

Article

Contagion Patterns Classification in Stock Indices: A Functional Clustering Analysis Using Decision Trees

Jorge Omar Razo-De-Anda ¹, Luis Lorenzo Romero-Castro ² and Francisco Venegas-Martínez ^{1,*} 

¹ Escuela Superior de Economía, Instituto Politécnico Nacional, Mexico City 11350, Mexico; jorgerazodeanda@gmail.com

² Facultad de Economía, Contaduría y Administración, Universidad Juárez del Estado de Durango, Durango 34000, Mexico; castroluislorenzo@gmail.com

* Correspondence: fvenegas1111@yahoo.com.mx

Abstract: This paper aims to identify the main determinants of the countries that present contagion during the period 2000–2021, based on the determination of the behavior patterns of 18 stock market indices of 15 of the main economies. To do that, first, the B-spline method and Bezier curves are used to smooth observations by minimizing the noise. Subsequently, the Functional Principal Component Analysis (FPCA) methodology is applied. Then, the K-means clustering algorithm is used to determine the main groups using the silhouette method and cross-validation, considering the sum of squares of the distances as the function to minimize. Finally, classification trees and macroeconomic and financial analyses are used to determine the rules of variables that give a direct explanation of the contagion (clustering) between the stock indices. The main empirical results obtained suggest that the most significant macroeconomic variables are the Gross Domestic Product, the Consumer Price Index, and Foreign Direct Investment, while in the financial aspect and the most representative are Domestic Credit and number of companies listed on the stock market. It is worth noticing that government spending does not have a significant effect at any time as a determinant of contagion. Finally, it is important to mention, and surprising, that Mexico's IPC was not clustered in the same group of US stock market indices anytime, despite the strong commercial relationship and the geographical closeness.

Keywords: functional principal component analysis; cluster analysis; contagion; stock index patterns

MSC: 62J12; 62P05; 91G15



Citation: Razo-De-Anda, J.O.;

Romero-Castro, L.L.;

Venegas-Martínez, F. Contagion

Patterns Classification in Stock

Indices: A Functional Clustering

Analysis Using Decision Trees.

Mathematics **2023**, *11*, 2961. [https://](https://doi.org/10.3390/math11132961)

doi.org/10.3390/math11132961

Academic Editors: Daniel Sevcovic

and Matthias Ehrhardt

Received: 20 April 2023

Revised: 26 June 2023

Accepted: 29 June 2023

Published: 3 July 2023



Copyright: © 2023 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license ([https://](https://creativecommons.org/licenses/by/4.0/)

[creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/)

[4.0/](https://creativecommons.org/licenses/by/4.0/)).

1. Introduction

Increased interactions among countries around the world, either due to the increased flow of capital among global economies or as a consequence of the increased exchange of goods and services among them, have brought many benefits to economic agents. In this sense, agents have at their disposal greater goods and services, including financial assets, derived from the increase in transactions by both national and foreign companies. While this situation has been beneficial for global economic growth, certain negative aspects have also emerged as a result of the interaction and dependence among economies; see, for example, Zhou et al. [1] In this setting, it seems that the consequences of unstable economic phenomena that a country suffers tend to spill over or spread to others. These consequences are reflected almost immediately in international trade transactions or in the prices of financial assets. Therefore, the picture of the macroeconomic and financial systems in one shot can help to explain and identify the contexts of economic crises, derived from external factors, such as war conflicts, pandemics, or policy decisions by local governments. There is some evidence of contagion across major stock markets during crisis episodes, as shown in BenMim and BenSaida [2]. On the other hand, Sheng Kao et al. [3] and Shun Kao et al. [4] also show evidence before and after the subprime crisis effects of contagion

among stock markets. More recent studies, such as the one by Okorie and Lin [5], analyze the effects of stock market contagion in the COVID-19 crisis. Finally, a recent study about the effects of contagion in several economic crises is revised in Yarovaya et al. [6]. That is why a stock index may serve as a thermometer for the economic and financial situation of a country.

The study period of this research covers from 2000 to 2021, with monthly observations for the case of the price of stock indices, while annual observations for the same time horizon are used for macroeconomic and financial variables. This research considers a sample of 15 of the major countries: the United States, Mexico, Brazil, Argentina, Chile, and Canada; Europe, the United Kingdom, France, Spain, Germany, Switzerland, and the Netherlands; and Asia, Japan, China, and Russia. In this sense, the study of Akhtaruzzman et al. [7] examines the effects of periods of economic and financial crises on the stock markets of the Chinese economy. It is also important to mention that many studies show that behavior patterns tend to intensify in periods of economic and financial crises, so special emphasis is placed on the analysis of periods marked in history by financial stress, such as the subprime crisis of 2008 and the recent economic crisis caused by the COVID-19 pandemic in 2020. Examples of these behavior patterns can be found in Mohti et al. [8], Zorgati et al. [9], and Banerjee [10], where these authors find empirical evidence that contagion among countries is aggravated in periods of financial and economic instability.

Due to the above, this investigation aims to identify, firstly, if there is a pattern among the stock market indices of some of the main world economies and, secondly, to verify if this pattern tends to intensify, displacing the stock indices to behave equally in periods of crisis, which could be translated as contagion. However, the above is still up for debate. Haile and Pozo [11] show that there are two types of channels to explain contagion, namely the fundamental channel and the investor behavior channel. More generally, Bekaert et al. [12] mention that there are six types of information transmission channels: (i) relations of the international banking sector at the country level; (ii) country-specific policy responses to the crisis; (iii) trade and financial links; (iv) information asymmetries and informational flows; (v) domestic macroeconomic fundamentals; and (vi) investor contagion. These authors test the wake-up call hypothesis, which states that a crisis initially restricted to one market segment or one country provides new information that may prompt investors to reassess the vulnerability of other market segments or countries; this triggers the spread of the crisis across markets and borders. Given the ideas above, we believe that the fundamental channel is a major contributor to contagion in a crisis episode. That is why we use the fundamental channel to seek the relationship between contagion patterns and the macroeconomic variables of a country. In this way, it could be taken into consideration so that a country in a specific pattern could emigrate to another with greater benefit from imposing changes that tend to modify its own macroeconomic variables. Some works, such as those by Gkillas et al. [13] and Ye et al. [14], show empirical evidence that macroeconomic variables can determine the size of the contagion. In this research, we take Altınbaş et al. [15] as a reference to examine contagion through the fundamental channel, considering macroeconomic variables that could potentiate this situation. Among the core economic variables that can have an impact on contagion are industrial production, the inflation rate, the exchange rate, and government expenditure. In the same way as Altınbaş et al. [15], we propose the use of machine learning tools because they are more efficient at determining the importance of the variables when the relationships are non-linear: We then proceed to use the combination of qualifying trees to determine the rules that must be considered for the determination of a group, and, additionally, we perform a multinomial regression with regularization of the RIDGE and LASSO types to determine which are the main groups. Unlike the analysis presented in Altınbaş et al. [15], we proceed to review the correlations over time with the help of the principal components in their functional version. That is why in this research, in addition to detecting contagion through stock indices, we focus on determining the possible financial and macroeconomic variables that can deepen it. The question that this research will answer is: What financial and

macroeconomic variables can reveal the strength or weakness of the contagion pattern among economies in periods of crisis?

Therefore, this investigation provides a direct explanation of the countries that exhibit contagion (grouping or clustering) of the stock market indices based on the analysis of the behavior pattern of 18 stock market indices of the main economies of the world using Functional Principal Components Analysis (FPCA), the K-means clustering algorithm, and classification trees.

This paper differs from the current literature in the following aspects: (1) it provides an efficient alternative to the treatment of observations unlike others since it has a robust analysis of the relationships of observations in a functional way over time; (2) it identifies the causes under which a behavior pattern shows up using cluster tools, the grouping of stock indices is performed by K-means considering the scores resulting from the FPCA; (3) it considers the recent crisis caused by the COVID-19 pandemic in 2020; (4) it identifies the determining macroeconomic and financial variables that explain the formation of each of the groups and the behavior patterns of their members by using classification trees which determines the most important variables for each of the clusters and the rules under each one is grouped derived from the combination of certain macroeconomic variables; finally (5) it performs a multinomial penalized regression model of the RIDGE and LASSO type as suggested by Friedman et al. [16] which has three objectives, confirm the result of the classification trees, determine the importance of each variable, assessing the impact on the probability for the determination of each group.

This research is organized as follows: Section 2 provides a detailed review of the specialized literature, emphasizing the different approaches and tools, the various methodologies used to study this phenomenon, highlighting their advantages and limitations, and the relationship between the financial system and economic activity; Section 3 examines the patterns among economies and determines relevant factors and the real link; Section 4 provides the methodologies that will be used for the investigation in detail: B-splines, Bezier curves, and FPCA; Section 5 states the main results that show empirical evidence of contagion via the behavior of the stock indices and how this intensifies in periods of economic instability by using K-means and sliding windows and using classification and econometric analysis; finally, Section 6 gives the conclusions and acknowledges limitations.

2. Contagion Literature Related to Financial Link: Different Study Approaches and Tools

The importance of the financial system lies in the funding activities carried out as well as the various indicators used as a basis for the decision-making of economic agents in relation to savings and investment. Direct consequences on the local and global economic systems due to the commercial and diplomatic ties between various countries can be reflected in the financial system. Therefore, transferred effects from one economy to another can be identified by variables that are typical of the financial system materialized in a profit or a loss. The term used in the literature to refer to transferred negative effects is called contagion, referring to the transmission of a disease or condition in a health context.

Some authors propose more clearly a formal contagion definition in the context of economic and financial systems, such as Pericoli and Sbracia [17] and Toribio-Dávila [18], who agree that contagion refers to any type of transmission of economic disturbances from one country to another, beyond those caused by fundamental links. In this sense, the World Bank integrates three types of fundamental links. Specifically, for the financial link case, defined as the relationship that exists between two or more economies through the capital market, we can find evidence on this in the studies of Rojas and Chamorro [19]. These authors try to explain the convergence of countries members of the Market Integration in Latin America (MILA) and its short- and long-term consequences using a Vector Autoregressive (VAR) Model. In the same context, Sosa and Ortiz [20] analyze the global financial crisis impact on the countries members of the North American Free Trade Agreement (NAFTA) through their stock indices from January 2003 to February 2015. For this purpose,

sliding windows and symmetrical and asymmetrical Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models were used. The main results show symmetrical volatility in indices and an increase in volatility after the subprime stock market crisis. The close relationship between the Mexico IPC index and the S&P was also evident, followed by the FTSE and the S&P. In the same way, Gavidia-Pantoja [21], through varying copulas, examines the relation between international financial crises and the behavior of Latin American markets (Mexico, Brazil, Colombia, Argentina, Chile), considering four financial crises periods such as the European, American, Asian, and Mexican. The conclusions show evidence of structural changes in volatility during periods of different crises. Similar results can be observed in Bucio-Pacheco et al. [22] and Díaz-Rodríguez y Bucio [23]. More recently, Santillán-Salgado et al. [24] analyzed the dependence between the Merval (Argentina), IPSA (Chile), IBOVESPA (Brazil), and IPC (Mexico) stock indices using dynamic copulas. The results show that the IPC index and the IBOVESPA index have greater dependence compared with the other indices, with values of 52.96% in the upper tail and 46.15% in the lower tail.

Piffaut and Miró [25] analyze the possible contagion that may exist among the main financial markets of Asia, the United States, and Europe from 1995 to 2016, using a Dynamic Conditional Correlation (DCC)-GARCH model and different cointegration tests such as those of Granger and Johansen. The results show that the S&P 500 is closely related to the European, Shanghai Composite Index, and Japan's Nikkei 225 indices, while Akhtaruzzaman et al. [26] use a similar methodology (VARMA DCC-GARCH model) examine how financial contagion occurs in financial and non-financial companies, specifically in G7 countries and the Chinese economy during the COVID-19 period, showing evidence that correlations between stock returns for both financial and non-financial companies increased considerably during the pandemic. The magnitude of the increase in correlation was greater for financial companies, implying that they have a greater impact on the transmission of financial contagion than non-financial companies. More recent papers dealing with risk contagion in financial crises can be found in Vortelinos et al. [27] and Uddin et al. [28]. In addition, Ji et al. [29] use a DCC-MGARCH model to study the contagion effect of financial markets in a crisis. More evidence on this topic can be found in Davidescu et al. [30], Bildirici et al. [31], and Ramírez et al. [32].

