Analysis Method of Structural Equation Modeling

A. V. Burkov[1], E. A. Murzina[2]

[1]*Applied Statistics and Informatics Department, Mari State University,*
Lenina 1, Yoshkar-Ola, 424002, Mari El, Russia
[2]*Department of Management and Law, Volga State University of Technology,*
Lenina 3, Yoshkar-Ola, 424002, Mari El, Russia

Annotation
The goal of this article is to define a set of factors which most strongly influence a probability of finding (opening) self-employment by graduates of high schools of the United States of America. Different methods for reducing the number of independent variables are reviewed and compared. First, it is the correlation analysis. This method was used for the assessment of these factors' influence on self-employment of graduates of high schools of the United States of America. Secondly, it is the classical factor analysis, and finally it is the Confirmatory Factor Analysis with application of Structural Equations. These methods were used to decrease the quantity of the variables influencing self-employment. Confirmatory Factor Analysis with application of Structural Equations was used for the confirmation of the received results. As the data base the information of the National Science Foundation of the United States of America was used. Moreover in this article the factors which have the most significant influence on self-employment of American graduates are described.

Keywords: Correlation analysis, Factor analysis, Structural equations, Higher education, Labor market.

1 Introduction

In the modern world education is one of the most important prerequisites of successful development of society. Sustained economic growth of any country and its competitiveness in the times of globalization of world economy are impossible without highly educated and highly skilled laborers.

When considering the tasks facing higher education and statistics of education, the development of a methodological basis of statistical analysis of higher education expert labor market and the employment of such specialists for jobs in their degree field seems potentially productive and extremely important. It is important for national and international studies.

The experience of the USA is interesting from the point of view that if in the Russian Federation the introduction of a two-stage system of higher education is still developing, in the USA the similar experience of training specialists is already obtained. To our mind, special attention should be paid to the criteria of education quality which is used in the USA. The USA is a recognized leader among other countries with market economy, it is a country with a highly developed private sector. A considerable share of concentration of medium-sized and small business is in many respects caused by the mentality: in the USA the choice of speciality is influenced by possibilities of employment, but the prestige of a business also influences the education one's gets.

The organization of business by graduates from American higher education institutions is influenced by the whole set of factors. Taking into consideration everything said before, at the first stage of our research it is necessary to define factors influencing the basis of a new business organized by American university graduates.

2 Research Methodology

The methods used in the research to find out factors which most strongly influence a probability of finding (opening) self-employment by graduates of high schools in the USA are the following: correlation analysis, Eigenvalue method, factor analysis. To prove the results the method of structural equations was used. The data base was the information of the National Science Foundation of the USA (http://www.nsf.org).

3 Findings and Discussion

Our goal is to define a set of factors which most strongly influence a probability of finding (opening) self-employment by graduates of high schools in the USA. We used the database on experts with higher education of the National Science Foundation of the USA. There are 447 parameters for each graduate. For the construction of a regression model we cannot use all of 447 parameters from the database, therefore, it is necessary to make a reduction of data which is carried out in several stages.

On the basis of expert estimation, the parameters which have no influence on the process of self-employment have been removed from the database. For example: ”Taking courses during last (reference) week“, ”Living in the USA during last (reference) week“ or ”Reason for working less than 35 hours a week“. At the given stage the number of parameters was reduced from 447 to 224.

Using the correlation analysis (pair correlation), from 224 parameters those which most strongly influence the probability of self-employment have been selected. As a result, 36 parameters have been selected with *the module* of *pair correlation coefficients* of 0,1 or more (Table 1).

The parameters with high autocorrelation have been removed from the selected 36 parameters: for example, parameters ”Age Group [5 year intervals]” and ”Year, date of birth [recoded for public use]” as they correlate with the "Age" parameter (the value of correlation is 0,99). Moreover, some variables can be grouped together: for example, ”Employer size“ or ”Type of educational institution [employer]” (Table 1). Hence, such parameters cannot be used as independent variables and should be removed. After the removal of highly correlated parameters and parameters that can be grouped together only 18 parameters were left.

It is noteworthy that at first there were 100 000 cases in the database, but after the removal of cases with missing values, there were 569 cases left in the database. To check the received representative sampling we compared histograms of distribution of the variable "Age", both for the general set (Fig.1) and for sampling (Fig.2).

So, the histogram of distribution of the variable “Age” for the general set does not considerably differ from the same histogram for sampling, it is possible to make a conclusion about the representative sampling.

On the basis of the remained 18 parameters the 3 and 2 factor models were received using factor analysis. The Kaiser criterion (Kaiser, 1960) based on Eigenvalues was used to find out the number of factors (Table2) and the Cattell’s scree test (Cattell, 1966).

