643.12	0.000148	0.000154	0.000154	0.000166	0.000085	0.000150	0.000150	0.000092	0.000166	0.000079	0.000092	1797	638.4
2000-01-13 00:00:00+00:00	.AEX	163017	0.003791	90016	0.000104	639.78	0.000138	0.000129	0.000129	0.000154	0.000033	0.000138	0.000104	0.000051	0.000154	0.000127	0.000051	1801	642.21
2000-01-14 00:00:00+00:00	.AEX	163016	0.012061	90045	0.000071	647.74	0.000104	0.000104	0.000104	0.000097	0.000050	0.000104	0.000071	0.000032	0.000097	0.000103	0.000032	1799	655.6

D EMVID Index Data Set Information

This section contains the metadata for the EMVID Index data set used in this research. Table 9 is a summary of the columns within the EMVID index data set. This table is used as a reference to explain the data-processing steps required in this research.

Table 9: A summary of the columns within the EMVID index data set.

Column	**Measure**
day	Day of Month
month	Month of Year
year	The Year
daily_infect_emv_index	Daily EMVID Measure

E EMVID Index Data Set Structure

The first 10 records from the EMVID index data set are illustrated in Table 10. This table is used as a reference to explain the data-processing steps required in this research.

Table 10: The first 10 records of the raw infectious diseases index data set.

day	month	year	daily_infect_emv_index
1	1	1985	0
2	1	1985	0
3	1	1985	0
4	1	1985	0
5	1	1985	0
6	1	1985	7.94
7	1	1985	0
8	1	1985	0
9	1	1985	0
10	1	1985	0

F Smoothing Trial

A smoothing trial exercise was conducted to investigate whether the order of the B-spline basis system makes a difference in the accuracy of the smoothing process. The realised volatility data were smoothed using a fourth, fifth and sixth-order B-spline across a range of smoothing parameters from $\lambda = 10^{-3}$ to $\lambda = 10^{4}$. Once the data had been smoothed, a random 5% sample was extracted, and the root mean squared error (RMSE) value was calculated for this sample. This step was done over 25 trials, and the RMSE was saved for each trial. A histogram of the RMSE trials was generated for each permutation of order and smoothing parameter across all six markets. The histograms are summarised in Figures 23 to 28. The RMSE is shown in log_{10} format, as the actual numbers were too small to fit elegantly on each plot. The red line on the X-axis represents the mean RMSE over the 25 trials.

It is evident, from the figures, that at low values of λ, the lower order B-spline basis system has a smaller RMSE value. Thus, the fourth-order B-spline basis system is the most accurate at low smoothing parameter values. However, as λ increases, the average RMSE value for each order B-spline is similar. Therefore, if a larger λ value is used in the smoothing process, there is little difference in accuracy between different B-spline orders.

