

CONFERENCE OF EUROPEAN STATISTICIANS

UNECE Work Session on Statistical Data Editing

(27 – 29 May 2002, Helsinki, Finland)

Topic (iv): Impact of new technologies on statistical data editing

**IMMEDIATE QUALITY CHECKS OF OUTPUT PRICES USING THE BLAISE-MODULE
FOR COMPUTER-ASSISTED TELEPHONE INTERVIEWING**

Contributed paper

Submitted by Statistics Belgium¹

Abstract: Statistics Belgium set up an outbound call centre for the collection of output prices by way of computer-assisted telephone interviewing; the software used is the CATI-module of Blaise. Telephone interviewing was chosen because it allows very rapid data collection, which is crucial for output prices information. Another limiting factor for data availability, however, is the time and effort spent in checking for errors. Blaise-CATI can solve this problem: newly introduced prices can immediately be evaluated against other information and, when falling outside preset boundaries, be verified with respondents still on line. A practically error-free input can be guaranteed this way. The present paper explains how automatic quality checking of input is implemented in the Statistics Belgium application of Blaise-CATI for collecting output prices and how these can be used to efficiently calculate output price indices, required by the Eurostat regulation for short-term statistics.

I. INTRODUCTION

1. The EU Regulation 1165/98 on *short-term statistics* prescribes the collection, amongst others, of output prices with a well-specified scope, form, level of detail, timeliness, etc. This regulation expresses the need of the European Central Bank to have access to rapid short-term statistics in order to assess economic development in the Member States in the context of a single European monetary policy (EMU and Euro).

2. Until the beginning of March this year, the official Belgian output prices system was not compliant with Regulation 1165/98. Moreover, the survey method was rather labour-intensive and costly. Because of limited resources, the new survey and calculation methodology needed had to combine the lowest possible cost with the highest possible quality. Having identified three possible ways of conducting a price survey, we opted for an outbound call centre rather than a classical paper form approach or an inbound call centre.

3. The advantages of an outbound call centre, its low response burden, high cost-efficiency, guaranteed data quality and very rapid results, more than offset its one main disadvantage, the complexity of the implementation process, which consists of developing methodologies and software, settling practical and organizational matters and training operators and IT support.

II. SURVEY FOR OUTPUT PRICES: REQUIREMENTS

4. The Output Price Index (OPI), also called Production Price Index (PPI) is one of the short-term indicators calculated for all industrial sectors and required by Eurostat, as prescribed by the EU Regulation 1165/98. The output price indices are referred to as variables 311 (domestic market), 312

¹ Prepared by Peter Slock (peter.slock@statbel.mineco.fgov.be) and Marc Debusschere (marc.debusschere@statbel.mineco.fgov.be).

(non-domestic market, EU and non-EU together) and 310 (total market). One of the main requirements for these short-term indicators is their rapid, timely availability: the output price indices must be available monthly within one month and 5 calendar days after the end of the reference month. Hence, a fast collection of definitive prices for all industry sectors is crucial for successful delivery of the output prices.

5. At the same time, the need to collect prices for calculating the *material costs indices (CCM indices, Costs of Construction Materials indices)* became apparent: EU Regulation 1165/98 also prescribes the delivery of these indices (variable 321) on a quarterly basis.

6. Hence, a new survey for collecting output prices and input prices of construction materials had to be started, meeting following requirements:

