The use of information from experts for agricultural official statistics

Marco Ballin, Riccardo Carbini, Maria Francesca Loporcaro, Massimo Lori, Roberto Moro, Valeria Olivieri, Mauro Scanu, ISTAT

1. Introduction

Expert information is a source of information widely used by the Italian National Statistical Institute (Istat) in order to satisfy the request of statistics on crops at NUTS3 level.

One of the characteristics of this source of information could be considered the opportunity to have information on phenomena that are difficult or prohibitively expensive to observe (e.g. estimate of very specific crops at a very disaggregated territorial level) as well as better timeliness with respect to traditional surveys. Furthermore, expert information can be easily integrated with traditional data sources.

In this framework, a prominent problem is the evaluation of the quality of statistics produced using expert information and to understand on what conditions this source can be fruitfully used in official statistics, in combination with traditional statistical sources (when necessary).

In fact, differently from current practices on “usual” survey strategies based on sample surveys or on administrative data, expert information is not complemented by a solid set of rules for the quality evaluation of the produced results.

Nowadays, the literature on elicitation (that is the process to gain information from experts) offers many alternatives (Garthwaite et al, 2005).

In this paper, after a review of the Italian experience in the use of this type of source, the role that a modern elicitation methods can play in the production of official statistics will be evaluated and a set of quality indicators that could be issued together with the produced statistics will be proposed.

2. Why using elicative methods?

Up to now, expert opinion has been widely used in Istat to produce short term statistics on crops. In short, data are provided by local authorities that collect experts evaluations on area and yield of different crops (January: estimates on areas for winter cereals, May: estimates on area for maize, June estimates on yields for winter crops ,…). Such data are used to accomplish for Eurostat regulations, gentlemen agreements, and are disseminated by the Istat web site.

\[\text{Address for correspondence: Marco Ballin, Istat, via Ravà 150 Rome, Italy, 00142 (e-mail ballin@istat.it)}\]
As it can be seen from the following graphs, the time series of these evaluations emulate “reasonably well” the estimates produced by traditional sample surveys every year (e.g. by the Farm Structure Survey, figure 1 and 2). Furthermore these statistics have some other good properties:

- First of all the timeliness (they are usually available before traditional statistics);
- Secondly, usually they are available with a great geographical detail because local authorities or experts supply evaluation at NUTS3 level;
- Thirdly, they are not expensive.

Figure 1. Common wheat area (national level) according to the Farm Structure Survey (sample survey) and aggregated evaluations from local authorities

![Common wheat area (ha)](common_wheat_area.png)

Figure 2. Maize area (national level) according to the Farm Structure Survey (sample survey) and aggregated evaluations from local authorities

![Maize area (ha)](maize_area.png)
Nevertheless, these evaluations cannot be considered true “statistical estimates” because they lack of quality indicators or a commonly accepted methodology. In other words we could have some difficulties in defining and filling up a quality report. Since they have some positive aspects, a useful tool could be added in the armoury of official statisticians at least to produce statistics on phenomena that are difficult or too expensive to observe with traditional tools. To this aim we can use the methods developed in the field of modern elicitation.

3. Organization of the elicitation process

The evaluation of expert’s opinion is a very delicate procedure. For this reason, great care should be given to the overall elicitation process, including the organisational aspects.

The importance of this activity is underlined in Garthwaite et al. (2005). The first step in an elicitation process is the “set up”, or in other words “preparing for the elicitation”. Garthwaite et al. (2005) write that this activity consists in “…selecting the expert(s), training the expert(s), identifying what aspects of the problem to elicit, and so on”. Apart from this, Garthwaite et al. (2005) focus mainly on the psychological side effects of the elicitation process (a fundamental aspect, that will be dealt with in Section 3.2), but do not specify explicitly any good practice for the set up stage. Protocols have been defined in this context. A very detailed protocol is in van der Sluijs et al. (2004) which is essentially based on the Stanford/SRI protocol (Spetzler et al., 1975) which is considered as the most used. In our context, we have applied a simplified protocol based on that in van der Sluijs et al. (2004), illustrated in Section 3.1. Section 3.2 shows the necessary euristics in order to overcome any bias during the elicitation process.

