PORTFOLIO SELECTION DURING CRISES USING PRINCIPAL COMPONENT ANALYSIS

Roxana IOAN[1]

DOI: 10.2478/tjeb-2020-0009

The current Covid-19 crisis, that debuted as an exogenous shock, has determined overreaction and herding among investors, as well as efficiency breaches and alterations within the Gaussian characteristics of returns' distributions by generating strong asymmetries. Under such circumstances, the – econophysics approached – correlation structure within the market has been affected to an uncertain extent. Within these conditions, the problem of optimal portfolio selection becomes a subject of interest even for professional investors who tend to seek refuge towards developed, mature economies and quality assets. The Principal Component Analysis manages to offer a considerably useful tool for the allocation problem, by reducing dimensionality and, at the same time, by being able to account for the "unobservable risk" hidden within the market. The method is able to provide well-diversified portfolios, by explaining the systematic risk component within the analyzed returns series. The paper analyzes the ability of the PCA method, used aside with a KMO Test, to provide high rates of return even during turbulent, extremely volatile periods, such as the current one. The results show that the PCA weighted portfolio manages to achieve a rate of return higher than the one generated by an equally weighted "market portfolio", proving therefore the robustness of the method.

Keywords: **Principal Component Analysis, Covid19 crisis, correlation matrix, Kaiser – Meyer – Olkin (KMO) Test**

JEL Classification: **C38, G11, D53**

[1] *Department of Economics and Economic Modelling, Faculty of Economics and Business Administration, West University of Timisoara, Romania*

DOI: 10.2478/tjeb-2020-0009

1. Introduction

The current Covid-19 crisis has dramatically affected the volatility, efficiency and global return of all capital markets. This crisis represents an unprecedented worldwide shock, not being contained inside a certain country, region or continent, but spread all over the world and affecting all categories of people. At the beginning, the Covid-19 crisis was regarded as an exogenous crisis in respect to financial markets, as it only influenced these markets at a psychological level, through the reactions of the investors. As the crisis evolution increased worldwide and governments drew restrictive measures, the crisis no longer affected the financial markets only from an exogenous point of view. Entire sectors of economic activity were and still are affected by restrictions, quarantine or major changes in the customers' behavior. Travels, tourism, restaurants, services, sales were all diminished and the global energy demand and consumption shrunk (Yoshino, Taghizadeh-Hesary & Otsuka, 2020). Morgan Stanley states on their website that the present Covid-19 crisis is accelerating de-globalization, as flows of goods, money and people are limited by restrictions. Under such circumstances, securities issuers from those fields were affected by this activity stagnation, causing high specific risk increases.

Currently, markets have to face overreaction and herding behaviors due to panic created by bad news concerning the economic crisis induced by restrictive measures, high unemployment and the downfall of entire activity sectors. Even professionals, as Morgan Stanley, who are most likely not to overreact to such environmental shocks state that they have to stay oriented to "quality companies with low debt capable of delivering steady earnings". Such erratic behaviors inside capital markets lead not only to allocational efficiency but also to informational efficiency breaches. During periods of time characterized by such herding behaviors, important changes occur within the correlation structures arising inside the markets, as well as within the uncertainty in the price formation process.

It is widely accepted that market returns tend to display Gaussian resembling distributions during non-turbulent times, often characterized by fat tails (Crescimanna & Di Persio, 2016). The fat tails are confirmed by others (Han, 2013) whose work shows that these are strongly connected with the performance of the issuer. In fact, herding behaviors such the ones created within the present pandemic crisis are believed to be the very cause of asymmetric distributions, characterized by evident fat tails (Crescimanna & Di Persio, 2016). The shocks generated by the sanitary crisis have caused high asymmetries within the daily returns series, determining the alteration of their Gaussian properties. Such changes are considered as important marks for determining crash moments (Benazzoli & Di Persio, 2016). Under such circumstances, those risk estimation methods having normality assumptions became unusable, as their results would have been seriously biased. In these cases, when many risk estimation methods tend to provide misleading results, classic portfolio selection methods have to be replaced with alternative selection methods, not relying on the existence of normality assumptions.

In this respect, the economic literature contains papers and studies that indicate the usefulness of Principal Component Analysis in portfolio management. Pasini (2017) emphasized the use of Principal Component Analysis both on homogeneous and heterogeneous groups of assets in order to optimize portfolios and obtain best returns. Avellaneda (2020) states that the method is suitable for markets containing asynchronous price information, such as the mature and developed markets, as within these types of markets stock prices seem not to move together as much as within emerging markets from poor economies (Morck, Yeung & Yu, 2000). Konarzewska (2013) proves the utility of PCA as an algorithm for portfolio components selection, as well as for optimal portfolio construction.