- The output prices must be collected monthly for the total industry sector, comprising nearly all NACE rev.1 codes in the range 10-41 (i.e. sections *C to E*); the corresponding indices must be calculated at the NACE 2-digit level of detail and for all Main Industrial Groups. The material cost prices must be collected quarterly for the NACE 45 sectors (construction sectors, section *F* of NACE rev.1) at the level of detail specified by C111 (CC nomenclature of types of construction: one-dwelling residential buildings) and C112 (two- and more dwelling residential buildings).
- The output price indices refer to a specific month and must be available within 1 month and 5 calendar days after the end of that reference month. The material costs indices refer to a specific quarter and must be available within 3 months after the end of that reference quarter.
- For the output price indices, prices have to be obtained both for the export (i.e. 'non-domestic') market and the inner (i.e. 'domestic') market, with as observation unit the *Local Kind-of-Activity Unit*. Hence, the survey consists of questioning a local production unit - called *respondent* hereafter - of some companies for one or several products produced by that *respondent* in that area of activity (see the NACE rev.1 activity nomenclature).
- As *indices* are required by the EU Regulation rather than prices, the interest of the survey does not lie in the collected prices themselves but in the price *evolution* which is observed from one month to another or with respect to a year in the past. Hence, the prices of certain products from certain producers (respondents) must be collected *monthly* and be checked as soon as possible on their continuity.
- No legal obligation exists in Belgium for the provision of output prices or material cost prices, as opposed to mandatory surveys such as Prodcum.

III. THREE POSSIBLE SOLUTIONS

7. The first possible solution is a classical survey conducted with paper forms to be filled in manually by the respondents *and* to be sent back by them. The only advantage of this approach is that most people are familiar with it, so the implementation of it is not really new and the way of working has already been time-tested considerably. Also, this is the way in which price information was collected in Belgium until now. Some people were in favour of this possible solution, because no complex routines had to be installed.

8. However, a lot of serious disadvantages are related to such a classical approach:

- Very costly: forms to be printed thousands of times and sent out by classical mail to all respondents; then the filled forms must be sent in (again postage costs and delays implied!) or faxed to a green number, some reminders need to be sent and some data even has to be collected by a visit to the respondent.

- Very slow: same reasons as for the high costs; so, it takes a long time until sufficient forms come in after being filled by the respondent.
- Burden on respondents: respondents already have plenty of forms to be filled in which are mandatory. As a result, they are pretty reluctant to fill in some additional forms on a voluntary basis, or they are fed up with it after a short time.
- Manual input into database: hence *slow* data input, which is error prone because of several manual steps involved; can be accelerated by scanner support.
- Quality checks necessary *after* input: again time-consuming.
- Problem of *timeliness* because of all time-consuming, consecutive steps as described above.

9. The second possible solution is a survey via an *inbound call centre*: the respondents are asked to phone at regular time intervals to a telephone menu system and to key in the required numbers. The advantages of this system are:

- Low burden on the respondents (but discipline required);
- Highly cost-efficient: no operators needed;
- Immediate checks of the input are possible;
- Automatic and immediate input into database.

10. For a voluntary survey, as is the case for the output price and material costs indices, it may be too much to expect respondents themselves to take the initiative to phone monthly to a metal voice menu system in order to declare some of their prices. Another disadvantage of an inbound call centre survey is that the reliability of the collected data fully depends on the user-friendliness of the menu system through which the respondents are guided and on the goodwill, discipline and patience with which the respondents work themselves through that menu system. Furthermore, a lot of IT implementation and maintenance effort is required in order to arrive at a good working automated phone survey system. Even when the system works properly and when the respondents phone their prices in on a regular basis in a conscientious way, some human intervention might still be needed to ensure continuing collaboration of respondents and to check data quality or to assist respondents in case of doubt.

11. Therefore, a third possible solution was examined: a survey conducted by an *outbound call centre*, run by operators who conduct the interview with the respondents completely by phone.

IV. OUR SOLUTION: OUTBOUND CALL CENTRE

12. In this alternative, the respondents are contacted by telephone with very precise questions. All phone calls are done in a computerized way from a central outbound call centre at Statistics Belgium head office in Brussels by using the *CATI* approach: Computer-Assisted Telephone Interviewing. The main advantages of this approach are:

- *Low burden on respondents*: 1 short phone call received per month, no forms to fill in
- More control over the data collection, as the operators actively contact the respondents at the moment that the data are required
- *Personal guidance* of the respondent by the call centre interviewer: the interviewer can directly respond to questions from the respondents, direct feedback
- Computer-assisted: respondent-specific questions appear in the right order and in great detail on the screen, guiding the operators through the interview
- Immediate input of the obtained data into the database
- Immediate cross-checks with previous data while respondent is still on line
- Very cost-efficient: few operators do inputting as well as checking
- Data quality practically guaranteed: immediate checks of input
- Very rapid results (due to efficient data intake and guaranteed quality of input).

