Implementation of Multivariate Statistical Analysis for Warning Forecasting

Zbigniew MIERZWA
PhD in economics, Financial University, Moscow, Russia
ZEMezhva@fa.ru
orcid.org/0000–0002–2488–6315

"In the analysis of economic forms, moreover,
neither microscopes nor chemical reagents are of use.
The force of abstraction must replace both."
Karl Marx, 1867
Preface to the First German Edition of "Capital. A Critique of Political Economy", Volume I

Abstract. Traditionally, for the purposes of forecasting socio-economic phenomena are used econometric methods (methods). Much less frequently for these purposes, we used the methods of multidimensional comparative analysis, including the Wroclaw method of taxonomy. This methodology allows not only classifying the analyzed objects, such as countries or regions but also, taking into account time, to determine the trajectory of the actual development. By modeling the numerical values of variables one can determine a desired or optimal path of development. The third method of application of Wroclaw taxonomy is a ranking of the studied objects about the level of development. The article presents the fundamentals of the Wroclaw taxonomy and basic methodological issues that arise in its application.
Keywords: Wrocław taxonomy; classification; path of growth; inequality; multidimensional comparative analysis.
JEL C53,C55,C61

Применение многомерного статистического анализа для конструкции предупреждающих прогнозов

Збигнев МЕЖВА
доктор экономических наук, Финансовый университет, Москва, Россия
ZEMezhva@fa.ru
orcid.org/0000–0002–2488–6315

Аннотация. Традиционно для целей прогнозирования социально-экономических явлений используются эконометрические методы (модели). Значительно реже для этих целей применялись методы многомерного сравнительного анализа, в том числе метод Вроцлавской таксономии. Эта методология позволяет не только классифицировать исследуемые объекты, например страны или регионы, но также, с учетом времени, определять траекторию фактического развития. Путем моделирования числовых значений переменных можно определить желаемую или оптимальную траекторию развития. Третьим способом применения Вроцлавской таксономии является ранжирование исследуемых объектов по уровню развития. В статье представлены основы Вроцлавской таксономии и основные методологические вопросы, возникающие при ее применении.
Ключевые слова: Вроцлавская таксономия; классификация; траектория роста; неравенство; многомерный сравнительный анализ.
JEL C53,C55,C61

INVOCATION

Michael Roberts, the first-class 'red' blogger, writes in his last post (https://thenextrecession.wordpress.com/2017/11/27/neoliberalism-works-for-the-world/):

"Marx was the first to note the tremendous boost to production that the capitalist mode of production delivered compared to previous modes. But as I have shown in previous posts, there is another side to capitalism's early years: the immiseration of the working class." And that is a different reality from neoliberal's claims.

And he continued: "The *empirical evidence* supports Marx's view that, under capitalism, **poverty** and **inequality** of income and wealth have not really improved under capitalism, neoliberal or otherwise. Any improvement in poverty levels globally, *however measured*, is mainly explained by in state-controlled China and any improvement in the quality and length of life comes from the application of science and knowledge through state spending on education, on sewage, clean water, disease prevention and protection, hospitals and better child development. These are things that do not come from capitalism but from the common weal."

So, Marx's prediction 150 years ago that capitalism would lead to greater concentration and centralization of wealth, in particular, the means of production and finance, has been borne out. Contrary to the optimism and apologia of mainstream economists, poverty for billions around the world remains the norm, with little sign of improvement, while inequality within the major capitalist economies increases as capital is accumulated and concentrated in ever smaller groups.

We would like to analyze the question what does it mean '*empirical evidence*' and does it count '*however measured*". It seems we ought to measure any economic and social phenomenon correctly. Isn't it?

Because one cannot **rewind history** and replay events after making small controlled changes, causation can only be ***inferred***, never exactly known.

DIAGNOSIS AND TREATMENT

The economy is a living organism. So, **the correct diagnosis is a *conditio sine qua non* of successful treatment**. That is, we ought to search ***causal relationships*** between **all** factors of social, political, and economic life of society. But it is, of course, impossible. However, we must try to do it anyway. From the point of contemporary statistical techniques, it will be the cruel torture. Nevertheless, we have no choice.

The world of economy is really very complex. The difficulty of grouping or ordering (by similarity, for example) grows exponentially with the number of objects to be classified and the number of dimensions on which they are being grouped. Thus, for very large samples and many variables, some shorthand methods such as clustering algorithms or formulas need to be devised. Otherwise, we will be unable to see the forest for the trees. Moreover, there are so many varieties of the cluster and taxonomic analysis that the novice, not to mention the expert, risks bewilderment.

It is hard to find a definition of the term *complex phenomenon* in specialist literature. We can describe the *complex phenomenon* as an abstract construct depicting the qualitative state of directly immeasurable real objects, described by a number (more than one) of *diagnostic* variables. The diagnostic characteristics which describe the studied problem change under the influence of various factors, including the ones of random character, and moreover they remain in interrelationships.

A problem is that if the number of dimensions is large, and the number of categories in each dimension is also large, the resulting typology or classification may contain a great many cells or types. For example, even if all dimensions are dichotomies, the formula for determining the number of cells is 2^m, where ***m*** is the number of dimensions. Thus, for five dichotomous dimensions the typology will contain only 2^5 or 32 cells, but for 12 dichotomous dimensions, the number of cells is 2^{12} or 4,096. If the dimensions are *polytomous* rather than *dichotomous*, as it is often the case in economic research, the number of cells expands much more rapidly. Because the number of types can be so large, researchers have often found it helpful to use partial or shorthand typologies. These can be formed either by constructing only a portion of the full typology or by first constructing the full typology and then selecting only certain types for use in the analysis (or by merging some types together). For example, if we wish to construct a typology from seven dichotomous variables, we may find it difficult to work with all of the 128 resulting types.

The question is **for what purpose** a classification is to be established. It means that a taxonomic system will reflect the purposes for which it is constructed. It is particularly important that the maker of a classification should have a clear idea of what he wants, and he/she should indicate what components are used as the basis for his classification.

Classification involves the ordering of cases in terms of their ***similarity*** and can be broken down into two essential approaches: ***typology*** and ***taxonomy***. The former is primarily conceptual, the latter empirical. Construction of a typology requires conceptualization along at least two dimensions. ***Taxonomy*** begins empirically, rather than conceptually, with the goal of classifying cases according to their measured similarity on observed variables. The principal approach here is the implementation of ***taxonomic*** or ***cluster analysis***. For example, ***Ward's hierarchical clustering method*** (a widely used agglomerative, objective, average linkage procedure) finds that the objects group themselves into several distinct clusters. Conceptually, what do these clusters represent? However, the cluster solution does not speak to the **conceptual meaning** of the clusters but instead confines itself to a demonstration of their empirical presence.

