ELLEN HANAK FREUD*

Making Better Sense of the Numbers on Developing Country Agriculture

INTRODUCTION

The dilemma facing analysts of agricultural sectors in developing countries is time-worn and seemingly intractable. To inform policy choices and orient agricultural research, one needs to be able to follow trends on strategic variables: output, land use, on-farm and post-harvest productivity, internal and external trade, consumption. Yet getting accurate read-outs of the basic numbers is fraught with difficulties. The official aggregate series, such as those furnished by ministries of agriculture and the FAO, are often unreliable; statistical services frequently lack the resources to do proper surveys and the data are sometimes subject to revision by the authorities for non-technical reasons. Independent surveys tend to be too limited in sample size, geographical coverage and frequency to serve as an adequate substitute.

Analysts tend to cope with the problem in one of two ways: either by using the official series 'for want of something better', or else by extrapolating, at least implicitly, from survey data or more qualitative field observations. Studies rarely go back and forth between survey and aggregate data, and they rarely make use of alternative indicators. As a result, contradictions and inconsistencies among the sources go unnoticed and unassessed. Each study tells a part of the story, but there is no sense of collective responsibility for getting the whole story straight.

This paper argues that we can do much better, as a profession, in making sense of the numbers, if we develop the reflex to navigate across data sources rather than staying boxed into a particular approach at a particular level of analysis. This means applying some simple principles of circumspection and cross-checking to the data we do have, on the one hand, and making a judicious use of surveys to fill in knowledge gaps, on the other.

Following a presentation of these data handling principles, the paper illustrates their application with an example from the groundnut subsector in Senegal. The conclusion discusses possible ways of improving data analysis capacity.

*Ellen Hanak Freud, CIRAD, Nogent s/Marne, France. The author thanks David Rohrbach and John Sanders for helpful comments on an earlier version.

PRINCIPLES OF ECLECTIC DATA HANDLING

Two principles need to guide data handling methods. First, the inherent quality of data is variable and needs to be assessed. Second, doubtful numbers need to be subjected to a rigorous cross-examination. The following checklist enumerates some simple methods for applying these principles.

How reliable are official data series?

At the minimum, annual series are available for crop area, yield and output, and for imports and exports. For agro-industries, other series like factory purchases and sales will also exist. These data are not equally accurate. Is the series an estimate or an accounting value? If it is an estimate, how good was the collection method? If accounting values were used, were there strong incentives to over-report or under-report?

Generally, the greatest level of uncertainty surrounds output estimates for crops consumed on the farm or sold through the informal sector, and yield estimates for smallholder crops. Food consumption and land-use data derived from these numbers are just as questionable. Trade data are usually more reliable, unless there are reporting biases or large quantities of unofficial trade. So are factory data for industrial crops, agricultural inputs and industrially processed foods. Output series for industrial crops are, as a consequence, reasonably accurate. Series on the apparent consumption of imported foodstuffs are more reliable than those for locally produced foods. Yield data will tend to be more accurate for estate crops, which can be obtained with a minimal amount of survey work, and for crops managed by project authorities if they have their own survey teams.

Having access to at least some reliable series concerning the problem to be analysed is invaluable for cross-checking the accuracy of more doubtful numbers.

What are the limits of occasional survey data?

For certain variables, statistical services only collect data at much longer intervals. These may include food intake, agricultural revenues, input use or stocks. Independent surveys are crucial tools for filling in knowledge gaps on these variables and on the annual series. Yet these types of surveys have their own problems. Cost considerations tend to severely restrict either sample sizes or the precision of the data-gathering methods, and frequently both. The resulting limits of the data need to be made explicit to avoid misinterpretation and extrapolation errors.

How accurate are the data-gathering methods? In most cases, surveys rely on respondent recall, which leaves more or less room for error, depending on the nature of the question. For example, consumption today is easier to recall than that of several years ago, farming practices are remembered better than labour requirements, output of crops sold can be checked, unlike amounts kept for home consumption. What are the likely biases in the responses caused by

the tendency to forget small transactions or to overstate consumption of certain items? Are there characteristics of the survey which could introduce biases, for instance strong seasonal effects or an unusual year? Is the sample size adequate to permit direct extrapolation?

Do alternative sources and indicators tell the same story?

