ESCWA STATISTICS

Regional Workshop on the Statistical Business registers for the Arab Countries 26 - 29 September 2016 Amman, Jordan

ورشة العمل الاقليمية حول سجلات الأعمال الإحصائية للبلدان العربية

عمان – الأردن، 26 - 29 أيلول/سبتمبر

Economic And Social Commission For Western Asia

Quality of Statistical Business Registers
UNECE Guidelines on Statistical Business Registers

جودة سجلات الأعمال الإحصائية

UNITED NATIONS

الاسكوا

ESCWA

Wafa Aboul Hosn, Section Chief, Economic Statistics,
aboulhosn@un.org

40 YEARS

Content

- SBR quality dimensions
- SBR specific characteristics
- Quality assessment of administrative registers
- Frame errors and their implications
- Metadata
- Quality assessment methods
- SBR quality indicators
  - Quality of input
  - Quality of processing
  - Quality of output
- Quality policy and improvement

SBR quality dimensions

Quality: the degree to which a set of characteristics fulfils requirements and fit for purpose.

Standard quality dimensions:

- Relevance: data meet current and potential needs of users
- Accuracy: information recorded corresponds to reality
- Timeliness: data provide a picture of the real world /least time lag

Particular quality dimensions to SBR

- Punctuality: When to expect the frozen frames and statistics
- Accessibility: possibility of internal users obtaining individual data by connecting to database, and external users obtaining aggregate tables from the NSI output
- Comparability: same concepts and methods over region and time
- Coherence: internal SBR data; coherence and with other registers creating and maintaining links (use of a common identification code)
- Consideration of cost: as cost can be a quality constraint

SBR's characteristics (distinct from standard surveys)

- Use of administrative data
- Heterogeneity of inputs
- Variability of inputs over time
- Technological requirements
- Primary output is micro-data (individual unit level not statistical aggregates)
- Heterogeneity of users
- Continuous updating of SBR data (not only by actual economic changes but by correction)

Quality assessment of administrative data or SBR

Takes into account the entire process of acquisition, loading and processing of administrative data.

The assessment can be in terms of the standard quality dimensions; or the quality framework developed by Daas et al (2009) based on the hyper dimensions: source, metadata and data

- Source: measures the extent to which information contained in a data source is exploited: frequency of delivery, relevance, relationship with administrative authorities…

- Metadata: focuses on conceptual and process related quality components of the source metadata, administrative regulations (administrative units, definitions of the data they provide, reference time periods, forms used in data acquisition)

- Data: focuses on quality indicators describing the quality of data input to the SBR, technical checks, accuracy, completeness, timeliness and integrability

Frame errors and their implications

- Errors in existence: due to false information regarding demographic characteristics of a particular unit; recorded as economically active but are not in the real world (over-coverage) or economically active but not present in the SBR (under-coverage)

- Errors in identification characteristics: errors in names addresses etc (difficult data collection due to problems locating and contacting units and impede use of SBR as tool to link and coordinate data from different sources

- Errors in stratification characteristics: errors in legal form, economic activity code, size class and geographic location. Results in inefficient sampling and sample allocation and inaccurate population estimates derived from SBR

Frame errors and their implications

Impact of errors on survey frames (which is a measure of accuracy of the SBR)

- **Under-coverage:** the frame does not include all units within the scope; reasons: omissions, errors in activity status, or stratification characteristics; results **in under-estimation**
- **Over-coverage**: the frame includes units that are out of scope; reasons: duplication, errors in activity status, errors in stratification characteristics; results in **over estimation**
- Errors in unit characteristics: i.e. stratification cause inefficient sampling, errors in contact data result in increased non-response and follow-up

Metadata

Source of data: alphanumeric code identifying the source used in assigning the value of a characteristic of a unit

Procedures used for attribution of unit characteristic values:
- Production process: descriptions of the processes for acquisition and processing of administrative and statistical data, SBR profiling and improvement surveys, production of frozen frames and survey frames, production of statistics, integration of administrative registers
- Updating and error correction process: descriptions of updating processes and determining whether changes reflect real world economic changes or corrections
- Reliability of data: include quality indicators and references to the production sources and processes
- Characteristic value updating history: date/period to which the value of a characteristic relates and date last updated in the register

Quality assessment methods

- User survey: to obtain the views of major groups of users regarding each of the quality dimensions
- SBR audit: systematic: independent and documented process for obtaining quality audit evidence and evaluating
- SBR improvement surveys and quality measurement surveys: to improve the quality of unit characteristics, by detecting and correcting errors and filling in missing values; measure the accuracy of the measurement of the quality of characteristics such as activity status, economic activity code and size
- Auditing clerical work: involves checking samples of clerical updates, checks based on regular random representative samples. Quality auditing function linked to documentation and training functions so that issues are identified resolved documented and covered in future staff training
- Macro-editing: monitoring changes in the SBR at an aggregate level, linked to the production and dissemination of frozen frames
- Defining and monitoring quality indicators

SBR quality indicators

Conceptual framework

➢ Key factors

- Time: reference date
- Scope: a particular type of unit and within that type a subset of all possible units
- Sub-population: defined in terms of size, region and economic sector
- Characteristic: of a set of units
- Criteria: to assess the quality of the characteristic, unit by unit; for each, it should be possible to assess whether the value is right/wrong, or where it lies on a scale of quality between 0 and 1

