Unit Error in Business Statistics: Towards a Research Programme

Boris Lorenc

together with Daniel Assoulin (SFSO), Wim Kloek (ESTAT),
Li-Chun Zhang (U Southampton)

Statistics Sweden

Statistiska centralbyrån

Outline

- Large impact of unit error → action needed

  ➢ *Two action directions proposed*

- Three approaches: conceptual → empirical → statistical

  ➢ Move from first to second

- Further research ideas

  - Optimal set of unit types: criteria, empirical support

  - Formal description of relations of selected unit types (ontology)

  - Probability distributions on unit characteristics

  - Micro-units (e.g. production factors)

  - Other: Cognitive analysis of the unit creation process in business registers; bottom-up profiling and top-down profiling; additive vs non-additive variables

  ➢ Research these ideas

Large impact of unit errors

- 9% of objects in Statistics Sweden's farm register were in error in a core register variable due to "errors in objects" (Wallgren and Wallgren, 2007:102)
- erroneous equating of enterprise to legal unit leads to
  - 38% of employees associated in business demography to SMEs (other than micro) in fact belong to enterprise groups (from Brion et al., 2014, p.6)
  - errors in produced statistics by size classes: 56% of the value added, 77% of fixed assets and 59% of export sales not allocated to large enterprises (from Brion et al., 2014, p.6)
  - total turnover of 19 large French groups with 30 enterprises and 1226 legal units being 30% higher than when account consolidation based on profiling is done (Brion et al., 2014, p.8)
- probability of error in assigning enterprises <10 empl. to NACE activity classes ranges from 3% to 98%, with an unweighted median of 37% (van Delden and Scholtus, 2014, p.19)

The unit error

- Error in produced statistics due to deviations, systematic and random, from correctly identifying units, their properties and their relations, leading to observation and data collection (or failure to do so) from not fully correctly identified, characterised and delineated, units and unit groups.
  - Propagates to all produced statistics in which the unit is involved
- Earlier/ongoing work
  - In the paper (Richter, Willeboordse, Struijs, Rainer, Archer, College, Pietsch, Lavallée, Carone, Teillet, Brion et al., van Delden et al., Zhang, etc etc etc)

Statistics Sweden

Statistiska centralbyrån

Three approaches

- Conceptual (verbal, rule based)
    - Introduces concepts, typologies, verbal regulations for identifying, relating, and characterising units, informal descriptions of causes of errors; no empirical (numerical) data
- Empirical
    - Collects data that relate to the concepts of the first approach: measures processes, estimates errors, etc; no general framework
- Statistical-theoretical
    - Based on conceptual understanding and collected data, builds statistical models for improving inference of interest, also enabling relating the unit error to the main body of statistical survey theory (e.g. the Total Survey Error approach – improving balance of efforts to control survey errors)

Statistics Sweden
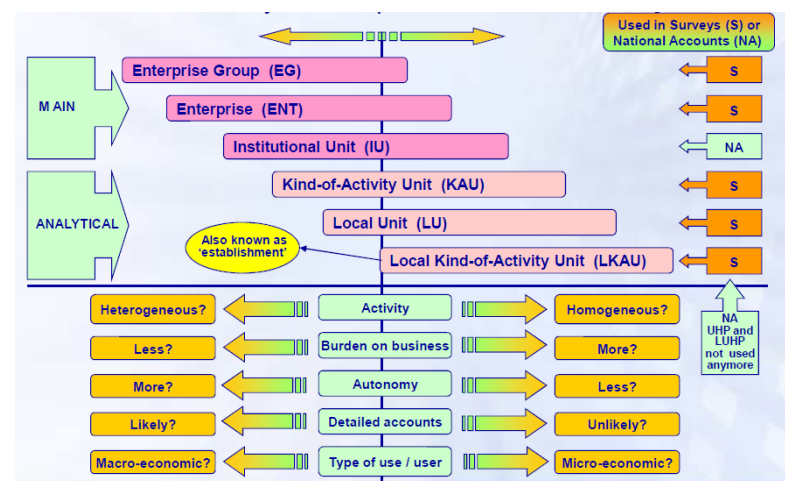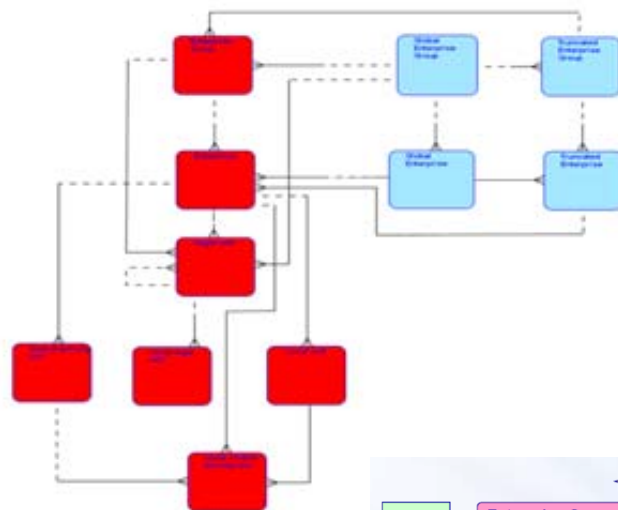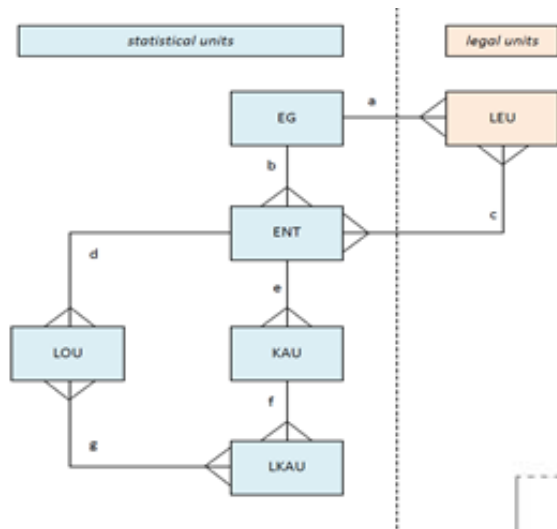
Statistiska centralbyrån