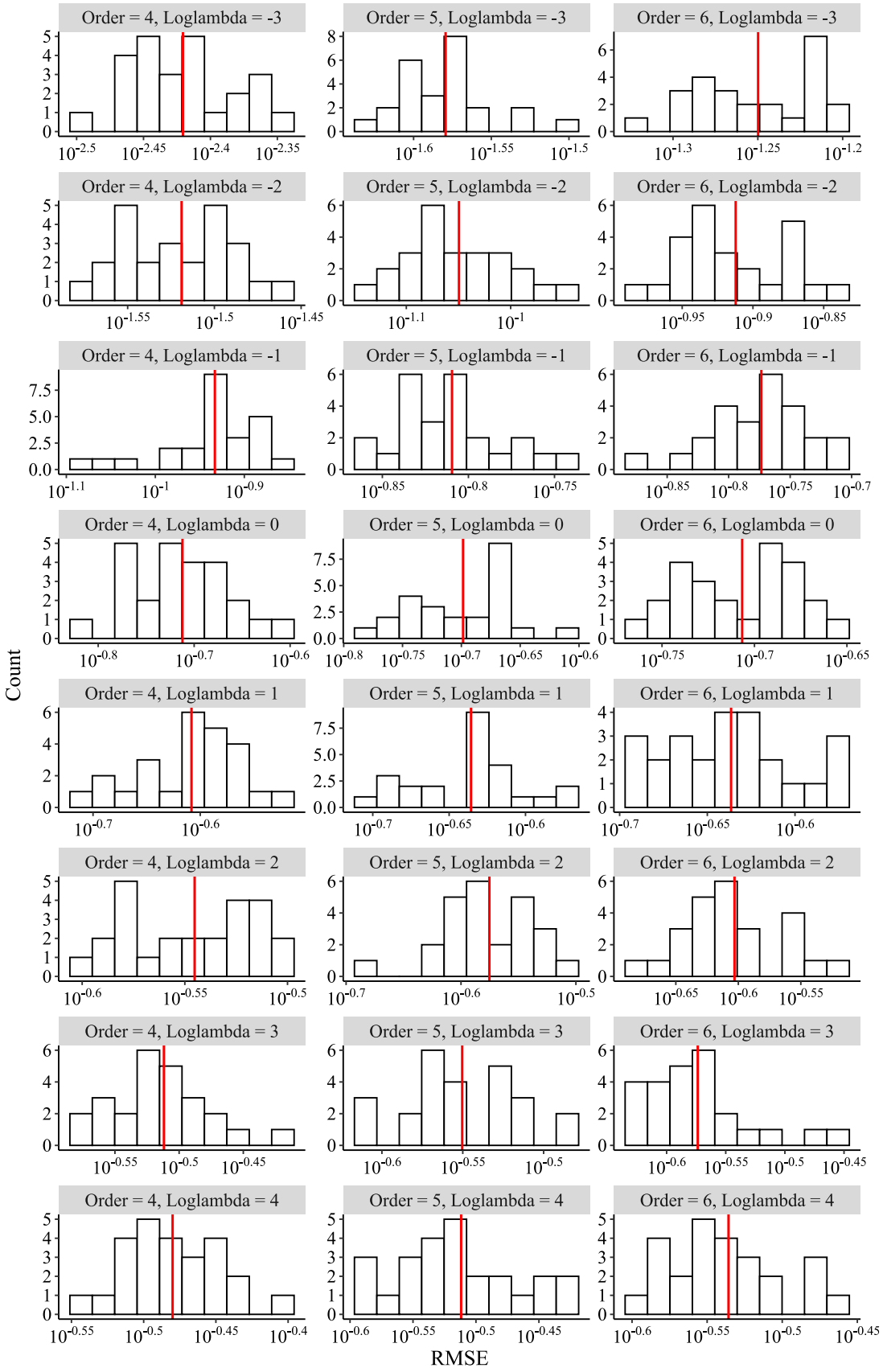


Figure 23: The RMSE of the smooth realised volatility data (5% sample) with different combinations of smoothing variables for Brazil.

