University for Business and Technology in Kosovo

UBT Knowledge Center

Oct 28th, 9:00 AM - Oct 30th, 5:00 PM

Assessing quality and use of economic data: case of Kosovo

Mentor Geci
University for Business and Technology, ubtstats@ubt-uni.net

Kaltrina Bunjaku
University for Business and Technology

Teuta Dervishi
University for Business and Technology

Armira Fetoshi
University for Business and Technology

Assessing quality and use of economic data: case of Kosovo

Mentor Geci, Kaltrina Bunjaku, Teuta Dervishi,
Armira Fetoshi, Mirora Limani

UBT – Higher Education Institution, Lagjja Kalabria, 10000 p.n.,
Prishtine, Kosovo (UBT STATS)
ubtstats@ubt-uni.net

Abstract: Quality of economic data is of key importance in real assessment of economic situation in a country. Among the challenges in Kosovo in terms of economic statistics is their quality and consistency. Study will examine the quantity and quality of economic data with particular emphasis on macroeconomic data. International statistical standards of the International Monetary Fund, European Central Bank / Eurostat are used as reference for comparison regarding the quantity and quality of economic data. For the study purposes, the economic data will be divided into sectors according to international standards. Official statistics producers will be used as data inputs as specified in the respective law in Kosovo. Data quality are reliability, transparency, accuracy are criteria to be used. Statistics revision policy or procedure as well as time-series and accessibility will be part of the study also. UBT STATS established in June 2013 deals with the collection, organization and publication of statistical data and emerged as a need of having a structured database consistent with international practices. An overview will be devoted to UBT STATS also as center for statistics, modelling, and economic forecasting.

Keywords: data quality; macroeconomic statistics; international statistical standards; data consistency; transparency; data gaps.

1. Introduction

Data quality is multidimensional concept starting from data collection, compilation, validation, publication. As contrary to the developing economies where statistical infrastructure is in place and functions according to the specified system, in case like Kosovo, developing a statistical framework is still issue in process.
In first part special focus is devoted to the theoretical background in terms of frameworks for data quality. Credible international standards are used, mainly those of International Monetary Fund and Eurostat. After describing the theoretical background, the assessment to the Kosovo statistics is made using UBT STATS framework, broadly consistent with the international standards. With the main focus on comparability and consistency, data in this study are grouped according to specific sector using international statistical standards, namely, government statistics, consumer price index, etc. Data are sourced from certain publishing institutions, further processed by UBT STATS.

2. Assessing the data quality – theoretical background

Data quality framework is a system for assessing data quality by developing certain standards and criteria. Countries use individual evaluation standards to assess data quality but also at international level such as International Monetary Fund, European Central Bank, Eurostat and other relevant institutions standards. The main focus of the theoretical aspect of the evaluation will be given to international standards and principles at the same time representing standards to be applied simultaneously to the relevant member states and that in certain cases represents structural criteria for membership. In terms of individual empirical papers also, most of them addresses issues such as data consistency, data quality, data reliability, etc.

At institutions level, International Monetary Fund has created separate standards in this regard. A data quality assessment framework is created including so many elements starting from top level (dimensions) down to the level of specific indicators. Based on International Monetary Fund standards (*IMF Data Quality Assessment Framework – IMF DQAF 2003*), data quality assessment is decomposed into : (i) *prerequisites of quality*, (ii) *assurances of integrity*, (iii) *methodological soundness*, (iv) *accuracy and reliability*, (v) *serviceability*, and (vi) *accessibility*.

Quality assessment framework (IMF DQAF 2003) in terms of *prerequisites of quality* starts with legal and institutional environment by specifying the responsibility for collecting, processing, and disseminating the statistics. In most of the countries it is specified by respective law on official statistics. In case of Kosovo, according to the Law on Official Statistics, there are three producers of official statistics, namely: Statistical Agency of Kosovo, Central Bank of the Republic of Kosovo, and Ministry of Finance, as well as other bodies as authorized from by Statistical Agency of Kosovo. Resources and relevance represents second and third element within the prerequisites of quality.

