CONTRIBUTION TO KNOWLEDGE OF LATENT DIMENSIONS EXTRACTION
IN SCIENTIFIC METHODOLOGY

Dobromir Bonacin, Zdravko Carev and Danijela Bonacin
Travnik, Bosnia & Herzegovina

Abstract
In this article it was presented totally new procedure and approach to latent dimensions analysis with suggestion that in the future we turn away thinking about refusing “unimportant” dimensions by any criterion. It was suggested that it is necessary to start data analysis (even with new methods preparation) that will take in account all data, meaning also all information in some data collected set. Seems that finally comes a time for fundamental reconsidering of known methods for hypothesis checking. Of course, old methods do not need to be ejected and forget but to determine their real methodological strength and to prepare new ones. Those new methods should start from maximal comprehensions and not from practicism and low level knowledge.

Key words: latent dimensions, criteria, structure, logic, All

Introduction
Determination of “real” number of latent dimensions in social and other scientific areas is a problem that is constantly researched and whose “final” solution everyone tends. So, many researchers tried, with different criteria and different starting conditions in data analysis, to establish some absolute criterion that can satisfy, if doesn’t all, then almost all (no matter how we define term “almost”) situations. It is, of course, natural wish that science can be brought into “standardized waters” where objectivity and repeatability could be maximized and researcher’s subjectivity on results minimized. In such context we most often talk about choosing adequate criterion that satisfies some mathematical, logical or even intuitive function of extremity of some value. So we have, among most popular criteria, egg. GK criterion (Characterized with fact that reasonably assumes that some factor cannot have less variability then one variable itself. That is well sustainable although one can find many reasons that can deny such starting attitude), then EB criterion (Characterized primarily by reasonable assumption that real number of factors is defined by common part of variability of variables included. Of course, neither this have to be strictly reliable in many situations,), then SCREE technique (This technique is essentially a transfer of mathematical tangential knowledge where we reject rest Eigen values of latent dimensions when sorted Eigen values come from non-linearity to linearity, so we can assume that further Eigen values and ‘their’ factors do not explain additional capital comprehensions, which is marked with very high level of speculation.), then percentage criterion (which assumes that the extraction is finalized when we exhaust some in advanced defined percentage of total variability. Of course, this is pure speculation, but it works sometimes.), etc. Any of those, or some other criteria, remains interested and keeps some specifics that in particular situations can make it adequate.

But we have extremely rare situations where all of those criteria agree about number of significant latent dimensions in some field of associative connections of indicators or variables, and sometimes groups or entities. For those reasons, problem of latent dimensions number, do not help us in realization of scientific aims or needs (knowledge) but unfortunately very often brings confusion that we cannot describe as small or trivial. Because of that problem, different researchers on virtually same samples, even with same variables application, obtain different latent dimensions, from facts connected with different attributes of virtually similar samples, through fact of different factors numbers (which results in different structures), to finally significantly different relations of such factors obtained.

All mentioned, in significant measure weeks thrust of scientific auditorium to such methods, and some even goes so far that generally denied any value to those methods, simply because there does not exist general, researches independent paradigm for application of some reliable model of latent dimensions extraction. Additionally, it is often doubtable is it anyway necessary to find ‘optimal’ oblique dimensions or simply orthogonal, and if there exists agreement, it definitely is not in method choice (varimax-maximax-quartimax... or oblimin-orthoblique-promax...) and a choice of method is mostly connected with specific researcher’s knowledge or even what method is technically available. Situation is not even better in choice of space where everything about latent dimensions will be done, and we have targeting in real space, image, Harris’... with results in all that spaces even slightly different. So, if we have only three orthogonal and three oblique solutions in three named spaces, and let we have only three named (or any other) criteria for latent dimensions extracting, it is easy to calculate that we have potentially (max.) 81 different solution.
Although, some of those solutions are highly congruent and with very similar structure, they can also be so different that we cannot be sure if it is a same project anyway. Obviously, this problem will not be solved in a way that for particular typical scientific disciplines or areas we suggest some solution, and especially in scientific way will not be solved in a manner that some solutions can be incorporated in commercial statistical packages and their marketing distribution will choose particular solution (or several solutions) integrated in selling products.