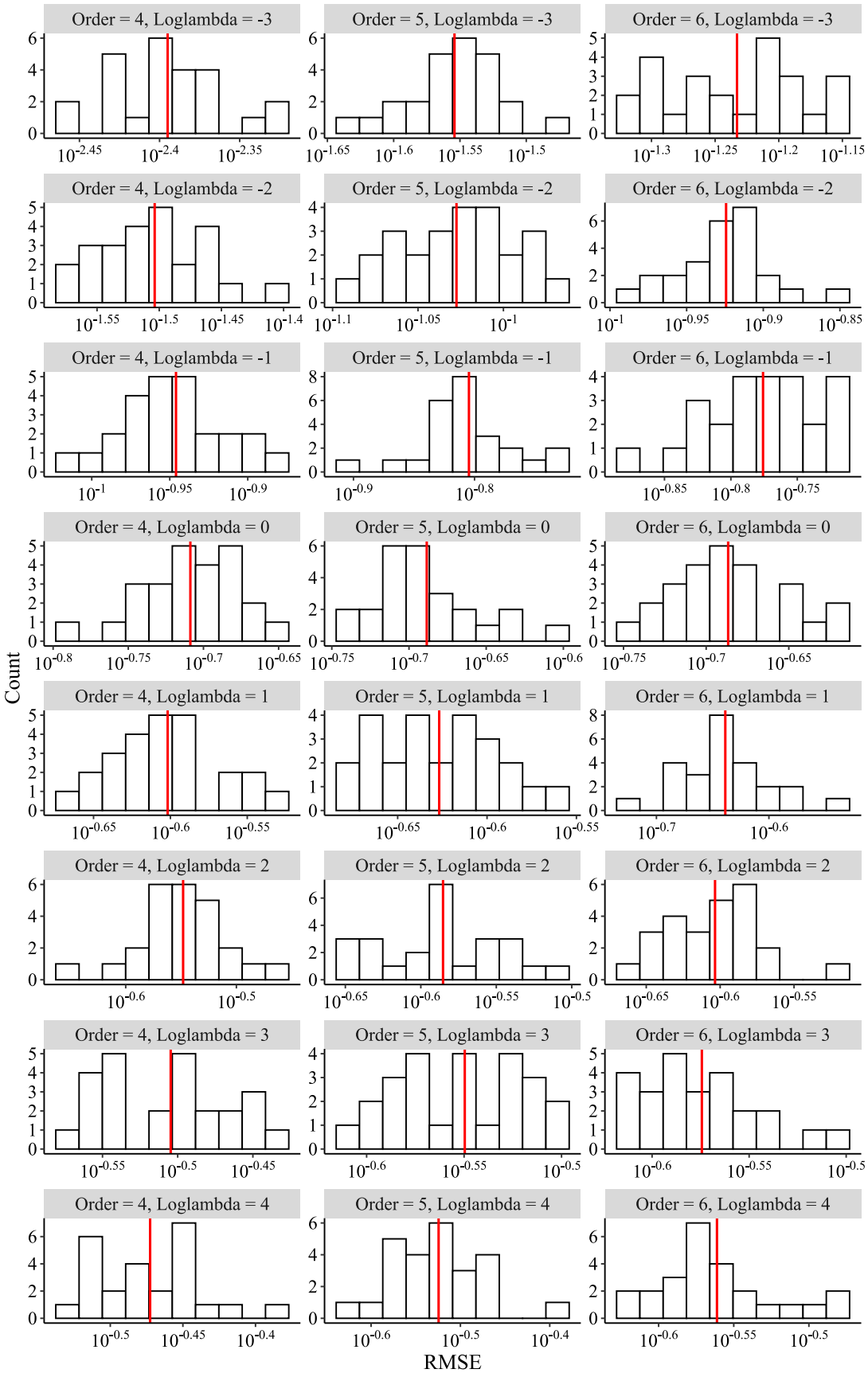


Figure 24: The RMSE of the smooth realised volatility data (5% sample) with different combinations of smoothing variables for China.

