CHALLENGES TO EUROPEAN STATISTICS' CREDIBILITY - WHY EUROPEAN STATISTICS IS STILL NOT TRANSPARENT AND ITS CREDIBILITY UNSATISFACTORY

Vasily M. Simchera[1], Ali Serhan Koyuncugil[2]

ABSTRACT

Despite the recognized catholicity and the credibility of European statistics in general (with some exceptions of implemented projects) and in individual European countries still raises number of issues and doubts.

At the same time following the Quételet law, which states the more crimes the less penalties, European statistics today as the most advanced, intellectual and informative one still accumulate doubts and issues towards itself. These doubts, issues and challenges grow in geometric progression and first of all it concerns its credibility, which appears to be its basis and nerve.

Why European statistics credibility is unsatisfactory and tends to go down? Why its transparency and people's confidence are down? At last, why it stirs up the social tensions instead of opposite action in the enlightened European society?

Is it because the European statistics regardless of its virtual image in real life still not transparent, practices double standards, "feeds" society with unnecessary methodologies instead of clear and available information and rather misleading than helping? Or is it because the modern statistics in general and almost everywhere by its nature is dependent and serves authorities and not the truth?

Is it possible to change such situation to the better, and if it is when and on which conditions?

---

[1] Professor, Academician of the Russian Academy of Economic Science, Head of Statistics Division of the Custom Union of the EurAsEC.

[2] The original version of paper can be found in Chapter 3 of the book "Surveillance Technologies and Early Warning Systems: Data Mining Applications for Risk Detection" edited by Ali Serhan Koyuncugil (Capital Markets Board of Turkey, Turkey) and Nermin Ozgulbas (Baskent University, Turkey). Published by Information Science Reference, Hershey - New York, USA (an imprint of IGI Global) http://www.igi-global.com/chapter/nature-scales-statistical-estimations-divergence/46804.

General condition of success for any research is the convergence of theoretical assumptions to the facts being observed, and vise versa the facts being observed to the theoretical assumptions. No matter whether we talk about inductive or deductive researches, determined or undetermined facts and theirs cause – effect connections, natural scientific or socio-economic researches, reliable or less reliable data – the condition always remains the same.

Convergence of theory and practice, forecasts and facts, their adequacy or inadequacy are set by identifications of existing in nature and familiar to science of theoretical and empirical distributions. Convergence at the level of the necessary and sufficient conditions in contrast to abstract ideals is being checked by known criteria of statistical agreement (or in case of its absence) by trial-and-error method and likelihood criteria and common sense. This is the way as many centuries ago the imaginations and theories is tested by facts and practices, and in turn facts and practices is tested by theory and imaginations. And there is nothing else that human mind could invent either in the past nor today. Where a theory relies on the facts and facts relies on theory it is possible to carry out a statistical experiment, which can and indeed gives a significant results with clear sense and paramount scientific and practical importance, and where does not – such an experiment is impossible, and there is no point to initiate this experiment as its results will be false. Unfortunately we should state here that a first case (rather in natural science than in public science) by various reasons, and mainly for a general reason of the limit of knowledge and resources of their realization, is restricted and ultimate, and each successful experiment is interpreted as unique success, while a second case, because of the ignorance of the law of the limited knowledge, is not restricted and infinite. As a result we have domination in a science of simple and mainly false surveys results which are worthless and insignificant, and as a general consequence – depreciation of the efficiency of the science and knowledge's, and total ignorance of them.