13. The only disadvantage of this approach is its organizational complexity. In order to implement an outbound call centre, a host of methodological, software, database, equipment and personnel problems

have to be solved and integrated into a smoothly co-operating whole. Accordingly, a professional outbound call centre is a big investment on the long term, which is not paid back by the returns of solely one survey. However, the many advantages cited above are not outweighed by this one disadvantage.

14. After careful consideration of the advantages and disadvantages of each approach, Statistics Belgium opted for an outbound call centre to collect output prices. Cost-efficiency and respondent-friendliness are significantly higher than in the classical method. Although the advantage over an inbound call centre is less obvious, we still considered the outbound call centre to offer a better trade-off between *data quality and timeliness* on the one hand and resources needed on the other hand. The most promising feature, in our opinion, is the possibility for *immediate verifications*, which guarantees the highest possible quality at the shortest possible delay. As the efficiency of the Call Centre is largely determined by the way in which it is computer-assisted, various existing software packages for computer-assisted telephone surveys were examined and compared.

15. This led Statistics Belgium to Blaise, a statistical data capture software package developed by Statistics Netherlands. In fact, Blaise-CATI integrates all aspects of telephone interviewing: management across and within surveys, appointments, interviews, data input, extensive quality checks, etc. Not only does Blaise-CATI support overall management of a survey (storing addresses and other relevant information on respondents, managing appointments and interview calendars, assigning respondents to operators, etc.), it also can guide an operator step-by-step through a specific interview and it allows immediate checks of the input against predetermined limit values or previous data. Responses are stored in a Blaise file, which can easily be converted to the DB2-database format we have been using for quite some time at Statistics Belgium.

16. Furthermore, the Blaise software was already used by Statistics Belgium in agricultural statistics for conducting the yearly agriculture census and other surveys using a sampling plan in the agriculture area. Blaise could easily deal with the nearly 63,000 respondents of the agriculture census, where every respondent filled in a form - electronically or on paper - containing some 400 variables, and allowed a fast publication of the census results, both final and preliminary ones. All those data were collected in a central database under Blaise. Reminders to respondents who did not answer yet or did not confirm their data, were done by telephone, where Blaise fully manages the list of phone calls to be made. Currently, a Blaise application is being developed for a *register of agricultural exploitations*, which must contain all administrative data, such as co-ordinates of the farm, contact people for surveys, all historic data with respect to the exploitations (since the start date of the exploitation until the end date of the activity), ... The register will serve as a basis for personalizing the forms by pre-printing all known information. There has been a quite positive experience from the application of the Blaise software in agricultural statistics for 3 years, therefore the step towards a wider use of Blaise at Statistics Belgium was not that big anymore.

17. So, a new unit 'Call centre' was founded at Statistics Belgium by Marc Debusschere, head of the Short-term Indicators unit, in September 1999. Initially, the new unit was simply attached to *Short-term Indicators*, as the call centre was set up in order to develop a survey of prices in services to be used for the construction material costs index. This survey was later extended to industry for the output prices. The set up of a call centre at Statistics Belgium had a double purpose: first of all to conduct a survey for collecting the prices of construction inputs and industrial output prices, next to serve as a piloting new survey approach for the whole of Statistics Belgium. Although the first applications of the call centre are in the area of new short-term indicators, at a later stage existing paper surveys of Statistics Belgium will be converted into phone surveys conducted by the Call Centre and new telephone surveys will be launched. In that way, the experience gained from our prices survey can be used for other applications. To do so, sufficient operators have to be available to expand the call centre.

