ЕКОНОМІКО-МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ

KMYTIUK Tetiana,
PhD (Economics), Associate Professor at the
Department Mathematical Modelling and Statistics
Kyiv National Economic University
 named after Vadym Hetman
54/1, Prospect Peremohy, Kyiv, 03057, Ukraine

ORCID: 0000-0001-5262-856X
kmytiuk.tetiana@kneu.edu.ua

КМИТЮК Тетяна,
к. е. н., доцент, доцент кафедри математичного
моделювання та статистики
Київського національного економічного
університету імені Вадима Гетьмана
проспект Перемоги, 54/1, м. Київ, 03057, Україна

ORCID: 0000-0001-5262-856X
kmytiuk.tetiana@kneu.edu.ua

FACTORS OF PRICING OF AGRICULTURAL PRODUCTS

ЧИННИКИ ЦІНОУТВОРЕННЯ СІЛЬСЬКОГОСПОДАРСЬКОЇ ПРОДУКЦІЇ

Introduction. *Price is the main factor that has a direct impact on the main results of the enterprises, especially agricultural. The pricing policy is able to maintain the competitiveness of the enterprise, to ensure its solvency and stability.*

Problem. *Pricing is carried out in a complex interaction of a set of internal and external Factors. The corresponding complexity of the phenomenon leads to the need to include many Factors in the model, which can negatively affect the result. The numerous number of Factors necessitate the use of methods that allow the identification of the most statistically significant Factors and assess their relationship with the resulting feature.*

The aim *of the article is to identify the main Factors and their influence on the formation and change of the price of agricultural products using Factor analysis.*

Methods. *Different methods such as mathematical and statistical (factor analysis, principal component), economic and statistical (multiple regression method based on factor scores) and generalization were applied in the article.*

Вступ. *Ціна є головним чинником, який безпосередньо впливає на основні результати діяльності підприємств, зокрема й аграрних. Цінова політика здатна зберегти конкурентоспроможність підприємства, забезпечити його платоспроможність і стабільність.*

Проблема. *Ціноутворення здійснюється в складній взаємодії внутрішніх і зовнішніх чинників. Відповідна складність явища потребує включення у модель багатьох факторів, які можуть негативно вплинути на результат. Велика кількість факторів вимагає використання методів, які дадуть змогу ідентифікувати найбільш статистично значущі чинники та оцінити їх взаємозв'язок із результуючою ознакою.*

Мета *статті – виявлення за допомогою факторного аналізу основних чинників, визначення їх впливу на формування та зміну ціни сільськогосподарської продукції.*

Методи: *математико-статистичні (факторний аналіз, головних компонент); економіко-статистичний (множинний регресійний на основі факторних оцінок); узагальнення.*

Results. *The application of Factor analysis and method of principal components determine the most significant Factors influencing the formation of agricultural products prices. The author used different dimensionality reduction methods and obtained three main components: supply, demand, and the solvency of the population. This helped to reveal the relationship between variables and price without multicollinearity problem.*

Conclusions. *The author used Factor analysis and the principal component method; it helped to take into account the peculiarities of the relationships between the indicators that characterize potato pricing, to eliminate multicollinearity between independent indicators, to get fewer common Factors than the original number of variables, to obtain a quantitative assessment of the latent variable. The use of orthogonal rotation helped to distribute the load on all three Factors more evenly, which simplified their further interpretation and to save 95 % of the total primary information after rotation and to reduce the dimensionality of the feature space, which is a high indicator for practical tasks.*

Keywords: Factor analysis, principal component method, multiple linear regression model, multicollinearity, latent variable.

Результати дослідження*. За факторним аналізом та методом головних компонент визначають найбільш вагомі фактори впливу на формування цін сільськогосподарської продукції. Використовуючи різні методи зменшення розмірності, отримано три основні компоненти: пропозиція, попит, платоспроможність населення, що дало змогу виявити зв'язок між змінними та ціною без проблеми мультиколінеарності.*

Висновки*. Використання факторного аналізу та методу головної компоненти надало можливість врахувати особливості зв'язків між показниками, що характеризують ціноутворення на картоплю, усунути мультиколінеарність між незалежними показниками, отримати менше спільних факторів, ніж вихідна кількість змінних, заробити кількісну оцінку латентної змінної. Використання ортогонального обертання сприяло більш рівномірному розподілу навантаження на всі три фактори, що спростило їх подальшу інтерпретацію та зберегло 95 % загальної первинної інформації після обертання та зменшення розмірності простих функцій, що є високим показником для практичних завдань,*

Ключові слова: факторний аналіз, метод головних компонент, множинна лінійна регресійна модель, мультиколінеарність, латентна змінна.

Introduction. The position of agriculture within the economy seems very important in developing and developed countries. The agricultural sector is very important for the structural transformation of the economy, the development of domestic production and competition increases employment in rural areas.

One of the biggest problems in the market of agricultural products is pricing, as well as solving a number of issues to improve product quality. Price is the main Factor that has a direct impact on the main results of the enterprise – profit and sales. In addition, the pricing policy maintains the competitiveness of the enterprise, ensures its solvency and stability.

Problem. The pricing system of agricultural products is a complex and dynamic multifactorial process based on a certain methodology, i.e. general rules of price formation related to micro and macroeconomic features of the economic system, as well as with different areas of pricing.

Prices for agricultural products are the most dynamic character of the agricultural market. They are sensitive to the influence of numerous economic, technological, social and political factors. The dynamics of price movements are associated with changes in quality, the volume of production

and supply in the market, the solvency of consumers and the level of consumption of this product, the volume of inventories, the general economic condition of enterprises and more.

Weather conditions also have a significant impact on the pricing of agricultural products. Potato prices depend on favorable weather conditions. If the year is productive, the prices for these products will be lower – this is due to the large supply; and vice versa, if the year is lean, the prices will rise (decrease in supply).

The peculiarity of pricing and prices for agricultural products is that most of its species show seasonal price fluctuations. In particular, potato prices are significantly reduced during the mass harvest and in the initial period after harvest. In the future, prices begin to rise and often reach their maximum with the approach of the next harvest season.

