

Niftiyev, Ibrahim

Working Paper

Understanding Principal Component Analysis (PCA) in the Azerbaijan Economy: Case Studies of Vegetable and Fruit Sectors

Suggested Citation: Niftiyev, Ibrahim (2021) : Understanding Principal Component Analysis (PCA) in the Azerbaijan Economy: Case Studies of Vegetable and Fruit Sectors, SSRN, Rochester, NY, <https://doi.org/10.2139/ssrn.3782426>

This Version is available at:

<https://hdl.handle.net/10419/230316>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Understanding Principal Component Analysis (PCA) in the Azerbaijan Economy: Case Studies of Vegetable and Fruit Sectors

Working Paper No. 4

Ibrahim Niftiyev^{1, 2}

¹ *PhD Candidate, Univesity of Szeged, Doctoral School in Economics, 6722 Kálvária sgt. 1., Szeged, Hungary.*

² *Centre for Studies on European Economy (AIM) - Azerbaijan State University of Economics (UNEC), AZ 1001 Istiglaliyyet str. 6, Baku, Azerbaijan.
ibrahimniftiyev@gmail.com*

This working paper has not gone through the final formal review and should be cited as a working paper. The main purpose of this working paper is to get ready the materials, concepts, and discussions for the final academic journal publication which is ongoing.

There is no funding or official support from any organization or person. All errors and expressions are those of the author.

Ibrahim Niftiyev

ORCID number: 0000-0003-3437-9824

Research Gate/ Google Scholar profile: Ibrahim Niftiyev



Szeged – 2021 (February)

Table of Contents

Abstract	3
1. Introduction	3
2. Data and Methods	5
<i>3.1. PCA Analysis of the Vegetable Sector</i>	<i>8</i>
<i>3.2. PCA Analysis of the Fruit Sector</i>	<i>11</i>
<i>3.3. Comparison of the sub-sectoral performance</i>	<i>14</i>
4. Discussion	15
5. Take-home Messages	16
6. Concluding Remarks	17
References	18

Abstract

Is it possible to apply Principal Component Analysis (PCA) in the Azerbaijan economy? Before blindly answering the question, careful examination of the collected data and the relevance of the data set to the mentioned analysis should be checked. PCA is a widely used multivariate dimension reduction tool that statistical analysis employs. Nevertheless, its application area is limited. Also, PCA analysis in macroeconomic studies is not much preferred and in Azerbaijan's case, there is not one. To understand this methods' relevance two case studies, namely vegetable and fruit sectors have been chosen. The aim was to quantify the sub-sectoral performance via vegetable and fruit sectors. The collected data set and PCA analysis on it were evaluated from multiple angles to outline the roadmap for future investigations. The author argues that in line with the method's theoretical expectations, Azerbaijan's vegetable and fruit sectors partially fulfill the planned goals of the analysis. This means PCA is a highly useful analytical tool and might produce valuable sub-sectoral performance evaluations in the case of Azerbaijan. According to the produced indices, the vegetable sector performed better than the fruit sector from 1999 until 2015. However, between 2015–2020 fruit sector outperformed the vegetable sector. A similar analysis can be organized among the other sub-sectors or between sectors in the case of the Azerbaijan economy. Furthermore, this working paper embodies experimentation results, not conclusions about the methods' validity and justification. Hence, this work compares three rotations of PCA analysis (Varimax, Quartimax, and Equamax) and constructed indices separately for vegetable and fruit sectors.

Key words: Azerbaijan economy, Principal Component Analysis (PCA) agriculture, agrarian sectors, vegetable sector, fruit sector, economic diversification, sub-sectoral performance, economic performance evaluation

JEL Classification: E01, C38, O13, Q11, Q18

1. Introduction

PCA analysis is a useful technique when the data set is large, and multiple variables should be analyzed (Bro & Smilde, 2014). Early work on PCA analysis of Jolliffe (1990) emphasized that if the correlation among the variables is high, then it can be reduced to find “a true dimension” of the data set that will provide the same information with the minimum amount of information loss. That reduction produces so-called “components” which help to understand more complex patterns among the multiple data series (Ringnér, 2008). Ringnér (2008) also mentions the independence of the components rather than being uncorrelated between them. A significant job can be done if the original variable quantity a can be reduced to b via newly constructed index variables, or components. Usually, PCA is a pre-analysis of the variables of interest and plays the role of an analytical bridge to conduct additional analysis.

This working paper essentially tests already existing methods that maintain the core principles of PCA analysis via the capabilities of SPSS software (version 23). Also, it should be noted that the current paper does not aim to test any hypotheses or answer a solid research question. It will be the main task of the upcoming papers which will address the concrete research objectives. My main goal with this working paper is to leave you with an idea of how to implement a PCA analysis when you can fulfill the main expectations of the method.

There are several benefits of PCA analysis in the case of the Azerbaijan economy. First, the PCA method is relatively more understandable for the internal audience compared to the more advanced econometrical methodologies. Second, especially in the case of agriculture, Azerbaijan's official statistics committee provides several closely related indicators that can be generalized or aggregated. Third, after the pioneered works, PCA can become a widely used and approached multivariate statistical tool to understand the Azerbaijan economy. All that being said, the main philosophy, implementation stages, and steps might give rise in the body of literature concerning the discussed PCA analysis.

Newly adopted state programs such as "Azerbaijan 2030: National Priorities on Socio-Economic Development" (prezident.az, 2021) require more detailed economic policy implementations. Without sectoral assessments, it is hardly possible to conduct them. Surely, before the analysis, proper multidimensional and trustworthy statistical data production is a must. Also, sub-sectoral performance evaluations might be useful to address possible Dutch disease effects or, in general, resource curse explanation of the lagging non-oil development. Several papers investigated the resource curse and Dutch disease phenomenon in the case of Azerbaijan to understand the lagging performance of non-oil sectors (Hasanov, 2013; Zulfigarov and Neuenkirch, 2019; Şanlısoy and Ekinci, 2019). A particular focus always should be paid to the non-oil tradable sectors, and even going further, non-oil sub-sectoral dynamics and performance should be analyzed. This was suggested by the previous works of the author (Niftiyev, 2020a; 2020b). The ultimate scale to slow down the Dutch disease or curse effects given the resource-driven development of the national economy is possible only via diversified regional economic growth. Otherwise, a barrage of economic risks poses a tragic fate in the long-term as the positive spillovers of the manufacturing (which means non-oil tradeable sectors in Azerbaijan's case) cannot be replaced with capital-intensive extractive industries (see Matsuyama, 1992).

