PN-ABN-004
Rapport ASPAP/DAI N° 203e

A Methodological Comparison of the National Household Budget and Consumption Survey and the Division of Agricultural Statistics Agricultural Household Revenue and Expenditure Survey

Division of Agricultural Statistics
MINAGRI

ASPAP: Agricultural Surveys and Policy Analysis Project

(Projet des Enquêtes Agricoles et Analyse
des Politiques Economiques du Secteur Rural)

Contract USAID n° 696-0126

by

Dr. Scott Loveridge[1]
Short-term Consultant, DAI/UM[2]

David J. Megill[3]
Short-term Consultant on Sampling, DAI[4]

Samuel Munyaneza
Statistician, DSA

Kigali, Rwanda
July, 1992

[1]Assistant Professor, Department of Agricultural and Applied Economics, University of Minnesota, St. Paul, MN 55108.

[2]University of Minnesota, sub-contractor, and Development Alternatives, Incorporated (DAI), 7250 Woodmont Ave., Suite 200, Bethesda, MD 20001, principal contractor for ASPAP, USAID Contract N°696-0126-C-00-7777-00.

[3]1504 Kenwood Ave., Alexandria, VA 22305, USA

[4]Development Alternatives, Incorporated (DAI), 7250 Woodmont Ave., Suite 200, Bethesda, MD 20001, principal contractor for ASPAP, USAID Contrat N°696-0126-C-00-7777-00.

This document translated from the original French ASPA/DAI Rapport N° 203 by Leonidas Murembya and Jaakko Kangasniemi. The translation was financed and coordinated by Michigan State University.

Introduction

To date, there have been only two national surveys in Rwanda on the economic activities of rural households: the National Household Budget and Consumption Survey (ENBC, l'Enquête Nationale sur le Budget et la Consommation des Ménages), conducted in 1983, and the Survey on Revenues and Expenditures of Agricultural Households (RD, L'enquête revenus et dépenses des ménages agricoles), conducted in 1990 by the Division of Agricultural Statistics (DSA). The objective of this report is to present a brief summary of the methodological differences between the two surveys. Both surveys have strengths and weaknesses. The object of this report is not to criticize the methods used in either survey, but rather to facilitate comparisons between the results of the two surveys. It should be noted that as the first national survey of its kind in Rwanda, the ENBC results helped DSA to conceive its revenue and expenditure survey.

Objectives of the Surveys

The objectives of the two surveys were different. The ENBC, a project of the Ministry of Planning (MINIPLAN), was primarily concerned with household consumption, whereas the RD of DSA, a project of the Ministry of Agriculture (MINAGRI), was primarily concerned with agricultural production and its utilization. These objectives greatly influenced the methods of data collection and, in turn, the results of the two surveys.

Target Population

The ENBC sample covered all rural households of sub-saharan african origin. In contrast, the RD covered only rural *agricultural* households. The RD sample was the more limited of the two surveys because of its objective of measuring agricultural production and utilization. Since non-agricultural rural households in Rwanda tend to be better off than agricultural households, there is reason to suspect that the ENBC target population is slightly richer than that of the RD.

Sampling

The two surveys had similar sample size and used similar sampling methods. (See DSA, November 1991 for a summary of the RD sample design, and MINIPLAN, 1986 for information on the ENBC sample design.) The RD sample had 1,248 households, of which 40 were eliminated because of death and other problems (refusal to participate, relocation, etc.). The ENBC sample had 1,170 households, out of which 270 households were selected for intensive data collection. During the analysis, ENBC analysts decided

to eliminate the information on the 900 households not subjected to intensive data collection. Thus the effective sample size of the ENBC was 270 households. We did a simulation to estimate how much the standard errors of household mean revenues found by the two surveys would be influenced by sample size assuming identical data collection methods (see the Appendix for the explanations of the method used in the simulation). The simulation showed that the bigger RD sample should result in a standard deviation 35% of that of the ENBC. The reported standard deviation for the mean revenue for the ENBC was 5,733 (MINIPLAN 1986, p. 119). Assuming an annual inflation rate of 4% in rural areas between 1983 and 1990, the MINIPLAN estimate for the standard deviation increases to 7,544. The RD standard deviation was computed to be 1,939, meaning that the RD estimate is 26% of the ENBC's standard deviation. Thus the ENBC has a higher variance than what would be expected based on the smaller sample size only. We suggest explanations for this below. Despite the large standard deviation of the ENBC, we found that the estimates on mean revenues were statistically different at the 95% level of significance.

