Вестник Томского государственного университета. Философия. Социология. Политология. 2022. № 69. С. 152–160.
Tomsk State University Journal of Philosophy, Sociology and Political Science. 2022. 69. pp. 152–160.

Original article
УДК 311.2
doi: 10.17223/1998863X/69/16

SELECTION OF FACTOR EXTRACTION METHODS IN COMPLICATED RESEARCH CONTEXTS: PRACTICE RECOMMENDATIONS

Anna N. Suleymanova[1], **Irina K. Zangieva**[2]

[1, 2] *National Research University Higher School of Economics, Moscow, Russian Federation*

[1] *asuleymanova@hse.ru*

[2] *izangieva@hse.ru*

Abstract. It is a common practice among social scientists to use "factor analysis" and "principal components analysis" interchangeably, even though PCA is not a factor extraction method, but a dimension reduction technique. Most of the recent studies with factor analysis rely solely on PCA or fail to specify which factor extraction method was used. Supposedly, it is caused by the lack of structured and comprehensive guidance on the selection of factor extraction methods. The aim of this study is to develop a theoretically and empirically justified algorithm of factor extraction method selection, depending on a combination of research context features, such as (a) sample size, (b) number of indicators specifying each factor, (c) size, (d) range of communalities, (e) presence of model error and (f) distribution of indicators. Seven factor extraction methods were studied: principal component analysis, weighted and generalized least squares method, maximum likelihood method, principal axis analysis, alpha-factor analysis, and image factoring. Theoretically justified algorithm was created and tested via statistical experiment with Monte Carlo simulation. Following the general outline of previous works' experimental designs, we specified factor loadings matrices for each research context with nonzero loadings, derived correlation matrices and produced 500 Monte Carlo simulated samples (3000 samples in total) per research context. Every factor extraction method was applied to every sample and the resulting factor loadings matrices and communalities were recorded and summarized. Four criteria of factor analysis extraction adequacy were applied: squared mean errors of factor loadings, squared mean errors and absolute mean errors of communalities, and number of Heywood cases. As a result we formulated four main recommendations: it is advised to use (1) principal axis analysis or alpha-factor analysis, if a model error is suspected, (2) maximum likelihood method or generalized least squares method, if the sample is large enough and indicators are normally distributed, or vice versa, if the sample is not large enough and distribution of indicators differs from normal, (3) maximum likelihood method, if the sample is large enough, but the indicators are not normally distributed, or if the indicators are normally distributed, but the sample size is not large enough and the communalities are small, (4) generalized least squares method, if the indicators are normally distributed and the communalities are large, but the sample size is not large enough.

Keywords: factor analysis, Monte Carlo simulation, factor extraction, principal component analysis

For citation: Suleymanova, A.N. & Zangieva, I.K. (2022) Selection of factor extraction methods in complicated research contexts: practice recommendations. *Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sotsiologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science.* 69. pp. 152–160. (In Russian). doi: 10.17223/1998863X/69/16

Научная статья

ВЫБОР МЕТОДА ФАКТОРИЗАЦИИ В ЗАВИСИМОСТИ ОТ ИССЛЕДОВАТЕЛЬСКОЙ СИТУАЦИИ: ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

Анна Наильевна Сулейманова[1], Ирина Казбековна Зангиева[2]

[1, 2] *Национальный исследовательский университет «Высшая школа экономики», Москва, Россия*

[1] *asuleymanova@hse.ru*

[2] *izangieva@hse.ru*

Аннотация. Исследование посвящено сравнению применимости семи основных методов факторизации в зависимости от количества индикаторов, их распределения, модельной ошибки, объема выборки, структуры общностей. По итогам статистического эксперимента составлены следующие рекомендации: а) в случае наличия модельной ошибки использовать метод главных осей или альфа-факторный анализ; б) при отсутствии модельной ошибки выбирать между методом максимального правдоподобия и обобщенным методом наименьших квадратов.