Professionalism, transparency as well as ethical standards represents main cornerstones regarding the *integrity assurance*. Professionalism in this regard is decomposed further into professional principles through sound statistical policies and practices, statistical techniques (models), dissemination, and what is very important proper interpretation and avoiding the misuse of statistics.

After the legal setup and assurance of integrity, *methodological soundness* prevails. First and most important element in terms of methodology is consistency of concepts and definitions in line with international accepted statistical frameworks. In other case, if there is difference in definitions of particular statistics in relation to international standards, data are not comparable to other countries and might cause improper interpretation of real situation. The particular importance is given also to the scope of statistics and proper sectorization in terms of institutional sectors of economy and as well as instruments. More developed financial systems are, more complex is to proper sectorize institutional sectors and instruments, as well as to measure it. Basis for recording is another element that fall within the methodology, like market pricing, accrual basis accounting, gross/net issue, are some of the times to be taken into consideration.

Accuracy and reliability measured through source data, sound statistical techniques, and sound validation of the data are key elements regarding the accuracy and reliability of the data. As producer of official statistics, assuring comprehensive and timelines in obtaining the data is key precondition. Assessing the source data in periodical and ad-hoc basis is inevitable process. Sample process, coverage, sampling and non-sampling errors should be assessed. Applying certain statistical techniques when it comes to data collected from sample (physical persons) is part of the statistical task. Apart from the data validation, statistical revision is common. Since the main goal of statistics is to represent the reality, statistical revision should be made. Developing statistical revision policy or guideline and having it published it makes the entire process more transparent and justifies the data revision. Worth mentioning one case that statistical revision is regular process, for example, every year due to the

audited financial statements from the external auditor, might be changes in the companies' balance sheet and income statement that should be reflected in produced and published statistics. This should be covered in statistical revision policy.

As IMF specifically states, statistics with adequate periodicity and timeliness, are consistent and follow a predictable revisions policy that in fact represents the *serviceability* of the data. IMF specifies also dissemination standards, namely, periodicity and timeliness.

Accessibility representing last dimension within the overall framework of data quality deals with presentation in a clear and understandable manner, forms of dissemination are adequate, and statistics are made available on an impartial basis. Key point in this picture is documentation of processes and making it accessible to external parties. By documentation of processes is meant documentation on concepts, scope, classifications, basis of recording, data sources, and statistical techniques is available, and differences from internationally accepted standards, guidelines, or good practices are annotated.

In terms of sectoral breakdown, according to the International Monetary Fund it includes seven statistical sectors, such as: national accounts statistics, consumer prices index, producer prices index, government finance statistics and public sector debt statistics, monetary statistics, balance of payments statistics and international investments position statistics.

Handbook on Data Quality Assessment Methods and Tools of Eurostat (Eurostat 2005d) aims at facilitating implementation of data quality assessment in the European System of Statistics by providing a detailed description of the data quality assessment methods. The Eurostat specifies quality indicators to be followed by the European System of Statistics across member countries that includes six such indicators, as follows: (i) *relevance*, (ii) *accuracy*, (iii) *timelines and punctuality*, (iv) *accessibility and clarity*, (v) *comparability*, and (vi) *coherence*.

Eurostat and European System of Statistics in 2011 published European Statistics Code of Practice for the national and community statistical authorities. It covers responsibilities for statistical authorities and quality standards. Nevertheless, in terms of authorities it specifies relevant issues are professional independence, mandate for data collection, adequacy of resources, quality commitment, statistical confidentiality, and impartiality and objectivity. In terms of quality standards it specifies principles as: *professional independence*, *mandate for data collection*, *adequacy of resources*, *quality commitment*, *statistical confidentiality*, *impartiality and objectivity*.

Except it is addressed at institutional level, there are also several authors approaching the data quality issues. Van den Berg et al. (1999) within the framework of the data quality uses Life Cycle Assessment as long term approach that is based on components such as: reliability and validity. In general, most of authorities addresses issues of quality, timelines, accuracy, reliability, etc. Zhu et al. (2014) concludes that looking ahead, data quality research will continue to grow and evolve. In addition to solving existing problems, the community will face new challenges. Research is needed to develop techniques for managing and improving the quality of data in these new forms.

