Life Safety and Security

open access

Health security

Minireview

DIVERSIFICATION IN HEALTH CARE INDUSTRY IN US MARKET: AN EMPIRICAL TIME-SERIES ANALYSIS

Alessandra Costa[1,2,3]

[1] Department of Economics, University of Messina, Messina, Italy
[2] Department of Biomolecular strategies, genetics and avant-garde therapies, I. E. ME. S. T., Palermo, Italy
[3] Louvain Institute of Data Analysis and Modeling in economics and statistics (IDAM), CORE, Louvain-La-Neuve, Belgium

Correspondence:
Alessandra Costa
e-mail: alessandracosta@iemest.eu

Received: December 20th, 2018
Revised: January 18th, 2019
Accepted: January 24th, 2019

Abstract

The US healthcare industry represents the fastest growing of the US economy: since it accounts for 17.2% of total GDP, there is no doubts about its attractiveness for investors and it could be useful to analyze the benefits that could arise from diversification portfolio's strategies.

Thus, by using high frequency time-series analysis, we focus on different indices and founds that are able to measure the stock price performances of healthcare industries and its sub-segments, in the US market. We show that models that take into account dynamic correlations among assets better performs respect to traditional models, and allows investors to minimize the variance of their financial portfolio, thus providing the usefulness of time series methods as management tool for investors.

Keywords

forecast, hfd, dcc, healthcare

Introduction

The US healthcare system has a particular configuration respect to the other industrialized countries. As reported by G. Rodic, Gleason and O. Rodic [1], the United States has no single national system of healthcare, but it could be defined as a hybrid system, in which private health insurance can be purchased from profit commercial insurance companies or nonprofit ones.

The last Census Report on healthcare insurance's coverage in the USA specify that private coverage includes health insurance provided through employer and that directly purchased by an individual from an insurance companies, while government coverage includes federal programs and individual state health plans for particular categories of population (like military insurance). In 2017, 91.2% of population had health insurance coverage by any health plan: among them, the 67.2% had private health insurance, and about the 37.7% had a public health coverage.

Healthcare accounts for a large portion of the US economy. As reported in Figure 1, it represents the 17.2% of GDP and it's about eight percentage points above the average of the industrialized members of the Organization for Economic Co-operation and Development (OECD).

During the last year, the health expenditure per capita of the USA reached $10209.4, even if the government seems to play a marginal role. Through a rapid comparison, considering the Government schemes, the average health expenditure is about $3697.2 against the private one, in the same year.

Although the US government and individual investors are an important component of the US healthcare system, it's important to underline that its main players are the profit maximizing entities (companies, private and public funds, etc.). Thus, from an investment point of view could be useful to analyze the main index of the entire healthcare sector in the US, the Vanguard Health Care ETF (hereafter, VHT)[1].

The definition of the healthcare industry, as the one dedicated to the study, treatment and prevention of illness, disease and injury, allows us

[1] The VHT Index includes healthcare stocks pulled from the 98% of the total US stock market capitalization.

to individuate its main sub-segments:

1. Healthcare equipment and devices;
2. Pharmaceuticals, biotechnology and other related life sciences;
3. Providers and services.

This classification is useful for designing a portfolio of assets of the profit maximizing players of the healthcare industry.

In our analysis we include an index for every component of each sub-segment of the healthcare industry, thus the portfolio is composed of the following seven indices:

1. the iShare NASDAQ Biotechnology ETF or IBB for the biotech segment;
2. the iShare US Healthcare Providers ETF or IHF, for manufacturers and distributors of the medical devices;
3. the iShare US Healthcare Providers ETF or IHF, for manufacturers and distributors of the medical devices;
4. the Invesco Dynamic Pharmaceutical Portfolio ETF or PJP, that expresses the pharmaceutical companies' performances;
5. The SPDR S&P Health Care Equipment ETF or XHE, for healthcare equipment and supplies companies;
6. the SPD S&P Healthcare Services ETF or XHS for services entities;
7. Apple Inc. or AAPL for the mobile health. The inclusion of this stock highlights the impact of big data on the healthcare industry, passing through the smart wearable technology market. Apple entered into the healthcare market in 2014, with its Healthapp and Healthkit, but the real success derives from the release of the ResearchKit and the Apple Watch. Since their introductions and the growing interest of media, Apple technologies are experiencing a continuous diffusion among the world and within the medical sector, so that the company had shift its researching activities from wellness to medical.

Given the traditional models of the GARCH family that are able to model the excess of kurtosis and to capture the leverage effect, new volatility models have been introduced into the econometric and the financial literature (from seasonality models to the multivariate ones, also with long memory). This paper applies the Dynamic Conditional Correlation Model to the VHT Index and to the seven segment-specific stock indices within the healthcare industry of the US market, and evaluate the benefits of diversifying the portfolio, that arise from correlation analysis.

The remainder of the paper is organized as follows. In Section 2 we include an overview of time series idea and methods used in public health and biomedical research. Section 3 describes the data and the model used in the empirical analysis. Section 4 presents the results of the application to the seven-segment specific stocks of the healthcare US industry and the VHT Index.

Overview

Time series methods are widely used in econometrics and finance: the econometricians are usually interested in determining how much one variable will change in response to a variation of others variable, and to forecast the size of the error. In this context, the GARCH models have been introduced and since their preliminary implementations, their useful characteristics have led to a growing interest of the other sciences. From a rapid search on PubMed, it's possible to identify what are the topics where time-series methods have been employed.

The diffusion of technology has increased the availability of the data, producing time-series. For example, time series microarray experiments are used to study dynamical biology processes, such as gene expression [2, 3]. Huynh-Thu VA et al [4] used time-series gene expression data to infer the casual relationship among genes, by evaluating the performance of GENIE3[2] method and its dynamic variant, dynGENIE3. Time series methods have been increasingly adopted in physiological studies. In this field of research, there are many practical applications, that range from the image analysis (PET, MRI) to the patient data monitoring.

Fritson et al [5] in their paper presented a new method in order to individuate specific correlations between sensory input and brain physiologic response, by transforming the latter in a time series. Richman and Moorman [6] developed a new measure of entropy using a non-linear estimation method for clinical cardiovascular time series. In the same way, time-series have been used in epidemiology studies, investigating the relationship between different types of exposures (such as pollution, weather or environmental variables) and health conditions

[2] The GENIE3 is a new algorithm for the inference of GNRs, using high throughput genome data.

or health outcomes (such as mortality rate, diseases, infections, hospital admissions or drug prescriptions. [7, 8, 9, 10].