In its simplest form, *classification* is merely defined as the **ordering** or **grouping** of objects into groups or classes on the basis of their ***similarity***. Statistically speaking, we generally seek to minimize within-group *variance*, while maximizing between-group variance. This means that we arrange a set of objects into groups so that each group is as **different** as possible from all other groups, but each group is internally as **homogeneous** as possible. By maximizing both within-group homogeneity and between-group heterogeneity, we make groups that are as **distinct** (non-overlapping) as possible, with all objects within a group being as alike as possible. These are general goals that specific classification techniques may alter somewhat.

However, classification is understood as the general process of grouping entities ***by similarity***, unfortunately, similarity has no definite meaning in economics. Classification can either be unidimensional, being based solely on a single dimension or characteristic or multidimensional, being based on a number of dimensions. When multidimensional, the dimensions are generally thought to be correlated or related. Unrelated dimensions generally would not be combined in a classification but could be. Dimensions are generally ***categorical data***, such as *nominal* or *ordinal* variables. However, *interval* and *ratio* variables can be used as well. In economics quantified cluster and taxonomic methods can use variables of all levels — nominal, ordinal, interval, or ratio.

Two characteristics distinguish ***typologies*** from ***generic classifications***. A typology is generally multidimensional and conceptual. Typologies generally are characterized by *labels* or *names* in their cells. The generic classification process is quite simple. The only basic rule is that the classes formed must be both ***exhaustive*** and ***mutually exclusive***. This means that if ***N*** objects are to be classified, there must be an appropriate class for each (exhaustivity), but only one correct class for each, with no case being a member of two classes (mutual exclusivity). Thus, there must be one class (but only one) for each of the ***N*** objects.

As just defined, classification is both a *process* and an *end result*. We may thus speak both of the processes of classification and of a classification so formed. As an end *result*, taxonomy is similar to a typology, and in fact, many people use the two terms interchangeably. Here we will reserve the term taxonomy for a classification of *empirical* entities. The basic difference, then, is that a typology is conceptual while taxonomy is empirical. Exceptions to this generally involve the subsequent identification of empirical cases for conceptual typologies, but not the conceptualization of taxonomies.

A classification is no better than the dimensions or variables on which it is based. If you follow the rules of classification perfectly but classify on trivial dimensions, you will produce a trivial classification. One basic secret to successful classification, then, is the ability to ascertain the *key* or *fundamental characteristics* on which the classification is to be based. As a case in point, a classification that they have four legs or two legs may produce a four-legged group consisting of a giraffe, a dining-room table, and a dancing couple. Is this what we really want? So, it is crucial that the fundamental or defining characteristics of the phenomenon be identified. Unfortunately, there is no specific formula for identifying key characteristics, whether the task is theory construction, classification, or statistical analysis. In all of these diverse cases, **prior knowledge** and **theoretical guidance** are required in order to make the right decisions. So, the **selection** of ***diagnostic variables*** is a first key issue in the diagnosis of economic health.

So, the first exercise is *data collection*. More difficult problems arise at this first stage than in any other subsequent analytical phases. The reliability of the data, their conceptual validity is the most difficult hurdles to dear. Conceptual invalidity and data unreliability are 'crimes' well enough. A 'mortal' sin does in no way absolve a venial one!

The selection of diagnostic variables is a particularly important and responsible process for it directly influences the final results of the study. It is crucial the diagnostic variables used in the study meet the requirements of *relevance*, *normativity*, and *explicitness*. The requirement of *relevance* demands that variables representing the most significant components of the analyzed phenomenon. The requirement of *normativity* denotes measures having either positive or negative influence on the analyzed phenomenon. The requirement of *explicitness* demands that the study uses variables which explicitly specify the relations between a phenomenon represented by a given measure and other phenomena.

The computer programs have been developed for the selection of an optimal subset of a set of possibly *informative, diagnostic* or *prognostic* variables. They can be equally useful for other discriminant analysis or pattern recognition problems involving variable selection. The approach is probabilistic; i.e., *diagnostic probabilities* are assigned to object on the basis of the values observed on the diagnostic variables. The statistical model used is largely based on the assumption of independence between the variables, but one model-parameter, the so-called '*global association factor*', is added in order to take dependency into account. The stepwise forward selection strategy of adding in each selection step a new variable to the set of already selected variables is used. The user may choose between numbers of selection criteria. Such a criterion is used in order to decide in each selection step which variable should be added. All criteria are based on measures of ***diagnostic*** or ***prognostic performance***.

It may be that only a few chief types are found to be really important for us, so that we may focus on these and neglect the remainder. Alternatively, it may turn out that a number of types (perhaps an unknown number) are needed, but not the entire typology. In such cases, it is common to utilize a shorthand typology by first constructing only ***key criteria types***, and then locating all other types of reference to these criteria.

For example, we could define as a ***criterion type***, the type with the highest diagnostic values on all dimensions. In this case, a criterion type is named ***pattern type*** and served as, for example, ***development pattern*** or ***taxonomic measure of development*** as in Wroclaw Taxonomy. Then other types could be measured in terms of their deviation or distance from this criterion. Often two polar types are used. ***Polar types*** are two extreme opposite types (such as the type scoring highest on all dimensions and the type scoring lowest on all dimensions) as, for example, in **TOPSIS** method (**T**echnique for **O**rder-**P**reference by **S**imilarity to **I**deal **S**olution), firstly proposed by Hwang and Yoon (1981). All remaining types would be intermediate to the polar types and could be located in terms of their *deviation* from these two cells. This allows a researcher to leave the majority of cells latent and to construct only those cells that have *representative types*, as measured by their deviation from a criterion type or polar types.

THE HOLY GRAIL GDP

The Holy Grail — different traditions describe it as a cup, dish or stone with miraculous powers that provide *happiness*, *eternal youth* or *sustenance in infinite abundance*.

However, many supporters of capitalism as the only and best system of the human social organisation are worried that capitalism does not (or no longer seems) to deliver ever-increasing living standards for the majority, but instead is producing ever greater inequalities of wealth and incomes, to such a point that it could provoke a backlash against the system itself.

Ideas on the links between economic growth and development during the second half of the 20th century had a formative influence. Gross Domestic Product (GDP) and economic growth emerged as leading indicator of national progress in many countries. However, GDP was never intended to be used as a ***measure of wellbeing***. At the same time, it is growing acceptance of the fact that monetary measures, such as GDP per capita, are inadequate proxies of development, and especially social one. In the 1970s and 80s development debate considered using alternative focuses to go beyond GDP, including putting greater emphasis on employment, followed by redistribution with growth, and then whether people had their basic needs met. These ideas helped pave the way for the ***human development*** (both the *approach* and its *measurement*).

Between 1967 and 1972 UNESCO has conducted research program "Toward a system of quantitative indicators of components of human resources development" and issued 20 research studies. That program has been continued by "Social science project on human resources indicators" program. A first selection of the papers and studies, relating to an earlier phase of the project, was published in 1972 by the Institute of Philosophy and Sociology of the Polish Academy of Sciences [Gostkowski, 1972]. The shift in emphasis in the work on social indicators was made in the next collection of papers prepared for the research project on indicators within the UNESCO Social Science Methods and Analysis Division titled "The use of socio-economic indicators in development planning" [UNESCO, 1976]. This collection has illustrated the change in direction, from indicators for international comparison to the 'operationalization' of indicators, i.e. for their use in planning and programming.