An equally important factor influencing prices is the level of solvency of the population. Public policy measures should be aimed not only at regulating prices but also at increasing incomes. The pricing process is regulated at the level of the economy through a system of economic levers by changing not the prices themselves, but the factors that affect them (the tax system, inflation, etc.).

Thus, pricing is carried out in a complex interaction of a set of factors of internal and external order. The relevant pricing factors will help to investigate the possible range of variations in product prices.

The complexity of socio-economic phenomena leads to the need to include in the model many factors. In practice, a linear multiple regression model is used to identify relationships between variables. Nevertheless, including a large number of factors in a multiple regression model can be harmful.

Numerous factor values necessitate the use of methods that allow the identification of the most statistically significant factors and assess their relationship with the resulting feature.

This method can be factor analysis, which takes into account fewer accepted factors than the number of observed variables.

Since the factor is the cause of the driving force of a process or phenomenon that determines its nature, the detection, evaluation and prediction of their impact on changes in agricultural performance by factor analysis is the main task of this study.

Analysis of recent research and publications. There are a lot of theoretical and methodological scientific works of many scientists are devoted to the principles of agricultural development. Praburaj L. discussed the role of the agricultural sector in the economy and further development of it [1] and noticed agricultural development is essential for the overall economic transformation of a country. The papers [2; 3] focus on the analysis of the agriculture data and finding optimal parameters to maximize the crop production using different data mining techniques. Several authors analyzed influencing Factors of the use of agricultural land [4]. R. Manishimwe et al. [5] outlined the prime Factors influencing the price structure for farm products.

Problems of a quantitative assessment of interdependencies between economic indicators have been studied by the scientists. Johnston R., Jones K., Manley D. studied the effectiveness of correlation-regression analysis in the process of solving problems related to the establishment of relationships between the characteristics of different species [6]. Daoud J. investigated the construction of the multiple linear regression equation [7]. Greene W. H. Presented the mathematical apparatus used in the process of constructing linear regression models [8]. Hansen B. analyzed variables in the regression model [9].

An important stage in building a regression model is the stage of selecting influential Factor features among numerous measured indicators. The adequacy of the constructed regression model depends on the quality of such a choice. Factor analysis was created in the early twentieth century for the needs of psychology. Subsequently, Factor analysis has become widespread in economics, medicine, sociology, and other sciences, which deal with a huge number of variables, of which it is usually necessary to identify the leading [10]. By means of the Factor analysis communication of one variable with another is established, the degree of this communication is defined, and the basic Factors underlying the specified changes are found out. Moreover, Factor analysis is used to identify the hidden structure of a set of variables [11]. Factor analysis is especially productive in the initial stages of research when it is necessary to identify any previous patterns in the study area. This allows you to make the next experiment perfect than experimenting on variables chosen randomly or randomly.

Principal component method (PCA) is an appropriate multivariate method that helps to reduce the dimensionality of a data set consisting of a large number of interrelated variables while maintaining as much variation as possible. The practical foundations of the methodology for determining the number of Factors are presented in [12–14], the rotation methods of Factor analysis such as orthogonal and oblique are presented in [15–18], and the use of Factor estimates in multiple regression analysis is in [19].

However, despite a significant number of theoretical and practical works, the issue of the practical application of these methods in a particular economic sphere, in particular in the agricultural sector, remains insufficiently resolved. An important problem is the choice of the most influential Factors scores that will be included in the model. Thus, the question of the choice and combination of approaches to solving this issue remains relevant.

In this study, we will demonstrate the application of a combination Factor analysis approach and principal components method to construct a multiple linear regression model that shows the relationship between independent economic variables and agricultural prices (such as potatoes), without existing multicollinearity, but without losing of the significant information.

The aim of the article is to identify the main Factors and their influence on the formation and change of the price of agricultural products using Factor analysis.

Methods. Mathematical, statistical economic, multiple regression and generalization methods were applied in the paper to study the influence of Factors on the pricing of agricultural products and obtain practical results. Factor analysis and the principal component methods were used to reduce numerous variables to a smaller number and create a set of Factors that are considered as uncorrelated variables for assessing the impact on price formation.

Results. The price of agricultural products is formed under the influence of many Factors, which can be expressed using economic and statistical methods, in particular the multiple linear regression model. This model is used to study the dependence of one variable on other variables. Factors included in multiple regression model are divided into dependent y and independent variables x_i. In this case, the variable x_i is called regressor (explained), and y is called resultant or explanatory variable.

In multivariate regression, we assume that the dependent variable y can be affected by more than one Factor. In this case, multiple linear regression analysis will be used to investigate the relations between price and Factors of traits of pricing. The multiple linear regression model is a function, which describes the relationship between the dependent variable y and regressors $x_1, x_2, \dots x_m$. It can be submitted as follows:

$$y = f(x_1, x_2, \dots, x_m) + \varepsilon = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \varepsilon, \qquad (1)$$

where β_m – is unknown model parameters to be evaluated;

ε – is the model's error term (also known as the residuals).

The influence of each Factor (each independent variable) on the dependent variable is characterized by partial regression coefficients (parameters).

The partial regression coefficient shows how many units the value of the dependent variable will change as the value of the corresponding Factor increases per unit, if the values of all other Factors remain constant. The commonly used summary statistics in multiple regression are the standard error of the recession; the coefficient of determination (R^2) and adjusted R^2; F – $test$, by means of which the significance of all parameters is checked simultaneously, and each separately – by means of t – $test$ [20]. In practice, MLRM include a large number of independent (explained) variables (x_i) where some of them might be slightly correlated with the dependent variables (y) or may be redundant because of high correlations (multicollinearity problem) with other independent variables and largely duplicate each other.

Significant multicollinearity between independent variables can lead to some inaccuracies in the statistical interpretation of regression coefficients, reliable modeling results, and analysis of functional relationships. Multivariate statistical techniques such as Factor analysis help to solve the problem of multicollinearity.

