Statistical Data Quality Control Study under the Conditions of Direct Reporting for Internal Statistics Package by Network

Li Keqin*

Hubei Provincial Bureau of Statistics, Wuhan, Hubei, China, sywy131@126.com, 430071

Abstract

In recent years, sometimes the statistical data of Chinese government has been questioned in quality while concerned. Then it expedites a revolutionary change, that is, to implement the construction of four major statistical projects with the core of direct reporting for internal statistics package by network. The implementation of direct reporting for internal statistics package by network has realized the unified design, unified arrangement and unified determination of survey plan design, task arrangement, unit determination and data acquisition as well as national direct data acquisition, synchronous sharing, high degree of inter-professional integration and highly controllable statistical production, promoted unified regulations, enhanced survey efficiency, reduced the burden on enterprises and improved data quality. However, compared with the traditional production mode of statistical data, there have been new forms of expression for data quality problems under the conditions of direct reporting by network. With the principal line of gradually advancing direct reporting for internal statistics package by network and the construction of four major projects, this paper starts with the analysis of the nature and reasons for data quality problems under the conditions of direct reporting by network, and puts forward the data quality control concept and specific measures in combination with the effective exploration and practice made by Hubei Provincial Bureau of Statistics in terms of controlling the direct reporting data quality according to the idea of suiting the remedy to the case, hoping to cause the attention of the majority of experts, scholars and statistical colleagues to discuss the fundamental strategies of statistical data quality problems and strive to promote the government statistics of "three improvements".

Key Words: construction of four major projects, data accuracy, statistical method system

1

1. Introduction

The Chinese government has basically established a modern statistical system adapting to the socialist market economy system, fully using international statistical standards for reference and meeting the needs of China's economic and social development. However, the statistical data quality problems have been closely linked with the statistical development like shadows. The heavy and hard statistical reform task is under unprecedented enormous pressure. Therefore, since 2009, the National Bureau of Statistics has determined and made strenuous efforts to promote the construction of four statistical projects with the core of internal statistics package reform, that is, to turn pressure into motivation, seize the "ox nose" of improving data quality, comprehensively enhance statistical capacity, better meet the needs of the party, government and the social public for statistics and fundamentally improve the statistical credibility of the government.

Through the unified organization of the National Bureau of Statistics in March 2012, 700,000 "three above" enterprises and real estate development and operation enterprises directly reported the statistical data to the data center of the National Bureau of Statistics successfully through the computer network (referred to as "internal statistics package reporting by network") according to the internal statistics package system and requirements and achieved a complete success in the Annual Report 2011 and the Monthly Report of February 2012. It can be seen from the smooth implementation of the direct reporting for the internal statistics package and the subsequent official operation for more than six months that some problems existing in the statistical data quality before have been resolved to some extent, but some new situations and new problems are also revealed at the same time and must draw great attention from the national statistical system.

Hubei Province is one of the few provinces in the country launching pilot projects of internal statistics package reform in the province since 2009. The reform of the Provincial Bureau of Statistics reflects the characteristics of "early intervention, comprehensive pilot, deep participation and model in development". Hubei Provincial Bureau of Statistics has carried out some useful exploration for the statistical data quality control problems under the conditions of direct reporting for internal statistics package. This paper attempts to make a general display of the data quality control research achievements under the conditions of direct reporting for internal statistics package for more than three years from three aspects.

2.Characteristics and Problem Analysis

2.1 Implementation of direct reporting for internal statistics package and its innovation characteristics

Through pilot, the National Bureau of Statistics convened the national internal statistics package working conference in September 2011 to deploy the comprehensive implementation of internal statistics package in the country from Annual Report 2011 and fixed report in 2012. The national direct reporting system of internal statistics package was opened on February 18, 2012. In the reporting time limit, 700,000 "three above" enterprises and real estate development and operation enterprises in the country successfully submitted the data of the Annual Report 2011 and the monthly report of February 2012 to the National Bureau of Statistics through the direct reporting system by network. The initial victory of the direct reporting by network marks the revolutionary transformation of the statistical production mode of Chinese government and reflects the monumental leapfrogging in the development process of Chinese statistics.

Fully implement the direct reporting for internal statistics package, establish the survey unit library of "three above" enterprises and real estate development and operation enterprises and basically form the survey unit library updating and maintenance mechanism timely updated

2

and maintained and ensuring that the unit information is true and reliable; standardize the name, connotation, calculation methods, classification standards and statistical codes of the statistical indicators and form the statistical survey system of the internal statistics package of "three above" enterprises and real estate development and operation enterprises; develop the software system with the functions of unified arrangement, acquisition, processing and storage meeting the statistical surveys of "three above" enterprises and real estate development and operation enterprises; establish the direct reporting system by network meeting the statistical surveys of "three above" enterprises and real estate development and operation enterprises. All direct reporting enterprises have installed CA authentication security system and the construction of four major statistical projects has made decisive stage achievements.