Other works (Yang et al., 2017) indicate the Kaiser – Meyer – Olkin Test as a good selection method, besides the PCA. This idea is also confirmed by the paper of Benny (2018), that indicates the KMO Test to be a proper measure for sample adequacy that should be used in order to determine if a factor analysis would be suitable for the chosen financial data.

Yang (2015) proves the utility of the method from a performance analysis point of view. Principal Component Analysis is also used as a performance measure by Tan (2012), who also emphasized the utility of the method in optimizing the Black-Litterman allocation model. His paper proposes a PCA based method able to improve the stability of the classic PCA analysis. The results obtained by Kozak et al (2017) show the ability of PCA to drastically reduce dimensionality within data, as within their results a very small number of principal components manage to capture almost the entire explained variance in an out-of-sample analysis. Nobi et al. (2013) conclude that a within their data (20 financial indexes, daily data between 2006 – 2011), a very small number of eigenvectors show a large part of the relevant information, showing thus that the reduction of dimensionality does not affect the overall quantity of information provided and, moreover, that the most relevant part of information is comprised by the first few principal components. The most important result of this paper is that the authors note that this effect is even more prominent and observable in times of financial crises, proving the PCA to be a method proper to use during such times.

Yaohao et al. (2019) propose an improvement of the Principal Component Analysis, by introducing certain nonlinear interactions to the covariance matrix used within the method, analyzing daily data from seven major financial markets. Their results show significant improvements in the Sharpe and Sortino ratios of the constructed portfolios.

2. Methodology and Data

As stated before, crises are characterized by changes occurring within the correlation structures existing inside the markets, as well as within the uncertainty of returns. These correlations are the feature that makes diversification possible, but a comprehensive correlation analysis when studying many assets is a difficult and time-consuming task (Yang, 2015). Instead, by using the Principal Component Analysis method, the complexity of such a problem is reduced. While preserving the variation given by the correlations (or covariances), the method retains a few uncorrelated principal components instead of analyzing hundreds of correlations (Jolliffe, 1986).

Principal Component Analysis (PCA) represents a multivariate analysis method able to deal with interrelated variables (Pasini, 2017), by reducing the data set's dimensionality. In fact, PCA finds uncorrelated linear transformations within the data, characterized by high variance (Boulesteix, 2005). The method's goal is actually to maximize the variance of that linear combination of analyzed variables. A main advantage of the method is that it can be used to heterogeneous groups of variables, such as the stocks analyzed within this paper, that come from different activity fields and that have been differently influenced or affected by the pandemics. As stated by Yang (2015), the most important advantage of PCA is that it manages to measure the concentration of the analyzed market indicating in the meanwhile the diversification possibilities.

PCA can be applied using either the covariance or the correlation of data (Jolliffe, 1986), depending on the type of data used. In this paper, we chose to use the correlation, as we consider this method much more informative and able to capture certain structures within the data (idea also confirmed by Yang, 2015).

The PCA method also solves the highly important question of the number of assets to be included in a well-diversified portfolio, as it identifies a sufficient number of assets that are representative for the whole data set. When applying the PCA method, one can use either one of two methods designed for choosing the proper number of principal components to be retained: Cumulative Variance and Kaiser's Rule (Pasini, 2017). The Cumulative Variance rule consists of retaining the first components cumulating a percentage of 80% – 90% of the total variation. So, the proper number of principal components to be retained is the smallest number of components that manage to exceed the target percentage. Kaiser's Rule takes into account the sizes of the variances of principal components, retaining those components that display variances higher than 1.

Before applying the PCA procedure, we consider appropriate to conduct a Kaiser – Meyer – Olkin (KMO) Test, in order to measure the sampling adequacy of the analyzed assets, the test providing information both overall and for each analyzed asset individually (Kaiser, 1970; Cerny & Kaiser, 1977; Dziuban & Shirkey, 1974). In fact, the KMO test's statistic summarizes information regarding the partial correlation. Generally speaking, a KMO statistic between 0.8 and 1 indicates that the sample of data was well chosen and it is proper for further PCA analysis. A KMO statistic below 0.6 indicates a sample that should be modified. A KMO statistic close to 0 indicates large partial correlations in respect to the sum of correlations. So the sample would be completely wrong chosen. Generally, a KMO statistic exceeding 0.8 is considered to be an indicator for the fact that a PCA analysis would be properly conducted on the selected assets. So, an asset selection, realized on a correlation basis, will be done using the KMO test. The PCA will furtherly provide a weights allocation method that will be used in order to create a portfolio intended to provide a return superior to the "market portfolio" return.

