Proceedings of the 7th WSEAS International Conference on Mathematics & Computers in Business & Economics, Cavtat, Croatia, June 13-15, 2006 (pp7-12)

Factor Analysis as the Instrument in Relaxing the Assumptions of the Classical Model

JOSIP ARNERIĆ, ELZA JURUN & SNJEŽANA PIVAC
Department of Quantitative Methods
University of Split, Faculty of Economics
Matice Hrvatske 31, 21000 Split
CROATIA

Abstract: This work builds up a complete procedure of using factor analysis as the instrument in the case of relaxing the assumption of the classical model. Paper is focused on the situation when the multicolinearity appears as the dominant problem. This problem is solved by grouping of performance indicators, not only by technical principles, but also according to fundamental postulates of business economic theory. The whole procedure is illustrated by a practical example. The example originates from the real need to analyse and compare the performance of all manufacturing enterprises in the Split-Dalmatian County in 2004. The data set consists of a wide range of performance indicators for 1744 manufacturing enterprises, among which twelve are selected as representative ones. As the entire basic set is the issue of our interest, we find that the enterprises are markedly heterogeneous in terms of the chosen indicators. Therefore previous to comparison they have to be made homogeneous. After such homogenization, using principal components method four factors have been extracted, i.e. all selected variables (performance indicators) have been meaningfully grouped in to factors: activity, liquidity, leverage, economic efficiency. The essential part of analysis is establishing of direct, indirect and overall effects of each independent variable on return on equity as chosen dependent variable.

Key-Words: factor analysis, orthogonal rotation, principal component method, multiple regression analysis, stepwise selection, multicolinearity, direct and indirect effects on dependent variable

1 Introduction

Central model of the paper is created in response to a real needs for performance analysis and comparison of all productive enterprises in Split-Dalmatian County. In 2004 there have been 1744 manufacturing enterprises in that area.

Knowing that the general problem of enterprises in transition is unfavourable structure of capital and liabilities as a source of financing the assets, it was necessary to homogenizated the entire set of manufacturing enterprises.

Starting modelling for this purpose all enterprises with:

- zero employees,
- zero equity and
- net profit zero and less than zero,

have been excluded from further analysis. So, after that homogenization the modelling has been continued with 405 manufacturing enterprises. This is supported by the economic theory as well as by practical experience of the countries in transition.

Among wide range of performance indicators twelve of them have been extracted as inevitable on different levels of decision making.

2 Stepwise variables selection

Subset of chosen performance indicators has been taken among items from the balance sheet, which is legally defined:

- X1 total asset turnover,
- X2 current asset turnover,
- X3 fixed asset turnover,
- X4 revenue per employee,
- X5 average daily revenue,
- X6 current liquidity,
- X7 fixed asset to long term liabilities,
- X8 equity to total asset,
- X9 revenues over expenses,
- X10 expenses per employee,
- X11 earnings per employee,
- X12 equity per employee.

As the most representative indicator of profitability - return on equity is defined as dependent variable of the model:

$$y_i = \beta_0 + \sum_{j=1}^{12} \beta_j \cdot x_{ij} + e_i,\ i = 1,2,...,405 . \qquad (1)$$

In determining direct relative effect of each independent variable on return on equity, multiple regression model with all variables is used.

Estimation output looks like it follows:

Table 1.
Multiple Regression Model Estimation

Model Summary [f]

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	,882[a]	,779	,778	,443223	,779	1416,711	1	403	,000	
2	,901[b]	,813	,812	,408166	,034	73,200	1	402	,000	
3	,924[c]	,853	,852	,361432	,041	111,680	1	401	,000	
4	,931[d]	,867	,866	,344479	,014	41,439	1	400	,000	
5	,933[e]	,871	,869	,340393	,003	10,661	1	399	,001	1,695

a. Predictors: (Constant), X5
b. Predictors: (Constant), X5, X9
c. Predictors: (Constant), X5, X9, X8
d. Predictors: (Constant), X5, X9, X8, X1
e. Predictors: (Constant), X5, X9, X8, X1, X12
f. Dependent Variable: Y

