

Sources of Agricultural Statistics - Part 2: Discussion

Ulf Jorner

Statistics Sweden

1. Some Introductory Remarks

The sources of agricultural statistics are extremely diverse. Other fields of statistics, like economic statistics or transportation statistics, may have very complicated data capture mechanisms. However, I think that agricultural statistics has the most varied data sources (with the possible exception of environmental statistics). As we have seen from the papers presented in this section, these sources range from personal and telephone interviews, postal surveys and bookkeeping, over various administrative records, eye estimates and field measurements to satellite measurements. And even this list is far from complete and very compressed. It would, for instance, be easy to distinguish a number of different kinds of administrative records: taxation, subsidies, sales and deliveries of goods, etc.

The opportunities and challenges to agricultural statistics with respect to this multiplicity of data sources are well illustrated by the papers in this session. I would like to congratulate all authors for very clear and thought-provoking papers. At the same time, this very variation in approaches poses some problems for the discussion. First, I must apologize to the authors of contributed papers that space and time limitations compel me to mention just a few of them. Further, I have tried to structure my discussion and concentrate on a few of the, in my opinion, most interesting aspects on data sources in agriculture:

- Users and uses,
- Quality and costs,
- Combination of sources, and
- Use of models.

Moreover, I have tried to avoid technical questions, however relevant they may be, to minimize the risk that the discussion bogs down in details.

2. Berkeley Hill: “Surveys for Farm Structure, Agricultural Accounts and Total Income of Agricultural Households”

The focus of this paper is very much on how the use of statistics affects the design of statistical surveys, in this case how the needs of the Common Agricultural Policy of the European Union demands economic statistics for agriculture. I think that the paper gives a very readable and well-structured overview of this complicated subject.

The need to harmonize the definitions, data collection, etc. for (at present) 15 countries is just one example of the level of complication. Another is the problem to define the variables under study, e.g. Farm Net Value Added, or the objects under study, e.g. agricultural household, or, indeed, the whole sector under study, e.g. the agricultural community. The fact that these definitions tend to change with shifts in emphasis on the political level — the recent shift in focus from agricultural community to rural community is one example — further adds to the complication.

The paper makes a very clear distinction between macro and micro approaches, with a corresponding difference in data sources. On the whole, I agree with the author that the macro level is rather well covered today. I also agree that much more remains to be done — and that a concentration of efforts should be made — on the micro level, although I am a little more pessimistic on the rate of progress being made. One reason for this may be that Sweden is mentioned as a country with good basic data!

There are two specific questions I would like to raise on this paper:

- ? Do you think that the data quality problems with the IAHS approach, not least the variation over time and between countries, are so large that the results may be questioned by the users?
- ? Apart from the “political” opposition to the extension of FADN to cover also non-farming incomes, there seems to be a certain duplication of efforts between IAHS and an extended FADN. Do you think both approaches should be pursued, or that there should be a concentration of efforts?

3. René Gomme: “Agrometeorological Crop Yield Forecasting Methods”

Again, we here have an excellent overview of different approaches, this time on crop forecasting. This field has always been of very great interest to decision-makers in agriculture. As is mentioned in another of the papers in this session, we have an official Swedish time series in this area that starts in the late 18th century.

Developments in this field seem to me to have created something of a lure, that we can do away with all the expensive and time-consuming field measurements and simply replace them with satellite and weather data. Consequently, I am pleased that this paper, already in the very first paragraph of the abstract, stresses the need to combine agrometeorological approaches with traditional yield statistics.

The author stresses robustness as an essential quality in choosing between different methods. I think that this is essential, and a consequence might be that we should not evaluate methods the way we have traditionally done, i.e. over a whole period of years. As the author points out, it is very easy to construct a method that works well for average conditions. Perhaps we should attach more weight to extreme years when we evaluate, especially as the working of the method is more crucial to the users for these years.

Another point that the author makes is that more data (longer time series, finer grid, more variables, etc.) does not guarantee better results. I quite agree with his conclusion that accurate selection of relevant data is more important.

A number of interesting methodological questions are also taken up in the paper. I will just note that in a number of cases, analogous questions exist in other statistical fields. Just two examples where a “cross-disciplinary” study may give new ideas: There are obvious similarities between the G→M and M→G pair of approaches and the micro and macro approaches (and the bridges between them) in economics. Also, the question of whether to use average or historic/simulated data for “future” data in crop simulation models is essentially the same as the question of how to impute missing observations in surveys.

Two final questions:

- ? While I agree with the concept to use “value-added” variables in the simulation models, the corresponding measurements are relatively complicated and expensive. Under a fixed budget, do you still think that fewer observations (in time and/or space) on more relevant variables are to be preferred? Is there any empirical evidence?
- ? Do you think that there is a tendency that agrometeorological methods supersede rather than complement traditional methods? And, what is the best traditional method?