Factor analysis is based on the assumption that the phenomenon under study, determined by a certain system of features that change consistently,

can be described using a smaller number of other latent variables, called Factors that explain the reasons for these changes. The number of Factors is much less than the number of the original variables [21].

There are two basic approaches to Factor analysis: common Factor analysis and principal component analysis (PCA). The goal of data reduction is typically achieved by the use of principal components analysis as opposed to common Factor analysis [22].

In common Factor analysis, the variance of measured variables is partitioned to common variance and unique variance. The Factors are estimated to explain the covariance among the observed variables x_i ($i = 1,2, \dots , n$), and the Factors are viewed as the causes of the observed variables as a linear combination of m Factors, $F_1, F_2, \dots, F_m$, and a uniqueness component, u_i. Instead, the principal components method reveals hidden (latent) main reason, which explain the correlation between the signs and are explained in terms of content. It is assumed that the features of x_i ($i = 1,2, \dots, n$) are indicators of such latent properties that are not directly measured. The main reason of the correlation of the features of the $j - th$ group is called the PC_j ($j = 1,2, \dots, m$) component, $m < n$. Indications that fall into different groups are uncorrelated, and the PC_j components are independent too. It will be combined two general approaches: principal components analysis and common Factor analysis (or principal Factor analysis). Principal component analysis was used primarily for data reduction, whereas common Factor analysis was used to understand the relations among a set of measured variables in terms of underlying latent variables.

The basic idea underlying Factor analysis is that n observed random variables, x, can be expressed, except for an error term, as linear functions of m (< n) hypothetical (random) variables or common Factors, that is if $x_1, x_2, \dots, x_n$ are the variables and $f_1, f_2, \dots, f_m$ are the Factors, then [23]:

$$
\begin{aligned}
x_1 &= \alpha_{11}f_1 + \alpha_{12}f_2 + \cdots + \alpha_{1m}f_m + u_1, \\
x_2 &= \alpha_{21}f_1 + \alpha_{22}f_2 + \cdots + \alpha_{2m}f_m + u_2, \\
&\quad \dots \\
x_n &= \alpha_{n1}f_1 + \alpha_{n2}f_2 + \cdots + \alpha_{nm}f_m + u_n,
\end{aligned}
\tag{2}
$$

where $\alpha_{ij,}\ i = 1,2, \dots, n; j = 1,2, \dots, m$ are the constants called the Factor loadings and $u_{j,}\ j = 1,2, \dots, n$ are error terms, sometimes called specific Factors.

It is extremely important to build an economically sound Factor model when applying the appropriate analysis and determining the impact of Factors on the pricing of agricultural products, as the incorrectly created model can lead to incorrect results. Research of the object of modeling and compilation of its mathematical description consists in establishing connections between the characteristics of the process, identifying its boundary and initial conditions and formalizing the process in the form of a system of mathematical relations.

Analysis of above literature, gave us to follow a detailed description of each step of applying Factor analysis.

Step 1. Standardization (normalization) of data is a necessary initial stage of data transformation using Factor analysis, i.e. reducing the dimensionality of the feature space, as variables can differ significantly in values. Standardized elements (z_{ij}) of which are determined by the formula:

$$z_{ij} = \frac{x_{ij} - \underline{x_i}}{\sigma_i}, \tag{3}$$

where $\underline{x_i}$ — is the average value of x_n;

$\overline{\sigma_i}$ — is the standard deviation of x_n.

In contrast to the input values, the normalized values do not depend on the units of measurement, they fluctuate around the same (zero) average, which greatly simplifies the analysis.

Construction of the correlation matrix of the system of variables by calculating the Pearson linear correlation coefficients shows correlation coefficients between variablesю

Step 2. Extracting Factors and calculating Factor loadings a_{in}, which are the main subject of interpretation. At this step, the methods of component analysis, main Factors and maximum likelihood are used.

In practice, PCA is most often used to isolate Factors. Its main idea is to select in a multidimensional number $X = (x_1, x_2, \ldots, x_n)$ groups of closely correlated variables and replace them without loss of information content by the main components $F = (f_1, f_2, \ldots, f_m)$.

The mathematical model of the PCA can be written in the form:

$$f_j = \beta_{11} z_1 + \beta_{12} z_2 + \cdots + \beta_{ij} z_i, \tag{4}$$

where f_j — is the main component ($j = 1, 2, \ldots, m$); β_{ij} — is the coefficient reflecting the contribution of the variable z_i to the main component f_j; z_i — is standardized source variable.

An algorithm for selecting the main components are presented to follow:

1. *Standardization of the initial variables*, leading to the fact that the variances of all standardized variables become the same (all standardized variables have the same information value) and the origin is transferred to the center of the data cloud.

2. *Linear transformation* of $Z = (z_1, z_2, \ldots, z_i)$ with the aim of constructing a new orthogonal number of principal components $F = (f_1, f_2, \ldots, f_m)$.

To carry out this transformation, it is necessary to calculate the coefficients of matrix A=$\{\beta_{ij}\}$. They are determined and based on the following requirements:

• Principal components must be linear combinations of variables $z_1, z_2, \ldots, z_i$;

• Principal components must be orthogonal;

• The first principal component should have the maximum sample variance (s^2), the second principal component should have the maximum sample variance with the first fixed, etc.: $s^2(f_1) \geq s^2(f_2) \geq \cdots \geq s^2(f_m)$;

• The total variance of the initial variables must be equal to the total variance of the principal components.

The calculation of the principal components is reduced to the calculation of eigenvectors and eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ of the correlation matrix of the initial data. In this case, the eigenvalues will be equal to the variances of the new variables $s^2(f_1) = \lambda_1 \geq s^2(f_2) = \lambda_2 \geq \cdots \geq s^2(f_m) = \lambda_m$ and the eigenvectors will coincide with the columns of the matrix A$=\{\beta_{ij}\}$: $\left(\beta_{1i} \ldots \beta_{mj}\right)^T$ i-th eigenvector corresponding to the eigenvalue λ_i.