In 1960s United Nations Research Institute for Social Development (UNRISD) has conducted program "Pioneering Social and Human Indicators of Development" with sub-program "Measuring Social Development". In 1970s program "Debating the Social and Political Dynamics of Modernization" with sub-program 'UN "Unified Approach to Development"'. In 1980s program "Promoting a Holistic and Multidisciplinary Approach to Social Development" with sub-program "Measurement and Social Indicators". In 1990s program "The Social Effects of Globalization" with major areas of research among others "Crisis, Adjustment and Social Change", "Environmentally and Socially Sustainable Development", and "The Future of the Welfare State". In 2000s the major areas of research are poverty eradication, the promotion of democracy and human rights, gender equality, environmental sustainability, the effects of globalization.

UNRISD has convened Meeting of Experts on Social Development Indicators which has been held 8–11 April 1991 in Rabat, Morocco to review work of last three decades in the field of social development indicators and explore ways of providing better information on which to base social policy decisions. These included the persistence of data problems that limit the usefulness of numerous key indicators; the need for lower-cost and innovative methods for collecting and managing social data; efforts to improve the applicability of existing indicators; and new initiatives to develop indicators for the analysis of diverse phenomena (environmental degradation, gender inequality, and poverty).

Currently research for Social Change–Transformations to Equity and Sustainability: UNRISD Strategy 2016–2020 sets out the main priorities and themes of UNRISD research within an overarching institutional framework that links research, communications, policy engagement, results and impact. The Institute's current research is organized in three programme areas: Social Policy and Development, Gender and Development, and Social Dimensions of Sustainable Development.

HUMAN DEVELOPMENT INDEX

The human development approach, developed by the economist Mahbub Ul Haq, is anchored in the Nobel laureate Amartya Sen's work on human capabilities, often framed in terms of whether people are able to "be" and "do" desirable things in life.

The first Human Development Report introduced the ***Human Development Index* (HDI)** as a measure of achievement in the basic dimensions of human development across countries [United Nations Development Programme, 2017]. The Human Development Index (HDI) is a ***summary measure*** of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living. The HDI is the geometric mean of normalized indices for each of the three dimensions.

In successive ***Human Development Indexes*** has been embodied one of the more important achievements of the human development approach, namely, a growing acceptance of the fact that monetary measures, such as GDP per capita, are inadequate proxies of development.

Human Development Reports (HDRs) have been released since 1990 and have explored different themes through the human development approach. The reports, produced by the Human Development Report Office (HDRO) for the United Nations Development Programme (UNDP), are ensured of editorial independence by the United Nation's General Assembly. Indeed they are seen as reports to UNDP, not of UNDP. This allows each report greater freedom to explore ideas and constructively challenge policies.

The HDI was created to emphasize that **people** and their capabilities should be the ultimate criteria for assessing the development of a country, not economic growth alone. The HDI can also be used to question national policy choices, asking how two countries with the same level of GNI per capita can end up with different human development outcomes. These contrasts can stimulate debate about government policy priorities.

The HDI simplifies and captures only part of what human development entails. It does not reflect on inequalities, poverty, human security, empowerment, etc. The HDRO offers the other composite indices as a broader proxy on some of the key issues of human development, inequality, gender disparity, and poverty. These somewhat crude measures of human development remains a simple unweighted average of a nation's longevity, education and income. Over the years, however, some modifications and refinements have been made to the index.

The 2016 Human Development Report is the latest in the series of global Human Development Reports published by the United Nations Development Programme (UNDP) since 1990 as independent, analytically and empirically grounded discussions of major development issues, trends and policies. Additional resources related to the 2016 Human Development Report can be found online at http://hdr.undp.org. Here you can find digital versions of the report and translations of the overview in more than 20 languages, an interactive web version of the report, a set of background papers and think pieces commissioned for the report, interactive maps and databases of human development indicators, full explanations of the sources and methodologies used in the report's composite indices, country profiles and other background materials. In archive one can find as well previous global, regional and national Human Development Reports.

However, human development and sustainable development should be treated as a *multivariate phenomenon* where multiple relations between multiple variables are examined simultaneously. The concepts of sustainability and sustainable development form the basis of long-term growth strategies. Its realization should help to build foundations for economic growth that also improves many social dimensions such as social inclusion, poverty, labor market situation, health.

TAXONOMIC MEASURES OF DEVELOPMENT

In modeling of socio-economic development, economic forecasting (prognosis), multidimensional comparative analysis one of the key questions is correctness of implementation the *distance* as a measure of taxonomic *similarity*, especially when variables are *composite aggregates*.

For any given *multivariate phenomenon* we have to take into account the possibility and reasonability of implementing:

Multiple-criteria analysis

Multiple-variable (multivariable) analysis (multifactor analysis)

Multiple-object analysis

Multidimensional analysis.

The taxonomic methods can be divided as follows:

1. Methods of ordering objects:

a) ***Linear ordering*** — the projection of multivariate space onto a straight line; it allows the establishment of the hierarchy of objects, that is the ordering from the object which is highest in the particular hierarchy to the object which is lowest in the hierarchy.

b) ***Non-linear ordering*** — the projection of multivariate space onto a plain; it does not allow for the establishment of a hierarchy of objects but only the establishment of similar objects to each given object.

2. Methods of a grouping of the studied objects:

a) ***Direct grouping methods*** — provide the groups of objects without moving them between groups in the subsequent stages of their grouping.

b) ***Iterative methods of grouping*** — the initial division into groups of objects, the choice of the function criterion of *goodness* of the grouping, the choice of the rules for the moving objects between groups which allows the increasing *goodness* of the grouping, the establishment of the rule which ends iteration.

3. The methods of choosing the representatives of objects and diagnostic variables:

a) Methods based on the matrix of the distance between the compared objects.

b) Methods based on the matrix of coefficients of correlations between potential diagnostic characteristics.

4. The aggregate methods (synthetic) of the construction of diagnostic variables:

a) Standard methods, which require the definition of the *standard (pattern) of development* and are based on the determination of the distance between the objects studied and the standard

b) Non-standard methods, which do not require the definition of development standard.

Therefore, there is a problem of choosing the optimal procedure for the analysis of the empirical data with specified statistical characteristics. Moreover, it is possible to combine all or some of above analyses into the complex method, for example, the most frequently applied *multidimensional multivariable analysis*. The more complicated analysis takes into account the dimension of time — multidimensional multivariable time-series.

In the selection of *diagnostic variables,* we can distinguish three of the most general groups of selection criteria for the factors of social and economic development: *substantive*, *formal* and *statistical*. The *substantive* selection of factors should include knowledge about the society, economy, finance, industries, and spatial economics and so on. The experience and intuition of the researcher are also essential. The most important problem of choosing diagnostic variables is the reliable existence of the values of their characteristics in a dynamic interpretation (continuous time-series). Each investigated ***object*** can be described by definite numbers of diagnostic variables.