It is almost always possible to find different types of data for the same phenomenon. If the story told by independent sources and methods is consistent, it is easier to place confidence in the numbers, especially if one of the sources is a relatively reliable aggregate series.

The most direct type of cross-checking is among different sources for the same variables, for instance independent field surveys and aggregate series on yields or output per farm household. Sometimes it is also possible to compare different types of indicators; for example, do household consumption data coincide with data on sales by the food industry? An extremely useful virtual indicator is obtained by putting the data into a less abstract form, to assess more directly whether the numbers make sense. For instance, translate the consumption data from per capita annual values into the number of weekly meals it implies, using local recipes and serving portions. Or calculate how many truckloads would be needed to move an estimated amount of informal trade. This type of 'reality check' helps reject estimates that are way off the mark.

Comparisons are often complicated by the use of different units of measurement: are the data expressed in net or gross yields, before or after processing? What are the loss rates or conversion rates used? Consistency checks obviously need to use a common base, and it sometimes takes digging to find out what the units are.

When there is inconsistency between uncertain aggregate data and independent survey results, the temptation may be to accept the latter as more reliable. One must first establish that the discrepancy does not stem from a sampling bias, especially given the small sample sizes of most surveys. Cross-checking can be a useful tool for verifying whether a sample is representative, especially when reliable aggregate series correspond to some of the survey questions.

Do the numbers add up?

Some fundamental relationships must hold among the basic aggregates of the agricultural economy. Production by definition equals the sum of its uses: local consumption, net exports and net stocks. The sum of area cultivated to different crops and fallows must equal the quantity of available agricultural land. Yields are generally a positive function of agricultural inputs and improved management practices. Consistency is not reassuring if some of the variables under consideration have been derived from these relationships. But it becomes a powerful analytical device when alternative sources can be exploited, particularly if some are known to be reliable.

AN APPLICATION TO THE SENEGALESE GROUNDNUT SECTOR

Groundnuts were the motor of the Senegalese economy throughout the colonial period and well into the 1970s, accounting for large shares of export earnings, rural employment and industrial output. Since the 1970s, their contribution to the formal sector has diminished radically, as sales to the oil industry have plummeted. The more modest decline in the official output series implies that the major change is the growth of an informal market for groundnut products (Figure 1). This widely held interpretation was bolstered by the fact that the parallel market became legal in the late 1980s, as a part of the liberalization programme begun earlier in the decade with the end of input credit programmes and cuts in seed distribution (Gaye, 1997).

Wishing to revive the subsector, the government and a major donor agency commissioned a study to verify the nature of the problem (CIRAD, 1997). By exploiting some basic accounting relationships in the oilseeds sector and a range of alternative data sources, including a light survey of groundnut farmers, the team was able, not only to raise doubts about the size of the parallel market, but to estimate its trajectory over the past 30 years. It also revised the series on production, re-estimated groundnut area and yields and, by extension, raised questions about the official trends in output of pearl millet, the other major crop in the farming system.

The principal accounting equation used was the requirement that output equal the sum of end-uses:

$$Q_G = S_R + S_{FS} + U_{FS} + U_{IS} + ST \tag{1}$$

where Q_G is groundnut output, S_R is farmer-held seed for the next season, S_{FS} is groundnut purchases by the oil industry for distribution as seed the next season, U_{FS} is groundnut purchases by the oil industry for processing, U_{IS} is

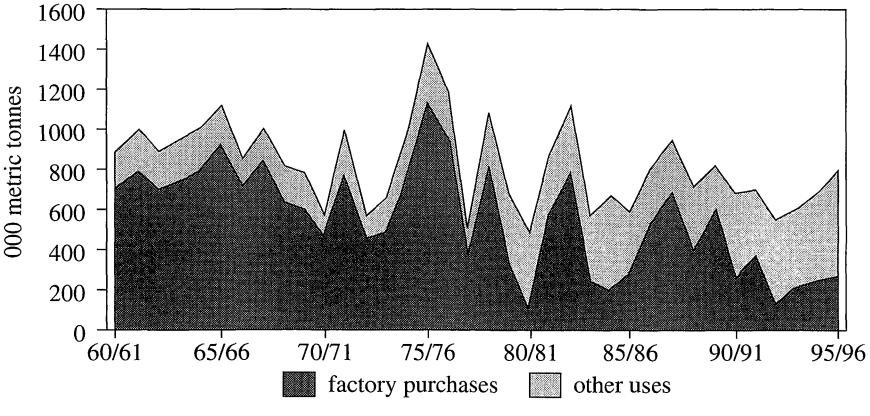

FIGURE 1 *Groundnut output estimates and factory purchases in Senegal*

groundnuts passing through the informal sector (market plus home consumption), and ST is farmer-held stocks other than seed, assumed negligible.

