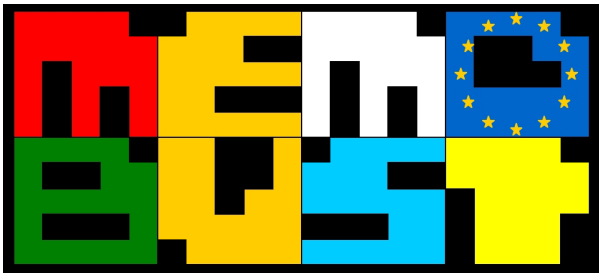

This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: Preliminary Estimates with Model-Based Methods

Contents

General section

1. Summary

For each survey, the standard process from collection to elaboration of survey data needs to be accomplished within a fixed period of time, i.e., the final estimates must be disseminated at the prefixed time *t*. In this context, direct estimators of the target parameters – based on the sampling units included in the Theoretical Sample (TS), selected by a probabilistic sampling design – are design unbiased and consistent; the sampling error depends on the variability of the phenomenon under study, on the planned sample size and on the effectiveness of the selection procedure. Direct estimators based on the Observed Sample (OS) – that is a subset of TS whose size depends on the total nonresponse rate – can be biased in function of the response process generating the OS.

We assign the term "preliminary" at the estimates computed using the statistical information available at time preceding the time *t*, on the basis of the OS denoted as Preliminary Sample (PS). The most straightforward practice in this situation is to apply the same estimation techniques utilised to produce the final estimates. Alternative estimation techniques should take under control the bias and the revision error, given by the difference between final and preliminary estimates. In order to test the quality of the preliminary estimator, the revision error should be evaluated for different survey occasions.

The main theoretical problem to be faced in a short-term preliminary estimation context concerns the possible self-selection of quick respondents, that can lead to biased estimates of the unknown population mean and variances. In the context of short-term business surveys – usually planned for estimating parameters such as indexes and their changes over time – one common method is based on the evaluation, for each design stratum, of the direct estimator of the index imputing the missing responses for the sampling units belonging to TS. Another type of procedure utilises the direct estimates of the design stratum indexes without imputation of the missing responses both in OS and in PS. These approaches can be based on imputation methods supposing no systematic differences between early and late respondents.

Preliminary estimation methods may be classified in function of the stage on which specific preliminary methods are applied. In fact, it is possible to identify methods that are acting:

- at the sampling design stage, by selecting a preliminary subsample of TS;
- at the estimation stage, in the following ways:
    1. by means of imputation techniques of missing data, that are applied to the non-respondent units in TS but not in PS;
    2. by means of weighting adjustment, i.e., modifying the sampling weights assigned to the units in PS in order to take into account non respondents in TS;
    3. by applying direct and indirect estimators, using known population totals of auxiliary variables and/or time series of preliminary and final estimates of the variable of interest.

The techniques based on the selection of a preliminary sample and the methods requiring imputation and weighting adjustment are generally based on unit level models. These models use disaggregated auxiliary information coming from survey data at previous times and/or administrative register data. For the methods in the last class the relation between the variable of interest and the auxiliary variables is usually formalised through domain level models in which the auxiliary information is expressed in terms of domain known totals or estimates. In the last class fall an estimation technique developed by Rao et al. (1989) in which preliminary estimates are computed assuming AR(1) models for final estimates and the revision error. This is the main specific model-based procedure used for the computation of preliminary estimation and it is described in this module.

2. General description of the method

In the context of a given sampling survey we mean as *preliminary estimate* the estimation of a parameter of interest obtained on the basis of a sub-sample of *quick respondent units* that is available within a time lag after the reference time point *t* (or end of the reference period) of the survey, while the correspondent *final estimate* is based on a final sample, including both *quick* and *late respondents*, observed within a time lag . The indicators measuring the statistical quality of a *preliminary estimation method* are based on the differences, evaluated at the different times (identifying the correspondent survey occasions) between preliminary estimates obtained by means of the method under study, and the corresponding final estimates. These differences are known as *revision errors*.

In this context, Rao et al. (1989) adopt a time series approach: let Y_t^P, Y_t and $Y_t^* = Y_t - Y_t^P$ be respectively the preliminary estimate at time *t*, the final estimates and the measurement errors in preliminary estimates at time *t*, $t = 1,\ldots,T$. Furthermore, Y_t and Y_t^* are supposed to follow an AR(1) process:

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0,\sigma^2\right) \tag{1}$$

$$Y_t^* = \psi Y_{t-1}^* + \varsigma_t, \qquad \varsigma_t \sim N\left(0,\sigma_0^2\right) \tag{2}$$

ε_t and ς_t assumed to be independent, while ϕ and ψ are autocorrelation coefficients ranging between -1 and 1.