For the *formal* criteria, the following issues should be included:

Measurability

Ensuring the comparability of the objects (diagnostic variables) in space and time

Complete data for all objects (diagnostic variables) and periods of the study.

The most important *statistical* criteria are:

A large spatial and spatial-temporal variability (coefficient of variation for the variable $v \geqslant 10\%$)

Asymmetric distribution

No excessive correlation.

The Polish *specialité de la maison* is **TMD — *taxonomic measure of development*** (**TMD**, Polish abbreviate — **TMR**) elaborated in 1967 by Zdzisław Hellwig (1925–2013), one of the greatest Polish experts in the field of statistics and econometrics [Hellwig, 1968, 1972, 1975, 1978, 1981]. Z. Hellwig has used the work of a group of mathematicians from Wrocław University headed by Hugo Steinhaus (1887–1972) for the elaboration of the *typology* of economic development [Florek *et al.*, 1952; Steinhaus, 1957]. From 1968 to 1974 he was a UNESCO foreign expert in Paris. This method has been used for *ranking* countries according to their level of development on the basis of several indicators in UNESCO ***Social Science Project on Human Resource Indicators***. His greatest achievements at that time are connected with the implementation of taxonomic methods and methods of multidimensional comparative analysis (**MCA**, Polish abbreviate –WAP).

At the beginning, Hellwig's work was based strictly in accordance with Steinhaus' graphic approaches involving linkage diagrams and multiple contour lines — taxonomic graphs [Hellwig, 1967, 1968]. The technique apparently remained undiscovered by English-speaking numerical taxonomists whose approach relied more upon numerical and computer techniques. However, a review of the method in 1964 drew attention to the work of the Wroclaw group and caused some researchers to adopt the Wroclaw techniques for the graphic representation of conventional taxonometric results. Nevertheless, although it is widely used in Poland, it is really hard to find it in any scientific paper written in English [Głodowska, 2016; Gostkowski, 1972, 1975; Jurkowska, 2014; Pawlewicz, 2005; Mesjasz-Lech, 2010; Pietrzak & Balcerzak, 2016; Sej-Kolasa, 2009].

These taxonomies are *linkage techniques*, with many variations. The *taxonomic units* (which in the present exercise are countries) are represented at first as a disjointed graph, the ***taxonomic graph***. The general algorithm consists in computing the distances for all pairs of units placed in the *n*-dimensional Euclidean space, finding the shortest distances and representing by a graph the units (*vertices* or *nodes*) connected to their nearest partner by links representing in *length these distances*. The connected graph, in this case, is the '*shortest spanning tree*'.

The first step consists in connecting each *vertex* (one may start with any of the vertices) with its nearest neighbor in the group. The 'link' is proportionate to the Euclidean distance between the two vertices or '*nodes'*. A disjointed graph of *first-order* concentrations is obtained when all the vertices have been connected. By inspection one next finds the shortest distances between pairs of nodes belonging to two different *concentrations*. Concentrations of the *second order* are thereby determined. The process is repeated until all the nodes have been connected to form one *single joint graph*. Clusters are formed by removing, after arranging n — l links in decreasing order the K longest ones, so that $K + 1$ clusters are formed. The number K is determined by the '*critical minimum distance*', which is defined as $C = \overline{C} + 2S_d$ where $\overline{C}$ is the mean and S_d the standard deviation of the minimum distances.

Today they are acknowledged not only in Poland but have gained international recognition. In these works, Z. Hellwig drew attention to the real possibilities of applying taxonomic methods of linear ordering to economic problems. He introduced the concept of ***synthetic measure*** — based on the concept of ***development pattern*** and commonly known today as ***Hellwig's economic development measure*** — and an algorithm for grouping objects into relatively homogeneous subsets. These ideas have inspired a number of scholars in various Polish institutions. The terms — also introduced by Z. Hellwig — of development measure, development pattern, development path and the optimal trajectory of development, aggregate and its information potential — are still employed by Polish scientists representing various fields of study.

In order to use this measure, a given multivariate phenomenon is decomposed to some economic aspects (called ***objects***), where each object describes a different part of the economic system. For each object, a subset of ***potential diagnostic variables*** is selected. The variables enable to characterize the selected object and allow describing it. Then, based on chosen diagnostic variables TMD is calculated, taking into account impact of all objects' variables of the examined economic phenomenon. TMD is interpreted as composite (synthetic) measure of phenomenon's development level. The application of TMD allows *ordering* analyzed objects (for example countries) based on the calculated level of development of the phenomenon. The use of TMD in economic analysis enables to assess the current situation of the objects under study and to make possible their ranking from the worst to the best. It needs to be stressed that at the beginning Z. Hellwig has tried to achieve ***typology*** of the analyzed phenomenon. In international comparative analyze it makes possible to describe static *socio-economic profiles* of the analyzed countries.

Multiple criteria analysis methods can be divided into two groups. The first group allows carrying out ordering of objects from the worst to the best from the perspective of analyzed complex phenomena. ***A taxonomic measure of development*** proposed by Z. Hellwig can be found in this group. The second group of methods allows classification of analyzed objects into *homogeneous subsets*, where the objects are characterized with similar values of the variables. In this group, one can find cluster analysis with Ward's method as an example.

The Wroclaw Taxonomy, along with other variations on the same theme, distinguishes itself from a large number of techniques which flourish in this field by sacrificing to simplicity and speed of calculations some of the information contained in the data and distance matrices. However, it was important in the time when computers on the stage of occurrence.

The most important advantage of the Hellwig's concept relates to its cognitive values in explaining economic reality and flexibility in its application. The tool can be used to analyze the most of economic phenomena that have complex nature. Even now, if however ease and speed of the calculations are given great weight because of inaccessibility to, or cost of using 'hardware', then the balance will be tipped in favor of Wroclaw taxonomy or allied techniques.

Current development of the concept of TMD concentrates on taking into account *spatial interdependence* in the design of the measure, and *time-spatial* interdependence as well. The other direction of development of taxonomy is the concept of taxonomies of structures.

Static and dynamic issues

Time is an important dimension (variable) in social and economic research. After all, the word 'development' implies taking into account time. Many classifications in social science are *synchronic* or *cross-sectional*, meaning that it occurs *at a single point in time*. Such synchronic or non-evolutionary relationships are sometimes termed static relationships. Although more difficult, it is also possible to utilize *diachronic classification* based, for example, on measures of change or on measures of *evolutionary resemblance*.