The implementation of direct reporting for internal statistics package by network has realized the transformation of survey plan design from independent design by various professions to unified design, the transformation of survey task arrangement from distributed arrangement by all professions to unified arrangement, the transformation of survey unit determination from self-determination by all professions to unified determination, the transformation of original data acquisition from indirect acquisition by the country and the reporting layer by layer by the statistical agencies at all levels to direct acquisition by the country and online synchronization and sharing by the statistical agencies at all levels, the transformation of the enterprise-oriented survey from professional division to high integration and simultaneous proceeding between professions and the transformation from difficult control of statistical process to highly controllable statistical production.

The innovation characteristics of the direct reporting for internal statistics package are mainly reflected in the following five aspects.

The first is to promote unified regulations. Uniformly regulate the index meanings, calculation methods, classification criteria and statistical coding, uniformly regulate survey unit library and survey system, uniformly regulate the work arrangements, data acquisition, transmission and process flow and behavior and generally eliminate the overlapping phenomena existing in survey items and indicators.

The second is to enhance the survey performance. The national survey tasks are directly arranged by the National Bureau of Statistics to the local statistical agencies and enterprises at all levels through the network, eliminating the report printing and issuing layer by layer. The enterprise data is directly submitted to the National Bureau of Statistics through the network, achieving the original data obtained by the country, provinces, cities and counties simultaneously by privilege and canceling the reporting layer by layer. The original data of enterprises is uniformly stored and managed by the country and the professions and regions share the original data according to the authority, greatly improving the data summarization, processing and application performance.

The third is to effectively reduce the enterprise statistics burden. There are a total of 450 indicators in the internal statistics package survey system of "three above" enterprises and real estate development and operation enterprises, 79 fewer than before; the repeated reporting of the same indicator by the enterprise is also eliminated. The enterprises accept all report tasks in once and uniformly participate in training, changing the practice that the enterprises undertake different professional reports from multiple ways and participate in different professional training respectively. The electronic statistical reports are filled in conveniently compared with the paper media reports and many statistical indicators can be automatically generated from the enterprise accounting financial and production business indicators, eliminating the step that the enterprises deliver the reports to the statistical agencies at basic level.

The fourth is to improve the data quality. The improvement of the data quality is safeguarded from the following six levels: first, achieve the comprehensive and accurate grasp of the basic conditions of the respondents on the whole and eliminate the acquired, bankrupt and eliminated enterprises in the survey unit library to effectively avoid the repeat and omission of respondents and ensure the authenticity of the reporting units. Second, achieve the country's centralized management of original data and reduce the possible interference of statistical data in intermediate links to some extent. Third, the indicators reported in the electronic reports will be automatically generated in other reports, ensuring the consistency of the indicators. Fourth, the logical audit relationship is set between the indicators, which are automatically audited by the system and the unsanctioned data cannot be reported, ensuring the connection and matching between the indicators. Fifth, the software system can automatically keep all modification traces of data by enterprises and statistical agencies at all levels to facilitate data query, review and acceptance and to provide a solid basis for determining statistics responsibility. Sixth, various regions and professions strictly execute unified statistical standards, effectively improving the comparability and applicability of statistical data.

2.2 Main factors affecting the data quality in direct reporting by network

2.2.1 Definition of statistical data quality in this paper

It can be said that the statistics is accompanied with topic of statistical data quality since its emergence and the statistical data quality has received much concern. However, there is still not an authoritative definition of statistical data quality internationally and different scholars have different researches on statistical data quality from different points of view. In the "Statistics Encyclopedia", the data quality is defined as "the degree to meet the data user demand and expectations" and its connotation is further decomposed into purposiveness, accuracy, timeliness, consistency, comparability, usefulness and clarity. The Chinese government tends to grasp the connotation if the statistical data quality from three basic principles on the practical level relative to all schools of thoughts on the theoretical level: Applicability, Accuracy and Timeliness.

It shall be emphasized that the above three principles are closely associated with each other and have some contradiction and conflict due to different statistical service functions and different concerns demanded by different users. Even for the same statistical data, different users will put forward different quality requirements: some may focus on accuracy and some may focus on timeliness. Therefore, as the producers of statistical information and providers of statistical service, the statistical agencies shall continue to weigh at various aspects of statistical data quality and find an optimal balance point under the restriction of all conditions to meet customer needs.

This paper discusses the statistical data quality problems of the Chinese government and more focuses on the accuracy of the statistical data, that is, the so-called statistical data inaccuracy or distortion. With respect to direct reporting for internal statistics package by network, the data quality discussed by us mainly refers to data accuracy and is mainly aimed at some problems affecting data accuracy appearing in the data production process.