The β_{ij} values are called Factor loadings. They represent the correlation coefficients between the original variables and the principal components. Factors include those variables for which $|\beta_{ij}| > 0{,}7$.

3. *Reducing the dimension of the space* $(f_1, f_2, \ldots, f_m)$ by cutting off uninformative variables. The decision on the number of remaining Factors is quite arbitrary, but there are some rules that allow you to rationally choose the number of Factors. As a rule, the number of Factors identified should be significantly less than the number of initial variables. To solve this problem, the following methods are used: the Kaiser criterion, scree plot test, the total percent variance explained.

Step 3. After determining, the number of components, to facilitate the interpretation of the analysis, rotate the components that have been determined. Factor rotation minimize the complexity of the Factor loadings to make the structure simpler to interpret. Factor loading matrices are not unique, for any solution involving two or more Factors there are an infinite number of orientations of the Factors that explain the original data equally well. Rotation of the Factor loading matrices attempts to give a solution with the best simple structure.

Step 4. Factor analysis and the principal component method are purely quantitative methods and allow you to select Factors.

Interpretation of Factor analysis involves the researcher examining which variables are attributable to a Factor and giving that Factor a name. For example, a Factor may have included four variables that all relate to demand. Traditionally, at least two or three variables must load on a Factor, so it can be given a meaningful interpretation. The naming of Factors is a subjective, theoretical, and individual process.

The main object of Factor analysis is to combine items with high loading in one Factor that can explain the majority of the process together.

The Factor score is also called the component score. This score is of all rows and columns, which can be used as an index of all variables and can be used for further regression analysis. With this Factor score, we will assume that all variables will behave as Factor scores and will move.

Factor scores can be considered as independent variables for predicting the price of potatoes using the following multiple regression model:

$$P = b_0 + b_1 FS_1 + b_2 FS_2 + \ldots + b_m FS_m + e, \qquad (5)$$

where P — is depend variable (average price);

b_0 — is regression constant;

$b_1\ b_2 b_m$ — are regression coefficients;

$FS_1, FS_2, \ldots, FS_m$ — are Factor scores;

e — is random error term.

The Factor analysis that is realized by means of a method of the main components is carried out. The procedures of the Factor analysis in this paper were presented with the help of a software system of statistical analysis and data processing STATISTICA.

Data set was consisted of data of explained variable (y) and the predictors variables (x_i), $i = 1 \ldots n$. The data were taken over a range of 31 years from January 1991 to December 2021 [25].

Explained variables are average retail prices of potatoes in Latvia. According to the above discussion, the predictors' variables were formed that the most influential on the price of potatoes:

• Average monthly and median wages and salaries in Latvia, (*euro*) (x_1);

• Sown area, (*thsd ha*) – area in which potatoes are grown, (x_2);

• Harvested crop production, (*thsd t*) – it is production harvested from basic, repeated and inter-row sown areas, (x_3);

• The average yield of potatoes, (*quintals per ha)* is calculated on the basis of the harvested crop potatoes in basic sown areas per 1 ha of spring productive area of the corresponding crop, (x_4);

• The output of potatoes per capita (*kg*), (x_5);

• Gross Domestic Product (*billion US dollars*), (x_6);

• Inflation rate in Latvia, (*%*), (x_7);

• Household consumption expenditure on food average per household member per month, (*euro*), (x_8).

Analysis of the system of statistical indicators will allow to identify qualitative and quantitative patterns of the studied independent variables. Some main statistical indicators are used; among them are the means, maximum and minimum values of variables, standard deviations, coefficients of variation, and others. *Table 1* shows the descriptive statistics for these variables of the model.

The large difference between the minimum and maximum values of variables is observed in *Table 1*: x_1 – average monthly and median wages and salaries, x_3 — harvested crop production, x_5 – the output of potatoes per capita, x_7 — inflation rate. These variables can have negative consequences when they are being included in the regression model. In this regard, these large differences must be fixed. To do this, these variables will be included

in the model in a transformed form, which will reduce the asymmetry of their distribution, as well as bring the distribution of the regression residuals closer to normal.

Table 1

Descriptive statistics of independent variables

Variables	Mean	Minimum	Maximum	Sample variance	Standard Deviation	Variation	Standard Error
x_1	340.355	36	841	60947.84	246.876	72.535	44.34
x_2	47.771	18.1	96.9	510.78	22.6	47.31	4.059
x_3	693.954	377.454	1271.7	55637.17	235.875	33.99	42.364
x_4	155.806	112	223.5	1079.33	32.853	21.086	5.9
x_5	304.097	199	496	5608.89	74.893	24.628	13.451
x_6	33.435	12.4	61.6	267.35	16.351	48.902	2.937
x_7	22.616	−1.2	155	2002.39	44.748	197.859	8.037
x_8	118.807	91	150	234.09	15.3	12.878	2.748

Source: own study based on the research results.

The coefficient of variation shows the percentage of standard deviation to the arithmetic mean of the sign. It is known that the greater the coefficient of variation, the smaller the homogeneous population and the less typical average for this population. It was found that the population is quantitatively homogeneous if the coefficient of variation does not exceed 33 % (Gill et al., 2003). We can see exceeding the limit on such variables: x_1 – average monthly and median wages and salaries, x_2 − sown area, x_3 − harvested crop production, x_6 – *Gross Domestic Product*, x_7 − inflation rate. This means that the size of the standard deviation is 73 %, 47 %, 34 %, 49 %, 197 % in accordance, of the size of the mean. This implies big differences among values of independent variables.

Standard deviation helps us assess how far the values are spread above and below the mean. A high standard deviation of x_3, x_4, x_5, x_8 shows that the data are widely spread and are less reliable.

The linear regression model assume that the predictor variable is independent and also have a correlation, have effect or impact, into the dependent variable. Pairwise correlations among the average price of potatoes data and predicted variables are showed in *Table 2*.