Units of Measurement

The absence of commonly used standard units is a problem survey researchers face in Rwanda and elsewhere. The solutions employed by the two Rwandan surveys to overcome the unit of measurement problem were different. The RD survey, with its interest in agricultural production, relied on standard buckets given to each household in the sample. This innovative solution was not possible for the ENBC due to its focus on the very small quantities of food prepared daily by the household. The ENBC solved the units problem by measuring the volume of each container used by the household. This necessarily results in some measurement errors, but for measuring very small quantities, it may be the best method available.

Production and Consumption

The ENBC survey estimated consumption by asking respondents to report on meals prepared. The enumerator visited the household once a day during seven days to record what was prepared. This seven-day excercise was repeated four times a year, yielding a total of 28 observed days per household per year. Purchases and sales were recorded for a period of 14 days. This was also repeated four times a year, yielding a total of 56 observed days per household per year. The ENBC collected additional information on purchases and sales with a questionnaire on transactions above 200 FRW during the preceeding three months. The agricultural production of the household was obtained by taking the difference between consumption and purchases. The ENBC's daily visits to record consumption and transactions were not well planned. Each enumerator was required to do a certain number of interviews within a three-month period, but no monthly requirements existed. Consequently the number of observations

was roughly equal between each three-month period, but most of the observations were obtained during the last month of each period. Since the analysis was done without reference to months, analysts effectively weighted some months of the year more heavily than others. Table 1 shows the number of households visited and household-days observed per month in the ENBC. Given the seasonality of agricultural production, this introduces a non-negligible bias in the production estimates. Moreover, consumption and transactions vary by season and these variations cannot be captured within the seven- or fourteen-day observation periods. Thus, this increases the estimated variance of household revenue in the ENBC.

Table 1
NUMBER OF HOUSEHOLDS AND HOUSEHOLDS INTERVIEWED BY MONTH
ENBC (Rural Survey)

	1982 Nov.	Dec.	1983 Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Total
HOUSEHOLDS	130	178	77	100	170	99	85	164	123	74	168	161	34	270
HOUSEHOLD-DAYS	1383	1708	516	1110	1610	1043	810	1556	1365	409	1593	1526	173	14807
% OF HH-DAYS	9.3	11.5	3.4	7.5	10.9	7.0	5.5	10.5	9.2	2.8	10.8	10.3	1.2	100.0

Source: MINIPLAN (1986, p. 25)

The RD survey used a direct method to measure production. Enumerators visited each household once a week during 52 weeks to record the harvest of the sixteen principal crops. Purchases and sales were recorded once a month, or a total of 12 times. The RD did not try to measure consumption directly, but estimated the availability of crops for each household by comparing production, acquisitions and transfers. The RD ignores own-consumption of minor crops and animal products when estimating total household revenue. ENBC estimated these to be 5 percent of the value of total food consumption.

A strong point of the ENBC was that it sought to measure expenditures on all goods used by the household, whereas the RD concentrated only on agricultural products and inputs. Thus, the ENBC is the only source of information on expenditures on clothes, school fees, taxes, etc.

Net Versus Gross Sales

It is well known that some households buy and sell same products during the agricultural year even though they are not trading. Loveridge (1989) found that 26% of agricultural households bought and sold sorghum, whereas 37% of agricultural households bought and sold beans. According to the standard explanation, these households sell after the harvest because they need money, and buy back later when they

need food or seeds. This may well be true, but Loveridge (1989, pages 50-51) suggests additional reasons why some households may buy and sell the same product during the year (in this case sorghum, but similar reasoning may apply to other crops, too):

1) A poor agricultural season followed by a good season (A family buys sorghum to make up shortages and then produces a surplus that it sells).

2) The family sells old low-quality sorghum and replaces it by buying higher-quality sorghum for consumption.

3) The family sells its harvest and buys higher-quality seeds for the next season.

4) The family miscalculated its sorghum needs for the year. Large quantities of sorghum are bought for the ceremonies related to unexpected births, marriages, etc.

5) The family harvests and sells sorghum when prices are high, and buys it when the main harvest pushes prices down.

6) The family does not want to risk losing its harvest in storage (storage and security problems).

In fact, one can consider agricultural markets as a banking system for the producer. When the producer has sufficiently high production, he or she can deposit the products in the market just as workers can deposit their salaries in a bank. When the producer has a need for agricultural products, he or she can get them from the market just like workers can withdraw money from their bank account. But just as loans and withdrawals from bank accounts are not considered to be part of workers' salary, it is equally wrong to think that each transaction by producers is part of their total revenue. What is important is the net position in transactions. If a producer has bought 2 kilos of beans and sold 10 kilos of beans, his net sales are 8 kilos, or the difference between purchases and sales. Since the RD concentrated on agricultural production, it considered farmers' net sales. In contrast, the ENBC considered only gross sales when computing household revenues (10 kilos of beans in our example). For this reason, revenues from the sales of agricultural products were larger in the ENBC than they were in the RD.