Source: According to FINA data base

Table 2.
Parameter Estimation by Stepwise Selection

Coefficients [a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	,242	,022		10,895	,000
	X5	1,55E-006	,000	,882	37,639	,000
2	(Constant)	-1,438	,197		-7,284	,000
	X5	1,46E-006	,000	,828	36,808	,000
	X9	1,566	,183	,192	8,556	,000
3	(Constant)	-1,776	,178		-9,994	,000
	X5	1,41E-006	,000	,803	40,028	,000
	X9	2,138	,171	,263	12,510	,000
	X8	-,870	,082	-,213	-10,568	,000
4	(Constant)	-2,025	,174		-11,655	,000
	X5	1,41E-006	,000	,800	41,847	,000
	X9	2,195	,163	,270	13,456	,000
	X8	-,860	,079	-,211	-10,958	,000
	X1	,121	,019	,118	6,437	,000
5	(Constant)	-1,994	,172		-11,602	,000
	X5	1,42E-006	,000	,806	42,470	,000
	X9	2,177	,161	,268	13,498	,000
	X8	-,773	,082	-,189	-9,408	,000
	X1	,108	,019	,105	5,716	,000
	X12	-1,58E-007	,000	-,064	-3,265	,001

a. Dependent Variable: Y

Source: According to FINA data base

It is evident that all variables are not statistically significant using stepwise method. Namely, only five variables (Table 2.) are left in the model satisfying condition for p value less than 0.05 to be entered and p value not greater than 0.10 to be removed from the equation. Also from correlation matrix (Table 3.) among all performance indicators it can bee seen that multicolinearity problem exists. By testing bivariate correlation coefficient, using one tailed test, for almost 30 correlation coefficients empirical significance is less than 0.05. Moreover Farrar-Glauber test has confirmed that multicolinearity appears as serious problem.

According that, determinant of correlation matrix is near singular (close to zero), which is evidence that it can not be accepted hypothesis that correlation matrix is identity matrix. In such cases, most appropriate statistical – mathematical procedure, for solving this problem, is to reduce many observed variables in less number of underlying variables which are called factors. Each factor represents linear combination of variables with similar characteristics by factor loadings or standardized weights. It means that factor analysis enables for "removed" independent variables to be indirectly regressed on dependent variable trough factors.

3 Factor Analysis

Factor analysis is used to find underlying variables or factors among observed variables. In other words, if multicolinearity exist among these variables, factor analysis can be used to solve it. In such way all direct and indirect effects of each independent variable on regresand variable can be measured. The procedure is taken up throught three stages:

- examination if correlation matrix can be factorized and if there exist high degree of common variance that can be explained,
- extraction of optimal factors (components) as linear combination of observed variables,
- orthogonal rotation of factors in order to maximize the relationship between the variables.

Basis of factor analysis is variance-covariance matrix of independent observed variables, i.e. it is assumed that variance between each two variables can be decomposed.

Table 3.

Correlation Matrix [a]

		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Correlation	X1	1,000	,602	-,007	,018	,007	-,085	-,012	-,039	-,059	,022	-,042	-,202
	X2	,602	1,000	-,035	-,008	,140	-,203	,126	,061	,027	-,015	,078	-,106
	X3	-,007	-,035	1,000	,007	-,005	,006	-,036	-,044	-,018	,008	-,011	-,013
	X4	,018	-,008	,007	1,000	,107	,034	-,016	-,048	-,008	,997	,515	,557
	X5	,007	,140	-,005	,107	1,000	-,015	,022	-,025	,282	,075	,447	,070
	X6	-,085	-,203	,006	,034	-,015	1,000	-,253	,457	,261	,004	,314	,209
	X7	-,012	,126	-,036	-,016	,022	-,253	1,000	-,346	-,133	-,007	-,095	-,077
	X8	-,039	,061	-,044	-,048	-,025	,457	-,346	1,000	,297	-,070	,211	,332
	X9	-,059	,027	-,018	-,008	,282	,261	-,133	,297	1,000	-,068	,643	,102
	X10	,022	-,015	,008	,997	,075	,004	-,007	-,070	-,068	1,000	,445	,547
	X11	-,042	,078	-,011	,515	,447	,314	-,095	,211	,643	,445	1,000	,415
	X12	-,202	-,106	-,013	,557	,070	,209	-,077	,332	,102	,547	,415	1,000
Sig. (1-tailed)	X1		,000	,443	,357	,447	,043	,403	,216	,117	,327	,200	,000
	X2	,000		,241	,433	,002	,000	,006	,109	,292	,382	,059	,016
	X3	,443	,241		,447	,462	,454	,232	,190	,360	,437	,414	,400
	X4	,357	,433	,447		,016	,249	,371	,166	,434	,000	,000	,000
	X5	,447	,002	,462	,016		,381	,328	,310	,000	,066	,000	,079
	X6	,043	,000	,454	,249	,381		,000	,000	,000	,468	,000	,000
	X7	,403	,006	,232	,371	,328	,000		,000	,004	,441	,028	,061
	X8	,216	,109	,190	,166	,310	,000	,000		,000	,079	,000	,000
	X9	,117	,292	,360	,434	,000	,000	,004	,000		,086	,000	,020
	X10	,327	,382	,437	,000	,066	,468	,441	,079	,086		,000	,000
	X11	,200	,059	,414	,000	,000	,000	,028	,000	,000	,000		,000
	X12	,000	,016	,400	,000	,079	,000	,061	,000	,020	,000	,000	