Table 2

Correlation matrix for the average price of potatoes data (y) and predicted variables

	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
y	1								
x_1	0.903	1							
x_2	−0.880	−0.908	1						
x_3	−0.833	−0.841	0.962	1					
x_4	0.787	0.917	−0.825	−0.678	1				
x_5	−0.778	−0.783	0.902	0.970	−0.6077	1			
x_6	0.938	0.968	−0.907	−0.863	0.867	−0.814	1		
x_7	−0.559	−0.518	0.736	0.721	−0.421	0.642	−0.487	1	
x_8	−0.307	−0.280	0.065	0.059	−0.277	0.075	−0.380	−0.461	1

Source: own study.

All variables, except x_8 – household consumption expenditure on food average per household member per month, have a high correlation with the dependent variable y. The correlation coefficients ranged from moderate ($r = 0.56$) to high ($r = 0.90$) by absolute mean.

In addition, some predictor variable shows a correlation between them. For example, the correlation coefficient between x_1 variable and the others variables from 0.52 to 0.92 by absolute mean, which is a strong correlation of independent indicators. Just only variable x_8 is the least correlated with all other independent variables.

These results indicate that, perhaps, the assumption of independence does not have enough support. Therefore, the final model would be of high variance and unstable.

In view of the previous remark, the conclusions drawn from the analysis of the scatter plot table will be more important. Finally, looking at the multicollinearity, it is shown the correlation between all predictors in the next graphs (*Figure 1*).
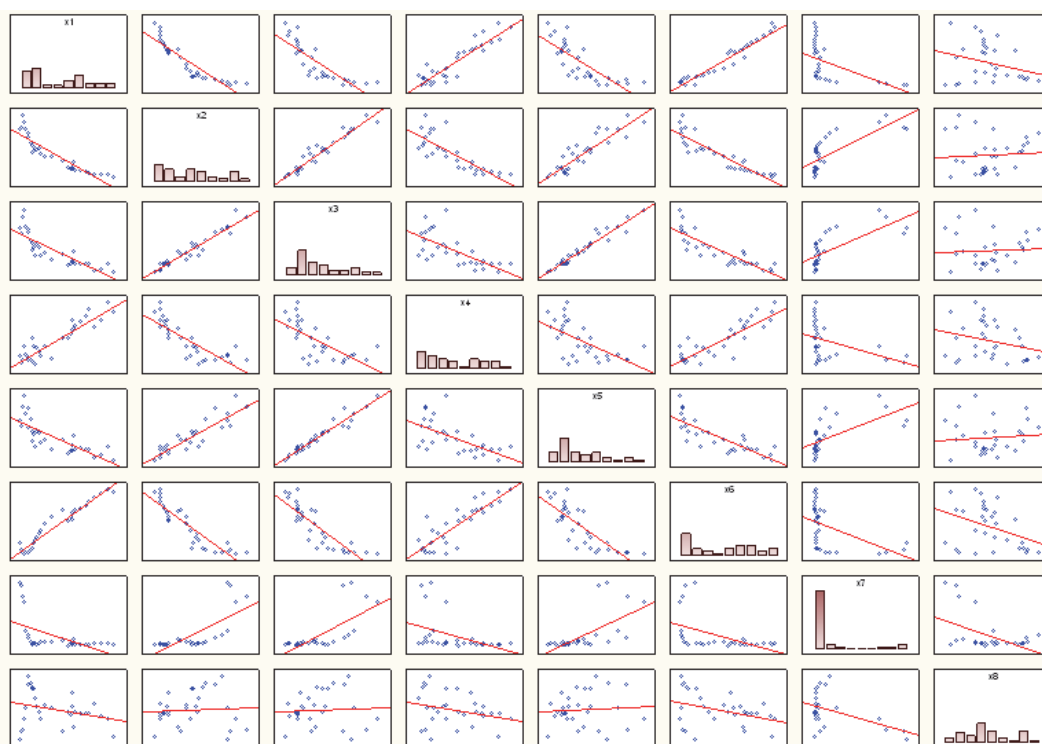

Figure 1. Matrix visualization of correlation between all variables

Source: own elaboration.

Unfortunately, it has shown many predictors correlated that means then can be calculated from others, and there are not independent. The data is rather dirty, outliers are visible, which means deviations from normality. The determinant of this correlation matrix is zero (0) means than exist multicollinearity.

As a result, the dataset and the predictor's variables have multicollinearity.

This problem can be tackled using a technique for dimension reduction, principal component analysis, which reorganize the dataset in components or dimensions independents.

All values of independent variables were normalized, according to the first step of the principal component method. Unfortunately, there is no main objective way to decide how many principal components are enough. This will depend on the specific field of application and the specific data set.

The number of Factors was set for a qualitative Factor analysis, which must be identified so that they fully describe the data and be significant. The number of Factors that should be limited in the future research was selected using the method of the scree plot that was introduced by Cattell (*Figure 2*). Cattell's Scree Plot is a graphical representation of the Factors and their corresponding eigenvalues. The $x - axis$ describes the Factors (components) and the eigenvalues are along the $y - axis$. We can see from the *Figure 2* the first component accounts for the greatest amount of variance, it has the highest eigenvalue. Then the eigenvalues continually decrease, resulting in a picture that is often called the "elbow" shape. The scree plot cutoff is quite subjective, requiring that the number of Factors is limited to those occurring before the bend in the elbow.
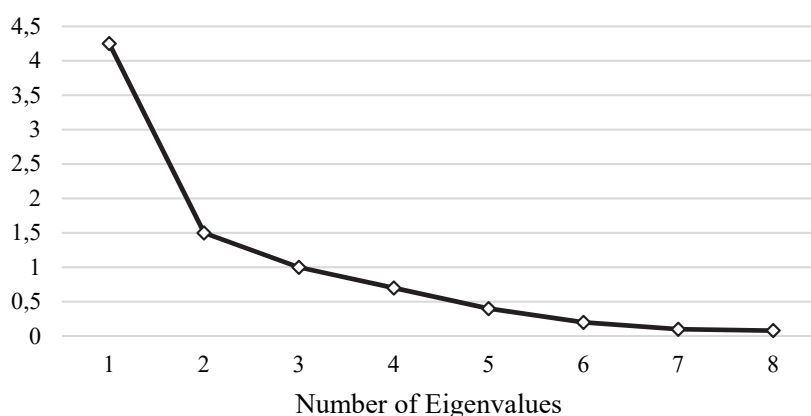

Figure 2. The screen test of the number of eigenvalues

Source: own elaboration.