Both U_{FS} and S_{FS} are known values. A first approximation of S_R was made by assessing the additional quantities needed to cover the estimated area planted in year $t + 1$, using the coefficient given by the extension service (120kg/ha.). U_{IS}, the residual value on the right-hand side of equation (1), is hence the official estimate of informal uses. This level leapt from 100 000 tonnes in the 1970s to over 300 000 tonnes in the early 1990s.

To determine whether this amount was reasonable, it was then necessary to get a reading on the three components of U_{IS}:

$$U_{IS} = O_A / 0.28 + C_G / 0.7 + X_{IS} \tag{2}$$

where O_A is artisanal groundnut oil, C_G is groundnut for human consumption (in whole nut or paste form), X_{IS} is net informal groundnut exports to neighbouring countries, and the denominators indicate average conversion ratios to unshelled groundnut.

Although there was much talk of informal exports in Senegal, oil industry surveys of informal trade and customs data on edible oil imports of neighbouring countries made it possible to estimate actual flows of X_{IS} at only 5000 tonnes per year, the level of sales to Mauritania. So the explanation for the growth in U_{IS} had to be an increase in local consumption. For groundnut consumed in whole or paste form, the informal market and farmer reserves are the only sources, so total consumption corresponds to C_G in equation (2). But for edible oils (C_O), other sources exist besides O_A, particularly sales of refined vegetable oil by the oil industry (O_{FS}):

$$C_O = O_{FS} + O_A \tag{3}$$

Since O_{FS} is a known value, O_A can be calculated as a residual of equation (3) once one obtains a good estimate of total oil consumption.

The consumption estimates were based on two surveys. The first, by a nutrition research institute, measured household intake over a five-day period in a number of regions in the late 1970s and early 1990s. The second, by the study team, was based on rural household recall of oil use patterns at different points in the 1995/96 agricultural year and calculations of groundnut quantities remaining for home consumption after sales and seed use. The results of these studies converged. To obtain the final estimates, the data were also subjected to a 'reality check' in terms of numbers of typical oil and groundnut-based meals that they permitted consumers to eat each week. These estimates generously allowed for a per capita consumption of oil and oil equivalents of 16kg per year, a third higher than levels in neighbouring countries. It is worth noting that a third source, a household expenditure survey done for the World Bank in 1994, had to be rejected, because its estimates of spending on edible oils came out lower than industry sales of refined oil.

Even so, the resulting estimates of C_O and C_G suggest that there has not been a substantial growth in the local market for either groundnuts or artisanal

TABLE 1 *Area, output and yield estimates for groundnuts in Senegal*

	1960–69	1970–79	1980–89	1990–95
Official area (ha.)	1 087 000	1 111 000	1 033 000	875 000
Official output (t.)	932 000	862 000	761 000	667 000
Official yields (kg/ha.)	858	776	736	762
Revised output (t.)	932 000	837 000	652 000	464 000
Implied yields* (kg/ha.)	858	754	630	530

Note: *Calculated with revised output and official area.

groundnut oil. Although consumption has increased with population growth, so too have sales of (mainly imported) refined edible oil. The estimated value of U_{IS} comes out at only 120 000 tonnes in the early 1990s. By extension, this means that total output has been in the range of 450 000 to 500 000 tonnes, some 40 per cent below the official estimates (Table 1).

This revision also implied problems in the two components used to construct the official output estimate, area and yields. Recent official yields were of the order of 800kg/ha. – as high as 940kg/ha. in 1995/96 – as much or more than average levels of the 1970s. There were good reasons to believe that yields had fallen over this period. Although climatic conditions had not further deteriorated, demographic trends implied a decline in fallows. Fertilizer use on groundnuts (available from the input industry) had almost completely ended with the reforms in the early 1980s, and the quality of the seed stock had declined with cutbacks in the seed distribution programme soon afterwards. But had yields fallen as low as the revised output data implied (530 kg/ha.), or were area estimates also overstated?

