

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Paris, France, 28-30 April 2014)

Topic (v): International collaboration and processing tools

**Editing and Imputation in the Memobust Handbook
on Methodology of Modern Business Statistics**

Prepared by Sander Scholtus and Leon Willenborg (Statistics Netherlands)

I. Introduction

1. The ESSnet project Memobust (Methodology of Modern Business Statistics) ran from January 2011 to March 2014. According to the project proposal, its main objectives were “the identification of best practices and the development of common methodology and ESS [European Statistical System] guidelines supporting the production of business statistics aiming at reducing respondent burden and fostering efficiency and integration of processes.” These objectives were made operational by the project goal of writing an updated handbook on methodology for business statistics. The project was supported by Eurostat; the actual work was done at the national statistical institutes of eight (later: seven) European countries.

2. The rest of this paper is organised as follows. In Section II, the Memobust project is briefly described. Section III describes the Memobust handbook. We discuss the topics that are covered by the handbook, its target group, its appearance, how it was written and reviewed, etc. We also discuss the role that the Memobust handbook can have in other activities. Section IV focuses on the description of the editing and imputation topics in the handbook. Some concluding remarks follow in Section V.

II. The Memobust project

3. The Memobust project was an ESSnet project that ran from January 2011 to March 2014. The national statistical institutes (NSIs) of seven countries participated throughout the project: Norway, Sweden, Poland, Hungary, Italy, Switzerland, and the Netherlands, which coordinated the project. In addition, the NSI of Greece participated only in the first half of the project.

4. The main operational objective of the Memobust project was to update the existing *Handbook on the Design and Implementation of Business Surveys* (Willeboordse, 1998). After about fifteen years, this handbook had become out of sync with statistical practice due to many developments that have taken place since its completion. Eurostat identified in particular the following areas where new methodological developments had to be taken into account: data collection, estimation, design of statistical processes, and data integration.

5. Willeboordse (1998) is a handbook in the traditional sense of the word, which means that it has a monolithic structure. From the outset of the project, it was decided that the new handbook should have a different, modular structure, to facilitate easy maintenance in the future. A consequence of this decision was that the new handbook had to be written mostly from scratch.

III. The Memobust handbook

A. Aim, scope, and intended readership

6. The purpose of the Memobust handbook is to aid those working in the area of business statistics. As such, the intended readership of the handbook is rather diverse. The handbook is primarily aimed at professionals who are active in the area of business statistics at (national or international) statistical institutes, including business survey managers, statisticians, and methodologists. It may also appeal to academic researchers who want to learn more about the techniques that are currently being applied to produce business survey data in practice. In particular, the handbook should be helpful to those who are new to (a particular area of) business statistics.

7. The prerequisites are modest. The technical level of many of the contributions has been deliberately kept low, with the aim of getting across the basic ideas behind a technique or methodology. For those who want to delve deeper into a particular topic, references are provided to more advanced, more detailed or more technical material. In principle, these references should be publicly accessible and written in English.

8. In principle, the scope of the Memobust handbook is restricted to describing those methods that are currently in use in the production of business statistics within the ESS, or that could potentially be used as such. In the former case it concerns methods that have been around for some time. In the latter case it concerns promising methods from recent research. Inevitably, the handbook also discusses some aspects that are not strictly methodological (e.g., related to process design and quality) and/or that are not restricted to the area of business statistics (e.g., some methods could be used also for social statistics). For a discussion of the differences between business statistics and other types of official statistics, see Kloeck (2011).

9. The Memobust handbook is *not* devoted to laws, regulations, official definitions of concepts, etc., although these are referred to occasionally.¹ Nor is the handbook intended to be prescriptive concerning the use of methods for the production of business statistics. Rather, the merits and demerits of different methods are described and compared.

10. The title of the handbook includes the word ‘modern’. It should be stressed that this is not so much a statement of fact as an appeal to keep the handbook up to date. This can only be achieved if the handbook lives up to its expectations and is valued by its users (see also Section V below).

B. Form and appearance

11. The name ‘Memobust handbook’ is somewhat misleading. In fact, this handbook is not a traditional book like its precursor Willeboordse (1998) and other existing handbooks on business statistics, such as Cox et al. (1995). Rather, it consists of an interconnected set of separate electronic documents, called modules.

