Models in official statistics

Kees Zeelenberg
with Bart Buelens and Peter Paul de Wolf

Statistics Netherlands

Contents

- What is model-based estimation?
- Examples of models in official statistics

- Guidelines
  - General principles
  - Criteria for the use of models

- Big data and models in official statistics

Estimation in official statistics

- Design-based estimation

➢ *Model-based estimation*
  - e.g. to estimate values for non-sampled persons

Examples of models in official statistics

- Seasonal adjustment
- Non-response correction
- Small-area estimation (SAE)
- Mixed-mode effects: correction
- Breaks in time-series due to redesigns: correction
- Time-series estimation of monthly unemployment
- Big data: Non-probability samples

So we are already using models in official statistics.

➢ *So do not be too principled*!

Models in official statistics

Generally:

1. *Statistical topic*, e.g. unemployment, GNP-growth
2. *Model* for estimating or improving statistical figure
3. *Data* for estimating model

Guidelines: general principles

- *Objectivity*
  - "6.1: Statistics are compiled on an objective basis determined by statistical considerations."

- *Reliability* and *accuracy*:
  - "12: European statistics accurately and reliably portray reality"

- *Transparency*:
  - "6.4: Information on the methods and procedures used is publicly available."

(Code of Practice for European Statistics)

Objectivity

- Data for estimation should be related to the statistical topic

  - Entities (households, enterprises, regions, …) should reflect those of the statistical topic
  - Nowcasting, but no forecasting

Reliability

- Failure of the model should not lead to failure in the statistical estimates or conclusion based on those
  - i.e. estimation procedure must be *robust* against model failure
  - Example: turning-points in time-series

- *No behavior* in model specification
  - Researcher may "rediscover" that behavior
  - Example:
    - Statistician: estimates consumption in budget survey as function of income
    - Researcher: "discovers" that consumption is function of income

Transparency

- Document model and analysis

- Publish model and analysis

- Explain role of model in table explanations

Guidelines for model selection

- Model selection: consider alternatives
- Model fit: test for fit and specification of model
- Robustness: do sensitivity analysis
- Stability: test
- Mean square error (MSE): report
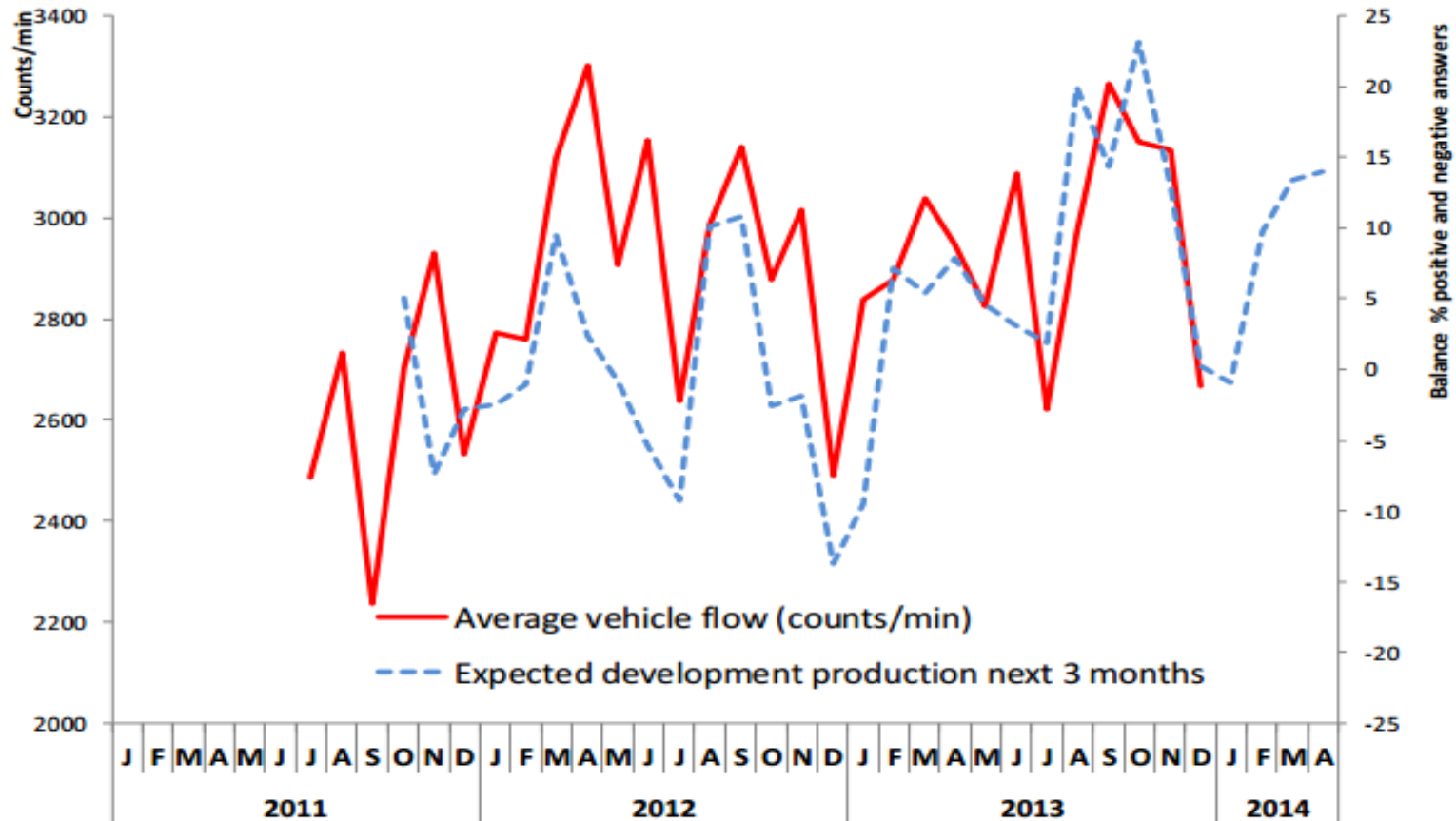- Assumptions: discuss plausibility

Future: Big Data

- *Big data come with big problems*

  - Very selective, not representative, linking difficult
  - Steps forward in official statistics need model-based estimation
  - So the use of models will expand

Example: traffic and output (Van Ruth, 2015)

Summary

- Examples show: we are already using models in official statistics. *So do not be too principled*!
- But :
  - We must remain *objective, reliable* and *transparent*

- *Future: Big data*
  - Steps forward in official statistics need model-based estimation

Models in official statistics

Kees Zeelenberg
with Bart Buelens and Peter Paul de Wolf