To get a reading on this, the study used data provided by surveyed farmers on three variables: quantities of seed used (S), output (Q_G) and plot size (A_G) Although plot size estimates tend to be imprecise in these farming systems, the first two variables are well known by farmers. As such, one key element of yields, the seed multiplication coefficient (Q_G/S), could be fairly reliably assessed. This figure came out at 5 to 1, down from 7 to 1 in the years when seed quality was higher and general input supply conditions better. Before translating this into yields (Q_G/A_G), the ratio for the seed density (S/A_G) was verified, as field work from the late 1980s had indicated this was increasing (Kelly *et al.*, 1996). The survey results found average densities of 140kg/ha., a slight increase over the extension norms of 120 kg/ha. This result was consistent with farmers' responses to qualitative questions on this practice. *In summary*, the survey's yield estimate of 700kg/ha. for the 1995/96 crop year also implied that there were 130 000 fewer hectares planted to groundnuts that year than shown in the official series. In some years, the area overestimate is probably even larger.

Such major revisions in the numbers also raise questions about what is happening to pearl millet, the other principal food crop in these systems. Pearl

millet accounts for roughly two-thirds of all cereal production in Senegal, and a far greater proportion of cereals in the groundnut basin, since rice, maize and sorghum are mainly grown in other regions. The official series show a long-term decline in millet output per rural inhabitant – moving from 170kg in the 1960s to just over 120kg in the 1990s – and implying a decline in on-farm consumption, marketed surpluses, or both. The data show no progression in either area or yields since the mid-1980s.

Yet various indicators suggest that the millet sector is far from moribund. For instance, per capita imports of rice and wheat have levelled off since the mid-1980s, there has been a rapid spread of millet-processing facilities in urban areas and of mechanical threshing equipment in the countryside, and there is a well established community of grain traders in major production and consumption areas. By the accounts of all observers (including the market information service which collects price and quantity information) the heart of the grain-belt is the Sine-Saloum region. This is the very region where groundnut output was grossly overestimated. It is thus likely that the 'missing' groundnut area is actually being farmed with millet, a positive 'flip side' to the downward trends in the groundnut sector.

UNFINISHED BUSINESS

Clearly, the identification of data problems such as those illustrated underscores the need to improve the quality of collection methods for the basic series. But this is easier said than done; at the time of the study in Senegal, the statistical services of the Ministry of Agriculture were in a third phase of a donor-funded project to improve crop estimates.

Until things improve, it is imperative to develop a planned effort to cross-check key variables. This means overcoming institutional barriers to making a better use of existing information. The compartmentalization of various statistical services and research programmes limits the reflex to cross-examine data of different types and the scope for data exchange. People need to know who is doing what, and to have the possibility to share information and debate its interpretation. One solution, being tried in several African countries, is to establish 'observatories' of strategic subsectors – small teams of analysts responsible for pooling information from different sources and making sense of it all.

More survey work is also necessary for understanding the trends. But what kind of surveys? When is informant recall adequate, and when must one take direct measurements? How big do samples need to be, and what kinds of sampling and extrapolation techniques are appropriate? To be sure, a theory base exists on these questions, but there is an unresolved tension, in practice, between statistical rigour and the available means. Time and cost considerations both imply opting for light surveys whenever possible, but in this activity there is still a lot of 'muddling through'. Statisticians, economists and farming systems specialists need to work together to establish pragmatic guidelines for raising the value-added in reliable data supply – one of the scarcest agricultural commodities that there is!

REFERENCES

CIRAD (1997), *L'arachide au Sénégal: Un moteur en panne*, Paris: Karthala-Cirad.

Gaye, M. (1997), 'Le phénomène du marché informel de l'arachide du Sénégal', *Arachide Infos*, **7**, 3–8.

Kelly, V., Diagana, B., Reardon, T., Gaye, M. and Crawford, E. (1996), *Cash Crop and Food Grain Productivity in Senegal*, International Development Paper no. 20, Michigan State University, East Lansing.