There are two Factors (PC) having eigenvalues ≥ 1 and elbow point also at the same point. The plot in *Figure 2* suggests that a useful model for these data may have 2 or 3 Factors. The scree plot begins to level off after Factor 3, so the focus should be on these first three concepts.

An alternative method to determine the number of principal components is to look at a Scree Plot, which is the plot of eigenvalues ordered from largest to smallest. The number of components is determined at the point, beyond which the remaining eigenvalues are all relatively small and of comparable size. This assumption is confirmed by the following graph (*Figure 3*).
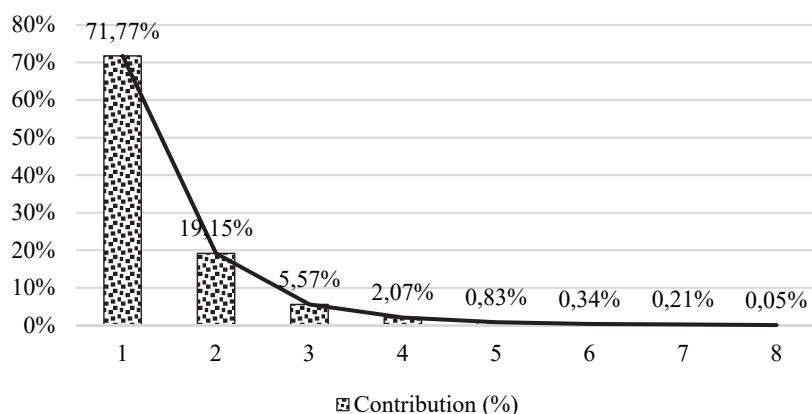
**Figure 3. Percentage contribution of each variable
to the eight principal dimensions, %**

Source: own elaboration.

The next necessary procedure for Factor analysis is the rotation of Factors. After turning, the load is redistributed to each Factor so as to reduce the difference between the loads of all Factors. The varimax method was chosen to rotate the Factor load matrix.

As you can see from *Table 3* (*a*) before the rotation, the main load of the input variables falls on Factor 1. After the rotation *Table 3* (*b*) the loads on all Factors are distributed more evenly, which simplifies their further interpretation. Moreover, marked loadings are higher than 0.7, which means the Factor extracts sufficient variance from that variable.

Table 3

Factor loads without and after rotation

Variable	Factor 1	Factor 2	Factor 3
x_1	**−0.952**	0.192	0.168
x_2	**0.987**	0.097	−0.003
x_3	**0.956**	0.140	0.241
x_4	**−0.853**	0.252	0.432
x_5	**0.908**	0.119	0.359
x_6	**−0.951**	0.256	0.025
x_7	0.684	0.660	−0.048
x_8	0.156	**−0.942**	0.204
Expl.Var	5.742	1.533	0.446
Prp.Totl	0.718	0.192	0.056

Variable	Factor 1	Factor 2	Factor 3
x_1	−0.557	0.094	**−0.807**
x_2	**0.748**	0.124	0.640
x_3	**0.890**	0.082	0.438
x_4	−0.303	0.066	**−0.939**
x_5	**0.926**	0.022	0.330
x_6	−0.637	0.202	**−0.725**
x_7	0.608	0.660	0.316
x_8	0.056	**−0.951**	0.216
Expl.Var	3.390	1.416	2.914
Prp.Totl	0.424	0.177	0.364

a) Factor Loadings (Unrotated)
   Extraction: Principal components
   (Marked loadings are > ,7000)

b) Factor Loadings (Varimax raw)
   Extraction: Principal components
   (Marked loadings are > ,7000)

Source: own elaboration.

Varimax rotation of the Factor axes helped to extract three Factors which accounted for 95 % of the total variance of the original eight variables (*Table 3* (b). Factor pattern coefficients of the rotated Factors show the

relative contribution of each trait to a particular Factor. The first Factor, which explained 42.4 % of the generalized variance was characterized by high positive loadings. The Factor 2 includes just only one variable and describes by 17.7 %. The variables had high negative loadings and contribute to 36.4 % of the variation which was more related to the third Factor.

Factor loadings is understood as correlation coefficients between variables and Factors. Thus, the x_1 variable negative correlates most strongly with Factor 3, namely, the correlation value is –0.807, the x_2 variable with Factor 1 (0.748), the x_8 variable negative correlates most strongly with Factor 3 (–0.951), etc.

Factors can be represented as three orthogonal coordinate axes. If on these axes we mark the values of the Factor loads corresponding to each of the eight normalized indicators, then we get 8 points, each of which characterizes one of the rows of the matrix. As a result, we have the following graph in *Figure 4*.
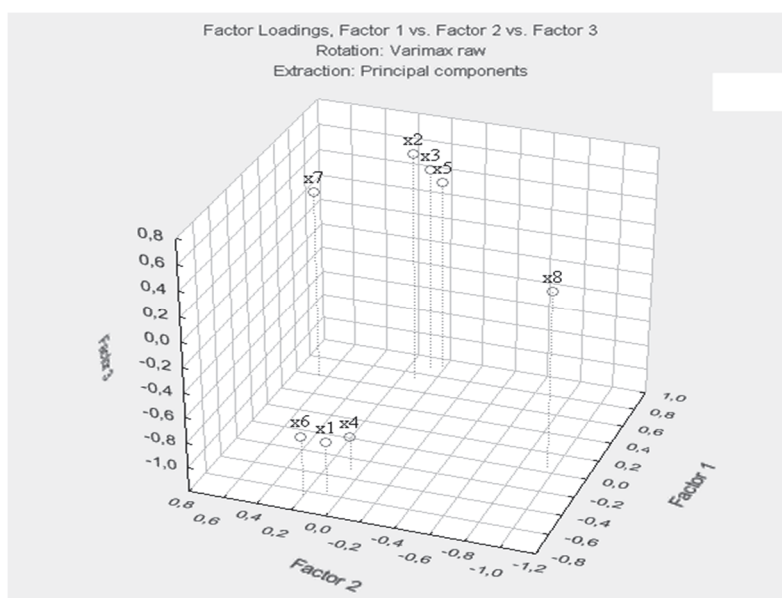

Figure 4. A visual view of Factor loads after varimax rotation

Source: developed by the author.