1: Move: Conceptual → Empirical

I. Consistency with methodology of other efforts towards quality improvement (6 sigma, process measurement, etc).

II. Establishment of a fact based decision making process regarding statistical units

III. Without empirical data, it is not possible to work towards integration of unit uncertainty into general survey methodology theory and practice, specifically the total survey error framework

2: Possible directions

- Optimal set of unit types: criteria, empirical support

- Formal description of relations of selected unit types (ontology)

- Probability distributions on unit characteristics

- Micro-units (e.g. production factors)

- Other

  - Cognitive analysis of the unit creation process in business registers (difference admin units/stats units?)
    - typology of errors?
  - Bottom-up profiling and top-down profiling: equivalence of outcomes?
  - Specification of additive variables vs non-additive variables

2: Possible directions

- Optimal set of unit types: criteria, empirical support

- Formal description of relations of selected unit types (ontology)

- Probability distributions on unit characteristics

- Micro-units (e.g. production factors)

- Other
  - Cognitive analysis of the unit creation process in business registers (difference admin units/stats units?)
    - typology of errors?
  - Bottom-up profiling and top-down profiling: equivalence of outcomes?
  - Specification of additive variables vs non-additive variables
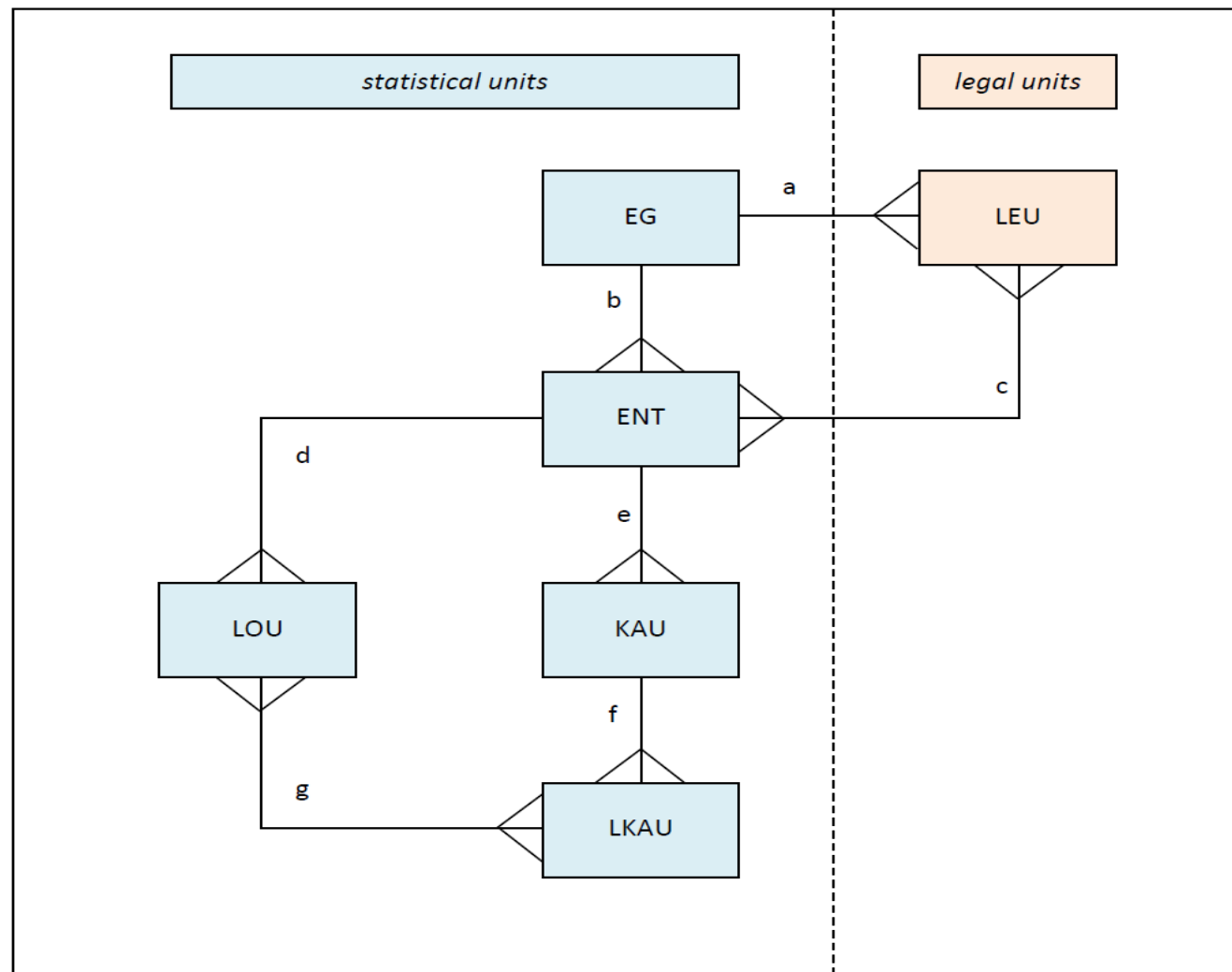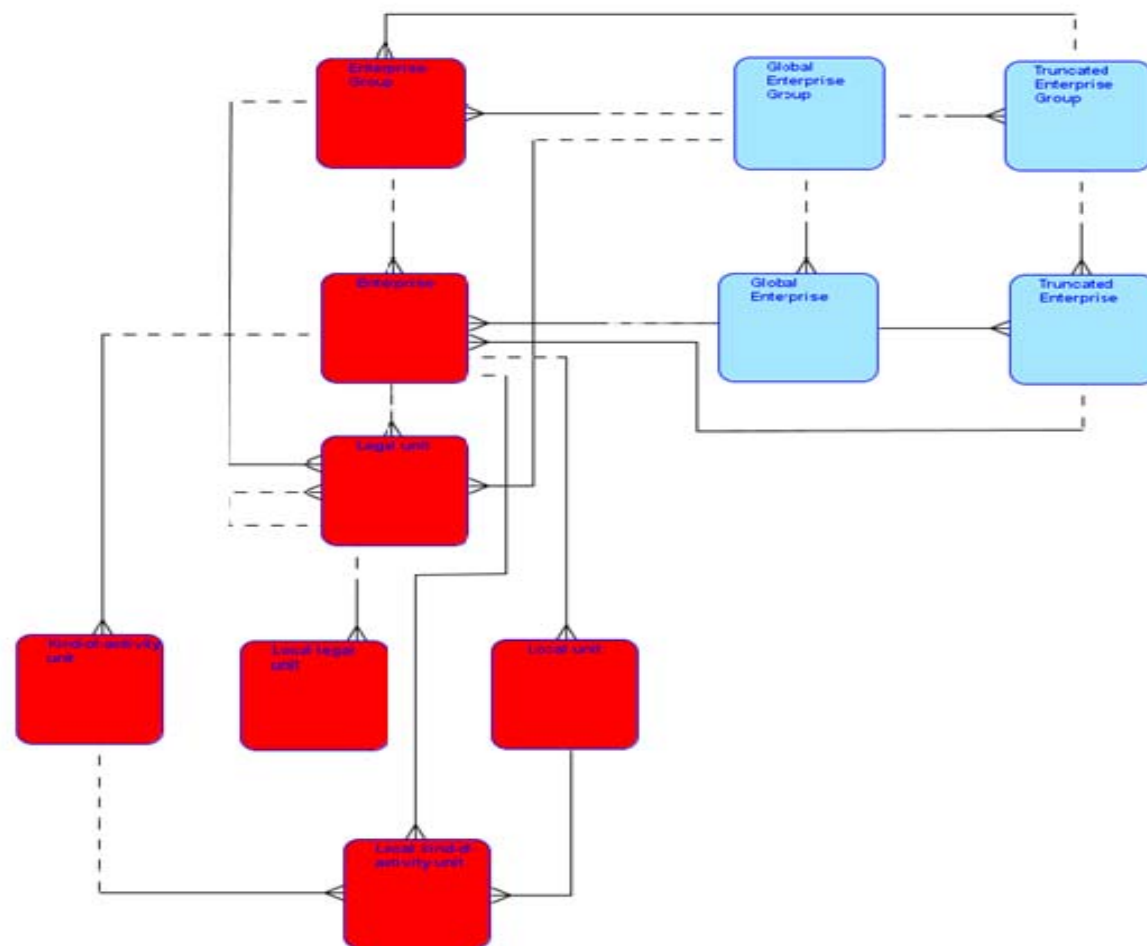
Thank you!

Units in business statistics: EC 696/93

Units in business statistics: EGR

Units in business statistics: TF on SU