In order to simplify the analysis, the present paper is only focused on stocks. We chose to study the selection of a well-diversified portfolio based on the 30 assets composing the Dow Jones Industrial Average (DJIA), a main market index in the United States of America and also seen as a main global market index. The reason behind choosing to study U.S. stocks is the intention of analyzing the performance of quality stocks, from a developed, mature economy, during crisis times. Moreover, Morgan Stanley, for example, advise investors against Emerging Markets investments, stating that their main focus should be towards issuers and economies with at least stable if not improving growth, capable to generate expected returns. At some point, they even stated to have built their portfolio with "domestic" assets, referring to U.S. securities. The time period taken into account for analysis starts at the beginning of 2020 (1st of January 2020), when the first news regarding the fact that a highly contagious disease was affecting China appeared in Europe and in the USA. The studied period ends at 30th of September, before the second wave of the pandemics hit the USA and Europe. We consider this time period containing the debut of the pandemics, the first lockdown applied worldwide and the relative stagnation in the summer to represent all the phases of this phenomenon known so far.

The entry data of the PCA method consists of the average value of daily log-returns, computed on a moving window of 30 days, thus providing a dynamic type of analysis, that always takes into account the most recent 30 days, instead of a static all-database study. In order to conduct the described procedure, R software was used.

3. Results and Discussions

As stated in the previous section, our study begins with a KMO test (results presented below), in order to select those assets with a KMO statistic larger than 0.8, indicating a non-large sum of partial correlations relative to the sum of correlations.

Table 1. KMO statistic for a 30 days moving window average (all 30 assets)

Kaiser-Meyer-Olkin factor adequacy														
Overall MSA = 0.91														
MSA for each item =														
V1	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**V10**	**V11**	**V12**	**V13**	**V14**	**V15**
0.91	0.78	0.86	0.89	0.92	0.89	0.88	0.94	0.94	0.92	0.93	0.95	0.93	0.93	0.84
V16	**V17**	**V18**	**V19**	**V20**	**V21**	**V22**	**V23**	**V24**	**V25**	**V26**	**V27**	**V28**	**V29**	**V30**
0.93	0.93	0.9	0.93	0.93	0.93	0.86	0.93	0.89	0.94	0.94	0.89	0.92	0.8	0.72

Source: own results, R software

The KMO test results presented in Table 1 indicate the future structure of the portfolio consisting on those assets exceeding a 0.8 KMO statistic value. The assets excluded from the portfolio are V2 (Amgen Inc.), V29 (Walgreens Boots Alliance, Inc.) and V30 (Walmart Inc.), as these have a KMO statistic lower than 0.8.

The remaining components are subjected to a new KMO test, in order to confirm the adequacy of our sampling. This time, all analyzed assets display KMO statistic values exceeding 0.8, as presented in Table 2.

Table 2. KMO statistic for a 30 days moving window average (selected 27 assets)

Kaiser-Meyer-Olkin factor adequacy													
Overall MSA = 0.92													
MSA for each item =													
V1	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**V10**	**V11**	**V12**	**V13**	**V14**	**V15**
0.93	0.9	0.91	0.93	0.9	0.86	0.93	0.93	0.92	0.96	0.94	0.94	0.91	0.86
V16	**V17**	**V18**	**V19**	**V20**	**V21**	**V22**	**V23**	**V24**	**V25**	**V26**	**V27**	**V28**	
0.92	0.93	0.93	0.94	0.96	0.95	0.87	0.95	0.89	0.93	0.93	0.89	0.94	

Source: own results, R software

As we have selected only those assets passing the Kaiser – Meyer – Olkin Test, we can proceed to the portfolio construction and its result evaluation. In this respect, furtherly, with the selected 27 assets we conduct the PCA analysis. Principal Components are presented in such an order that each successive principal component explains a decreasing amount of variation within the analyzed data. Table 3 presents the standard deviation, proportion of

DOI: 10.2478/tjeb-2020-0009

variance and cumulative proportion of variance, listed accordingly to the order of importance. Figure 1 presents the variances of the first principal components. Figure 2 shows the proportion of variance of the first principal components.

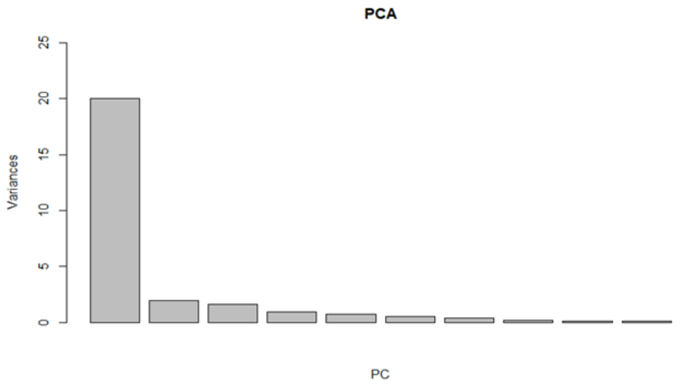
Both the Cumulative Variance and Kaiser's rule point out that PC1 explains a high enough proportion of the total variation, as the Proportion of Variance for PC1 is 0.7423 and the variance of PC1 is 20.0426 (the square of 4.4769), in comparison with the following components. The proportion of variance of PC1 is almost 10 times higher than the proportion of variance explained by PC2 (0.74230 for PC1 compared with 0.07146 for PC2). The variance of the first principal component is much higher than the variance of the following component (20.0426 for PC1 and 1.9293 for PC2). Of course, at a closer look we could conclude that the first two or even the first three principal components should be selected, but by analyzing the much higher proportion of variance of the first principal component we can see that this one is much more informative that any of the other components. From these points of view, we can conclude that the information provided by the first principal component is sufficiently explanatory for being able to contribute to creating a well-diversified, profitable portfolio.