Thus, this working paper's structure is as follows. The next section explains collected data, an explanation of the variables, and the methods that have been incorporated. Moreover, the next section visualizes the collected data via Gaussian distribution and box plots. Section three reports the results, concerning the experimentation of the PCA methods (especially rotation techniques). Section 4 provides brief take-home messages which were formed throughout the experimentation. The last section gives concluding remarks that encompass the chief findings of the current investigation.

2. Data and Methods

Firstly, the variables have been visualized in terms of Gaussian distribution to understand the normality, skewness, and kurtosis. Secondly, box-plot figures demonstrate another angle of the distribution, also allowing for the outlier detection, as the PCA analysis is sensitive to outliers (Jolliffe & Cadima, 2016). There are various methodologies and rich strategies to identify the connection among the variables, to detect outliers, and to show a bigger picture for multivariate data (Bro & Smilde, 2014).

Probability models also can create principal components that incorporate latent variable density models - being closely related to statistical factor analysis (Navas et al., 2008). Nevertheless, this paper's experimentations do not include any probability-based PCA analysis, though it provides an even wider horizon to consider for the Azerbaijani economy. Table 1 reports the variables of interest, their short name (which was used in the software), and the levels of the measurement.

Table 1. Analyzed variables, levels of measurement and their sources.

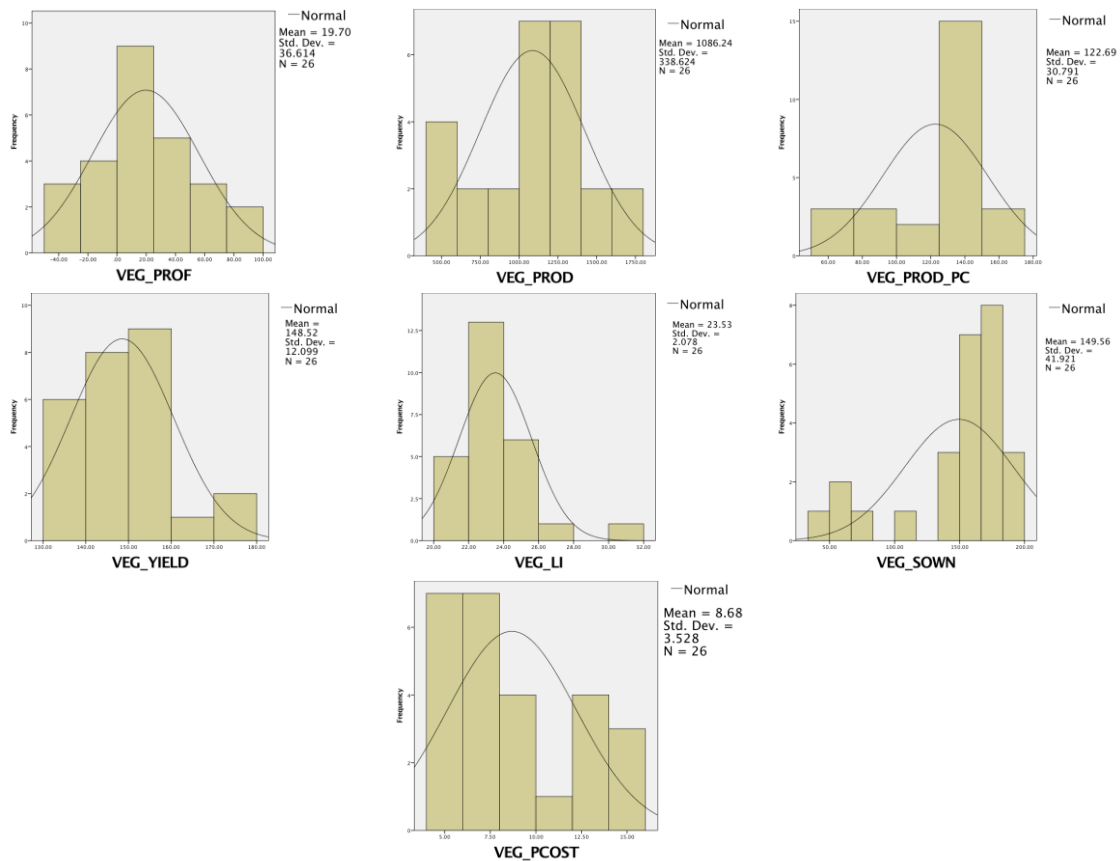
N	Short name	Original name	Measurement	N	Source
1.	VEG_PROF	Profitability of the vegetable products	in %	26	State Statistical Committee of the Republic of Azerbaijan
2.	VEG_PROD	Production of the vegetable products	thsd. tons	26	
3.	VEG_PROD_PC	Per capita production of the vegetable products	kg	26	
4.	VEG_YIELD	Yield of the vegetable products	100 kg/ha	26	
5.	VEG_SOWN	Sown area of the vegetable products	in ha	26	
6.	VEG_LI	Labor input of the vegetable products	person/hour per 100 kg of vegetables in agricultural enterprises	26	
7.	VEG_PCOST	Production cost of the vegetable products	in AZN, per 100 kg of vegetables in agricultural enterprises	26	
8.	FR_PROF	Profitability of the fruit products	in %	26	
9.	FR_PROD	Production of the fruit products	thsd. tons	26	
10.	FR_PROD_PC	Per capita production of the fruit products	kg	26	
11.	FR_YIELD	Yield of the fruit products	100 kg/ha	26	
12.	FR_CULT	Cultivated area of the fruit products	in ha	26	
13.	FR_LI	Labor input of the fruit products	person/hour per 100 kg of vegetables in agricultural enterprises	22	
14.	FR_PCOST	Production cost of the fruit products	in AZN, per 100 kg of vegetables in agricultural enterprises	26	

Gaussian distribution of the vegetable variables in Figure 1 shows a relatively normal distribution of vegetable profitability, vegetable production, and per capita production.

Vegetable yield, labor input in the vegetable sector, and production cost data are skewed to right, while sown are is skewed leftwards. Figure 2 illustrates the same data in box plot style. This was useful in the outlier detection.