3. Quality assessment framework for Kosovo – UBT STATS approach

As depicted in the previous section, there are several frameworks and criteria used for quality assessment of data. However, worth mentioning that IMF's but also Eurostat and other frameworks are addressing broadly same issues starting from legal basis up to the dissemination elements. Kosovo is internationally classified as developing economy based on main economic indicators (GDP per capita, etc.) and quality of data is a debating issue. UBT STATS that functions as a statistical center within

the UBT that among others tasks deals with the data quality issue developed its framework based on international statistical standards.

In order to have better view of the UBT STATS approach for data quality assessment, in the table below will be shown international standards along with the approach used for assessing the data quality for Kosovo statistics.

Table 1. Data quality assessment frameworks

| IMF *(Data Quality Assessment Framework)*	Eurostat *(Handbook on Data Quality Assessment Methods and Tools of Eurostat)*	Eurostat and ESS *(European Statistics Code of Practice)*	UBT STATS
• *Prerequisites of quality;* • *Assurances of integrity;* • *Methodological soundness;* • *Accuracy and reliability;* • *Serviceability;* • *Accessibility.*	• *Relevance;* • *Accuracy;* • *Timelines and punctuality;* • *Accessibility and clarity;* • *Comparability;* • *Coherence.*	• *Professional independence;* • *Mandate for data collection;* • *Adequacy of resources;* • *Quality commitment;* • *Statistical confidentiality;* • *Impartiality and objectivity.*	• *Legal basis;* • *Quantity and quality;* • *Timelines;* • *Relevance;* • *Revision policy;* • *Accessibility;* • *Data organization;* • *Consistency;* • *Time series.*

There are 12 statistical modules covered so far by UBT STATS, namely: national accounts statistics, government finance statistics, labor market statistics, labor force survey, enterprise survey, consumer price index, import price index, production price index, energy statistics, hotelier statistics, education statistics, and health statistics. So, there will be matric presentation using the above mentioned framework into the statistics modules as presented above. In order to have better view of the quality assessment of the data done by UBT STATS, the radar presentation will be used. Five level grades are used, starting from 1 to 5, where grade 5 represents best grade.

Graph 1. UBT STATS data quality assessment

National accounts statistics	Government finance statistics
Labor market statistics	Labor force survey
Enterprise survey	Consumer price index

