WORLD FOOD SURVEY

William H. Wigton

Statistical Reporting Service

U. S. Department of Agriculture

1977

Abstract

In recent years, much literature has been written about the importance of knowing amounts of food on a worldwide basis. It is not the intent of this author to present or develop those arguments--they have been well documented. However, it is the intent to spell out a reasonable procedure to estimate world food with present technology at reasonable costs. The advantages and costs of this procedure will be presented. If interests are generated, details can be provided.

At a later point in time, if LANDSAT were to become available in real time, substantial improvements could be possible using LANDSAT in another way.

INTRODUCTION

The Statistical Reporting Service, USDA, makes crop, livestock, income, and price estimates throughtout the year. The crop and livestock estimates are based on a nationwide June Enumerative Survey (JES) that is a probability survey of land parcels selected from the total land area in the U.S. Since every parcel of land in the U.S. has a known chance of being selected, all areas, whether selected or not, are represented by the sample. Inferences about the total U.S. are possible and can be quite good. The accuracy of these estimates depends on the survey design (sample selection techniques, estimation procedures) sample size, and quality of enumeration.

The estimates are generated from an area sampling frame (the sampling frame is the universe broken down into its component parts with direction how to get each part, if selected. An area sampling frame is a sampling frame of all land areas. The concept is simple: "Divide the total land to be surveyed into N small blocks without any overlap or omission; select a random sample of n blocks (called segments); obtain the desired data . . . and estimate the population value, multiply the sample totals by N/n." [1] The most efficient frame is a frame where the universe of interest has been divided into equal parts with respect to the characteristic to be estimated so that only a few segments will need to be selected in order to accurately estimate the entire population. The reason that a sampling frame would be cost effective for worldwide estimates is that it takes nearly as large a sample to estimate a commodity such as wheat for the State of

Kansas as it does for the U.S., or the whole world, for that matter. [2]

U.S. AREA FRAME

Generally speaking, there are two types of area frames. The first area frame developed was called the master sample of agriculture. It was developed in the 1940's to obtain information about the farm sector. The sampling strategy called for dividing the total area in the rural community into blocks--each block with the same number of households.

The selection procedure is to number the blocks, select the random numbers from a table, and interview the households inside the blocks with the selected numbers. For demographic type data, the strategy works well. Many Southern states still use this master sample concept although it is being replaced as quickly as possible.

The frame strategy that is replacing the master sample is called land use area frame, instead of putting the same number of households into a block, we try to equalize the blocks with respect to land area.

The first step is to divide the total land area into huge pieces of land--each piece of land being as homogeneous as possible. For example, one homogeneous category might have from 75 to 100 percent of the land cultivated for crops. A second would have 50 to 75 percent of the land cultivated. Once all the land has been divided into these large areas, we assign the areas to strata by putting similar amounts of cultivated land together. These large areas are divided into smaller sampling units--as equal in size as possible--and are assigned numbers for sample selection. The interview strategy is that all land inside the selected sampling unit must be accounted for.

The estimates that are generated from the area sampling frame have coefficients of variation (CV) of 2 to 6 percent at state levels for most important items, and 1 to 3 percent at U.S. levels ("Coefficient of Variation" (CV) is a term used by sampling statisticians to estimate relative precision of sample estimates. The CV of any estimate is equal to (standard error of the estimate)/(value being estimated)).

ANALYSIS OF USDA DATA

If we assume for a moment that an area sampling frame was available on a worldwide basis similar to the two types of frames in the U.S., we can use U.S. data to compute sample size

requirements and costs for the world or the individual countries, depending on what is needed. Sample size requirements needed to estimate corn, wheat, and soybeans to obtain a CV of 2 percent, 5 percent, and 10 percent can be found in Tables 1 and 2.

These results can be interpreted in several ways:

1. The land use sampling frame is more efficient in the long run because sample size requirements are greatly reduced.

2. Both of the above frames are general purpose frames and they obtain information on many crops at the same time. Information on any one crop is inexpensive.

3. Sample size requirements agree to a certain extent. For example, the land use area frame requires samples of size between 500 to 700 to get estimates with CV'S of less than 2 percent error.

PROCEDURES

What is needed to make estimates on a worldwide basis is first, an area sampling frame for the major exporting and importing countries; then, an area frame for the rest of the world. Once this is accomplished, samples can be selected to estimate wheat, corn, rice, income, livestock, etc. Sampling strategies will be designed to optimize information received for a fixed cost.

PRIORITIES

Priorities should be set. Food economists must decide which crops need to be estimated and to what degree of accuracy. It must also be decided whether a true worldwide estimate is necessary or whether a survey that covers some lesser percent of the total acreage is sufficient.

ENUMERATION COSTS

Each of the above points will have an impact on the total cost. If an estimate for wheat on a worldwide basis is all that is needed, then a very efficient sampling frame can be developed. Once the frame is available, an accurate estimate of wheat could be produced very cheaply--800 segments would seem to be sufficient.