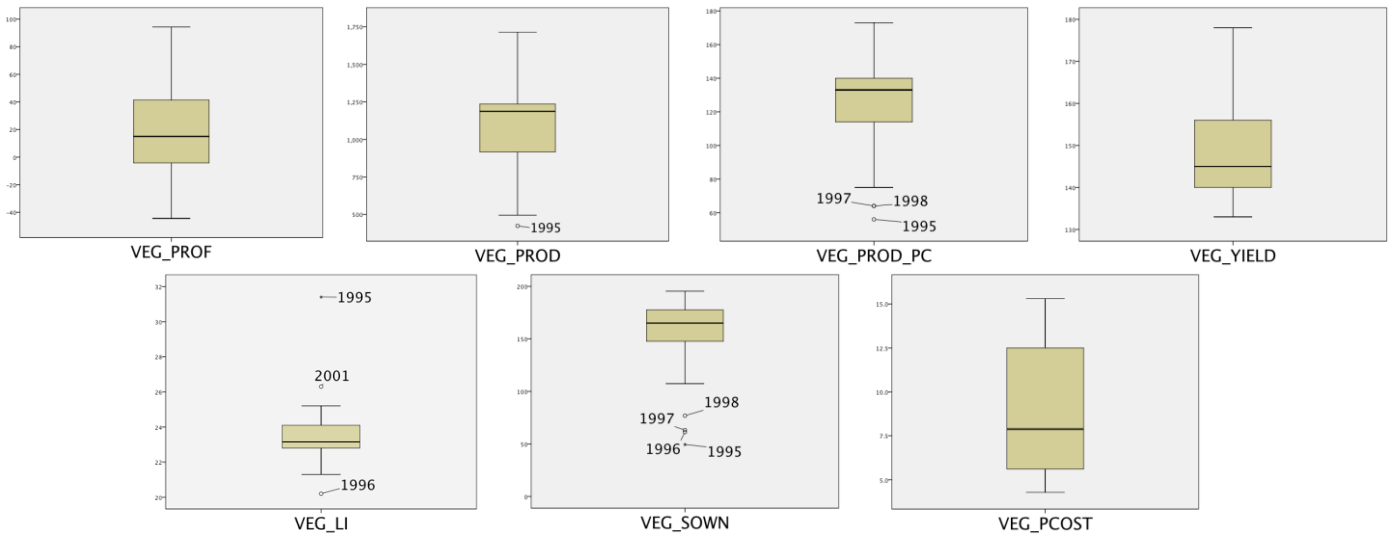
Excluding vegetable profitability, yield, and production cost, other variables had outliers mainly covering the late 1990s and early 2000s. The same goes with Figure 3 and Figure 4 which indicate the variables related to the fruit sector. While fruit yield has a skewed distribution to the right, cultivated area and labor input shows skewed data to the left. The nature of the outliers in fruit sectors is almost the same, just the year of 2017 is substantially below compared to the other data points. As a recommendation for further work, normality tests (i.e. Shapiro-Wilk test) should be more compact to outline the data distribution. Lastly, as an outlier replacement strategy, winsorizing seems expedient because of its practical and simple nature in the case of few outliers.

Figure 1. Gaussian distribution of the vegetable variables, 1995–2020.



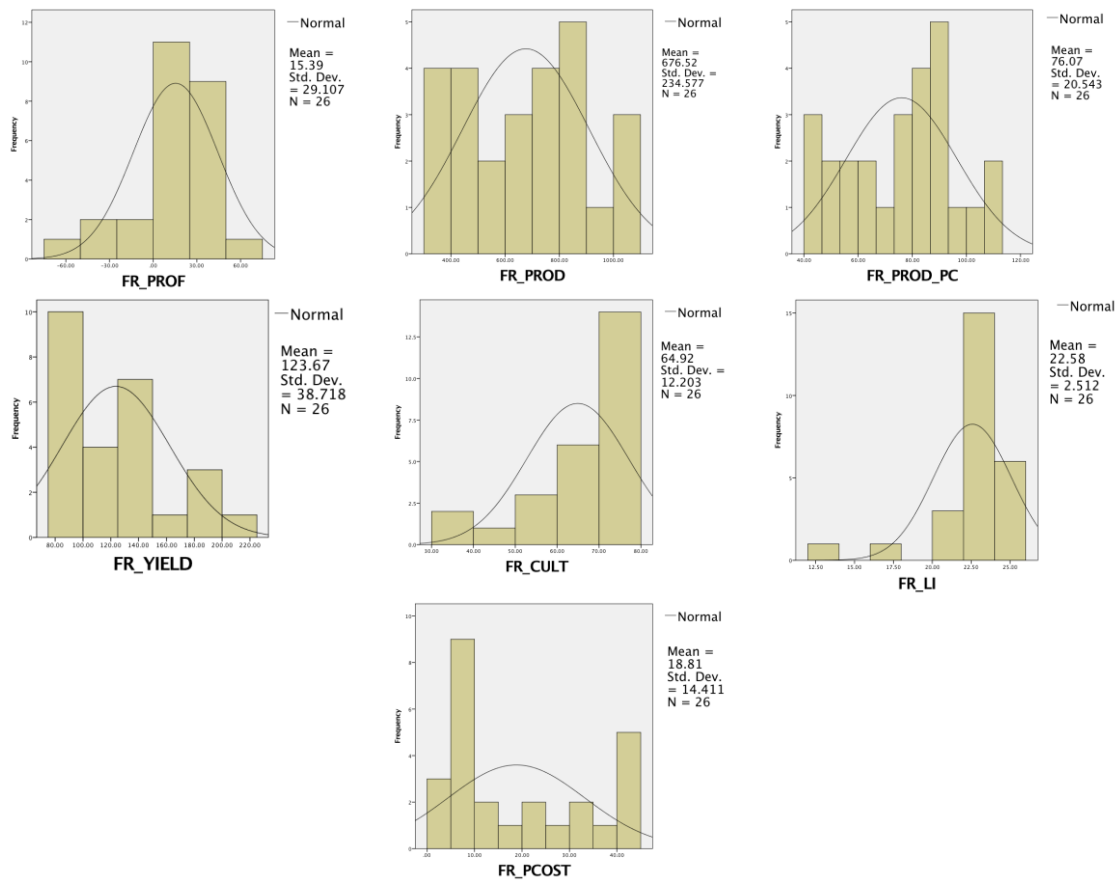
Source: SSCRA (2021)

Figure 2. Box plot visualization of the vegetable variables, 1995–2020.