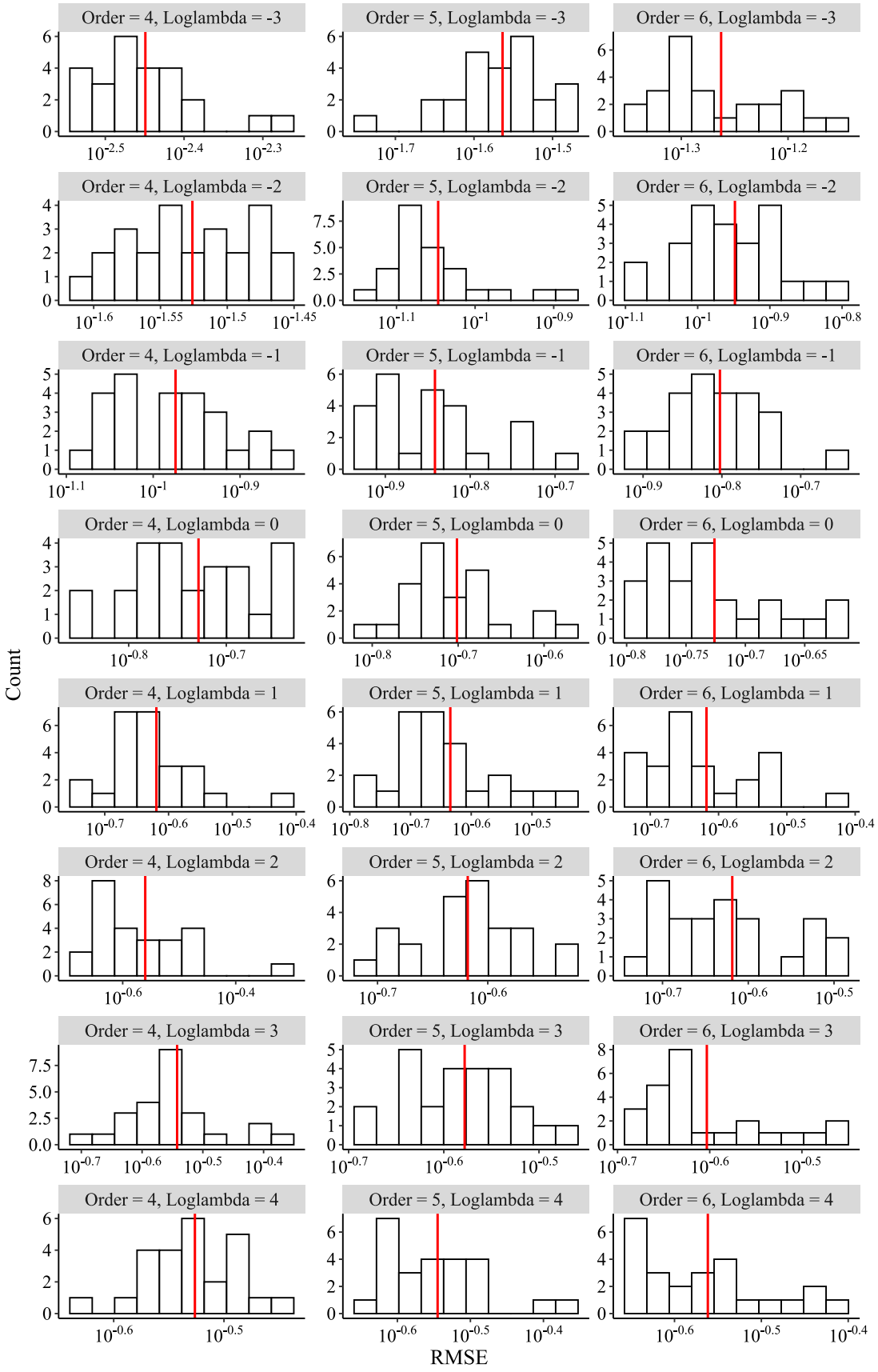


Figure 25: The RMSE of the smooth realised volatility data (5% sample) with different combinations of smoothing variables for India.