Table 1 Correlation ot the parameters

Name	DESCRIPTIONS	INCLUDED	GROUP	CORRELATIONS
AGE	Age	X		0,11
AGEGR	Age Group [5 year intervals]		X	0,11
BIRYRP	Year, date of birth [recoded for public use]			-0,11
EMSEC SM	Employer sector [Summary code]		X	0,31
EMSEC PB	Employer sector [recoded for public use]		X	0,29
HDACY R	Academic year of highest degree			-0,14
DGRYR	Year of highest degree	X		-0,14
HDACY 3	Year of highest degree [3 year intervals]		X	-0,14
HDAY5	Year of highest degree [5 year intervals]		X	-0,14
HSYR	Year of receiving high school diploma	X		-0,11
ACDRG	Type of degree valid during the week of Oct. 1		X	-0,12
ACTRD T	Activity, Research, Development, and Teaching	X		-0,14
ACTTC H	Activity, Teaching	X		-0,16
EMED	Employer is an educational institution	X		-0,25
EMSIZE	Employer size		X	-0,49
EDTP	Type of educational institution [employer]		X	0,33
FPTIND	Full-time/part-time status including all jobs during the reference week	X		-0,13
NEDTP	Type of a non-educational institution [employer]		X	-0,57
NRFAM	Reason for working outside the highest degree field: family-related reasons	X		0,10
NROCN A	Reason for working outside the highest degree field: a desired job is not available	X		-0,10
WAACC	Work activities on principal job: accounting, finance, contracts	X		0,16
WAPRS M	Summarized primary work activity		X	0,15
WASVC	Work activities in the principal job: professional services	X		0,13
WASAL E	Work activities in the principal job: sales, purchasing, marketing	X		0,15
WATEA	Work activities in the principal job: teaching	X		-0,14
NEWBU S	New business	X		0,25
BAACY R	Academic year of first bachelor degree			-0,11
BAYR	Year of first bachelor's degree	X		-0,11
MRDAC YR	Academic year of most recent degree			-0,14
MRYR	Year of most recent degree	X		-0,14
MR3YR	Year of most recent degree [3-year intervals]		X	-0,14
MR5YR	Year of most recent degree [5-year intervals]		X	-0,14
D2AYR	Degree award date based on academic year			-0,12
D2YR	Year of second highest degree	X		-0,12

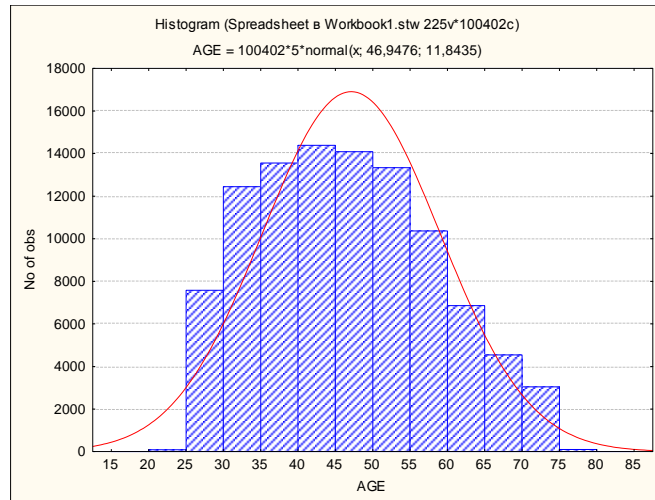| D3AYR	Degree award date based on academic year			-0,10
D3YR	Year of third highest degree	X		-0,10

Fig. 1 Histogram of distribution of the variable "Age" for the general set
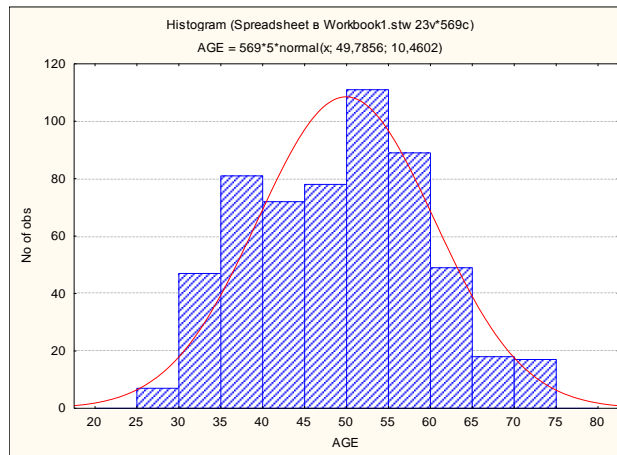

Fig.2 Histogram of distribution of the variable "Age" for sampling

Table 2 Eigenvalues of factors

Eigenvalues Extraction: Principal components

Eigen value	% Total Variance	Cumulative Eigenvalue	Cumulative %
6,372488	35,40271	6,37249	35,40271
2,925216	16,25120	9,29770	51,65391
1,334197	7,41221	10,63190	59,06611
1,118176	6,21209	11,75008	65,27820
1,062368	5,90204	12,81244	71,18025

The scree plot of Eigenvalues is given in Fig.3.