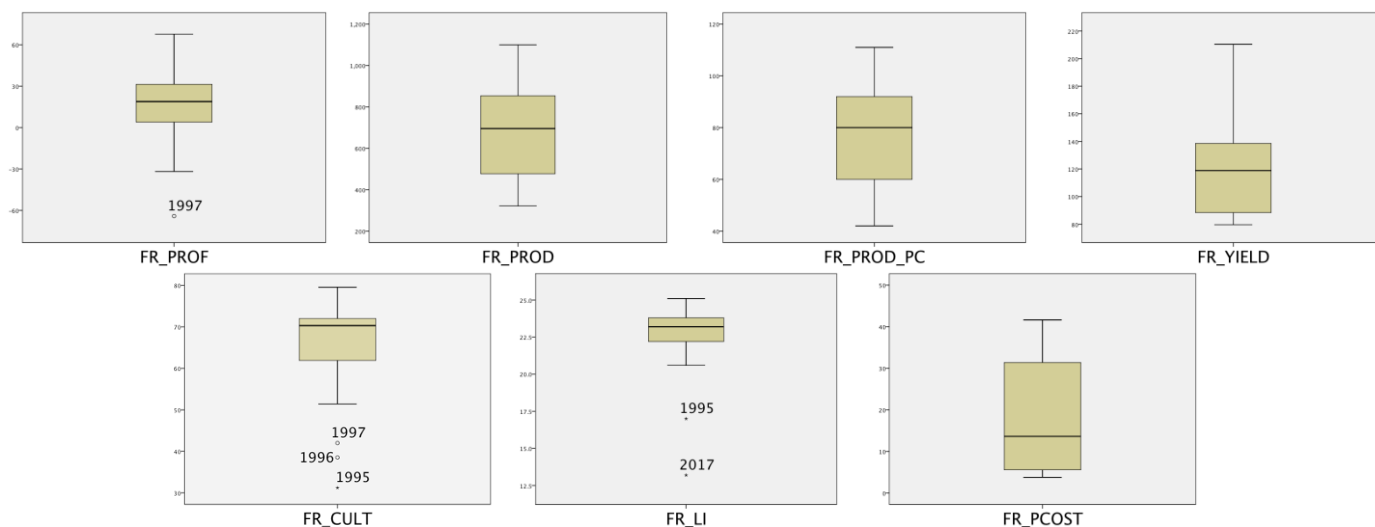


Source: SSCRA (2021)

Figure 3. Gaussian distribution of the fruit variables, 1995–2020.



Source: SSCRA (2021)

Figure 4. Box plot visualization of the fruit variables, 1995–2020.

Source: SSCRA (2021)

3. Results

3.1. PCA Analysis of the Vegetable Sector

Before the PCA analysis is being conducted, Kaiser-Meyer-Olkin Measure of Sampling Adequacy or simply KMO and Bartlett's Test of Sphericity must be done to understand if the variable count and sampling size allows applying PCA or not. The recommended value for the KMO test is greater than 0.500 according to some sources (Field 2005a; 2005b) but also the highly used rule of thumb is >0.400 . In our case (see Table 2), the KMO value is 0.637 which is greater than both 0.400 and 0.500 indicating a moderate level of adequacy. Then, as expected, Bartlett's Test of Sphericity shows high significance indicating the fact that at least one correlation among the variables is significant. Thus the PCA analysis is relevant.

Table 2. KMO and Bartlett's Test of vegetable variables.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.637
Bartlett's Test of Sphericity	Approx. Chi-Square	205.068
	df	21
	Sig.	0.000

Current experimentations opted to extract a fixed amount of components to reduce 7 variables. Table 3 shows the communality values for the vegetable variables. A communality explains the degree of the variance for the particular extraction that is being accounted for by the other variables. As a rule of thumb, a value greater than 0.200 is acceptable for the PCA analysis. However, some sources and papers raise this value up to 0.500 (Dogbegah et al. 2011). In light of the current work and also any upcoming works this aspect of PCA analysis must be kept in the mind. From Table 3, we can see that

vegetable production, per capita production, vegetable yield, and sown area in the vegetable sector scored the highest values, while vegetable profitability, labor input, and production cost could not achieve higher values. Nevertheless, they still preserve the adequacy in terms of the widely accepted rule of thumbs.

As the number of the extracted components was fixed from the very beginning, the scree plot's role is diminished in this study. From the collected data the relation was obvious enough to assume this. Nevertheless, Figure 5 illustrates Eigenvalues for the vegetable sector and 2 components seem to fit the theory (as a cut-off value Eigenvalue is 1 and a value of less than one is not acceptable in terms of component number).

Figure 5. Scree plot of the vegetable variables.

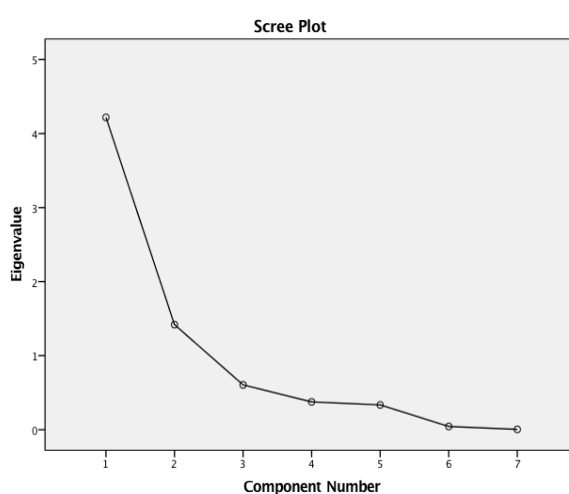


Table 3. Communalities information of the vegetable variables.

Communalities		
	Initial	Extraction
VEG_PROF	1	0.714
VEG_PROD	1	0.926
VEG_PROD_PC	1	0.907
VEG_YIELD	1	0.908
VEG_SOWN	1	0.950
VEG_LI	1	0.675
VEG_PCOST	1	0.558

The next step of the PCA analysis is understanding the total variance explained by the principal components. This step presents a piece of crucial information about the overall performance and reliability of the PCA analysis and should be handled very carefully. From Table 4 we can see that percentage of total variance explained in component 1 was 60.262%, while component 2 explained 20.263% of the variance that the first component failed to explain. The cumulative percentage of total variance explained based on the extraction sums of squared loadings is 80.524 % which can be considered as a good indicator of the performance of the analysis.

Table 4. Total Variance Explained in vegetable variables.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Var.	Cum. %	Total	% of Var.	Cum. %	Total	% of Var.	Cum. %
1	4.218	60.262	60.262	4.218	60.262	60.262	4.062	58.035	58.035
2	1.418	20.263	80.524	1.418	20.263	80.524	1.574	22.489	80.524

Then, the rotation is the most essential step in the PCA analysis. Early works (Norusis 1988) and recent publications emphasized its role in the PCA. In line with the possibility that SPSS software allows applying the rotations like Varimax, Quartimax, Equamax, Promax, and Direct Oblimin, this working paper utilized the first three methods as orthogonal rotation methods to conduct and compare the results. Promax and Direct Oblimin methods which are oblique rotations can be also applied, but just for simplicity and practicality, this working paper focused only on the first three methods. Component matrix without rotation is Table 5, while Table 6 reports the results of all three mentioned rotations.