**From the beginning**

We should start from idea about what are latent dimensions anyway. But, as a part of problem solving, we will first agree that sample of objects from analyzed population is representative i.e. hypothetically enormous (egg. bigger then 10000 from 30000 population) and that mathematical (statistical) degree of transparency (transfer of conclusions) from sample to population is known. In that manner we have solved problem of objects – they are representative and we will not (in this case) discuss more about it. Latent dimensions are nothing else but relations based on primarily common variability of starting indicators/variables. Obviously, if no common variability exists between variables, no relations will be, and no latent dimensions we can extract, no matter of common variability type (usually correlations, covariance etc.). All such associations will be of zero value and no multidimensional space will exist, and, of course, no latent dimensions too. Well, we can say that latent dimensions are hypothetical mathematical constructs performed from relations of primary indicators or variables. That is an important reason to take care of what variables will be included because factor (latent dimensions) structure depends of their number and type. So, let we suppose that we have sample with a large number of variables (egg. 500) that describes early mentioned 10000 entities. Now, we have ensured that variables sample is also representative and with very high level of transparency according to fact that with included variables we cover whole (or almost whole) researching space. And let suppose we prepare initial data of all standards for data collecting, and in accordance with Measurement theory, and so we will not discuss that topic further any more. Now, still only remains that, in some chosen metric (real, image, Haris’ etc.) we extract latent dimensions, supposing that number of those dimensions will be a smaller than of initial variables. If no starting variable was (coincidentally or not) a linear combination of one or more others in model, and initial matrix is positively definite, all conditions are realized for Full extraction profile - meaning for extraction of maximal number of latent dimensions, that is in most cases (and most metrics) equal to number of initial variables. So, one can conclude that we mix a little our data, but in sense of dimensionality we got nothing.

In other words, if the number of latent dimensions is the same as initial variables, then we are still in some space inside some metrics, but still in 500 dimensions. This problem we call Problem of dimensionality and most often it has been trying to solve above eigen values that are essentially common distance (or better – deepness) that some potential latent dimensions are present in some space. And, if some eigen value is lower, than we conclude that it is not especially deeply included in that higher common space, and according to that, its latent dimension is not pregnant enough with interesting information so that dimension for further procedures we exclude. In that manner, obviously for further procedures we keep lower number of new dimensions, and because they are combined from initial (measurable) variables we call it latent (not directly measurable). With this procedure our research space is usually significantly reduced and from starting m variables (dimensions) we obtain less number of k latent (dimensions) where it worth rule (intuitive but not mathematically) that number of latent dimensions should be max. m / 3. It is clear that in case of starting 500 variables k could be over 160, so we did not made some big step in wanted direction, and we are still in too dimensioned space and far from research study simplification. On “this place” mentioned extraction criteria “enter game”, although, as we can see, in real complex situations do not have any serious role.

“**They are all perfect**”

In an extraordinary movie “The last samurai” one of main characters, except war strategies (in accordance with Japanese Buddhism) was occupied with beauty of cherry flower, searching for “perfect” flower, suggesting us that even if such searching lasts whole life, it will not be life spent in vain. In last movie sequences that hero dies and in the last moments on Earth he had saw enormous number of cherry flowers with comprehension that “they are all perfect”. In the same way, at Middle ages scientists and alchemists search for perfect metal, perfect elixir, perfect knowledge, perfect women, perfect weapon,..., and modern statistical alchemists search for perfect factors in factor analyses. Somebody has to get idea earlier that any of those factors (latent dimensions) is perfect. In the moment when we decide (by any criterion) for dimensionality reduction we choose some trees (or combination) from existing wood, supposing that they are more important, and some smaller and weaker we reject supposing they are irrelevant. By that act we lose some information for sure, and part of information lost is as bigger as we reject more dimensions, and if they content unknown information for us. So, with dimensions extraction, according to modern science, we are in quite big confusion, because, from one side we have ‘butterfly effect’ (with which we do not agree, but it is irrelevant now), so only one move of butterfly wing in Myanmar can, depending of conditions,
produce an earthquake in Canada, suggesting that we can forget or superficially reject nothing. And in our case we reject 25, 30, 45, even more then 50 % of total information in our data, keeping only “most important significant” dimensions. How can we know that such few rejected percentages are not crucial for understanding completeness of complex data we dispose? What are our rights to reject it only because they are signed with some smaller number (eigen value)? If that is as it is, why don’t we delete number egg 4 because there are many numbers bigger than 4? Where is the end of such rejection? Will we start to kill people that are “less important” in our social or financial latent space? Let suppose that from 500 initial variables we got 500 latent dimensions (as usually). Any of those 500 latent dimensions means something, no matter that means a little to us. What rules we determine if we deny some parts? Are that (in that case) rules at all? So, if today “main” global things are wars, politics and recession, do we have the right to say to a little girl: “Give us your doll to burn it, because in World of wars and recession it worth’s nothing!” In total model of World occupied with wars, recession and billions, such doll could be accompanied to some latent dimension with minimal eigen value, small enough so we can reject as irrelevant. Such world will continue functioning as far as we postulate principles of war, leadership of money, billions and destruction. And dolls will disappear. In catharses of his last breath Japanese samurai comprehends that all cherry flowers are perfect (no matter what they are) and our evaluations about what is perfect and what is not are only subjective superficial perceptions of world in which we exist and which we are integral part. Let transfer just a small part of that knowledge in other scientific or knowledge areas, without dying let us remove the veil and we will see that every doll is perfect and every factor (latent dimension) is perfect and significant because it describes important specific manifestation for itself, and no other latent dimension describes it in such way. It is unique! By rejecting such (‘small’) latent dimension we reject something we do not understand but exists, as well as something that contributes totality of world in which we are, so everything should be kept. Otherways, anybody who wants can easily determine criteria as he wants, rejecting everything out of those criteria. And, as far as we know about people through history, it can be very, very dangerous.