The correction of a general situation perhaps requires not only changes in the existing unsatisfactory market treatment, and correction of the negligible treatment towards fundamental researches and labor-intensive experimental results, but raising the systematic level of the knowledge production itself and its application according to the exact form of the identified processes and events in a way their endogen necessity in the world around us. This is the cause of scientific experiment stagnation and further the science as a whole. We should not criticize the external circumstances but think on how to clear a science itself from futile imitation accumulated during centuries, reconsideration of the statistical experiment basis – this is what one should start from and what indeed can help to correct the unsatisfactory situation in the modern science. This means that phenomena and events in the world around us their content, dynamic and structure should not adapt to a format of scientific experiment that is its usual paradigms, algorithms and interpretations but in contrast – the format of scientific experiment itself its set of ideological potential must constantly alter and adapt to a world around us, to catch and produce a future tendencies its fast going and

dominantly differently directed and therefore contradictive changes and give it a shape of essential construction which helps not only for better understanding but also transform our world in efficient and reasonable ways.

Conceiving the situation this way one should start from fundamental basis of the modern scientific experiment, its theoretical hypothesis the basis of which is multivariate statistical distributions and their approximating functions and laws. Depending on how full and certain these functions and laws reflect the structure and dynamic of modern world, the tendencies of its alteration, so this is a degree which define how these functions and laws are applicable today in order to influence the modern events in a constructive way, providing each time the possibilities for effective decision making. The constructive answer to this question demands reconsideration of the whole variety of the existing types of univariate and multivariate distributions their inventory, adaptation and identification applicable to a modern problems solving for production, labor and life. The first step towards obtaining such an answer is typology of present (or in any case most commonly used) distribution functions and laws of observed phenomena and their systematization with regard to problems solving of multivariate distributions as most important and significant ones. There are a lot of distribution functions, but they are separated and can not be united. Part of these functions (linear, normal, power-series, exponential and others distribution functions) acceptable on the level of required and sufficient conditions are approximating observed empirical facts, and have clear substantial interpretation of parameters and sense of the results being received on their basis and these functions are widespread. The other part (the most part) of these functions (logistic, maximum likelihood, and in particular nonlinear and nonparametric distribution functions) are less proved theoretically, and inadequately or even does not reflect the existing empirical distributions and require as a rule robust work while their identification and appliance of them to the problem solving of multivariate analysis. There is also one, some kind of off-balance part of multivariate distribution functions, aimed towards chaotic, partly robust and partly non parametric and further fuzzy distributions, which as a rule lack of any theoretical basis and badly or no way approximating the respective empirical data observed and demand a development of substantially different approaches to their construction. Here we can attribute the synthesis combinatorial problem of distribution function, constructing the unknown hybrid functions on the basis of the existing known ones. In view of exceptional complexity these two tasks are just mentioned here. The overcoming of the existing gaps between theoretical and empirical multivariate distribution functions assumes the typology's ground and representation of these distributions and functions by their resolving powers which are defined by attributes and criteria of necessary and sufficient equivalence. Nowadays one of the most popular ways of using multivariate distributions is data mining. Therefore, data mining has the same gaps mentioned.