a. Determinant = 2,30E-006

Source: According to FINA data base

Variance of independent variables can be divided into common variance (communality), which explains their intercorrelation, and specific variance, which can not be explained. Unexplained variance usually includes error variance caused by measurement error.

For testing if correlation matrix can be factorized usually is used Bartletts test, examining determinant of its matrix (Table 4.). Hypotheses in this case are set up as:

$$H_0 : ... R = I$$

$$H_1 : ... R \neq I$$

where: R is correlation matrix and

I is identity matrix.

It can be accepted alternative hypothesis if empirical significance level is less than 0.05, i.e. correlation matrix is not identity matrix. Therefore, it can be factorized.

Table 4.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,568
Bartlett's Test of Sphericity	Approx. Chi-Square	5182,870
	df	66
	Sig.	,000

Source: According to FINA data base

Kaiser-Meyer-Olkin indicator above 0.5 is satisfactory, i.e. there exist high degree of common variations between variables that can be explained.

In this paper correlation matrix will be factorized using principal component method.

The main question is which variables will be grouped into which factors?

It is assumed that the 12 observed variables (the X_i) that have been measured for each of the k subjects (enterprises) have been standardized and represented in following form:

$$\begin{aligned} X_1 &= a_{11}F_1 + ... + a_{1m}F_m + e_1 \\ X_2 &= a_{21}F_1 + ... + a_{2m}F_m + e_2 \\ &... \\ X_{12} &= a_{121}F_1 + ... + a_{12m}F_m + e_{12} \end{aligned} \quad (3)$$

The F_j are the m common factors, the e_i are the 12 specific errors, and the a_{ij} are the $12 \times m$ factor loadings. The F_j have mean zero and standard deviation one, and are generally assumed to be independent. The e_i are also independent and the F_j and e_i are mutually independent of each other.

Table 5.

Eigenvalues Values of Correlation Matrix and Total Variance Explained

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,037	25,312	25,312	3,037	25,312	25,312	2,645	22,039	22,039
2	2,027	16,892	42,204	2,027	16,892	42,204	1,944	16,196	38,235
3	1,723	14,361	56,565	1,723	14,361	56,565	1,852	15,435	53,670
4	1,305	10,874	67,438	1,305	10,874	67,438	1,652	13,769	67,438
5	1,013	8,446	75,884						
6	,802	6,683	82,567						
7	,690	5,751	88,318						
8	,611	5,095	93,413						
9	,337	2,812	96,225						
10	,300	2,500	98,725						
11	,153	1,274	100,000						
12	3,12E-005	,000	100,000						

Extraction Method: Principal Component Analysis.

Source: According to correlation matrix in Table 3.

In matrix form system of equations in expression (3), for $m < k$, can be written as:

$$X_{12\times 1} = A_{12\times m}F_{m\times 1} + e_{12\times 1}, \tag{4}$$

which is equivalent to:

$$R = AA^T + \text{cov}(e), \tag{5}$$

where $R_{12\times 12}$ is correlation matrix of $X_{12\times 1}$. Since the errors are assumed to be independent, $\text{cov}(e)$ should be a 12×12 diagonal matrix. This implies that:

$$Var(X_i) = \sum_{j=1}^{m} a_{ij}^2 + Var(e_i), \quad \forall i. \tag{6}$$

The sum of X_i's squared factor loadings is called its communality (the variance it has in common with the other variables through the common factors). The i^{th} error variance is called the specificity of X_i (the variance that is specific to variable i).