Finally, time-series analysis has been employed to evaluate the effectiveness of healthcare plans and their links with economic performances.

Even if this brief review shows how time-series methods have been used for designing medical experiments, this paper uses time-series analysis for investigating the healthcare industry from an economic point of view, and attempts to assess the possibilities of portfolio diversification strategy.

In fact, even if economists and econometricians usually work with time-series methods in order to examine the co-movement relationship between different variables across markets and countries, in various microeconomics and macroeconomics areas, they haven't paid much attention on the co-movements in the healthcare industry. With respect to this sector, previous studies are limited and the existing researches are based on wavelet analysis, usually in an international framework of analysis. For example, Chen et al [11] explored the dynamic relationships among DE, UK, and US healthcare sector returns for 1992-2013, and Chen [12] used the wavelet analysis to analyze the OECD healthcare financing and its correlation with macroeconomics variables.

Therefore, this study investigates the co-movements of healthcare industry within the US market, considering the particular public-private mix of entities and the predominant role of private insurance companies and financial market. It covers a different and more recent period of analysis, (from 2010 to 2018), due to the availability of daily data and concentrates on financial healthcare indices. Moreover, the estimation involves the Factor DCC model.

Data and empirical model

The empirical analysis is based on a series of daily log returns of a seven-composite portfolio assets and the VHT Index.

Data has been extracted from Yahoo Finance, and relates to the years 2010 to 2018 (8 years). The included stocks are the iShare NASDAQ Biotechnology ETF (IBB), the iShare US Healthcare Providers ETF (IHF), the iShare Dow Jones U.S. Medical Devices Index Fund (IHI), the Invesco Dynamic Pharmaceutical Portfolio ETF (PJP), Apple Inc. (AAPL), the SPDR S&P Health Care Equipment ETF (XHE), the SPD S&P Healthcare Services ETF (XHS). The dataset runs from 31, December 2009 to 28, November 2018, providing a total of 17.925 observations (2242 observations for every index). Table 1 reports the descriptive statistics for the time-series included in our analysis.

Figure 2 shows daily variation of each series, considering the adjusted close prices.

Those data are able to represent the healthcare industry in its complex configuration. All the data were coded using MATLAB and WinRats 10.0.

Instead of using daily adjusted close prices of each asset, the logarithmic returns are created and they are rescaled through multiplication by 100, in order to avoid for small variances and small parameters too.

From Figure 3 it can be visually noted that every index in our analysis doesn't exhibit trend or seasonality components and this often indicates that time series are stationary.

Table 2 confirms stationary of all the time series through the Augmented Dickey and Fuller (ADF) test, by selecting 13 lags (an approximation of the cube root of the number of observations as the order of the test).

In order to decide a model for the mean, the @varlagselect procedure is run and the minimum big lag length is one, as confirmed by the Akaike criterion and the Hannan and Quinn algorithm, so that 1 lag is chosen. Then, a test for the presence of multivariate arch effect is run, showing that the maximum significance (p 0.000*), and this implies that the covariance matrix is not constant but is has some 2nd order dependence. In a sense, this represent a useful guideline for the choice of the model. Because of covariance matrix is not constant and the correlations in portfolio matters, the basic idea is to look at the component of our composite portfolio and estimate a Dynamic Conditional Correlation (DCC) model, with linkage to the VHT index. Then, we will verify the implications of covariance modelling with DCC for portfolio management theory.

The DCC model, introduced by Engle [13], imposes a GARCH dynamics on both the conditional correlations and the conditional variances.

It allows correlations to be time-varying, through the following parametrization:

$$\Sigma_t = D_t R_t D_t$$

Where the variance covariance matrix Σ_t is decomposed into a diagonal matrix D_t and the time varying correlation matrix R_t.

For the specification of conditional variance, Engle suggested a covariance matrix, that's a matrix weighted average of Bollerslev CCC model [14] and a diagonal BEKK [15].
Consequently, for the correlation matrix, the simplest specification seems to be the exponential smoother:

$$p_{i,j,t} = \frac{\sum_{s=1}^{t-1} \lambda^s \varepsilon_{i,t-s}\varepsilon_{j,t-s}}{\sqrt{(\sum_{s=1}^{t-1} \lambda^s \varepsilon_{i,t-s}^2 \varepsilon_{j,t-s}^2)}} = [R_t]_{i,j}$$

and correlation could be constructed through the exponential smoothing, or, alternatively, through a GARCH (1,1) model:

$$q_{i,j,t} = \bar{p}_{i,j} + \alpha\left(\varepsilon_{i,t-1}\varepsilon_{j,t-1} - \bar{p}_{i,j}\right) + \beta(q_{i,j,t} - \bar{p}_{i,j})$$

Where $\bar{p}_{i,j}$ is the unconditional correlation

between $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$ and the correlation estimator is:

$$p_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}q_{j,j,t}}}$$

Given the DCCE specification, we get:

$$r_t | I_{t-1} \sim N\ (0, D_t R_t D_t)$$

$$D_t^2 = diag\ \{w_i\} + diag\ \{x_i\} \odot r_{t-1}r'_{t-1} + diag\ \{\lambda_i\} \odot D_{t-1}^2$$

$$\varepsilon_t = D_t^{-1} r_t$$

$$Q_t = S \odot (\iota\iota' - A - B) + A \odot \varepsilon_{t-1}\varepsilon'_{t-1} + B \odot Q_{t-1}$$

$$R_t = diag\ \{Q_t\}^{-1} Q_t\ diag\ \{Q_t\}^{-1}$$

The first equation represents the assumption of normality, that allow us to consider the likelihood function. The second equation means that each asset follows a univariate GARCH model, while the fourth equation represents the generalized DCC of Engle, that enables to overcome the original problem of equal dynamic for all correlations, even if the number of parameters dramatically increase.