V. THE IMPLEMENTATION PROCESS

18. Setting up an outbound call centre in Statistics Belgium for the collection of output prices using Blaise-CATI was quite a complex matter, mainly because so many different aspects, most of them new to us, were involved: software development, technical installations, organising a new unit, IT and

operational training, methodological development, etc. All of these had to be tackled more or less simultaneously and integrated into a smoothly running whole. The implementation of Blaise for the output price indices and material costs survey required quite some programming effort. In fact, the call centre application had to be developed in such a way that the operators, who are not IT specialists, can use the software without having to deal with any programming hitches or limitations. The whole implementation process consists of 3 big parts: preparation, installation and production.

A. Preparation

19. First of all, some people from the IT division visited the call centre at CBS-Heerlen of Statistics Netherlands, where quite some experience with Blaise-CATI was present already. So, Statistics Belgium obtained a good idea about how a computer-assisted call centre operates and it received a lot of very valuable practical information. As a result of that visit, Statistics Belgium decided to adopt Blaise-CATI as the software for our outbound call centre. Next, some IT people followed courses of Blaise in order to learn to program and to apply Blaise. Then they had to develop the initial application screens in Blaise to be used by the call centre operators.

20. In addition to the changes in survey procedures, the survey methodology and the database architecture were also updated for the output price indices, in order to obtain one global, homogeneous storage and publication system for all short-term indicators. Here, the experience from other short-term indicators calculated by Statistics Belgium, such as the Production Index and the New Orders Index, helped a lot. DB2 database tables were created to store both raw price data and output price indices.

21. Also, a representative sample had to be drawn from the total population of industrial products and producers. The products and then the respondents making those products (i.e. the producers) were selected by some random PPS procedure (Probability Proportional to Size) from the Prodcop population (for the output price indices). Here, 'Size' is based on Prodcop survey data of the year 1999, such as the values of supplied products per 8-digit Prodcop code and per respondent. The selection procedure led to a sample of 784 products covering all 223 4-digits NACE classes. It represents 26% of the products in the Prodcop-database but it ensures coverage of 75% of the total value of the production. For each product a certain number of transactions are selected. In the context of our survey, *transaction* is the key concept to deal with: it is the combination of one Prodcop product group (determined by its 8-digit Prodcop code) and one respondent producing in that product group. The central idea of the survey is to collect every month one domestic output price and one output price for the export market per transaction. So, a transaction is considered as the observation unit in the survey. In average, 54% of the transactions are selected, ranging 10% to 100%. In total, this corresponds to a coverage of 73% of the value of the transactions under study.

B. Installation

22. Also simultaneously, the Call Centre was equipped technically: telephone connections, PC's and other equipment were installed. The Blaise-CATI software was bought and installed. Call centre operators were selected or appointed, and trained in telephone interviewing techniques and in the use of Blaise-CATI.

23. Next, Blaise databases were created for respondents and price data. A separate database table was foreseen for every year of the survey, and the respondents were free, up till January the 1st of this year, to give their prices in Euro or Belgian francs.

24. Then, the IT team developed, in narrow collaboration with the operators, specific and detailed interview screens for the material costs and output prices survey at the same time, as user-friendly as possible. This also involved connections with the databases, both for data input and for extraction of previous data against which to check new data. For instance, a newly introduced price is compared automatically with the corresponding price noted for previous month, and if the new price differs more than 10% from the previous one (in whatever direction), a warning is displayed and the operator should check the new price with the respondent at line. Of course, operators had to familiarize themselves with

jointly using telephone and PC. In order to facilitate this new way of working, head sets with microphones were bought.

25. Later on in the project, IP phones were installed. In this way, the interviewers don't have to dial the phone numbers themselves anymore: Blaise automatically looks up the contact person at the respondent company and dials the corresponding phone number from its address book.