Therefore, at the end of the axis there are variables that have large loads only in connection with this Factor and, therefore, characterize it. The variables at the origin have little stress due to both Factors. Variables away from the axes are associated with both Factors. If a Factor cannot be clearly defined in terms of its relationship with the original variables, then it should be marked as undetectable or general (common for all variables).

As a result, the following three principal components were obtained.

The first principal component correlates with the following indicators $F1$: Sown area (x_2), harvested crop production (x_3), the output of potatoes per capita (x_5). We can call this component supply

$$f_1 = 0.748x_2 + 0.890x_3 + 0.926x_5.$$

The second principal component correlates with the following indicators $F2$: Household consumption expenditure on food average per household member per month (x_8). We call this component demand

$$f_2 = -0.951x_8.$$

The third principal component correlates with the following indicators $F3$: Average monthly and median wages and salaries (x_1), the average yield of potatoes (x_4), Gross Domestic Product (x_6). We call this component solvency of the population

$$f_3 = -0.807x_1 - 0.939x_4 - 0.725x_6.$$

The eigenvalues were examined to measure the amount of variation retained by each principal component and determine the number of principal components to be considered. The eigenvalues and the proportion of variances (i.e., information) retained by the principal components (PCs) are presented in *Table 4*.

Table 4

Data on the proportion of total variance

Value	Extraction: Principal components			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative, %
1	4.234669	71.77410	4.234669	71.77410
2	1.455577	19.15709	5.691347	90.93119
3	1.049910	5.57364	6.741257	96.50483

Source: own study based on the research results.

The cumulative percentage of the total variance shows that the three components describe 96.5 % of the initial data set. The first component explains 71.77 %, the second – 19.16 %, and the third – 5.6 %. This means that 4.5 % of the data variance is accounted for by other Factors not taken into account. This confirms the correct choice of the number of Factors. The choice of the number of Factors also meets the Kaiser criterion. All values from the eigenvalue column are larger than 1.

Some of the calculated individual Factor scores of the generalized Factors of the components are presented in *Table 5*.

Table 5

Individual Factor scores

Variables	Factor 1	Factor 2	Factor 3
x_1	0.8935	2.4840	0.9947
x_2	0.4282	2.2236	1.5206
x_3	1.8196	1.5238	0.5835
x_4	2.9266	0.5961	–0.6121
x_5	1.4302	–0.3068	0.3226
x_6	0.9883	–1.6046	0.6010
x_7	1.4858	–1.5018	0.1934
x_8	0.7306	–1.4774	0.5944
x_9	–0.5422	–1.0540	1.4244
x_{10}	0.6568	–1.3689	–0.0391
x_{11}	0.1962	–1.2403	0.3859
x_{12}	–1.1799	–0.4248	1.7693

Factor Loadings (Varimax raw)
Extraction: Principal components

Source: own study based on the research results.

Let's construct a regression, where the dependent variables will be the values of the Factors. "Average price of potatoes" (*Y*) indicate as the dependent variable, and all the others as independent ones. The results have been obtained in the form of a table of final regression, which are presented in *Table 6*.

Table 6

Results of final regression model

	Coefficients	Standard Error	t-Stat	P-value
Factor 1	–0.098	0.011	–8.633	3.2E–09
Factor 2	0.018	0.011	*1.566*	*0.12891*
Factor 3	–0.101	0.011	–8.866	1.6E–09

Constant = 0.282, R = ,923, R^2 = ,852, Adjusted R^2 = ,836
F = 51.865 p < ,00000 Stand. error: ,06243

Source: own study based on the research results.

Multivariate multiple linear regression analysis results based on FA and PC scores showed that the overall model was statistically significant – the determination coefficient shows what proportion of the response variance is explained by the influence of predictors in the constructed model. The closer R^2 is to 1, the better the model. In our case, R^2 = 0.852, it means that predictors explained the influence on depend variable by 85.2 %.

Therefore, MLRM model based on Factor analysis and principal component scores for investigating the relations between average price and Factors can be written as:

$$y = 0.282 - 0.098F_1 + 0.018F_2 - 0.101F_3.$$

Therefore, as we can see from the equation, Factors such as F1 and F3 have a negative impact on unit pricing. However, with an increase in the Factors F2, the price of the product will increase.

Conclusions. Theoretical review showed that the pricing of agricultural products, namely potatoes, depends on different micro and macroeconomic features. Price is sensitive to the influence of numerous economic, technological, social Factors, such as inflation, demand, changes in quality, the volume of production, supply in the market, the solvency of consumers, the level of consumption of the product, the volume of inventories, the general economic condition of enterprises, weather conditions and other. To identify the relationship between variables and price is possible using a multiple linear regression model. However, the inclusion of a large number of independent variables in the model can lead to a large correlation between independent variables, which will negatively affect its significance. We decided to solve this problem with the help of Factor analysis, using the method of principal components. Use of Factor analysis and the principal component method helped:

• to take into account the peculiarities of the relationships between the indicators that characterize the process of potato pricing;
• to eliminate multicollinearity between independent indicators;
• to get fewer common Factors than the original number of variables;
• to get three main Factors, using different methods of reducing, such as the Kaiser criterion, scree plot test, total percent variance explained;

• to identify three principal components: first F1 – supply; second F2 – demand; third F3 – solvency of the population;

• to obtain a quantitative assessment of the latent variable;

• the use of orthogonal rotation helped to distribute the load on all three Factors more evenly, which simplified their further interpretation;

• to save 95 % of the total primary information after rotation, which is a high indicator for practical tasks, after reducing the dimensionality of the feature space.