One of the disadvantages of the copula methodologies is that commonly the analysis is carried out in pairs, where the dependence of A to C can be biased by the behavior of B, resulting in the same effect that cannot be observed. Additionally, dependence is commonly measured in queues; however, the pattern of contagion behavior can occur even in different quartiles of the distribution. In the case of econometric models that seek to estimate the effect on the mean, they have the same problem. Some efforts to solve these situations have been made through the implementation of dynamic correlation models and their direct effect on conditional volatility. However, these models can be difficult to calibrate and represent complexity for the estimation of parameters by increasing the number of variables in models of simultaneous equations such as those of multivariate ARCH families.

To solve the above problem, an alternative is to identify factors or patterns among the series, considering their dynamic correlation. Additionally, reducing the dimensionality of the problem allows for greater efficiency in terms of computational resources for estimation and calibration. A tool that meets these characteristics is the principal component. The usefulness of this technique can be summarized in two central points. First, it allows representing in an efficient way a small-dimension mapping of the system. In this sense, principal components analysis is the first step to identifying possible "latent" or unobserved variables. Secondly, it allows, from the original variables, the construction of new uncorrelated variables. Carrion-i-Silvestre and Villar [33] applied this technique to obtain several channels of contagion among countries. The results showed evidence of financial contagion after the 2007 crisis in the stock markets of the 21 most industrialized countries, such as Germany, the United States, France, Japan, and the United Kingdom.

Despite the benefits, one disadvantage is that the traditional principal components are static, so the analysis of correlations does not change over time. However, this is solved in Ramsay and Silverman [34] with the functional version of this technique. The main idea is to replace vectors with functions, matrices with compact linear operators, covariance matrices with variance operators, and scalar products in vector space with scalar products in integrable-square functional space (Lin Shang [35] p. 14). Some attempts to implement functional analysis can be found in Dewandaru et al. [36]. Considering a functional approach through wavelets, the authors show evidence of financial contagion among the capital markets of Hong Kong, Australia, and Japan during the 12 major global crises. One problem with this methodology is related to the issue of separation into factors. Functional Principal Component Analysis makes it easier to analyze the various forms of contagion transmission between markets. Likewise, this tool eliminates the problems associated with the omission of relevant variables and the estimation of simultaneous equations. So far, there are few studies in the literature that seek to determine patterns of contagion from the use of traditional and functional principal components. On the other hand, despite the vast literature that indirectly evidences contagion around the world and in recent years, there are very few authors who focus on explaining the variables that promote this phenomenon.

3. Contagion Patterns among Economies and Economic Activity

The World Bank mentions another type of fundamental link: the real link. This is defined as those derived from relationships associated with international transactions or associated with foreign direct investment flows. Within the literature are studies that demonstrate the strong relationship between the Financial System and economic activity. Hassan et al. [37] provide evidence of the positive relationship between financial development and economic growth in countries classified by geographical areas for low-income countries and an opposite relationship for high-income economies, using variables such as the growth rate of the Gross Domestic Product (GDP) per capita as a proxy variable of economic growth, and on the side of financial development the variables of credit granted by the banking sector as a percentage of GDP, domestic credit of the private sector as a percentage of GDP and, finally, monetary aggregates (M3) as a proportion of GDP. In the same context, De la Cruz-Gallegos and Lizárraga [38] intend to verify the possible relationship between the banking system and economic growth in Mexico. For this purpose, the sectoral components of the Global Index of Economic Activity (IGAE, for its acronym in Spanish) were used as economic growth. On the side of the banking sector, credit provided to the same sectors of the IGAE (primary, secondary, and tertiary sectors) was used. Among the conclusions, the authors show a long-term positive relationship with economic growth. Lezama-Palomino et al. [39] tried to verify the existence of a relationship between economic growth and the behavior of the Stock Exchange in Colombia using variables such as GDP growth, the General Index of the Colombian Stock Exchange (IGBC for its acronym in Spanish), Trading volume on the IGBC (VOLIGBC), Profitability of Market Concentration (MC), and the Consumer Price Index (CPI). One of the main results was that stock market behavior in Colombia positively affects economic performance. Landa-Díaz and Silva-Barrón [40] also examine the magnitude of the impact of financial development on economic growth in the cases of Argentina, Chile, Mexico, Colombia, and Brazil. This study considers variables like GDP per capita, Foreign Direct Investment, exports, imports, human capital, and economic growth, while with respect to financial indicators, market capitalization, stock turnover index, total value of bank assets, administrative cost, total value of shares traded on the stock exchange, and domestic credit to the private sector granted by the bank are incorporated. In summary, there is evidence showing how GDP growth and financial development are related. In addition, international trade drives the expansion of economic growth. Summarizing, there is evidence that shows how GDP growth and financial development are related. In addition, international trade drives the expansion of economic growth.

For the European region, Bernádez-Castrejón [41] shows evidence that the main European stock indices are good predictors of economic activity in the period from 1981 to 2020. Moreover, Vidal-Avello [42] explores the problem of financial contagion, emphasizing the macroeconomic determinants that encourage it. The data show the three main variables related to financial crises: GDP growth, government debt, and investment. Specifically, they find that among the variables that determine an increase in financial contagion probability stand out the GDP growth and GDP per capita. Furthermore, variables such as investment and foreign direct investment generate a decrease probability in contrast. Therefore, it seems that the financial system could be used as a variable to explain economic activity in crisis periods.

4. Methodology

The contagion in this case will be defined by the tendency of the stock market indices to be grouped into different sets under a functional data analysis. Fewer groups means that the contagion pattern among the countries tends to strengthen, while the greater the variability between the stock indices, the more the contagion pattern weakens. The second step is to identify the factors that support this pattern for a certain period of time. The following section describes the methodology used to analyze contagion and the determinants that drive this phenomenon. First, the B-spline method and the Bezier curves are described, which are used for the smoothing of observations. Subsequently, the methodology of FPCA is detailed. Finally, classification trees are used to determine the variables that drive each group.

4.1. B-Spline Method and Bezier Curves

For the application of FPCA, the first step is to transform the observations into functions. For this purpose, the B-Splines methodology is applied. The main objective of the B-spline smoothing estimator in a regression lies in the efficient adjustment of the data while maintaining a degree of smoothness. The basis of the smoothing methodology starts with the Bezier curve, which instead of considering lines between control points considers spline functions of degree K . In their simple form, splines are regression models that integrate predictors, being a set of basic functions (orthonormal eigenfunctions) that are used to force the regression line to alternate direction at some point within the support range of the function describing the following nonparametric regression configuration:

$$y(x_i) = f(x_i) + \varepsilon_i \quad i = 1, 2, \dots, p \tag{1}$$

where $f(\cdot)$ is a function that has a continuous second derivative. In some cases, the function value $f(x_i)$ can be approximated by the multiplication of a set of basic functions $\phi(x_i) = [\phi_1(x_i), \dots, \phi_K(x_i)]$ and its coefficients $\beta = (\beta_1, \dots, \beta_K)$, therefore $f(x) = \beta\phi'(x) = \phi(x)\beta$ and

$$y(x_i) - \beta\phi'(x_i) = \varepsilon \tag{2}$$

where a standard measure of goodness of fit for data is given by $\sum_{i=1}^p [y(x_i) - f(x_i)]^2$ and a natural measure of smoothness is the integral of the square of the second derivative of the function, given by $\int_{x_1}^{x_p} [D^2 f(x)]^2 dx$. An overall performance indicator is formed as:

$$\sum_{i=1}^p [y(x_i) - f(x_i)]^2 + \lambda \int_{x_1}^{x_p} [D^2 f(x)]^2 dx \tag{3}$$

The minimization of the above equation grants smoothed functions of the data. A non-negative value of the smoothing parameter λ determines the trade-off between goodness of fit and smoothness of the function.

4.2. Functional Principal Component Analysis

As a second step, identify the countries that are related considering a metric given by principal component scores based on multivariate correlation. According to Chávez-Chong et al. [43], FPCA is an extension of the classic Principal Component Analysis, in which principal components are represented by functions and not by vectors. FPCA finds the set of orthogonal principal component functions that maximize variance across each component. The functional principal component $\phi_1(x)$ represents the variance of the scores of the principal components, it is maximized subject $\|\phi_1^2(x)\| = \int_{x_1}^{x_p} \phi_1^2(x)dx = 1$ such that:

$$\beta_1 = \int_{x_1}^{x_p} \phi(x)f(x)dx, \tag{4}$$

The successive functions of the principal components can be obtained iteratively by subtracting the first k from the functions of the principal components $f^0(x) = f(x)$ for each $1 \leq k \leq K < n$, i.e.,

$$f^k(x) = f^{k-1}(x) - \beta_k\phi_k(x) \tag{5}$$

The availability of a sample of N curves allows for investigating the way they vary among themselves. In this sense, the idea of the correlation and variance of matrices in the multivariate context is the correlation and covariance functions or surfaces $\rho(s, t)$ and $\sigma(s, t)$. The value $\rho(s, t)$ defines the correlation between the values $x(s)$ and $x(t)$ on a sample or population of curves and similar for $\sigma(s, t)$. (Ramsay et al. [44], p. 56). Two basis systems and a matrix of coefficients are used for a single object. A bivariate correlation surface estimate is described as:

$$r(s, t) = \sum_k^K \sum_\ell^L b_{k,\ell}\phi_k(s)\psi_\ell(t) = \phi'(s)B\psi(t) \tag{6}$$

where $b_{k,\ell}$ are eigenvalues, $\phi_k(s)$ is an orthonormal eigenfunction for variation over s and $\psi_\ell(t)$ is an orthonormal eigenfunction for variation over t . To strengthen the analysis of patterns, the K-means algorithm is proposed as a clustering tool. Every observation x_i is assigned to a given group since the sum of the distance of the squares of the observation of its assigned center is minimized. The Total Variation (TV) within the group is defined by:

$$TV = \sum_{k=1}^K W(C_k) = \sum_{k=1}^K \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{7}$$

Finally, the optimal number of groups can be defined by three different methods: the elbow method, silhouette method, and gap statistics.

4.3. Classification Trees and Penalized Multinomial Regression (RIDGE and LASSO)

Once the clusters are determined, the Classification Trees (CT) tool is applied. The objective is to determine the rules or characteristics, as well as their reference values, that led the countries of the same group to behave in this way during periods of crises. According to Gámez-Martínez et al. [45], p. 34, the classification tree methodology can work with continuous variables and categorical variables. The construction of the tree is done during the learning phase, which can be simplified into the following steps that are applied iteratively:

1. Each node is divided according to a test that is raised based on the value of some of the defined characteristics. In the case of binary analysis, false or true, the variables that meet the conditions are assigned to the child nodes and the remaining ones to the other. It is necessary to propose a measure of impurity. In this sense, the measures of Entropy and the Gini Index used by the CART algorithm stand out. This process is performed iteratively.

2. The partitioning node process is stopped when an established condition is met. There are several possibilities, among others, to stop the division of the nodes when they are pure, when their size is less than a certain threshold, or in the case that they exceed a certain level of purity considered in terms of the proportion of the majority class, or to use the gain of information or reduction of impurity as a criterion to stop the growth of the tree.

In line with the above, there is abundant literature on the use of decision trees in finance. Some examples can be found in Kaminski et al. [46] and Chang et al. [47]; both documents show the advantages of decision trees over other methodologies that follow the same spirit. In the same way, there are other works where macroeconomic indicators are incorporated to face financial problems; an example of this can be found in Xia et al. [48], where decision trees are used to study credit ratings and predict credit risk.

The decision trees are an efficient and dynamic methodology that can be used with other machine learning methodologies, as was carried out in this investigation. The above nexus can be seen in Deng et al. [49], where decision trees are used with other tools for studying the Chinese capital market. Recent works where decision trees are used in finance topics can be found in Yu and Zhao [50] and Kristóf and Virág [51].

Finally, in the specific case of contagion, Chevallier [52] presents a study where the main object is investigating financial contagion in the COVID-19 pandemic by examining 31 stock markets.

Decision trees are an efficient and dynamic methodology, so much so that they can be used with other machine learning methodologies, as was carried out in this research. The above link can be seen in Deng et al. [49], where decision trees are used with other tools to study the Chinese capital market. It is also worth mentioning that recent work using decision trees in financial issues can also be found in Yu and Zhao [50] and Kristóf and Virág [51]. Finally, in the specific case of contagion Chevallier [52] presents a study where the main objective is to investigate financial contagion in the COVID-19 pandemic by examining 31 stock markets.