It is possible to review data mining under two separate headings as Statistics and Information Technologies. At the beginning of data mining concept discovered in 1990's IT perspective mostly underlined but in 2000's analytical or statistical point of view of data mining has been becoming most integral part of data mining. Hastie et al. (2001) emphasized the statistical view of data mining with statistical learning concept. Rao (2001) linked statistics and data mining with more strong ties and mentioned data mining as a future of statistics. On the other hand, Moss and Atre (2003) mentioned the difference between Classical Statistics and data mining. Data mining mostly interested in big data sets. Therefore, finance is one of the most suitable implementation areas of data mining because of the huge data produces with transactions. Kovalerchuk and Vityaev (2002) inspected the implemantation domains of finance in data mining with examples and emphasized that stock prices, currency rates and bankruptcy predictions, claim management, customer profiling and money laundry some of the implemantation domains of data mining. In addition, data mining successfully applied to financial performance and distress prediction. Koyuncugil and Ozgulbas (2006a) emphasized the problems of Turkish SMEs and suggested financial profiling as a first step of solutions. Then, the authors determined the financial profiles of 135 SMEs listed in Эstanbul Stock Exchange (ISE) according to 2004 ISE data. Koyuncugil and Ozgulbas (2006b) defined a financial performance measure for Turkish SMEs with 2004 ISE data. Koyuncugil and Ozgulbas (2006c) defined the factors which effected financial failers of ISE listed SMEs with 2000-2005 data via CHAID (Chi-Square Automatic Interaction Detector) decision tree algorithm which is one of the most update data mining methods. Ozgulbas and Koyuncugil (2006) and Ozgulbas et al. (2006) determined weak and strenght sides of SMEs in financial mean, financial performance level with 2000-2005 data via data mining. One of the most efficient facilities of data mining is Early Warning Systems because of its definition. Data mining aims to discover hidden relations, covered patterns and then use them for prediction of future behaviours. This definition of data mining makes it the most efficient tool for Early Warning Systems. Therefore, many recent studies in early warning domain has been using data mining. Koyuncugil (2006) developed an early warning system for manipulation and insider trading detection in Stock Exchange Markets and proved that the system works successfully with real data. Simchera (2003a) is given examples of systematic errors, which are made at improper identification of direct and reversed numbers modules, recursive and discursive rates, direct and reversed exchange rates as well as effective interest rates, annuities and bills with recourse, and also numerous examples on collecting and publishing various estimates 9such as GDP, life standards etc.) on the basis of the same methods and estimates with the application of different methods. The module increase direct number for example consumer price index with value of 1.25 is 0.25, whilst module of reversed number for example index of consumer inflation would be 0.2 (1/1.25), in everyday practice it is considered to be equal to 0.25 module, which is not right and in turn illustrates as demonstrative as possible

the most spread example of typical routine "misspell" in statistics that is in other words – to judge anything with the same value and present the same thing with different numbers. And from this quite small "misspell" grows fraud in statistics. Keen and Smith (2007) are given with exceptional arguments the examples of fraud in the estimates of British tax calculations; in the working paper of Ruhashyankiko and Yehoue (2006) and in working paper by Mauro (2002) authors argue the corruption in the private sector of economy and its effect on economic growth reduction. The persuasive estimate of debt dynamics and imbalances are given in the working paper by Meredith (2007) and measurement of financial market liquidity is considered in the working paper of Sarr and Lybek (2002). Booth et al. (1999) based on excellent models the authors illustrate the insolvent insurers estimations for portfolio investments, funds and exchange rates, derivatives, pension funds. In the work of Vaitilingam (2007) there are examples of such estimate, in particular examples ob divergence of estimates for actuarial indices, based on sample observations (covering 30 companies - Dow Johns Indices up to 500 companies – Standards and Poor's Index) and mass observations (covering thousands of companies (400 million companies all over the world). In general the methodology of distortion and publishing of unreliable estimates and explanations of cause and sources of fraud in statistics are given in significant work by OECD (2003, 2008).

What we need is to look at the nine types of distributions within two categories (category of linear and category of nonlinear distributions). The adequate choice of which is the first condition for an efficient minimization of fundamental estimation discrepancies in the modern statistics and arising on these basis deep delusions and undisguised lies.

Without such identification all statistical estimations (and further expectations, forecasts etc.) are shifted and have only illustrative meanings. Nowadays, estimation word coincides another concept different from statistics. This new concept is data mining. There are a lot of definitions of data mining. Because data mining is an evolutionary area. One of the most common definition: 'Data mining is the process of extracting previously unknown, valid and actionable information from large databases and then using the information to make crucial business decisions (Cabena et al., 1997).' Data mining is not a single step analysis. Data mining is a sequential multi analysis and multi task process. But, mainly the core of the whole data mining process is called data mining.