From the Table of Eigenvalues it is clear, that the greatest share of variance is being described by the first two factors, the same fact is confirmed by the plot, it has excesses on the second and third points (Fig.3). Hence, it is most logical to consider the 2 and 3 factor models. Of the factor analysis methods we selected the principal component analysis with various variants of rotation of axes for the 3 factor and 2 factor models. For rotation of axes the following methods were used: Unrotated, Varimax raw, Varimax normalized, Biquartimax raw,

Biquartimax normalized, Quartimax raw, Quartimax normalized, Equamax raw and Equamax normalized.
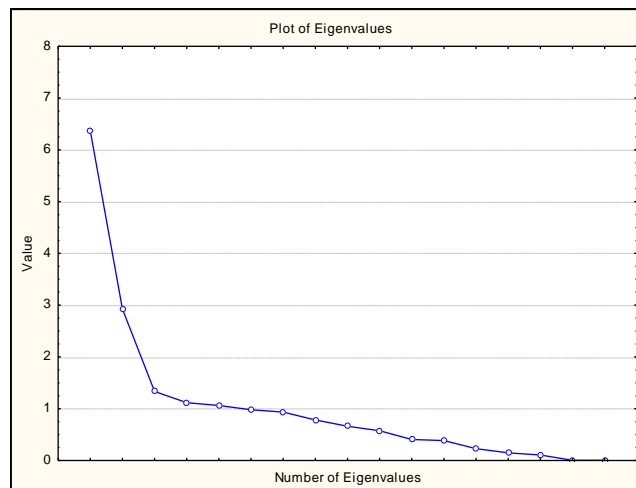

Fig.3 The scree plot of Eigenvalues

As the criterion of of the rotation of axes method, the values of factor loadings were used. The method with the greatest values of factor loadings is considered the best. Factor loading is considered to be high if its value is 0,5 or more. The variable joined in the factor which has the greatest factor loading. Moreover, the best model should have a minimal correlation between factors.

After analyzing models with various variants of rotation of axes, we came to the conclusion, that the optimal method is Varimax normalized, both for the 3 factor and 2 factor models. Factor loadings for the given models are given in Table 3 and Table 4 (the most significant loadings are in bold type).

Relying on factor loadings, it is possible to assign variables to those factors. Let's have a closer look at the received models and describe them in detail.

The received 3 factor model describes dependence of probability of self-employment on the following 3 factors: "Experience", "The Attitude to education and science" and "Business characteristics".

The factor «Experience» describes work experience of a graduate and is linear approximation of the following characteristics of a high school graduate: "Age", "Year of highest degree", "Year of receiving a high school diploma", "Year of first bachelor's degree", "Year of most recent degree", "Year of second highest degree", "Year of third highest degree". It is noteworthy that factor loadings of variables for this factor are higher than 0,9, that shows good approximation of variables by the given factor.

The factor "The Attitude to education and science" shows the attitude to education and science, how much the activity of a graduate is connected with education and science, it is linear approximation of the following characteristics: "Activity, Research, Development, and Teaching", "Activity, Teaching", "Employer is an educational institution", "Work activities in the principal job: teaching". Factor loadings for this factor are not so unequivocal if compared with the previous ones, but all their values are high and not lower than 0,6.

Table 3 Factor Loadings for the 3 factor model
Factor Loadings (Varimax normalized)
Clusters of loadings are marked; they determine the oblique factors for hierarchical analysis

	Factor 1	Factor 2	Factor 3
AGE	**0,952996**	0,061468	0,020654
DGRYR	**0,921701**	-0,082739	0,060049

Table 4 Factor Loadings for the 2 factor model
Factor Loadings (Varimax normalized)
Clusters of loadings are marked; they determine the oblique factors for hierarchical analysis