Table 5. Component Matrix of the vegetable variables.

	Comp.	
	1	2
VEG_PROF	0.834	0.134
VEG_PROD	0.960	-0.064
VEG_PROD_PC	0.950	0.065
VEG_YIELD	0.135	-0.943
VEG_SOWN	0.865	0.449
VEG_LI	-0.613	0.547
VEG_PCOST	0.746	-0.049

Interestingly, Varimax and Equamax produced the same results while Quartimax showed slightly different results. Component 1 can be named as “productivity and profitability”, while component 2 reflects the cost side of the vegetable sector. That is why high loadings to component 1 can be found from VEG_PROF, VEG_PROD, VEG_PROD_PC, VEG_YIELD, and VEG_SOWN.

The component transformation matrix which is depicted in Table 7 does not provide any meaningful interpretations. It is presented just for the sake of information about the rotation.

Table 6. Rotated component matrices of the vegetable variables.

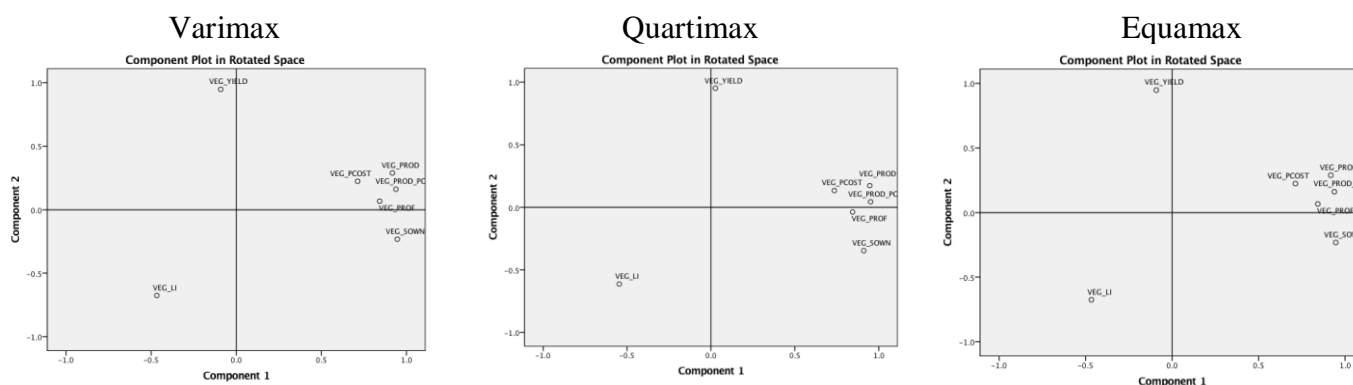
Rotated Component - Varimax			Rotated Component - Quartimax			Rotated Component - Equamax		
Comp.			Comp.			Comp.		
1 2			1 2			1 2		
VEG_PROF	0.842	0.067	VEG_PROF	0.844	-0.038	VEG_PROF	0.842	0.067

VEG_PROD	0.918	0.289	VEG_PROD	0.946	0.174	VEG_PROD	0.918	0.289
VEG_PROD_PC	0.938	0.161	VEG_PROD_PC	0.951	0.044	VEG_PROD_PC	0.938	0.161
VEG_YIELD	-0.091	0.948	VEG_YIELD	0.026	0.952	VEG_YIELD	-0.091	0.948
VEG_SOWN	0.946	-0.232	VEG_SOWN	0.911	-0.347	VEG_SOWN	0.946	-0.232
VEG_LI	-0.467	-0.676	VEG_LI	-0.547	-0.613	VEG_LI	-0.467	-0.676
VEG_PCOST	0.713	0.224	VEG_PCOST	0.735	0.134	VEG_PCOST	0.713	0.224

Table 7. Component transformation matrices of vegetable variables.

Component	1	2	Component	1	2	Component	1	2
1	0.972	0.236	1	0.993	0.114	1	0.972	0.236
2	0.236	-0.972	2	0.114	-0.993	2	0.236	-0.972

Figure 6 visually represents the loadings. The general pattern and structure exhibit a similarity among all rotated component matrices. VEG_LI loaded negatively to both components. Other variables clustered around the high and positive values of Component 1 and low and positive values of Component 2. Similarly, VEG_YIELD gets small but positive values from Component 1 while very high values are attributed from Component 2.

Figure 6. Loading plots of the vegetable variables.

3.2. PCA Analysis of the Fruit Sector

The obtained values of the KMO sampling adequacy test is 0.746 which is a good result. Bartlett's Test of Sphericity also fulfills the expectation by being highly significant (see Table 8).

Table 8. KMO and Bartlett's Test of fruit variables.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.746
Bartlett's Test of Sphericity	Approx. Chi-Square	337.129
	df	21
	Sig.	0.000

Interestingly, the communality values of the fruit sector are higher than the vegetable sector (See Table 9). For instance, from Table 3 it can be seen that VEG+LI and VEG_PCOST have low values; however, commonalities in fruit variables are significantly high not dropping down from 0.769.

Table 9. Communalities information of the fruit variables.

	Initial	Extraction
FR_PROF	1	0.782
FR_PROD	1	0.992
FR_PROD_PC	1	0.991
FR_YIELD	1	0.937
FR_CULT	1	0.883
FR_LI	1	0.769
FR_PCOST	1	0.844

Compared to the total variance explained in the vegetable sector (80.524, see Table 4), two components accounted for 88.553 % cumulative explanation of the seven original variables.

Table 10. Total Variance Explained in vegetable variables.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Var.	Cum. %	Total	% of Var.	Cum.%	Total	% of Var.	Cum. %
1	4.691	67.015	67.015	4.691	67.015	67.015	4.41	63.005	63.005
2	1.508	21.538	88.553	1.508	21.538	88.553	1.788	25.548	88.553

Similar to the case of the vegetable variables, the role of the scree plot is low in fruits variables too. Nevertheless, Figure 8 informs about the recommended choice of the components to reduce the original variable count. As this study fixed two components to apply a dimension reduction process, the eigenvalue-based choice does not conflict with this action.

The component matrix is given in Table 11 below.

Figure 7. Scree Plot of the fruit variables.

