UDC 338.4

VARIABILITY ANALYSIS OF REGIONAL INDICATORS OF
KAZAKHSTAN'S REAL SECTOR OF ECONOMY

Dagmirzaev O.A.
S. Seifullin Kazakh Agro Technical University,
Nur-Sultan, Kazakhstan

Abstract. The study's main purpose is to analyze the variability of regional indicators of Kazakhstan's real sector of economy by leveraging Factor Analysis. Processed indicators encompass 1999–2019, i.e. 21 years. The number of regions is 14. Input information was compressed as follows: mean values of indicators of each region in 21 years were calculated and then one final table was created.

Results of Factor Analysis. All industries of real sector of economy may be lined up in four groups according to common interregional variability criterion: 1) **mining industry**; 2) **manufacturing industry**; 3) **agriculture** (crop and livestock breeding); 4) **transport** and **communication**. According to Factor Analysis methodology, it is considered that there is a certain factor behind the indicators of each selected group, i.e. there is an objective reason. The factors are independent, therefore, the indicators of different groups do not have statistical relations among each other.

The indicators of mining and manufacturing industries have certain regional orientations, i.e. few areas have either mining or manufacturing sectors developed. Agriculture is also a region-dependent sector of the real economy. Significant part of regional variability of livestock sector and almost total interregional variability of population have common factor behind (total variances **0.57** and **0.94**). In other words, livestock breeding is predominantly developed in regions with a denser population. Transport and communication sectors are not bound to certain regions, they flourish in heavily populated areas (total variances **0.69**; **0.85**; **0.61**).

Interregional fluctuations of resource indicator, i.e. capital investments, almost totally match with similar fluctuations in mining and construction (total variances **0.94**; **0.70**; **0.97**). As for capital investments directed to other sectors, their interregional variability is negligible. What is this talking about? Capital investments, except for mining and construction, are pinpoint investments, i.e. allocated to certain regions only.

Keywords: real sector of economy, regions of the country, variability, multivariate statistical analysis, Factor Analysis, variance, correlation coefficients.

Introduction.

The basis of Kazakhstan's national economy is made of manufacturing, agriculture, construction, transport, and communications. Being a unitary state, country consists of 14 regions and 3 cities of republican subordination. The level of economy defines the real sector. At the same time, everything is done locally. Each region has its own uniqueness in terms of economic potential and nature-climatic conditions.

It is assumed that the indicators of real economy are objectively interrelated and regions influence them. To research this, we chose one multivariate statistical analysis method, namely Factor Analysis. General idea of it is as follows: relations among the analyzed indicators may be explained by unmeasurable (hidden or latent) factors. The purpose of this study is to identify and interpret such factors by studying the regional variability (changeability) of indicators of the real economy.

Analysis of publications.

In Kazakhstan, more than half of GDP growth comes from mining both directly and indirectly [1]. The main problem is a delay in diversification of economy, no advancement in sectors other than raw materials. In 2021, the economy is expected to grow up to 3.2% due to the revival of domestic activity, the recovery of global oil demand, and the continued implementation of budget support measures [2].

Analytical works on country's economy or selected industries are usually limited to studying the dynamics of indicators only [3,4]. They may give some important information about the nature and patterns of data, but do not give the full picture of interrelations. It is more interesting to study the hidden part of cause-effect relations among indicators by using Factor Analysis. We have experience in conducting such a study in relation to indicators of the exclusively agricultural sector of the economy of Kazakhstan [5].

We analyzed an extensive list of sources [6-10] to explore Factor Analysis, which has 2 models: Principal Components and Principal Factors. In most cases, these two methods lead to rather similar results, but in fact, they are

different. Principal Components model studies the total variability of the variables, and the Principal Factors focus on the part of indicator's variability that is common for other variables too.