A useful feature of DCC model is the two-step estimation, because the log-likelihood can be written as the sum of a mean and a volatility part and a correlation component. the log-likelihood consists of two parts: a volatility one and a correlation component, so that:

$$L(\Theta, \phi) = L_v\ (\Theta) + L_c(\Theta, \phi)$$

Where:

$$L_v(\Theta) = -\frac{1}{2}\sum_{t=1}(nlog(2\Pi) + \log|D_t|^2 + r_t' D_t^{-2} r_t)$$

and

$$L_c(\Theta, \phi) = -\frac{1}{2}\sum_{t=1}(\log|R_t| + \varepsilon_t' R_t^{-1}\varepsilon_t - \varepsilon_t'\varepsilon_t)$$

This specification simplifies the estimation and the results will be consistent.
Within this framework, we insert the preliminary step of factor extraction from the dataset.

Results and discussion

Following the parametrization discussed above, we estimate a DCC model on the matrix of log returns, for all the assets. We include in the estimation process the VHT index.
Moreover, since the goal of this paper is to analyze the benefits that investors could get from diversification strategy, we preliminary estimate a Factor GARCH model. For the latter, we decide to extract one common factor. Figure 4 shows the factor loading, that's the contribution of each series to the factor. It differs largely across series, and the greater contribution comes from the IBB and the PJP indices. The estimates also account for the heteroskedasticity of the idiosyncratic component. As literature suggests, a univariate GARCH model is applied to each series of disturbance, in order to get the conditional variance matrix of the idiosyncratic component, showed in Figure 5. We don't model the conditional mean of the idiosyncratic part, but if it's modeled an ARMA model should be used.
From the first step of estimation, we also note, as reported in Figure 6, that market and conditional volatilities follow the same direction of the common factor and the only difference is due to the range of variation for each series; it's possible to note that series that are more closed to the factor are the IBB and PJP indices, that are the more market-based sub-segments of the healthcare industry.
On the other hand, we fit the DCC model with lagged symmetric innovations and lagged correlations, DCC (1,1), as determined before by @ varlagselect procedure. The DCC parameters are α (0.0139) and β (0.9831), and show a persistence of 0.997.

Figure 7 plots the correspondent conditional volatilities. As we can see, they are time-varying in ways that are typical of GARCH models and we can individuate different spikes in the graph.
The first one occurs in the middle of 2011 (August), when Wall Street crashed, after the credit rating downgrade by S&P of US sovereign debt and it affected all the assets included in our analysis. Although its magnitude is relatively small respect to the previous one, another spike can be individuated in 2014, while another period of fluctuations occurs between 2015 and 2016.
This latter is, of course, influenced by the presidential campaign of Donald Trump and his plan for replacing the Affordable Care Act (ACA), commonly called Obama Care, even if there are no relevant spikes in 2017, when the American Health Care Act (AHCA), or Trumpcare, was approved.
The time-varying correlations (vs. VHT) are plotted in Figure 8. From the graph it's clear that no significant changes and volatility clustering occurred in the exam period. It can be seen that the correlations are positive, for all the assets, except for the AAPL, but this result could be justified from the fact that Apple officially entered in the healthcare market in 2014. Empirical results in this section show that all the sub-segments of the healthcare industry are strongly correlated. Moreover, the correlations are higher for IBB, IHF and PJP respect to the other assets of the analysis.
Finally, the out of sample evaluation is carried out within a portfolio optimization framework. As the healthcare industry is one of the fastest growing industry in the USA economy, there is no doubts about its attractiveness for investors. It also clear that they often are searching for high and stable returns, so that with this condition, we investigate the benefit from investors that could arise from diversification in the US healthcare industry. In the classic asset allocation problem, the weights are chosen in order to minimize the portfolio variance; empirically, this implies the estimation of the mean and the covariance of the asset returns. We consider the classic problem of determining the Global Minimum Variance of the portfolio

$$w^{GMV} = \min w' \sum w$$

s.t $w'1=1$, with w the vector of weights and $\sum$ the variance covariance matrix.
For the latter, we use two different specifications: the sample variance and the DCC variance-covariance matrix. Table 3 reports the weights obtained from the estimation.
The first column reports the weights in case of sample variance -covariance, the second one in the case of DCC estimation and the last one corresponds to equal weights for all the asset in the portfolio. Immediately it can be noted that we have different ponderations for the weights, even if some similitudes persist. In fact, in the first two columns, we can easily see that for some assets the weights are negative: this happens for IBB, IHF, IHI and PJP in the case of sample variance-covariance, while with DCC negative weights are associated with IBB, IHF, AAPL, XHE indices. In order to understand the meaning of the estimation results, we have to underline that we've constructed a composite portfolio with the main US healthcare assets, so that it could be more appropriate to interpret the weights in term of "exposure" to a given asset class. This implies that we have to take into account of them when we aggregate multiple positions in order to evaluate the net exposure to the risks. This is confirmed by the comparison of the performance of individual assets and those of the simulated portfolio. We report, for the former the standard deviations and for the latter, the GMV portfolio based on sample covariance, GMV portfolio based on DCC estimation and 1/N portfolio.
DCC shows smaller volatility of returns, confirming that diversifying is a good strategy, particularly when time-series methods allow to exploit the pattern of dynamic correlations between the assets.

Conclusion

This paper examines the volatility and the co-movements of the VHT index and other seven assets, which express the performances of the sub-segments of the US healthcare industry, using DCC-GARCH framework of Engle, and modeling the correlations with the market through the leading common factor. This provides evidence in favor of the portfolio diversification also in the healthcare industry, and traders and governments could use such information to evaluate their strategies. Governments could base the introduction of healthcare reforms considering their hypothetical impacts on the stability of the returns of financial assets present in the industry. Moreover, since the healthcare system of US is a hybrid one, with DCC model

(based on time varying mean and covariances of returns), is possible to better individuate the connection between plans for guarantee high quality healthcare and private strategies intended to maximize shareholder's value.
Our finding also indicate that all the sub-segments of healthcare sector move together in the same direction, so that co-movements exist, and this result highlights the importance for traders of taking into account the time varying properties of the industry's stock returns co-movements in the portfolio maximization problem.

Table 1. Descriptive Statistics

Stock	N. Obs.	Sample mean	Variance	Standard Error
VHT	2244	0,057718	0,918182	0,958218
IBB	2244	0,06202	2,197886	1,482527
IHF	2244	0,064256	1,228745	1,108488
IHI	2244	0,013989	6,74184	2,543981
PJP	2244	0,067341	1,428431	1,195170
AAPL	2244	0,096014	2,612162	1,616218
XHE	2244	0,054600	1,037906	1,018777
XHS	2244	0,092512	4,647733	2,15586

Table 1. Descriptive Statistics. Table 1 includes marginal statistics for time-series stocks.