12. There are currently 105 modules in the handbook. Within the handbook, modules are clustered into topics. Each topic covers a specific part of the methodology of business statistics. Examples of topics include “Sample Selection”, “Data Collection”, and “Weighting and Estimation”. A full list of topics in the handbook is given in Table 1 below, along with the number of modules within each topic.

¹ Concurrently to the Memobust project, another Eurostat project called the Framework Regulation Integrating Business Statistics (FRIBS) is producing a *Manual on European Business Statistics* that focuses on the legislative and conceptual side of business statistics in the ESS.

Table 1: List of topics in the Memobust handbook.

Topic name	Number of modules
Introduction	1
General Observations	6
User Needs	1
Overall Design	1
Repeated Surveys	1
Questionnaire Design	4
Statistical Registers and Frames	7
Derivation of Statistical Units	1
Dynamics of the Business Population	1
Sample Selection	7
Data Collection	7
Response	2
Micro-Fusion	13
Coding	8
Statistical Data Editing	8
Imputation	7
Weighting and Estimation	14
Quality Aspects	2
Macro-Integration	7
Seasonal Adjustment	3
Statistical Disclosure Control	2
Dissemination	1
Evaluation	1

13. At the time of writing, the modules of the Memobust handbook are available as PDF files on the project's website, which is part of a larger web portal concerning all ESSnet projects:

<http://www.cros-portal.eu/content/memobust>.

Apart from the handbook itself, other project information can be found there as well. It is likely that the contents of the handbook will be moved to a different location after the project has finished.

14. The main advantage of the modular form of the handbook is that it allows continuous updating of the handbook. These updates may include the modification of existing modules, the addition of new ones, and the deletion of obsolete ones. This updating can be done locally, affecting only a small part of the handbook, while leaving the bulk intact.

15. There are two types of modules in the handbook: themes and methods. Roughly speaking, theme modules are less specific and more verbal pieces that aim to discuss a common point in a general, non-technical way. They point out, for instance, what certain techniques have in common, why they are used, etc. In particular, nearly all topics contain a so-called 'main theme module' that introduces the subject treated in that topic. On the other hand, a method module has a more narrow focus. It provides a detailed description of a particular method.

16. The distinction between theme modules and method modules helps the handbook to serve the broad spectrum of readers mentioned above. Theme modules should appeal to the entire intended readership, and in particular to statisticians and managers. Most method modules should appeal mainly to methodologists, as the name suggests. For that reason they are allowed to be a bit more technical than theme modules. However, they still should provide a practically relevant overview. For further details they should guide the interested reader to the existing relevant literature where additional methodological or technical details of the methods described can be found.

17. Both themes and methods have a standardised format. They have been written using templates that were especially designed for the handbook. The use of templates gives a common structure to the modules in the handbook. It ensures that certain information is provided about all methods and themes,

and that this information can be found at fixed places within the modules. It does not mean that a single author has to provide all the information required. Information could be added or modified later by other experts.

18. There are several ways to access the information in the handbook. First of all, the electronic modules are stored on the Memobust website in a hierarchical structure (i.e., organised by topic). This structure should provide sufficient information for a reader to find a module on a particular subject. Another option is to use the glossary which provides access to relevant modules on the basis of key words. Finally, the modules in the handbook contain many cross-references to each other. This makes it possible to navigate within the handbook, without reverting to the glossary or the hierarchical structure.

C. The writing and reviewing procedure

19. Within the project, most modules were written by one country and reviewed by two other countries. This internal review procedure was very useful, both as a form of quality control and as a means to obtain descriptions in the handbook that extend beyond the experiences of a single NSI. In addition, a topic leader was appointed for each topic, to coordinate the structure of the topic (its modules) together with the other countries involved (authors and reviewers).

20. An editorial board was also set up, to oversee and assist the whole procedure. Towards the end of a successful writing and reviewing procedure, each module went to the editorial board for final approval. As a rule, the editorial board accepted the module after minor revisions. A particular task of the editorial board was to look at harmonisation between topics, i.e., to stimulate the use of the same terminology throughout the handbook and to avoid overlapping or even contradicting descriptions in modules written by different authors. Finally, towards the end of the project, a preliminary version of the handbook was validated by experts from Eurostat and NSIs not involved in the project.

21. In the project description, Eurostat requested that the terminology (terms and definitions) in the handbook should comply with the Generic Statistical Business Process Model (GSBPM; see Vale, 2009). Accordingly, we used the GSBPM as a basis for the division of the handbook into topics. Each module also contains a reference to the corresponding GSBPM phase (if relevant).² In addition, use was made of the SDMX Metadata Common Vocabulary (SDMX, 2009) where this was possible.

