RESEARCH AND DEVELOPMENT FOR STATISTICS IN FOOD AND AGRICULTURE: THE BAECON EXPERIENCE[1]

Mamerto D. Damasco[2]
David C. Besa

INTRODUCTION

Statistical data collection on a sampling basis as employed in the Philippines, is the universally accepted technique for amassing reliable data that is cost efficient and easily outputted in a relatively short time. We classify such sample surveys on food and agriculture into three main types: 1) household interview surveys, 2) farm book-keeping survey and 3) objective area and yield surveys.

Of course, as in any activity carried out by people, none of these survey types is devoid of technical and administrative flaws. First: the accuracy of the results of the household interviews is conceptually "inferior" to those of the other two types because of non-sampling error. If household interviews could provide precise and accurate results and could be conducted at reasonable cost, such surveys could provide adequately straightforward as well as structural statistics. Such an ideal enumeration method, however, does not exist. Second: farm book-keeping projects provide data for almost all types of farm and household characteristics but may be subject to serious bias in the sample, mainly because of the problems of selection of random cooperative and knowledgeable sample farmers. And third: while objective area and yield surveys may greatly improve the quality of

---

[1]Paper presented at the Annual Convention of the Philippine Statistical Association held at the Central Bank of the Philippines, June 29, 1979.

[2]Mr. Damasco is Officer in Charge, Office of the Assistant Director for Statistical and Administrative Operations of the Bureau of Agricultural Economics; Mr. Besa is head of the Statistical Research and Development Section, Statistics Division. Both are involved in directing R & D programs of the Bureau.

area and yield data they cannot provide structural statistics because the survey data are not obtained directly from the farm operator who could furnish information on other desirable farm and household characteristics.

It is thus of practical value to bear in mind that these three types of agricultural surveys, when undertaken individually, can provide only partial requirements for data of the quality needed in the formulation of agricultural plans and programs. In other words, too much should not be expected of any one type of agricultural survey conducted; and the role that any type plays in the agricultural data collection system should be well understood.

The Bureau of Agricultural Economics through its statistical research and development arm aims to make improvements on each of these types of operations, and likewise aims to operationalize and use these three types as links that should lead to the improvement of the quality of data available on food and agriculture in this country.

ORGANIZATION OF R & D AT BAECON: A PERSPECTIVE

Organized efforts to undertake statistical research activities at BAEcon were initiated as early as 1969. In that year, BAEcon implemented in Nueva Ecija a USAID-assisted project known as Counterpart Project 70169, to evolve, test and demonstrate improved techniques in agricultural data collection. A consultant on statistics from the USDA was engaged as project consultant for a period of six months. Experiments on crop cutting and objective area measurement for interview checks were undertaken. Other important elements of survey operations which were developed and tested in Nueva Ecija were the following:

1) Good questionnaire design
2) Simplified and well organized manual of instructions
3) Objective method of selecting field data collectors
4) Training of interviewers, enriched with oral and written practice exercises, as well as field practice
5) Effective programming of field activities
6) Closer field supervision
7) Better mobility of supervisors and interviewers

The project (CP 70169) made possible the training of six key officials of the Bureau in the United States: one attended a year's course on statistical research and development in 1973 at USDA; two completed a 5-month R & D-oriented course at USDA, also in 1973; and two obtained Master of Science degrees in Agricultural Economics and Statistics in 1974. Likewise, all key provincial and regional personnel underwent a 3-week training on all the phases of survey work tested in Nueva Ecija in 1969-1970.

In 1973, the BAEcon Statistics Division was expanded with the creation of the R & D Section. However, mainly due to the initial lack of personnel, the new section, until about the end of 1977, was functionally and physically lodged with the Survey Plans and Designs Section of the same Division. The personnel of the fused units were primarily engaged in the planning, organization and administration of the BAEcon statistical surveys so that very little of their time was devoted to statistical research activities. Studies were confined to the improvement of the design of the Integrated Agricultural Surveys (IAS), which led to the splitting of the IAS in 1976 into two major survey operations: 1) Rice and Corn Survey and 2) Livestock and Poultry Survey. The latter survey employed a multiple frame which provided more precise estimates. Double sampling was applied to the RCS sample for the surveys on farm prices and wages, which subsamples were enumerated simultaneously with the RCS samples for purposes of economy.