Table 2. Dickey-Fuller Unit Root Test

	Dickey-Fuller Unit Root Test
Stock	T-Statistic
VHT	-14.0381**
STW.PA	-14.0247**
IBB	-13.8904**
IHF	-13.7515**
IHI	-13.4439**
PJP	-13.6954**
AAPL	-12.1745**
XHE	-13.2281**
XHS	-13.7845**

Table 2. Dickey-Fuller Unit Root Test. This table shows the result of the ADF test for unit root for each stock in our analysis. All the series are stationary, and the results are also confirmed through the KPSS test for trend stationary.

Table 3. Simulated portfolio's weights

	w_SampeCov	w_DCC	w_Eq
VHT	1,51654	0,92788	0,12500
IBB	-0,47956	-0,43467	0,12500
IHF	-0,33106	-0,36200	0,12500
IHI	-0,32114	0,36458	0,12500
PJP	-0,12292	0,27272	0,12500
AAPL	0,08640	-0,01819	0,12500
XHE	0,26468	-0,17588	0,12500
XHS	0,38706	0,42556	0,12500

Table 3. Simulated portfolio's weights. The table shows the different weights obtained from the estimation. The first column considers the weights estimated by the sample variance-covariance matrix; the second column reports weights obtain through DCC variance-covariance matrix and the last one the uniform weights. It can be noted that only for IHF and IBB indices, the weights are negative in the first two estimations, even if their size are slightly different, while for IHI, PJP, AAPL and XHE, the sign of the weights are the opposite.

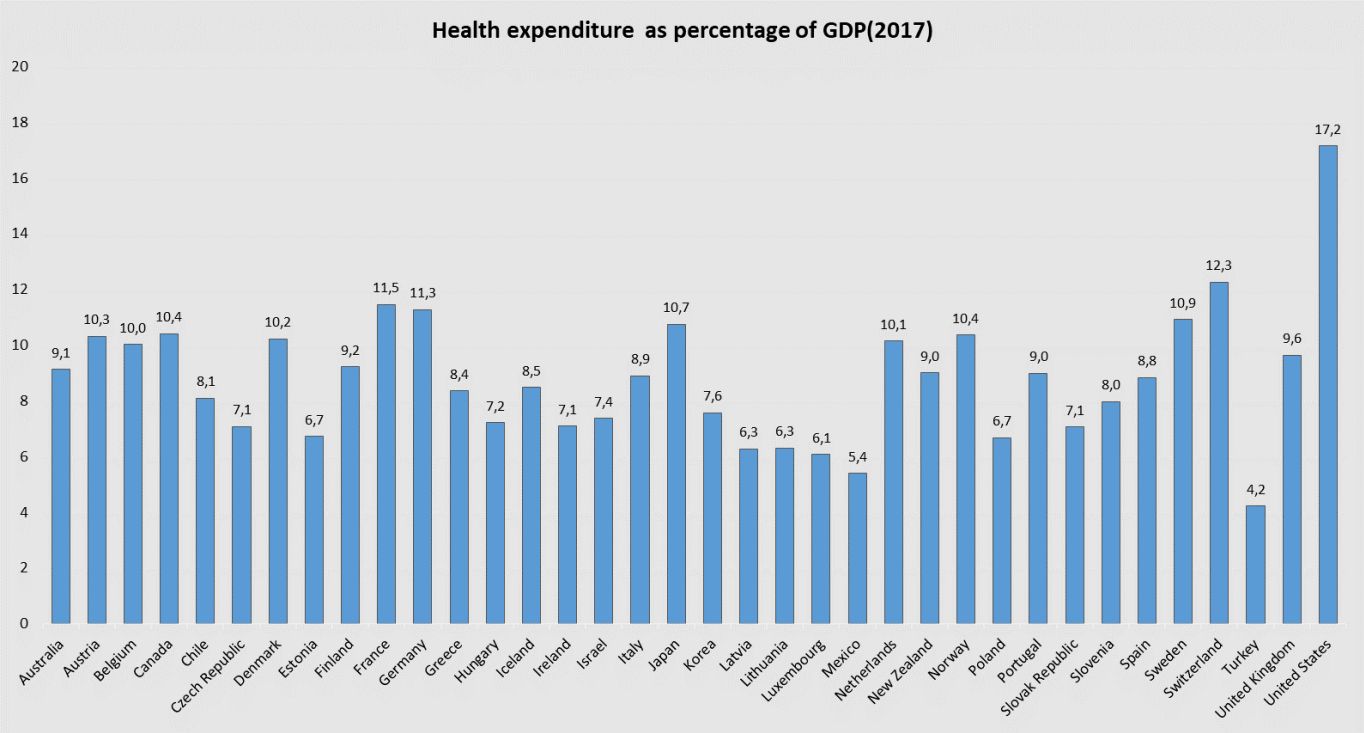

Figure 1. Health Expenditure (% GDP). Figure 1 shows the health expenditure, as percentage of GDP, for all the country. US spends 17.2% of GDP, the highest value respect to all the other countries in the world.