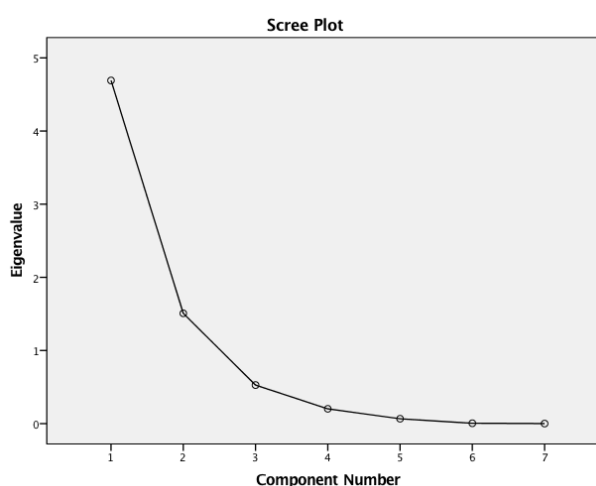


Table 11. Component matrix of fruit variables.

	Component	
	1	2
FR_PROF	0.736	0.491
FR_PROD	0.992	-0.092
FR_PROD_PC	0.996	-0.009
FR_YIELD	0.860	-0.444
FR_CULT	0.809	0.479
FR_LI	-0.047	0.875
FR_PCOST	0.883	-0.255

Same with the case of vegetable sector, fruit sector's Varimax and Equamax rotations are identical, while Quartimax showed different results (see Table 12).

Table 12. Rotated component matrices of the fruit variables.

Rotated Component Matrix - Varimax			Rotated Component Matrix - Quartimax			Rotated Component Matrix Equamax		
Comp.			Comp.			Comp.		
1	2		1	2		1	2	
FR_PROF	0.557	0.687	FR_PROF	0.701	0.539	FR_PROF	0.557	0.687
FR_PROD	0.975	0.206	FR_PROD	0.996	-0.026	FR_PROD	0.975	0.206
FR_PROD_PC	0.953	0.287	FR_PROD_PC	0.994	0.057	FR_PROD_PC	0.953	0.287
FR_YIELD	0.953	-0.169	FR_YIELD	0.888	-0.386	FR_YIELD	0.953	-0.169
FR_CULT	0.63	0.697	FR_CULT	0.775	0.531	FR_CULT	0.63	0.697
FR_LI	-0.305	0.822	FR_LI	-0.105	0.87	FR_LI	-0.305	0.822
FR_PCOST	0.918	0.019	FR_PCOST	0.897	-0.196	FR_PCOST	0.918	0.019

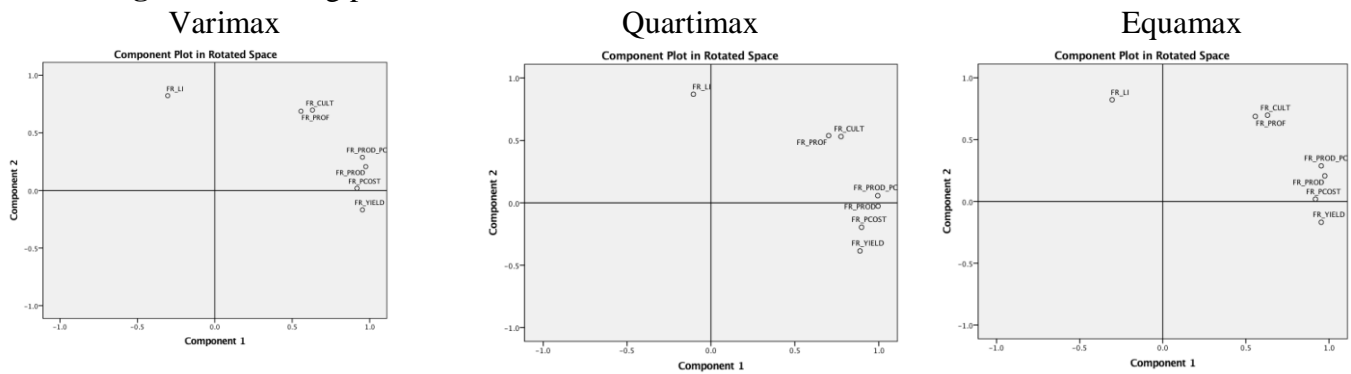
Component transformation matrix which is depicted in Table 7 does not provide any meaningful interpretations. It is presented just for the sake of information about the rotation.

Table 13. Component transformation matrices of fruit variables.

Component	1	2	Component	1	2	Component	1	2
1	0.955	0.297	1	0.998	0.066	1	0.955	0.297
2	-0.297	0.955	2	-0.066	0.998	2	-0.297	0.955

When visually represented as in Figure 8, it can be seen that the loadings are concentrated in the high and positive part of Component 1, also some variables get relatively high values from both of them. Correspondingly, FR_LI also deviated from the main cluster of the loadings.

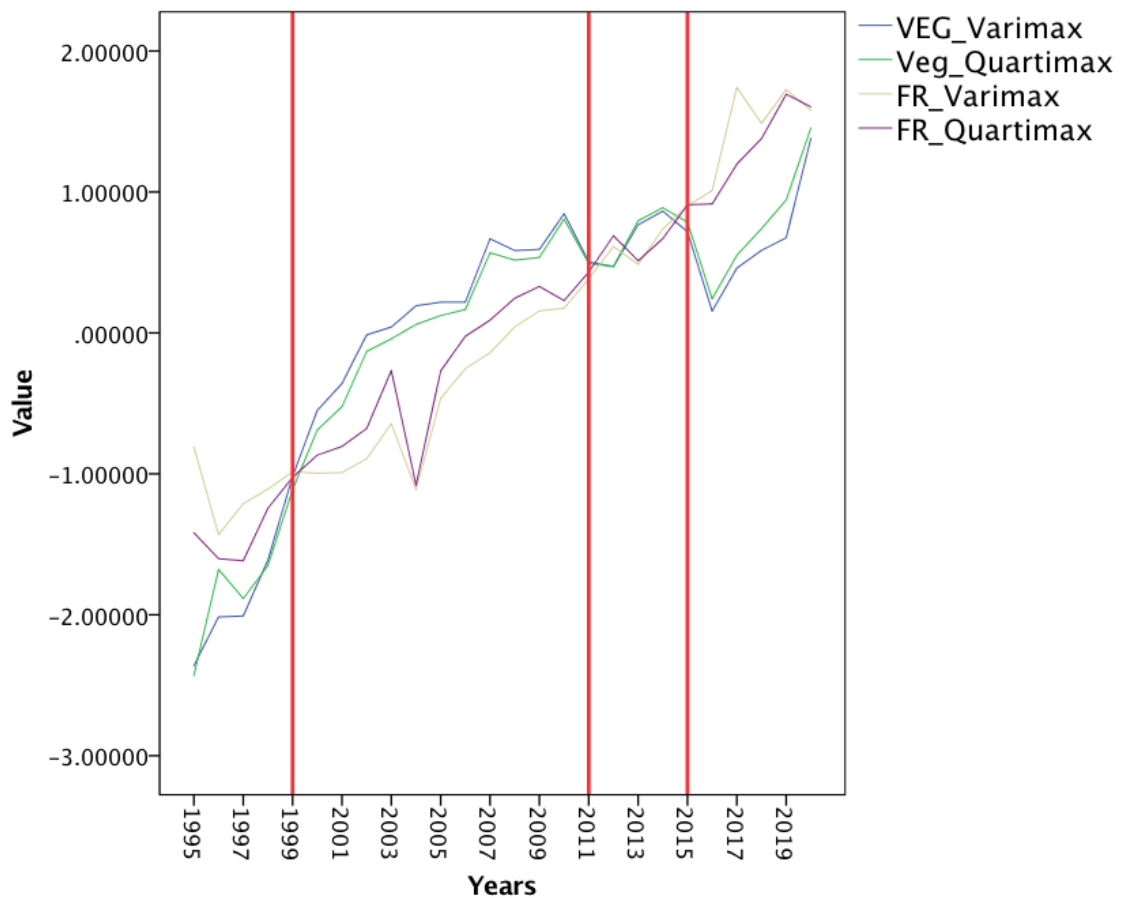
Figure 8. Loading plots for fruit variables.