During the interim period (1973-1977), as more funds became available for the hiring of more personnel, a continuing upgrading program for personnel to be assigned in the RDS was undertaken. In the summer of 1977, an agreement was forged between BAEcon and UPLB for the latter to develop and administer a statistical research-oriented training program (which was funded by BAEcon under its 4-year Ford Foundation supported personnel development program). Seventeen BAEcon personnel completed the summer course. Also under the BAEcon Ford-Foundation supported program, four are presently completing their thesis for a Master of Science in Statistics at UPLB. Another technician will be returning within three months with a Master of Science in Statistics from Kansas State University. These trained and experienced statisticians form the personnel nucleus of the RDS which was finally made operational early in 1978 as a distinct unit of the Statistics Division.

INITIAL EXPERIENCES GAINED FROM OBJECTIVE AREA AND YIELD SURVEYS

Area measurement

This activity was undertaken in Nueva Ecija (under CP 70169) to demonstrate the feasibility of supplementing household surveys with objective area surveys, the latter designed to check the results of the former. However, due to limited time and resources, only sixty purposively selected farms for the entire demonstration province were covered. The original plan to employ double sampling in the selection of samples to be covered was not carried out. The results, therefore, may not be truly reflective of conditions obtaining in the whole area but should only provide a numerical indication of response bias arising from the interviews.

It may be observed from the results obtained for the 60 farms that on the assumption that physical measurement gives the correct area, the farmers under-report the area of their farm by about 2.87 percent (Table 1). However, this statement is not true for all types of

Table 1. Palay area as reported through interview and as measured by type of crop and variety harvested, Nueva Ecija, 1970

Item	*Number of Farms Covered*	*Farm Area (Ha.)*		
		Reported	*Measured*	*Error (%)*
Lowland, irrigated	33	62.71	67.48	7.04
a. HYV	12	24.55	26.81	8.43
b. Traditional	21	38.16	40.67	6.17
Lowland, not irrigated	27	58.90	57.73	2.03
a. HYV	10	23.50	22.67	3.66
b. Traditional	17	35.40	35.06	0.97
ALL FARMS	60	121.61	125.21	2.87

Source: Unpublished BAEcon Report on CP 70169.

palay crops. Farmers whose farms were not irrigated overestimated their farm area by 2.03 percent. This could provide indications that response errors are random in nature. Thus, if errors are not systematic as recorded during the experiments, it may be assumed that if we look at the expanded results of the household survey for the whole province, the total response errors would possibly become very small.

Since some farmers report their farm area "via the seeding rate" or on the basis of the quantity of seeds sown, it is highly probable that the correct seeding rate for high-yielding varieties at that time had not been accurately established, which may explain the larger errors for farms devoted to high-yielding varieties both for each type of palay crop. Table 2 shows that the quantity of seeds for irrigated farms is less than that for non-irrigated farms, which is rather obvious since most, if not all, of unirrigated palay crops were planted without deliberate concern for distancing wherein the density of planting is normally heavier than those of irrigated farms where the straight line row method is more prevalent. The results show an average of 19 gantas per hectare and 22 gantas per hectare of lowland irrigated and lowland non-irrigated farms, respectively.

Table 2. Farm area and quantity of seeds used by variety and type of farm, Nueva Ecija, 1969-1970

Item	*Number of Farms Covered*	*Average Farm Size (has.)*	*Total Area as Measured*	*Quantity of Seeds Used (gantas)*	
				Total	*Per hectare*
Lowland, irrigated	33	2.04	67.48	1,261	19
a) HYV	12	2.32	26.81	458	17
b) Traditional	21	1.94	40.67	803	20
Lowland, not irrigated	27	2.14	57.73	1,293	22
a) HYV	10	2.3	22.67	533	24
b) Traditional	17	2.06	35.06	760	22
ALL FARMS	60	2.09	125.21	2,554	20

Source: Unpublished BAEcon Report in CP 70169.

Crop Cutting Experiments

A logical approach to the measurement of production of a given crop like palay which is grown by millions of farmers in scattered holdings is to consider the two components of production – the area and the yield separately, and measure or estimate each. Multiplying two figures gives the total output. In the Philippines, this approach is not practicable, hence, has not been adopted. Filipino farmers could more easily and conveniently give total production figures rather than the yield rate. Yield estimates of crops in the country are then derived from the reported production and area.