	Factor 1	Factor 2
AGE	**0,954366**	-0,012021
DGRYR	**0,916927**	0,135735

HSYR	**0,956896**	0,061130	0,020929
ACTRDT	-0,044594	**0,653206**	-0,280236
ACTTCH	0,049863	**0,902174**	0,011445
EMED	0,023532	**0,801382**	-0,049644
FPTIND	0,117082	0,168499	**0,363400**
NRFAM	-0,057464	0,163233	**0,525861**
NROCNA	0,092724	0,064859	**0,288000**
WAACC	0,019644	-0,346718	**0,524974**
WASVC	0,039137	-0,049291	**0,275367**
WASALE	-0,003421	-0,279992	**0,580865**
WATEA	0,030087	**0,821298**	0,144960
NEWBUS	-0,077575	-0,102705	**0,364370**
BAYR	**0,977493**	0,059618	0,033453
MRYR	**0,923060**	-0,084538	0,059542
D2YR	**0,952719**	0,025186	0,047984
D3YR	**0,955046**	0,078714	0,070306

HSYR	**0,958248**	-0,011453
ACTRDT	0,022482	**0,700040**
ACTTCH	-0,105162	**0,874567**
EMED	-0,068952	**0,790894**
FPTIND	**0,149831**	-0,079367
NRFAM	0,014094	**0,046620**
NROCNA	**0,114510**	0,004039
WAACC	0,031704	**0,453740**
WASVC	0,053400	**0,110096**
WASALE	0,016311	**0,399886**
WATEA	-0,088993	**0,767415**
NEWBUS	-0,060479	**0,176327**
BAYR	**0,979465**	-0,006301
MRYR	**0,918140**	0,137439
D2YR	**0,953615**	0,029312
D3YR	**0,960593**	-0,017869

Finally, the third factor « Business characteristics » describes a new business where a graduate is employed, it is linearization of characteristics: "Full-time/part-time status including all jobs during the reference week", "Reason for working outside the highest degree field: family-related reasons", "Reason for working outside the highest degree field: a desired job is not available", "Work activities in the principal job: accounting, finance, contracts", "Work activities in the principal job: professional services", "Work activities in the principal job: sales, purchasing, marketing", "New business". Factor loadings for the variables of this factor are not so significant, they all exceed two times the values of factor loadings for other factors and for some variables do not exceed 0,15.

The received 2 factor model describes the dependence of probability of self-employment on the following 2 factors: « Experience and environment conditions » and « Business characteristics ».

The first factor « Experience and environment conditions » describes work experience of a graduate and work conditions. It is linear approximation of the following characteristics: "Age", "Year of highest degree", "Year of receiving a high school diploma", "Full-time/part-time status including all jobs during the reference week", "Reason for working outside the highest degree field: a desired job is not available", "Year of first bachelor's degree", "Year of most recent degree", "Year of second highest degree" and "Year of third highest degree". Practically all factor loadings of variables for the given factor have high value of more than 0,9, except for factor loadings for parameters "Full-time/part-time status including all jobs during the reference week" and "Reason for working outside the highest degree field: a desired job ia not available", the value of their factor loadings does not exceed 0,15. This fact shows the low influence of these parameters on this factor.

The second factor "Business characteristics", as well as the third factor in the 3 factor models characterizes the business in which a graduate is employed. The given factor is linearization of the following parameters: "Activity, Research, Development, and Teaching", "Activity,

Teaching", "Employer is an educational institution", "Reason for working outside the highest degree field: family-related reasons", "Work activities in the principal job: accounting, finance, contracts", "Work activities in the principal job: professional services", "Work activities in the principal job: sales, purchasing, marketing", "Work activities in the principal job: teaching", "New business". The situation with factor loadings for this factor is similar to the first factor of the given model. Factor loadings change within 0,8-0,04, and the parameter "Reason for working outside the highest degree field: family-related reasons" has the least influence on the given factor.

Below there are given Tables of correlations between factors (Table 5 and Table 6), both for the 3 factor and the 2 factor models.

Table 5 Correlations between factors for the 3 factor model
Correlations between oblique factors
(Clusters of variables with unique loadings)

	Factor 1	Factor 2	Factor 3
Factor 1	1,000000	0,033756	0,088356
Factor 2	0,033756	1,000000	-0,182457
Factor 3	0,088356	-0,182457	1,000000

Table 6 Correlations between factors for the 2 factor model
Correlations between oblique factors
(Clusters of variables with unique loadings)

	Factor 1	Factor 2
Factor 1	1,000000	-0,022367
Factor 2	-0,022367	1,000000

The level of correlation between factors in the both models is low enough, this confirms their importance. In the 3 factor model the correlation between factors is 0,03, 0,08 and -0,18 accordingly, and in the 2 factor model the correlation between factors is -0,02.