On the other hand, the process takes place as 'Knowledge Discovery' as well. Data mining step of the knowledge discovery process mostly means modified multivariate statistical analysis methods. These methods automized, scaled to analyze huge data sets via modification. Therefore, it is possible to define data mining as an evolution of statistical methods via Information Technologies and automated processes. Data mining mainly has two different point of view: 1. Statistical, 2. Information Technologies. Data mining from statistical point of view calls 'Statistical Learning'. Statistical learning let us know what data tells

instead of subjective sayings. Data mining, automatic extraction of predictional strategic knowledge is mostly based on multivariate statistical methods. Discovery process aims to extract valuable knowledge from hidden, covered, unknown patterns or relations. Patterns usually imply two concepts as similarities and dissimilarities. Profiles, specifications, identifications and unique properties can be determined by dissimilarities or divergence. In addition, divergences can be play a key role for determination of early warning signals for anomalies, errors and fraud. Therefore, divergence is not one of the concerns of only statistics but data mining too. Yet there is also more powerful reason for divergence of the existing statistical estimations which are discrepancies in the conceptual scaling reflecting different ideologies of perceptions of the same phenomena. Especially in relief such discrepancies project at the joint of various world view sciences and epochs sometimes containing in the same definitions and classifications and even in symbols and measures completely different indicative view and value sense. The phenomenon in question comes out visually and instructively at their most on the example of widening the gaps between statistical information and socio-economic disinformation which are in temporary crisis circumstances go beyond all sensible apogees, turning even old time lie into the definite value. What has really happened here, what causes, except sheer self-interest which give rise (this process continues) to such terrible condition of socio-economic information? From technological point of view we have to deal today what on the one hand the capacity and flows of economical information for the period of globalization (1991-2008) would increase tenfold. On the other hand – for a mentioned period the credibility and quality of published data and consequently a quality of decisions made of their bases decreased drastically. As a result some countries and world as a whole instead of expected transition from manual control to automatic one have got into manipulative management. In total of published data the volume of primary data has significantly reduced a lot of valuable and demanded information are consumed by commercial classified information and corruption, we have lost many sources and spheres of primary data, the level of data comparability had fallen, virtually the possibilities for testing them for convergence, precision and credibility are taken out of public access. Information is dominantly collecting for information itself that is why it works with reduced efficiency. The Internet is a convincing example which shows that it servicing itself and its providers for 97, 5% and only for the rest 2, 5% it is commercial network and information service. The understanding of that the people and the worlds nations need not information typologies but information itself is substituted and lost. How to turn the contemporary situation for the better, to overcome the accumulated information gaps, misbalances, obstructions and through this to provide freed space and resources for collection, processing and distribution of credible information on the basis of which one only can provide crucial improvement of total socio-economic situation in the countries to organize the real transition from today's reforms to the reforms of forthcoming efficient transformation. Simchera (2003b) is given general answer to this question in his

work. Below with use of additional information we give following concise answer to this question. There are a lot of economical estimations but they are uncoordinated and mainly doubtful and incomparable. The gaps between existing indicators (due to conceptual differences of their understanding and coverage, material divergence in prices, exchange rates etc.) reach sometimes multiple values and strike off any possibility to its wise application. All socio-economic parameters gathered and published today are pretty often characterizing mainly not real situation in the world but represent only calculation effect, and in many cases even less – that is statistical calculations errors. On the basis of such deliberately incorrect and even false data one should not make that conclusions and assumptions which are made in modern statistical science. The science loses even more in its potential due to data manipulating practices which became rather standard than exception, substitution of one parameters for the others, aberration of not only size and meaning but the sense of the phenomena in question, that is what confirm the today's crisis estimations. There is no other brunch of the science where measurements would be more contradictive, estimations divergent and results just useless for a practice, than it is in the economic science. This is why it is logical not accidental that much of published data should be taken out of 'circulation' as useless. Consequently one should not understand all of the above as total result which shows that with this level of divergence, incorrectness there is less point to calculate and publish statistical data than not to. This fully concerns to calculations and publications of market statistical data which not only distort but also misrepresent all normal perceptions about real situation in modern economy, and virtually turn all published estimations into economical phantoms and threaten of credibility loss to all conclusions and values of modern science. Subjective evaluations, interpretations and conclusions generally can be very far away from what data tells us objectively. In addition, the data can be directed, dirty, missing, unstandardized and uncomparable. Therefore, there are some necessary steps for acceptable decision making process: 1. Data must be realiable. 2. Analysis must be suitable for the aim (s) and the data. 3. The evaluation of the results of the analysis must be correctly 4. Conclusions must be in an objective manner In case of missing at least one of the necessary steps given above, then, the gap between estimations and the realizations will be statistically or scientifically unacceptable.