On the other hand, the crop cutting experiment aims to demonstrate an objective method of measurement to determine actual palay yield at harvest. Two methods of crop cutting were used: 1) area method and 2) line method. A total of 22 sample parcels (pitaks) were covered – 11 samples for each method used.

Table 3 shows that the area method appears to be more efficient than the line method. Errors of the computed production estimates of the sample parcels based on the sample cuts taken as a whole, ranged from 0.01 per cent to 1.32 per cent for the area method; errors ranged from 0.86 per cent to 26.9 per cent for the line method. Summing up the results for each method, the errors for area and line methods were recorded at 0.17 per cent and 1.71 per cent, respectively.

In view of the small number of purposively selected sample parcels covered, the results of the experiments should not be taken as fully reflective of the relative efficiency of both methods employed. In fact, as observed during the field operations, either method provided satisfactory results depending on the stand of the crop, the size of sample parcel, method of planting, and the manner the crop cutting procedures were implemented. For instance, the very large error of 26.9 per cent recorded for one sample parcel cut by line method, was discovered, after careful checking, to have been caused by carelessness of the survey teams; three bundles of palay harvest from the sample cuts were missed during the threshing. Since the experiments were undertaken purposively as training exercises for BAEcon technicians, the records for this particular sample cut were not corrected to dramatize the extreme need for carefulness in executing all phases of the crop cutting exercises.

Table 3. Results of crop-cutting experiments in palay, by variety, by type of planting and farm, Nueva Ecija, 1969-1970

Variety and type of farm	AREA METHOD: Number of sample parcels	AREA METHOD: Production from sample parcels: As computed[a]	AREA METHOD: Production from sample parcels: Actual[b]	AREA METHOD: Error	LINE METHOD: Number of sample parcels	LINE METHOD: Production from sample parcels: As computed[a]	LINE METHOD: Production from sample parcels: Actual[b]	LINE METHOD: Error
	Number	Kilogram	Kilogram	Percent	Number	Kilogram	Kilogram	Percent
1. Lowland, irrigated	10	23,251.44	23,252.62	0.01	11	36,409.50	37,041.68	1.2
a) High Yielding	9	20,636.39	20,653.85	0.09	10	35,528.59	35,837.03	0.86
b) Ordinary	1	2,615.05	2,598.77	0.63	1	829.91	1,204.65	26.96
2. Lowland, not irrigated	1	3,688.40	3,640.50	1.32				
a) High yielding								
b) Ordinary	1	3,688.40	3,640.50	1.32				
TOTAL	11	26,939.84	26,893.12	0.17	11	36,408.50	37,041.68	1.71

[a]Computed harvest based on sample plots

[b]Total harvest sample parcels.

Source: Unpublished BAEcon Report on CP 70169

CURRENT DIRECTIONS OF THE BAECON R & D PROGRAM

On List Frame Improvement (Barangay Frame)

Auxillary information associated with every sampling unit of the population is crucial in planning efficient sampling strategies. If its linear relationship with the variable of interest is sufficiently high, a sizeable reduction of sampling error could be achieved through its use as a concomittant variable in regression-type estimators, as a measure of size in pps sampling, or as a stratification variable in stratified sampling.

The major problem of sampling statisticians particularly those who are engaged in statistics for agriculture, is the paucity of auxillary data. Because of this, very often the sampling statisticians are forced to settle with the less efficient sample design, or by allocating part of their resources for the collection of such data. It was under these circumstances that the Barangay Screening Survey (BSS) was devised and introduced nationwide in 1971 by BAEcon. The BSS was designed to collect barangay level data through interview of barangay captains or a group of knowledgeable farmers residing in the barangay. The primary purpose of the survey was to collect and store the barangay data systematically for easy retrieval in the preparation of the design of agricultural surveys.

a. Earlier findings on the BSS data

How reliable are BSS data? This had long been asked even before the BSS was introduced. The answer to the question however has never been fully satisfactory. Nevertheless, the BSS was conducted nationwide mainly due to the very encouraging results of the Nueva Ecija Pilot Project. According to findings, the BSS data, when used as a stratification variable, would significantly improve the efficiency of a sample survey.[3] It was mainly the use of the BSS data as a stratification variable in the design of the BAEcon Integrated Agricultural Survey (IAS) which brought about reliable statistics on rice and corn.