The *notion of resemblance* is associated with that of '*proximity*', which is similar to distance. For example, we speak of the gap between countries, of how one country is out-distancing another, etc. The ordinary two-dimensional graph where values of two variables are represented on orthogonal axes was among the first elementary attempts to use distances and construct a picture from two sets of data. The three-dimensional model is as far as we can get in a three-dimensional world to illustrate in a concrete manner *trends* and *relative movements* of variables. The two-dimensional graphs are however much more in favor because of ease of reproduction in books, the three-dimensional graphs reproduced in perspective on a flat surface being most times unsatisfactory. The *n*-dimensional space extrapolates the 3-dimensional one, and loses its 'spatiality', its physical meaning, while retaining the *mathematical properties* of *Euclidean metrics*, with the ordinary algebraical measures of the distances, angles, etc.

Groupings in a Euclidean space, whatever the dimensions, are results of a set of techniques which are circumscribed by the term ***cluster or taxonomic analysis***. They are numerical techniques based on notions of ***similarity***, ***resemblance*** or ***proximity*** of the variables to be classified or units to which the variables correspond. These notions of similarity, etc. find quantitative expressions in methods of quantification which are:

a) coefficients of similarity

b) correlation coefficients, regression coefficients, factor analysis and related techniques, like factor analysis of correspondence

c) distance-based, both *Euclidean* distance, like the one proposed in Wroclaw Taxonomy and *non-Euclidean* distances like the **I-distance** proposed by Branislav Ivanovic and the **Generalized Distance** proposed by Mahalanobis.

In the Wroclaw Taxonomy, the chosen metric is the Euclidean one, the *unit of measurement* is, in fact, the standard deviation, and the standardization process converts at the same time the vectors x and y to $x-\bar{x}$ and $y-\bar{y}$ respectively.

Interpretation of diagnostic variables

One of the important goals of every applied science is, beyond doubt, the establishment of *connections* and *mutual relations* between facts and processes which constitute the object of a given scientific observation. This is by no means a simple task since we have to deal not only with direct, clear, strong and durable connections, mainly of causal character but also with indirect, weak, unstable, changeable and stochastical relations which are difficult to examine and to identify. Whereas the connections embraced by the first group can be expressed in functional terms, the elements of the second group require some specific methods of description, an outstanding role amongst which play statistical methods. This is due to the fact that ***stochastical relationships*** are created very often either by intermediate or illusory connections between phenomena involved.

Due to the shape of the relationship between explanatory variable (called ***predictor*** or more traditionally — independent variable) and explained variable (called ***predictand*** or more traditionally — dependent variable or outcome variable) in econometrics is used following types of explanatory variables: stimulants, nominants, dis-stimulants, and *neutrals*.

As we said above, prior knowledge and theoretical guidance are required in order to make the right decisions concerning the selection of diagnostic variables**.** In Wroclaw taxonomy, it is also needed to divide diagnostic variables into stimulants, nominants, dis-stimulants, and *neutrals* on the basis of the *types of preferences*. It is this stage of work where prior knowledge and theoretical guidance are required in order to make the right decisions. In Wroclaw Taxonomy the stimulants are defined as variables that have a stimulating effect on the level of development of the phenomenon studied and therefore are desirable as their highest values. Dis-stimulants are variables acting to hinder the development of the phenomenon and therefore high values are not desirable. Nominants are variables that have a stimulating effect on the level of development (as stimulants) yet to a certain point (or span), called nominal, and above and below that point (span) the character of dis-stimulants. Neutrals are variables indifferent to an explanatory variable or with very weak dependence.

Dis-stimulants can be transformed into stimulants by calculating the inverse of each value according to the formula:

$$x'_{ij} = \frac{1}{x_{ij}}. \quad (i = 1, 2,\dots, n); (j = 1, 2, \dots, m) \qquad (1)$$

It is also admissible to use statistical measures for the initial selection of variables, such as coefficient of variation, correlation, and asymmetry. However, the point is that in the final set of variables, the information is not replicated.

Choice of predictors

The main value of the Hellwig's proposals relates to its cognitive values in explaining economic reality, methodological simplicity and flexibility in its application. The tool can be used to analyze most of the economic complex phenomena. However, there are two main limitations on the application of TMD in economic research. The first objective limitation is the *availability of statistical data*. The second one relates to *researcher's knowledge and experience*, which should allow to concretize properly an analyzed phenomenon and then to express its multi-dimensionality using single measurable economic aspects. In the case of the first problem cognitive values of the tool is not fully utilized or the tool cannot be applied. In the case of the second limitation, the cognitive values of the tool are used improperly, which can lead to serious cognitive errors.

In many cases, taxonomic researches are as a matter of fact very similar to statistical and econometric modeling. So, the problem is of special importance when dealing with any type of economic models. The three crucial questions which one faces are the following:

1. How to make the optimal selection of the set of variables (called here ***predictors***) which play the role of 'independent' variables?
2. How to fix the number (*n*) of predictors which should be introduced?
3. How 'to weight' the influence of predictors on the predictand(s), i.e. the 'dependent' variable?

Unfortunately, in many empirical implementations of Wroclaw Taxonomy, these issues are neglected [Borys, 1978; Grabiński et al., 1983; Grabiński et al., 1989; Kolenda, 2006; Malina & Zeliaś, 1997; Nowak, 1990; Panek, 2009; Pluta, 1977, 1986; Zeliaś, 2000].

The problem of weighting in multidimensional comparative analysis

The variables are not weighed, in the computation of the Euclidean distances for the Wroclaw Taxonomy. This distinguishes Wroclaw distance from the I-distance and Generalised distance of Mahalanobis. The problems of weighting have been raised in the UNESCO Project on Human Resources Indicators [Hellwig, 1969]. There can be no standard methods of weighting the variables which are used for computing the development distances. The choice of weights will depend on the purpose of the researcher: he may give more weight to variables relating to 'welfare' aspects than to 'production' aspects. There is no uniform development pattern, and most variables move in and out of importance as development proceeds along. The importance of a relatively high value for a variable might be crucial at a level of development and not at all important at another level when 'high' values for other variables become in turn of greater importance.

The main question here is the choice of the endogenous and exogenous criterion of the *relative importance* of variables. In this place, the following question may be put forward: are all variables, equally important in judging about achieved, say, level of economic growth or are some of them more and some less important? Before answering this question an agreement should be made as to what *criterion* we will be ready to accept in order to be able to distinguish between 'more' and 'less' important variables. There are two possible ways of selecting such a criterion. The first consists in accepting as a criterion one of the variables $X_1, X_2, \dots, X_n$ (say the variable X_n). In this case, we will speak about ***endogenous*** criterion. The second is equivalent to selecting some additional variable, say X_{n+1}, and letting it play the role of the criterion of the relative importance of the variables $X_1, X_2, \dots, X_n$. This is the case of an ***exogenous*** criterion.