SBR quality indicators

Quality assessment mechanisms and criteria

- Use of external information source: a value of a unit is correct if it is sufficiently close to a reference value from an external source, it focuses on compliance and is a proxy for accuracy when the real value is not known; the compliance rate is a practical substitute for the reliability rate when the latter cant be measured
- SBR improvement survey: quality survey (expensive)
- Internal consistency: a value is correct if it is consistent with the other characteristics of the same unit, based on plausibility criteria
- Temporal consistency: based on a comparison of the quality of a characteristic with its values in previous time periods to identify impossible or implausible changes from one period to another
- Quality without witness: assessment without having a specific reference value; can use the date on which the value was most recently updated (the more recent, the better), the name of the information source, the methodology adopted

SBR quality indicators

Divided into three groups corresponding to processing phases:

- Quality of input
- Quality of processing (throughput)
- Quality of output

SBR quality indicators

Quality of SBR inputs

- SBR has many input sources, mostly administrative. The qualities of the input sources affect output quality
- Quality assessment includes accessibility and clarity of administrative data. Data format and content precisely reported and data accompanied by metadata, examples and advice
- Other indicators: time lag and completeness of characteristic
- Problems that can arise: changes in the source not known to the SBR (such as changes in registration or cancelation rules, classification criteria, administrative control processes). Big changes should be detected and reviewed to determine underlying cause. Main objective is to check the stability of the sources
- Percentage/proportion of variations (per characteristic)
- Indicator for over/under coverage

SBR quality indicators

Quality of SBR processes

SBR processes divided into three phases

➢ Phase 1: Integration of input data from administrative sources (to create clusters referring to the same unit), divided into two sub-phases

- First sub-phase: analyze within a source
  Possible quality indicators:
  ✓ Number of duplicates as a proportion of the total number of supplied record (a decrease over time indicates an increase in quality of the source)
  ✓ Number of new records during a given reference period

- Second sub-phase: link between sources (aim is to identify the set of administrative data records relating to each individual legal unit; typical base used is the taxation register)
  Possible quality indicators:
  ✓ Number of clusters of records in period (t) linked to legal units
  ✓ Number of clusters of records in period (t) not linked to legal units
  ✓ Number of clusters of records in period (t-1) not linked to legal units
  ✓ Under-coverage indicator (due to time lag in the registration of units in the taxation register)

SBR quality indicators

Quality of SBR processes

➢ Phase 2: Assignment of values of characteristics (and identification of active units in year t, because frozen frames and subsequent survey frames contain only active units)

Possible quality indicators:
- ✓ Active status concordance rate
- ✓ Inactive status concordance rate
- ✓ Activity status discordance rate

Sources available for determining /checking activity status:

- Small/medium size enterprise survey- checks for active and inactive units
- Foreign trade survey- checks for active units
- Bankruptcy database- checks for inactive units

SBR quality indicators

Quality of SBR processes

➢ Phase 3: Editing procedures (implementation of the editing and imputation rules used in final identification of the frozen frame for a particular reference period t

Quality indicators measure the errors produced by each rule. Quality checking plan consists of a set of separate quality project modules executed sequentially
✓ Cleaning (rules determining the exclusion of some units from further checks)
✓ Deterministic (if/then clauses causing automatic changes in the values of relevant characteristics whenever the condition occurs)
✓ Errors (rules generating error warnings whenever the condition occurs)
✓ Other useful indicators: looking at trends, measuring the increase or reduction in the number of units that fail each type of edit over time

Quality policy and improvement

❑Quality policy includes

- the decision to measure the quality of SBR
  - ✓ a by-product of measuring SBR quality is the generation of metadata and paradata allowing a better knowledge of the state, content, structure and processes of SBR
- the decision to communicate information about SBR quality
  - ✓ presenting to users quality dimensions and associated indicators
  - ✓ Informing users about significant events related to SBR maintenance (availability of new administrative data, next frozen frame, changes in important units, in classification)
- the decision to initiate a program of improving SBR quality (should be based on a careful analysis of costs and benefits)

Quality policy and improvement

❑ General approach to improving quality

Iterative process based on

- Construction of a set of quality and performance indicators for SBR inputs, processes and outputs
- Setting quality and performance targets
- Defining quality assessment tools

❑ Improving timeliness

- By systematically applying updates available from relevant administrative sources
- By reducing the time required to apply updates so that the changes can be quickly reflected in the survey frames

❑ Improving completeness (constant investigation of new ways to extend the SBR by collecting information about enterprises from additional resources)

Quality policy and improvement

❑ Improving coverage (reducing under-coverage and over-coverage, especially duplication of units)

❑ Improving quality reports (aim should be to deliver short sub-annual quality reports to alert users to significant recent changes in the SBR)

❑ Providing survey report; provide functionality that:
- Allows a survey to monitor its population
- Provides a frame on the basis of which efficient sampling schemes can be designed, panels monitored, and the results from sample surveys can be grossed up to population estimates
- Provides the information for assembling mailing lists, dispatch of questionnaires, monitoring responses, and contacting units in the event of nonresponse.

Thank you

شكراً

Economic And Social Commission For Western Asia