A study[4] covering the seven provinces of the Ilocos region and three provinces of Central Luzon (Bataan, Bulacan and

---

[3]Unpublished report on the 1969-1970 USAID-NEC Pilot Project.

[4]David, Isidoro P. Non-Sampling Errors in Agricultural Surveys, The Philippine Statistician, January-June 1979, Vol. XXVII.

Tarlac) investigated the reliability of the BSS data. To throw more light on the findings of this study, let us consider some important background information about the data used.

The data compared with the BSS was derived from the house-to-house enumeration of households known as the BAEcon Household Screening Survey (HSS). All household heads in a barangay were asked to report the area of parcels of land regardless of location i.e., inside or outside the sample barangay used for: a) palay only, b) corn only, c) palay, corn and other crops (alternate and or intercropped) and d) other crops only. The individual responses were aggregated for the barangay, which summaries were then compared with the corresponding BSS data.

On the other hand, the BSS data includes only crop areas which are located inside the barangay; unlike the HSS, it does not include areas which are located outside operated by farmers living in the barangay. The concepts used in both surveys, therefore, are not comparable, particularly on 1) areas of crop land and 2) inventory of livestock and poultry. The authors believe that this is one of the major reasons for big differences and low correlation between the BSS and HSS data, particularly on crop areas as found by David. The following findings (of David) therefore would seem to need further verification:

1) Gross over-reporting for crop area.
2) Number of households understated by eight percent in Ilocos region and overstated by 16 percent in Central Luzon.
3) Considerable errors in livestock number ranging from 23 to 40 percent in Ilocos and from 20 to 104 percent in Central Luzon.
4) Low correlation between the BSS and HSS data.

Summarily, if the concepts used in both surveys were entirely comparable, the above findings would seem to indicate the lack of knowledge of the barangay captains on the data asked for in the BSS.

b. *Conceptual sources of errors in the BSS data*

Concerned with the existence of large errors in the BSS data, the Bureau reviewed the system and tried to identify the sources of errors.

Conceptually, the total error in the BSS data consists of the error of response and the interviewers' errors. The errors of the BSS may result from one or more of the following:

a) *Faulty perception of the definitions used in the survey.*

   This would generally result in double counting in the case of crop areas. For example, when the respondent is asked about land used for palay only (or any other crop) he may include land used for palay which is alternately planted with corn; and when land used for palay and corn only is asked, the same area will likely be reported again. Thus an area used alternately for three different crops during the year may be magnified twice the effective area.

b) *Lack of knowledge about the data asked which leads the barangay captain to make "guesses" during the interview.* The errors introduced by this type of respondent result in a low correlation of the BSS data and the HSS data.

c) *Sudden shifting of crop areas in the barangay which is unknown to the BSS respondents.* The respondent's natural reaction without prior knowledge of such changes is to give a response based on his past observations, which may not hold true at the time of interview.

CURRENT FINDINGS ON THE 1976 BSS DATA

In the recent evaluation made on the validity of the BSS data, the Bureau considered the heavily sampled special surveys conducted in 1978. One is the Coffee Survey which was undertaken in eight towns of Cavite involving 103 sample barangays, and the Garlic Survey in Ilocos Norte and Ilocos Sur covering 90 and 19 sample barangays, respectively. Following the procedure employed by David, the HSS coffee and garlic areas were compared with the corresponding BSS data.

Tables 4 to 6 show the differences between the BSS and HSS data categorized according to barangay sizes. The negative and positive differences of the two independent estimates indicate under-reporting and over-reporting, respectively. The following highlights may be gleaned from the tables:

1. The BSS coffee area in Cavite was accurately reported with a minimal under-reporting of less than one percent; garlic area

was over-reported by 5.2 percent and 22.8 percent in Ilocos Sur and Ilocos Norte, respectively.