Ключевые слова: факторный анализ, симуляция Монте-Карло, извлечение факторов, метод главных компонент

Для цитирования: Suleymanova A.N., Zangieva I.K. Selection of factor extraction methods in complicated research contexts: practice recommendations // Вестник Томского государственного университета. Философия. Социология. Политология. 2022. № 69. С. 152–160. doi: 10.17223/1998863X/69/16

Introduction

Despite the longstanding and widespread usage of factor analysis and large number of factor extraction methods, in social sciences the statement “factor analysis was applied” often means that it was, in fact, dimension reduction with principal component analysis. According to publications in the field of social sciences, indexed in the Scopus database in 2017, principal component analysis is used at least twice as often as actual factor extraction methods altogether: among 1574 articles with “factor analysis” in keywords, abstract, or title, only 133 mention principal axis, maximum likelihood or least squares methods; alpha-factor analysis and image factoring were not mentioned once; 110 articles mention principal component analysis. The rest of the articles do not specify, which method of factor extraction was applied.

At the same time, textbooks and manuals on quantitative data analysis often limit themselves to discussion of the factor extraction process (factorization), but rarely elaborate on the benefits of each method regarding the properties of data input and on how the results might differ [1; 2]. The reason for the neglect of all other factorization methods might be the lack of a comprehensive set of best practices for their selection.

These methods were repeatedly compared in different research circumstances, but, generally, the comparison usually covers two or three of them applied to a very specific type of data [3–7]. In this article we compare six factor extraction methods (unweighted least squares, or ULS; generalized least squares or GLS, maximum likelihood method, or LM; principal axis method or PAX; alpha-factor or AF; Image

Factor analysis or IF) and one dimension reduction method (Principal Component Analysis or PCA) in regard to a broad spectrum of possible, real world data types. We intend to organize existing recommendations and build a theoretically and empirically confirmed algorithm of factor extraction method selection depending on a pool of input data characteristics that were determined based on existing research and known properties of methods: (1) sample size; (2) indicators to factors ratio; (3) communalities size; (4) communalities range; (5) whether there is a significant chance of model error and (6) distribution of indicators. We call a certain combination of these characteristics a *research context.*

The main goal of the research is to develop an algorithm for the selection of factor extraction method depending on the research context. We compare methods based on their *contextual performance,* assessed by the occurrence of Heywood cases, and mean squared errors of both factor loadings and communalities produced by each method.

Related work

Factorization methods are, in their sense, mathematical instruments with a certain algorithm, which means that, in every research context, the choice of the method should be supported by theoretically justified and empirically tested recommendations. Yet the best practice for these kinds of methods so far is a segmental set of insights on the performance of each separate method.

Acito and Anderson [3. P. 228–230] compared Alpha-factor (AF), Image Factor (IF) analyses, and PCA with Monte Carlo simulations in situations, when the quality of PCA results declined: a) sample size is small, b) there are less than 6 indicators for one factor and c) communalities size differs significantly. In these situations, AF and IF were more effective, than PCA. Mislevy determined that ML is preferable, when a small number of factors is extracted on the large number of indicators. GLS, on the contrary, is preferable, when many factors are extracted on a small number of indicators. Both methods are applicable, if the numbers of both indicators and factors are small [7. P. 20–21]. Generalized least squares (GLS) is more appropriate in situations, when the sample size and communalities sizes are small, compared to unweighted least squares (ULS) and maximum likelihood (ML) [8. P. 292]. According to Fabrigar [9. P. 272–275], ML is the most appropriate method in cases when the distribution of indicators is close to normal; otherwise, they advise to analysts to use PAX. By MacCallum [10. P. 84–86] it is determined that the sample size and ratio indicators – to-factors have little effect on ML results if communalities are high. In this case, factor loadings always recovered almost perfectly. However, if some communalities are small, indicators to factors ratio and sample size start to influence the quality of recovery of factor loadings.