Valuation of Own Consumption

To estimate total household revenues, one must convert the unmarketed part of production into monetary values. This can be done by using the prices observed in commercial transactions, but there is more than one price to choose from. The price received by producers is seldom the same as the price paid by the consumer, since consumers also need to pay for the storage and transportation services provided by

merchants. Given the ENBC's consumption orientation, its analysts chose to use consumer prices to value own consumption. In contrast, the RD's production orientation drove its analysts chose to value own consumption by using producer prices. Both methods are easily defensible, but one must recognize that they lead to different results. Because own consumption is the most important source of revenue for agricultural households, this difference between the two methods considerably influences the level of estimated revenues.

Summary and Conclusions

The ENBC and the RD used very different methods for collecting and analyzing data on the revenues of agricultural households. The ENBC was oriented toward consumption, whereas the RD was oriented toward production. This difference in orientation explains many of the differences between the two methods. Table 2 presents a summary of the differences between the two surveys. The experience from the ENBC helped the RD analysts to design their survey. Both surveys were well designed, conducted, and analyzed to serve the objectives of the ministries involved. However, the differences in method are so large that direct comparisons between the two surveys to analyze trends in the rural economy of the country are not recommended.

Table 2
SUMMARY OF METHODOLOGICAL DIFFERENCES
ENBC-RD

	ENBC	RD	PROBABLE EFFECT OF THE DIFFERENCE ON THE ESTIMATE OF MEAN REVENUE
YEAR COLLECTED	1983	agricultural year 1990	unknown
POPULATION	rural households	agricultural households	ENBC higher
SAMPLE	270 households	1208 households	RD more precise
NUMBER OF HHs SURVEYED MONTHLY	variable--74 to 178	stable--1208	RD more precise
UNIT OF MEASUREMENT	containers used by households	standardized bucket	RD more precise
PRODUCTION	imputed	direct measurement, 52 weekly visits	RD more precise
TRANSACTIONS	14 daily visits during a short period four times a year and four quarterly visits	12 monthly visits	unknown
CONSUMPTION	direct measurement, 7 daily visits for a week, four times a year	imputed	unknown
NUMBER OF PRODUCTS COVERED IN ESTIMATING OWN CONSUMPTION	all	16 principal crops	ENBC 5% higher and more precise
SALES	gross	net	ENBC higher
VALUE OF OWN CONSUMPTION	consumer price	producer price	ENBC higher

Bibliography

Division des Statistiques Agricoles, "Enquête Nationale Agricole 1989: Production, Superficie, Rendement, Elevage et Leur Evolution 1984-1989," Publication No. 22, Ministére de l'Agriculture, de l'Elevage et des Forêts, November 1991.

Loveridge, Scott, "Importance du Haricot et du Sorgho dans le Système de Production des Cultures Vivrières au Rwanda: Disponibilités Actuelles et Projections pour l'Avenir," Division des Statistiques Agricoles, Ministère de l'Agriculteur, de l'Elevage et des Forêts, République Rwandaise, 1989.

Ministère du Plan, "Enquête Nationale sur le Budget et la Consommation des Ménages, Milieu Rural (Nov. 1982 - Déc. 1983) Volume 2: Structure du Budget des Ménages," République Rwandaise, November 1986.

Appendix

Comparing the RD and ENBC Estimates of Mean Household Revenue

The main text of this document compared the methods of measuring household revenue used by the two surveys: the RD (conducted in 1990) and the ENBC (conducted in 1983). In this appendix, we ask whether the difference between the two estimates of household mean revenue is statistically significant given the sizes of the two surveys, under the assumption that the data was collected with the same method. The difference between household mean revenues can be expressed as follows:

$$\bar{x}_1 - \bar{x}_2,$$

where:

$\bar{x}_1$ = estimated mean revenue per household 1983 (ENBC)

$\bar{x}_2$ = estimated mean revenue per household 1990 (RD)

In general, the variance of the difference between two estimates from different samples can be expressed as:

$$Var(\bar{x}_1-\bar{x}_2) = Var(\bar{x}_1) + Var(\bar{x}_2) - 2 \times Cov(\bar{x}_1, \bar{x}_2),$$

where:

$Var(\bar{x}_1)$ = *variance of* $\bar{x}_1$

$Var(\bar{x}_2)$ = *variance of* $\bar{x}_2$

$Cov(\bar{x}_1, \bar{x}_2)$ = *covariance between* $\bar{x}_1$ *and* $\bar{x}_2$

Since the two surveys are independent, the covariance between the two estimates should equal zero. The variance of the mean revenue difference is therefore simply the sum of the two variances.