2. Under-reporting tends to be very apparent in barangays with very small crop areas, and over-reporting in barangays with very large crop areas.
3. Cumulative under-reporting for barangays with small areas is recorded at 23.5 percent of the HSS-based coffee areas in Cavite; cumulative over-reporting is registered at about 23 percent. This could explain the low total error in (1) of less than one percent.
4. The cumulative under-reporting on garlic areas is placed at 37 percent and 39.5 percent in Ilocos Sur and Ilocos Norte, respectively; over-reporting at 42.2 percent and 62.4 percent, respectively.

The analysis of the Ilocos garlic data seem to agree with David's findings that BSS crop areas are generally over-reported. (Incidentally, David also made use of Ilocos BSS and HSS results.) However, the degree of over-reporting when *individual crop* are considered is very much smaller than what he earlier reported. This finding would seem to indicate strongly that double counting of area which is alternately planted to two or more crops really exists in the BSS data; and this is amplified during the aggregation process for all other crops. To illustrate: suppose a piece of land is cultivated to three different crops in succession during the year. If the BSS collection methods are not properly understood by the interviewer or the respondent, it is probable that the same piece of land could be reported three times, i.e., one for each of these crops. Thus, when individual crops are considered, as was done in this recent BAEcon study, the error may become small. When the total area for all crops is taken, for this example of three crops, the errors of the BSS data may consist of twice the physical area plus three times the error of reporting the physical area.

How useful is the BSS data in the planning of agricultural surveys? The usefulness of BSS data for purposes of sampling does not depend entirely on whether it is significantly different from the "true" data. Even if the relative difference are large, for as long as their correlation is high, the BSS data could still prove as an effective device for reducing sampling errors in agricultural surveys. Analysis showed that the correlation between the BSS and the HSS coffee

Table 4. Differences between the 1976 BSS coffee areas by size and those of the 1978 BSS sample barangays in Cavite province

Size[a]	Number of Barangays	Area		Dif.	Cum.
		BSS	HSS		
Less than 1.00	13	3.25	300.12	-296.87	-296.87
1.00 – 20.00	34	298.66	828.09	-529.43	-826.30
20.01 – 40.00	15	416.90	738.36	-321.46	-1147.76
40.01 – 60.00	10	502.00	625.03	-123.03	-1270.79
60.01 – 80.00	7	512.00	417.19	95.32	-1175.47
80.01 – 100.00	5	474.00	236.83	237.17	-938.30
100.01 – 150.00	11	1,385.00	1,082.11	302.89	-635.41
150.01 – 200.00	4	765.00	423.51	341.49	-293.92
Over 200.00	4	1,026.50	756.41	270.09	-23.83
TOTAL	103	5,383.82	5,407.65		

[a]Based on the BSS data.

Table 5. Differences between 1976 BSS garlic area by size and those of HSS of samples in the 1978 Special Survey on Garlic, Ilocos Norte

Size	Number of Barangays	Area		Dif.	Cum.
		BSS	HSS		
Less than 1.00	8	0.25	69.71	-69.46	-69.46
1.00 – 5.00	4	17.00	38.79	-21.79	-91.25
5.01 – 10.00	1	10.00	9.50	+ 0.50	-90.75
10.01 – 25.00	3	39.00	48.95	- 9.95	-100.70
Over 25.00	4	220.00	105.12	114.88	14.18
TOTAL	19	286.25	272.07		

Table 6. Differences between 1976 BSS garlic area by size and those of HSS of sample barangays in the 1978 Special Survey on Garlic, Ilocos Norte

Size	*Number of Barangays*	*Area*		*Diff.*	*Cum.*
		BSS	*HSS*		
Less than 1.00	33	2.85	145.83	-145.98	-142.98
1.00 – 5.00	29	82.00	188.55	-106.55	-249.53
5.01 – 10.00	10	92.00	120.03	-28.03	-277.56
10.01 – 25.00	8	155.00	86.30	+68.70	-208.86
Over 25.00	10	530.00	161.10	368.90	+160.04
TOTAL	90	861.85	701.81		

area in Cavite is 0.49; the garlic areas at 0.58 and 0.47 for Ilocos Sur and Ilocos Norte, respectively. What is really bothering is the existence of highly significant over-reporting and under-reporting of areas for barangays of extremely small and large sizes of croplands. This could distort the correlation of the BSS and HSS data.