The findings about the properties of factor extraction methods can be schematized as follows (table 1):

Theoretically justified advice for the best selection of factor extraction methods, which we are going to verify empirically on the base of statistical experiment, are as follows:

- use PCA if the data follows every prerequisite for factor analysis,
- use ULS If (1) indicators distribution differs from normal, (2) sample size is small, (3) communalities are small, (4) there is model error and (5) there are few indicators and a lot of factors

Table 1. **Factor extraction methods features**

Factor extraction method	Is the method sensitive to the aspect of research context?					
	Distribution of indicators	Sample size	Communalities size	Communalities range	Number of indicators	Model error
Principal component analysis	Yes	Yes	Yes	Yes	Yes	Yes
Principal axis method	No	Yes	Yes	Yes	Yes	Yes
Unweighted least squares	No	No	No	No	No	No
Generalized least squares	No	Yes	Yes	No	No	Yes
Maximum likelihood	Yes	Yes	Yes, but large and consistent communalities lower the sensitiveness to sample size and number of indicators		Yes	Yes
Image factoring	Yes	No	Yes	No	No	Yes
Alpha-factor analysis	Yes	No	Yes	No	No	Yes

• use GLS if (1) indicators distribution differs from normal, (2) the range of communalities is wide and (3) there are few indicators and a lot of factors.

• use PAX if the only deviation from ideal data properties is non-normal distribution of indicators.

• use AF or IF if the only deviation from the ideal data properties is small sample size.

• use ML if the only deviations from ideal data properties are small sample size and there are few indicators and a lot of factors.

This hypothesis can be summarized and visualized in the following scheme of theoretically substantiated selection algorithm for factor extraction methods depending on research context.

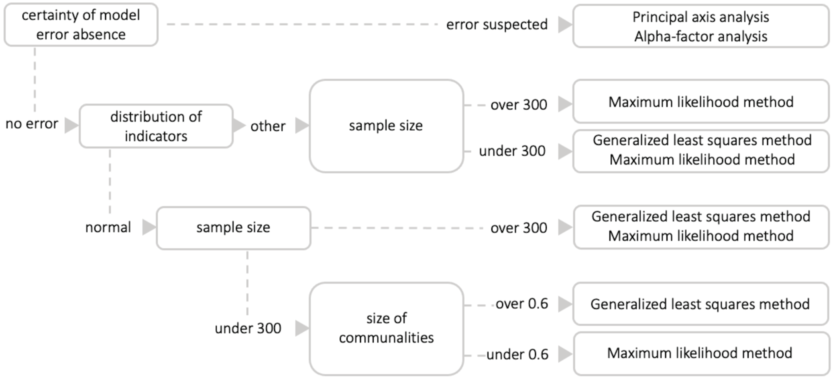
certainty of model error absence
error suspected
Principal axis analysis
Alpha-factor analysis
no error
distribution of indicators
other
sample size
over 300
Maximum likelihood method
under 300
Generalized least squares method
Maximum likelihood method
normal
sample size
over 300
Generalized least squares method
Maximum likelihood method
under 300
size of communalities
over 0.6
Generalized least squares method
under 0.6
Maximum likelihood method

Fig. 1. Theoretically substantiated selection algorithm for factor extraction methods depending on research context

Design of experiment

We determined six aspects of research context that might affect the contextual performance of factor extraction methods: (a) presence of model error, (b) indicators to factors ratio, (c) distribution of indicators, (d) sample size, (e) size of communalities, and (f) range of communalities.

According to best practices the threshold parameters for the chosen aspects of research context are: no less than 200 cases in a sample [9, 11], communalities no smaller than 0.6, communalities range no wider than 0.3, 6 indicators per factor and normal distribution of indicators is assumed [12]. We also consider model errors in the form of mild cross-loadings. Below- and above threshold parameters of the models were specified as follows (table 2):

Table 2. **Specification of research context hypothetically best for use of each factorization method**

	PCA	ULS	PAX	GLS	AF & IF	ML
Indicators to factors ratio	6/1	3/1	6/1	3/1	3/1	3/1
Communalities size	>0.6	<0.6	>0.6	>0.6	>0.6	>0.6
Communalities range	<0.3	>0.3	<0.3	>0.3	>0.3	<0.3
Model error	No	Yes	No	No	No	No
Sample size	N = 5000	N = 50	N = 5000	N = 5000	N = 50	N = 50
Distribution of indicators	N (5; 1.73)	Bi (10; 0.3)	Bi (10; 0.3)	Bi (10; 0.3)	N (5; 1.73)	N (5; 1.73)

Following De Winter [12] general outline, we specified factor loadings matrices for each research context with nonzero loadings, derived correlation matrices and produced 500 Monte Carlo simulated samples (3000 samples in total) for each research context, according to the summary in *table 2*. Simulations were conducted via R library.