Table 3. Principal Components' Standard deviation, Proportion of Variance and Cumulative Proportion of Variance

Importance of components:										
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**	**PC9**	**PC10**
Standard deviation	4.47690	1.38901	1.27897	0.94641	0.83130	0.70155	0.63652	0.41900	0.37294	0.31283
Proportion of Variance	0.74230	0.07146	0.06058	0.03317	0.02560	0.01823	0.01501	0.00650	0.00515	0.00362
Cumulative Proportion	0.74230	0.81378	0.87436	0.90754	0.93310	0.95136	0.96637	0.97290	0.97802	0.98165
	PC11	**PC12**	**PC13**	**PC14**	**PC15**	**PC16**	**PC17**	**PC18**	**PC19**	**PC20**
Standard deviation	0.27595	0.26939	0.23240	0.20630	0.19959	0.19162	0.17556	0.16734	0.14786	0.13527
Proportion of Variance	0.00282	0.00269	0.00200	0.00158	0.00148	0.00136	0.00114	0.00104	0.00081	0.00068
Cumulative Proportion	0.98447	0.98716	0.98920	0.99073	0.99221	0.99357	0.99471	0.99575	0.99656	0.99723
	PC21	**PC22**	**PC23**	**PC24**	**PC25**	**PC26**	**PC27**			
Standard deviation	0.13127	0.12388	0.11010	0.10410	0.09738	0.07271	0.06622			
Proportion of Variance	0.00064	0.00057	0.00045	0.00040	0.00035	0.00020	0.00016			
Cumulative Proportion	0.99787	0.99844	0.99889	0.99930	0.99964	0.99984	1.00000			

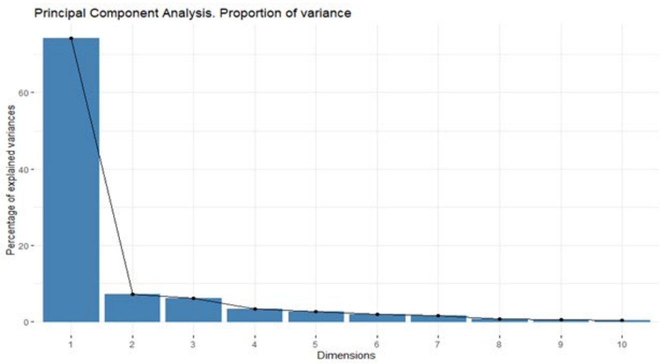
Source: own results, R software

DOI: 10.2478/tjeb-2020-0009



Source: own results, R software

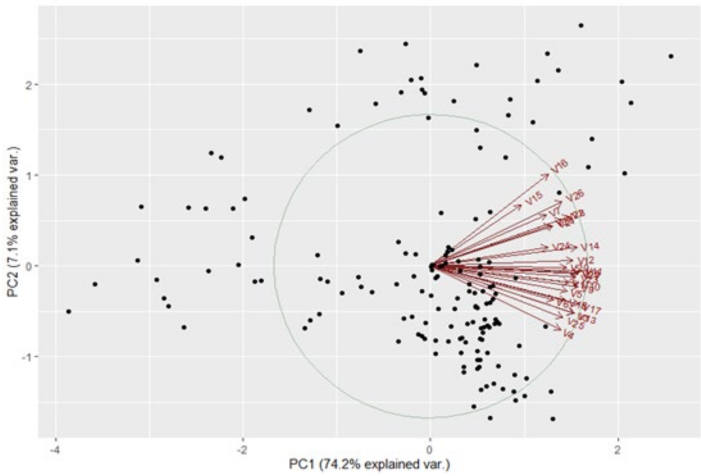
Figure 1. Main Principal Components' Variance