We now proceed as Friedman et al. [16] did and a linear regression model with regularization is carried out. Given the categorical nature of the dependent variable, the RIDGE and LASSO-type regularization regression models are used for the multinomial case (Regularized Multinomial Regression). In this case, the penalized likelihood function is determined by the following expression:

$$\max_{\{\beta_l, \beta_{0l}\}} \left[\frac{1}{N} \sum_{i=1}^N \log p_{gi}(x_i) - \lambda \sum_{l=1}^K P_{\alpha}(\beta_l) \right] \tag{8}$$

where

$$\log p_{gi}(x_i) = \sum_{l=1}^K y_{i,l}(\beta_{0,l} + x_i \beta_l) - \log \left(\sum_{l=1}^K e^{\beta_{0,l} + x_i \beta_l} \right) \tag{9}$$

Equation (9) shows an expression involving the probability of one of the levels of the categorical variable. The usual optimization algorithm for multinomial regression (Newton’s Method) can be somewhat complicated and resource-intensive; instead, we proceed to separate the algorithm by estimating the probability of each state in the categorical variable.

It is important to mention that for this point λ is established as the value that minimizes the classification error function in a multiple cross-validation process considering 10 n -folds. Each λ for each fold test has an increment of 0.01, taking values in the interval [0, 100]. It is important to mention that the objective function has the same form for both models (RIDGE and LASSO). The penalty term will be first a quadratic function and subsequently an absolute value penalty function, as in Friedman et al. [16].

5. Data, Empirical Results, and Their Discussion

This section first uses the FPCA approach to carry out the contagion analysis of the stock indices for 15 major economies. From the United States, Mexico, Brazil, Argentina, Chile, and Canada; from Europe, the United Kingdom, France, Spain, Germany, Switzerland, and the Netherlands; and from Asia, Japan, China, and Russia. The reference stock indices are listed in Table 1. The data for stock indices was obtained from Yahoo Finance, and the data for macroeconomic and financial variables from the World Bank.

Table 1. Stock indices; main world economies.

Country	Stock Index
America	
United States	S&P 500, Dow Jones and Nasdaq
Mexico	IPC
Brazil	IBOVESPA
Argentina	MERVAL
Chile	IPSA
Canada	TSX
Europe	
United Kingdom	FTSE 100
France	CAC
Spain	IBEX
Germany	DAX
Swiss	SMI
Netherlands	AEX
Asia	
Japan	Nikkei
China	Hang Seng and SSE
Russia	RTSI

Next, the stock market contagion in the countries is examined through the formation of groups among the stock market indices, considering their dynamics over time.

5.1. Observations Smoothing and Dynamic Correlation: Functional Approach

One of the main requirements to perform principal component analysis is a relative high correlation between variables. Therefore, as a starting point, the simple correlation between the stock indices selected for this study is presented. Table 2 shows the highest and lowest correlations of some of the stock indices. In this context, it is evident the high correlation that the Argentina Merval index keeps with two of the most important US stock indices, such as the S&P 500 and the Nasdaq, with a correlation of 95.6% and 96.3%, respectively. In the same way, a high correlation is observed between the IPC index (Mexico) and the Canada TSX index, with a value of 90.2%. In the case of the lowest correlations, the little relationship between the Japan Nikkei index and the Spain IBEX index is notorious, with values of only 23.5%. On the other hand, it is important to highlight the negative relationship of -4% between the IBEX index and the S&P 500. A similar case occurs between the Russia RTSI index and the Netherlands AEX index, with a negative relationship of -8.7% .

Table 2. Correlation Analysis. Author’s own elaboration with data from Yahoo Finance and R-4.3.1 software.

High Correlation (Top 5)		Low Correlation (Top 5)	
S&P 500 and Merval	0.956	S&P 500 and IBEX	0.041
DAX and S&P 500	0.939	Nikkei and IBEX	0.235
DAX and Dow Jones	0.938	IPC and CAC	0.152
Merval and Nasdaq	0.963	RTSI and Nasdaq	0.156
IPC and TSX	0.902	RTSI and AEX	0.087

According to the functional core component methodology, the first step of this technique is to obtain functional observations. To this end, a smoothing is realized through the B-splines and the Bezier curve, which aim to minimize the noise of the data. This procedure is applied to all indices used in this study. The results can be appreciated in Figure 1, which includes monthly observations from January 2000 to December 2021.

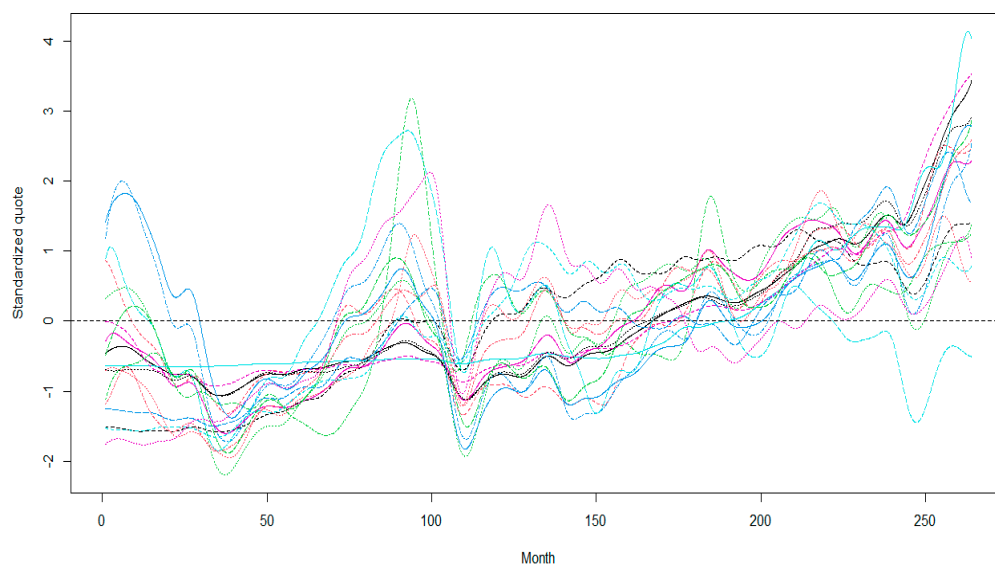


Figure 1. Smoothing of stock indices from January 2000 to December 2021. Author’s own elaboration with data from Yahoo Finance and R-4.3.1 software.

As can be observed in Figure 1, it seems that there is a generalized pattern of behavior among the indices. The black solid line represents the average function considering all stock indices. It is important to note that, among the functions, there are periods in which variability is appreciated, such as the period marked by observations 50 and 100. However, it seems to identify a pattern of generalized behavior among the indices from observation 100 (referring to the year 2008), in which clearly the variability between functions decreases, providing evidence of similar behavior among them. A similar behavior tends to strengthen in observation 50, which coincides with the period of the “Dot.com” financial crisis. This seems to indicate that the pattern of generalized behavior tends to strengthen in crisis periods, as can be seen in the two periods of economic turbulence mentioned above.

To reinforce the analysis, Figure 2 presents the correlation surface. It is possible to appreciate that the highest correlation occurs between 50 and 250 months (years 2004 and 2020, respectively) with a positive value greater than 0.5. This would suggest that the effects observed during observation 50 (2004) could be similar to COVID-19 crisis effects.

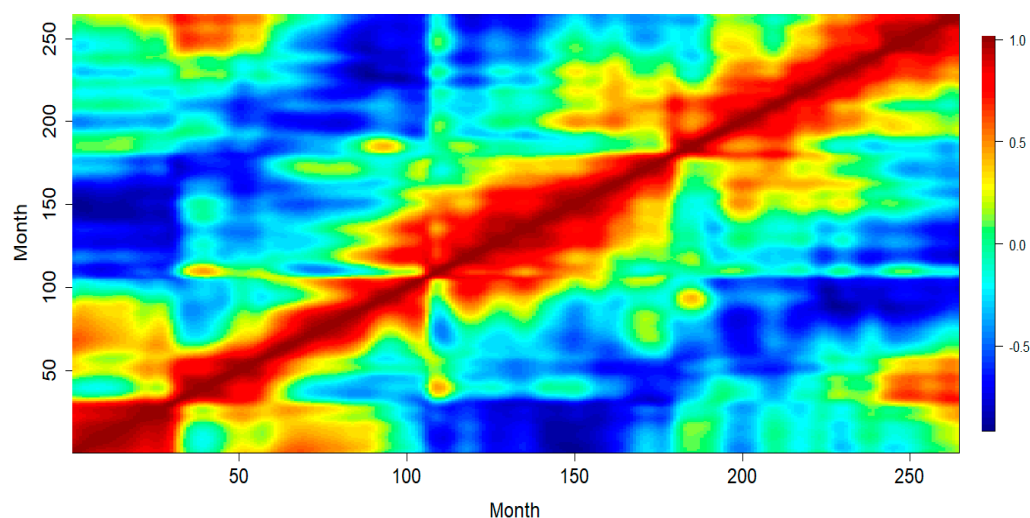


Figure 2. Variance and covariance correlation surface (dark red: high correlation; dark blue: low correlation). Author's own elaboration with data from Yahoo Finance and R-4.3.1 software.

5.2. Functional Principal Components

This section describes the application of FPCA in terms of eigen-decomposition analysis functions corresponding to the covariance and variance of functions. Figure 3 presents the extraction of the three main functional components. The first of these captures 41.3% of the variability, while the second and third capture, 32.2% and 9.7%, respectively. In cumulative terms, it represents 85% of the variation of all functions. The solid line refers to the average of the stock indices in the period 2000–2021. In this sense, it can be seen how the variability of components 1 and 2 with respect to the average seems to coincide with periods of financial crises (month 1 corresponding to January 2000, month 100 corresponding to the year 2008, and month 250 corresponding to the year 2020). The first and second components give us 75.5% of the variance explained. Hence, it is important to mention that for this study, only the first two components are considered.

5.3. Behavior Pattern of Stock Indices through *k*-Means

From the correlation analysis, it is possible to appreciate some existing relationships between stock indices. The most evident refers to the indices of Argentina (Merval) and the US (Nasdaq, S&P500, and Dow Jones). However, in order to provide a formal cluster analysis that also considers dynamic correlations over time, the K-means algorithm is used, which provides a more robust analysis of the different groups into which stock indices are grouped. Based on the scores resulting from the FPCA, the K-means cluster algorithm determines six main groups using the silhouette method and cross-validation, considering as a minimization function the sum squares of the distances for the complete study period, from January 2000 to December 2021. The K means algorithm is selected to generate the clusters by iterating through 10,000 different initial configurations for the random selection of the centroids, which is sufficient to ensure convergence due to the small number of scores resulting from the functional principal components. The groups generated by the average centroids of the 10,000 configurations are taken as the final result. This is efficient in terms of speed of convergence and calculation of groups for small samples, compared with other clustering algorithms such as J-Means and its derivatives and Gaussian Mixtures, which, despite being more accurate, require more additional resources compared with K means.

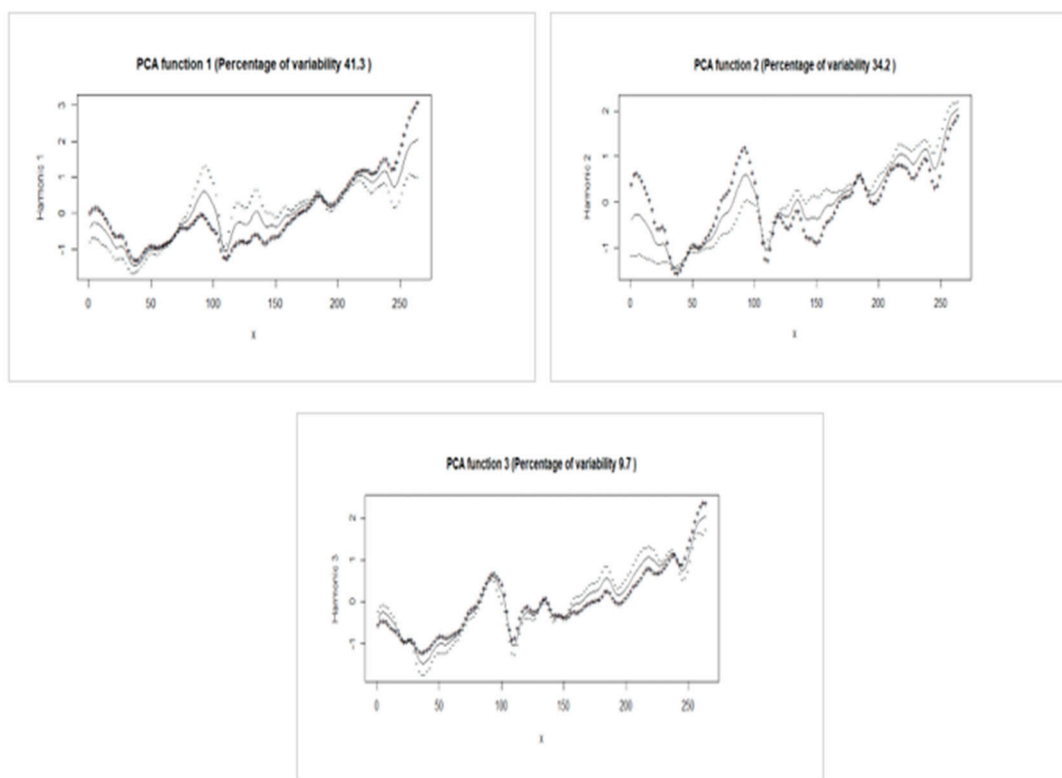


Figure 3. Three principal components. Percentage of explained variance (solid lines represent the average of the stock indices). Author's own elaboration with data from Yahoo Finance and R-4.3.1 software.