If to consider the correlation between probability of self-employment and the received factors (Table 7 and Table 8), we can come to the following conclusions. First, taking into consideration the 3 factor model it is obvious that the third factor has the greatest influence on probability of self-employment, it is followed by the first factor and then the second one. Secondly, with the growth of the first or third factors, the value of probability decreases, and with the growth of the second factor, the probability increases. Thirdly, in the 2 factor model the both factors have equally negative influence on probability of self-employment.

Table 7 Correlation between probability of self-employment and the received factors for the 3 factor model
Correlations
Marked correlations are significant at $p <, 05000$
N=569
(Casewise deletion of missing data)

	Factor 1	Factor 2	Factor 3
SELFEMPL	-0,20	0,11	-0,37

Table 8 Correlation between probability of self-employment and the received factors for the 2 factor model
Correlations
Marked correlations are significant at $p <, 05000$
N=569
(Casewise deletion of missing data)

	Factor 1	Factor 2
SELFEMPL	-0,22	-0,22

If to compare the 3 factor and the 2 factor models, it is possible to draw the following conclusion: though the correlation between factors in the 2 factor model is less than in the 3 factor model, from the point of view of the explanation of factor value, the 3 factor model is better. We will verify the results of the factor analysis using structural equations.

Reliability of the received factors was verified by the confirming factor analysis using structural equations. Using the structural equations for both the 3 factor and the 2 factor models, we employed the method of maximum likelihood estimation together with the method of least squares, with absence of correlation between factors and the residuals since low correlation between factors was given above (Table 5 and Table 6). The data for the analysis was the matrix

of correlations between parameters. the results of the analysis of the constructed structural equations were the following.

Table 9 The confirming factor analysis using structural equations for the 3 factor model

Model Estimates

	Parameter Estimate	Standard Error	T Statistic	Prob. Level
(Experience)-1> [AGE]	**-0,980**	**0,007**	**-1,409422E+02**	**0,000**
(Experience)-2> [DGRYR]	**1,000**	**0,000**	**8,086041E+03**	**0,000**
(Experience)-3> [HSYR]	**0,980**	**0,007**	**1,416687E+02**	**0,000**
(Experience)-4> [BAYR]	**0,985**	**0,005**	**1,880929E+02**	**0,000**
(Experience)-5> [MRYR]	**1,000**	**0,000**	**8,033911E+03**	**0,000**
(Experience)-6> [D2YR]	**0,988**	**0,004**	**2,397669E+02**	**0,000**
(Experience)-7> [D3YR]	**0,977**	**0,008**	**1,200892E+02**	**0,000**
(Teaching)-8> [ACTRDT]	**0,814**	**0,082**	**9,930843E+00**	**0,000**
(Teaching)-9> [ACTTCH]	**1,000**	**0,000**	**1,475717E+09**	**0,000**
(Teaching)-10> [EMED]	**0,930**	**0,033**	**2,829592E+01**	**0,000**
(Teaching)-11> [WATEA]	**0,951**	**0,023**	**4,060509E+01**	**0,000**
(Business)-12> [FPTIND]	0,013	0,271	4,958209E-02	0,960
(Business)-13> [NRFAM]	0,085	0,270	3,138192E-01	0,754
(Business)-14> [NROCNA]	-0,193	0,263	--7,332709e-01	0,463
(Business)-15> [WAACC]	**0,813**	**0,275**	**2,961953E+00**	**0,003**
(Business)-16> [WASVC]	0,169	0,265	6,393056E-01	0,523
(Business)-17> [WASALE]	**0,806**	**0,273**	**2,948677E+00**	**0,003**
(Business)-18> [NEWBUS]	0,272	0,255	1,066804E+00	0,286
(DELTA1)-19 (DELTA1)	**0,040**	**0,014**	**2,915270E+00**	**0,004**
(DELTA2)-20 (DELTA2)	0,000	0,000	2,041707E-01	0,838
(DELTA3)-21 (DELTA3)	**0,040**	**0,014**	**2,915235E+00**	**0,004**
(DELTA4)-22 (DELTA4)	**0,030**	**0,010**	**2,913469E+00**	**0,004**
(DELTA5)-23 (DELTA5)	0,000	0,000	3,872706E-01	0,699
(DELTA6)-24 (DELTA6)	**0,024**	**0,008**	**2,912069E+00**	**0,004**
(DELTA7)-25 (DELTA7)	**0,046**	**0,016**	**2,916457E+00**	**0,004**
(DELTA8)-26 (DELTA8)	**0,338**	**0,133**	**2,533464E+00**	**0,011**
(DELTA9)-27 (DELTA9)	**0,000**	**0,000**		
(DELTA10)-28 (DELTA10)	**0,135**	**0,061**	**2,217215E+00**	**0,027**
(DELTA11)-29	**0,097**	**0,04**	**2,168875E**	**0**