3.3. Comparison of the sub-sectoral performance

This sub-section presents the comparison of the produced indices based on the seven variables for vegetable and fruit sectors (Fig. 9). According to the results, vegetable sector caught up and rapidly increased its performance since 1995. Starting from 1999 vegetable sector exceeded fruit sector until 2011. Between 2011 and 2015 vegetable sector’s performance was volatile but fruit sector exhibited a stable upward trend. Since 2015, fruit sectors outperformed vegetable sector; however, vegetable showed recovery while 2020 challenged fruit sector.

Figure 9. Produced indices to measure sub-sectoral performance of vegetable and fruit sectors, 1995–2020.



4. Discussion

Since the collapse of the Soviet Union, the agriculture sector has gone through a series of peaks and troughs in Azerbaijan. The painful transition process and oil-led development created multiple hardships in front of the long-term sustainable development of agriculture and agrarian sectors. That is why several studies argued about the relevance of the natural resource curse doctrine and Dutch disease hypothesis in the Azerbaijan economy. It can be argued that the sub-sectoral performance of Azerbaijan's national economy mirrors the main macroeconomic and sector-specific realities. A proper performance evaluation might shed light on the micro realities to consider by the policy- and decision-makers. Thus, a specialized body of literature is highly beneficial to extend the understanding of sub-sectoral performance via available statistical analysis in the case of Azerbaijan.

It can be hardly said that Azerbaijan has a solid competitive advantage or productivity-driven agriculture sector. Nevertheless, the government has spent an enormous amount of resources to support the farmers and employees in agriculture, as they employ the biggest share of the labor force since the late 1990s. At first, vegetable and fruit sectors show similar developmental trends but the current working paper identified periodic differences in the performance on the sub-sectoral level via the PCA. The performance response of the fruit sector to the 2015–2015 commodity price downturns and national currency devaluation is noteworthy. Unlike the fruit sector, the vegetable sector's sectoral performance sharply declined in 2016. Despite it started to catch up with the fruit sector, this event presents an intriguing aspect after the absolute predominance of the vegetable sector over the fruit sector.

In a broader context, both vegetable and fruit sectors had their own ups and downs since 1999. However, the vegetable sector's remarkable performance gets challenged starting from 2011. External sustainability of the sectoral dynamics also should be considered in future analysis in parallel with the sub-sectoral performance. Vegetables and fruits are being mainly exported to the Russian and European markets. Trade signals also should be included in the analysis in order to understand the flip side of the phenomenon. Because if there are trade shocks, domestic producers would answer it as decreased sown or cultivated areas, diminished labor inputs and etc. in the case of the vegetable and fruit sectors.

Two devaluations of the national currency have occurred in 2014. Figure 11 highlights the aftereffects of the devaluation of the national currency which is different in the case of the vegetable and fruit sectors. This analysis brings to the surface the difference between vegetable and fruit sectors but the underlying drivers of these different reactions should be deeply investigated to address proper institutional and economic policy responses to keep the stable performance.

5. Take-home Messages

The working paper's insights suggest the following take-home messages when PCA is used in the case of the Azerbaijan economy. Please, note that these suggestions are based on the current working paper's experimentation results. The steps of the analysis may vary from case study to case study, as well as the data set. SPSS provides a set of extensions that can be applied in order to have wider implications.

- Despite sub-sectoral performance was evaluated based on the vegetable and fruit sectors, any other analysis that could employ PCA must take into account the impact of outliers as the PCA is sensitive to the outliers. Azerbaijan economy experienced a speedy transition period, responded to the negative oil price shocks, and showed supply volatility in agricultural exports. The probability of the potential outliers is significantly high. Proper outlier diagnostics or structural break tests are a must; otherwise, biased results will be inevitable;
- It is highly recommended that the pattern matrix also should be included in the analysis as it can show the significance of the loadings. As a rule of thumb, the values higher than 0.300 (0.200 if the variable is scale) shows statistically significant loadings which are handy to understand if the emerging loadings are happening because of certain rules or just out of coincidence;
- Some studies may look forward to utilizing Explanatory Factor Analysis (EFA) both in the case of vegetable and fruits sectors and in the case of related studies in the Azerbaijan economy to determine if any latent variables could influence the statistical analysis;
- The analysis of the component correlation matrix also is expedient to make the more grounded choice of the rotation method;
- According to the discussions of this study on the Research Gate platform (Research Gate, 2021), in this study the sample size of 27 and the variable count of 7 is expedient. To explain, further PCA should be sensitive to the sample size and the variable count. In other words, the variable-to-component ratio (v:c) is an important feature in the evaluation of PCA results. The application of pattern matrix can be informative to understand if the sample size and v:c is on the desired level or not;
- Multicollinearity problem poses a serious bias in the PCA, so the best scenario is to have a low level of correlations (see Stevens, 2002);¹
- Before the conduction of the PCA, one must identify the reasons behind the analysis. In the current working paper's case, the main reason was the conduction of the sub-sectoral performance valuation based on the identical seven original variables in the matter of vegetable and fruit sectors. Also, the produced indices are also handy to utilize in the regression, cluster, or ranking analysis. However, not always the rationale behind the choice to apply PCA is evident.

¹ However, this aspect is quite debatable.