26. At the next stage, first contacts were established with respondents in the sample, one NACE 4-digit activity at a time, in order to ask them if they were willing to co-operate. Those first contacts were usually made via the business federation(s) for that activity type. As an incentive, the federations or respondent companies were promised monthly feedback on price evolutions in their own NACE 4-digit activity. Those agreeing to participate were then asked for help in identifying a very specific contact person and backup person for the survey and in further narrowing down product specifications (because in most cases even 8-digit Prodcom products are still categories composed of many different products and product types) to an absolutely individual product and type.

27. Finally the survey went into production for the first 4-digit NACE activities: each month, in principle during the third week, contact persons receive a telephone call asking for the price of one or more of their very specific products selected for the survey. Based upon this information, output price indices for those 4-digit NACE activities are calculated and sent back by fax or e-mail to the respondents having contributed.

C. Production

28. All these implementation steps caused the installation of the call centre to be a time-consuming, lengthy process. Finally, on the 26th of April 2000, the Call Centre of Statistics Belgium was officially inaugurated. From then on, it was possible to manage a telephone survey completely by computer, to call hands free and to introduce data directly and interactively on PC.

29. In the course of the year 2000, the survey for the prices of construction inputs and next for industrial output prices as well, was started. Although the participation to the survey is voluntary, some respondents could be convinced pretty easy to enter into the survey, because of the minimal effort required from them and the 'quid pro quo' principle: in return for their participation, the respondent companies were promised to gain access to monthly output price indices for all industrial sectors at the NACE-4 level of detail or even finer. Since March the 5th of this year, this detailed index information (NACE-4 level) is available to the respondents via a secret website of Statistics Belgium.

30. Of course, the call centre interviewers still had to get familiar with this new way of working: how to ask the questions, what to do in case of one or several refusals from the respondents, how to obtain *and* retain the *goodwill* of those respondents, how to perform immediate quality checks, how to correct data of current and previous months. Actually, all people involved in the project had to acquire experience with a wholly new unit, software, methodology and survey method. This means that databases, survey screens, ways of establishing first contacts with respondents and so on still had to be adapted fairly regularly. By the end of last year, most of the companies which had to be selected by the survey sample, had been contacted, and some 50% of them was prepared to participate to the monthly survey, which is quite a success for a voluntary, monthly survey.

31. After extensive analysis of the survey data by some statisticians and a methodologist of Statistics Belgium, it became clear that the initial survey results not always complied with our high demands of quality with respect to our output and material costs prices. Hence, some additional quality checks were elaborated gradually during the implementation of the survey by an IT support specialist under the guidance of our quality manager. Globally seen, there are 2 kinds of quality checks implemented with respect to the value filled in for a certain field:

- *Hard* checks: the interviewer can't proceed in the survey screen, because the current value entered for that field is invalid or because that - mandatory - field has not been filled in. This always concerns an *error*, and happens in case of a value which is inconsistent with respect to the value which

was previously filled in for that field or for a related parameter, or it happens in case of a code filled in for a key field of the database table but not having the required format or falling outside the range of allowed values. Also the presence and uniqueness of a key value can be assured at this point by such a 'hard' check.

- *Soft checks*: the interviewer gets a *warning* from the software, denoting a possible error for the filled value. It happens in case of a value which is highly unlikely with respect to previously entered values or related parameters, and which needs at least to be checked. Again, the interviewer can and must check immediately with the respondent at the line whether this value reflects the real situation at the enterprise. Here, the interviewer has the possibility to deny the warning and to proceed without modifying the suspect value supplied. This kind of checks can be applied for non-crucial fields, such as non-key fields of which the correctness does not have a major influence on the index calculations to be performed from the output prices collected.