		5	+00	,030
(DELTA12)-30 (DELTA12)	**1,000**	**0,00 7**	**1,368791E +02**	**0 ,000**
(DELTA13)-31 (DELTA13)	**0,993**	**0,04 6**	**2,173207E +01**	**0 ,000**
(DELTA14)-32 (DELTA14)	**0,963**	**0,10 1**	**9,498401E +00**	**0 ,000**
(DELTA15)-33 (DELTA15)	0,338	0,44 7	7,571400E -01	0 ,449
(DELTA16)-34 (DELTA16)	**0,971**	**0,09 0**	**1,082963E +01**	**0 ,000**
(DELTA17)-35 (DELTA17)	0,351	0,44 0	7,979269E -01	0 ,425
(DELTA18)-36 (DELTA18)	**0,926**	**0,13 8**	**6,699364E +00**	**0 ,000**

For the 3 factor model the results are described in Table 9 (the significant facts are in bold type), and the normal probability plot of residuals is given in Fig.4. From Table 9 it is obvious that the ways for the factors "Experience" and "The Attitude to education and science" are significant as they have high value of T-statistics and low probability. Hence, it is possible to conclude that the considered model precisely describes the set forth above factors. However, we can observe that the factor "Business characteristics" is not so well described by the given model, not all ways for this factor have high T-statistics and low probability. For additional verification of the importance of the model the normal probability plot of residuals (Fig.4) has been constructed. From the plot it is obvious that the residuals of model are situated close to the straight line of normal distribution that confirms the importance of the constructed model. On the whole we can conclude that the results received, using the diagram of ways for the 3 factor model, coincide with the results of the factor analysis for this model which proves their correctness.

Table 10 The confirming factor analysis using structural equations for the 2 factor model

Model Estimates

	Param eter Estima te	Stan dard Erro r	T Statistic	P rob. L evel
(Experience)-1> [AGE]	**-1,000**	**0,00 0**	**- 2,083354E+05**	**0 ,000**
(Experience)-2> [DGRYR]	**0,980**	**0,01 0**	**1,020890E +02**	**0 ,000**
(Experience)-3> [HSYR]	**1,000**	**0,00 0**	**3,205600E +12**	**0 ,000**
(Experience)-4> [FPTIND]	0,254	0,22 7	1,117756E +00	0 ,264
(Experience)-5> [NROCNA]	0,136	0,23 8	5,715618E -01	0 ,568
(Experience)-6> [BAYR]	**0,997**	**0,00 2**	**6,361259E +02**	**0 ,000**
(Experience)-7> [MRYR]	**0,980**	**0,01 0**	**1,000551E +02**	**0 ,000**
(Experience)-8> [D2YR]	**0,995**	**0,00 3**	**3,952639E +02**	**0 ,000**
(Experience)-9> [D3YR]	**0,997**	**0,00 1**	**6,970720E +02**	**0 ,000**
(Business)-10> [ACTRDT]	**0,814**	**0,08 2**	**9,930843E +00**	**0 ,000**
(Business)-11> [ACTTCH]	**1,000**	**0,00 0**	**3,545235E +02**	**0 ,000**
(Business)-12> [EMED]	**0,930**	**0,03 3**	**2,829592E +01**	**0 ,000**
(Business)-13> [NRFAM]	0,027	0,24 2	1,124848E -01	0 ,910
(Business)-14>	**-0,588**	**0,15**	**-**	**0**