In Factor Analysis, the variance of each variable is conditionally equated to one. Unit variance is represented by the sum of commonality and specificity. Commonality is part of the variance due to the action of factors, i.e. objective reasons. Specificity is a part of the variance due to the specificity of only this variable and observation errors. In other words, the commonality is the total contribution of all factors to the variance of the variable, and the specificity is the difference between the total variance and commonality. There is no consensus on how many factors to examine. As a rule, the Kaiser criterion is used to rationally choose the number of factors [11]: if a factor does not distinguish a variance equivalent to at least the variance of one variable, then it may be ignored.

The results of the research.

Regional real economy is considered as a complex system, evaluated by numerous indicators and examined by Factor Analysis. Primary statistical information was derived from National Statistics Bureau of the Strategic Planning and Reforms Agency of the Republic of Kazakhstan [12–25]. The data of Turkestan region and Shymkent city were combined [25, 26]. Indicators in Tenge were converted to US dollars ($) using average annual exchange rates [27]. The list does not include the cities of Almaty and Nur-Sultan (due to difference in the structure of the economy).

The values of regional indicators of the real sector of economy for 1999–2019 were used, i.e. 21 years. The number of regions is 14. First, the initial information was compressed, for which the following algorithm was used: the average values of the indicators for each area (for 21 years) were calculated and then one final table was created. The STATISTICA 10 software package was used for statistical data processing [28].

The computer program yielded a matrix of factor loads containing correlation coefficients between variables and factors. If we square the factor loads, we get a matrix of common variances of variables and factors, in other words, the commonality between variables and factors (in fractions of one). For convenience purposes, we showed total commonality of factors and variables in percentage.

To improve the structure of factorial loads, we used the orthogonal method of rotation, namely varimax of the original space. When interpreting the analysis results, the **0.5** level was used as the threshold value for total variance. This means that we are only interested in those types of relations that are responsible for more than 50% of the variability. As for the number of factors, we decided to limit ourselves to three factors. Moreover, the validity of the third factor was checked each time by the Kaiser criterion [11].

The following regional indicators of the real sector of the economy were taken into statistical analysis (in million USD per year): **v1** - the volume of manufacturing production (goods, services); **v2** - gross output of agricultural products (services); **v3** - the scope of completed construction work (services); **v4** - gross output of transport services; **v5** - the scope of communication services. Two more variables were added: **p1** – capital investments during the year (in million $); **p2** - population at the end of the year (in thousands).

During data collection, it turned out that data for **v4** and **v5** (gross output of transport and communication services) are only available for 8 regions [12–25]. Therefore, the following hypothesis was proposed: since transport and communication sectors serve people, then **v4** and **v5** will be closely correlated with the population size of the territories, i.e. with indicator **p2**. To test the hypothesis, Factor Analysis was performed, for which we considered the data of 8 areas to be sufficient. The hypothesis is confirmed if **v4**, **v5** and **p2** stand out as a separate group of variables with statistically significant commonality. Then, in next step, it is possible to exclude the indicators **v4** and **v5** from the list of variables, which will make it possible to conduct a full-scale Factor Analysis using data from all 14 regions.

The first part of the study: testing the hypothesis that regional indicators of transport and communications are closely correlated with the population size of the territories.

The average values of indicators for 8 regions for 21 years were used as an input for Factor Analysis according to the scheme "all indicators + 3 factors" (Table 1). We focused on analyzing the regional variability of **v4, v5**, and **p2** (gross output of transport, scope of communication services, and population).

Variables **v5** and **p2** revealed common variances with the second factor: **0.69**; **0.85**; **0.61** (in fractions of one). This is a sign that the majority of variability of gross output of transport and communication services and population is one factor. Common variability means that the variables are interdependent. Conclusion: the indicators **v4** (gross output of transport services) and **v5** (scope of communication services) of the regions are in close statistical relationship with population size of the territories.

The second part of the study (the main part of the work).

We excluded gross output of transport and communication services from the list of indicators since they are the variables with statistically significant commonality. We divided the agriculture's gross output into two categories: a) gross output of crop breeding; b) gross output of livestock breeding.