ON NON-SAMPLING ERRORS

Area Measurement

Rationale of study. The demonstration project undertaken in Nueva Ecija on the errors in crop area estimates obtained through interviews was confined to the determination of the magnitude of response errors. Because of the constraints mentioned in earlier sections, there was no attempt made on putting to advantage the high correlation between the interview area and the corresponding measured area. To be able to take advantage of such relationship in enhancing the accuracy of the interview data, the measured area must be taken from a sub-sample of the total sample of surveys employing the interview method. The expanded experiment on area measurement undertaken in 1978 in 11 major rice-producing provinces (described below) was designed to provide experiences on how the objectively gathered data can be used to improve the reliability of estimates derived from household interview surveys, (see Expansion of Objective Area Surveys), as well as in providing more information on the nature of non-sampling errors obtained at the provincial level.

BAEcon's objective area measurement survey. The survey was originally planned to cover 12 provinces, one from each of the 12 regions of the country. However, in the actual implementation of the survey, South Cotabato was dropped from the survey. About 1,200 sub-sample farms of the Rice and Corn Survey (RCS) were selected at random from the list of RCS sample households with crops expected to be harvested during the survey period. (The survey was confined to such farms with rice expected to be harvested because the Crop Cutting Survey was linked to this survey, i.e., palay farms to be measured shall also be crop-out.) Because of manpower problems, only Pangasinan and Isabela were able to cover most of the sample farms selected for the survey.

Methods used in area measurement. Steel tapes (30m) and Rangematics were used in measuring the sample farms. The latter is a simple optic intrument capable of measuring linear distances of more than 95 percent accuracy when properly calibrated. The use of this instrument, however, is allowed only in measuring distances where it is extremely difficult to use the steel tape, or in cases where the farmer does not allow dragging of the steel tape over the rice plants. To minimize errors in measuring distances with the rangematic, two sets of independent measurements are taken from opposite ends of the distance taken by two technicians with the same instrument. If their average reading varies by more than 1.5 meters for distances more than 50 meters, another pair of readings are taken. The mean of their average readings should be within the specified limits. For distances of less than 50 meters the allowed error is less than one meter. The farm is divided into several triangles and every corner is staked. Corners requiring the use of the rangematic are staked with poles long enough to be clearly visible from either end. The distances between corners are then measured in accordance with the procedures discussed above. The area of each triangle is obtained by using Hero's formula.

Evaluation of the area measurement survey results. Out of the original sample of 1,200, only 678 were measured. Of this number actually received, only 637 survey returns were analyzed. The remaining 41 were not included due to glaring errors, either in the measurement of area and/or in the accomplishment of the survey forms. The results are summarized in Tables 7, 8 and 9.

The salient findings are as follow:

a) The errors in Pangasinan, Nueva Ecija, Misamis Occidental and North Cotabato are not different from zero.
b) Results in Laguna and Zamboanga del Sur indicated significant under-statement of area at $\alpha = .05$.
c) Areas in the remaining five survey provinces were significantly over-stated at $\alpha = .01$.
d) There are varying degrees of errors ranging from negative values in the Mindanao provinces and Laguna to positive values in all the remaining survey provinces in Visayas and Luzon.
e) There is very high correlation (ranging from .90 to .99) between the interview area and the measured area in all provinces.

Table 7. Relationship between Interview (I) and Measured (M) areas by province

Province	*Number of Farms*	*I*	*M*	$\frac{I-M}{M} \times 100$	*Average Farm Area*	*Correlation*
Pangasinan	203	276.65	275.35	0.5 NS	1.36	0.95
Isabela	105	143.23	136.05	5.3 **	1.30	0.97
Nueva Ecija	27	50.42	49.02	2.9 NS	1.82	0.98
Laguna	38	40.29	45.48	-11.4 **	1.20	0.95
Camarines Sur	24	13.82	13.21	4.6 *	0.55	0.96
Ilocos	39	55.26	48.92	13.0 **	1.25	0.97
Bohol	61	43.00	36.32	18.4 **	0.60	0.99
Leyte	29	22.36	19.22	16.3 **	0.66	0.96
Zamboanga del Sur	52	56.35	60.93	-7.5 *	1.17	0.90
Misamis Occidental	14	8.64	8.87	-2.6 NS	0.63	0.98
North Cotabato	45	50.23	50.67	-0.9 NS	1.13	0.93
TOTAL	637	760.25	744.04	2.2	1.17	