Deciding on form and parameters of distributions, we considered the data from 8th wave of European Social Survey that could be factorized: pseudo-interval attitudinal variables with 11 gradations that address the topics of trust to social institutions, satisfaction with social institutions and life, attitudes towards minorities. Distributions were analyzed across countries. For every variable by every country Kolmogorov-Smirnov normality test was conducted, proving that none of the variables were normally distributed. Consequently, variables were divided into two groups: those with distribution close to normal and those with distribution significantly unlike normal. For the former sample means and standard deviations were computed and averaged. These values ($\mu = 5$ и $\sigma = 1.73$) were used as parameters of normal distribution for respective research contexts. For the latter, the most resembling to real variables distribution was picked from the available in Simulation module, which turned out to be binomial distribution with parameters $p = 0.3$ and $n = 11$.

To pick out a seed for every simulated sample we used a true random sets generator, based on atmospheric noise, and recorded every seed for the experiment to be repeatable.

Every factor extraction method was applied to every sample and the resulting factor loadings matrices and communalities were recorded and summarized. The following criteria for comparison of factor extraction methods adequacy, common for similar research, were applied:

- Squared mean errors of factor loadings [10. P. 92–93; 12; 13. P. 35]. This criterion shows how much sample factor loadings differ from model ones and varies from 0 to 100%. If the error is 0%, it means that the sample matrix is identical to the model. These estimations were averaged for every method in every research context. According to this criterion, the most contextually appropriate method is the one that produced the smallest average squared mean error in the given research context.
- Squared mean errors [8. P. 287] and absolute mean errors of communalities. The former is analogous to squared mean error of factor loadings, the latter, however, serves the purpose of assessing the inclination of the methods to over- or underestimate the communalities.

• Number of occurred Heywood cases, when the communality is greater than 1 [8, 12].

We compared the contextual performance of methods by the following rule: in each research context the best method is the one with the least mean square error for both factor loadings and communalities. We consider the difference in errors significant if the remainder is greater than 5%. If two or more methods have similar performance in given research situation, we consider the one without inclination to over- or underestimate the communalities the best.

Experimental results

Based on the experimental results we specified the theoretically justified algorithm to four basic rules (Fig. 2).

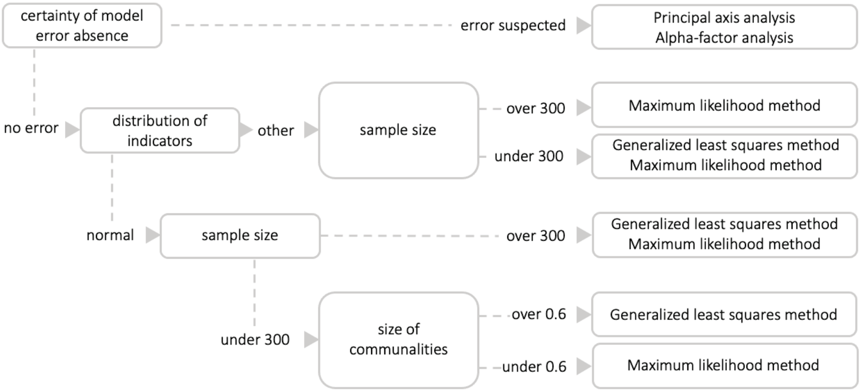
certainty of model error absence
error suspected
Principal axis analysis
Alpha-factor analysis
no error
distribution of indicators
other
sample size
over 300
Maximum likelihood method
under 300
Generalized least squares method
Maximum likelihood method
normal
sample size
over 300
Generalized least squares method
Maximum likelihood method
under 300
size of communalities
over 0.6
Generalized least squares method
under 0.6
Maximum likelihood method

Fig. 2. Algorithm of selection of the most contextually appropriate factor extraction method

Model error suspicion divides all methods into two groups: if there is a significant chance of model error, it is advised to use PAX or AF; if one is sure that there are no substantial cross-, it is advised to use ML or GLS. In case of the latter one needs to consider three more aspects of research context: distribution of indicators, sample size and communalities size.