Rewriting models (1) and (2) in state space form and ignoring sampling errors, they obtain the following final preliminary estimate for the period t+1 by means of Kalman filter (see for instance Harvey, 1984)

$$\hat{Y}_{t+1} = \alpha\left(\phi Y_t\right) + \left(1-\alpha\right)\left(Y_{t+1}^P - \psi Y_t^*\right), \tag{3}$$

where $\alpha = \sigma^2/\sigma_0^2$.

The preliminary estimate (3) can be viewed as a weighted average of the final estimate of the previous period *t* and the preliminary estimate for time $t+1$ adjusted for the previous measurement error.

Two alternative ways to obtain starting preliminary estimates Y_t^P can be used and both of them will be described in the next section.

2.1 Particular cases and extensions

Whenever any auxiliary information is available at current time, an extension of the basic previous method is possible by introducing in the model the auxiliary information correlated with the target variable of interest. In that case, the AR(1) model assumed for the final estimates (1) can be generalised in the following way:

$$Y_t = \phi Y_{t-1} + \sum_{k=1}^{P} \beta_k X_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0,\ \sigma^2) \tag{4}$$

Another extension can be obtained by introducing other previous estimates that can be considered highly correlated with the estimates at current time. For example, in the case of monthly estimates it is reasonable to assume that the final estimates at time *t* is both correlated to the final estimate at time *t*-1 and with the final estimate at time *t*-12. In this case, the following model can be assumed:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-12} + \varepsilon_t, \qquad \varepsilon_t \sim N(0,\ \sigma^2) \tag{5}$$

This model is known in the literature as a seasonal autoregressive model (see Choi and Varian, 2009).

A mixed extension of model (4) and (5) can be further considered, assuming the following model:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-12} + \sum_{k=1}^{P} \beta_k X_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0,\ \sigma^2) \tag{6}$$

Moreover dummy variables can be introduced into the previous model whenever specific domains estimations are required.

3. Preparatory phase

4. Examples – not tool specific

5. Examples – tool specific

6. Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

7. References

Bolfarine, H. and Zacks, S. (1992), *Prediction Theory for Finite Populations*. Springer-Verlag, New York.

D'Alò, M., Gismondi, R., Solari, F., and Naccarato, A. (2006), Estimation in Repeated Business Surveys using Preliminary Sample Data. *Atti della XLIII Riunione Scientifica SIS, 14-16 giugno, Torino*.

Choi, H. and Varian, H. (2009), Predicting the Present with Google Trends. Draft report Google Inc.

De Sandro, L. and Gismondi, R. (2004), Provisional Estimation of the Italian Monthly Retail Trade Index. *Contributi-Istat*, 24/2004.

Harvey, A. C. (1984), Dynamic Models, the Prediction Error Decomposition and State-space. In: D. F. Hendry and K. F. Wallis (eds.), *Econometrics and Quantitative Economics*, Blackwell, Oxford, 37–59.

Lamberti, A., Naccarato, A., and Pallara, A. (2004), Improving Timeliness of Short-term Business Statistics through State-space Modelling of Preliminary Survey Data. *Proceedings of the European Conference on Quality and Methodology in Official Statistics*. http://q2004.destatis.de/download/Table-of-contents_Programme.pdf.

Matei, A. and Ranalli, M. G. (2011), Adjusting for nonignorable nonresponse using a latent variable modeling approach. *Proceedings of the LVIII Meeting of the ISI, Dublin, 21-26 August 2011*. http://www.stat.unipg.it/~giovanna/papers/AMatei_GRanalli_isi11.pdf

OECD, Short-Term Economic Statistics (STES) Timeliness Framework. http://www.oecd.org/std/short-termeconomicstatisticsstestimelinessframework.htm

Papageorgiou, H. and Vardaki, M. (2002), Trade-off between Timeliness and Accuracy. Doc.Eurostat/A4/Quality/02/Timeliness/Timeliness and Accuracy.

Rao, J. N. K., Srinath, K. P., and Quenneville, B. (1989), Estimation of Level and Change using Current Preliminary Data. In: Kasprzyk, Duncan, Kalton, and Singh (eds.), *Panel Surveys*, John Wiley & Sons, New York, 457–485.

