The use of imputation in Danish RTI

By Søren Kühl Andersen

STATISTICS DENMARK

Content

- New publication schedule -> More imputation
- Extend
- Methods
- Simulation studies on revisions

New publication schedule

- First publication at t+20 - t+22
  - Provisional figures only at three main commodity groups
  - 4 NACE aggregates to Eurostat
- Provisional detailed industry figures one month later
  - Commodity groups revised
  - 13 NACE aggregates to Eurostat
- Final figures on all levels on month later

New publication schedule

What does that mean?

- We shortened the production time by approximately 10 days
  - Satisfies users (media, businesses, banks)
  - Lower response rates
  - New deadlines
  - More imputation
- Provisional figures are revised twice
  - Late data is incorporated to a larger extend
- Only one publishing day per month
  - Easier to organize
  - But now three months are handles simultaneously

The extend of imputation

- No mass-imputations
  - Only where necessary
- Just over 100 units are kept out of grossing up calculations
  - They contribute only with own TO
  - Imputed values needed in case of non-response!
- Why are those 100 units special?
  - We do not wish to or cannot include them in regular estimation methods
  - They are large
  - Some are artificial units
    - Collection of more enterprises
    - Split of one enterprise into more units
    - A few units with main activity outside RT

The extend of imputation

- Small industry groups without any responding units
  - Imputation of three units
  - Happens rarely
- Results from simulation studies:
  - 8-20 units needed imputation
  - In reality it will be fewer (4 for May 2012)

Methods – currently in use

1. Carry forward the growth rate of unit from same month last year

$$T_t = T_{t-1} \cdot {}^{T_{t-12}}/_{T_{t-13}}$$

Advantages:

- Suited for units with own seasonal pattern (often the case with large units)
- Unaffected by possible errors or outliers among other responses

STATISTICS DENMARK

Methods – currently in use

Disadvantages

- T_{t-1}, T_{t-12} and T_{t-13} all must be available
- Unsuitable for months including Easter
- Not useable in case of special occurrences around $t - 12$ (structural changes etc.)
- Does not take into account trend changes among similar enterprises

Methods – currently in use

2. Apply growth rate for actual month for units in same industry and size group

$$T_t = T_{t-1} \cdot \sum_{i \in M} T_t{}^{i} \,/\, \sum_{i \in M} T_{t-1}{}^{i},$$

Where M is the matched sample set of units from same industry and size group, responding in both t and $t-1$.

STATISTICS DENMARK

Methods – currently in use

Advantages:

- Can be used if there is no data on the unit from $t - 12$ or $t - 13$
- Can be used in “Easter months”
- Unaffected by structural changes in previous year

STATISTICS DENMARK

Methods – currently in use

Disadvantages:

- Does not take into account units' special seasonal pattern
- Affected by outliers/errors among similar units
- A sort of "mean imputation", which is bad!

Methods – currently in use

3. Carry forward the average of the three previous months

$$T_t = T_{t-1} + T_{t-2} + T_{t-3}/3,$$

adjusted for non-response.

Only to be used when method 1 and 2 cannot be used (no data from T_{t-1} or no data on similar units).

STATISTICS DENMARK

Methods – current practice

Now in production:

Methods prioritized depending on month:

In Mar, Apr and May:

2., 1., 3.

In other months:

1., 2., 3.

Output:

Imputed values for manual checking

List of units still needing manual imputation (rare)

Methods

Other methods to consider:

- A mix of 1. and 2. (when 1. is not possible) – use growth rate of similar units, but adjust for the units' historical derivation.
- A more general version of 1., using for example $T_{t-2} \cdot {}^{T_{t-12}}/_{T_{t-14}}$ or average of similar growth rates
- A more general version of 2., using for example T_{t-2}, T_{t-3} or T_{t-4} or average growth rates for similar units in more periods
- Additional "safety nets" to method 3., using for example older VAT data.

STATISTICS DENMARK

Methods

Implementing further methods would require:

- A database cleaned for outliers due to structural changes
- A lot of programming
- Real-time simulation-based choosing of best method in each individual case

Why not?

Not enough time – and current practice showed good results!

Simulation studies on revisions

- Simulation:
  - First estimates produced at t+17-t+19
  - Comparing to estimates produced on t+47-t+49 and t+77-t+79
- Imputation:
  - Current practice
  - Different priorities
  - More units for imputation (all of biggest size group, all)

Simulation studies on revisions

Results:

- Response rate compared to t+28 drops from around 85 to 70
- 8-20 of the special units needed imputation, regularly very big ones
- Revision from 1st to 3rd published figures between -0.2 and 0.3 – bigger on commodity groups
- Slight positive bias around 0.1, equal among commodity groups

Simulation studies on revisions

- First early publication: May figures on 22nd June
  - Response rate over 80!
  - Only 4 imputed values (hereof two large petrol station companies)
- New deadline (pushed from 11th to 9th)
- New reminding schedule:
  - 1st reminder on 10th
  - Deadline 16th
  - 2nd reminder by 17th (e-mails, letters and **phone to the special units**)

STATISTICS DENMARK