If we denote the variable-criterion ***C*** then formulation of the "problem of weights" shows very clearly that the selection of the sequence of numbers of weights depends heavily on the selection of criterion ***C*** and this, in turn, is a matter of an arbitrary decision. One cannot, therefore, expect that the problem of weights is liable to the unique, optimal solution unless criterion ***C*** has been properly defined. But even if this were already done, there is still much room left for arbitrariness because a selection of weights can be performed in many different ways, for example:

1. Weights based on regression coefficients	Weights determined by means of endogenous criterion
2. Weights based on the concept of capacity of information	
3. Weights based on the factor analysis method	Weights determined by means of exogenous criterion
4. Weights based on the correlation matrix	
5. Weights based on the coefficient of variation	

FORMAL CHARACTERISTICS OF WROCLAW TAXONOMY

The study of the level of socio-economic development can be conducted by the use of two taxonomic methods: the Hellwig's pattern method and the non-pattern method. We present only their basic assumptions. However, we have to take into account there seems to be no limit to the range of varieties of possible metric spaces [Walesiak, 2016b].

1. The set of output data has been assembled to form the so-called ***observation matrix***:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} \quad (2)$$

where:

m — number of diagnostic variables (columns) (j = 1, 2, …, m)

n — number of objects (rows) (i = 1, 2, …, n)

x_{ij} — value of the j^{th} diagnostic variable in the i^{th} objects (i = 1, 2,…, n)

2. Because diagnostic variables have, as a rule, different physical dimensions they cannot be directly compared. To enable such a comparison, the variables have to be normalized by eliminating the effect of units of measurement. It is a prerequisite in Wroclaw Taxonomy to achieve the *comparability* of ***all*** final diagnostic variables. This entails, among others, the necessity to strip variables of their natural units in which the diagnostic variables are expressed as well as to normalize variables. It requires their range of variability to be smoothed.

The analyzed variables are standardized as follows:

$$z_{ij} = \frac{x_{ij} - \overline{x}_j}{s_j}; \text{ (i = 1, 2,…, n); (j = 1, 2, …, m)} \quad (3)$$

where:

z_{ij} — normalized value of the j^{th} variable for the i^{th} object

x_{ij} — value of the j^{th} variable for the i^{th} object

$\overline{x}_j$ — arithmetic mean of variable X_j

s_j — standard deviation of variable X_j.

where:

$\overline{x}_j$ — arithmetic mean of variable X_j, where

$$\overline{x}_j = \left(\frac{1}{n} \sum_{i=1}^{n} x_{ij} \right), \quad (4)$$

S_j — standard deviation of variable X_j, where

$$S_j = \left[\frac{1}{n} \sum_{i=1}^{n} \left(x_{ij} - \overline{x}_j \right)^2 \right]^{1/2} \quad (5)$$

Variables' differentiation has been determined for each initial diagnostic variable, with the variation coefficient being the main criterion. The coefficient is calculated according to the formula:

$$v_j = \frac{S_j}{\overline{x}_j} \text{ (j = 1, 2,…, m),} \quad (6)$$

with $\overline{x}_j$ and S_j as above.

3. The above standardization produced a matrix of standardized values *Z*:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2m} \\ \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nm} \end{bmatrix} \quad (7)$$

Table 1. **The Value of TMD for the European Union Countries between 2000–2013**

Country\year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Austria	0.382	0.374	0.391	0.399	0.444	0.426	0.426	0.381	0.387	0.418	0.447	0.503	0.468	0.462
Belgium	0.387	0.372	0.376	0.370	0.387	0.366	0.354	0.323	0.320	0.351	0.372	0.389	0.368	0.356
Bulgaria	0.020	0.005	0.009	0.019	0.037	0.057	0.063	0.041	0.057	0.072	0.067	0.087	0.062	0.075
Cyprus	0.278	0.273	0.282	0.270	0.264	0.257	0.249	0.225	0.216	0.228	0.254	0.260	0.215	0.195
Czech Republic	0.231	0.240	0.251	0.258	0.261	0.259	0.259	0.244	0.243	0.257	0.256	0.282	0.249	0.241
Denmark	0.432	0.421	0.440	0.437	0.448	0.442	0.440	0.403	0.406	0.426	0.432	0.475	0.439	0.424
Estonia	0.170	0.147	0.155	0.159	0.191	0.201	0.221	0.197	0.174	0.148	0.160	0.260	0.249	0.256
Finland	0.414	0.396	0.408	0.403	0.419	0.413	0.407	0.374	0.388	0.410	0.427	0.492	0.433	0.397
France	0.389	0.388	0.384	0.388	0.377	0.359	0.350	0.318	0.323	0.348	0.361	0.391	0.374	0.389
Germany	0.382	0.366	0.363	0.360	0.369	0.335	0.330	0.311	0.322	0.358	0.382	0.424	0.408	0.411
Greece	0.169	0.176	0.182	0.200	0.183	0.170	0.160	0.136	0.136	0.153	0.135	0.097	0.067	0.059
Hungary	0.196	0.203	0.215	0.226	0.220	0.200	0.193	0.157	0.153	0.167	0.166	0.183	0.165	0.156
Ireland	0.365	0.357	0.369	0.395	0.406	0.418	0.417	0.373	0.345	0.319	0.337	0.375	0.348	0.357
Italy	0.274	0.275	0.274	0.264	0.272	0.246	0.246	0.227	0.219	0.234	0.242	0.181	0.234	0.214
Latvia	0.057	0.063	0.080	0.083	0.104	0.115	0.129	0.133	0.126	0.056	0.056	0.012	0.102	0.127
Lithuania	0.066	0.066	0.083	0.083	0.084	0.104	0.109	0.115	0.101	0.082	0.096	0.249	0.170	0.182
Luxembourg	0.341	0.363	0.350	0.349	0.374	0.356	0.339	0.317	0.292	0.309	0.355	0.457	0.362	0.349
Malta	0.319	0.315	0.317	0.339	0.337	0.326	0.331	0.280	0.278	0.300	0.303	0.404	0.323	0.302
Netherlands	0.451	0.433	0.443	0.453	0.452	0.436	0.448	0.412	0.410	0.431	0.443	0.471	0.450	0.430
Poland	0.129	0.117	0.108	0.101	0.097	0.082	0.085	0.100	0.112	0.161	0.180	0.205	0.189	0.183
Portugal	0.225	0.217	0.206	0.208	0.199	0.,174	0.149	0.115	0.113	0.138	0.153	0.181	0.144	0.131
Romania	0.008	0.010	0.030	0.032	0.041	0.043	0.051	0.043	0.022	0.006	0.009	0.009	0.022	0.032
Slovakia	0.155	0.155	0.165	0.156	0.153	0.132	0.136	0.132	0.137	0.183	0.197	0.249	0.198	0.196
Slovenia	0.248	0.254	0.256	0.251	0.265	0.239	0.253	0.224	0.228	0.244	0.249	0.282	0.255	0.242
Spain	0.257	0.262	0.257	0.271	0.266	0.276	0.258	0.225	0.172	0.169	0.188	0.189	0.164	0.157
Sweden	0.394	0.372	0.388	0.396	0.402	0.389	0.389	0.369	0.377	0.392	0.418	0.462	0.429	0.428
United Kingdom	0.395	0.389	0.403	0.426	0.433	0.431	0.430	0.382	0.378	0.381	0.376	0.419	0.374	0.364

Source: Głodowska, A. (2016). Multidimensional analysis of social convergence within the European Union countries. *Chinese Business Review*, 15(3), p. 108.