Three of the seven examined methods (ULS, PCA and IF) did not turn out to be the most contextually appropriate for corresponding research contexts. This result is substantial in terms of PCA, as the most employable factor extraction method in sociological research (especially considering the fact that PCA is not a factor extraction method, but a method for dimension reduction), not being the most contextually appropriate in PCA context, which is a combination of rules of thumb for input data for factor analysis: no less than 6 indicators for every factor, normal distribution of indicators, sample size no smaller than 300 cases, lack of model error, communalities size no less than 0.6 and communalities range no more than 0.3).

Summary of PCA contextual performance in all six research contexts can be specified as follows: (table 3).

Table 3. **Summary table for contextual adequacy criteria for PCA**

	MSE for communalities	MSE for loadings
PCA context	14%	3%
ULS context	27%	22%
PAX context	15%	4%
GLS context	21%	11%
AF & IF context	29%	17%
ML context	27%	10%

Results highlight that, out of 6 considered contexts, PCA is the most appropriate method for PCA research context. Likewise, equally well within measurement accuracy it scored in PAX context that differs from PCA context only by distribution of indicators (which is normal for PCA context and deviates from normal for PAX context). Therefore, if the only considered or accessible method is PCA, it is advised to strictly adhere to every prerequisite to input data characteristics and monitor the size and range of resulting communalities. The only admissible deviation from requirements might be different from normal distribution of indicators, which only impairs the contextual performance of PCA by 1%, under otherwise equal conditions.

Discussion

We investigated the properties of six factor extraction methods against each other and PCA. As a result we theoretically composed, empirically tested and adjusted an algorithm for selection of an appropriate factor extraction method depending on research context combined of a) the number of indicators on every factor (threshold requirement – no less than 6 indicators); b) distribution of indicators (normal against any other); c) whether there is a chance of model error; d) sample size (threshold requirement – 300 cases); e) size of the communalities (threshold requirement – 0.6) and f) range of the communalities (threshold requirement – 0.3). The theoretically substantiated algorithm was derived from mathematical model and previous research analysis. It was then tested on simulated data matching the stated research contexts.

As a result of the experiment, only for two research contexts theoretically substantiated algorithm turned out to be correct: ML is indeed appropriate in contexts where the only deviations from ideal input data for factor analysis are sample size and number of indicators. Partly proved to be true the appropriateness of GLS for contexts where every prerequisite is violated, except for sample and communalities size, but equally appropriate in this context turned out to be ML. The rest of the algorithm proved to be inaccurate: in contexts where every prerequisite is violated, the most appropriate factor extraction methods are PAX and AF; in the rest contexts the most appropriate methods are either GLS or ML, or both turned out to be equally appropriate. We speculate that the possible cause might be the large number of criteria that make up the research context. Previous research only combined three or less criteria [3, 8, 13–15]; here, on the contrary, we combined six of them. Presumably, their combinations, as opposed to separate criteria, determined the contextual performance of factor extraction methods.

Theoretically substantiated algorithm was improved: the number of significant aspects of research context was narrowed down to four (sample and communalities size, model error, and distribution of indicators), and only four of examined factor extraction methods were contextually appropriate in given research contexts. Overall, the recommendations on method selection might be put as follows:

1. PAX and AF are advised for use when a model error is suspected;
2. ML and GLS are equally advised for use when (a) indicators are distributed normally and sample size is over 300 cases and (b) both indicators are not normally distributed, and sample size is less than 300 cases (upon the condition that model error is ruled out and irrespectively of communalities size);
3. ML is advised to use when (a) indicators are not distributed normally and sample size is over 300 cases and (b) indicators are distributed normally, but sample size is less than 300 cases and communalities are less than 0.6 (upon the condition that model error is ruled out);
4. GLS is advised to use when indicators are normally distributed and communalities size is over 0.6, but sample size is less than 300 cases (upon the condition that model error is ruled out).