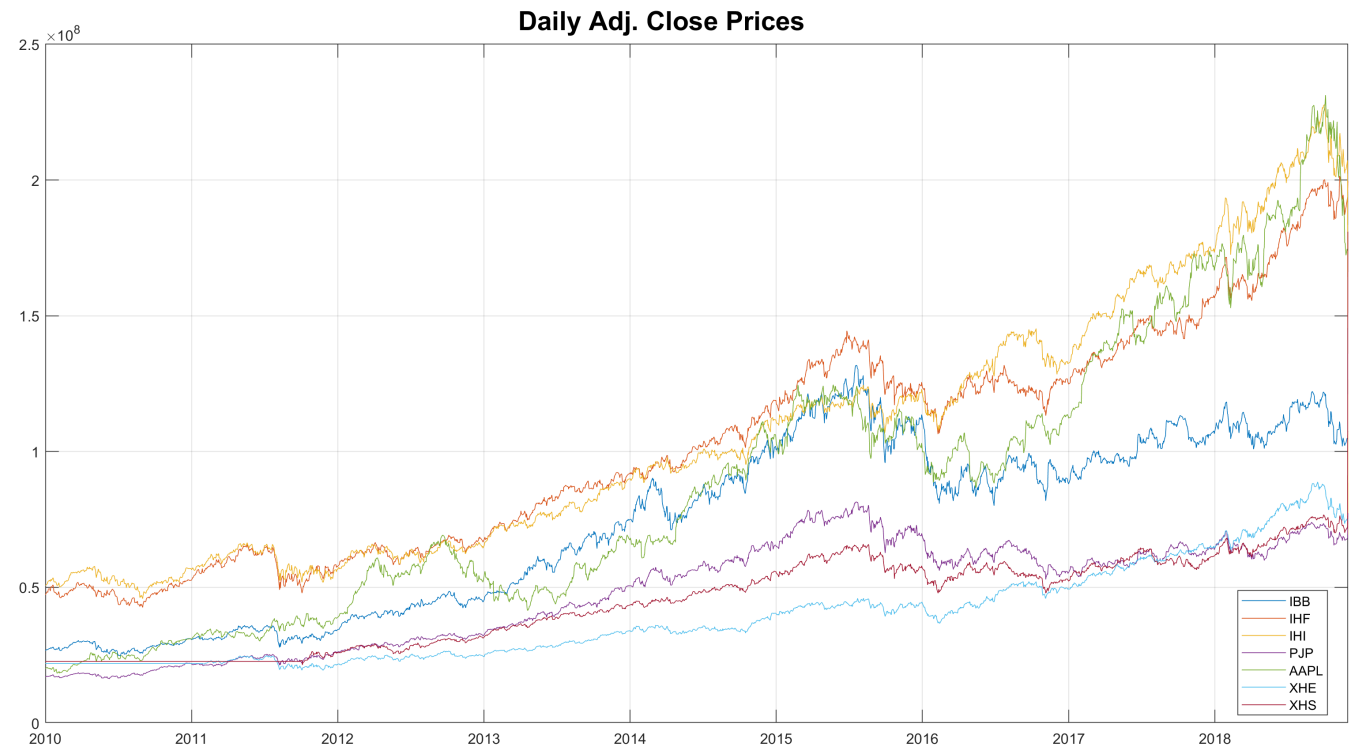

Figure 2. Daily Adj. Close Prices. Figure 2 shows daily adjusted close prices for the series included in the seven-portfolio assets, and the bottom part displays the daily time-series for the VHT index.

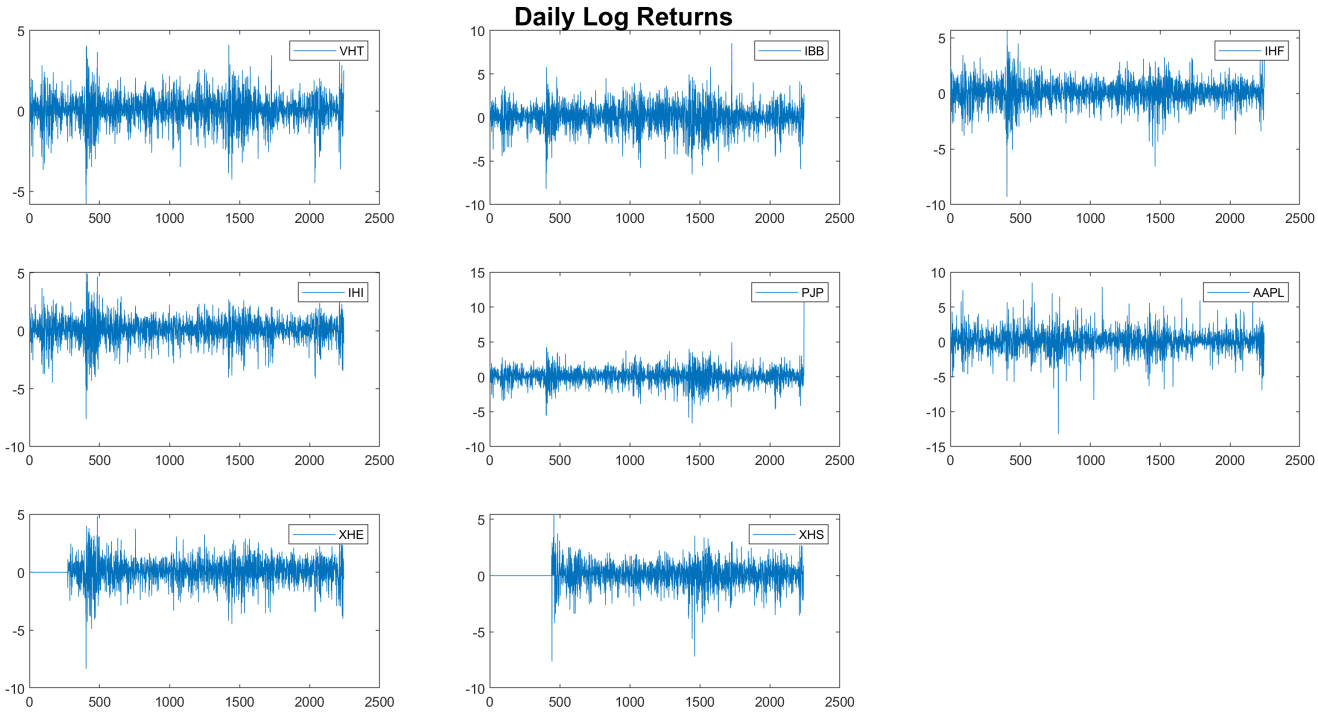

Figure 3. Daily Log Returns. Figure 3 shows daily log returns of each series included in the analysis.

