

ASPAP/Miniplan Trip Report

9 January to 20 February 1989

Agricultural Surveys and Policy Analysis Project (ASPAP)

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

USAID Contract No. 696-0126

Nicholas Minot*
Short term consultant, DAI/MSU**

Kigali, Rwanda
February 1989

* Visiting Specialist, Department of Agricultural Economics,
Michigan State University, East Lansing, Michigan, 48824 USA

** Michigan State University, subcontractor, and Development
Alternatives, Incorporated, 623 Ninth St. NW, Washington, D.C.
20001, principal contractor for ASFAP, USAID Contract No.
696-0126-C-00-7777-00.

The terms of reference for the trip were defined as follows:

1. Work closely with Dr. Jim Ansoanuur and ASPAP staff to review data processing and analysis carried out for the rural data of National Household Budget and Consumption Survey (ENBC), plan processing and analysis activities for the urban data of the ENBC, and identify the most policy-relevant types of analysis for the urban data.
2. Provide recommendations for computer hardware and software purchases for Miniplan.
3. Provide recommendations for short-term training of Miniplan personnel.
4. Provide additional analyses of the rural budget data.

This report is divided into six sections: a summary of activities during the assignment, recommended priorities for urban ENBC data processing, a work plan for the urban ENBC data processing, recommendations for hardware and software needs, recommendations for training, and recommendations for future assistance.

1. SUMMARY OF ACTIVITIES

1.1 Activities at Michigan State University, 9-13 January

The first week of this assignment was spent at Michigan State University working on two tasks: 1) comparative testing of two software packages (FOCUS and SPSS) and 2) carrying out additional analysis of the rural ENBC data. For the comparative testing, we started out by creating equivalent 3MB data files in FOCUS and SPSS. Then, five similar operations typical of those required for processing budget survey data were performed with each. The two packages were compared for processing speed and ease of programming. The results are summarized in section 4 and presented in detail in Annex G.

With regard to the additional analyses of the rural ENBC data, some 80 runs were done, concentrating on the characteristics of households relying on different types of income. The results highlight the overwhelming importance of food production for own-consumption: for half the households, this represents the majority of net income. Less than 5% of the households depend to this degree any one of the following categories: cash crop sales, food crop sales, beer brewing, or other artisanal activity. For the rest of the households, no single activity accounts for the majority of net income. For coffee producers, the value of final consumption is higher than average, but caloric intake is somewhat lower. As could be expected, households with important non-agricultural income tend to have smaller farms, although the direction of causality is not clear.

1.2 Activities in Rwanda, 17 Jan - 20 Feb

During my four weeks in Rwanda, most of my time was devoted to working with Dr. Jim Ansecauur, training him in the use of FOCUS, the structure of the urban data files, and the different stages in processing the data. During the last ten days, Munyaneza Samuel participated full-time in these activities. With regard to FOCUS training, before I arrived Jim had made a careful review of the FOCUS user manual, so we moved almost immediately from simple exercises to actual processing tasks. These tasks include producing verification listings, checking data in the questionnaires, and generating output tables, as well as creating and modifying FOCUS files. He now has a sense of the programming requirements for a given task, although identifying the tasks which need to be done will take more time and experience.

In terms of data processing activities, we have concentrated on the systems for classifying transactions, for weighting the results, and for valuing transactions in kind, as well as covering the procedures for cleaning the data and creating the summary file. Jim Otto had started the **classification** system, but around 7000 transactions remained to be categorized. All but 270 transactions (0.25% of the total) are now categorized, although further work will be needed to correct and finetune the classification. The spatial **weighting factors** for the ENBC/Urban data and temporal weighting factors for BQ and BR have been calculated and installed in the main data file. A program to calculate the temporal weighting factors for CA (consommation alimentaire) has been prepared as well. With regard to the **valuation** of transactions in kind, a price file has been created, based on cash purchases in BQ. Using this price file, 98% of the transactions in kind in BQ have been valued (the rest will be valued "manually"). We have also reviewed the procedures for **cleaning**, although the sophisticated data entry and verification programs written by Jim Otto have allowed much of this work to be shifted to Ministry personnel. And finally, a preliminary **summary file** of cash and in kind transactions has been created, although the final version must await the finalization of the classification system and the integration of budget files. It should be noted that this schedule is highly accelerated relative to the time required to do similar tasks with the rural data: the weighting system alone took a month to create in 1987.

