Tilastokeskus

Selective editing in Finnish SBS

Tilastokeskus

Basic principles

- Selective editing is used to separate influental observations from non-influental observations
  - Influental -> Interactive treatment
  - Non-influental -> automatic treatment / no treatment
- Influental means units that have high contribution to the estimates at the level of usage
- Local scores are at first calculated on invidual variables of the observation
- Global score (unit level score) is calculated from local scores
- Observations with global score over the set limit are considered influential

Tilastokeskus

Score functions

- A score function usually evaluates either *influence* or *risk*
- The *influence* factor quantifies the relative influence of a record on a publication estimate
- The risk factor quantifies either
  - The extent to which a record violates edits
  - The extent to which it deviates from anticipated values
- Local Score = influence * risk
- Global score is a function of local scores

There is not a right or wrong way to generate score functions.
There is numerous ways. Just be creative, but keep it simple!

Tilastokeskus

Finnish case (1/3)

- Global score for an observation is given on basis of three categories of scores
  - Absolute limits
  - Influence of the observation
  - Influence of errors (failed edits)
  - Changes from previous year
- Global scores are used to distinguish what observation are
  - Manually edited (over 100 points)
  - Automatically edited (under 100 points)

Tilastokeskus

Finnish case (2/3)

- Local Scores
  - Absolute limits
    - Personnel over 100
    - Turnover over 50 mil €
    - Deviance from edit rule (error) over 1 mil €
  - Influence of the observation
    - Turnover, Local Score = Weight * Turnover / sum(Turnover)
    - Personnel, Local Score = Weight * Personnel / sum(Personnel)
  - Influence of errors (failed edits)
    - Error in profit and loss account
      - Local Score = Weight * Error in euros / sum(Turnover)
    - Error in balance sheet
      - Local Score = Weight * Error in euros / sum(Balance total)

Tilastokeskus

Finnish case (3/3)

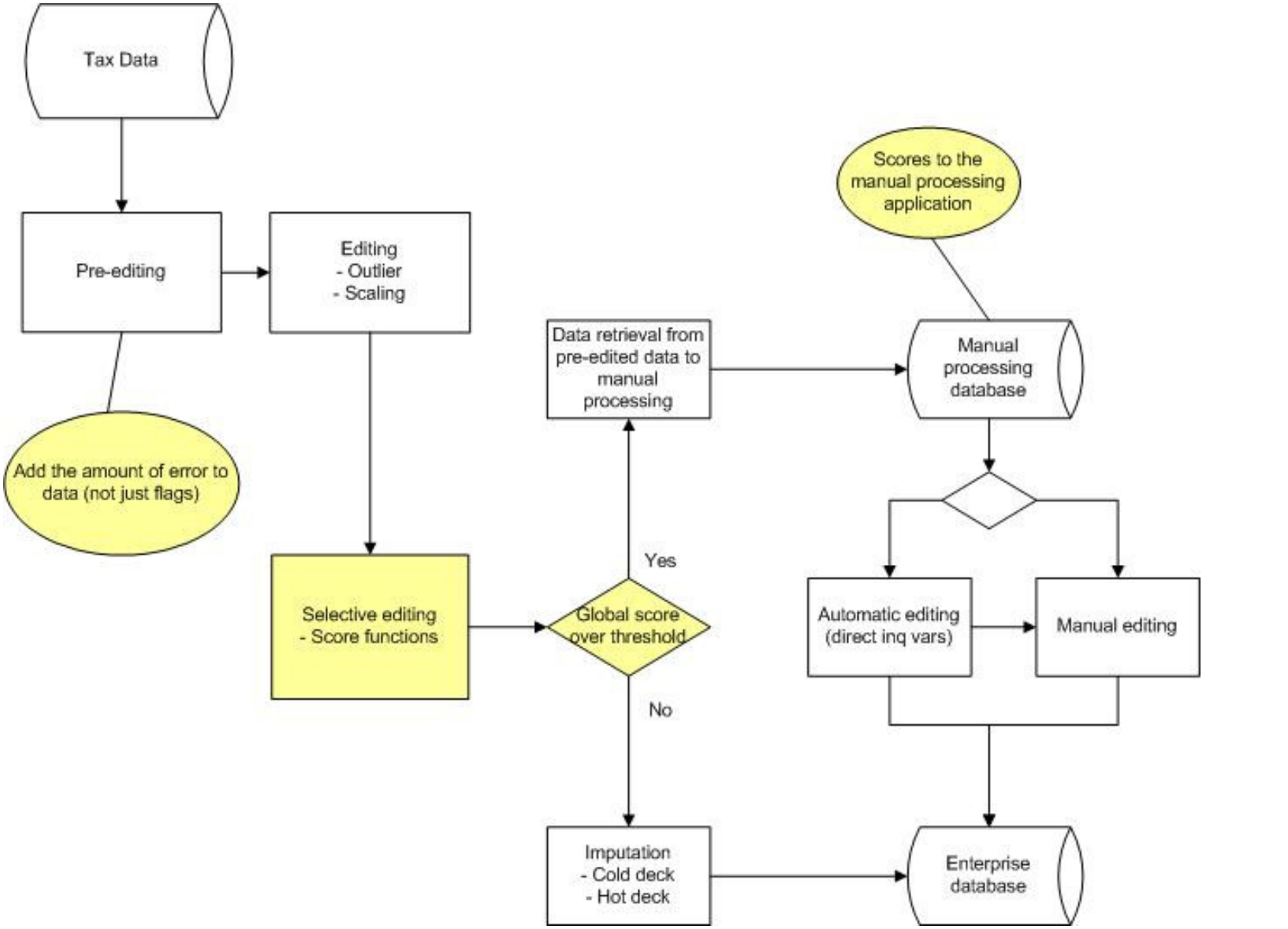
- Local Scores (continued)
  - Changes from previous year
    - Turnover, Weight * Change in euros / sum(Turnover)
    - Balance total, Weight * Change in euros / sum(Balance total)
- Global score
  - Sum of local scores
  - Max function is used to get a single score from a group of local scores
    - For example, from three different local scores on Error in profit and loss account, largest one is only used, not the sum of the three local scores

Tilastokeskus

Simplified example

- Company's turnover is 15 million euros, last year 10 million, in a NACE class, which has turnover of 325 million euros
- Local score 1 = 3000 * (15-10) / 325 = 46 points
- Same company had an error of 200 000 euros in profit and loss account
- Local score 2,4 = 150000 * 0,2 / 325 = 92 points
- Local score 2,5 = 150 000* 0,1/325 =46 points
- Company has 150 employee's. Local score 3 = 100
- Global score = 46+92+100 = 238 points

Selective editing in E&I process



Tilastokeskus

Making a simple selective editing procedure

- Define what are the classifications and their level of aggregation used in the publication/other important use
- Define what are the important variables or groups of variables in the observation
- Define the local score functions for these variables
- Define the calculation of global score function
- Define the cut-off threshold
- Define if there are some other practical considerations for giving scores, define what observations you would like to manually treat in any case
  - Give them local scores that exceeds the cut-off threshold (or more)
- Define the weights for individual local scores
- Test and balance!