Notice first, in Figure 4, that Group 1 consists of the UK's FTSE 100 stock index, Germany's DAX, Switzerland's SMI, and Japan's Nikkei. The second group is made up of the three main stock indices of the United States, S&P 500, Dow Jones, Nasdaq, and Merval of Argentina (consistent with correlation analysis, with one of the highest values). In the same way, for this group, the observations are closer to each other. The third group is made up of three Latin American economies: Mexico, Chile, and Brazil (IPC, IPSA, and IBOVESPA, respectively), in addition to China's Hang-Seng and Canada's TSX. The fourth group is made up of the Netherlands with the AEX stock index and the French CAC index. Moreover, Shanghai, with the SSE stock index, and Russia, with the RSTI, make up the fifth group. Finally, the country that is separated from the others is the IBEX of Spain.

Although the general analysis of the whole period shows some relationships that cannot be appreciated with a simple correlation analysis, it does not allow observing the evolution of the cluster through time. In order to strengthen the analysis of patterns, we proceed to incorporate the methodology of sliding time windows with a sample size of 12 monthly observations. In addition, the above allows for observing the evolution of the clusters and identifying patterns of generalized behavior in crisis periods.

5.4. Stock Indices Contagion through K-Means and Sliding Windows

It is important to remember that this research is guided mainly by the definition of contagion, which occurs when the transmission channel presents some variation after a shock in the financial market and worsens in periods of crisis. In this sense, this research seeks to expose the intensification of transmission channels through the behavior patterns of stock market indices and demonstrate through empirical evidence their increase in a crisis context. Therefore, two important periods of major economic turbulence are considered: the subprime crisis of 2008 and the economic crisis caused by COVID-19 in 2020. Figure 5 shows the stock indices clusters for the 2008 subprime crisis, where the pre- and post-

crisis behavior is additionally included. First of all, it is possible to appreciate that for 2005 (pre-crisis), the stock indices are grouped into six sets, showing a greater dispersion among them, where in most cases, the indices that integrate the group are from financial markets sharing the same geographical area or integrate the same economic block, such as the case of Mexico’s IPC, which is associated with other Latin American indices such as Brazil’s IBOVESPA and Chile’s IPSA. The same argument applies to European countries, associating the CAC of France with the AEX of the Netherlands. It is important to highlight the fact that the most important stock indices in the United States (S&P500, Dow Jones, and Nasdaq) coincide with the Argentine Merval again. Secondly, for 2007 and 2008, the stock indices were grouped into only two clusters, showing a significant decrease in dispersion; this is empirical evidence of the intensification of the behavior pattern among stock indices. Finally, it is observed in the graph of 2011 (post-crisis) how the stock indices tend to separate, forming four groups.

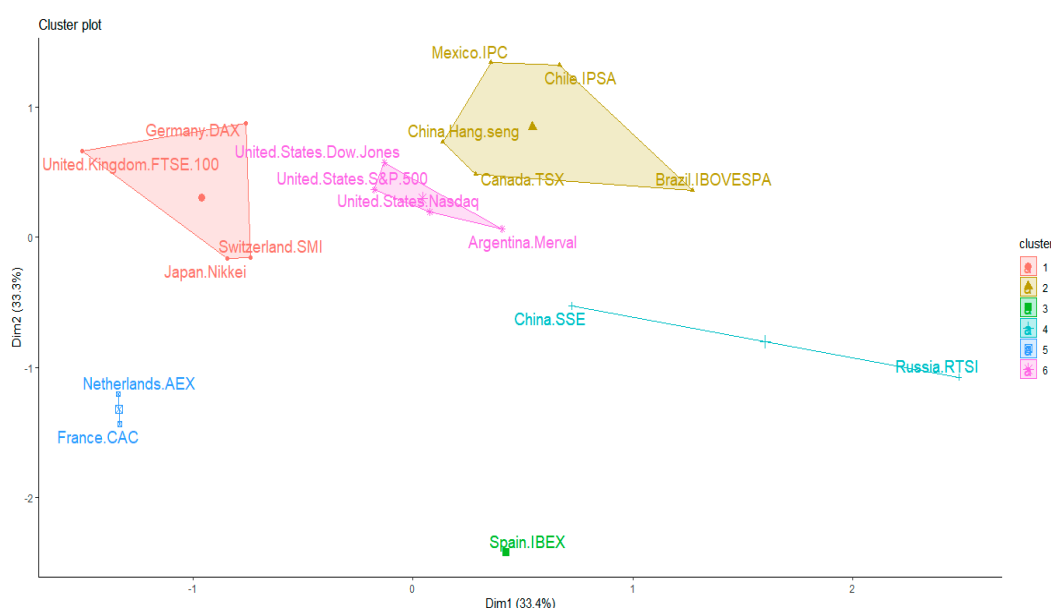


Figure 4. K-means clustering, stock market indices from 2000 to 2021. Author’s own elaboration with data from Yahoo Finance and R-4.3.1 software.

For the case of the 2020 economic crisis, the increase in the contagion pattern among the indices seems to share the same dynamics as the 2008 crisis, showing its intensification starting in 2019, as can be seen in Figure 6. It is also possible to observe again two clusters, where one of them is made up of the IBEX of Spain, the RTSI of Russia, and the SSE of China, like the group of 2008, while the rest of the countries in the sample integrate the remaining group. For 2020, the intensification of the clustering pattern among the indices can be observed. A point to highlight from the analysis of this period is that the three most important indices in the United States were grouped with the Marvel of Argentina, while in the second group, indices such as the Mexican IPC and the Chilean IPSA are integrated from Latin America. On the other hand, European indices such as the French CAC and the UK FTSE. For 2021, the cluster seems to maintain the same structure as in 2020, with the only difference that the French index and the AEX of the Netherlands become part of the group.

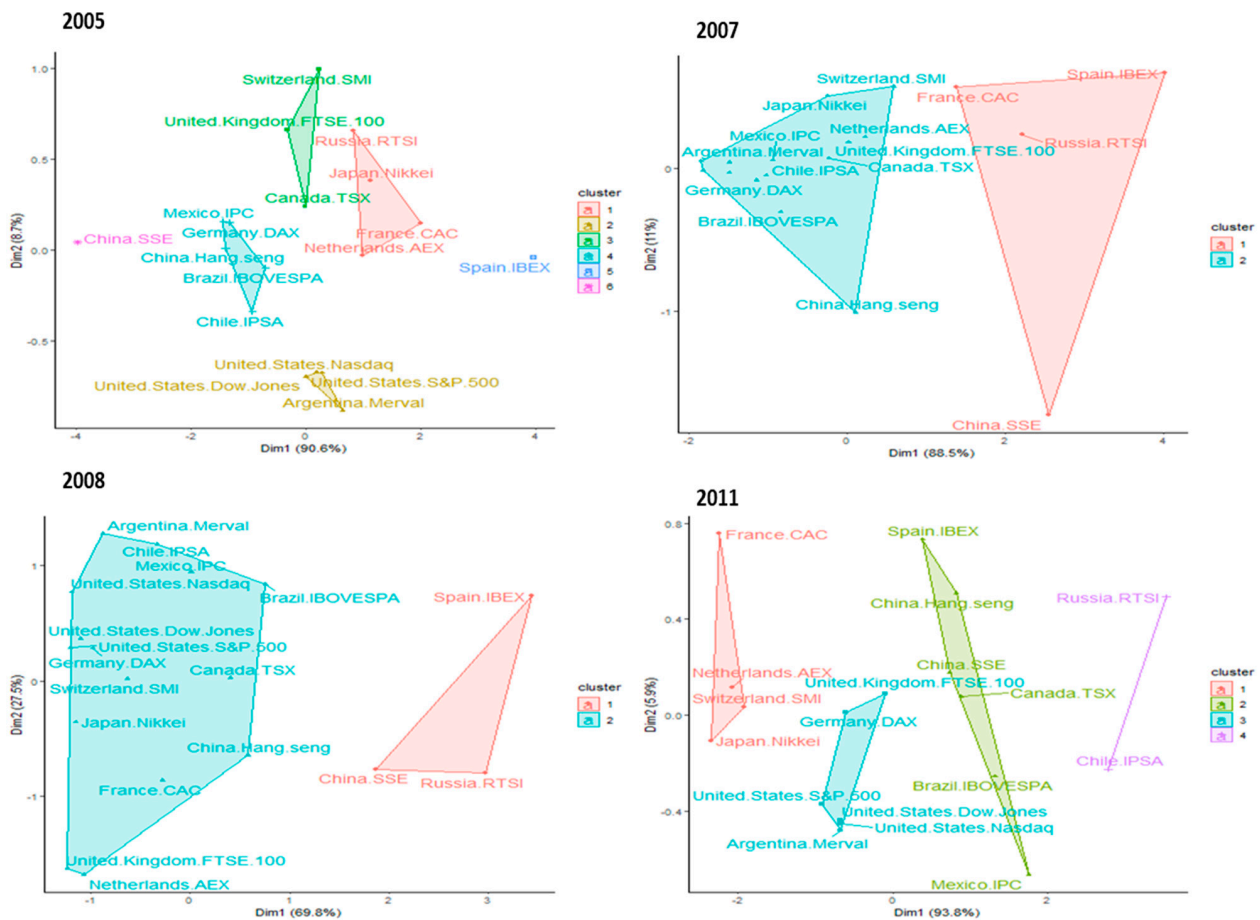


Figure 5. Stock index clusters by K-means. (2005, 2007, 2008, 2011). Author’s own elaboration with data from Yahoo Finance and R-4.3.1 software.

In order to make the analysis more robust among the stock indices for both crisis periods, the behavior of the stock indices returns in the first group for 2008 is presented in Figure 7. For October 2008, there were significant falls in all indices, with the Merval of Argentina and the IBOVESPA of Brazil suffering the largest decreases with -46% and -28% , respectively.

For the case of 2020, Figure 8 shows that the behavior pattern among the indices for the first group is similar to the last one, being the Merval and IBOVESPA of Argentina and Brazil, who presented the most considerable decreases for March of that year, with falls of -36% for both. In addition, the Argentine index shows the largest fluctuations compared with the others. For the second group, the behavior between the indices intensifies for March and November, respectively, highlighting the fact that the Spanish IBEX and the Russian RTSI present the largest decrease for March, but they also show the greatest recovery by the end of 2020.

5.5. Macroeconomic and Financial Variables as Determinants of Stock Market Contagion

From the results of the previous section, empirical evidence is shown of the existence of a pattern among stock market indices that seems to intensify in a crisis period. However, the variables that favor the formation of these groups and what are the characteristics that drive each of the countries to associate are unknown at this point. The objective of this section is to identify the macroeconomic and financial variables that promote contagion as well as the extraction of important rules for grouping in periods of financial and economic crises. For this purpose, classification trees are used, considering financial variables related to economic growth in accordance with the literature. Gross Domestic Product, Foreign

Direct Investment, Domestic Credit, Gross Fixed Capital Formation, Consumer Price Index, BIG MAC (Real exchange rate, proxy), Net Exports, Government Expenditure. Additionally, a variable for the size of the origin market of the index is included to avoid biases in the analysis (Companies listed in the capital market).