D. Possible uses of the handbook

22. The basic purpose of the handbook is to aid those working in the area of business statistics. This is where we position our main target group: employees working at a (national or supra-national) statistical institute, such as methodologists, statisticians, and managers. This group includes professionals who already have some experience working in this area and want to increase their knowledge of certain topics. A special group within statistical offices consists of novices, who are new to official statistics – or at least to business statistics – and want to have a quick overview of the area, before delving into one or more topics. The handbook should provide them with an overview of all topics that are important at the moment, initial information about these topics, and pointers to the literature.

23. A second possible use of the handbook is as a source of information for users of business statistics. Some users of statistical data produced by an NSI might be interested to learn more about a specific method that has been applied to the data they analyse. The NSI could guide them to a relevant section of the Memobust handbook, either through personal contacts or through hyperlinks in instructions to data files.

24. The Memobust handbook could be used as a basis for training courses on business statistics, for various target groups, and for courses ranging from general (aimed at giving a good overview) to specialised (aiming to treat a topic in some depth). Within the project, a general course on methodology

² It should be noted that the Memobust handbook used version 4.0 of the GSBPM, which was the most recent version available at the start of the project. Version 5.0 of the GSBPM was released in December 2013, too late to be taken into account in the present handbook.

for business statistics was held in November 2013, based on a selection of topics from the handbook. This course was primarily aimed at potential users of the handbook from ESS member states not directly involved in the project. Overall, the course was well received by the participants.

25. As a fourth point, the Memobust handbook could be used as a starting point to list in a systematic way for which topics useful software tools are already available (and describe these tools) and for which topics such tools do not yet exist. Using the information in the handbook, it should be easy to identify methods that lack suitable software tools. It would then be most desirable if open source tools were developed for these methods.

26. Finally, the handbook could be used to organise the (European) discourse on statistical methods used in business statistics. As mentioned above, the handbook does not intend to promote certain methods or demote certain others. In fact, the suitability of a method can be judged only when it is applied in a particular context. In this sense, methods by themselves are ‘neutral’. The handbook could facilitate discussions within the ESS about the merits and demerits of certain methods, the conditions under which they can be applied, possible alternatives, etc. In this way, the handbook could act as a means for further harmonisation and standardisation of methodology for business statistics in the ESS.

IV. Editing and imputation in the Memobust handbook

27. As can be seen in Table 1, two topics in the Memobust handbook are “Statistical Data Editing” and “Imputation”. Tables 2 and 3 below list the modules of which these topics consist.

28. Within the topic “Statistical Data Editing”, the main theme module introduces the problem of improving the quality of observed data to the point that they can be used for statistical purposes, by checking for errors and amending them. The remaining modules discuss methods that can be used in this context; this includes both methods that actually perform amendments to the data (Deductive editing; Automatic editing; Manual editing) and methods that can be used to organise the data editing process in an efficient manner (Selective editing; Macro-editing).

Table 2: List of modules in the topic “Statistical Data Editing”.

Module name	Module type
1. Statistical data editing – Main module	Theme
2. Deductive editing	Method
3. Selective editing	Theme
4. Automatic editing	Method
5. Manual editing	Method
6. Macro-editing	Theme
7. Editing administrative data	Theme
8. Editing for longitudinal data	Theme

Table 3: List of modules in the topic “Imputation”.

Module name	Module type
1. Imputation – Main module	Theme
2. Deductive imputation	Method
3. Model-based imputation	Theme
4. Donor imputation	Theme
5. Imputation for longitudinal data	Theme
6. Little and Su method	Method
7. Imputation under edit constraints	Theme

29. The two final modules in this topic are devoted to special issues: editing administrative data and editing for longitudinal data. Administrative data are used more and more in the production of business statistics. These data require an editing process that differs in some respects from the typical editing process for traditional survey data. Finally, editing in the context of longitudinal data is discussed in a separate module because most business surveys have a longitudinal aspect, if only because the largest business units are usually observed in every survey round.