The Gap Between Estimations And Realizations With Numerical Examples

Here are examples illustrating gaps in economical estimations, which directly point to the sources of bad informational situation in the world and from our point of view due to this fundamental reason – situation of modern economical science. According to official statistics GNP growth in the USA in constant prices

of 2000 (with regard to statistical calculation errors) in 2007 in comparison with 2000 accounted for 1,18 times (increase 2,3% per year); (in 1991-1999 – 1,18) for the period 1981-2007 – it would be 1,93 points (2,4% per year); national wealth's estimation of growth are 1,16; 1,18 and 1,72 points respectively. These are actual estimations of growth. And here are the figures – phantoms which substitute the quite humble figures against their background of real rates of the USA economical growth. The growth of the wide spreading Dow Jones Index (share rates of 30 largest companies in the world) for the same period accounted for 1,22; 3,8 and 7,1 times (the points fixed for this index at the end of 1990 were 2892, and at the end of 1999 – 10 991; 2007 – 13 368; June 2009 – 8 501), the Standard and Poor's Index (500 companies) – 1,3; 2,1 and 19,0 times (the points fixed for this index at the end of 2003 – 1 109; 2007 – 1 479; June 2009 – 923) respectively, and the growth for Nasdaq (5 000 technology companies) – only from the middle of 1996 has exceeded 153 times (the points fixed for this index at the end of 2003 – 2 010; 2007 – 2 654; June 2009 – 1 796). In other G-7 countries the gaps between real rates and phantom – rates are approximately the same. In England in particular with the GNP growth for the period 1991-1998 in 1,25 times increase (for the period 1981-1998 - 1,49 times increase) the stock exchange FTSE index showed growth of 2,7 and 9,1 times. The exception was France where while GDP growth for the period 1991-1998 was 1,13 times its stock exchange CAC – 40 index showed growth of "just" 2,7 times. Even if we take estimations of GDP's of G-7 countries in current prices, the gaps between them and stock exchange increases would stay fantastic and destroy representations of any dignity. In addition, here are some others kind of phantoms. The real indicators of the USA's GDP with most favorable calculations – today would be $14,0 (in 2003 – $11,0; 1990 – $5,8) bill. doll. Indicators of national wealth are $28,3, $24,8 and $14,8 bill. doll. respectively; considering real financial assets $39,5, $33,2 and $22,4 bill. Doll (U.S. Census Bureau, 2009). The capitalization value of USA's companies (price multiplied by quantity of issued shares) on financial markets estimated, before crisis 2008, for $200-$215 trill., annual turnover of shares – $100-$120 trill. The crisis only for a last half of 2008 and first half of 2009 has devaluated these assets to 40-50% and yet it is going to devaluate them to 30-35% in forthcoming months, equalizing them with estimations of real assets, which turns all fund market in the USA into a phantom. The phantoms are also modern indicators for banks assets. The real assets for 1000 largest banks are estimated for $20 (in 2003 - $12,7) trill. The derivatives (fictitious capital) of all world's banks exceed $400 (in 2003 - $100) trill., while their equity capitals account for only $30-$25 and $19,5-$18,5 (in the USA -$14-$10) trill. respectively. On the assumptions of estimations of U.S. Federal Deposit Insurance Corporation (2007) consolidated real assets of commercial banks (at the end of 2007 – 7 282 banks in the USA), investment institutions or investment banks (total number 1 251 at the end of 2007) and credit unions (8 101 at the end of 2007) excluding mutual benefit societies which would not exceed $14 till. (to be precise $13.792,5 bill.) in