3.1 Set up stage

The set up stage consists of the following steps:

- Problem definition: the purpose is to define clearly and unambiguously the objective of the elicitation process. It is necessary to define a unit-scale which is familiar to the expert. Once defined the problem, the organizer of the elicitation process should be able to specify a set of questions related to the objective of interest.
- Selection of experts: it should be clearly described who are the experts on the objective of interest. This set of experts should be able to represent all the points of view. It should also be clear how to contact the experts.
- Selection of facilitators: preference should be given to facilitators with knowledge of the objective of interest. Additionally, a sound statistical background is advisable.

The previous steps have been considered in the application of the elicitation process, as in the following description.

**Problem definition** – The objective of interest was threefold

1. Quantity of saffron produced in Italy in 2007
2. Extension of surface used for the production of saffron in Italy in 2007
3. Forecasted quantity of saffron that will be produced in Italy in 2008.
The usual unit/scale for these questions (applied also in this elicitation process) are Kilograms for quantities and Hectares for surfaces. The questionnaire considers specific questions for the objectives of the elicitation process, and additionally other “control questions”.

**Selection of experts** – The elicitation process has been conducted on one expert. This is an expert of the consortium of saffron producers in one of the two most important areas for the production of saffron in Italy.

**Selection of facilitators** – One facilitator has been chosen in the Istat regional office of the chosen consortium. This facilitator has a wide experience in the agricultural sector. Furthermore he has basic knowledge of methodological statistics.

### 3.2 Euristics

**Definition** - Cognitive psychology relates to the question “how people make a decision?” or “how come to judgments and solve problems?”. Tversky and Kahneman (1971, 1983) examined the kinds of instinctive processes that people employ in this area of human behavior. They discovered mental strategies used under uncertainty. The researchers called these strategies with the term *heuristics*, and they defined them as intuitive judgements or simply common sense based on mental operations. This field of psychological research is very useful in an elicitation process, because of the involvement of decision making processes in experts’ opinion making. Heuristics are quick and easy to use informal methods to help solving a problem. In psychology, heuristics are simple, efficient, learned rules that work well under most circumstances. In certain cases they lead to severe errors and systematic cognitive biases.

In this paragraph we first review the main heuristics as well as their influence over the thought process. Secondly we describe a training program for experts and facilitators whose aim is to familiarize them with the strategies of the human thought process.

**Well know heuristics** (i.e. people mental strategies)
- Availability
- Representativeness
- Anchoring and adjustment

**Availability** – Availability is used by people to judge the probability of events or the frequency of class membership. This assessment is based on how easily an example can be brought to mind with the help of frequency and probability (examples from large classes are usually recalled better and faster than examples from less frequent classes, and likely occurrences are easier to imagine than unlikely ones). Other factors are ease of imagining or emotional impact (the ease of imagining or the vividness and emotional impact of that example becomes more credible than the actual statistical probability). The most common bias associated with these heuristics is illusory correlation. The phenomenon of seeing the relationship between two events when no such relationship exists.
Representativeness - Representativeness is used by people to assess the probability that A belongs to a class B. In this heuristic, objects of similar appearance are assumed to belong to the same class. This heuristic is applicable to the assessment of single event probabilities for non-repeatable events, that are typically the focus of an elicitation exercise. Several biases have been attributed to the representativeness heuristic:

- The conjunction fallacy
- Base rate neglect
- Insensitivity to sample size
- Confusion of the inverse
- Insufficiently regressive predictions: translating rather than predicting.

Tversky and Kahneman (1971, 1983) describe extensively these biases.

Anchoring and adjustment: People often estimate an unknown quantity by adjusting an initial value (anchor), until a final value is reached. People tend to remain too close to the initial value, although new data is given.

Research in Cognitive Psychology continues studying heuristics and thought processes.

In an elicitation process two training programmes are directed respectively to facilitators and experts in order to illustrate the heuristics and to make facilitators and experts aware of the many biases that can occur in judgment under uncertainty.