I have prepared several documents are available for consultation after my departure. In July 1988, I prepared a 20-page guide to processing the urban budget data, covering data cleaning, classification, weighting, integration, and the creation of a summary file. During this trip, I prepared a detailed description of the data processing carried out during this assignment along with a list of programs used (Annex A). In addition, I documented the principal files and programs from the rural data to serve as a reference for similar files and programs for the urban data. Included in this documentation were a list of rural data files (Annex B), a description of the variables in these files (Annex C), an annotated list of the main programs used for weighting, valuing, and integrating the rural data (Annex D), and an annotated list of the programs used to create the 80-odd tables in Volume 3 of the rural ENBC results (see Annex E). Many of the

programs and file structures can be used with little modification for the urban data set. And finally, I compiled a six-page quick reference guide to the principal commands and procedures (see Annex F).

I have also been involved in a number of other activities while in Rwanda. First, at the request of the Director of Surveys, I provided initial orientation to Dr. Randall Schnefp, a UNICEF consultant who just arrived for a 6-12 month tour to analyse the ENBC anthropometric data. Second, I assisted Nsengimana Elie with several bugs in the data entry program for the food consumption questionnaire. The program should be ready by the end of February. Third, upon the request of the Director General of Statistics, I reviewed the files left by Christophe Muller to determine if they were useable. Unfortunately, the files are useless as a result of the complete lack of documentation, the large number of missing files, and the rudimentary state of what remained (see Annex I).

2. PRIORITIES FOR URBAN DATA ANALYSIS

The ENBC contains a wide range of information concerning the urban and rural economy: expenditures, income, home production, food consumption and nutritional indicators, demographic composition of households, daily activities, and household assets. The core of the rural budget and food consumption results have been published (Volumes 2, 3, and 4), as have some partial demographic results from the urban areas (Volume U2). Because resources are limited, results will be slow coming, and priorities must be set. Although such decisions are somewhat subjective, the following reflects my opinion of the demand for results by government agencies and development organizations.

1. The most important results yet to be published are those of the urban budgets based on the 300 "daily" households and the questionnaires BQ, BR, and CA. The demand for the urban budget data (particularly the composition of expenditure) is clearly the greatest. The Bureau de Prix of Miniplan is already using preliminary figures for constructing new price indices. The Banque Nationale Rwandaise has repeatedly asked when urban expenditure data would be available. And several development organizations have expressed an interest in the expenditure data for marketing studies.
2. Next in importance are the urban food consumption results based on the questionnaire CA and the anthropometric data in both urban and rural areas. This topic is of particular importance to UNICEF and other agencies involved in nutrition and social well-being.
3. The analysis of household assets in rural and urban areas (questionnaires Q6 and PA) follow in importance. Although there is little explicit demand for this data, it would provide a useful complement to the budget data and might provide indices for easy identification of low-income households.

Date base	Situation as of Feb. 10, 1989	February - May 1989	June - August 1989	Sept. - December 1989	January - June 1990	June - December 1990
Rural budget (Q2, Q4, Q5)	Vol 3 printed and distributed, Vol 4 printed but not distributed	Distribution of Volume 4	Presentation of rural results at ENBC seminar			Possible report comparing rural and urban results
Urban daily budget (BQ)	Data entered, basic verification done, weighting system done, classification and valuation mostly done	Complete classification system and valuation of transactions in kind	Produce summary file from BQ and BR, provisional tables excluding autoconsumption, presentation at ENBC conference	Valuation of CA, incorporation into summary file, run 80 tables based on programs for Vol 3 (rural data), possible cleaning and fine-tuning of classification system	Econometric analysis, commentary for budget report, analysis of special topics	
Urban retrospective (BR)	Data entered for 300 hh, basic verification in progress	Classification system adapted to new transactions, valuation				
Urban food consumption (CA)	Data entry program written, being tested	Data entry started and completed	Classification, creation of conversion coeff. file		Preparation of tables for food consumption and nutrition report	Commentary for report, review, publication
Rural anthropometric (Q5)	Data entered, UNICEF consultant started processing	Tables produced and report started	Preparation of publication, review, printing			
Urban anthropometric (CA)	(see urban food consumption)	Data entry started and completed	Preliminary tables, start report if contract extended	Preparation of publication if contract extended		
Rural household assets (Q6)	Data entered, no processing	No activity	No activity	No activity	No activity	No activity
Urban household assets (PA)	Outline of file structure and data entry program defined	Jim Otto writes data entry and verification programs	Data entry, no processing	No activity	No activity	No activity
Urban struct & activities (SA)	Data entered but no processing, partial results published as U2	No activity	No activity (unless Simonet returns)	No activity	No activity	No activity