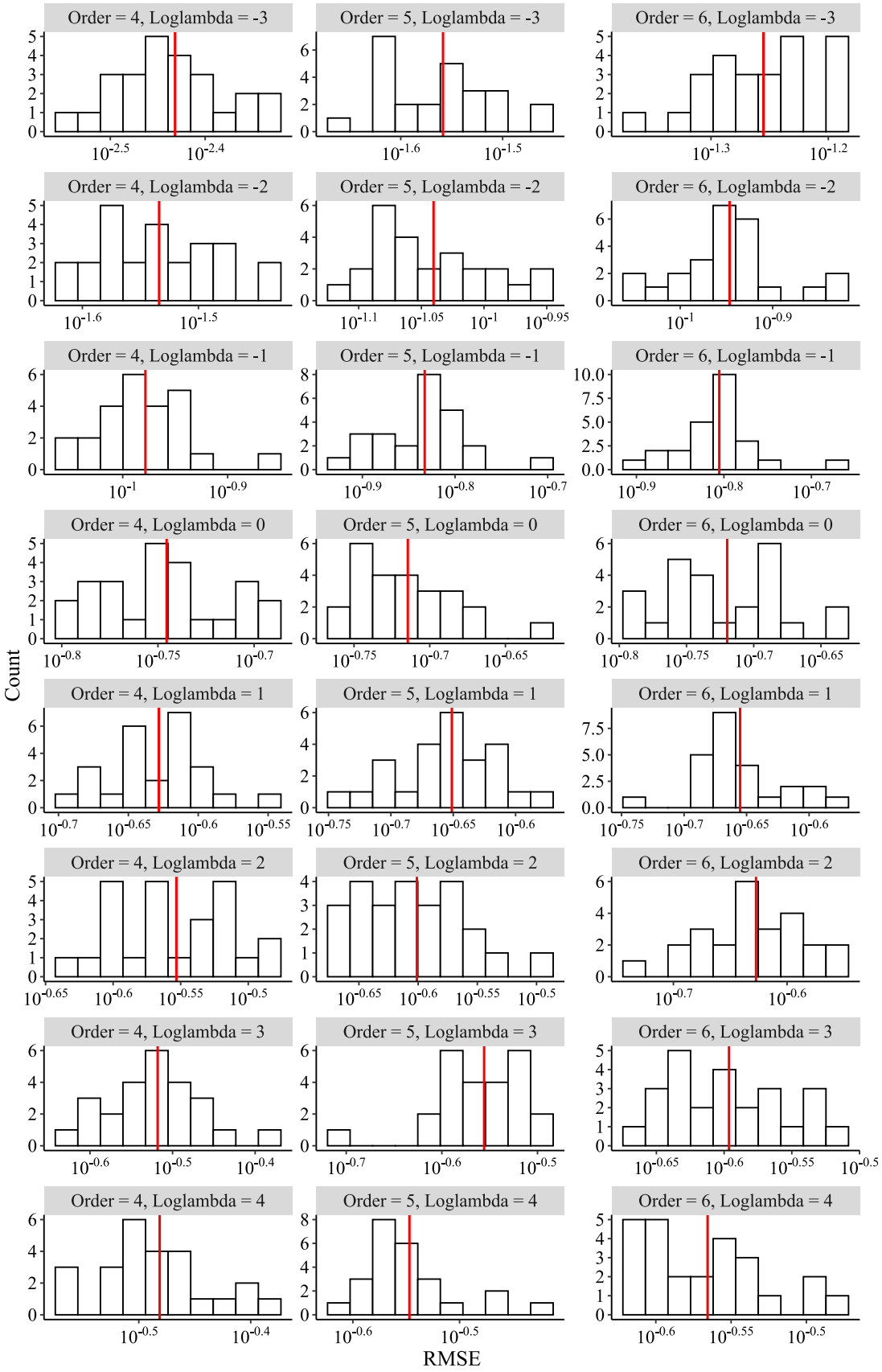


Figure 26: The RMSE of the smooth realised volatility data (5% sample) with different combinations of smoothing variables for Eurozone.