4. Such normalized data can be used for the construction of Hellwig's pattern model. There are many methods of constructing the model. We can define the model as an abstract object characterized by the maximum values of each normalized variable:

$$z_{01}, z_{02}, \ldots, z_{0m}$$

where for:

– stimulants $z_{0j} = \max_i \{z_{ij}\}$, (8)

– distimulants $z_{0j} = \min_i \{z_{ij}\}$. (9)

The Euclidean distance between objects and the identified 'pattern of development' was calculated using the below formula:

$$d_{0i} = \sqrt{\sum_{j=1}^{m} \left(z_{ij} - z_{0j}\right)^2}, \qquad (10)$$

where d_{0i} is the Euclidian distance of i object from pattern model.

5. The next stage is a determination of a taxonomic development measure. The resulting values of d_{0i} are used to calculate the value of Hellwig's synthetic measure of development, as follows:

$$TMD_i = 1 - \frac{d_{0i}}{d_0}, \text{ where:}$$

$$d_0 = \bar{d}_{0i} + 2s_d.$$

TMD_i is a taxonomic measure of development for i object

$\bar{d}_{0i}$ is an arithmetic mean of the Euclidean distance

s_d is the standard deviation of the Euclidean distance

$TMD_i \in [0;1]$.

Because TMD values are between 0–1, it means that values closer to the 1 present a higher level of development of the object. As an example, in table 1 we present results of such calculations conducted by A. Głodowska (2016).

INSTEAD CONCLUSION

It is too early to summarize. We have to present proposition associated with the time and space. A dynamic taxonomy is still in its infancy. Some studies have been based on historical series and on a static view of countries. It seems if the analysis is repeated over time, it is possible to achieve a dynamic view (movements) of the economic, technological, and social performance of each country, which can provide important information to evaluate the effectiveness of economic and research policies. However, it does not facilitate the identification of the performance and strategic *behaviour* of countries. A full accounting of time and space improves the analysis based on the multivariate approach [Młodak, 2006]. Taxonomic methods seem to be extremely useful for the spatial studies. The other important question is the modelling (forecasting) of socio-economic phenomena with the use of taxonomic methods [Cieślak, 1976, 2001; Grabiński, 1984; Heilpern, 2014]. There are also some unsettled questions, as, for example, a generalization of distance measure [Walesiak, 2016a], normalization of diagnostic values [Kukuła, 1996, 1999, 2000]. In recent times taxonomic methods are used in the financial sector, for example, in the analysis of stock exchange [Łuniewska & Tarczyński, 2006, 2007; Jajuga, 2000] to group similar investment vehicles and to construct sectorial stock market indices.

References

1. Borys, T. (1978). Metody normowania cech w statystycznych badaniach porównawczych [Standardization methods of variables in comparative statistical researches]. *Przegląd Statystyczny, 2*.
2. Cieślak, M. (1976). *Modele zapotrzebowania na kadry kwalifikowane* [The models of the demand for qualified personnel]. Warszawa, Poland: PWN.
3. Cieślak, M. (Ed.). (2001). *Prognozowanie gospodarcze. Metody i zastosowanie* [Economic forecasting. Methods and Application]. Warszawa, Poland: PWN.
4. Florek, K., Łukaszewicz, J., Perkal, J., Steinhaus, H. I., Zubrzycki, S. (1952). Taksonomia Wrocławska [Wroclaw taxonomy]. *Przegląd Antropologiczny*, Vol. XVII.
5. Głodowska, A. (2016). Multidimensional analysis of social convergence within the European Union countries. *Chinese Business Review*, *15*(3), 103–114. doi: 10.17265/1537–1506/2016.03.001
6. Hellwig, Z. (1967). *Procedure of Evaluating High-level Manpower Data and Typology of Countries by means of the Taxonomic Method*. Toward a system of quantitative indicators of components of human resources development. Study III, SS/2836/1711. UNESCO, Methods & Analysis Unit, Department of Social Sciences, Paris.
7. Hellwig, Z. (1968). Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju i strukturę wykwalifikowanych kadr [Application of the taxonomic method for the typological division of countries according to their level of development and structure of qualified personnel]. *Przegląd Statystyczny, 15*(4), pp. 307–327.
8. Hellwig, Z. (1968). *On the optimal choice of predictors*. Toward a system of quantitative indicators of components of human resources development. Study VI, COM/WS/95. UNESCO, Methods & Analysis Unit, Department of Social Sciences, Paris.
9. Hellwig, Z. (1969). *On the problem of weighting in international comparisons*. Toward a system of quantitative indicators of components of human resources development. Study VII, SHC/WS/234. UNESCO, Methods & Analysis Unit, Department of Social Sciences, Paris.
10. Hellwig, Z. (1972). *The Selection of a Set of 'Core' Indicators of Socioeconomic Development*. Social Science Project on Human Resources Indicators. Study XXI, SHC/WS/274. UNESCO, Methods & Analysis Unit, Department of Social Sciences, Paris.
11. Hellwig, Z. (1972). *Approximative methods of selection of an optimal set of predictors*. Toward a system of quantitative indicators of components of human resources development. Study XVI, SHC/WS/215. UNESCO, Methods & Analysis Unit, Department of Social Sciences, Paris.
12. Hellwig, Z., & Gospodarowicz, A. (1975). Zastosowanie analizy porównawczej w badaniach międzynarodowych [The use of comparative analysis in international studies]. *Z prac ZDSE*, Vol. 83. Warszawa, Poland: GUS.