Table 1

Total variances according to the scheme "all indicators + 3 factors"

Indicators	Total variances in fractions of one			Sum of total variances	Contribution of indicator to the total variance, %
	Factor 1	**Factor 2**	**Factor 3**		
v1	**0.83**	0.03	0.13	0.99	14.1
v2	0.16	0.03	**0.77**	0.96	13.8
v3	**0.92**	0.03	0.03	0.98	14.1
v4	0.15	**0.69**	0.07	0.91	12.9
v5	0.09	**0.85**	0.04	0.98	14.1
p1	**0.95**	0.01	0.04	1.00	14.2
p2	0.14	**0.61**	0.14	0.89	12.6
Sum of total variances	**3.24**	**2.24**	**1.22**	**6.70**	
Contribution of factor to the total variance, %	**46.3**	**32.1**	**17.4**		**95.8**
Source: The result of Factor Analysis of regional economic indicators of 8 regions of Kazakhstan within 1999–2019 [12-25], performed by the author of the article					

List of analyzed variables: **v1** - manufacturing output (goods, services); **v2** - gross output of crop breeding; **v3** - gross output of livestock breeding; **v4** - volume of completed construction work (services); **p1** – capital investments; **p2** - population size at the end of the year.

Table 2 shows the average values of indicators for all 14 regions in 21 years.

Table 2

Average values of all regions indicators within 21 years (data 1999 - 2019)

#	Regions	Average values of indicators of the real sector of the economy for 21 years ($ million)				Average capital investments ($ million)	Average population (thousand people)
		v1	**v2**	**v3**	**v4**	**p1**	**p2**
1	**Akmola region**	1093	642	292	278	604	746
2	**Aktobe region**	4512	205	322	484	1603	758
3	**Almaty region**	2013	837	720	650	1401	1789
4	**Atyrau region**	14565	47	99	1448	5661	527
5	**East Kazakhstan region**	3707	420	581	492	1048	1420
6	**Jambyl Region**	869	307	259	292	615	1045
7	**West Kazakhstan region**	5305	143	215	411	1214	619
8	**Karaganda region**	5639	245	344	538	1260	1358
9	**Kostanay region**	1986	737	424	237	614	900
10	**Kyzylorda region**	3488	157	84	333	899	687
11	**Mangistau region**	7566	4	25	600	1606	490
12	**Pavlodar region**	4108	211	241	353	1026	753
13	**North Kazakhstan region**	523	896	337	123	369	619
14	**Turkestan region**	2670	690	545	727	1498	2500
Source: The result of statistical processing of regional economic indicators of 14 regions of Kazakhstan within 1999–2019 [12-25], performed by the author of the article							

To display data in a more concise form, we performed statistical processing of Table 2. Processing results: a) average regional means; b) standard deviations and coefficients of variation of the obtained means; c) the correlations among means of regional indicators. Table 3 shows the values of obtained results (measurement unit of **p2** is thousand people, other indicators are in million $). As you can see, the interregional variation of the mean of all indicators ranges from **61** to **87%**. The largest interregional fluctuation is observed in the mean of capital investment

(coefficient of variation is **93%**). Regarding the paired correlation coefficients analysis, this tool has limited capacity to adequately interpret the relations among indicators when their number is more than two.

Table 3

Results of statistical processing of indicators' average values for all regions

Indi-cators	Average of regional means	Standard deviation of the obtained means	Variance of the obtained means, %	Paired coefficients of correlation					
				v1	**v2**	**v3**	**v4**	**y1**	**y2**
v1	**4146**	**3607**	**87**	1.00	**-0.69**	-0.49	**0.86**	**0.92**	-0.31
v2	**396**	**305**	**77**		1.00	**0.74**	-0.36	-0.44	0.48
v3	**321**	**197**	**61**			1.00	-0.10	-0.28	**0.78**
v4	**498**	**322**	**65**				1.00	**0.96**	0.13
p1	**1387**	**1293**	**93**					1.00	-0.13
p2	**1015**	**571**	**56**						1.00
Source: The result of statistical processing of regional economic indicators of 14 regions of Kazakhstan within 1999–2019 [12-25], performed by the author of the article									