32

4. The demographic analysis in the rural and urban areas is useful to the extent that it allows us to relate demographic variables with economic, agricultural, and nutritional ones. However, as a demographic data base, it is neither the largest, nor the most recent available in Rwanda.

With regard to the types of analyses to be done with the urban budget and food consumption data, the first step would be to repeat the analysis done for the rural areas, including the composition of food and non-food expenditures, the relative importance of different sources of income, the influence of household characteristics on household budgets, and econometric analysis of demand. As mentioned, the programs to produce the 80-odd tables of Volume 3 covering these topics can be used with little modification on the urban budget data.

As for more specific types of analysis, a few might be mentioned: 1) A comparison of standards of living in the urban and rural areas; 2) an analysis of economic, social, and demographic factors related to low caloric intake and poor nutritional status; 3) a study of the importance of imported goods in the budgets as a function of income and urban/rural residence, useful in estimating the impact of exchange rate adjustments; 4) a comparison of the level of official imports and that of consumption for selected goods, providing hints as to the size and nature of "informal" trade; and 5) a study using estimated income elasticities to predict changes in the composition of urban, rural, and total demand associated with alternate assumptions about economic growth and urbanization.

3. WORK PLAN FOR URBAN ENBC DATA PROCESSING

The activities necessary in cleaning, processing, and analyzing the urban data can be identified without much difficulty. And the order of activities naturally follows the discussion of priorities. However, establishing a calendar for these activities is much more risky without knowing the amount of human resources the Ministry will allocate to the effort. For reasons discussed in section 4, I assume, for the purpose of this work plan, that no Ministry personnel will participate in the data processing and analysis and that Drs. Jim Ansoanur and Randall Schnepf will be the only ones working on these tasks, along with short-term assistance through the ASPAP project.

The daily budget data (questionnaire BQ) and the retrospective budget data (questionnaire BR) have been entered for the 300 "daily" households. Basic verification and cleaning, based on programs written by Jim Otto and run by Nsengimana Elie, have been completed, and similar cleaning is underway for BR. Some of the BR questionnaires for the other 900 "retrospective" households have been entered, although this was discontinued on my recommendation because the data are probably useless (similar data for the rural area were never used because the information for each household is so incomplete). The classification system for BR and BQ is essentially complete, although some corrections remain and the categories may need to be modified in order to adjust for differences between urban and rural budget patterns. The system for both spatial and temporal weighting has been

developed and installed. Almost all the transactions in kind in BQ have been valued. A preliminary summary file with transactions in cash and in kind has been created. Elie is writing the data entry program for the food consumption questionnaire (CA) based on detailed specifications prepared by Jim Otto. Data entry will begin as soon as the program is finished, probably in two weeks or so.

4. RECOMMENDATIONS FOR COMPUTER HARDWARE AND SOFTWARE

The recommendations for computer related acquisitions will be brief since Jim Otto is more qualified to provide advice in this area. However, several comments are in order.

Regarding hardware, there are currently 12 IBM-compatible computers at the Centre de Calcul of the Ministère du Plan (see Annex H for a list). Most, though not all, are at the disposition of the Direction Generale de la Statistique. The utilization rate seems quite high: at any one time, all or almost all the computers are in use (several machines are used for an evening shift of data entry). Another problem is that the urban budget data set is much larger than the rural data set. For example, the files containing the daily urban budget data occupy 10 Mb, five times the space for the corresponding rural data. This means that some runs take over ten hours. Furthermore, such large files require more working disk space, and we have already tried programs that exhausted the 8 Mb of available disk space.