VI. FUTURE DEVELOPMENTS

32. Now that the survey is fully operational, one of the major challenges left is shortening the delay between the reference month, for which prices have to be obtained, and the moment at which enough output prices are available for that month so that reliable output price indices can be calculated from them. Once those output prices are available for a certain month, the output price indices can be published within half a day later, thanks to our automated chain of calculations, controls and database transfers. Our goal now is to provide output price indices within one week after the end of the reference period, this is one month earlier than currently required by the Eurostat regulation. All normal dissemination products (press releases, graphs, tables, electronic files, web pages, etc.) will then also be available within the same delay.

33. Our call centre has already assisted in several other short surveys. Currently, it is examined what's the most opportune way of organizing the call centre within the whole of Statistics Belgium: or the current call centre has to be expanded to one, large call centre which performs surveys for all statistical divisions together; or every statistical division gets its own call centre, just performing its own surveys with interviewers specialized in the sectors concerned by that statistical division. Another question to resolve is whether to further foster this telephone way of doing surveys or to broaden the scope of data collection towards general *data collection centres*, which also deal with electronic forms on internet or CD-ROM.

VII. CONCLUSIONS

34. The 3 main lessons learnt from the output prices pilot survey are:

- Support from an experienced and dedicated IT team is crucial for a successful launch and follow-up of a CATI survey: the Blaise software needs some customisation so that it's tailored to the specific survey environment in which it has to operate.
- The presence of a CATI survey manager in the call centre during the whole duration of the survey is necessary not only to control the way of questioning by the call centre operators but also and mainly for guiding and coaching the operators when interviewing a respondent: the survey manager has to distribute the workload over the operators, co-ordinate those people in their daily work, assist them with background information on the purpose and contents of the survey, ... It has to be stressed that the call centre operators now have a complex task, which is much more than just taking routine interviews and keying in values. The interviews performed in the call centre now bring together in one unit all of the following activities, which were previously done by several units or at least by several people with quite different tasks: questioning the respondents, assisting the respondents, keying in the values directly into the Blaise database, directly asking the respondents about wrong or doubtful input while they're still at the line (when the software detected an anomaly in the data entered), correcting data, transferring data to the DB2 database. Hence, highly qualified and flexible people should be hired for interviewing in an outbound call centre.

- The initial efforts required for setting up a CATI survey are huge and should not be underestimated, but the rewards obtained from this approach are even larger and so these efforts are paid back, certainly on the long term. So, for a survey which has to be conducted once, it's not really worth doing all these efforts, whereas for a survey which is conducted repeatedly, on a regular basis (e.g., monthly for the output prices, quarterly for the material cost prices) during a longer period, it really pays off: although it takes quite some time (i.e. several months) in order to get the survey fully operational, the repeated follow-up questioning after that is a pretty easy and promising way of conducting surveys.

35. Contacts with other National Statistical Institutes, such as the ONS in the UK, Statistics Canada and Statistics Netherlands, showed the many assets of using Blaise for computer-assisted interviewing as well as some useful applications, and were a stimulus to Statistics Belgium to continue the development of the Blaise-CATI system for our survey, despite of the various practical difficulties to overcome in the beginning.

Acknowledgement

First of all, we want to thank our colleagues involved in the output price indices project for developing the Blaise system for the output and material cost prices and for dealing with the daily Blaise problems we faced during last 2 years: the 4 operators of the call centre, Koen Hooyberghs for launching the material costs and output prices survey in the call centre and doing the follow-up, Christophe Arickx for his enthusiastic database and other support and people from our IT division for providing various IT support regarding Blaise. Also, we'd like to thank very much our colleague Dominique Perreaux, who launched Blaise at Statistics Belgium by implementing it for quite some surveys in the domain of agriculture and who provided us with some valuable ideas.

References

Marc Debusschere, *Collection of prices by a Call Centre using Blaise-CATI*, Statistics Belgium, March 2001.

Dominique Perreaux, *Statistiques agricoles – Principales utilisations de Blaise* ('Agricultural statistics, main ways of utilisation of Blaise'), internal document of Statistics Belgium, February 2002.