Royall, R. M. (1992), Robustness and Optimal Design Under Prediction Models for Finite Populations. *Survey Methodology* **18**, 179–185.

Tam, S. M. (1987), Analysis of Repeated Surveys Using a Dynamic Linear Model. *International Statistical Review* **55**, 63–73.

Valliant, R. (1999), Uses of Models in the Estimation of Price Indexes: a Review. *Proceedings of the Survey Research Methods Section*, American Statistical Association.

Specific section

8. Purpose of the method

The method is used for the preliminary estimation of the target variable, with the aim to obtain the estimates relying on statistical information available at time preceding the time *t*, i.e., on the basis of only a set of quick respondents which define the so-called preliminary sample.

9. Recommended use of the method

1. The time series of final and preliminary estimates should be long enough.

10. Possible disadvantages of the method

1. When the time series of preliminary and final estimates is short the estimation of model parameters can be very unstable.

11. Variants of the method

1. As an alternative to Rao et al. (1989) a time series methods based on the treatment of unit non-response may be applied. In this case, the late response is treated as nonresponse but in order to avoid biased estimates, the self-selection of quick respondents mechanism should not be considered as totally random. Difference between early and late respondents must be considered. In this framework, it could be interesting describing the process of self-selection of quick respondent by means of a latent variables, which can be interpreted as the ability to respond quickly. This predisposition can be used in order to estimates the quick response probabilities using a logistic model, allowing at the same time to deal with the non-ignorable auto-selection of preliminary respondent can be treated allowing . For more detail on the method see Matei and Ranalli (2010).
2. The preliminary estimation may be treated also using the dynamic linear model. For detailed information on these type of models see Harvey (1984) and Tam (1987). An application of this type of modelling in the context of preliminary estimation can be found in Lamberti et al. (2004).
3. Finally, some small area methods, for instance synthetic type estimator and modified GREG can be easily adapted also for the preliminary estimation, especially when the preliminary sample size is too small for computing reliable estimates.

12. Input data

1. Time series of final estimates.
2. Time series of final estimates of the revision errors.

13. Logical preconditions

1. Missing values
    1. Not applicable.
2. Erroneous values

1.

3. Other quality related preconditions

    1.

4. Other types of preconditions

    1.

14. Tuning parameters

1. Not applicable.

15. Recommended use of the individual variants of the method

1. The variants of the method can be used when additional auxiliary information or correlated estimates are available.

16. Output data

1. Ds-output1 = preliminary estimates of the target parameter.

17. Properties of the output data

1. The model-based preliminary estimates should guarantee a lower revision error than the direct estimates.

18. Unit of input data suitable for the method

Time series of Preliminary and Final Estimates at previous times.

19. User interaction - not tool specific

1. Choice of auxiliary covariates and/or estimates.

20. Logging indicators

1. Not applicable.

21. Quality indicators of the output data

1. Time series of revision errors.
2. Quality assessment of the result.
3. Model diagnostics to evaluate the model fitting when model-based estimators are applied.

22. Actual use of the method

1.

Interconnections with other modules

23. Themes that refer explicitly to this module

1.

24. Related methods described in other modules

1. Weighting and Estimation – Preliminary Estimates with Design-Based Methods

25. Mathematical techniques used by the method described in this module

1. Matrix algebra
2. Kalman filter

26. GSBPM phases where the method described in this module is used

1. 5.6 Calculate aggregates

27. Tools that implement the method described in this module

1. No software tools are still available.

28. Process step performed by the method

Estimation of target parameters on the basis of information collected on quick respondents.

Administrative section

29. Module code

Weighting and Estimation-M-Preliminary Estimates Model-Based

30. Version history

Version	Date	Description of changes	Author	Institute
0.1	01-03-2012	first version	Michele D'Alò, Claudia De Vitiis	ISTAT
0.2	26-06-2012	second version	Michele D'Alò, Claudia De Vitiis	ISTAT
0.3	30-09-2013	final version	Michele D'Alò, Claudia De Vitiis	ISTAT
0.3.1	13-11-2013	revisions based on review by Editorial Board	Michele D'Alò, Claudia De Vitiis	ISTAT
0.3.2	13-11-2013	preliminary release		
1.0	26-03-2014	final version within the Memobust project		

31. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	26-3-2014 13:33