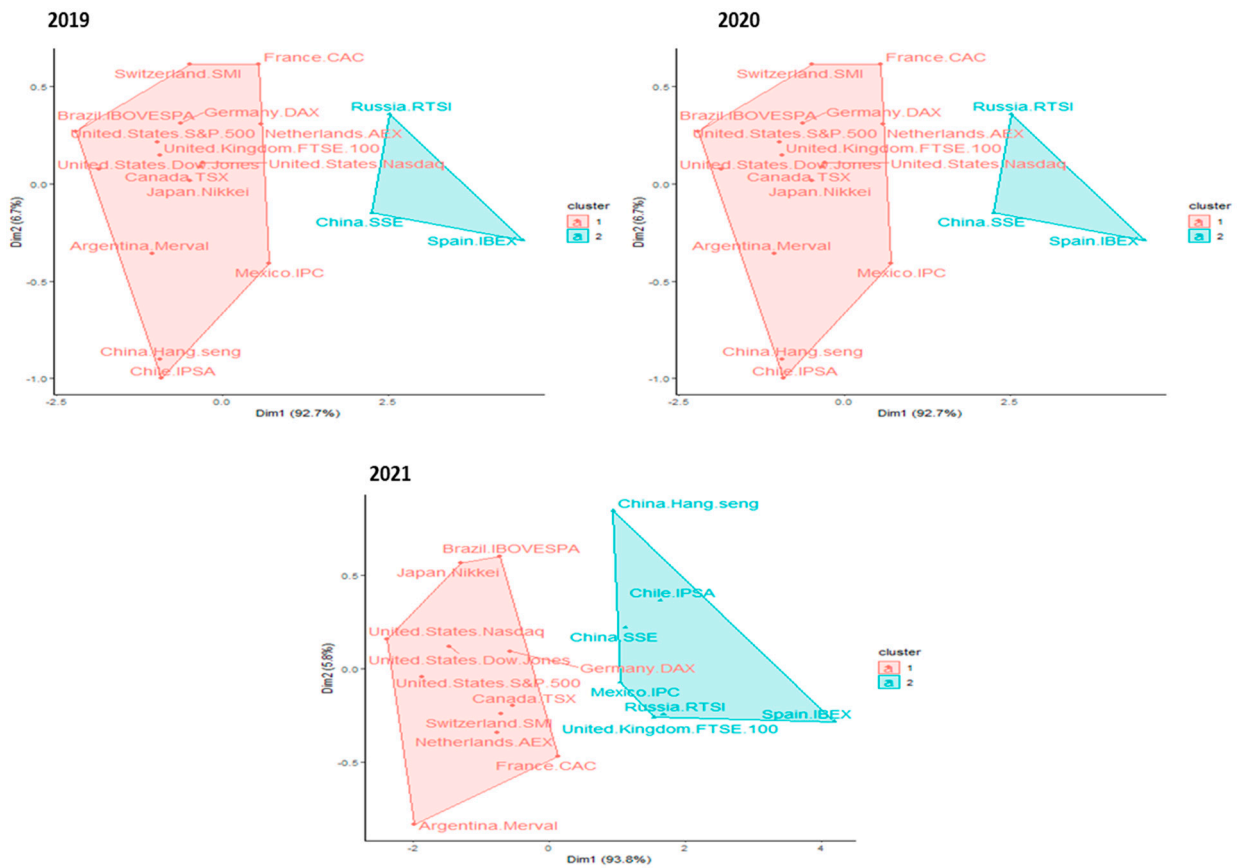


Figure 6. Stock index clusters by K-means (2019, 2020 and 2021). Author’s own elaboration with data from Yahoo Finance and R-4.3.1 software.

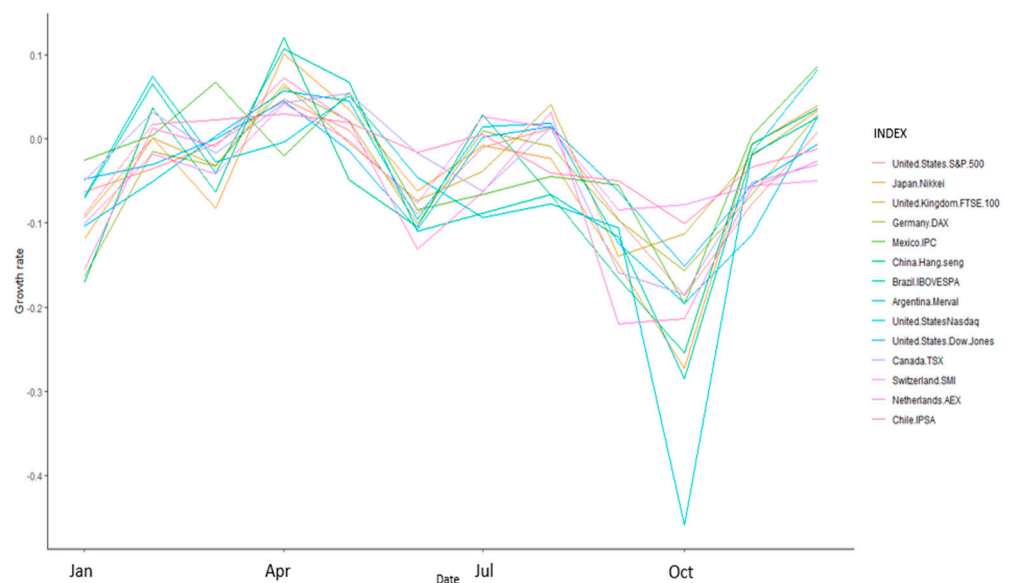


Figure 7. Stock indices behavior (growth rate). Group 1 for 2008. Author’s own elaboration with data from Yahoo Finance and R-4.3.1 software.

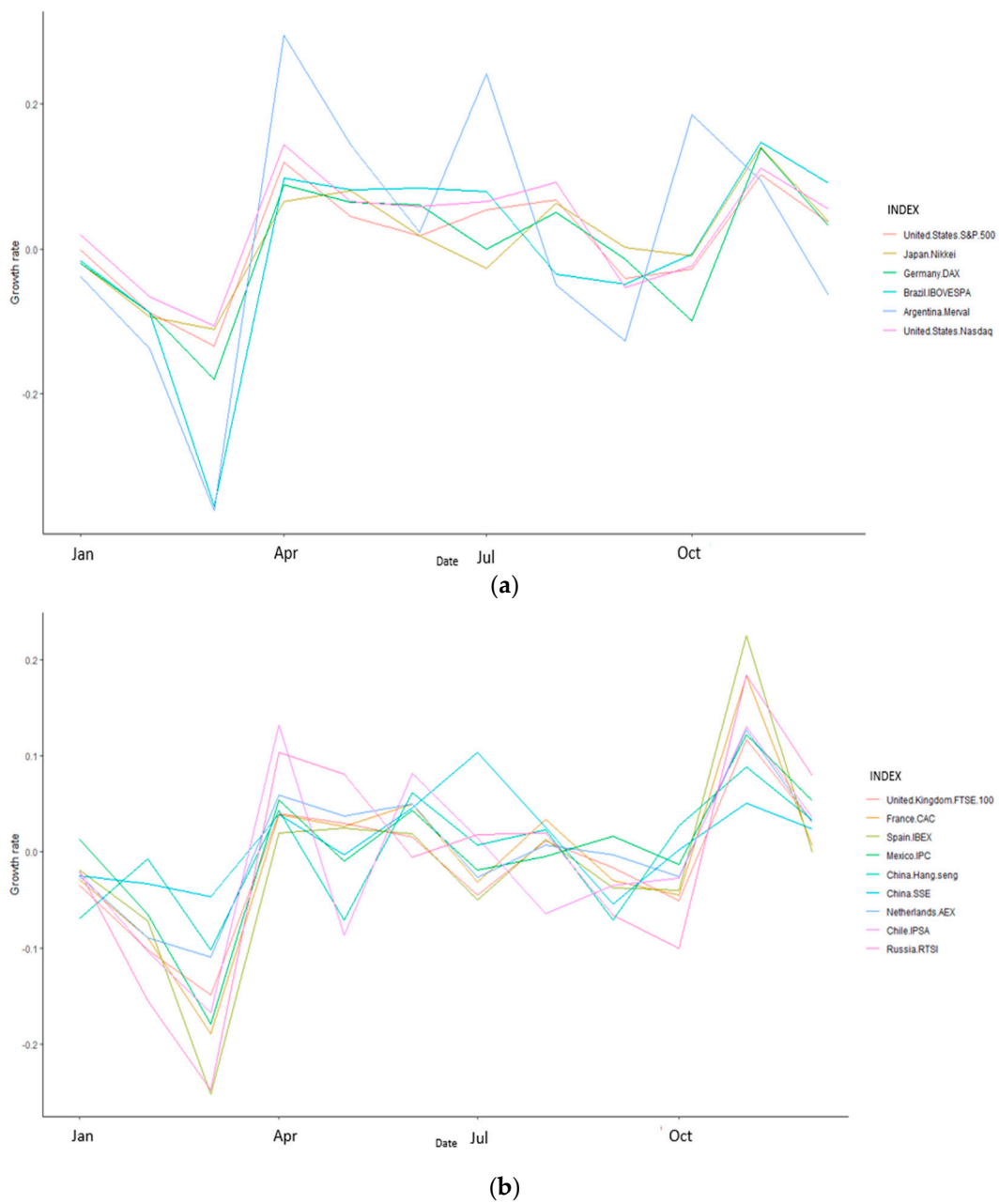


Figure 8. Stock indices behavior (returns). COVID-19, (a) first group and (b) second group. Author’s own elaboration with data from Yahoo Finance and R-4.3.1 software.

Figure 9 integrates the variables for the 2007 groups, where the number of groups into which the stock indices are divided is reduced to two. Gross Domestic Product (GDP) seems to have an important weight as the main node, appearing also at the end node.

Table 3 presents the conditions seen in the classification tree for 2007 (Figure 9). It is important to point out the fact that, again, the Gross Domestic Product (GDP) is presented as one of the variables with the greatest weight for 2007. From this analysis, it is highlighted that for group 1, the two rules start from the assumption that countries must register a high Gross Domestic Product above \$1,275,551,818,841.76 million dollars. The same situation can be observed for the second group, where the first rule starts with having registered a GDP above \$1,691,268,576,062.40 million dollars.

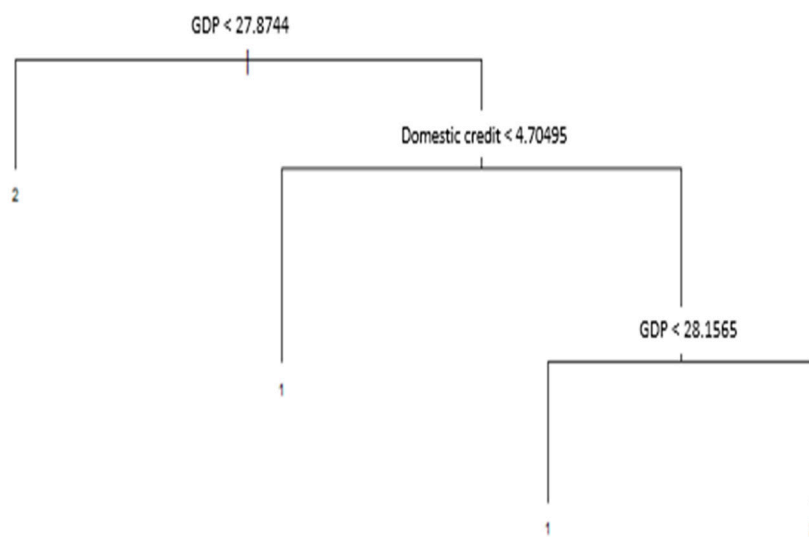


Figure 9. Determinant variables of Groups 1 and 2 (2007). Author’s own elaboration with data from the World Bank and R-4.3.1 software.

Table 3. Classification tree rules for 2007. Author’s own elaboration with data from the World Bank and R-4.3.1 software.

Group 1	
Rule	Conditions
1	GDP > \$1,275,551,818,841.76 and Domestic credit < 110.49%
2	\$1,275,551,818,841.76 < GDP < \$1,691,268,576,062.40 and Domestic credit > 110.49%
Group 2	
Rule	Conditions
1	GDP < \$1,275,551,818,841.76
2	GDP > \$1,691,268,576,062.40 and Domestic credit > 110.49%

For the year 2008, it can be seen in Table 4 that the number of conditions for groups 1 and 2 are two for both. For this year, the variables that determine both groups are the BIG MAC Index (BIG MAC) and the companies that are listed on the stock exchange as shown in Figure 10.

Table 4. Classification Tree Rules for 2008. Author’s own elaboration with data from the World Bank and R software.