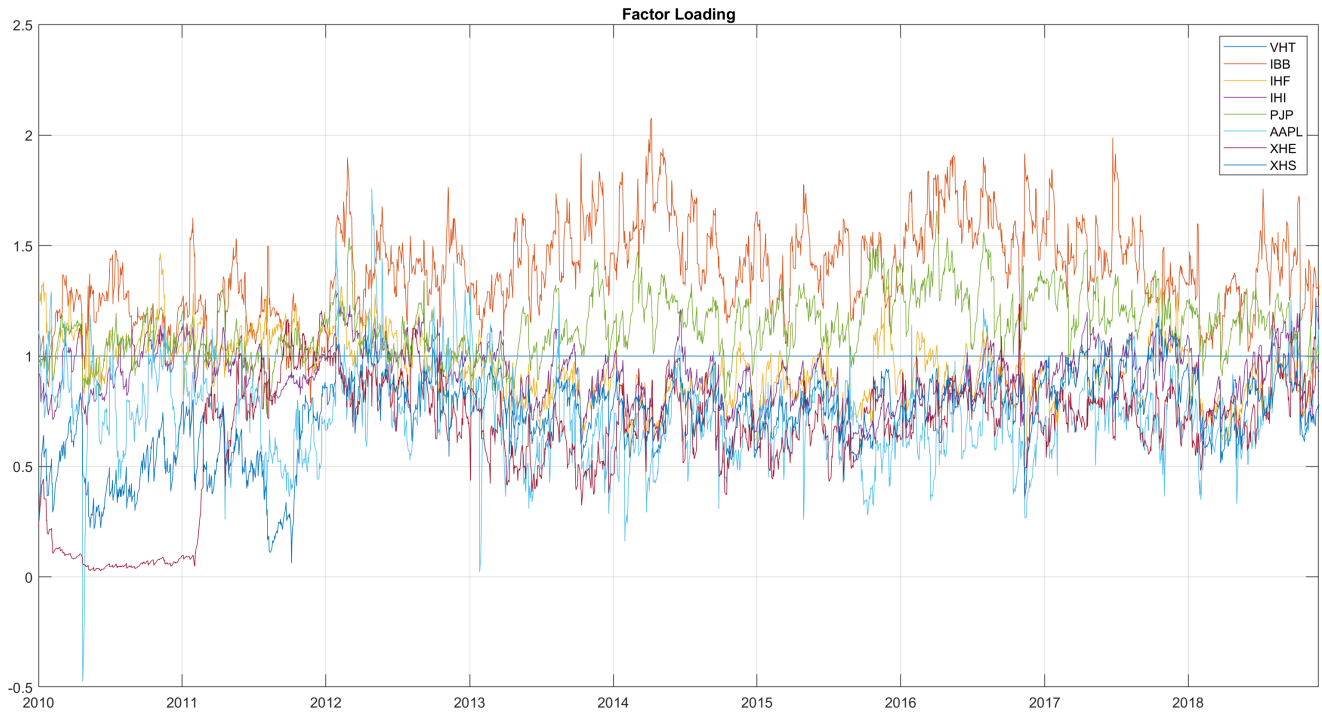

Figure 4. Factor Loading. The figure represents the contribution of each series to the determination of the common factor.

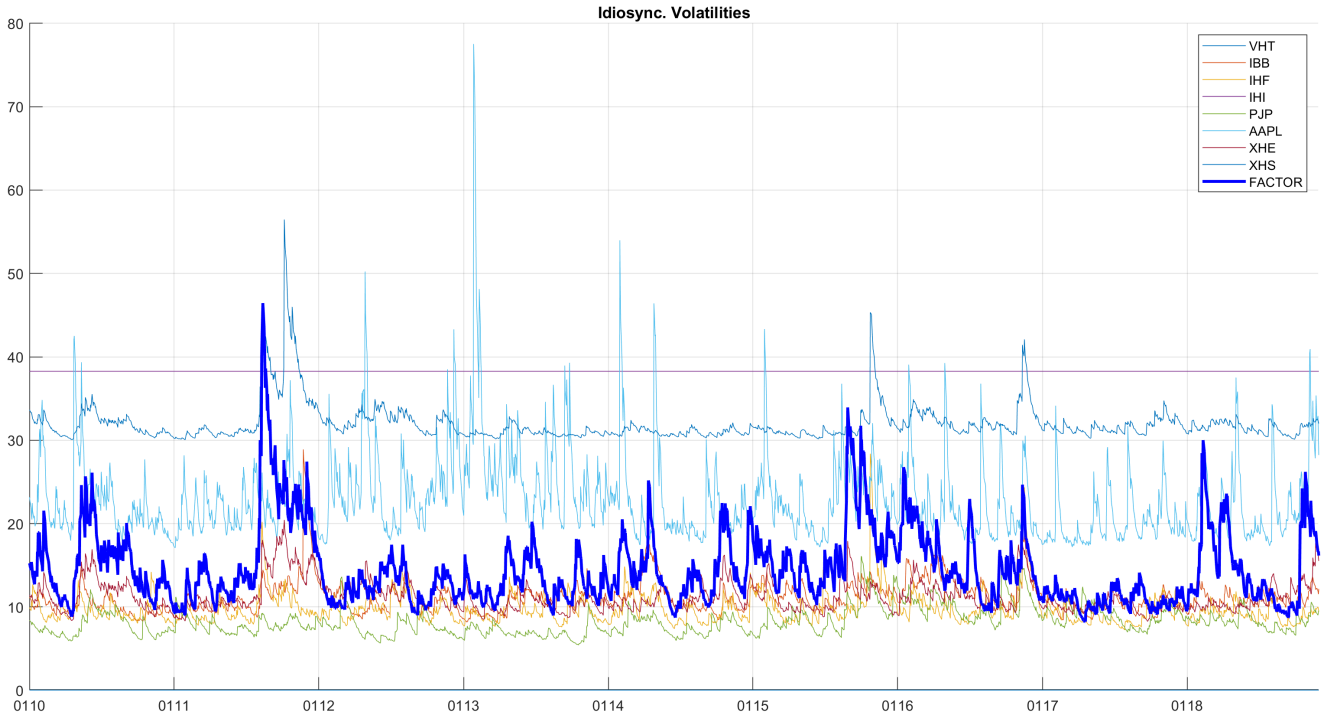

Figure 5. Idiosyncratic Volatilities. Figure 5 shows the idiosyncratic volatilities estimated via the Factor GARCH model, on VHT Index.