Source: own results, R software

Figure 2. Proportion of Variance

PCA is conducted by an eigen decomposition on a square matrix which, as stated before, in this paper is the correlation matrix of stocks' log-returns. Using eigen decomposition, we will obtain the eigenvectors as well as the eigenvalues (variances of each principal component). Of course, every eigenvalue, corresponds to a certain eigenvector. In the context of asset allocation, such as the present analysis, PCA can be used in order to decompose a returns matrix into statistical factors. The obtained factors are considered to be latent factors which represent an amount of unobservable risk somehow "hidden" within the assets. Thus, using these factors in the allocation process, actually constitutes a method of portfolio diversification. The PCA eigenvectors values, squared, can be furtherly used as weights as they represent asset weights corresponding to each "principal component" portfolio. The number of "principal components" portfolios is the actual number of principal components. The variance of each "principal component" portfolio represents its corresponding eigenvalue. The eigenvectors can take values between – 1 and 1, and therefore the PCA method is compatible with short selling within the markets (unlike Markowitz optimal portfolio selection, for example).



Source: own results, R software

Figure 3. Correlation Matrix

Table 4. PC1 coefficients and their squares (portfolio weights)

ASSET	Symbol	PC1 coefficients	Weights
V1	AAPL	0.197965	0.039190
V3	AXP	0.204274	0.041728
V4	BA	0.185187	0.034294
V5	CAT	0.194730	0.037920
V6	CRM	0.175705	0.030872
V7	CSCO	0.164608	0.027096
V8	CVX	0.197311	0.038931
V9	DIS	0.210475	0.044300
V10	DOW	0.210575	0.044342
V11	GS	0.214952	0.046204
V12	HD	0.202842	0.041145
V13	HON	0.205462	0.042215
V14	IBM	0.208516	0.043479
V15	INTC	0.129389	0.016741
V16	JNJ	0.167222	0.027963
V17	JPM	0.211144	0.044582
V18	KO	0.191668	0.036737
V19	MCD	0.209787	0.044011
V20	MMM	0.193551	0.037462
V21	MRK	0.174538	0.030464
V22	MSFT	0.187158	0.035028
V23	NKE	0.207019	0.042857
V24	PG	0.167148	0.027938
V25	TRV	0.188849	0.035664
V26	UNH	0.187632	0.035206
V27	V	0.210013	0.044106
V28	VZ	0.171832	0.029526
	Sum of weights		1

Source: own results, R software

The first principal component is regarded as the "market component", (Pasini, 2017), within which each stock displays the same contribution. This "market component" is actually that latent force driving assets to most often "move together". It is therefore considered to represent the largest source of systematic risk. From this perspective, the values of the first eigenvector should normally have the same sign. Figure 3 confirms this assumption as stocks display the same sign. At the same time, all coefficients of the first component are positive, as shown in Table 4, meaning that the portfolio to be constructed will only contain Long positions. As the rest of the eigenvectors represent other sources of risk (activity sector, business style, etc.), they will contain positive as well as negative values.

Using the values of the first principal component, squared, we obtain the actual weights that will be further used in order to build the portfolio. These weights are presented in Table 3. After obtaining the weights for creating a portfolio, we will furtherly construct the portfolio with the 27 selected stocks. In order to evaluate the performance of this portfolio, we will compare its rate of return with the rate of return of a "market portfolio", i.e. an equally weighted portfolio replicating the structure of a main market index, such as Dow Jones Industrial Average.

In order to conduct such a comparison, we will assume an investment of $100,000 in each of the built portfolios. As the previously analyzed period of time lasts until 30.09.2020, we will consider this specific date the date at which the investments are initiated. The portfolios will be afterwards evaluated after 2 weeks, on the 14.10.2020. As we are conducting a merely theoretical study, several simplifying assumptions will be considered. First, we will consider the market as being perfectly liquid, meaning that any placed order will be immediately executed, at the order price. So, we consider that the allocational efficiency of the market exists and there is no slippage. Second, we consider that trading a single asset is possible, even if in reality trading usually implies packs of stocks. Moreover, we even consider the possibility to buy non-integer numbers of assets in order to emphasize the specific contribution of the weights to the general return of the two portfolios. Third, we do not take into consideration any financial intermediation or management costs.

The results obtained, presented in Appendix 1 and Appendix 2 reveal a better result for the PCA weighted portfolio, in comparison with the equally weighted "market portfolio". The PCA weighted portfolio displays a rate of return of 3.06% for the analyzed period of time (two weeks), corresponding to an annualized rate of return of approximately 79.5%. At the same time, the equally weighted DJIA replicating portfolio provides a rate of return of 2.49%, or an annualized return of 64.85%. This could only mean that the KMO Test had a positive input and eliminated those assets that lowered the total return of the portfolio, as these assets are included in the second portfolio, the "market portfolio" and, apparently, they contribute to a lower total return. On the other hand, the PCA weights seem to attribute a higher importance to the proper assets and to contribute to a better result than in the case of the equally weighted portfolio. In this situation, we can conclude that our two-step method, a selection with the KMO Test and an allocation with the PCA method, manages to obtain a return better than the one given by the “market portfolio”. At the same time, the method doesn’t affect the diversification level, as the KMO test only eliminates 3 assets from a total of 30 assets (10%). Therefore, the number of assets of the two portfolios in not significantly different, so that the level of diversification may be considered similar, as well as the total risk of these portfolios.