On the basis of Table 2 (with the average values of indicators across all 14 regions for 21 years), a Factor Analysis was performed according to the scheme "all indicators + 3 factors" (Table 4). The share of the third factor in total variance is **22.2%**, which exceeds the contribution of one indicator's variance to the total variance on average **15-16%** (the third factor is justified by the Kaiser criterion) [11].

Table 4

Total variances according to the scheme "all indicators + 3 factors"

Indicators	Total variances in fractions of a unit			Sum of total variances of indicator	Contribution of indicator to the total variance, %
	Factor 1	**Factor 2**	**Factor 3**		
v1	**0.76**	0.06	0.15	0.97	16.2
v2	0.09	0.10	**0.79**	0.98	16.3
v3	0.02	**0.57**	0.33	0.91	15.2
v4	**0.95**	0.02	0.02	0.99	16.5
p1	**0.96**	0.01	0.01	0.99	16.5
p2	0.00	**0.94**	0.03	0.97	16.2
The sum of the total variances from the factor	**2.79**	**1.70**	**1.33**	**5.82**	
Contribution of factor to the total variance,%	**46.4**	**28.4**	**22.2**		**97**
Source: The result of statistical processing of regional economic indicators of 14 regions of Kazakhstan within 1999–2019 [12-25], performed by the author of the article					

The **first** factor explains the interregional variability of: **v1** - the volume of manufacturing production (goods, services); **v4** - the volume of completed construction work (services); **p1** – capital investments. In other words, the regional fluctuations of: a) manufacturing; b) the construction; c) capital investments, subject (mainly) to one factor with total variances **0.76**; **0.95**; **0.96**.

The **second** factor highlights the variances of gross output of the livestock and the population (**0.57** and **0.94**). Comment: A significant part of the interregional variability of the livestock sector and practically all of the interregional variability of the population indicator is explained by one factor. Simply saying, this means livestock production is more developed in regions with a denser population.

Third factor explains the interregional variability of **v2** crop production: total fractional variance is **0.79**. It does not influence the livestock sector significantly (the commonality is only 0.33).

Preliminary conclusions. Interregional variability of indicators: manufacturing (plus construction) and agriculture are unrelated sectors. Some regions are predominantly industrial, and some are agricultural. It appeared that crop and livestock sectors are gravitating towards different factors. In other words, regions with developed crop production and well-developed livestock production are, as a rule, different territories.

It appeared that branches of the real economy may be grouped into three according to interregional variability criterion: 1) manufacturing and construction; 2) agriculture; 3) transport and communication.

The conclusions dictate the need to disclose the manufacturing sector, it is the sum of: a) mining (full name: mining and excavation); b) manufacturing; c) power supply and gas supply; d) water supply and sewerage system.

To obtain a complete picture of the relations, we considered it necessary to perform the Factor Analysis with additional inclusion of all structural indicators to the list.

Here is a list of analyzed variables: **v1** - the volume of production in the mining industry; **v2** - the volume of production of the manufacturing industry; **v3** - the volume of the electricity and gas supply industry; **v4** - the volume of the water supply and sewerage system; **v5** - gross output of the crop breeding industry; **v6** - gross output of the livestock industry; **v7** - the volume of completed construction work (services); **p1** - capital investments during the year; **p2** - population size at the end of the year.

We performed the preparatory work for calculating average values of new indicators for all 14 regions for 21 years and filled out the matrix of initial data (as in the case of Table 2). The results are shown in Table 5. **Three factors** explain about **90%** of total variance of the researched indicators. The contribution of each factor to the total variance of indicators is approximately **30%**. The contribution of each indicator to the total variance is about **9-10%** on average.