Import price index

Production price index

Energy statistics

Hotelier statistics

Education statistics

Health statistics
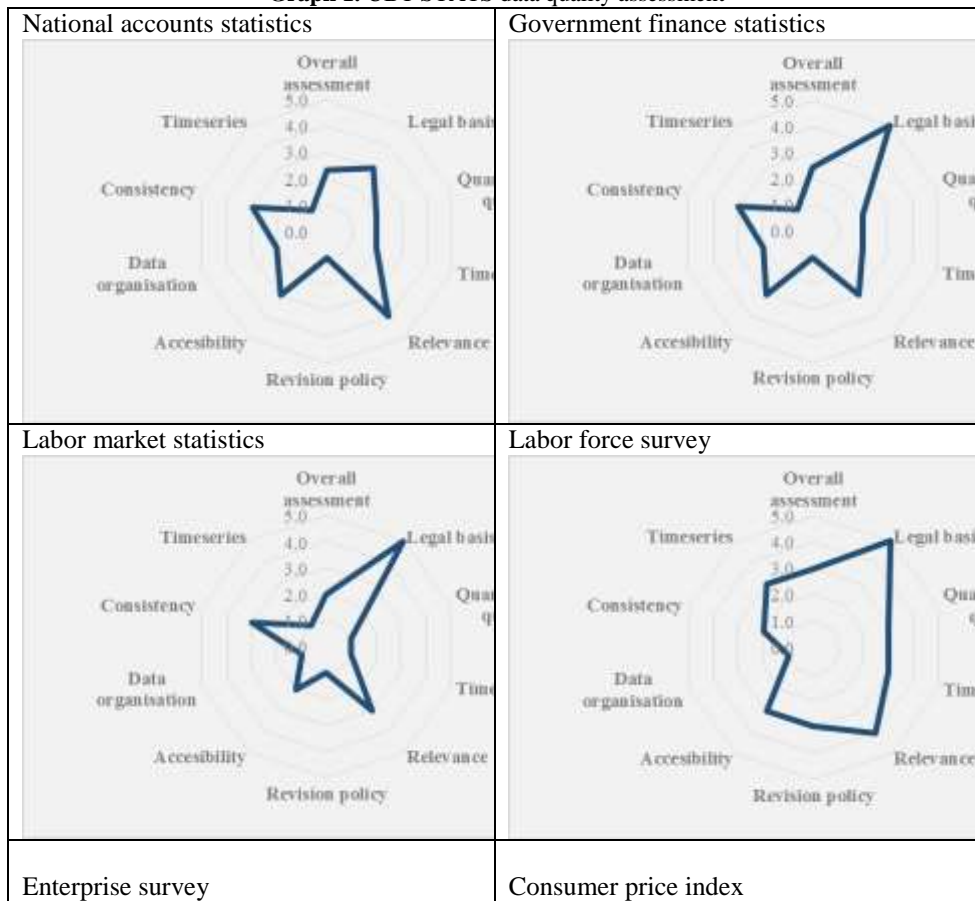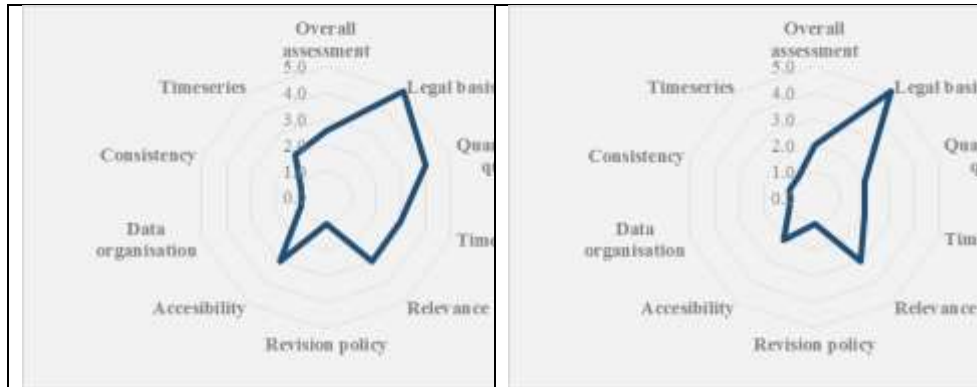
To this end, as presented in the graphs, most of the issues in terms of data quality are: no time series, lack of consistency, absence of revision policy, frequent revision of the data, frequent change in the presentation format, no time series available, lack of information or no information for methodological changes, frequent changes in the presentation format, delays in reporting, absence of publication calendar, totaling issues, data permutation, numbers repetition issues across several periods, accessibility issues, etc. In order to have comprehensive view of all processed modules, below is shown aggregated presentation using simple average by components.

Graph 2. UBT STATS data quality assessment – overall assessment

The results are based on the three years research activity and assessment done by UBT STATS. As shown in the assessment made above, there are lot of issues that UBT STATS addresses in this regard. However, if we would like to rearrange by weight there are three components that needs urgent steps from competent authorities, such as: (i) data organization, (ii) times series, and (iii) consistency.

4. UBT STATS – short overview and its activities

UBT STATS operates as a center of competence for Statistics, Analysis and Forecasting. UBT STATS main value is in the organization of information in the form that they are more structured, easily understandable, and comparable in time and make possible their use for purposes of research purposes.

The planned activities for the period 2016/2017:

Promotion of outputs through the publication on websites. Preparing periodic reports by compiling reports on the latest developments in specific areas which will be published regularly on the website of UBT and in any electronic written press. Reports will be focused on specific topics including professional commentaries by UBT STATS.

Individual data processing of institutions. Institutions that are used as source of data only publish data at the aggregate level and not for institutions. Missing information in the Kosovo market are those on individual institutions.

Development of research with the public. Considering available resources, UBT STATS plans to utilize students to do various research through surveys. They may include all fields but the focus will be based on actual themes that emerge. A case study may be expectations on the movement of prices, which compiled often as the survey in other countries regarding expectations on the development of prices.

Incorporating research conducted by students on specific topics. Using data from internal research by students of UBT is a very good source of information we plan to incorporate on UBT STATS. This refers to subjects which have as part of their curriculum conducting research as part of the final assessment.