Reference

1. Kim, J.O. & Mueller, W. (1978) *Factor analysis: Statistical Methods and Practical Issues*. Beverly Hills, CA: Sage.

2. Harman, H. (1976) *Modern Factor Analysis*. Chicago: The University of Chicago Press.

3. Acito, F. & Anderson, R. (1980) A Monté Carlo Comparison of Factor Analytic Methods. *Journal of Marketing Research*. 17(2). pp. 228–236.

4. Browne, M. (1968) A comparison of factor analysis techniques. *Psychometrika*. 33(3). pp. 267–334.

5. Costello, A. & Osborne, J. (2005) Best practices in Exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*. 10(7). pp. 1–9.

6. Keith, T., Caemmerer, J. & Reynolds, M. (2016) Comparison of methods for factor extraction for cognitive test-like data: Which overfactor, which underfactor? *Intelligence*. 54. pp. 37–54.

7. Mislevy, R. (1986) Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*. 11(1). pp. 3–31.

8. Ihara, M. & Okamoto, M. (1985) Experimental comparison of least-squares and maximum likelihood method in factor analysis. *Statistics & Probability Letters*. 3. pp. 287–293.

9. Fabrigar, L., MacCallum, R., Strahan, E. & Wegener, D. (1999) Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*. 4(3). pp. 272–299.

10. MacCallum, R., Widaman, K., Zhang, Sh. & Hong, S. (1999) Sample size in factor analysis. *Psychological methods*. 4(1). pp. 84–99.

11. Marsh, H., Balla, J. & McDonald, R. (1988) Goodness-of-Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size. *Psychological Bulletin*. 103(3). pp. 391–410.

12. De Winter, J. & Dodou, D. (2016) Common Factor Analysis versus Principal Component Analysis: A Comparison of Loadings by Means of Simulations. *Communications in Statistics: Simulation and Computation*. 45(1). pp. 299–321.

13. Briggs, N. & MacCallum, R. (2003) Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*. 38(1). pp. 25–56.

14. Nylund, K., Asparouhov, T. & Muthén, B. (2007) Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*. 14(4). pp. 535–569.

15. Coughlin, K. (2013) *An Analysis of Factor Extraction Strategies: A Comparison of the Relative Strengths of Principal Axis, Ordinary Least Squares, and Maximum Likelihood in Research Contexts that Include both Categorical and Continuous Variables*. Graduate Theses and Dissertations. [Online] Available from: http://scholarcommons.usf.edu/etd/4459.2013 (Accessed: 26th October 2022).

Information about the authors:

Suleymanova A.N. – Master in Sociology, senior lecturer, Department of Sociological Research Methods, School of Sociology, Faculty of Social Sciences, National Research University Higher School of Economics (Moscow, Russian Federation). E-mail: asuleymanova@hse.ru

Zangieva I.K. – Candidate of Sociological Sciences, Associate Professor, Department of Sociological Research Methods, School of Sociology, Faculty of Social Sciences, National Research University Higher School of Economics (Moscow, Russian Federation). E-mail: izangieva@hse.ru

The authors declare no conflicts of interests.

Сведения об авторах:
Сулейманова А.Н. – магистр социологии, старший преподаватель кафедры методов сбора и анализа социологической информации, департамента социологии, факультета социальных наук Национального исследовательского университета «Высшая школа экономики» (Москва, Россия). E-mail: asuleymanova@hse.ru
Зангиева И.К. – кандидат социологических наук, доцент кафедры методов сбора и анализа социологической информации, департамента социологии, факультета социальных наук Национального исследовательского университета «Высшая школа экономики» (Москва, Россия). E-mail: izangieva@hse.ru

Авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию 20.08.2022;
одобрена после рецензирования 20.09.2022; принята к публикации 27.10.2022
The article was submitted 20.08.2022;
approved after reviewing 20.09.2022; accepted for publication 27.10.2022