2.2.2 Main factors affecting statistical data quality under the conditions of direct reporting for internal statistics package by network

Since the full implementation of direct reporting for internal statistics package by network, some data quality problems have also been exposed. For example, the data authenticity and accuracy remains to be further enhanced; some data in direct reporting by network is still unmatched, divergent and inconsistent with relevant index data. Under the current conditions of direct reporting for internal statistics package by network, the main factors affecting

statistical data quality include:

The first is the low degree of adaptability of some enterprises. As the main body of the market economy, some enterprises mainly focus on the production and operation and pay high attention to the issues closely related to the vital interests of the enterprises such as finance and tax. In contract, they fail to fully understand the performance of statistical obligations according to the law and fail to fully implement, resulting in the mismatching between the statistical information quantity and quality, leading to data distortion or acquisition difficulties and increasing the investigation cost of the statistical department.

The second is the inadequate training and guidance strength of statistical departments at basic level for enterprises. The uneven quality of the enterprise statistical personnel, frequent replacement of some enterprise personnel, substandard original records, statistical ledgers and other basic information of enterprise and difficulties of the statistical departments at basic level in ensuring more resources invested in enterprise training and guidance in respect of manpower and material resources result in lack of proficiency of statistical personnel in some enterprises in business and substandard statistical data sources, affecting the authenticity and accuracy of statistical data quality to some extent.

The third is that the statistical systems and methods need to be further improved. In recent years, especially since the implementation of direct reporting for internal statistics package by network, China has done a lot of work in improving and perfecting the statistical systems and methods, but there are still some aspects not adapting to the requirements of social and economic situation development. For example, the statistical survey content cannot better reflect the requirements of the scientific development; the statistical survey is overly dependent on comprehensive reports with more survey content and less information content; the survey items are too many and the basic level burden is heavy; the overlapping surveys exist and the division of labor and cooperative relationship shall be further straightened out.

3. Strategies and Measures

3.1 Concept of data quality control in direct reporting by network

The new situation and new problems appearing in the statistical reform with the core of internal statistics package shall be faced up to and actively solved in line with the principle of being practical and realistic, daring to undertake and being tough and aggressive to steadily push forward the reform. The data quality in the direct reporting by network shall be controlled in line with the following principles:

The first is to adhere to the principle of harmonious promotion of the construction of four major projects. The second is to adhere to the principle of strengthening the statistical grassroots work. The third is to adhere to the principle of strengthening the legal system safeguard of statistical work. The fourth is to adhere to the principle of better serving the respondents.

3.2 Control measures for data quality in direct reporting by network

First, further nurture and promote statistical core value system. The statisticians must have the spirit. The most fundamental is to adhere to the statistical core value and take truth and credibility, scientific rigor, innovation and enterprise, service and dedication as the common value orientation and goal pursuit of all statisticians for the statistical career and as the spiritual pillar and ultimate belief of all statisticians.

Second, further deepen and expand the statistics achievements of the construction of four major projects. First, take pains to further optimize the statistical system design, uniformly

define survey indicators, determine survey methods, develop survey system, design the software program, complete system assessment and actively and steadily put into effect; second, take pains to further reinforce the basic unit directory, increase the storage audit frequency in strict accordance with the statistical standards and the storage audit procedures and intensify the review of units in database; third, take pains to further improve the software and direct reporting system by network, gradually improve the overall function of the software and direct reporting system by network towards the target of more stable system, more user-friendly interface, more convenient use and safer information and meet the needs of various statistical surveys; fourth, take pains to further expand the professional scope of the internal statistics package and prudently and reliably incorporate more professions in the scope of direct reporting for internal statistics package.

Third, further consolidate and strengthen statistical grassroots work. Adhere to the principle of promoting statistical business foundation while developing statistical grassroots construction and conduct harmonious promotion of statistical grassroots construction.

Fourth, further plan and promote the modernization of statistical means. The most important is to vigorously promote the wide application of modern information technology in statistical work and to make full use of computer technology, network technology, space technology and other kinds of modernization promoting statistical survey means.

Fifth, further improve and strengthen statistical legal construction. Actively improve the "Rules for the Implementation of Law of Statistics" and local statistical laws and regulations, departmental and local statistical regulations and other legal systems with the general outline of promoting, implementing and executing the "Law of Statistics" and fully implement statistical legal popularization planning.

Sixth, further value and build harmonious statistical interaction.

References

Wang Sheng, Research on Problems in Statistical Data Quality and Countermeasures, "Modern Economic Information", Article No.: 1001-828X(2010)12-0206-03

Yang Chaoying, Weng Liyu, Discussion on Statistical Data Quality Problems, "Journal of Fujian Agriculture and Forestry University (Philosophy and Social Sciences)", 2003, 6 (1): 46-48

Ren Yufeng, On Statistical Data Quality and its Developing Methods, "Statistics and Management", 2010 (01), 15-16

Luo Jianmei, Reflections on How to Improve the Quality of Chinese Government Statistical Data, "Economic Research Guide", 2009 (19), 06-07