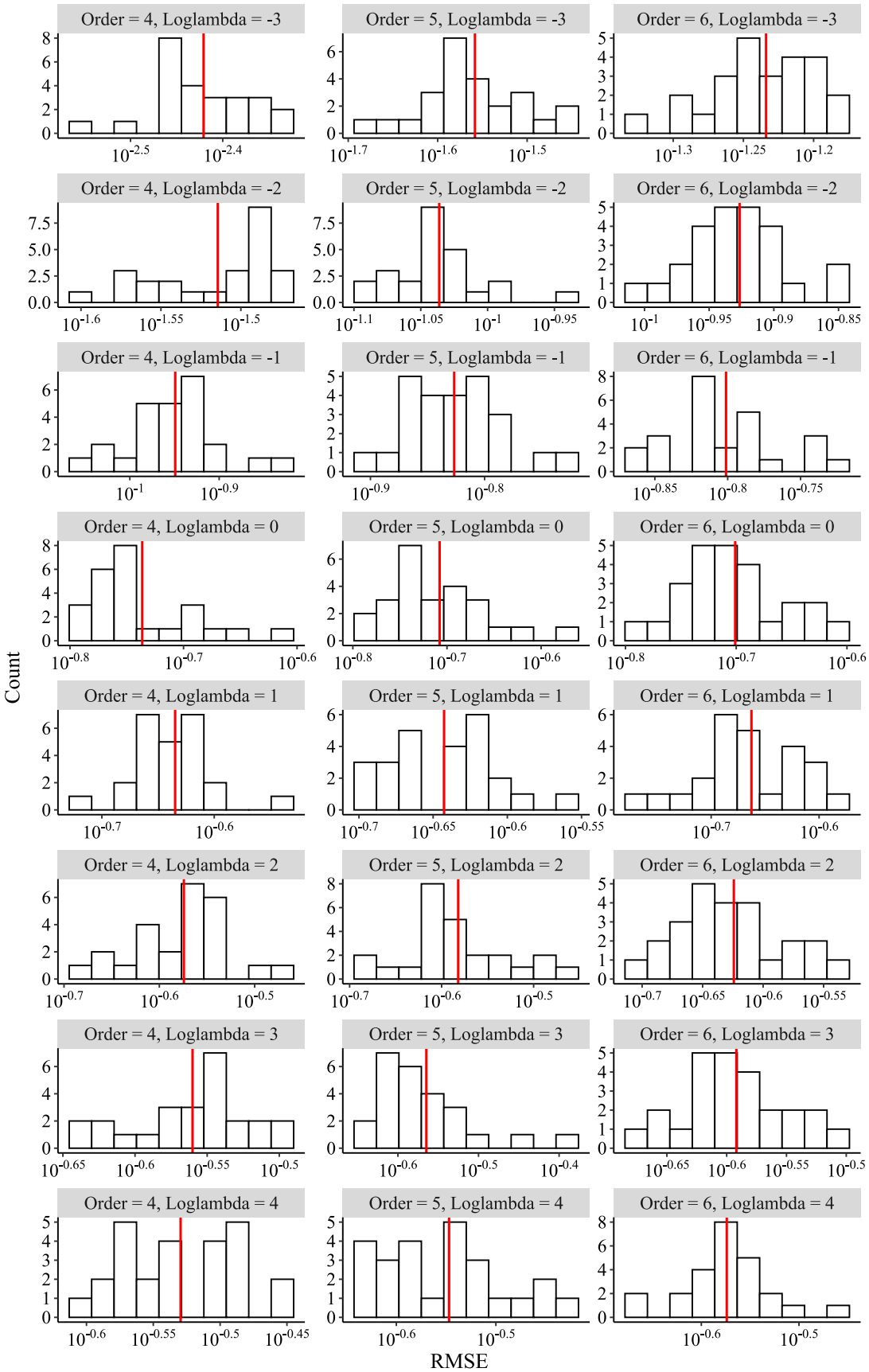


Figure 27: The RMSE of the smooth realised volatility data (5% sample) with different combinations of smoothing variables for the United Kingdom.