Now factors can be extracted from correlation matrix by solving characteristic equation as follows:

$$det(R - \Lambda \cdot I) = 0 \tag{7}$$

By solving above equation we gets eigenvalues of correlation matrix $\lambda_i \quad i = 1,2,...,k$, where k is number of variables.

Each eigenvalue shows part of variance that can be explained by each factor. So, usually each eigenvalue is expressed relatively on the number of variables, as:

$$\frac{\lambda_i}{trR} \quad i = 1,2,...,k \tag{8}$$

It can be seen that trace of correlation matrix equals k variables, because all diagonal elements of correlation matrix are ones.

By Keiser criteria it is necessary to extract only factors with eigenvalues greater than one, which cumulative explains more than 60% of total variance. In our case optimal number factors to extract is four factors (Table 5.).

After estimation of factor loadings it is necessary to examine their significance. In empirical research usually factor loadings (given in standardized units) above ± 0.04 are statistically significant, because they explain more than 16% of variance. At the end it is used Kaiser Varimax method of orthogonal rotation factor axis to get more meaningful grouping of variables and to ensure independence between factors (Table 6.).

Table 6.

Rotated Component Matrix [a]

	Component 1	2	3	4
X4	,974			
X10	,973			
X12	,706			
X8		,829		
X6		,717		
X7		-,646		
X9			,791	
X11	,484		,768	
X5			,750	
X1				,887
X2				,887
X3				,396

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Source: According to FINA data base

From Table 6. it is evident that variables X4, X10 and X12 are grouped into factor 1. All this variables belong to the same category of economic indicators - **economic efficiency indicators**.

Variables X6, X7 and X8 are grouped into factor 2 - category of **liquidity indicators**.

Variables X5, X9 and X11 are grouped into factor 3 - category of **leverage indicators**.

Variables X1, X2 and X3 are grouped into factor 4 - category of **activity indicators**.

4 Direct and Indirect Effects Estimation

After calculating factor scores for each linear combination, the same are used in regression with return on equity as dependent variable:

$$\hat{y}_i = \phi_1 \cdot F_1 + \phi_2 \cdot F_2 + \phi_3 \cdot F_3 + \phi_4 \cdot F_4, \qquad (9)$$

where in this model:

$\hat{y}_i$ is expected standardized value of return on equity (dependent variable),

ϕ_i is estimated parameter for factor i and

F_i is adequate factor score.

Table 7.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,841[a]	,708	,705	,54324404	1,767

a. Predictors: (Constant), FA4, FA3, FA2, FA1

b. Dependent Variable: Z

Source: According to FINA data base

Table 8.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	285,954	4	71,489	242,241	,000[a]
	Residual	118,046	400	,295		
	Total	404,000	404			

a. Predictors: (Constant), FA4, FA3, FA2, FA1

b. Dependent Variable: Z

Source: According to FINA data base

Table 9.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
	FA1	-,034	,027	-,034	-2,645	,045
	FA2	-,267	,027	-,267	-9,868	,000
	FA3	,779	,027	,779	28,832	,000
	FA4	,168	,027	,168	6,222	,000

a. Dependent Variable: Z

Source: According to FINA data base

In Tables 7., 8. and 9. complete multiple regression diagnostics with incorporated factor analysis results is shown. It is evident that in this model all economic-theoretical, econometrics and statistical criteria are significant. Especially, by meaningful grouping of variables the problem of multicolinearity is solved.

Simultaneously, estimated parameters are remained consistent. Even without testing it is obvious from factor correlation matrix (Table 10.), that multicolinearity disappears.

Table 10.

Coefficient Correlations [a]

Model			FA4	FA3	FA2	FA1
1	Correlations	FA4	1,000	,000	,000	,000
		FA3	,000	1,000	,000	,000
		FA2	,000	,000	1,000	,000
		FA1	,000	,000	,000	1,000

a. Dependent Variable: Z

Source: According to FINA data base

5 Conclusion Remarks

Indirect effects are calculated by appropriate factor parameters ϕ_i and adequate factor loading from rotated component matrix. It is obvious that indirect effects are not negligible (as X11 - earnings per employee in the Table 11.). Exactly this proves that any variable must not be removed from the model, as multicolinearity factor, because its indirect effects on the dependent variable can be very significant. Result of using standard statistical-econometric methods (as stepwise technique is in this example) is excluding a numerous variables with significant influence on dependent variable. Even more, multicolinearity becomes a barrier for specification of any influence of "removed" variables.