4. Conclusion

Within a highly volatile market, affected by imitation, overreaction and herding behaviors, the correlation structure of the market is altered to an uncertain and dynamic extent. During such times, even professional investors tend to migrate towards “safe” conditions: mature economies, developed countries, historically well-performing assets or highly capitalized issuers. The allocation problem and the goal of an optimal portfolio selection become highly debated themes as does the maintenance of a certain target return.

During such turbulent periods, the Principal Component Analysis and the Kaiser – Meyer – Olkin Test prove to be useful tools for allocation, appropriate not only for Long, but also for Short transactions. The PCA’s main advantage is that it manages to reduce dimensionality, but, at the same time, to provide a well-diversified portfolio. By using the eigenvector of the first principal component, the PCA method conducts towards the construction of a market-oriented portfolio. Despite its many advantages, one should also consider the fact that at some point, the method tends to ignore additional sources of risk by practically ignoring the other principal components, so that a certain proportion of information is lost, as the market is the only risk source considered. On the other hand, this could also mean that the method only considers what seems to be important. Our results have already proven the fact that the proportion of variance displayed by the first principal component is much higher than the ones of the following components, so, indeed, the importance of the following components is small as they can only explain, together, a third of the proportion already explained by the first component taken individually.

Our results also confirm the positive values of the first principal component eigenvector, proving that all analyzed assets are indeed influenced by the market, they all display positive dependencies to the general evolution of the market. In our study, the rate of return of the portfolio created by using the PCA along with the KMO test manages to exceed the general market return, measured by an equally weighted “market portfolio”. As this happens during a crisis time, marked by a serious contraction of the economic activity, we can conclude that the method proves its own utility, as well as robustness. The proposed combination of methods not only “beats the market”, but also keeps the portfolio at a good diversification level, at the number of assets included in the portfolio is not significantly different from the number of assets in the "market portfolio". Moreover, the portfolio created by the proposed method is a portfolio that "follows" the market, as it is constructed based on the first principal component. So, the portfolio proposed by this method is not in any danger to miss reaction if the market would suddenly fall (as this is a main problem for the portfolios containing assets anti-correlated with the general evolution of the market). We conclude that, this way, the method manages to construct a portfolio with an acceptable level of total risk: a well-diversified portfolio, that follows the evolution of the market.

Although the PCA weighted portfolio provides a high rate of return, periodic reallocation is strongly recommended, as the sources of risk are not static variables, but dynamic ones, as is the correlation structure existing within the market, especially during periods of time characterized by high volatility and uncertainty. So, in this sense, the method should by applied on a weekly basis, allowing it to account for the most recent information available within the market and providing the investor with sufficient reaction time.

Acknowledgement

This work was cofinanced from the European Social Fund through Operational Programme Human Capital 2014-2020, project number POCU/380/6/13/125015 "Development of entrepreneurial skills for doctoral students and postdoctoral researchers in the field of economic sciences"

References

Avellaneda, M. (2020). Hierarchical PCA and Applications to Portfolio Management. Revista Mexicana de Economía y Finanzas Nueva Época, Vol. 15, No 1, pp. 1-16, DOI: https://doi.org/10.21919/remef.v15i1.446

Benazzoli, C. & Di Persio, L. (2016). Default contagion in financial networks, *International Journal of Mathematics and Computers in Simulation*, 10, pp. 112-117.

Boulesteix, A.-L. (2005). A note on between-group PCA, *International Journal of Pure and Applied Mathematics*, Volume 19 No. 3, 359-366, 2005.

Cerny, C.A., & Kaiser, H.F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research*, 12(1), pp. 43-47.

Crescimanna, V. & Di Persio, L. (2016). Herd Behavior and Financial Crashes: An Interacting Particle System Approach, *Journal of Mathematics*, art. No. 7510567, 2016.

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? *Psychological Bulletin*, 81, pp. 358-361.

Han, L., (2013), Empirical study on stock's capital returns distribution and future performance, TigerPrints, 5-2013, Clemson University

Jolliffe, I.T. (1986). Principal Component Analysis, Second Edition. Springer Texts in Statistics, Springer New York, 2002.

Kaiser, H.F. (1970). A second generation Little Jiffy. *Psychometrika*, 35, pp. 401-415.

Konarzewska, I. (2013). Optimal stock portfolio – Application of multivariate statistical analysis. Acta Universitatis Lodziensis, Folia Oeconomica 286, 2013, pp. 253-267.