The problem of disk space can be dealt with relatively inexpensively by purchasing 40-Mb "hard disk" card for the IBM AT being used for ENBC data processing. An even greater contribution to processing capability would be the purchase of a micro-computer with a 386 chip, which would run several times faster than the 286-based IBM AT now used. Other than the 386-based computer, any other computer purchases should be directed toward the lower performance machines, such as the IBM PS/30. This is because several IBM ATs are being used mainly for word processing. By purchasing relatively inexpensive machines to be allocated to word processing, IBM ATs would be freed up to meet any increased processing demand.

Regarding software, one issue concerns the word processing needs. In spite of a training session in Word Perfect, Wordstar (version 3.40) continues to be used exclusively. I worked with Nsengimana Elie and Nkurunziza J.M.V. to configure Wordstar for the laser printer. Although Wordstar is better at taking advantage of the laser printer's capabilities, Wordstar does work and no new software is needed for the time being.

Another important issue concerns the choice between FOCUS and SPSS. As mentioned above, I worked with MSU staff to run some comparative tests using SPSS and FOCUS. The following tentative conclusions were reached:

1. FOCUS is between two and five times faster than SPSS in producing reports and summary files. This advantage is even greater when using variables from different files.

2. FOCUS files are considerably smaller than equivalent uncompressed SPSS files, but only somewhat smaller than compressed SPSS files.
3. File creation and management tasks are probably more difficult with FOCUS than SPSS due to the former's hierarchical way of organizing data.
4. Producing simple, unformatted cross-tabulation from variables in the same file requires fewer commands in SPSS, although FOCUS is probably easier to use for more complicated procedures, e.g. when formatting is required, when several steps are required, or when variables from different files are used. The syntax in FOCUS seems more natural and more flexible, and fewer intermediate steps must be specified.

It is true that few people in the Ministry have become proficient in FOCUS in spite of various training efforts (see section 5), but this does not necessarily indicate that the software is difficult to learn. First, Miniplan has only a few people with the educational background conducive to learning data base management software. Second, few Ministry personnel are involved in data processing so that the trainees do not have a chance to use what they have learned. For this same reason, the SPSS training at Miniplan has not resulted in improved capacity.

I would recommend continuing to process the ENBC data in FOCUS for several reasons. First, the ENBC is a complex data base requiring the kind of processing for which FOCUS is more powerful. It involves data at various levels and the frequent use of variables from different files. Second, the ENBC files are relatively large, particularly in the urban data set where one of the six questionnaires yields a 10 Mb file. With such files, the speed of FOCUS is a major advantage. Third, a large number of programs have been written for the rural processing and analysis which are applicable in the urban processing, such as the programs to produce the 80-odd tables for Volume 3.

5. RECOMMENDATIONS FOR SHORT-TERM TRAINING

In order to evaluate training needs, it is useful to first consider the existing human resources. Currently, the Direction des Enquêtes has around 40 employees, including a large number of enumerators (roughly 25) and data-entry and coding people (about 10). On the other hand, there are only two at the A2 (4-year post-secondary) level and one at the A1 (3-year post-secondary) level.

The Direction is capable enough in the areas of data collection, coding, and data entry, with a large team of agents with experience on a number of major surveys. With regard to computer programming, the Direction has relied on the former Direction Générale de l'Informatique. Mungalibanje Jean Bosco is experienced with FOCUS and other packages, but is often busy with administrative tasks (the reorganization of the Ministry may leave him more available for programming). Nsengimana Elie is fairly strong in file management in FOCUS and has some experience in report generation. Samuel has some experience with report-generation in FOCUS.

The Direction is the weakest in analytical capacity. Samuel is the only one available for data processing and analysis, although being the only AO other than the Director, he is frequently drawn into other tasks such as survey design and formulation of work plans. For example, he was to participate in the ENBC analysis and received intensive training from me in June and July 1988. However, as soon as I left in August, he was diverted to work on the Enquete Nationale sur l'Emploi and yet another national survey (Enquete Nationale Adequation Formation-Emploi).