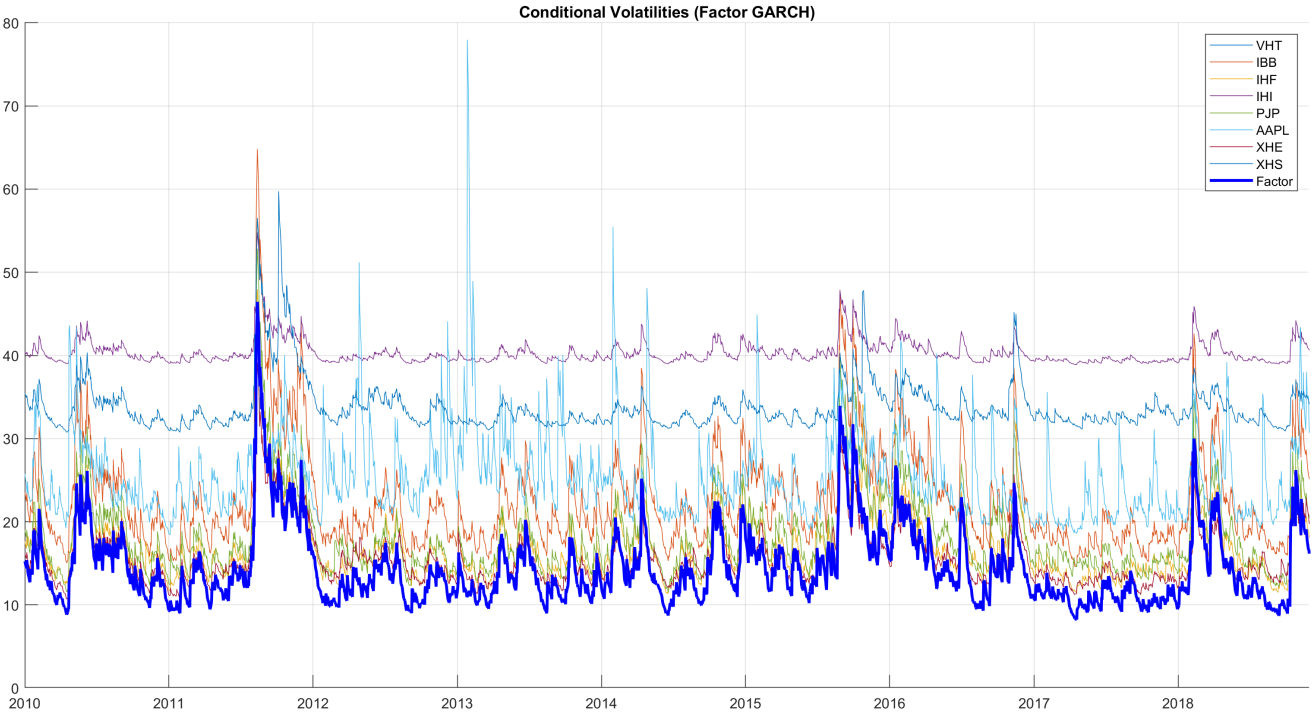

Figure 6. Conditional Volatilities (Factor GARCH). Figure 6 shows the co-movement of conditional volatilities and common factor, obtained from the first stage of estimation.

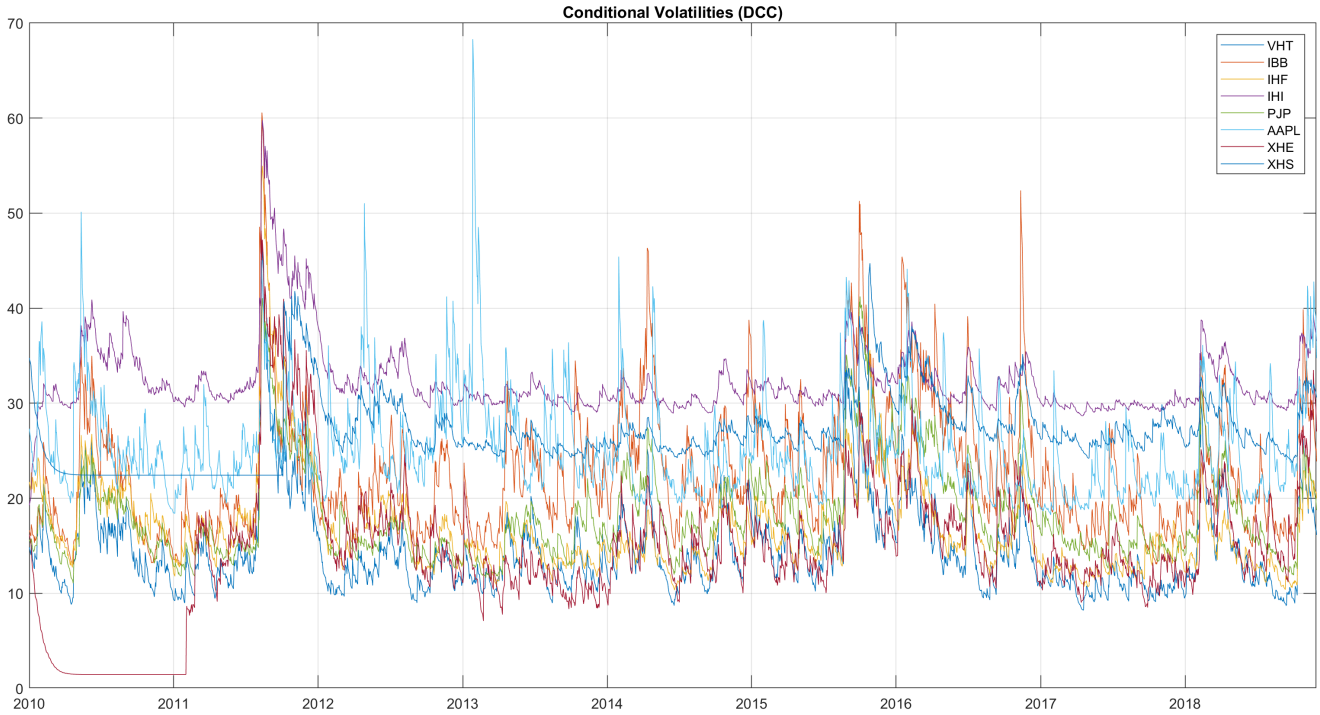

Figure 7. Conditional Volatilities (DCC). The figure underlines the typical GARCH patterns for conditional volatilities and individuates the main economic events that determine spikes in volatility series.