Group 1	
Rule	Conditions
1	BIG MAC > 1.09939794 and Companies < 2923
2	BIG MAC > 1.09939794, Companies > 2923 and Companies > 3684
Group 2	
Rule	Conditions
1	BIG MAC < 1.09939794
2	BIG MAC > 1.09939794, Companies > 2923, and Companies < 3684

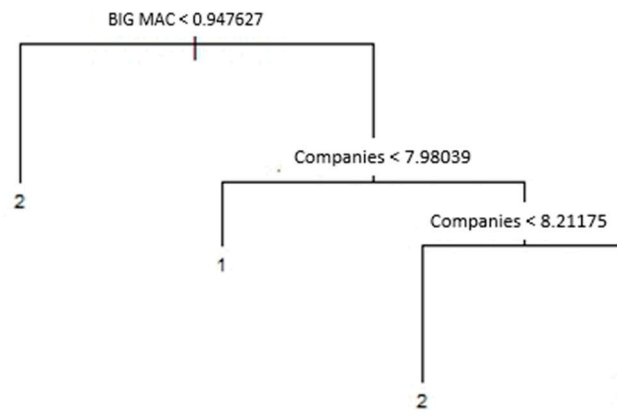


Figure 10. Determinant variables for Groups 1 and 2 (2008). Author’s own elaboration with data from the World Bank and R-4.3.1 software.

For 2011, the indices tend to separate to form 4 groups; it is important to note that the main variables that determine each group change are the BIG MAC index as a global node and Foreign Direct Investment (FDI). For this year, only the determining variables of groups 2 and 3 will be described because they are the groups that integrate the largest number of stock indices; see Figure 11.

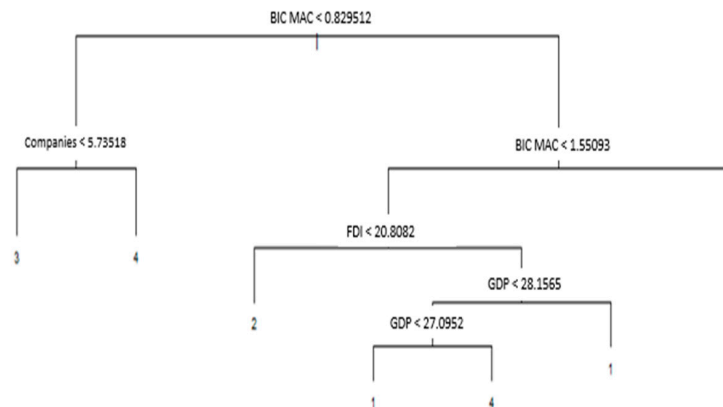


Figure 11. Group determinant variables (2011). Author’s own elaboration with data from the World Bank and R-4.3.1 software.

Group 3 is integrated by Mexico, Brazil, Spain, China and Canada, with only one rule: BIG MAC < 2.29 and companies listed on the stock market < 309, as shown in Table 5.

Table 5. Classification Tree Rules for 2011. Author’s own elaboration with data from the World Bank and R-4.3.1 software.

Group 2	
Rule	Conditions
1	BIG MAC > 4.72
2	BIG MAC < 4.72 and FDI < \$1,088,536,532.15
Group 3	
Rule	Conditions
1	BIG MAC < 2.29 and Companies listed on the stock market < 309

For 2019, the most representative variable is the Consumer Price Index (CPI), followed by the number of companies listed on the stock market and GDP in the third and fourth

nodes, as shown in Figure 12. For both 2019 and 2008 (Figure 5), it is important to note that the variable that appears as the main node is the Consumer Price Index (CPI). From above, it seems that this variable increases in importance before and after crises. Figure 12 presents the determinant variables of groups 1 and 2 in 2019, and Table 6 provides the corresponding classification tree rules.

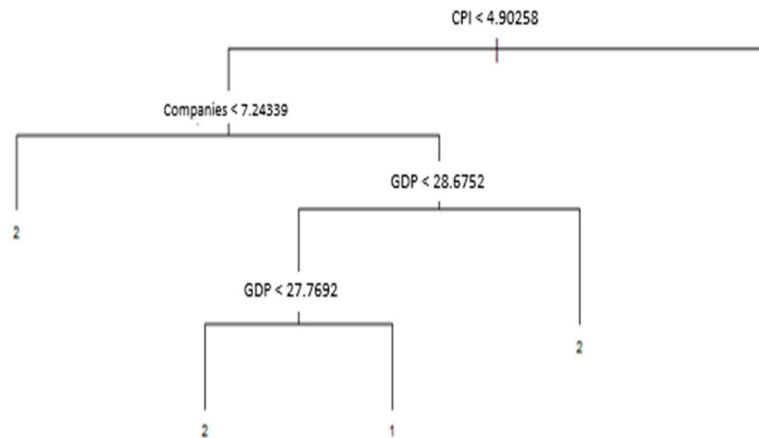


Figure 12. Group determinant variables (2019). Author’s own elaboration with data from the World Bank and R-4.3.1 software.

Table 6. Classification Tree Rules for 2019 Author’s own elaboration with data from the World Bank and R-4.3.1 software.

Group 1	
Rule	Conditions
1	CPI < 134.63, companies < 1399
2	CPI < 134.63 and Companies > 1399 and GDP > \$2,841,064,722,124.72
3	CPI < 134.63 and Companies > 1399 and GDP < \$1,148,180,923,182.96
Group 2	
Rule	Conditions
1	CPI >134.63
2	CPI < 134.63 and Companies > 1399 and GDP > \$1,148,180,923,182.96

Finally, in 2020 the most significant variables are Foreign Direct Investment (FDI) and Gross Domestic Product (GDP). Figure 13 shows the determinant variables of groups 1 and 2 in 2020, and Table 7 provides the corresponding classification tree rules.

Based on World Bank data, for 2020, the countries that integrate group 1, such as the United States, Canada, Brazil, Argentina, among others, will register a Gross Domestic Product that starts at \$389,288,056,265.33 (Argentina) and with a maximum value of \$20,893,746,000,000.00 million dollars (United States). Additionally, the countries that form part of group 1 for 2020, register an average amount of GDP of \$2,189,079,922,468.12 million dollars.

Derived from the classification trees, in Tables 5–7, it can be argued that the most important variables that determine each of the groups, in most cases, are Gross Domestic Product, Foreign Direct Investment, and the Consumer Price Index. In addition, the particularity of the countries belonging to group 1 in 2008 is that they tend to register an average production higher than the countries that integrate group 2, with average

values of \$2,068,151,373,874.05 and \$1,810,766,998,833.75 million dollars, respectively. For 2020, the average GDP changed between both groups given that the economies of group 2 registered an average value of \$1,485,745,587,141.41 and group 1 registered an average of \$2,189,079,922,468.12 million dollars. Additionally, for both 2008 and 2020, each group integrates an economy with a GDP with very large values, such as the case of the United States and China, with \$14,712,844,084,000.00 and \$20,893,746,000,000.00 in the case of the United States and \$4,594,307,032,667.98 and \$14,722,730,697,890.10 for the Chinese economy. Due to the above, for the calculation of the average of each group for both 2008 and 2020, the values of the US and China were not considered, in order to avoid bias in the data.

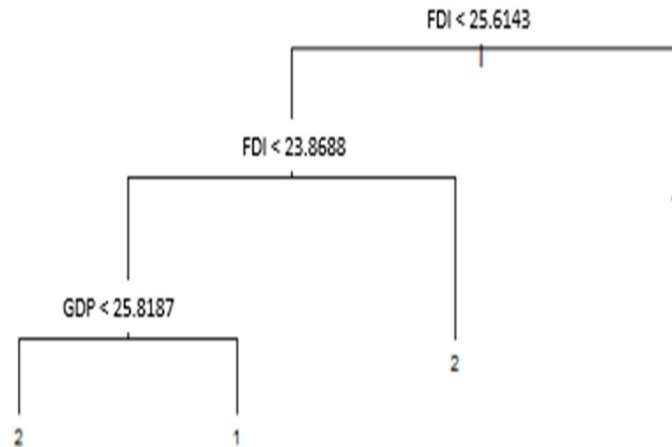


Figure 13. Group determinant variables (2020). Author’s own elaboration with data from the World Bank and R-4.3.1 software.

Table 7. Classification Tree Rules for 2020. Author’s own elaboration with data from the World Bank and R-4.3.1 software.

Group 1	
Rule	Conditions
1	FDI > \$133,091,140,846.40
2	FDI < \$23,232,081,879.27 and GDP > \$163,274,717,087.92
Group 2	
Rule	Conditions
1	FDI > \$23,232,081,879.27
2	FDI < \$23,232,081,879.27 and GDP < \$163,274,717,087.92

Next, to conclude the analysis, we carry out an econometric analysis throughout the periods of crisis or financial turbulence. In order to make the empirical results obtained with classification trees more robust, we identify both the importance and magnitude of each macroeconomic variable when determining the probability of belonging to a group. As mentioned previously, given the categorical nature of the dependent variable, we use a multinomial regression in its traditional version and additionally with RIDGE and LASSO-type regularization. This is a very valuable and important tool since it allows policymakers to identify strategies and decisions on macroeconomic variables to abandon a contagion pattern that implies an imminent crisis.

The econometric model presented below uses a dependent categorical variable that indicates the group to which the stock market indices belong after the application of K-means. The independent variables included are Gross Domestic Product (GDP), Consumer

Price Index (CPI), Foreign Direct Investment (FDI), Domestic Credit (DC), BIG MAC index (BIGMAC), Gross Fixed Capital Formation (GFCF), Net exports (EXP), Government expenditure (GEX), Number of companies (COMP), and Shares traded (SHARES). Periods that have at least two groups are selected (2007, 2008, 2012, 2014, 2019, and 2020). In this sense, the multinomial logit model converges to an original logit.

As can be seen in Table 8, the significant variables that can explain the change from one group to another are Gross Domestic Product, Gross Fixed Capital Formation, Net exports, Government expenditure, Number of companies, and Shares traded. It is important to mention that the aforementioned variables are related to economic growth, which may imply a relationship between the index groupings and economic growth. Next, an individual analysis is carried out by using a LASSO and RIDGE regression for each period to identify, in particular, the conditions that affected the contagion patterns.

Table 8. Results of the multinomial regression model (selected periods).

Variable	Coefficient	Z-Statistic	p-Value
GDP	4.1395	2.2226	0.0262
CPI	−1.4795	−1.0681	0.2855
FDI	0.3523	1.3162	0.1881
DC	0.8133	0.7098	0.4778
GFCF	−4.7518	−2.2575	0.0240
BIGMAC	0.6430	0.4447	0.6565
EXP	2.9238	2.0967	0.0360
GEXP	−4.5190	−2.5571	0.0106
COMP	1.6598	1.7018	0.0888
SHARES	−1.9020	−1.9548	0.0506

The first case analyzed is the 2008 crisis. As can be appreciated, there are two groups where the performances of the average cumulative returns of the stock market indices belonging to each cluster are presented in Figure 14. The cumulative trend of negative returns can be seen for both groups; however, group 2 has the worst performance with a negative average cumulative return of approximately 75%.

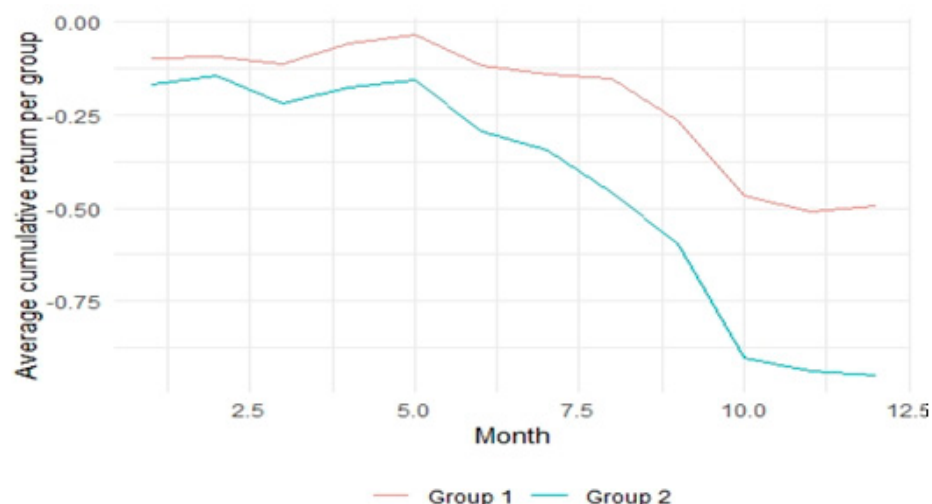


Figure 14. Average cumulative return. Cluster comparison (2008).

The RIDGE multinomial model (Table 8) shows a positive relationship between group 1 and macroeconomic variables, with the exception of Government Expenditure as a proportion of Gross Domestic Product (GDP). This suggests that a marginal increase in GDP or Foreign Direct Investment (FDI) generates an increase in the probability of being in group 1 (the cluster with the best performance throughout 2008). Additionally, Table 8 also shows the results of the regularized regression model of the LASSO type, in which the only significant variable is GDP. It is important to mention that GDP is one of the variables selected by the classification tree model as one of the variables found in the rules for group classification. From our econometric analysis, it can be inferred that the variable that has the greatest impact in determining whether a country belongs to one group or the other is the GDP, with a marginal coefficient of 0.00067.