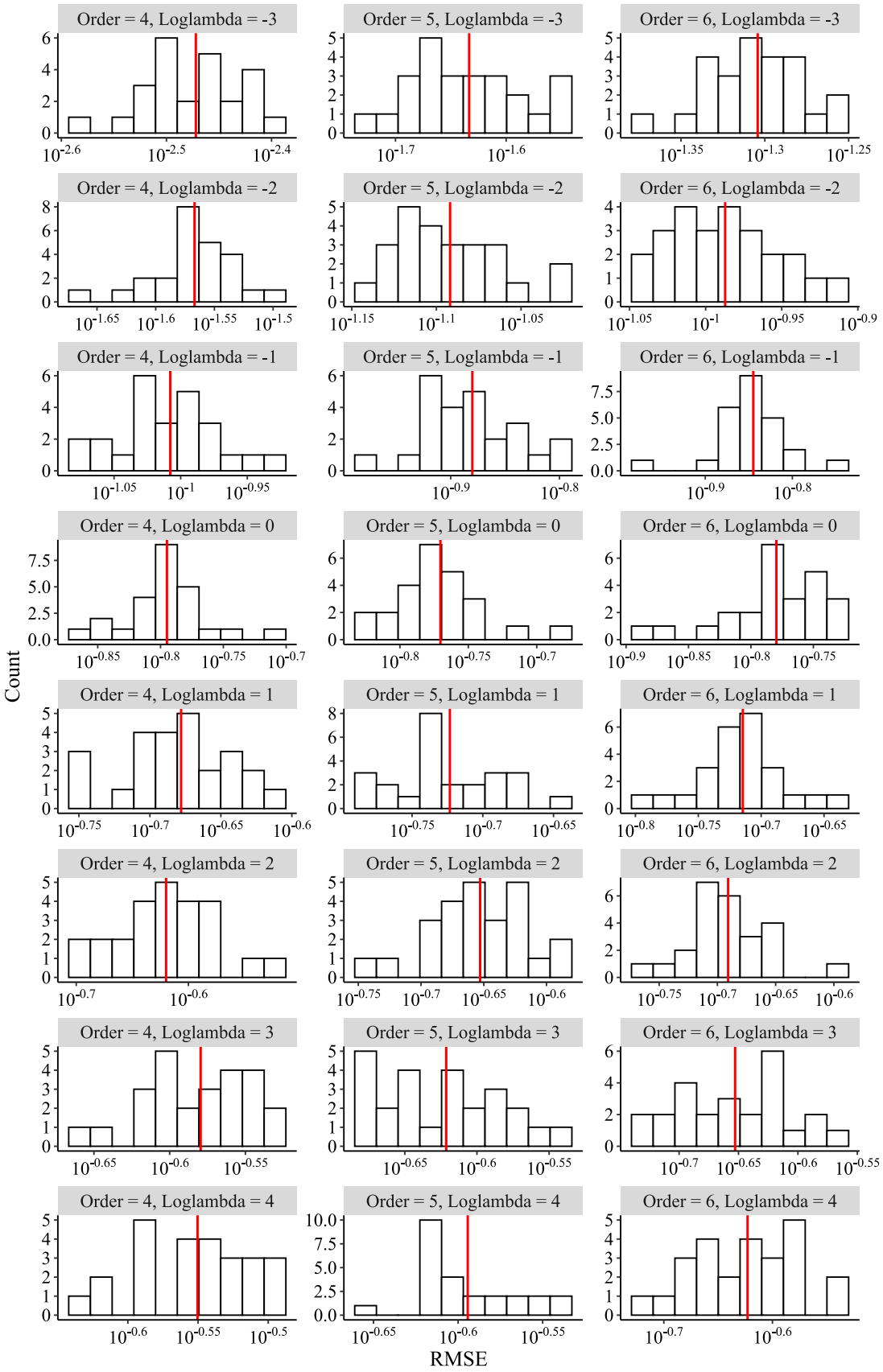


Figure 28: The RMSE of the smooth realised volatility data (5% sample) with different combinations of smoothing variables for the United States.

G Functional Regression

This section contains the additional results for the functional regression analysis. The functional regression was done for each of the 22 years. The functional regression results for 2000 - 2021 are given in the full dissertation. The functional regression result for 2008 is in Figures 29.