- Due to the fixed component approach of this working paper, the role of the scree plot was low. In the future, together with the scree plot, and the number of iterations (Maximum Iterations for Convergence) should be carefully regulated.
- The variable must behave simply, not complexly. It means that if the variable "X" gets high loadings from both components (in the case of two-component extraction), then the variable shows a complex structure that is hard to interpret within the PCA. To solve this problem, one might recode the variable, remove it or process it in a way that boils down to the simple structure in the PCA.
- Qualitative research methods also might accompany PCA. PCA is not a cause-effect-type of investigation and usually serves as a pre-investigation of the dataset. The vegetable and fruit sector in Azerbaijan illustrates highly volatile export indicators that are linked to the domestic sub-sectoral performance. That being said, focus group discussions, in-depth interviews, or case studies can be more conclusive compared to quantitative methods, which usually fail to establish a broader perspective.

6. Concluding Remarks

This working paper demonstrated the initial experimentation results about the possibility of the PCA in the Azerbaijan economy employing vegetable and fruit sectors as case studies. For the time being, measurable enhancement of the body of the literature that explores and evaluates the sub-sectoral performance is seldom concerning the Azerbaijan economy. The current working paper suggests that PCA can be a good statistical tool to fill the gap. However, it should be noted that this paper experimented with the methodology rather than tested any solid hypothesis. The steps that were described can be re-configured and adapted or even extended in line with the requirements of future research. Therefore, as SPSS gives a wide range of opportunities to employ PCA, new parts and steps can be included to have a more comprehensive and holistic methodology. Indeed, this methodology could contribute to the everchanging and rapidly expanding studies about the Azerbaijan economy.

The experimentation found that Varimax and Equamax produced the same results in the PCA, while Quartimax authored slightly different outcomes. In the PCA, the Direct Oblimin rotation method is more suggested (a consideration for future research). Moreover, according to the produced indices, four critical periods are evident: up to 1999, 1999–2011, and 2011–2015 and after 2015. The vegetable sector performed well in 1995–1999 and 2015–2020 period, while the fruit sector showed better results in 1999–2015 which is a higher time range compared to the vegetable sector. Because vegetable exports take higher share in non-oil exports in Azerbaijan, fruit sectors' share should be analyzed in relation to the sectoral performance to find economic problems that decrease efficiency, productivity, profitability.

It is desirable to have more scientific papers in the field of economics about the Azerbaijan economy that employs PCA. PCA can be a reliable source of aggregation among

macroeconomics studies. However, a literature review fails to show any serious works that already did so. On the positive side, as this working paper and experimentation on it demonstrates, there is a high probability in near future to read such papers across the body of literature regarded in the Azerbaijan economy. At least, the current work can be used as a source to conceptualize the main markers of the PCA in Azerbaijan's context. Hence, PCA showed a strong inclination to explain the Azerbaijan economy if the assumptions and the main steps are properly designed. PCA can produce quite handy indices that easily suit the other statistical and econometrical methodologies, such as correlation and regression reducing the risk of the multicollinearity or serial correlation in the residuals (as the researchers tend to use multivariate regression models, they usually prefer to use the variables which are close to each other, creating an enabling environment to experience typical issues in regression analysis). All in all, as a dimension reduction way, PCA fulfilled the main expectations of the current research and showed high promises for further works.

References

- Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, 6(9), 2812–2831.
- Dogbegah, R., Owusu-Manu, D., & Omoteso, K. (2011). A principal component analysis of project management competencies for the Ghanaian construction industry. *Australasian Journal of Construction Economics and Building*, 11(1), 26.
- Field, A. (2005a). *Discovering Statistics Using SPSS for Windows*, London, Sage Publications..
- Field, A. (2005b). Factor Analysis Using SPSS: Theory and Application, Available at: <http://www.sussex.ac.uk/users/andyf/factor.pdf>
- Hasanov F. (2013) Dutch disease and the Azerbaijan economy. *Communist and Post-Communist Studies*, 46, 4, pp. 463–480.
- Jolliffe, I. (1990). Principal component analysis: A beginner's guide - I. Introduction and application. *Weather*, 45, 375–382. <https://doi.org/10.1002/j.1477-8696.1990.tb05558.x>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Matsuyama K. (1992) Agricultural productivity, comparative advantage, and economic growth. *Journal of economic theory*, 58, 2, pp. 317–334.
- Navas, N., Romero-Pastor, J., Manzano, E., & Cardell, C. (2008). Benefits of applying combined diffuse reflectance FTIR spectroscopy and principal component analysis for the study of blue tempera historical painting. *Analytica Chimica Acta*, 630(2), 141–149.
- Niftiyev, I. (2020a). Descriptive Analysis of Employment in Azerbaijan: Possibilities of the Dutch Disease. *Journal of Economic Sciences: Theory & Practice*, 77(1), 100–112.
- Niftiyev, I. (2020b). The De-industrialization Process In Azerbaijan: Dutch Disease Syndrome Revisited. *Proceedings of the 4th Central European PhD Workshop on Technological Change and Development*, 357–396. Available at: <https://www.econstor.eu/handle/10419/227485> Accessed: 01.02.2021.

- Norusis, M. (1988) *The SPSS guide to data analysis for SPSS-X*, Chicago, SPSS Inc
- Prezident.az (2021). Azərbaycan 2030: sosial-iqtisadi inkişafa dair Milli Prioritetlər, Available at: <https://president.az/articles/50474>, Accessed: 02.02.2021.
- Research Gate (2021). Principal Component Analysis. An index production? Available at: https://www.researchgate.net/post/Principal_Component_Analysis_An_index_production?ec=topicPostOverviewAuthoredQuestions&sg=k7woZmhBOZbfp0Qpr_X9jfLEaRu63xdQ_nQXIX8xGDA4oigNXhmMrG-iFQh3o7Ee15tBMu6nfYXpyndB Accessed: 01.02.2021
- Ringnér, M. (2008). What is principal component analysis? *Nature Biotechnology*, 26(3), 303–304. <https://doi.org/10.1038/nbt0308-303>
- Şanlısoy S., Ekinci R. (2019) Azərbaycan Ekonomisinin Hollanda Hastalığı Açısından Değerlendirilmesi. *Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 26(2), 595–608.
- SSCRA – The State Statistical Committee of the Republic of Azerbaijan (2021). Agriculture, forestry and fishing. Available at: <https://www.stat.gov.az/source/agriculture/?lang=en> Accessed: 15.01.2021.
- Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Zulfigarov F., Neuenkirch M. (2019) *Azerbaijan and Its Oil Resources: Curse or Blessing?* Universität Trier, Research Papers in Economics No.11/19, Available at: <https://www.econstor.eu/bitstream/10419/207043/1/1667010700.pdf>, Accessed: 10.05.2020.