In what follows, the same analysis is carried out as before for the periods of 2011 and 2020. For the first case, four groups can be seen (Figure 15), where the average accumulated returns of the members of each cluster are very similar to each other. With respect to the pattern of contagion, despite the negative trend of average accumulated returns for all groups, it is cluster 3 that presents a better performance.

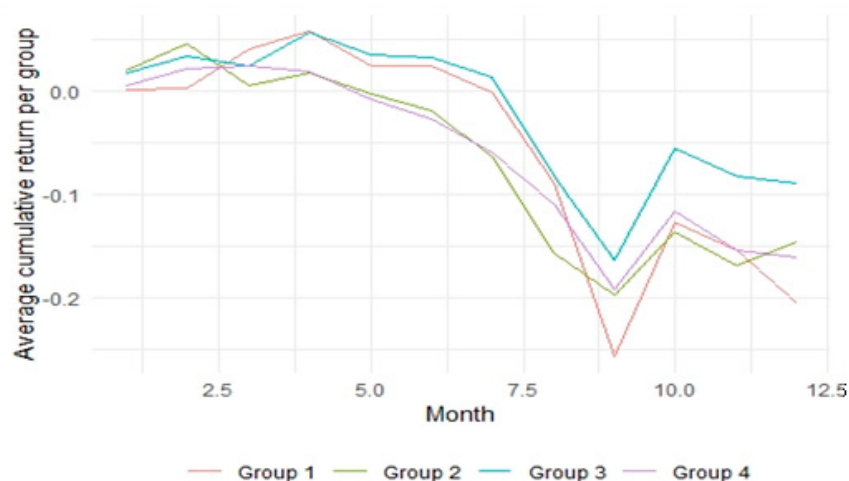


Figure 15. Average cumulative return. Cluster comparison (2011).

For group 3 (Table 9), the RIDGE model shows that the GDP, FDI, the number of listed companies, the Big Mac index, and the Consumer Price Index (CPI) are the ones with the largest coefficients. In contrast, for the LASSO model for group 3, the most important macroeconomic variables are GDP and FDI with a positive sign, which contribute to increasing the probability of belonging to this group with a marginal increase of the aforementioned variables, while the CPI has a negative and inversely proportional relationship with the probability that a stock index belongs to group 3. It is important to mention that group 2 was the second with the best performance throughout 2011. The variables with a positive sign that have the highest importance according to the LASSO model are the Big Mac and Government Expenditure, while the effect with a negative sign is reflected in the FDI and the number of companies listed on the stock exchange. This makes sense since one way to increase the probability that a stock index belongs to group 3 is to increase GDP and FDI, while another policy would be to decrease inflation (CPI). The variables in terms of significant importance coincide with the analysis carried out with the classification tree.

Table 9. Results of the RIDGE and LASSO multinomial regression models (2008).

	GDP	FDI	Domestic Credit	Companies	CPI	BIC MAC	Exports	GFCF	GOV EXP (%GDP)
RIDGE regression									
Group 1	0.00067	0.00034	0.00042	0.00015	0.00265	0.00023	0.00040	0.00058	−0.00124
Group 2	−0.00067	−0.00034	−0.00042	−0.00015	−0.00265	−0.00023	−0.00040	−0.00058	0.00124
LASSO regression									
Group 1	0.02429	0	0	0	0	0	0	0	0
Group 2	−0.02429	0	0	0	0	0	0	0	0

Finally, the results for the 2020 analysis are presented in Figure 16 where it can be seen that the performance of the stock market indices that make up group 1 is deficient compared with the same indicator of group 2.

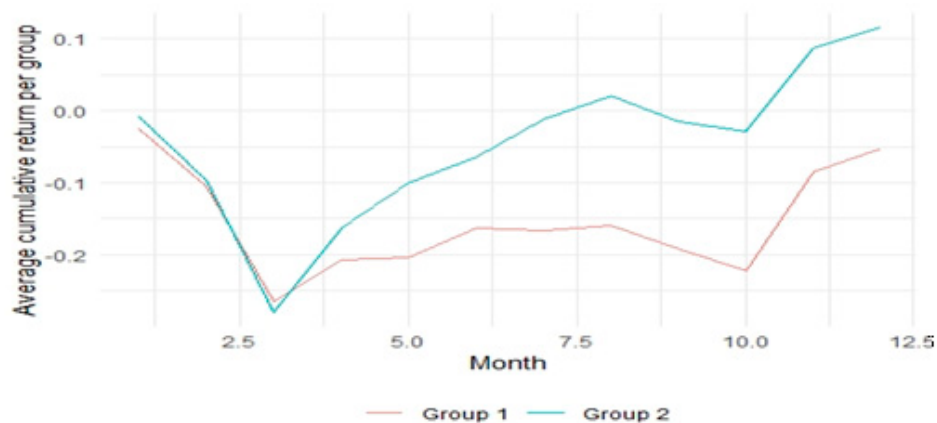


Figure 16. Average cumulative return. Cluster comparison (2020).

According to the results obtained of LASSO and RIDGE multinomial regression models, shown in Tables 10 and 11, respectively, increasing the GDP improves the probability that the stock market index can be integrated into group 2, where the second most important variable is the CPI, again with an inverse relationship (negative sign), meaning that a decrease in inflation helps to increase the probability of belonging to group 2.

Table 10. Results of the RIDGE and LASSO multinomial regression models (2011).

	GDP	FDI	Domestic Credit	Companies	CPI	BIC MAC	Exports	GFCF	GOV EXP (%GDP)
RIDGE regression									
Group 2	0.03006	0.06470	0.14704	−0.12792	0.49732	0.53000	0.06403	0.02800	0.26360
Group 3	0.05722	0.03348	0.01327	0.06779	0.93324	0.08871	0.03768	0.00273	0.24787
Group 4	0.02716	0.03121	0.13376	0.06012	0.43592	0.61871	0.02635	0.02526	0.01572
LASSO regression									
Group 2	0	−0.17455	0	−0.16634	0	1.9799	0	0	0.08004
Group 3	0.23696	0.05155	0	0	−3.2138	0	0	0	0
Group 4	0	0	0	0	0	−2.2828	0	0	0

Table 11. Results of the RIDGE and LASSO multinomial regression models (2020).

	GDP	FDI	Domestic Credit	Companies	CPI	BIC MAC	Exports	GFCF	GOV EXP (%GDP)
RIDGE regression									
Group 1	−0.03137	0.01944	−0.02089	−0.03852	0.39646	−0.05167	0.01168	−0.01367	0.14407
Group 2	0.03137	−0.01944	0.02089	0.03852	−0.39646	0.05167	−0.01168	0.01367	−0.14407
LASSO regression									
Group 1	−0.17129	0	0	0	1.94668	0	0	0	0.59491
Group 2	0.17129	0	0	0	−1.94668	0	0	0	−0.59491

Finally, it is interesting to note that for this last analysis, the determining variables of the 2020 groups differ for the most part with respect to the results of the decision trees. It is important to mention that the classification error rate is 50%, while the classification tree is 80%.

6. Conclusions

The FPCA offers a quite opportune treatment for the analysis of contagion determining the behavior pattern of the main worldwide stock indices since this methodology, unlike others, considers the time factor and performs a better dynamic correlation analysis of the observations. In this sense, the main conclusions obtained through FPCA and K-means are, first, that the stock indices behave fairly similarly in crisis periods, such as the 2008 subprime crisis and the 2020 COVID economic crisis. Secondly, the correlation map works as a measure of contagion between the stock indices and also as an early warning indicator. Thirdly, it is possible to appreciate the strong relationship among Latin American stock indices such as Mexico’s IPC, Brazil’s IBOVESPA, and Chile’s IPSA. It is also important to highlight the close relationship between Argentina’s Merval and the Dow Jones, Nasdaq, and S&P 500, especially during economic crises periods. Moreover, the IBEX, RTSI, SSE, and Hang-Seng indices seem to present an isolated pattern in K-means analysis, specifically in 2008, showing a lower tendency to present a contagion pattern than the rest of the indices in this study. Furthermore, the RTSI and the IBEX growth rates presented the most marked decreases in March 2020 and the largest recoveries for November in the same year, corroborating the strength between them and the isolated pattern in comparison to the remaining indices. It is important to mention that one of the drawbacks of this research was not being able to include additional economies due to the lack of data in any of the variables, such as India, which is one of the most important emerging economies.

Derived also from the FPCA and the sliding windows clusters, contagion patterns get stronger before the crisis period, so this methodology could be used as an early warning crisis system through the early observation of variation among financial system indicators such as stock market indices. Furthermore, the contagion pattern strength can be observed after this phenomenon ends; for instance, in 2008, the stock indices began to separate until 2011, three years after the crisis. The same argument applies to the recent COVID-19 crisis, given that stock indices did not seem to diminish the contagion pattern until the last data presented in this research. It is important to mention, and surprising that Mexico’s IPC was not clustered in the same group of United States stock market indices anytime, despite the strong commercial and diplomatic relationships that these countries share in addition to their geographical closeness. That was the reason the need arose to investigate which macroeconomic indicators are possible contagion determinants of stock indices.

Regarding classification trees, the determining variables with greater recurrence in periods of crisis are characterized by GDP, CPI, Exports, and FDI. The above supports the results of various authors where the relationship between economic growth and the financial system is observed. It is not surprising that these variables turn out to be the main determinants of the groups, given the nature of these indicators, where the stock

market depends to a large extent on the total production of the countries and the prices in their commercial markets, as well as the confidence of foreign countries to invest, which corroborates the fundamental hypothesis. One possible application is to have a tool that makes it possible to identify which macroeconomic variables need to be modified to allow a country to move from a group with a pattern to another that is more beneficial. This could be a way out and a recovery outlet for countries that have fallen into a financial crisis as a result of contagion.

Finally, a point to highlight is that government spending included as one of the original macroeconomic variables in the classification tree analysis does not have a significant effect at any time as a determining of contagion, indicating that the stock markets of the different countries do not depend on their government's performance and where they are closer to efficiency and self-regulation.

Future research may consider including a regularized regression model in its panel version to take advantage of the cross-sectional characteristics of the sample. Additionally, it will be recommended that the regularization be carried out for each of the variables by means of an adaptive LASSO regression based on the weights of a ridge regression. Another possible alternative is to explore grouping tools based on distributions (Gaussian Mixtures) to define a statistical structure for each of the groups. Perhaps the application of Fuzzy type tools will provide more flexibility to clusters. **Author Contributions:** Conceptualization,

data gathering, simulations, numerical tests, methodology, formal analysis, investigation, writing original draft preparation and writing, review and editing, L.L.R.-C., J.O.R.-D.-A. and F.V.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Instituto Politécnico Nacional.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The simulation and test results data presented in this study are available on request from the corresponding author.