Three Factors whose eigenvalues greater than 1 were selected as independent variables for multiple regression analysis. It was possible to build a significant multiple linear regression model to explore the influence independ variables on price, without multicollinearity problem, using a mixed technique of Factor analysis and principal component method for dimension reduction of variables.

Future research should focus on a detailed study of the relationship between other micro- and macroeconomic indicators and the price of agricultural products. Secondly, this study focuses on only three main Factors without taking into account other ancillary Factors, so it is possible to increase the observation data, and identify more Factors. It is also recommended to conduct research using other economic and mathematical methods to better explain the relationship between variables and Factors. A comprehensive model of the various Factors may provide a better explanation of the case.

Conflict of interest. The author certifies that she has no financial or non-financial interest in the subject matter or materials discussed in this manuscript; the author has no association with state bodies, any organizations or commercial entities having a financial interest in or financial conflict with the subject matter or research presented in the manuscript.

REFERENCES

1. Praburaj, L. (2018). Role of Agriculture in the Economic Development of a Country. *Shanlax International Journal of Commerce*, *6 (3)*, 1-5. https://doi.org/10.5281/zenodo.1323056 [in English].
2. Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of big data. *Big Data*, *4*, Article 20. https://doi.org/10.1186/s40537-017-0077-4 [in English].
3. Issad, H. A., Aoudjit, R. & Rodrigues, J. P. C. (2019). A comprehensive review of Data Mining techniques in smart agriculture. *Engineering in Agriculture, Environment and Food*, *12 (4)*, 511-525. https://doi.org/10.1016/j.eaef.2019.11.003 [in English].
4. Cintina, V. & Pukite, V. (2018). Analysis of influencing Factors of use of agricultural land. *Research for rural development, 1,* 181-187. https://doi.org/10.22616/rrd.24.2018.028 [in English].
5. Manishimwe, R., Niyitanga, F., Nsabimana, S., Kabayiza, A., & Mutimawurugo, M.-C. (2019). Socio-economic and Institutional Factors Influencing the Potato (Solanum tuberosum L.) Production at Smallholder Farmers Level in the Gicumbi District in Rwanda. *Tropicultura*, *37*. https://doi.org/10.25518/2295-8010.604 [in English].
6. Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Qual Quant*, *52*, 1957-1976. https://doi.org/10.1007/s11135-017-0584-6 [in English].

7. Daoud, J. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, *949*. https://doi.org/10.1088/1742-6596/949/1/012009 [in English].

8. Greene, W. H. (2017). *Econometric analysis (8-th ed.)*. New York University [in English].

9. Hansen, B. (2022). *Econometrics.* Princeton University Press [in English].

10. Noora, S. (2021). Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*, *9*, 4-11, https://doi.org/10.1007/10.12691/ajams-9-1-2 [in English].

11. Salmensuu, O. (2017). Macroeconomic Trends and Factors of Production Affecting Potato Producer Price in Developing Countries. *The Journal of Developing Areas, 55*, 91-105 [in English].

12. Chen, Y., & Li, X. (2022). Determining the number of Factors in high-dimensional generalized latent Factor models, *Biometrika*, *109 (3)*, 769-782. https://doi.org/10.1093/biomet/asab044 [in English].

13. Reijer, den A. H. J., Jacobs, P. A. M., & Otter, P. W. (2021). A criterion for the number of Factors. *Communications in Statistics – Theory and Methods*, 50:18, 4293-4299. https://doi.org/10.1080/03610926.2020.1713376 [in English].

14. Choi, I., & Jeong, H. (2019). Model selection for Factor analysis: Some new criteria and performance comparisons. *Econometric Reviews, 38 (6)*, 577. https://doi.org/10.1080/07474938.2017.1382763 [in English].

15. Ricolfi, L., & Testa, S. (2021). Factorial Invariance and Orthogonal Rotation, Multivariate Behavioral Research, *56:5*, 787-804, https://doi.org/10.1080/00273171.2020.1770571 [in English].

16. Yamashita, N., & K. Adachi (2020). Permutimin: Factor Rotation to Simple Structure with Permutation of Variables, Multivariate Behavioral Research, 55:1, 17-29, https://doi.org/10.1080/00273171.2019.1598331 [in English].

17. Hazhar, B., & Dler, K. (2019). An Application of Factor Analysis to Identify the Most Effective Reasons that University Students Hate to Read Books. *International Journal of Innovation, Creativity and Change, 6 (2)*, 251-265 [in English].

18. Jennrich, R. I. (2018). Rotation. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*, 279-304, Wiley Blackwell. https://doi.org/10.1002/9781118489772.ch10 [in English].

19. Tahtali, Y. (2019). Use of Factor scores in multiple regression analysis for estimation of body weight by certain body measurements in Romanov Lambs. *Peer J. 7,* Article e7434. https://doi.org/doi: 10.7717/peerj.7434 [in English].

20. Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research: Design and interpretation (2-nd ed.)*. Sage Publications, Inc. [in English].

21. Bartholomew, D. J. (2011). *Factor Analysis and Latent Variable Modelling*. International Encyclopedia of Statistical Science. Springer [in English].

22. Bryant, F. B., & Yarnold, P. R. (1995). *Principal-components analysis and exploratory and confirmatory Factor analysis.* In L. G. Grimm & P. R. Yarnold (Eds.). Reading and understanding multivariate statistics. American Psychological [in English].

23. Jollife, I. T. (2002). *Principal Component Analysis*. 2-nd ed. New York: Springer-Verlag. https://goo.gl/SB86SR [in English].

24. Cattell, R. B. (1966). The scree test for the number of Factors. *Multivariate Behavioral Research, 1,* 245-276 [in English].

25. Official statistics of Latvia (2005-2021). *Average retail prices of selected commodity (euro per 1 kg, if other – specified) 2005M01 – 2021M07* [Data set]. August 11, 2021. https://data.stat.gov.lv/pxweb/en/OSP_PUB/START__VEK__PC__PCC/PCC010m [in English].