The training program for facilitators is conducted by specific trainers. It consists of three modules. The first module deals with heuristics and the main associated biases. Facilitators and trainers are involved in questions related to the elicitation of expert’s opinions. Trainers and facilitators will be also asked to search better case by case tools (e.g. interview, questionnaire).

The second module consists of the basics of social interview. It deals with elements of social interaction, communication strategies, and interviewer duty, to involve and motivate experts in their elicitation task.

The third module consists in role-playing, i.e. reproducing the interview of a facilitator with an expert in order to experience the elicitation process. Role-playing is useful to examine difficulties and problems that could happen.

The training program for experts is conducted by facilitators. The facilitators meet the experts before the elicitation process begins. This training program is organized in two modules. The first one shows heuristics and biases. In the second module a simulation of an elicitation process is done, to explore experts’ easy-to-use strategies in judgment. For instance, the expert can be asked about his/her opinion (estimate) on the temperature of the room, or something similar (like something about his/her hobbies).

In our elicitation process, we considered only a simplified facilitator training, consisting essentially in the first module. We submitted the questionnaire to the facilitator in order to get his overall impression as well as any ideas on how to improve it.

4. Elicitation

Elicitation is the most important phase of the elicitation process. In fact an obstacle to elicitation is that experts cannot easily describe their knowledge on the subject.
Specific summaries have to be elicited from the expert. In our application the agricultural expert may not have a mathematical or statistical background, so it is preferable to avoid statistical terminology such as mode, median, average and to translate these concepts into common language. It is easy for particular summaries of the distribution, such as the mode, the maximum and the minimum. The mode is the most probable value, and this definition requires only a common-sense knowledge of probability. Maximum and minimum are common terms.

Moreover to define the form of the distribution representing the expert uncertainty and to confirm his/her summaries, a second method has been used, the fixed interval method. In our opinion, combining the two methods strengthens the elicitation.

In general the elicitation phase is similar to a survey, but there are relevant differences. There is not a set of respondents but one expert (or a set of experts), who have to summarize their knowledge on some aspects of the topic. Experts are not only asked to give summaries but also an auto-evaluation of their confidence in their summaries. Moreover the role of feedback in the elicitation process is systematically adopted to improve the estimates.

The main operations carried out in the elicitation phase were:
1. Interview;
2. Transformation of expert opinions into probability distributions;
3. Feedback;
4. Production of final results.

The first experience of the use of the elicitation method was carried out in the first months of 2008.

The facilitator carried out the interview using a questionnaire to elicit the quantity of saffron produced in Italy in 2007, the extension of surface used for the production of saffron in Italy in 2007, and the forecasted quantity of saffron that will be produced in Italy in 2008. The questionnaire was designed in order to elicit three estimates for each phenomenon of interest: the minimum, the maximum, the modal value (as the most probable value) for each phenomenon of interest.

Moreover the elicitation technique called fixed interval method was used for each phenomenon in order to

1. obtain information about the distribution shape (right skewed, left skewed, symmetrical, rectangular, unimodal, bimodal...)
2. check the estimates in order to avoid any misunderstandings with the expert

The fixed interval technique requires the expert to distribute ten “X” into five intervals of equal length. The extremes of the five fixed intervals are given by the subdivision of the interval between the maximum value and the minimum value previously elicited (Figure 3).

Figure 3. Elicited quantities for Italian production of saffron in 2007.

<table>
<thead>
<tr>
<th>INTERVALI</th>
<th>I intervallo</th>
<th>II intervallo</th>
<th>III intervallo</th>
<th>IV intervallo</th>
<th>V intervallo</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>105</td>
<td>110</td>
<td>115</td>
<td>120</td>
<td>125</td>
</tr>
<tr>
<td>XX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>125</td>
</tr>
<tr>
<td>XX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>125</td>
</tr>
<tr>
<td>XXXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
After checking the compatibility of the elicited estimates, the next step is the transformation of expert opinions/elicited values into probability distributions. In our application the beta distribution has been chosen to represent expert opinion. The beta distribution is a very flexible distribution: it has a finite range, can be unimodal (and sometimes bimodal), uniform ($\alpha = \beta = 1$), and can be either symmetric or skewed.