the pre-crisis year, including assets of commercial banks - $11 176,5 bill., and credit unions - $753,4 bill. (National Credit Union Administration, 2007). And these quite credible estimates in comparison with analogues estimates on assets of some separately taken American banks are also appearing to be phantoms. The allied assets of JP Morgan accounted for $97,5 trill. (62,5% over the world's GDP), while its real equity capital accounted for $2,5 trill. (2,6% of the assets). Respectively estimations of Goldman & Sachs are $50 trill. (83,3% of world's GDP) and $1,5 trill (3% of assets). Even the assets of largest bank holding HSBC (Hong Kong-Shanghais Bank, Corp.) are phantoms, the fictitious assets of which (in 2000 accounted for $6,5, and in 2008 - $108 trill.) exceeds the equity capital for almost 10-50 times more (Bank of Russia, 1999). According to market rates virtually all modern transnational companies including American pride – Ford and biggest world's bank – HSBC which real capitalization in good times would exceed trillion of US dollars – are bankrupts, whilst according to leading rating agencies they still have higher ratings than actually solvent companies in China, Brazil, India or Russia. Against the background of real GDP's volume accounted for about $60 trill. in 2008, it is absurd how estimates of fictitious capital look like, the guaranteed part of which in the same year would overrun $6*10^{14}$ that is $600 trill., while unguaranteed part $3*10^{15}$ that is $3 quadrillion, including the USA with $175 and $900 trill. At last the phantoms are the inadmissibly different by module, scope and contents indicators of GDP, inflation, national wealth in various countries. Finally, it is just ridiculous how in the form of phantoms – marginal's are presented today the audition estimations for capitalization of world's leading companies which have fallen drastically at once to 7-10% from the initial value following the world's crisis. The mixing of own capitals with attracted ones, real values with fictitious values, the parameters mentioned above and many other parameters of financial assets cause quite serious concerns and according to specialists' opinion require not only cardinal revision but also international legislative control. The above examples illustrate not only destructive but also demoralizing role of market representations about fundamentals of modern economical life and processes of economic developments of different countries and nations. They cause apprehension in a whole world and persuasively point at necessity of fundamental revision of all modern practices of economical estimations, and a new approach to the principals of economic measurements. Especially loud voices for this decision are coming out from the USA which is a motherland of financial pyramids and financial bubbles the burst of which can destroy not only the US but the rest of contemporary financial and economical system. When Alan Greenspan was speaking at the White House on 5th of April, that were unheard appeals to study statistics of world's financial markets, scrutinize and reveal the consequences of the growth of financial bubbles in the world's economy. The way out from this situation one can see in revision of core structure of a whole variety of market expectations their conventional agreements and standardization of their methods for comparative calculations. For provision of

practical realization of this way out the authorities should preliminary to take concrete solutions on the following basis:

- on the basis of the international standards of calculation and account one should organize the inventory and fundamental revaluation of all international and national assets and on this base one should take all fictitious assets out of balance turnover which dilute real representations in economy and excludes in principle the possibilities to coordinate it on a fair basis;

- instead of floating exchange rates one should introduce fixed ones which reflect real interaction of prices by all line of produced and consumed products, services and capitals;

- one should find corporative indices and fund market ratings as useless and cancel them due to they distort situation with global market estimations;

- on the principally new basis should be constructed and introduced into international turnover the single world's indices, which would representatively reflect the dynamics of world's fund and foreign exchange rates markets;

- on the basis of transparency and free accessibility, public watch and personal responsibility, with compensation of damage and lost profits one should reconsider the principles of IMF and World Bank work, and in case of failure to implement the liabilities – one should develop conditions to seize their activity as the international institutions – regulators of the international financial relationships;