4. Lars Hagblad: “Crop Cutting versus Farmer Reports - Review of Swedish Findings”

As the two earlier papers, this gives an interesting and well-structured comparison of methods. This paper, however, focuses on empirical comparisons. These have been carried out in Sweden during a rather long period, first to build up a crop-cutting system and then to find means to reduce costs with alternative methods. As a matter of fact, the Swedish crop statistics seem to have made more or less a full circle, beginning with eye estimates and ending up with farmers’ reports.

These empirical results again illustrate the point that most methods work under normal conditions, but not under adverse conditions, as in this case the 1992 drought. The evaluation seems to have taken due regard of this.

Another fact brought forward very well by this paper is that there are interesting differences between results for the whole country and smaller parts. Of course, we will normally expect larger random variations for smaller areas. However, the performance of a method on smaller areas may in some cases be an important consideration when choosing methods.

Another interesting, if not surprising, observation is that in some cases farmers may have an interest in over- or under-reporting. In the Swedish material, the potato producers seem to have an interest in keeping the harvest figures low. The differences in market mechanisms between crops thus seem to influence the choice of methods.

I would like to pose one question on the future of Swedish crop statistics and a more general question:

- ? After all this experimentation, what will be the future system in Sweden? Have you any plans to use, for example, crop simulation models?
- ? There has been some discussion on the performance of crop cutting versus farmers’ reports. Traditionally, crop cutting has been considered good but expensive. About ten years ago, there were some reports that crop cutting would be inferior on both accounts. On the basis of Swedish experience, what is your opinion?

5. Gordon Reichert, et al.: “Statistics Canada’s Near Real-Time Crop Condition Assessment Program (CCAP) Utilizing NOAA AVHRR Data - Remote Sensing, GIS and the Internet”

This paper presents, in a very interesting way, an exciting and very user-oriented program, in which 160 expert reporters have been replaced by the use of advanced technology. What I find especially interesting is the use of modern technology not only to collect the data, but also to make it quickly and readily available to users. Of course, this is nothing new as the concept of Geographic Information System (GIS) has been used in statistics for a relatively long time. However, I think that the Canadian program goes further than most other comparable programs in its user-orientation.

On the input side, I think that I as a Swede, with much more heterogeneous agriculture to describe, can only envy Canada the ability to use rather coarse satellite data, not even having to distinguish between different crops. When it comes to the cloud problem, we seem to have the same problems (and solutions). I think that this is a good example of a situation when more data, in this case measurements for several days, is indeed essential to good quality output.

As for quality, the short time series available makes it difficult to assess the method, especially as there seem to be few extreme years during the period. I was a little puzzled that the forecast does not seem to be calibrated to the level of definite figures, but on the other hand I am not sure what years are included in the calibration period. However, the method seems to score high on a quality component that has historically been somewhat overlooked by statisticians, i.e. timeliness.

Because of the user-orientation of this program, I think that the most interesting questions raised by this paper are connected with its use:

- ? How many users does the program have, who are they and what use do they make of the output?
- ? Does the program make money, i.e. does it cover its development costs and running costs?

6. Geoff Heffernan, et al.: “Data Collection for Environmental Monitoring of the Agricultural Industry”

This paper deals with another important aspect of agriculture, namely its impact on environment. This impact can be assessed in a number of ways, and this paper focuses on the collection of data on farm level and gives a very clear picture of the possibilities and problems of this approach.

It seems to me, interest in the connection between agriculture and environment has gone through a few phases since the issue was first brought to general attention about 30 years ago. First, agriculture was definitely a villain. Then, there was a shift in interest towards other industries. In the last few years, I think that there is a renewed interest in agriculture’s influence on environment, but a more balanced one. While the negative influence, of course, still is in focus, there is also an interest in the positive role agriculture plays, and an interest in questions like biodiversity.

The authors concentrate on data collection by interviews and mail questionnaires. It is obvious that many of the questions asked may be regarded as sensitive by respondents. As an example, I assume that some of the questions deal with practices that are covered by laws or regulations. It would be interesting to know if these are the types of questions that have been most problematic.

However, I have two other questions:

- ? What is the relationship between Agricultural Environmental Monitoring and Environmental Accounts? Do the environmental accounts use your data directly or indirectly?
- ? You mention the need to validate the data collected from farmers. What other sources of data do you use (e.g. administrative data, sales data, etc.)? Do you use such sources only to edit individual forms, or also to calibrate final results?

**7. Koji Kitamura and Fumio Nakayama: “New Estimating Method for Normal Yield of Rice”
Bal B.P.S. Goel: “Maize Yield Forecast Based on Cob Size”**

I have included these two contributed papers in my commentary because they provide an excellent contrast in modeling, even if both may still be well suited in their respective environment.

The first paper presents a method to describe the changes in normal yield that uses a very complicated model, incorporating principal component analysis, polynomial regression and spline functions. The second uses a very simple regression model with “natural” variables and relationships.

While I must admit that I personally prefer simple models as far as possible — one advantage is that in their very simplicity they show that they are approximations — I would like to pose a question from the users point of view:

- ? Is it important that the users understand our models and methods (and not only accept the results)?