The **first** factor explains the interregional variability of mining and construction (total variances **0.70** and **0.97**). This group also includes a resource indicator, namely capital investments with a total variance of **0.94**. The **second** factor controls the interregional variability of: a) manufacturing; b) power supply and gas supply; c) water supply and sewerage system (total variances **0.80**; **0.83**; **0.76**). The **third** factor is associated with interregional variability of agriculture: crop and livestock production (total variances **0.63** and **0.89**). A resource indicator is also adjacent to this; the population size has **0.80** total variance.

Table 5

Total variances according to the scheme "all indicators + 3 factors"

Indicators	Total variances in fractions of a unit			Sum of total variances of indicator	Contribution of indicator to the total variance, %
	Factor 1	**Factor 2**	**Factor 3**		
v1	**0.70**	0.02	0.25	0.97	10.8
v2	0.00	**0.80**	0.11	0.91	10.1
v3	0.04	**0.83**	0.00	0.87	9.7
v4	0.10	**0.76**	0.03	0.89	9.8
v5	0.11	0.06	**0.63**	0.79	8.8
v6	0.03	0.01	**0.89**	0.93	10.3
v7	**0.97**	0.01	0.00	0.99	11.0
p1	**0.94**	0.00	0.02	0.97	10.7
p2	0.00	0.03	**0.80**	0.84	9.3
Sum of total variances from the factor	**2.89**	**2.52**	**2.73**	**8.14**	
Contribution of factor to the total variance, %	**32.1**	**28.0**	**30.4**		**90.5**
Source: The result of statistical processing of regional economic indicators of 14 regions of Kazakhstan within 1999–2019 [12-25], performed by the author of the article					

In the previous analysis (Table 4), the crop and livestock sectors gravitated towards different factors according to the criterion of interregional variability (we explained it by the fact that livestock farming is more tied to regions with a higher population density). In the last analysis, the picture is different. Let's try to figure it out. In table 5, the number of factors was limited to three units. It appeared that crop breeding and population were influenced by the third factor merely due to the lack of a degree of freedom. In case of adding the fourth factor, a statistically significant part of the variances of livestock breeding and population will go to a new factor.

Conclusion.

All industries of real sector of economy of Kazakhstan may be lined up in four groups according to common interregional variability criterion: 1) **mining**; 2) **manufacturing**; 3) **agriculture** (crop and livestock); 4) **transport** and **communication**. According to Factor Analysis methodology, it is considered that a certain factor stands behind the indicators of each selected group, i.e. there is an objective reason. The factors are independent, therefore, the indicators of different groups do not have statistical relationships with each other.

The indicators of mining and manufacturing industries have certain regional orientations, i.e. few areas have either mining or manufacturing sectors developed. Agriculture is also a region-dependent sector of the real economy. Significant part of regional variability of livestock sector and almost total interregional variability of population have common factor behind. In other words, livestock breeding is predominantly developed in regions with a denser population. Transport and communication sectors are not bound to certain regions, they flourish in heavily populated areas.

Interregional fluctuations in resource indicator, i.e. capital investments almost totally match with similar fluctuations in mining and construction. As for capital investments directed to other sectors, their interregional variability is negligible. What is this talking about? Capital investments, except for mining and construction, are pinpoint investments, i.e. allocated to certain regions only.
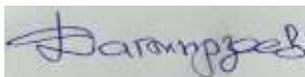
Information about the author:
Dagmirzaev Omirbek, candidate of technical sciences, S. Seifullin Kazakh Agro Technical University, Nur-Sultan, Kazakhstan; https://orcid.org/0000-0002-7766-8012

References
[1] Forbes Expertise dated June "Economy of Kazakhstan: what's next - stability or reforms?" (2019). URL: https://forbes.kz/process/expertise/ekonomika_kazahstana_chto_dalshe_-_stabilnost_ili_reformyi/

[2] World Bank in Kazakhstan (2021). URL: https://www.vsemirnyjbank.org/ru/country/kazakhstan/

[3] Kazakhstan in figures (2017). Ministry of National Economy of the Republic of Kazakhstan, Statistics Committee, p. 32.