Although Ministry personnel have participated in the programming, survey analysis has been the almost exclusive domain of foreign technical assistance. Without counterparts for these consultants, valuable opportunities for training have been lost. This is true for all the major surveys over the last few years: the Enquête Démographique, Enquête Nationale des Entreprises, and the Enquête National Budget-Consommation. In addition, the data have often been underexploited or at least subject to long delays before publication.

In summary, there is a severe imbalance between solid data collection capacity and almost non-existent analytical capacity. In the long term, training should be oriented toward data processing and analytical capacity, but in the short term, **the lack of Ministry staff is a more constraining factor than training per se.** The Direction des Enquêtes should have at least 2-3 AOs on its staff working on survey processing and analysis. Foreign advisors should be assigned to work with Ministry staff, rather than being considered substitutes for Ministry staff. Even with this level of Ministry investment in processing and analysis, **it would probably be useful to reduce the volume of data collected,** both in terms of sample size and questionnaire size. For example, the 1983-85 ENBC involved over a dozen questionnaires, some of which may never be analyzed. Another example is the 1981 Enquete Demographique, in which only a third of the sample of 60,000 households could be entered into the computers and analyzed due to data processing and storage constraints. **Survey data collection should never be started until the resources for its analysis have been identified.** This may seem obvious, but data collection for the Enquete Nationale sur l'Emploi will be finished in March, and there is still no indication who will carry out the analysis.

With regard to short-term training, it is probably not feasible to think of training analysts in the short-term due to the lack of professional staff. Samuel is currently working with Dr. Ansaonuur on the urban ENBC and could use further training in FOCUS, although he is leaving for studies in the United States in May. More feasible would be to train several people to be assistant programmers, perhaps to the level of Elie. First, they would be able to oversee data entry and verification, including adapting (and eventually writing) data-entry programs. This role is currently filled by Elie for the ENBC, though he may be leaving for training in May. Second, they could assist foreign advisors in producing tables from the data, much as Jean Bosco did on the Enquête Démographique and Elie did with the demographic data of the ENBC. About three or four people should be trained at this level, with Jean Bosco providing technical

support or possibly overall supervision. Halyamutu Adéodat and Mutagorama Javan, data entry supervisors, would be obvious candidates.

6. RECOMMENDATIONS FOR FUTURE ACTIVITIES

There is a clear need for short-term training of assistant programmers, as mentioned in the previous section. There is also a need for assistance in the preparation of a data entry program for the urban household assets data. These activities will need to be carried out sometime between February and May of 1989. Jim Otto would be the best candidate for this work, having had experience with the data entry and processing of the ENBC data.

There may also be a need for further assistance in the processing and analysis of the urban budget data. Although Dr. Ansoanuur has made a great deal of progress during this trip, some topics, such as the system for valuing transactions in kind and the processing of the food consumption data, have only been touched. If Dr. Ansoanuur feels another trip would be productive, I may be available in June or July. Such a trip would be even more useful if it coincided with a proposed seminar on the results of the ENBC. I would present the rural budget data (Volume 3), Niyibizi Bonaventure would discuss rural food consumption (Volume 4), and preliminary results from the urban budget data and the rural anthropometric data could also be presented.

AID/W

**ANNEX A: DESCRIPTION OF PROCESSING OF URDAN ENBC DATA
CARRIED OUT IN JANUARY - FEBRUARY 1989****1. Classification**

In January, there were about 7000 transactions not yet classified. After several rounds of corrections in DEFCOMP, the reasonable combinations of product (PSC) and type of transaction (NATURE) have been classified. Around 270 transactions remain unclassified, consisting mostly of illogical combinations of product and type of transaction that should be checked in the questionnaires.

The first step is to complete the classification for these 270 transactions. Each product-type combination must be either 1) classified by adding it to DEFCOMP with the corresponding account number or 2) corrected in BUFILE using BUCORR so that the new combination of product and type corresponds to an existing combination in DEFCOMP.