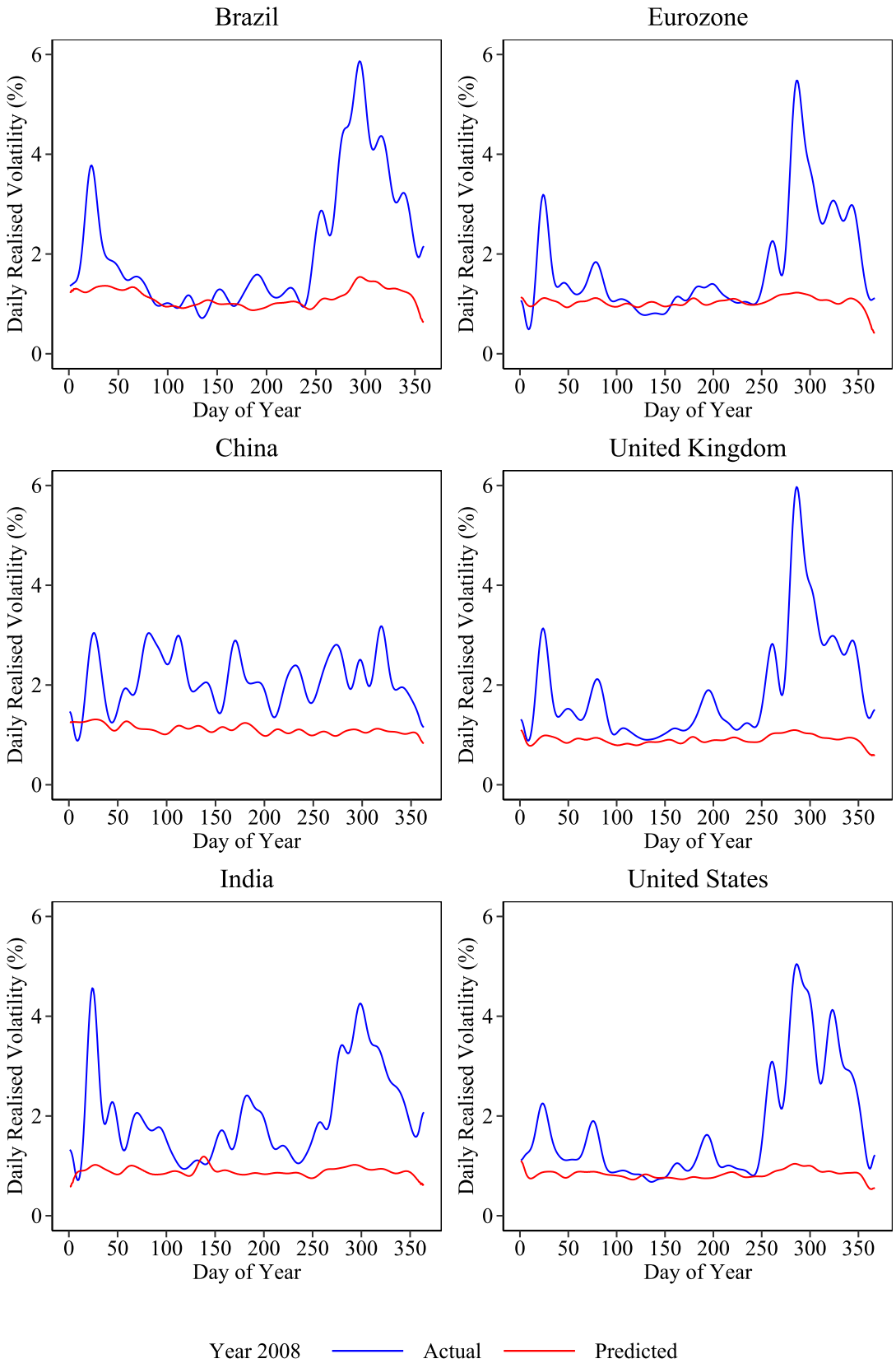


Figure 29: The actual (blue) and predicted (red) daily realised volatility for all six markets during the year 2008.