[4] Rudert D., Musaeva M. (2019). Agricultural sector of Kazakhstan: problems and development prospects // Transformation of the economy of Kazakhstan / Konrad Adenauer Foundation, p. 207-232.

[5] A. A. Nurpeisova, Sh. U. Niyazbekova, O. A. Dagmirzaev and others (2020). «STATISTICAL ANALYSIS OF VARIABILITY OF DATA OF THE AGRARIAN SECTOR OF KAZAKHSTAN ECONOMY». – BULLETIN OF NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN, ISSN 1991-3494, Volume 4, Number 386, 165 – 172. **https://doi.org/10.32014/2020.2518-1467.116**

[6] Lawley D.N., Maxwell A.E. (1967). Factor Analysis as a statistical method, p. 144.

[7] Kim, J.O., Mueller C.W., Klekka W.R. (1989). Factorial, discriminant and cluster analysis, p. 215.

[8] Kim, J.O., & Mueller, CW (1978). Factor Analysis: Statistical methods and practical issues. Beverly Hills, CA: Sage Publications.

[9] Kim, JO, & Mueller, CW (1978). Introduction to Factor Analysis: What it is and how to do it. Beverly Hills, CA: Sage Publications.

[10] Electronic textbook on statistics StatSoft (1999-2021). Principal components and Factor Analysis. URL: http://www.statsoft.ru/home/textbook/modules/stfacan.html

[11] Kaiser, H.F. (1960). The application of electronic computers to Factor Analysis. Educational an Psychological Measurement, 20, 141-151.

[12] Main socio-economic indicators of Akmola region. URL: https://stat.gov.kz/region/247783/dynamic

[13] Dynamics of the main socio-economic indicators of Aktobe region 1991–2019. URL: https://stat.gov.kz/region/248875/dynamic

[14] Dynamics of the main socio-economic indicators of development of Almaty region for 1991–2019. URL: https://stat.gov.kz/region/250502/dynamic

[15] The main socio-economic indicators of Atyrau region for 1991–2019. URL: https://stat.gov.kz/region/252311/dynamic

[16] The main socio-economic indicators of East Kazakhstan region 1991–2019. URL: https://stat.gov.kz/region/264990/dynamic

[17] Main socio-economic indicators of Zhambyl region. URL: https://stat.gov.kz/region/255577/dynamic

[18] Dynamics of the main socio-economic indicators of West Kazakhstan region. URL: https://stat.gov.kz/region/253160/dynamic

[19] The main socio-economic indicators of Karaganda region. URL: https://stat.gov.kz/region/256619/dynamic

[20] Dynamics of the main indicators of SED of Kostanay region 1991–2019. URL: https://stat.gov.kz/region/258742/dynamic

[21] Main socio-economic indicators of Kyzylorda region. URL: https://stat.gov.kz/region/260099/dynamic

[22] Dynamics of the main socio-economic indicators of development of Mangystau region. URL: https://stat.gov.kz/region/260907/dynamic

[23] Dynamics of the main indicators of the SED of Pavlodar region 1991–2020. URL: https://stat.gov.kz/region/263009/dynamic

[24] The main socio-economic indicators of the North Kazakhstan region. URL: https://stat.gov.kz/region/264023/dynamic

[25] Dynamics of the main indicators of socio-economic development of Turkestan region. URL: https://stat.gov.kz/region/20243032/dynamic

[26] The main socio-economic indicators of Shymkent city 1991–2019. URL: https://stat.gov.kz/region/20242100/dynamic

[27] Main socio-economic indicators of the Republic of Kazakhstan. URL: https://stat.gov.kz/official/dynamic

[28] New features of STATISTICA 10 (1999-2021). URL: http://statsoft.ru/products/new-features/STATISTICA10.php

Dagmirzaev Omirbek