NS – Not Significant at $\alpha = .05$
* – Significant at $\alpha = .05$
** – Significant at $\alpha = .01$

Table 8. Relationship between Interview (I) and Measured (M) area by tenure

Tenure	*Number of Farms*	*I*	*M*	$\frac{I-M}{M} x\ 100$	*Corre-lation*
Owned	205	265.83	256.11	3.8 *	0.98
Partly Owned	65	82.41	79.97	3.1 NS	0.95
Tenanted	247	242.22	239.99	0.9 NS	0.92
Leased	95	134.55	134.35	0.2 NS	0.95
Rented	25	35.24	33.62	4.82 NA	0.94
TOTAL	637	760.25	744.04	2.2	

* – Significant at $\alpha = .05$
NS – Not Significant at $\alpha = .05$

Table 9. Relationship between Interview (I) and Measured (M) areas by size of farm

Size	*Number of Farms*	*I*	*M*	$\frac{I-M}{M} x\ 100$	*Corre-lation*
Less than 0.50	169	60.41	50.08	20.6 **	0.73
0.50 – 1.00	199	159.44	146.44	8.9 **	0.59
1.01 – 1.50	96	122.18	118.34	3.2 NS	0.32
1.51 – 2.00	68	114.49	121.78	-6.0 *	0.26
2.01 – 2.50	45	102.21	98.67	3.6 NS	0.44
Over 2.50	60	201.52	208.73	-3.4 *	0.93
TOTAL	637	760.25	744.04	2.2	

** – Significant at $\alpha = .01$
* – Significant at $\alpha = .05$

Such relationship indicates considerable reduction of non-sampling errors of rice area estimates in each of the survey provinces when the measured areas of farms are used as concomittant variables in regression type estimators.

f) Only the *owned farms* show presence of significant errors. Although *rented* farms show higher relative error than the *owned,* the error of the former was found insignificant mainly due to extreme over-statement of the relatively small-size farms under this tenure classification.

g) The extreme size classifications contain larger relative errors. The size groups covering less than 1/2 hectare and 1/2 to 1 hectare contain positive errors (overstatement of area), as high as 20.6 percent and 8.9 percent, respectively, with both being significant at $\alpha = .01$. The group of farms consisting of 1-1/2 to 2 hectares, and the over 2-1/2 hectares registered negative errors (understatement of area) significant at $\alpha = .05$.

The relatively high overstatement of interview areas among the small size farms is traced mainly to the tendency of farmers to report area to the nearest quarter, third, half, etc., of a hectare. This observation was also noted in the study of 182 farmers in Manaoag, Pangasinan[5] where the degree of over-reporting of interview area was found to be lower in large farms. Only the extreme size categories of farms indicate high correlations while the middle classifications are characterized by the opposite values.

FUTURE BAECON R & D PLANS AND PROGRAMS

a. Expansion of Objective Area Surveys

One of the priority projects being planned at BAEcon is the development of a system that would make use of double sampling with regression estimate as developed by Neyman[6]. The theory says that if X and Y are highly correlated, and $\bar{y}$ is a sample mean estimated from a sample of size n, and $\bar{x}$ and $\bar{y}$ are likewise means

[5]International Rice Research Institute 1975. Annual Report Los Banos, Laguna, pp. 415-416.

[6]J. Neyman, "Contributions to the Sampling of Human Population", J. Anes. Stat. Assn., 33(1938) pp. 101-106.

estimated from sample size m which is a subsample of the original sample, then the estimate of $\overline{X}$ which is more reliable than $\overline{x}'$ is:

$$\overline{X}'' = \overline{X} + b\,(\overline{y} - \overline{y}')$$

where b is the estimate of the regression coefficient of X on Y from the m sub-sample. The variance of $\overline{x}''$ is given approximately by:

$$\text{Var}\,(\overline{x}') = \frac{S^2x}{m} \quad (1 - P^2\ (1 - \frac{m}{n}))$$

where S^2/m is the approximate variance of X, and P is the coefficient of correlation. The condition under which double sampling with regression estimate is worth using is:

$$P = \frac{2/C_1\, C_2}{(C_1 \ + C_2)}$$

where C_1 and C_2 are the unit cost of obtaining information about the X'm and Y's, respectively. Based on BAEcon cost estimates for obtaining data about X (say cost of measuring the area of a farm) and Y (corresponding cost of interview of area), the above equation would yield P equal to 0.60, very much less than those obtained in eleven provinces covered in the 1978 area measurement survey. Thus a subsample of 20 percent of the original sample and a correlation coefficient of about 0.90 would make double samples 1.39 more efficient than estimating X directly from a sample of size n. Hence considerable gain in accuracy may be achieved by the use of double sampling as described above.

Use of Computed Stratification Variables in Design of Surveys

It has been shown that the degree of correlation of the BSS and HSS data is sufficiently high for individual crops. The plan is to make use of such relationship to improve the design of ongoing BAEcon surveys employing the BSS data as a frame. To do this, regression models of the BSS and HSS data shall be derived and these models will be used in generating the stratification variable for a particular crop survey. A potentially useful model could be one derived by regressing the HSS crop area on the corresponding BSS area and the number of farm households.

Currently being pursued is an empirical investigation using the data from the Special Surveys on coffee and garlic to assess if there is any gain in the efficiency of a sample design by applying the above-mentioned principle. The result of this study could be useful in a wider application to all types of BAEcon surveys.

Construction of Area Frame

Surveys using an area frame are generally considered more efficient for crop and livestock statistics than surveys using a list frame. An area frame is less subject to change; hence, the burden of frequent updating is minimized.

One reason why area sampling has not been developed and employed in the Philippines is due to the lack of materials such as, uptodate large scale topographic maps showing permanent natural and artificial features that can be used as boundaries of area sampling units. Another major constraint is the difficulty to demarcate sufficiently small and unambigously defined sampling units. For example, a sample segment of a size of about one square kilometer of agricultural area would likely embrace a good number of small farms anywhere between 50 to 150. This poses serious operational problems.

The rapid development of remote sensing technology and photo interpretation offers an opportunity for BAEcon to develop an area frame for surveys on food and agriculture. Landsat frames each covering an area of 185 km x 185 km could now be readily procured from NASA. Land use classification could feasibly be extracted from the landsat frame by computer and this could be used to update existing old maps.

The technology needed to do this job shall be developed under a BAEcon-based USAID/USDA-assisted project which is expected to be implemented before the end of this year in Central Luzon as a pilot area. The ultimate goal of this project is to enable the construction of an area frame for the whole country within a period of 2 to 3 years, and to develop a system of extracting agricultural information from landsat imagery taken on the same at an interval of 18 days.

Expansion of Crop Cutting Surveys

There appears to be evidence that production reported in an interview are biased downward. A study of the rice production of some

68 farmers in Laguna[7] indicated that without deeper probing the farmers do not report about 4.9 percent of total production.

If this finding in Laguna is also generally true in other provinces, the reported total rice production of the country could be significantly understated. And if this finding is combined with the upward biased reporting of harvest area in some provinces, this would make the computed yields even more biased.

Crop cutting surveys if properly planned and implemented could solve the problem. As mentioned earlier, BAEcon has implemented a crop cutting survey together with an area measurement survey starting in 1978. Many problems were met during the actual implementation of the project. The major problems encountered are those related to the lack of resources (financial, manpower and special crop cutting facilities).

Another specific problem is the operational difficulty arising from the demand of crop cutting samples to be watched constantly to ensure that crop cutting is done at the right time. Because of the lack of adequate manpower assigned to do the job, in several instances sample farms are already harvested before the survey team could visit these farms to be crop cut.

The root of the above problems could be traced to inadequate budgetary support; hence, BAEcon has requested more funds for this particular activity starting in budget year 1980.

The authors wish to state that BAEcon management is continually endeavoring through R & D efforts to improve the reliability of statistics on food and agriculture in this country. The term reliability as used here refer to:

1. reliability of measurements, which is the way the statisitcian uses the term;
2. reliability in the way the data system operates; and
3. conceptual reliability.

The authors look forward to the further improvement of the data series on food and agriculture towards and along these three directions.

---

[7] International Rice Research Institute 1967 Annual Report pp. 271-272.