The beta distribution parameters were determined in a simple way, using the formulas for the approximation of beta distribution to a triangular distribution defined by the three elicited quantities: maximum, mode, minimum. The fitted distribution was compatible with the method proposed by Gilless and Fried (2000), a method that use in addition the 90th percentile (obtained using the fixed interval method). Then, in the feedback phase the fitted distribution was sent to the expert to evaluate the distribution and the “fiducial interval” or credible interval. At the end of this process, we obtained the final results (Figure 4).

**Figure 2. Fitted distribution for Italian production of saffron in 2007**

![Fitted distribution](image)

The beta distribution is semi symmetric, the mean is 119.2, and the 95% narrowest fiducial interval is 111.1 – 126.8.

5. Evaluation of elicitation results

The production of a report concerning the quality of statistics should be mandatory for every statistical process and should be a must for official statisticians. To help to this aim Eurostat is working on guidelines for filling up a quality report containing description and indicators comparable across countries and among statistical processes.

It is useful to remind the terms indicated in Eurostat’s guidelines to assess the quality of a statistical output:

- **Relevance**: outputs meet current and potential users’ needs.
- **Accuracy and Reliability**: outputs accurately and reliably portray reality.
- **Timeliness and Punctuality**: outputs are disseminated in a timely and punctual manner.
- **Accessibility and Clarity**: outputs are presented in a clear and understandable form, disseminated in a suitable and convenient manner, made available and accessible on an impartial basis, and accompanied by supporting metadata and guidance.
- **Coherence and Comparability**: coherence means that outputs are mutually consistent and can be used in combination; comparability is an aspect of coherence that means that outputs referring to the same data items are mutually consistent and can be used for comparisons across time, region, or any other relevant domain;

as well as the terms proposed to describe the quality of a statistical process:
- **Sound methodology**: sound methodology, including adequate tools, procedures and expertise underpins quality statistics.
- **Appropriate statistical procedures**: appropriate statistical procedures are implemented from design through data collection to data validation and evaluation.
- **Non-excessive burden on respondents**: the reporting burden is in proportion to the needs of the users and should not be excessive. It is monitored over time and targets are set for its reduction.
- **Cost effectiveness**: resources are effectively used.

In the guidelines, besides showing that it is possible to define a wide range of quality reports according to their objectives, the level of detail, the producer or user orientation, and the perspective of process or output, it is shown how the different type of statistical processes (e.g. sample surveys, administrative sources, etc) influence some parts of the report.

In the following, it is proposed, whenever needed or possible, some guidelines that the report should contain in the case of statistics based on elicitation of expert information.

**Relevance** - As for the other statistical processes a content-oriented description of the whole statistical output should be given.

In the case of elicitation of expert information some analogies can be found with the case of administrative processes.

In fact, as in the case of administrative databases, the knowledge of experts in terms of population, concepts and definitions of variables is fixed in advance and could be different from the ideal ones requested for statistical purposes. Given that the users should be aware of definitions of the main concepts used by elicited experts, an evaluation about the difference between expert definitions and those requested by the statistical process should be supplied. This evaluation could be furnished by the experts too.

**Accuracy** - In the case of traditional processes the difference between the estimates produced by every statistical process and the true values of the target populations are due to
- variability;
- bias.
The first one is due to the random component of the statistical process (e.g. sample); while the second one is the effect of non-sampling errors of the statistical process in terms of: coverage; measurement; nonresponse; processing.

The random component can be evaluated by indicators as sampling variance; different approaches for its estimation are available in the literature and usually a numeric evaluation can be given using only survey data. On the opposite, numerical evaluation of bias is usually more difficult and most of the times an ad hoc survey is needed.