- under the auspices of UN one should establish an International commission on causes qualifications and regulation of world's financial crisis consequences with the functions to determine a size of material loss and moral damage caused by the monetary authorities of the US and other countries – satellites due to failure to take the necessary measures for prevention of their negative consequences and rights to discover in accordance with international law all persons guilty for breaking the rules of financial business leading to a global default, mass bankruptcies which exceed all calculated losses of almost every country and nation in all world wars. The presented examples – are documentary approved facts of the informational distortion. However in today's practice the economy has to deal not with distorted information but with outspoken call to world's economical society what is in legal parlance called fraud, in the form of forgery, which is done by betrayal of trust – the deeds which are qualified as crime and penalties are severe. This could be a decision.

Conclusion and Future Studies

To overcome the differences in modern economic conditions and eliminate the existing unjustified in many cases multiple and even by a factor of ten divergences in economical estimations which distort economical representation

and excluding possibilities of their efficient practical use in the development of the international standards one should organize and implement special work on the correction of fundamental divergences in the economic estimations, restitution of incomparable parameters, standardization of methods of their calculations and publications on comparable basis and degrees of credibility claimed beforehand. In other words one should create a single number of standards of economical information and economical indicators based on total conventional decisions. All information in the world, all economical or rather all other indicators must be built and published on this basis. There is no country which must not and cannot be exception unless it claims to be included into the existing borders of the single world socio-economic space. The conditions and consequences of its activity and life first of all should be transparent and explicit in estimations. A "non-transparent" country cannot and has no moral rights to claim transparency from other countries and first of all from its own citizens. In this connection a total international standardization for all multiplicity of mutually demanded socio-economical parameters shall not mean the automatic unification of every possible other multiplicities, which are characterizing an objectively existing national flavour and specificity of development for every single country. On the contrary the modern standardization should initially take into account the necessity for an existence, maintenance and multiplication of various numbers of national indicators considering them as fundamental origins and solid ground for their own development and existence. That is why what has been done today in the sphere of international standardization of national indicators is only the first level not the top of the enormous work which must be done in every single country and in a whole world in behave of increase of their convergence and guaranteed provision for that level of the mutual subject understanding in the world which is only possible on the basis of the international standardization of the national indicators. Thus it is rather proposed than excluded that all preceding work of the International Statistical Institute (ISI), UN's League of Nations and other international organizations which has been made during XIX-XX centuries in the field of international standardization with regard to the stated position should be reconsidered and all used in the modern world international standards should be re-standardized. Perhaps against the background of the statistical standards flaws revealed by a crisis the international organizations should initiate and perform this very work. And data mining can be the main tool for this work. The big picture given above shows the need of clean, comparable and standardized definitions instead of directed ones for acceptable estimations and reliable conclusions. Clean, comparable data and standardized definitions lead us the datawarehouse logic or data mining infrastructure or basis. Indicators, proportions, ratios, rates and the other kind of statistics can be the first signals in case of evaluating them in a proper way. We should note that all the raw data, processed data must be clear, transparent and has equal definition for reliable comparison. Therefore, existing data stock can be collect with the same definition and standarts. Actually, UN has standard definitions for indicators, methodologies

and data preparation. Eurostat (European Union Statistical Office) has standards as well for European Union member and candidate countries. Furthermore, Eurostat statistical standards are mostly a subset of UN standards and widely there are no contradiction between UN and Eurostat standarts. Therefore, it is easy to merge UN and Eurostat data with some little justifications. UN country data infrastructure can become the base for unified statistical indicators datawarehouse. Coordinator can be the ISI as mentioned above for much more updated scientifically standards. Furthermore, practice of counteraction against legalization of statistical divergence can be use for occupational objectivity.

Therefore, practice of disclosure, critical assessment and correction of statistical divergence being revealed. The effective tools of counteraction are law on legal responsibility and legal protection for statisticians (currently not in use).