[WAACC]		**9**	**3,700352E+00**	**,000**
(Business)-15> [WASVC]	-0,164	0,236	-- 6,967788e-01	0 ,486
(Business)-16> [WASALE]	**-0,530**	**0,174**	**- 3,040265E+00**	**0 ,002**
(Business)-17> [WATEA]	**0,951**	**0,023**	**4,060509E +01**	**0 ,000**
(Business)-18> [NEWBUS]	-0,346	0,214	- 1,617683E+00	0 ,106
(DELTA1)-19 (DELTA1)	**0,000**	**0,000**	**2,061573E +00**	**0 ,039**
(DELTA2)-20 (DELTA2)	**0,040**	**0,019**	**2,103603E +00**	**0 ,035**
(DELTA3)-21 (DELTA3)	**0,000**	**0,000**		
(DELTA4)-22 (DELTA4)	**0,936**	**0,115**	**8,127443E +00**	**0 ,000**
(DELTA5)-23 (DELTA5)	**0,981**	**0,065**	**1,515202E +01**	**0 ,000**
(DELTA6)-24 (DELTA6)	**0,006**	**0,003**	**2,068245E +00**	**0 ,039**
(DELTA7)-25 (DELTA7)	**0,040**	**0,019**	**2,104467E +00**	**0 ,035**
(DELTA8)-26 (DELTA8)	**0,010**	**0,005**	**2,072333E +00**	**0 ,038**
(DELTA9)-27 (DELTA9)	**0,006**	**0,003**	**2,067659E +00**	**0 ,039**
(DELTA10)-28 (DELTA10)	**0,338**	**0,133**	**2,533464E +00**	**0 ,011**
(DELTA11)-29 (DELTA11)	**0,000**	**0,000**		
(DELTA12)-30 (DELTA12)	**0,135**	**0,061**	**2,217215E +00**	**0 ,027**
(DELTA13)-31 (DELTA13)	**0,999**	**0,013**	**7,562196E +01**	**0 ,000**
(DELTA14)-32 (DELTA14)	**0,655**	**0,187**	**3,508443E +00**	**0 ,000**
(DELTA15)-33 (DELTA15)	**0,973**	**0,078**	**1,253796E +01**	**0 ,000**
(DELTA16)-34 (DELTA16)	**0,719**	**0,185**	**3,888715E +00**	**0 ,000**
(DELTA17)-35 (DELTA17)	**0,097**	**0,045**	**2,168875E +00**	**0 ,030**
(DELTA18)-36 (DELTA18)	**0,881**	**0,148**	**5,966712E +00**	**0 ,000**

Let's now consider the 2 factor model. The parameters of the 2 factor model are given in Table 10 and the normal probability plot of residuals is given in Fig.5. As well as in the factor analysis, the factor "Experience and environment conditions" is well enough described by the diagram of ways, practically all the ways have high T-statistics and low probability. The exceptions are only the ways for the parameters "Full-time/part-time status including all jobs during the reference week" and "Reason for working outside the highest degree field: a desired job is not available". The second factor is not so well described by the given model. The ways for the parameters "Reason for working outside the highest degree field: family-related reasons", "Work activities in the principal job: professional services" and "New business" have low T-statistics and high probability. The normal probability plot of residuals for this model is practically similar to the normal probability plot of residuals for the 3 factor model - this proves that the model conforms to real data. On the whole, the model confirms the results of the factor analysis.

Considering all the results described above we can draw the conclusion that the received structure of correlation between the parameters and the factors which include them coincides with the structure of correlation of the factor analysis for the 3 factor and the 2 factor models. Hence,

the results of the confirming factor analysis using structural equations, on the whole, confirm the results of the factor analysis.
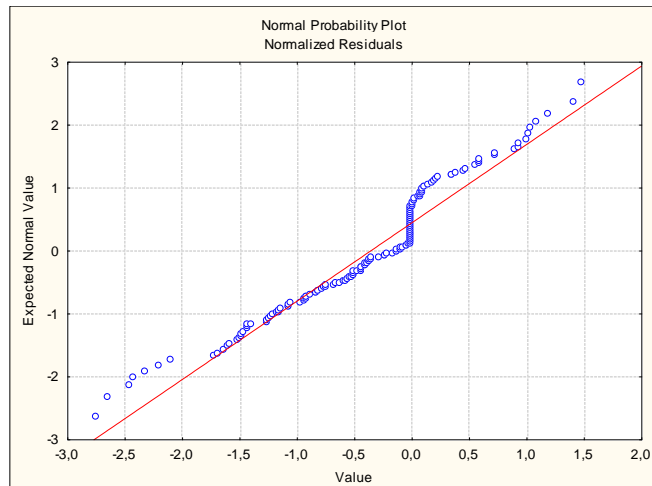

Fig. 4 The normal probability plot of residuals for the 3 factor model
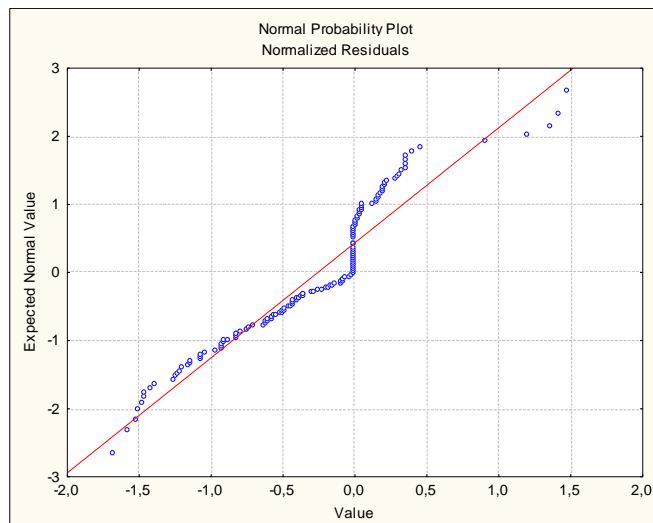

Fig. 5 The normal probability plot of residuals for the 2 factor model

Further on, on the basis of the received factors the regression models showing dependence of probability of self-employment on given factors will be constructed. For it various methods are to be used to find out the best regression model for binary responses.