In the case of elicitation of expert opinion indicators of variability should regard:
- heterogeneity of the elicited distributions from more than one experts (this should emulate the idea of variance);
- fiducial interval (this should replace confidence intervals);

On the bias side the report should contain indicators on the following aspects
- effect of feedback (this should emulate the indicators on effect of re-interview)
- assessment of expert knowledge by means of control/seed variables (this should replace the indicators for bias)

Furthermore, as in the case of traditional processes, the report should contain some indicators that give some hints on the main source of random and non random errors. For this reason, besides a description of the elicitation process, some indicators as the following ones could make users aware about the measurement errors and consequently about the use of these statistics:
- self assessment of expert knowledge and of his/her sources of information (by expert),
- assessment of how the questions in the questionnaire are phrased (by expert),
- meta-information on the interview (by facilitator).

**Coherence and Comparability** - Coherence and comparability are properties of the statistical outputs that account for their joint use. They can be evaluated in terms of concepts, definitions, target populations, etc. used by the processes at hand. There are different types of coherence and comparability, such as: comparability over time, over region, coherence between sub-annual and annual statistics, with other statistics, and so on. This distinction can be saved for the case of elicitation of experts information too. Nevertheless it could be useful to point out two general cases of interest (for the sake of simplicity we consider the case of two processes):
- one of the two processes is a traditional one;
- the same elicitation processes is used for two (groups of) experts to produce statistics referred to different geographical domains.

In the first the problem concerns mainly coherence because users must be aware about the extent to which outputs from the two processes can be reliably used in combination. Hence users should be informed about the concepts, definitions, target populations, etc. used by both processes. The difference between them should be underlined. For example, as it will be exemplified later, coherence could refer to trade statistics estimated both using an administrative process (i.e. the current official statistics) and evaluated by the expert (Do the two methods refer the same period of time? Do they use the same classification? Do they use the same unit measure?).

In the second case, it seems better to refer to comparability. Assuming that experts and facilitators have the same elicitation tools (questionnaire, training, etc.), the problem is to make users aware about the resources used by each expert (e.g. availability of local administrative databases) as well as the experts confidence in such sources (e.g. degree of coverage). Furthermore, users should be aware about
the concept used by the experts (e.g. are they referring to the administrative geographical areas or are they referring to some other definitions? Are they sharing the same definitions? Etc.).

**Other quality components** - For Timeliness, Punctuality, Accessibility and Clarity the elicitation process does not need any specification. Finally it should be noted that Confidentiality is still an open problem and must be treated appropriately, because data providers are experts that can be easily identified

Referring to the saffron application, three indicators were calculated:
1. Effect of feedback;
2. Fiducial interval;

The effect of feedback was positive, as the expert confirmed the elicitation result after the production of the first report, containing the fitted distribution and the fiducial intervals.

The fiducial interval is an indicator of the confidence of the expert about the elicited value. In this case it shows that the narrowest 95% interval for Italian production of saffron was 111.1 – 126.8, and the mean value was 119.2.

The third indicator is a self assessment of expert knowledge and of his/her sources of information. The index (sum of points/max) is equal to 81,2% (figure 5).

**Figure 5. Self assessment**

![Self assessment diagram]

---

**6. Future developments**

Elicitation of experts opinions has been seldom applied in an official statistics context. We believe that this procedure can be an additional source of information useful to the official statistician as well as to the users of official statistics, which has the benefit to be defined in a well established scientific framework.

The example of elicitation process described in this paper can be improved. First of all, more than one expert can be contacted. In this case, there is the methodological problem to combine the elicitations of two or more experts opinions. Some procedures are already shown in Garthwaite *et al* (2005).
Secondly, once experts opinions have been elicited, it is possible to combine the elicited distributions on experts opinions with the sample data, whenever data are available. This corresponds to the use of appropriate Bayesian procedures for the estimation of the aggregates of interest. We believe that experts opinions can be a valuable source of information also in other contexts than agriculture. In general, we believe that the elicitation of experts opinions, as well as of the public opinions, is important for National Statistical Institutes. This knowledge allows to compute any discrepancies between data and the opinions (of experts or of the public). As a matter of fact, this is a measure of the “strength” of the message disseminated by statistical offices by data.

References