Kozak, S., Nagel, S. & Santosh, S. (2017) Shrinking the Cross Section. Technical Report, *National Bureau of Economic Research: Cambridge,* MA, USA, 2017.

Morck, R., Yeung, B. & Yu, W. (2000), The information content of stock markets: why do emerging markets have synchronous stock price movements? *Journal of Financial Economics.* Vol. 58, Issues 1–2, 2000, pp. 215-260, https://doi.org/10.1016/S0304-405X(00)00071-4

Nobi, A., Maeng, S.E., Ha, G.G. & Lee, J.W. (2013) Random matrix theory and cross-correlations in global financial indices and local stock market indices. Journal of the Korean Physical Society. 2013, 62, pp. 569–574, 10.3938/jkps.62.569.

Pasini, G. (2017). Principal Component Analysis for stock portfolio management. *International Journal of Pure and Applied Mathematics,* Volume 115, No. 1, 2017, 153-167, ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version), http://www.ijpam.eu, doi: 10.12732/ijpam.v115i1.12

Tan, J. (2012). Principal Component Analysis and Portfolio Optimization. *SSRN Electronic Journal*. 10.2139/ssrn.2213687.

Benny, V. (2018). The impact of behavioral factors influences on individual investors in stock market investment. 5. pp. 118-126, St. Thomas College, Thrissur, Project Ayurveda Patents.

Yang, L. (2015). An Application of Principal Component Analysis to Stock Portfolio Management. Department of Economics and Finance, University of Canterbury. Retrieved from https://ir.canterbury.ac.nz/bitstream/handle/10092/10293/thesis.pdf

Yang, L., Rea, W. & Rea, A. (2017). Impending Doom: The Loss of Diversification before a Crisis. International Journal of Financial Studies, 2017, 5, 29, https://doi.org/10.3390/ijfs5040029.

Yaohao, P., Melo Albuquerque, P.H., Ferreira do Nascimento, I. & Freitas Machado, J.V. (2019). Between Nonlinearities, Complexity, and Noises: An Application on Portfolio Selection Using Kernel Principal Component Analysis, Entropy, vol. 21, pp. 376, 2019, DOI: 10.13140/RG.2.2.15768.24328

Yoshino, N., Taghizadeh-Hesary, N. & Otsuka, M. (2020). Covid-19 and Optimal Portfolio Selection for Investment in Sustainable Development Goals, *Finance Research Letters*, art. No. 101695, ISSN 1544-6123, https://doi.org/10.1016/j.frl.2020.101695.

https://finance.yahoo.com/

https://www.morganstanley.com/im/it-it/intermediary-investor/insights/articles/addressing-the-coronavirus-impact-emerging-markets-equity-portfolio.html

Appendix

Appendix 1. The Componence of Dow Jones Industrial and the correspondence with the R variable names

Symbol	Company Name	Variable
AAPL	Apple Inc.	V1
AMGN	Amgen Inc.	V2
AXP	American Express Company	V3
BA	The Boeing Company	V4
CAT	Caterpillar Inc.	V5
CRM	salesforce.com, Inc.	V6
CSCO	Cisco Systems, Inc.	V7
CVX	Chevron Corporation	V8
DIS	The Walt Disney Company	V9
DOW	Dow Inc.	V10
GS	The Goldman Sachs Group, Inc.	V11
HD	The Home Depot, Inc.	V12
HON	Honeywell International Inc.	V13
IBM	International Business Machines Corporation	V14
INTC	Intel Corporation	V15
JNJ	Johnson & Johnson	V16
JPM	JPMorgan Chase & Co.	V17
KO	The Coca-Cola Company	V18
MCD	McDonald's Corporation	V19
MMM	3M Company	V20
MRK	Merck & Co., Inc.	V21
MSFT	Microsoft Corporation	V22
NKE	NIKE, Inc.	V23
PG	The Procter & Gamble Company	V24
TRV	The Travelers Companies, Inc.	V25
UNH	UnitedHealth Group Incorporated	V26
V	Visa Inc.	V27
VZ	Verizon Communications Inc.	V28
WBA	Walgreens Boots Alliance, Inc.	V29
WMT	Walmart Inc.	V30