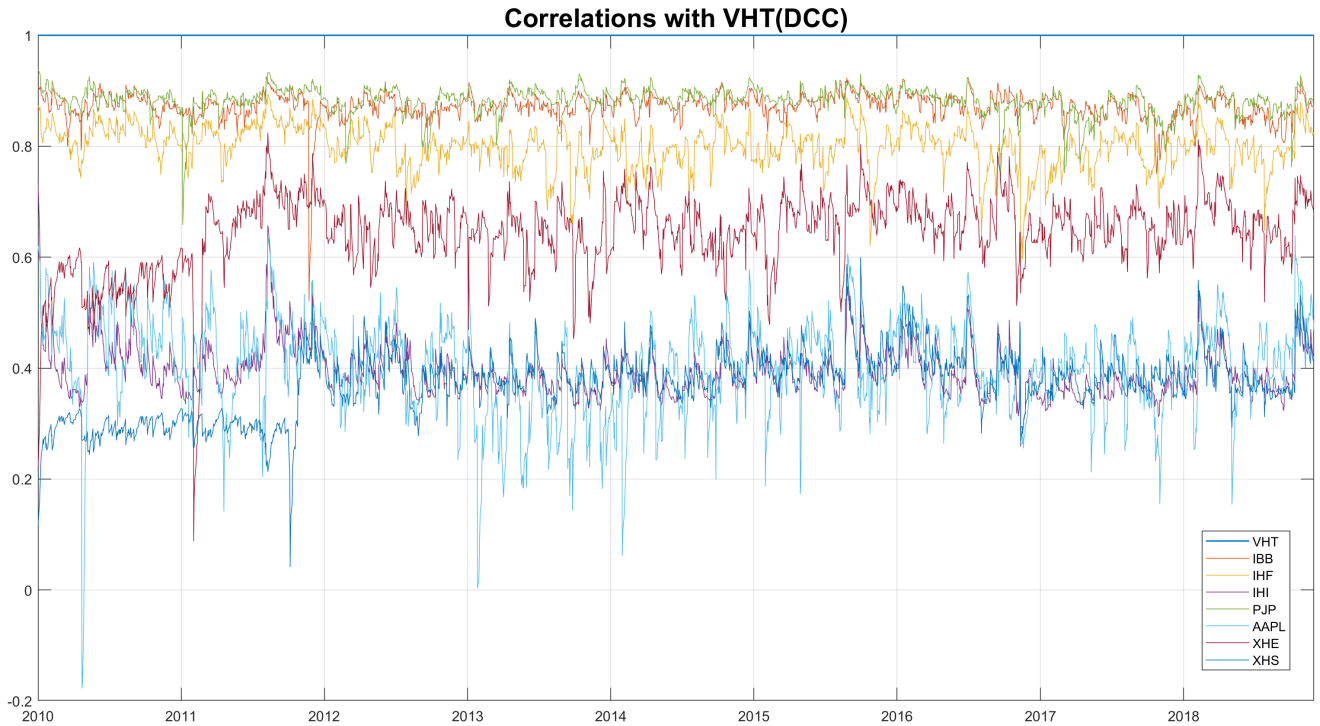

Figure 8. Correlations with VHT (DCC). Figure 8 evidences the correlation with VHT Index, showing the strongest correlations with IBB and PJP indices.

References

[1] Ridic, Goran & Gleason, Suzanne & Ridic. Comparisons of Health Care Systems in the United States, Germany and Canada. Materia socio-medica 2012; 24.

[2] Filkov, Vladimir & Skiena, Steven & Zhi, Jizu. Analysis Techniques for Microarray Time-Series Data. Journal of computational biology: a journal of computational molecular cell biology 2002; 9: 317-30.

[3] Ernst, Jason & Bar-Joseph, Ziv. STEM: a tool for the analysis of short time series gene expression data. BMC bioinformatics 2006; 7. 191.

[4] Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. PLOS ONE 2010; 5(9): e12776.

[5] Friston, K. J., Jezzard, P. and Turner, R.. Analysis of functional MRI time-series. Hum. Brain Mapp 1994; 1: 153-171.

[6] Richman, Joshua & Moorman, Joseph. (2000). Physiological Time-Series Analysis Using Approximate Entropy and Sample Entropy. American journal of physiology. Heart and circulatory physiology 2000; 278. H2039-49.

[7] William Checkley, Leonardo D Epstein, Robert H Gilman, Dante Figueroa, Rosa I Cama, Jonathan A Patz, Robert E Black. (2000). Effects of El Niño and ambient temperature on hospital admissions for diarrheal diseases in Peruvian children. The Lancet 2000; Volume 355, Issue 9202, Pages 442-450.

[8] Guijarro R, Trujillo-Santos J, Bernal-Lopez MR, de Miguel-Díez J, Villalobos A, Salazar C, Fernandez-Fernandez R, Guijarro-Contreras A, Gómez-Huelgas R, Monreal M. Trend and seasonality in hospitalizations for pulmonary embolism: a time-series analysis. J Thromb Haemost, 2015.

[9] Preti, Antonio & Lentini, Gianluca. Forecast models for suicide: Time-series analysis with data from Italy. Chronobiology International 2016; 33: 1-12.

[10] Hu, Y & Y Guan, M & Zhao, D & Chen, L & G Chen, F & Song, J.. Time-series analysis of association between SO(2), NO(2) pollution and daily emergency room visits in Shijiazhuang. Zhonghua yu fang yi xue za zhi Chinese journal of preventive medicine 2017; 51: 358-361.

[11] Chen, Mei-Ping & Chen, Wen-Yi & Tseng, Tseng-Chan. Co-movements of returns in the healthcare sectors from the US, UK, and Germany stock markets: Evidence from the continuous wavelet analyses. International Review of Economics & Finance, Elsevier 2017; vol. 49(C): pages 484-498.

[12] Chen, Wen-Yi. Co-Movement of Healthcare Financing in OECD Countries: Evidence from Discrete Wavelet Analyses. Romanian journal of economic forecasting 2016; 19: 40-56.

[13] Engle, Robert F.,May. Dynamic Conditional Correlation: A Simple Class of Multivariate GARCH Models. UCSD Economics Discussion Paper 2000; No. 2000-09.

[14] Bollerslev, Tim. Modelling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. The Review of Economics and Statistics, MIT Press 1990; vol. 72(3), pages 498-505.

[15] Engle, Robert F., and Kenneth F. Kroner.. Multivariate Simultaneous Generalized Arch. Econometric Theory 1995; vol 11, no. 1 (1995): 122-50.