13. Hellwig, Z. (1978). Zastosowanie wieloczynnikowej analizy porównawczej do oceny działalności gospodarczej przedsiębiorstw [The application of multivariate comparative analysis to assess the economic activities of enterprises]. in *Materiały konferencyjne „Metody taksonomiczne i ich zastosowanie w badaniach ekonomicznych"* [Conference proceedings "Taxonomic methods and their application in economic research"]. Szklarska Poręba, Poland, 25.10.1978.
14. Hellwig, Z. (1981). Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych [Multidimensional comparative analysis and its application in studies of multivariate business objects]. In W. Welfe (Ed.). *Metody i modele ekonomiczno-matematyczne w doskonaleniu zarządzania gospodarką socjalistyczną* [Economic-mathematical methods and models in improving management of the socialist economy]. Warszawa, Poland: PTE.
15. Gostkowski, Z. (Ed.). (1972). *Toward a system of human resources indicators for less developed countries: a selection of papers prepared for a UNESCO research project*. Warsaw, Poland: Ossolineum, Polish Academy of Sciences, Institute of Philosophy and Sociology.
16. Gostkowski, Z. (1975). The evolution of developmental gaps between rich and poor countries, 1955–65: a methodological pilot study. *International Social Science Journal*, Vol. XXVII, 1, pp. 38–52.
17. Grabiński, T., Wydymus, S., & Zeliaś, A. (1983). *Metody prognozowania rozwoju społeczno-gospodarczego* [Methods of forecasting of socio-economic development]. Warszawa, Poland: PWE.
18. Grabiński, T. (1984). Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych [Multidimensional comparative analysis in the study of the dynamics of economic phenomena]. *Zeszyty Naukowe Akademii Ekonomicznej w Krakowie*, Seria specjalna: Monografie, 61.
19. Grabiński, T., Wydymus, S., & Zeliaś, A. (1989). *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych* [Methods of numerical taxonomy in the modeling of socio-economic phenomena]. Warszawa, Poland: PWN.
20. Heilpern, S. (2014). Multivariate measures of dependence based on copulas. *Mathematical Economics*, *10*(17), pp. 17–32.
21. Hwang, C., & Yoon, K. (1981). *Multiple attributes decision making methods and applications*. Berlin, Germany: Springer.
22. Jajuga, K. (Ed.) (2000). *Metody ekonometryczne i statystyczne w analizie rynku kapitałowego* [Econometric and statistical methods in the analysis of capital market]. Wrocław, Poland: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
23. Jurkowska, B. (2014). The Federal States of Germany — Analysis, and Measurement of Development using Taxonomic Methods. *Oeconomia Copernicana*, *5*(3), pp. 49–73. doi:10.12775/OeC.2014.019.
24. Kolenda, M. (2006). *Taksonomia numeryczna. Klasyfikacja, porządkowanie i analiza obiektów wielocechowych* [Numerical taxonomy. Classification, rating and analysis of multivariate objects]. Wrocław, Poland: Wydawnictwo Akademii Ekonomicznej.
25. Kukuła, K. (1996). Problem wykorzystania nominant w wielowymiarowej analizie porównawczej [The problem of using a nominant in the multivariate analysis comparative]. *Zeszyty Naukowe Akademii Rolniczej w Krakowie*, Seria: Ekonomika, *25*.
26. Kukuła, K. (1999). Metoda unitaryzacji zerowanej na tle wybranych metod normowania cech diagnostycznych [The method of zero unitarization and the other normalization methods of diagnostic variables. *Acta Scientifica Academiae Ostroviensis, 4*, 5–31.
27. Kukuła, K. (2000). *Metoda unitaryzacji zerowanej* [The method of zero unitarization]. Warszawa, Poland: PWN.
28. Łuniewska, M., & Tarczyński, W. (2006). *Metody wielowymiarowej analizy porównawczej na rynku kapitałowym* [Methods of multidimensional comparative analysis on the capital market]. Warszawa, Poland: PWN.
29. Łuniewska, M., & Tarczyński, W. (2007). *Statystyczna analiza potencjału ekonomiczno-finansowego spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie* [Statistical analysis of the economic and financial potential of companies listed on the Warsaw stock exchange]. Wrocław, Poland: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
30. Malina, A. and Zeliaś, A., 1997: O budowie taksonomicznej miary jakości życia [The Construction of Taxonomic Measure of Quality of Life]. In: K. Jajuga, K. & M. Walesiak (Eds.). *Klasyfikacja i analiza danych. Teo-*

ria i zastosowania [Classification and analysis of data. Theory and application] (pp. 238–265). Taksonomia, Vol. 4. Wrocław, Poland: Prace Naukowe Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.

31. Mesjasz-Lech, A. (2010). A comparative analysis of the development of sustainable energetic resources in Poland with relation to other EU countries. *Advanced Logistic Systems,* Vol. 4.
32. Młodak, A. (2006). *Analiza taksonomiczna w statystyce regionalnej* [Taxonomic analysis in regional statistics]. Warszawa, Poland: Difin.
33. Nowak, E. (1990). *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych* [Taxonomic methods in classification of socio-economic objects]. Warszawa, Poland: PWE.
34. Panek, T. (2009). *Statystyczne metody wielowymiarowej analizy porównawczej* [Statistical methods of multivariate comparative analysis]. Warszawa, Poland: Wydawnictwo Szkoły Głownej Handlowej.
35. Pawlewicz, K. (2005). Differences in development levels of urban gminas in the Warmińsko-Mazurskie voivodship in view of the main components of sustainable development. *Bulletin of Geography. Socio-economic Series*, *29*, pp. 93–102.
36. Pietrzak, M. B., & Balcerzak, A. P. (2016). Assessment of Socio-Economic Sustainability in New European Union Members States in the years 2004–2012. In M. Papież & S. Śmiech (Eds.). *The 10th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings* (pp. 120–129). Kraków, Poland: Foundation of the Cracow University of Economics.
37. Pluta, W. (1977). *Wielowymiarowa analiza porównawcza w badaniach ekonomicznych. Metody taksonomiczne i analizy czynnikowej* [Multidimensional comparative analysis in economic research. Taxonomic methods and principal component analysis]. Warszawa, Poland: PWE.
38. Pluta, W. (1986). Wielowymiarowa analiza porównawcza w modelowaniu ekonometrycznym [Multidimensional comparative analysis in econometric modeling]. Warszawa, Poland: PWN.
39. Sej-Kolasa, M. (2009). Taxonomic Methods in Environmental Management, *Economics & Sociology*, *2*(2), pp. 114–122.
40. Sneath, P. H. A. (1995). Thirty years of numerical taxonomy. *Systematic Biology, 44*(3), pp. 281–298.
41. Sneath, Peter H.A., & Sokal, Robert R. (1963). Numerical taxonomy. The principles and practice of numerical classification. San Francisco, CA: W.H. Freeman and company.
42. Steinhaus, H. (1957). Mathematical Snapshots. Oxford, England: Oxford University Press.
43. United Nations Development Programme. (2017). Human Development Report 2016. Human Development for Everyone. New York, NY.
44. UNESCO. (1972). *Distance-based analysis, numerical taxonomy and classification of countries according to selected areas of socio-economic development*. Social Science Project on Human Resource Indicators. Study XIX. SHC/WS/237.UNESCO, Methods & Analysis Unit, Department of Social Sciences. Paris.
45. UNESCO. (1976). The use of socio-economic indicators in development planning. Paris, France: The Unesco Press.
46. Walesiak, Marek. (2016a). *Uogólnidna miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R* [Generalized distance measurement GDM in multivariate statistical analysis with the implementation of R programme]. 2nd enlarged ed. Wrocław, Poland: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
47. Walesiak, M. (2016b). Wybór grup metod normalizacji wartości zmiennych w skalowaniu wielowymiarowym [Selection of normalization methods of the values of variables in multidimensional scaling]. *Przegląd Statystyczny,* Vol. LXIII, 1.
48. Zeliaś, A. (2000). *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym* [Taxonomic analysis of spatial differentiation of living standards in Poland in the dynamic aspect]. Kraków, Poland: Akademia Ekonomiczna w Krakowie.