Acknowledgments: We, all authors, are very grateful to the four reviewers for carefully reading the document and for the many valuable and pertinent comments and suggestions that improved the final version of this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhou, Z.; Lin, L.; Li, S. International stock market contagion. A CEEMDAN wavelet analysis. *Econ. Model.* **2018**, *72*, 333–352. [[CrossRef](#)]
2. BenMim, I.; BenSaïda, A. Financial contagion across major stock markets: A study during crisis episodes. *N. Am. J. Econ. Financ.* **2019**, *48*, 187–201. [[CrossRef](#)]
3. Kao, Y.S.; Zhao, K.; Ku, Y.C.; Nieh, C.C. The asymmetric contagion effect from the U.S. stock market around the subprime crisis between 2007 and 2010. *Econ. Res. Ekon. Istraz.* **2019**, *32*, 2422–2454. [[CrossRef](#)]
4. Kao, W.S.; Kao, T.C.; Changchien, C.C.; Wang, L.H.; Yeh, K.T. Contagion in international stock markets after the subprime mortgage crisis. *Chin. Econ.* **2018**, *51*, 130–153. [[CrossRef](#)]
5. Okorie, D.I.; Lin, B. Stock markets and the COVID-19 fractal contagion effects. *Financ. Res. Lett.* **2021**, *38*, 101640. [[CrossRef](#)]
6. Yarovaya, L.; Brzeszczyński, J.; Goodell, J.W.; Lucey, B.; Lau, C.K.M. Rethinking financial contagion: Information transmission mechanism during the COVID-19 pandemic. *J. Int. Financ. Mark. Inst. Money* **2022**, *79*, 101589. [[CrossRef](#)]
7. Akhtaruzzaman, M.; Abdel-Qader, W.; Hammami, H.; Shams, S. Is China a source of financial contagion? *Financ. Res. Lett.* **2021**, *38*, 101393. [[CrossRef](#)]
8. Mohti, W.; Dionisio, A.; Ferreira, P.; Vieira, I. Contagion of the subprime financial crisis on frontier stock markets: A copula analysis. *Economies* **2019**, *7*, 15. [[CrossRef](#)]
9. Zorgati, I.; Lakhali, F.; Zaabi, E. Financial contagion in the subprime crisis context: A copula approach. *N. Am. J. Econ. Financ.* **2019**, *47*, 269–282. [[CrossRef](#)]
10. Banerjee, A.K. Futures market and the contagion effect of COVID-19 syndrome. *Financ. Res. Lett.* **2021**, *43*, 102018. [[CrossRef](#)]
11. Haile, F.; Pozo, S. Currency crisis contagion and the identification of transmission channels. *Int. Rev. Econ. Financ.* **2008**, *17*, 572–588. [[CrossRef](#)]

12. Bekaert, G.; Ehrmann, M.; Fratzscher, M.; Mehl, A. Global Crises and Equity Market Contagion—Working Paper Series 1381, European Central Bank 2011. Available online: <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1381.pdf> (accessed on 14 January 2023).
13. Gkillas, K.; Tsagkanos, A.; Vortelinos, D.I. Integration and risk contagion in financial crises: Evidence from international stock markets. *J. Bus. Res.* **2019**, *104*, 350–365. [CrossRef]
14. Ye, W.; Jiang, K.; Liu, X. Financial contagion and the TIR-MIDAS model. *Financ. Res. Lett.* **2021**, *39*, 101589. [CrossRef]
15. Altınbaş, H.; Pacelli, V.; Sica, E. An empirical assessment of the contagion determinants in the Euro Area in a period of sovereign debt risk. *Ital. Econ. J.* **2021**, *8*, 339–371. [CrossRef]
16. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1–22. Available online: <https://pubmed.ncbi.nlm.nih.gov/20808728/> (accessed on 9 May 2023). [CrossRef]
17. Pericoli, M.; Sbracia, M. A primer on financial contagion. *J. Econ. Surv.* **2003**, *17*, 573–575. [CrossRef]
18. Toribio-Dávila, J.J. El contagio económico y financiero: Conceptos básicos. In *La Crisis en Europa: ¿Un Problema de Deuda Soberana o Una Crisis del Euro?* 1st ed.; Méndez-de-Andés-Fernández, F., Ed.; Fundación de Estudios Financieros; Universidad de la Rioja: Logroño, Spain, 2012; pp. 29–38. Available online: <https://dialnet.unirioja.es/servlet/articulo?codigo=3928905> (accessed on 19 January 2023).
19. Rojas-Mora, J.; Chamorro-Futinico, J. Dinámica y volatilidad en los índices bursátiles de los países de la Alianza Pacífico. *Panor. Económico* **2017**, *24*, 71–84. [CrossRef]
20. Sosa, M.; Ortiz, E. Global financial crisis volatility impact and contagion effect on NAFTA equity markets. *Estocástica Finanz. y Riesgo* **2016**, *7*, 67–88. Available online: <http://zaloamati.azc.uam.mx/handle/11191/4946> (accessed on 13 May 2023). [CrossRef]
21. Gavidia-Pantoja, L.A. *Contagio Entre los Mercados Bursátiles Latinoamericanos. Una Aproximación a Través de Cópulas Cambiantes en el Tiempo*; Pontificia Universidad Católica del Perú (PUCP): Lima, Perú, 2017; manuscript in preparation. Available online: <https://www.researchgate.net/profile/Luis-Gavidia/project/Latin-American-Stock-Market-Linkages-A-time-varying-copulas-approach/attachment/59ee81d54cde2617ef868f49/AS:552774123388928@1508803029676/download/Seminario+de+Tesis+Parcial.pdf?context=ProjectUpdatesLog> (accessed on 6 February 2023).
22. Bucio-Pacheco, C.; Gutiérrez, R.; Sosa-Castro, M.M. Contagio vía copulas dinámicas en los mercados de capitales del TLCAN de 2000 a 2016. *EconoQuantum* **2018**, *16*, 65–80. [CrossRef]
23. Díaz-Rodríguez, H.; Bucio, C. Contagio bursátil en los mercados del TLCAN, países emergentes y el mercado global. *Rev. Mex. Econ. Finanz.* **2018**, *13*, 345–361. [CrossRef]
24. Santillán-Salgado, R.J.; Gurrola-Ríos, C.; Jiménez-Preciado, A.L.; Venegas-Martínez, F. La dependencia del Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores (IPC) con respecto a los principales índices bursátiles latinoamericanos. *Contaduría Adm.* **2018**, *63*, 1–18. [CrossRef]
25. Piffaut, P.V.; Miró, D.R. Integración, contagio financiero y riesgo bursátil: ¿qué nos dice la evidencia empírica para el periodo 1995–2016? *Cuad. Econ.* **2016**, *39*, 138–145. [CrossRef]
26. Akhtaruzzaman, M.; Boubaker, S.; Sensoy, A. Financial contagion during COVID-19 crisis. *Financ. Res. Lett.* **2021**, *38*, 1–16. [CrossRef] [PubMed]
27. Vortelinos, D.I.; Gkillas, K.; Tsagkanos, A. *Integration, Contagion and Risk Contagion in Financial Crises: Evidence from International Stock Markets*; Working Paper; Lincoln International Business School: Lincoln, UK, 2017. Available online: https://efmaefm.org/OEFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2017-Athens/papers/EFMA2017_0340_fullpaper.pdf (accessed on 2 February 2023).
28. Uddin, G.S.; Yahya, M.; Goswami, G.G.; Lucey, B.; Ahmed, A. Stock market contagion during the COVID-19 pandemic in emerging economies. *Int. Rev. Econ. Finanz.* **2022**, *79*, 302–309. [CrossRef]
29. Ji, X.; Wang, S.; Xiao, H.; Bu, N.; Lin, X. Contagion Effect of Financial Markets in Crisis: An Analysis Based on the DCC-MGARCH Model. *Mathematics* **2022**, *10*, 1819. [CrossRef]
30. Davidescu, A.A.; Manta, E.M.; Hapau, R.G.; Gruiescu, M.; Vacaru (Boita), O.M. Exploring the Contagion Effect from Developed to Emerging CEE Financial Markets. *Mathematics* **2023**, *11*, 666. [CrossRef]
31. Bildirici, M.E.; Salman, M.; Ersin, Ö.Ö. Nonlinear Contagion and Causality Nexus between Oil, Gold, VIX Investor Sentiment, Exchange Rate and Stock Market Returns: The MS-GARCH Copula Causality Method. *Mathematics* **2022**, *10*, 4035. [CrossRef]
32. Ramírez-Silva, R.A.; Nacional, I.P.; Cruz-Aké, S.; Venegas-Martínez, F. Volatility Contagion of Stock Returns of Microfinance Institutions in Emerging Markets: A DCC-M-GARCH Model. *Rev. Mex. Econ. Finanz.* **2018**, *13*, 325–343. [CrossRef]
33. Carrion-i-Silvestre, J.L.; Villar, O. Dependencia y contagio financiero en los mercados bursátiles durante la gran recesión. In Proceedings of the XIV Encuentro de Economía Aplicada, Huelva, Spain, 2–3 June 2011. Available online: <https://archivo.alde.es/encuentros.alde.es/anteriores/xiveea/trabajos/c/pdf/148.pdf> (accessed on 12 January 2023).
34. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*, 2nd ed.; Springer: New York, NY, USA, 2005; pp. 147–172. [CrossRef]
35. Lin Shang, H. Visualizing and Forecasting Functional Time Series. Ph.D. Thesis, Department of Econometrics and Business Statistics Monash University, Melbourne, Australia, 2010. Available online: https://bridges.monash.edu/articles/thesis/Visualizing_and_forecasting_functional_time_series/4546333/1 (accessed on 21 March 2023).
36. Dewandaru, G.; Masih, R.; Masih, A.M.M. Contagion and interdependence across Asia-Pacific equity markets: An analysis based on multi-horizon discrete and continuous wavelet transformations. *Int. Rev. Econ. Finanz.* **2016**, *43*, 363–375. [CrossRef]

37. Hassan, M.K.; Sánchez, B.; Yu, J.S. Financial development, and economic growth: New evidence from panel data. *Q. Rev. Econ. Financ.* **2011**, *51*, 88–100. [[CrossRef](#)]
38. De la Cruz-Gallegos, J.L.; Lizárraga, J.Á.A. Crecimiento económico y el crédito bancario: Un análisis de causalidad para México. *Rev. Econ. Fac. Econ. Univ. Autónoma Yucatán* **2011**, *28*, 10–32. [[CrossRef](#)]
39. Lezama-Palomino, J.C.; Laverde-Sarmiento, M.Á.; Gómez-Restrepo, C.A. El Mercado De Valores Y Su Influencia En La Economía: Estudio Del Caso Colombiano 2001–2013 (The Stock Market and its Impact on the Economy: A Colombian Case Study 2001–2013). *Rev. Int. Adm. Finanz.* **2017**, *10*, 29–37. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039736 (accessed on 3 May 2023).
40. Landa-Díaz, H.O.; Silva-Barrón, T. Impacto del desarrollo financiero en el crecimiento económico de América Latina. *Contaduría Y Adm.* **2021**, *66*, 1–21. [[CrossRef](#)]
41. Bernárdez-Castrejón, J.M. Relación Entre la Bolsa Europea y el Crecimiento Económico. Master's Thesis, Repositorio Institucional de la Universidad Complutense, Madrid, Spain, 2021. Available online: <https://eprints.ucm.es/id/eprint/67138/> (accessed on 27 March 2023).
42. Vidal-Avello, I. Determinantes del Contagio de Crisis Financieras. Master's Thesis, Universidad de Talca, Facultad de Economía y Negocios, Maule, Chile, 2019. Available online: <http://dspace.otalca.cl/handle/1950/11975> (accessed on 19 January 2023).
43. Chávez-Chong, C.O.; Sánchez-García, J.E.; De-la-Cerda, J. Análisis de componentes principales funcionales en series de tiempo económicas (Analysis of principal functional components in economic time series). *GECONTEC. Rev. Int. Gestión Conoc. Tecnol.* **2015**, *3*, 13–14. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2737561 (accessed on 28 January 2023).
44. Ramsay, J.O.; Hooker, G.; Graves, S. *Functional Data Analysis with R and MATLAB*; Springer: New York, NY, USA, 2009; p. 56. Available online: <https://mobt3ath.com/uplode/book/book-61855.pdf> (accessed on 13 May 2023).
45. Gámez-Martínez, M.; Alfaro-Cortés, E.; Alfaro-Navarro, J.L.; García-Rubio, N. Árboles de clasificación para el análisis de gráficos de control multivariantes. *Rev. Mat. Teoría Apl.* **2009**, *16*, 34. [[CrossRef](#)]
46. Kamiński, B.; Jakubczyk, M.; Szufel, P. A framework for sensitivity analysis of decision trees. *Cent. Eur. J. Oper. Res.* **2018**, *26*, 135–159. [[CrossRef](#)] [[PubMed](#)]
47. Chang, Y.C.; Chang, K.H.; Wu, G.J. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Appl. Soft Comput.* **2018**, *73*, 914–920. [[CrossRef](#)]
48. Xia, Y.; He, L.; Li, Y.; Fu, Y.; Xu, Y. A dynamic credit scoring model based on survival gradient boosting decision tree approach. *Technol. Econ. Dev. Econ.* **2021**, *27*, 96–119. [[CrossRef](#)]
49. Deng, S.; Wang, C.; Wang, M.; Sun, Z. A gradient boosting decision tree approach for insider trading identification: An empirical model evaluation of China stock market. *Appl. Soft Comput.* **2019**, *83*, 105652. [[CrossRef](#)]
50. Yu, J.; Zhao, J. Prediction of systemic risk contagion based on a dynamic complex network model using machine learning algorithm. *Complexity* **2020**, *2020*, 6035372. [[CrossRef](#)]
51. Kristóf, T.; Virág, M. EU-27 bank failure prediction with C5. 0 decision trees and deep learning neural networks. *Res. Int. Bus. Financ.* **2022**, *61*, 101644. [[CrossRef](#)]
52. Chevallier, J. COVID-19 pandemic and financial contagion. *J. Risk Financ. Manag.* **2020**, *13*, 309. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.