Oscillations

If we suppose that from totality of our surround (including us too) on some way we can exclude some (more or less) recognizable part and call it entity. Although it can be a part of other entities, if it is in some way complete, acts on its total surround. Question of all questions is: “What is a way it acts on surround?” Million of answers were generated but the answer should be only one: “Acts in that way so it tries to make all entities in surround equal to itself!”

Of course, it cannot succeed because Universe is too big for understanding, for including in him, etc., so it changes, vanish, destructs, develops, but still, if exists, tries the same. Let’s also suppose that every entity in affection on surrounding cannot change its acting every now and then, because it wouldn’t be itself, so from that it directly follows that as long as he intends the same, he will be acting in globally identical way. Finally, let us suppose (only for discussion), that such acting is oscillation emitting of specific frequencies towards the surroundings (as it is). It is obvious that (egg. with human) such oscillations is very big, but probably a finite number. In a manner of the previous discussion about latent dimensions it isn’t difficult to prove that every latent dimension represents a description of one entity acting frequency towards the surroundings. This is exactly the aim of this article. By rejecting some ‘irrelevant’ latent dimensions, we reject some oscillations of minor power but which characterizes exactly that entity and are its undividable part. It should be explored if we may do such procedures only because our cognitive apparatus cannot manage such number of information.

Methods, results and discussion

In methods it was enough to take any bigger amount of data set with n entities and m variables (egg. n > 680, and m > 25) and process data with latent dimensions extraction procedure (egg. Hotelling, Guttman, Joreskog...) to produce principal components (in the first step).
Then the procedure minimizes variables vectors deviations from factor vectors by rotation (e.g., Orthoblique). In forward steps entities values on (orthoblique) factors for the next level were taken, until on some level one single latent dimension was extracted. That last dimension by logic of procedure settles on very top of such structure. In the end, all data of all projections of all factors from analyses from all successive levels were sorted according to that final dimension because it represents that space in whole, with, of course, essential of kept variations. What is 'perfect' says: "In all situations, we always got a set described by figures 1-5 (the rest are not presented because of space but the rule is completely preserved) and that is just unbelievable. Here is presented an approximation of real data by polynomial trend of fifth level on main components.

Already at first sight it is clear that this is polynomial function of 5th level that describes sine (or cosine) function. That way we can present all 27 principal components from the example. It can be seen that the last (27th) only a set of positions that are defining ‘up and down’ function, and those are minimal oscillations which are in such system, model or entity possible at all. Of course, that doesn’t mean they are completely insignificant and that we can reject them easily. Here are presented Principal components, but the exact same image is given by Varimax positions, and after minimal transformations any other oblique solution as Orthoblique etc. And of course that tested hypothesis that taxonomic structures act the same way, which was approved with polar taxons model used (Kaiser and Harris with Momirović expanded – 1987, coded and modified by Bonacin - 2004).

Based on everything mentioned it should be highlighted that the time for new methods preparation has come, methods that will take in account all information in data set without losses of seemingly ‘irrelevant’ but essentially real existing information.

References

PRILOG POZNAVANJU EKSTRAKCIJE LATENTNIH DIMENZIJA
U ZNANSTVENOJ METODOLOGIJI

Sažetak
U članku je prikazan pristup analizi latentnih dimenzija na potpuno novi način uz prijedlog da se ubuduće otklene promišljanja o odbacivanju "nevažnih" latentnih dimenzija po bilo kojem kriteriju. Predloženo je da se započne analiziranje podataka (ako treba i pripremom novih metoda) koje će uzimati u obzir sve podatke, a time i sve informacije koje se nalaze u skupu prikupljenih podataka. Čini se da je došlo vrijeme da se poznate metode za dokazivanje hipoteza temeljito preispitaju, ali ne zato da bi ih se zaboravilo i odbacilo, već da bi se utvrdila njihova stvarna metodološka snaga i pripremile nove. Te nove metode bi trebale polaziti od maksimalnih spoznaja, a ne od praktizma i spoznaja nižih razina.

Ključne riječi: latentne dimenzije, kriteriji, struktura, logika, Sve