To obtain more information about other commodities, compromises will need to be made in stratification and sample

Table 1.--Master sample frame sample size requirements to obtain specified accuracies.

Area	Crop	Percent		
		2	5	10
Country or World	Corn	1054	169	42
	Wheat	1825	292	73
	Soybeans	1836	294	73

Table 2.--Land use area frame sample size requirements to obtain specified accuracies

Area	Crop	Percent		
		2	5	10
Country or World	Corn	500	80	20
	Wheat	631	102	26
	Soybeans	463	75	19

allocation. To estimate wheat, corn, soybeans, cotton, and livestock on a worldwide basis, 1500 segments may be needed to obtain a specified CV.

In the U.S., it costs $60 to enumerate a single segment. It could easily cost $100 if a worldwide survey were developed. This means it would cost $150,000 to obtain estimates of these commodities on a worldwide basis. On the other hand, if good estimates are needed on an individual country basis, the total costs would increase.

COST OF FRAME

In theory, the U.S. is divided into 3 million parcels of land that could be sampling units. In practice, it is not done. What is done, is to divide the total area into somewhat larger parcels (about 10 times the sampling unit size) and a sample of these is selected to be divided further into sampling units. A frame for Illinois using this procedure costs $20,000 to $25,000. A cheaper frame could be constructed; however, a frame that is developed with much care is well worth its cost because the sample size requirements are much smaller to obtain a given accuracy. Table 3 shows estimated funds needed to construct the land use frames. The cost may depart somewhat from these figures, but I have tried to be conservative. About 5 percent of these costs are materials and the rest labor.

SATELLITE DATA

An area frame ties in beautifully with satellite data such as LANDSAT imagery in the following ways. One critical step in using multi-spectral scanner computer compatible tapes is to identify and estimate the amount of reflected energy for each crop. Area frame data can provide unbiased estimates of crop reflectances in the whole image since the data are selected from the entire image.

Another equally critical step in making use of satellite imagery is to be able to convert classified points of satellite imagery of a crop into acres of a crop. By observing how known probability data is classified or more accurately, misclassified, adjustments can be made in total image classification so that unbiased acreage estimates can be vastly improved without increasing the sample size. Without the area frame, it is nearly impossible to remove biases created during classification. [3]

SUMMARY

Obviously, the assumption has been made that ground data collection is possible in foreign countries. At least, it can be expected that nations will cooperate and collect their own information in most of the world. If this is true, let us use the best available, most cost effective techniques in the large majority of the important countries. This will limit the guesswork to 20 percent or less.

Either for political or ecomonic reasons, some countries may not wish to cooperate in this endeavor. However, the countries that do have the sampling frame system would be able to forecast and monitor the crops early in the crop year and throughout the growing season. The participating countries could have crop

Table 3.--Estimated costs to construct a land use area frame for various countries.

Area	Total hectares	Agriculture hectares	Time to construct area frame
	Thousands	Thousands	Man-years
Africa			
Egypt	100,000	3,000	2
Ethiopia	122,000	80,00	4
Morocco	44,660	15,000	3
South Africa	122,100	110,000	8
North America			
Canada	997,000	70,000	8
Mexico	200,000	100,000	8
South America			
Argentina	278,000	170,000	8
Bolivia	109,860	30,000	4
Brazil	850,000	130,000	8
Asia			
China	960,000	327,000	8
India	330,000	200,000	8
Indonesia	191,000	30,000	4
Europe			
Bulgaria	11,000	5,000	2
Czechoslovakia	12,000	7,000	2
France	55,000	32,000	3
Hungary	9,300	7,000	2
Italy	30,000	17,000	2
Poland	31,000	20,000	2
Spain	50,500	34,000	3
Yugoslavia	25,600	15,000	2

estimates early and could be negotiating for export-import reserves.

Any system that depends entirely on satellite data and has no other alternative will not be dependable since clouds are usually quite prevalent, especially over good agriculture areas. The exception may be certain wheat areas, but a good wheat estimate is

only a partial answer since many of the major grains are substitutable.

The area sampling frame is a reasonable, practical, accurate, cost effective way to produce estimates for a state or a nation. This is even more true for the world, since the sample size requirements do not increase as N, the population total number of sampling units increases. Further, almost anything can be estimated with an area sampling frame from crop acreages and yields to livestock and farm income. In addition, the area frame is necessary to fully utilize satellite imagery.

References

[1]Houseman, Earl E., Area Sampling for Agricultural Statistics. Statistical Reporting Service, USDA, 1975.

[2]Cochran, W. G., Sampling Techniques. Second Edition, p. 23.

[3]Von Steen, Donald H. and Wigton, William H., Crop Identification and Acreage Estimation Utilizing LANDSAT (ERTS) Imagery. Statistical Reporting Service, USDA, 1975.