When computing the variance of each estimate, we must take into account different aspects of the sample design. The computer software PC CARP was used to compute the variance for the RD data, and takes into consideration the stratification and the size of the groups included in the sampling method. It also computes the design effect (DEFF), which is defined as the variance of the estimate from the actual survey divided by the estimate from the same survey with simple random sampling. The DEFF is a measure of the relative efficiency of the sampling method. The household mean revenue estimate by the RD survey in 1990, as well as the corresponding variance and DEFF of this estimation, calculated with PC CARP are as follows:

$\bar{x}_2$ = 47,199 FRW

$Var(\bar{x}_2)$ = 3,759,760

$DEFF(\bar{x}_2)$ = 1.8627

The standard deviation of $\bar{x}_2$ is equal to the square root of its variance, or 1,939 and the corresponding standard error (standard deviation divided by the estimate) is 4.1 percent.

We estimated the variance of the estimated mean revenue per household for the ENBC by creating a sub-sample of the 1990 RD survey that simulates the ENBC sampling method. The ENBC used a national sample of 90 groups with three households in each. Since the 1990 RD survey consisted of 77 DR ("Districts de Recensement", census districts) with four subgroups of four households in each, we randomly selected one subgroup from each DR and randomly eliminated one of the four households. Some of the selected subgroups only had three households because the fourth one had dissolved. In these cases, the entire subgroup was retained for the analysis. Thus, the subsample used to estimate the ENBC variance consisted of 77 groups of three households, one in each census district of the 1990 RD survey. The revenue data for these households was used to estimate the design effect (DEFF) for the ENBC sampling method. The DEFF calculated using PC CARP of our subsample is 1.2362. Under the assumption of similar stratification and group selection, the DEFF based on 77 census districts should be the same as that based on 90 districts.

The variance for a household mean revenue estimate from any sampling method can be expressed as:

$$Var_{PS}(\bar{x}) = \left(\frac{\sigma_x^2}{n_{PS}}\right) \times DEFF_{PS}$$

where:

$Var_{PS}(x)$ = variance of mean revenue from a given sampling method (PS)

σ_{x2} = variance of the population for household revenue

n_{PS} = sample size in a given sampling method

$DEFF_{PS}$ = Design effects of a given sampling method on the estimate for household mean revenue

The ratio of the variances of the estimates of mean revenue per household by ENBC and by 1990 RD can be expressed as:

$$\frac{Var_{ENBC}(\bar{x})}{Var_{RD}(\bar{x})} = \frac{\left(\frac{\sigma_x^2}{n_{ENBC}}\right) \times DEFF_{ENBC}}{\left(\frac{\sigma_x^2}{n_{RD}}\right) \times DEFF_{RD}} = \left(\frac{n_{RD}}{n_{ENBC}}\right) \times \left(\frac{DEFF_{ENBC}}{DEFF_{RD}}\right) =$$

$$= \left(\frac{1208}{231}\right) \times \left(\frac{1,2362}{1,8627}\right) = 3.47,$$

where the subscripts ENBC and RD indicate the two surveys.

The variance of the estimate of the mean revenue per household with the ENBC sampling method can be estimated by multiplying the corresponding variance of the 1990 RD survey by this ratio:

$$Var_{ENBC}(\bar{x}) = 3.47 \times Var_{ENRD}(\bar{x}) = 3.47 \times 3,759,760 = 13,046,366$$

This corresponds to a standard error of 7.7 percent with reference to the mean revenue per household of the 1990 RD survey. Applying this standard error to the (inflation-

adjusted) ENBC estimate of the mean revenue per household of 71,534 yields a standard deviation of 5,508 and a corresponding variance of 30.339,364. This estimate of the variance is more appropriate because it is adjusted to the values compared.

Based on the above definitions, the estimated difference between the ENBC and 1990 RD estimates of the mean revenue per household is:

$$\bar{x}_1 - \bar{x}_2 = 71,534 - 47,199 = 24,335$$

The variance of this estimated difference can be calculated using the formulas specified above in this appendix:

$$Var(\bar{x}_1 - \bar{x}_2) = Var(\bar{x}_1) + Var(\bar{x}_2) =$$

$$= 30,339,364 + 3,759,760 = 34,099,124$$

The standard deviation for the estimated difference is the square root of the variance, or 5,839. The 95% confidence interval equals:

$$24,335 \pm 1.96 \times 5,839$$

The lower and upper bounds for the confidence interval are 12,891 FRw and 35,779 FRW, respectively. Since zero is not included in the interval, we can conclude that the two estimates of the household mean revenues are statistically different (based only on sampling errors). However, the bias caused by the differences in data collection methods of the two surveys is probably larger than the sampling error.