Indirect effects must not be ignored, because of their significant total impact. From the Table 11. it is obvious that the total effect can contain the higher proportion of the indirect than of the direct effects.

Table 11.

Total Effects Estimation

Indicators	Direct Effect	Indirect Effect	Total Effect	Rank
X5	0,806	0,584	1,390	**1**
X9	0,268	0,616	0,884	**2**
X8	-0,189	-0,221	-0,410	**4**
X1	0,105	0,149	0,254	**5**
X12	-0,064	0,119	0,055	**10**
X10	0	-0,033	-0,033	**12**
X4	0	-0,033	-0,033	**11**
X6	0	-0,191	-0,191	**6**
X7	0	0,172	0,172	**7**
X11	0	0,598	0,598	**3**
X2	0	0,149	0,149	**8**
X3	0	0,067	0,067	**9**

Source: According to FINA data base

This is especially relevant in the cases of analyzing total effects of the highest ranking variables, such as in this case X9, X1 and X12.

Furthermore, the additional advantage of factor analysis is the fact that all information are included in research through only a few factors.

This paper reveals how relaxing the assumptions of the classical model can be solved by meaningful grouping of variables using factor analysis.

Even more factor analysis helps to specify all kinds of effects of each explanatory variable on dependent one, which was basic aim of this paper.

References:

[1] Anderson, R. E.; Black, W.; Hair, J. F.; Tatham, R. L., *Multivariate Data Analysis,* Prentice Hall, 1998

[2] Anderson, T. W., *An Introduction to Multivariate Statistical Analysis (Wiley Series in Probability and Statistics),* Wiley-Interscience, 2003

[3] Arnerić, J., Jurun, E., Comparison and Ranking Model of Performance Indicators, *Annals of DAAAM for 2004,* Vienna, 15th International Symposium, 2004, pp. 15-16.

[4] Arnerić. J., Jurun. E., Positioning Enterprises by Aggregate Performance Indicator, *Enterprise in Transition,* Split-Bol, 6th International Conference, 2005, pp. 26-28.

[5] Arnold, J. A., Hope, A. J. B., *Accounting for Management Decisions,* Prentice Hall, 1990

[6] Basilevsky, A. T., *Statistical Factor Analysis and Related Methods: Theory and Applications,* John Wiley an Sons Inc., 1994

[7] Belak, V., *Menadžersko računovodstvo*, Računovodstvo, revizija i financije, 1995

[8] Brooks, C., *Introductory econometrics for finance,* University Press, 2002

[8] Chiang, A. C., *Osnovne Metode Matematičke Ekonomije,* Mate, 1994

[9] Fulgosi, A., *Faktorska analiza,* Školska knjiga Zagreb, 1988

[10] Giffi, A., *Nonlinear Multivariate Analysis,* Wiley, 1996

[11] Green, P. E.; Carroll, J. D., *Mathematical Tools for Applied Multivariate Analysis,* Wiley, 1976

[12] Halmi A., *Multivarijatna analiza u društvenim znanostima,* Alinea Zagreb, 2003

[13] Halsey, R. F.; Subramanyam, K. R.; Wild, J. J., *Financial Statement Analysis,* McGraw-Hill/Irwin, 2005

[14] Harman, H. H., *Modern factor analysis,* University of Chicago, 1976

[15] Johnson, R. A.; Wichern, D. W., *Applied Multivariate Statistical Analysis,* Prentice Hall, 2002

[16] Jolliffe, I.T., *Principal Component Analysis,* Springer, 2002

[17] Karatzas, I.; Shreve, S E., *Methods of Mathematical Finance,* Springer 2001

[18] Thompson B., *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications,* American Psychological Association, 2004

[19] Thurstone, L. L., *Multiple Factor Analysis,* University of Chicago, 1974

[20] Žager, K. & Žager, L., *Analiza financijskih izvještaja,* Masmedia Zagreb, 1999