References

[1] Brown M.B., and Forsythe A.B(1974). "Robust Tests for the Equality of Variances". *Journal of the American Statistical Association*, No.69, pp. 364-367.

[2] Dallal G.E. and Wilkinson L(1986). "An analytic approximation to the distribution of Lilliefor's test statistic for normality". *The American Statistician*, Vol.40, No.4, pp. 294-296

[3] De Leeuw J. and Van Rijckevorsel J(1980). *HOMALS and PRINCALS - Some generalizations of principal components analysis.* In: Data Analysis and Informatics, E. Diday et al, eds. Amsterdam: North-Holland.

[4] Dziuban C.D. and Shirkey E.C(1974). "When is a correlation matrix appropriate for factor analysis? ", *Psychological Bulletin*, No. 81, pp. 358-361.

[5] *Encyclopedia of Statistical Sciences* 4. NY: Wiley, pp. 608-610.

[6] Frigge M., Hoaglin, D.C., and Iglewicz B(1987). "Some implementations for the boxplot". *In: Computer Science and Statistics Proceedings of the 19th Symposium on the Interface*, R. M. Heiberger and M. Martin, eds. Alexandria, Va.: American Statistical Association.

[7] Glaser, R. E(1983). *Levene's Robust Test of Homogeneity of Variances*.

[8] Harman H.H(1976). *Modern factor analysis*, 3rd ed., Chicago: University of Chicago Press.

[9] Hendrickson A.E. and and White P.O (1964). "Promax: A quick method for rotation to oblique simple structure". *British Journal of Statistical Psychology*, No. 17, pp. 65-70.

[10] Hoaglin D.C., Mosteller, F., and Tukey J.W. (1983). *Understanding robust and exploratory data analysis*. New York: John Wiley & Sons, Inc.

[11] Hoaglin D.C., Mosteller, F., and Tukey J.W.(1985). *Exploring data tables, trends, and shapes*. New York: John Wiley & Sons, Inc.

[12] Hoyle, R.H. (Ed.). (1995). *Structural Equation Modeling. Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage.

[13] Iзaлls A. (1987). *Eigenvalue techniques for qualitative data*. Leiden: DSWO Press.

[14] Jennrich R.I. and Sampson P.F. (1966). *Rotation for simple loading*. Psychometrika, Vol.31, No.3,pp. 313-323.

[15] Joreskog K. G. (1977). Factor analysis by least-square and maximum likelihood methods. In: *Statistical Methods for Digital Computers*, volume 3, K. Enslein, A. Ralston, and R.S. Wilf, eds. New York: John Wiley & Sons, Inc.

[16] Kaiser H.F. (1963). *Image analysis*. In: Problems in Measuring Change, C.W.Harris, ed. Madison, Wisc.: University of Wisconsin Press.

[17] Lilliefors H.W. (1967). "On the Kolmogorov-Smirnov tests for normality with mean and variance unknown". *Journal of the American Statistical Association*, No.62, pp. 399-402.

[18] Loh W.Y. (1987). "Some Modifications of Levene's Test of Variance Homogeneity". *Journal of Statistical Computation and Simulation*, No.28, pp. 213-226.

[19] Rao C.R. (1955). "Estimation and test of significance in factor analysis". *Psychometrika*, Vol.20, No.2, pp. 93-111.

[20] Rummel R.J. (1970). *Applied factor analysis*. Evanston: Ill.: Northwestern University Press.

[21] Theil H. (1953b). *Estimation and simultaneous correlation in complete equation systems*. The Hague: Central Planning Bureau.

[22] Tukey J.W. (1977). *Exploratory data analysis*. Reading, Mass.: Addison-Wesley.

[23] Velleman P.F. and Hoaglin D.C. (1981). *Applications, basics, and computing of exploratory data analysis*. Boston: Duxbury Press.

[24] Wilkinson J H. (1965). *The algebraic eigenvalue problem*. Oxford: Clarendon Press.

[25] Yalyalieva, T.V. and Murzina, E.A. (2015) , "The system of parameters efficiency of financial supervision". *Advances in Systems Science and Application* Vol.15 No.4, pp.384-391.

Corresponding author
Elena A. Murzina can be contacted at: elena.murzina@gmail.com.