Incorporating information from conferences and similar. Besides many useful information from students' research, it is planned also to incorporate information from conferences organized by UBT.

Development of new modules and periodic update of information. In order to make possible the preparation of periodic reports, regular updating of information is necessary. The update is an activity that consumes time and resources also because of disorganized inputs.

Consolidation of tables and variables. In certain modules there are many tables repeated and have almost the same story. Already after the first phase of data collection has been completed, remains to be done consolidation of them so each of them to have its own story. Consolidation means elimination of unnecessary and useless items.

References

1. Abdullah, N., Ismail, S. A., Sophiayati, S., & Sam, S. M. (2015). Data Quality in Big Data: A Review. *International Journal of Advances in Soft Computing & Its Applications*, *7*(3).

2. Bergdahl, M., Ehling, M., Elvers, E., Földesi, E., Körner, T., Kron, A.,. & Sæbø, H. V. (2007). Handbook on data quality assessment methods and tools. *European Commission, EUROSTAT, Wiesbaden*.

3. Cao, H., Hill, D., Juster, F. T., & Perry, M. (2005). Enhancing the Quality of Data on Income and Wealth. *Michigan Retirement Research Center Research Paper No. WP*, *101*.

4. Carson, C. S., Khawaja, S., & Morrison, T. K. (2004). *Revisions policy for official statistics: a matter of governance* (Vol. 4). International Monetary Fund.
5. Eckerson, W. W. (2002). Data quality and the bottom line. *TDWI Report, The Data Warehouse Institute*.
6. European Central Bank: Transparency [https://www.ecb.europa.eu/ecb/orga/transparency/html/index.en.html]
7. Eurostat: G20 Data Gaps Initiative (DGI) – background [http://ec.europa.eu/eurostat/statistics-explained/index.php/G20_Data_Gaps_Initiative_(DGI)_%E2%80%93_background]
8. Friedman, T., & Smith, M. (2011). Measuring the business value of data quality. *Gartner, Stamford*, *464*.
9. Haug, A., Zachariassen, F., & Van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, *4*(2), 168-193.
10. Instituto Nacional de Estadistica: Revision policy of the National Statistics Institute (2015)
11. International Monetary Fund: Data Quality Assessment Framework (DQAF), Washington D.C. (2012)
12. Jmp. Statistical Discovery: Compare Incomplete or Mismatched Data [http://www.jmp.com/support/help/Compare_Incomplete_or_Mismatched_Data.shtml]
13. Juster, F., & Smith, J. (2004). *Improving the Quality of Economic Data: Lessons from the HRS and AHEAD*. EconWPA.
14. Krippendorff, K. (2004). Measuring the reliability of qualitative text analysis data. *Quality & quantity*, *38*, 787-800.
15. Mazzi, G. L., Cannata, R. R.: A proposal for a revisions policy of Principal European Economic Indicators (PEEIs). European Commission. Eurostat (2008)
16. Meeker, W. Q., & Hong, Y. (2014). Reliability meets big data: opportunities and challenges. *Quality Engineering*, *26*(1), 102-116.
17. Morgan, S. L., Waring, C. G.: Guidance on Testing Data Reliability. Office of the City Auditor (2004).
18. Stecklov, G., & Weinreb, A. (2010). *Improving the quality of data and impact-evaluation studies in developing countries*. Inter-American Development Bank.
19. The Statistics Newsletter. G-20 Data Gaps Initiative (DGI): a second phase by the IMF and the OECD (2015)
20. van den Berg, N. W., Huppes, G., Lindeijer, E. W., van der Ven, B. L., & Wrisberg, M. N. (1999). Quality assessment for LCA. *CML Report*, *152*.
21. Wada, H., Fekete, A., Zhao, L., Lee, K., & Liu, A. (2011, January). Data Consistency Properties and the Trade-offs in Commercial Cloud Storage: the Consumers' Perspective. In *CIDR* (Vol. 11, pp. 134-143).
22. Zhu, H., Madnick, S. E., Lee, Y. W., & Wang, R. Y. (2014). Data and Information Quality Research: Its Evolution and Future.