Source: Yahoo Finance, 30.09.2020

DOI: 10.2478/tjeb-2020-0009

Timisoara Journal of Economics and Business TJE&B

Appendix 2. Evaluation of the PCA weighted portfolio

Symbols	Total investment	Close Price 30.09.2020	PCA Weights	Nr of stocks BL	Portfolio value 30.09.2020	Close Price 14.10.2020	Portfolio value 14.10.2020
AAPL		115.81	0.0392	33.84	3919.01	121.19	4101.07
AXP		100.25	0.0417	41.62	4172.79	104.81	4362.60
BA		165.26	0.0343	20.75	3429.42	163.24	3387.50
CAT		149.15	0.0379	25.42	3791.97	163.61	4159.60
CRM		251.32	0.0309	12.28	3087.22	261.83	3216.33
CSCO		39.39	0.0271	68.79	2709.59	39.89	2743.98
CVX		72.00	0.0389	54.07	3893.14	72.95	3944.51
DIS		124.08	0.0443	35.70	4429.95	126.59	4519.56
DOW		47.05	0.0443	94.24	4434.18	48.83	4601.93
GS		200.97	0.0462	22.99	4620.44	211.23	4856.32
HD		277.71	0.0411	14.82	4114.48	287.09	4253.45
HON		164.61	0.0422	25.65	4221.46	173.47	4448.68
IBM		121.67	0.0435	35.74	4347.90	125.94	4500.49
INTC	100000	51.78	0.0167	32.33	1674.15	53.55	1731.38
JNJ		148.88	0.0280	18.78	2796.31	148.10	2781.66
JPM		96.27	0.0446	46.31	4458.20	100.22	4641.12
KO		49.37	0.0367	74.41	3673.68	50.12	3729.49
MCD		219.49	0.0440	20.05	4401.07	227.62	4564.09
MMM		160.18	0.0375	23.39	3746.20	168.40	3938.45
MRK		82.95	0.0305	36.73	3046.36	80.51	2956.75
MSFT		210.33	0.0350	16.65	3502.82	220.86	3678.19
NKE		125.54	0.0429	34.14	4285.67	127.66	4358.05
PG		138.99	0.0279	20.10	2793.83	144.04	2895.34
TRV		108.19	0.0357	32.96	3566.41	112.03	3692.99
UNH		311.77	0.0352	11.29	3520.57	321.85	3634.40
V		199.97	0.0441	22.06	4410.56	202.20	4459.74
VZ		59.49	0.0295	49.63	2952.62	58.43	2900.01
Total value					*100000.00*		*103057.67*
% towards 30.09.2019				*3.06%*			

Source: own results

Appendix 3. Evaluation of the equally weighted "market portfolio"

Symbols	Total investment	Close Price 30.09.2020	Equal Weights	Nr of stocks BL	Portfolio value 30.09.2020	Close Price 14.10.2020	Portfolio value 14.10.2020
AAPL	100000	115.81	0.033	28.78	3333.33	121.19	3488.18
AXP		100.25	0.033	33.25	3333.33	104.81	3484.95
AMGN		254.16	0.033	13.12	3333.33	237.65	3116.80
BA		165.26	0.033	20.17	3333.33	163.24	3292.59
CAT		149.15	0.033	22.35	3333.33	163.61	3656.50
CRM		251.32	0.033	13.26	3333.33	261.83	3472.73
CSCO		39.39	0.033	84.62	3333.33	39.89	3375.65
CVX		72.00	0.033	46.30	3333.33	72.95	3377.31
DIS		124.08	0.033	26.86	3333.33	126.59	3400.76
DOW		47.05	0.033	70.85	3333.33	48.83	3459.44
GS		200.97	0.033	16.59	3333.33	211.23	3503.51
HD		277.71	0.033	12.00	3333.33	287.09	3445.92
HON		164.61	0.033	20.25	3333.33	173.47	3512.75
IBM		121.67	0.033	27.40	3333.33	125.94	3450.32
INTC		51.78	0.033	64.37	3333.33	53.55	3447.28
JNJ		148.88	0.033	22.39	3333.33	148.10	3315.87
JPM		96.27	0.033	34.62	3333.33	100.22	3470.10
KO		49.37	0.033	67.52	3333.33	50.12	3383.97
MCD		219.49	0.033	15.19	3333.33	227.62	3456.80
MMM		160.18	0.033	20.81	3333.33	168.40	3504.39
MRK		82.95	0.033	40.18	3333.33	80.51	3235.28
MSFT		210.33	0.033	15.85	3333.33	220.86	3500.21
NKE		125.54	0.033	26.55	3333.33	127.66	3389.62
PG		138.99	0.033	23.98	3333.33	144.04	3454.44
TRV		108.19	0.033	30.81	3333.33	112.03	3451.64
UNH		311.77	0.033	10.69	3333.33	321.85	3441.11
V		199.97	0.033	16.67	3333.33	202.20	3370.51
VZ		59.49	0.033	56.03	3333.33	58.43	3273.94
WBA		35.92	0.033	92.80	3333.33	35.91	3332.41
WMT		139.91	0.033	23.82	3333.33	143.94	3429.35
Total value					*100000.00*		*102494.34*
% towards 30.09.2019				*2.49%*			

Source: own results