30. Imputation for item non-response (as opposed to correction of observed values) is discussed separately in the topic “Imputation”. Within this topic, the main theme module introduces the problem of missing data, outlines potential solutions, and discusses aspects of imputation that are not related to particular imputation methods (e.g., the selection of auxiliary variables, the inclusion of a disturbance term, the use of sampling weights, etc.) There are separate modules devoted to common imputation methods (Deductive imputation; Model-based imputation; Donor imputation). As in the editing topic, there is a separate theme module on the use of imputation for longitudinal data, as well as a module on the method of Little and Su that can be used in this context. Finally, there is a module on possible solutions for the problem of imputation in the presence of edit rules that should be satisfied by the data.

31. For both topics, Statistics Netherlands acted as topic leader and all modules were written by project members from Italy and The Netherlands. Sweden and Hungary acted as reviewers for the topic “Statistical Data Editing”; in addition, the main theme module was also reviewed by Italy. For the topic “Imputation”, Norway and Sweden acted as reviewers.

32. In addition to the fifteen modules listed in Tables 2 and 3, the handbook also contains a module on “Editing during data collection”, as part of the topic “Questionnaire Design”. This module discusses the incorporation of automated edit checks and warnings into electronic questionnaires. It was written by the Central Statistical Office of Poland (the topic leader for “Questionnaire Design”). The Netherlands and Sweden acted as reviewers for this module.

33. It should be noted that the aim of the Memobust project was *not* to make substantial new contributions to methodological research, but rather to take stock of the current knowledge on methodology for business statistics and to summarise this knowledge in one place and in an accessible format. In writing the Memobust handbook, it was therefore considered important to refer to existing work as much as possible, in the form of available specialised handbooks, manuals, etc., both within and outside the ESS. In particular, for the description of data editing and imputation, extensive use could be made of EDIMBUS (2007) and De Waal et al. (2011), and many references to these works were included in the handbook. In addition, UNECE (2000) was used as a source for some of the terms in the Memobust glossary. Finally, these topics contain references to 23 different working papers presented at the Work Sessions on Statistical Data Editing over the years, as well as the volumes published by the UNECE that came out of these work sessions.

V. Conclusion

34. The Memobust project formally ended in March 2014. The Memobust handbook, on the other hand, is supposed to live on. Of course, this will happen only if the handbook is considered valuable by its intended readership. During the project, various dissemination activities were undertaken to promote the handbook, including the above-mentioned training course, an invitation to various member states of the ESS not involved in the project to review a preliminary version of the handbook, and presentations about the handbook at several methodological conferences. In the near future, the member states of the ESS will be asked to formally endorse the handbook.

35. Assuming for now that the handbook will find a purpose within the European statistical community, the question remains how to maintain the handbook in the future. In fact, the handbook has to be maintained and updated to keep track of new developments in methodological research; otherwise, it may quickly become outdated. As mentioned above, the modular structure of the handbook was designed to facilitate easy maintenance of the handbook. It has also been suggested to change the appearance of the handbook, by transforming the current set of PDF files to a wiki environment. This

would add more flexibility for making future improvements to the contents of the handbook. However, none of this is going to happen by itself; some things have to be organised.

36. If the Memobust handbook is to be open to new contributions and updates, some form of supervision is needed to guard the quality of what is written. This could be done, for instance, by setting up a centre of competence, supported by Eurostat, and appointing a moderator for each topic. If this turns out to be infeasible, an alternative could be to perform periodic updates of (parts of) the handbook through dedicated projects.

References

- Cox, B.G., D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge, and Ph.S. Kott (eds.) (1995). *Business Survey Methods*. New York: John Wiley & Sons.
- De Waal, T., J. Pannekoek, and S. Scholtus (2011). *Handbook of Statistical Data Editing and Imputation*. John Wiley & Sons, Hoboken, New Jersey.
- EDIMBUS (2007). *Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys*. Available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPM_EDIMBUS.pdf.
- Kloek, W. (2011). What makes business statistics different? Paper presented at the European Establishment Statistics Workshop, Neuchâtel, 12–14 September 2011.
- SDMX (2009). *SDMX Content-Oriented Guidelines, Annex 4: Metadata Common Vocabulary*. Available at: <http://www.sdmx.org>.
- UNECE (2000). *Glossary of Terms on Statistical Data Editing*. United Nations, Geneva.
- Vale, S. (2009). *Generic Statistical Business Process Model; Version 4.0 – April 2009*. Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS).
- Willeboordse, A. (ed.) (1998). *Handbook on the Design and Implementation of Business Surveys*. Office for Official Publications of the European Communities, Luxembourg.