The second step is correcting the existing classifications. This involves reviewing the listing produced by DEFCVER of type-product-account number to search for errors in classification. For example, a number of transactions involving salary for a given task (TYPE=064) are incorrectly classified as salaries (COMPTE=54300) instead of services (COMPTE=52XXX). These errors should be identified on the listing, then corrected in DEFCOMP using FILETALK.

The final step is to modify the budget categories themselves in response to the number of observations in each. Budget categories with fewer than a specified number of transactions (perhaps 100) should be collapsed into more aggregated categories. Likely candidates for aggregation include agricultural purchases (COMPTE=21XXX) and agricultural sales (COMPTE=51XXX). Other categories have so many transactions that they could be broken into smaller categories. Candidates for disaggregation include other meat (COMPTE=11520), milk and eggs (COMPTE=11540), revenue from other services (COMPTE=52700), and other salaries (COMPTE=54300). These modifications should be made keeping in mind the need to preserve comparability between rural and urban data sets, the size of the summary file, and the size of the tables to be produced for the report.

After a few rounds of corrections and modifications, the account number in BUFILE, BCPT, can be updated using the program CPTMOD. If this program is run on all the BUFILE files together, the save file will probably use up the disk space. Thus, it is better to run it with BQTMOD, which runs any chosen program seven times once for each of the BUFILE files. The program DEFCVER produces a listing of all the product-type combinations in order of the account number associated with that combination. It is useful for verification of the classification system.

2. Weighting system

The spatial and temporal weighting factors have been calculated in URECH and transferred to BUFILE. The spatial weighting system was based on the description of the sampling method by Roy (1984). The weights have two components: the first reflecting the probability of choosing a given district, the second reflecting the probability of choosing a given household within the district. In the first stratum (Nyarugenge), the districts were chosen with probabilities proportional to the 1983 population, so that the first component of the weights is calculated by the ratio of total population to population in the district chosen. In contrast, the districts for the second stratum were chosen with equal probability, so the first component of weights is calculated by the ratio of the total number of districts to the number selected (60/20). In both strata, the second component of the weighting factor was calculated as the district population in 1984 divided by the number chosen. The former figure came from the data collection files. The raw numbers used for these calculations are stored in URECH, a file initially created by URECHMD1. The calculation of spatial weights, PONS, was done with URECHMD2. It should be noted that the effective sample size is 297, as three households were not surveyed any of the four passages for questionnaire BQ.

The estimated urban population (from the sum of the spatial weights) is 55,971 households. This seems reasonable, but there appears to be an over-representation of Ruhengeri city, which appears significantly larger than Butare according to the weights.

The temporal weights were calculated according to the same formula used for the rural areas. The number of days surveyed in BQ was not obtainable from the file, so verification in the questionnaires was done for any household-passage that had less than 15 days with transactions. In the processing of this verification, seven unentered questionnaires were found and given to Elie for data entry. The weights are calculated assuming the existence of these data in the file. The calculation was done by the programs URECHMD4 and URECHMD5. PONTP is the temporal weight for using BQ data exclusively or for the small transactions if BQ and BR data are used jointly.

The reference period for BR was calculated by the difference between the reference date and the interview date. Verification in the questionnaires was done for any reference period less than 60 days or more than 115 days. The calculation was done by the programs URECHMD6, URECHMD7, and URECHMD8. The variable PONTG is the temporal weight for large transactions in BQ and BR if the two data sets are used jointly.

In order to get the two temporal weights into BUFILE, it was necessary to restructure the files using REBUILD. This is because PONTP and PONTG were originally at the level of the sampling unit (EU) instead of at the level of the household-passage.

The program URECHMD9 will calculate the weighting factors for the questionnaire CA (consommation alimentaire), although it can't be

run until the CA data has been entered. At that point, a save file can be produced with the number of days surveyed for each household-passage. This save file is used by URECHMD9 to calculate the temporal weights.

3. Valuing transactions in kind

A price file, PRIXURB, has been created based on cash transactions found in BQ, and all but 60 of the transactions in kind in BQ have been valued. The prices are calculated on the basis of weighted quantities and values for consumption purchases. Averages prices are available at the city-passage level, at the city-annual level, and at the national-annual level. These averages were calculated by PRIXUMD1, PRIXUMD2, and PRIXUMD3, respectively.

The valuation uses two join commands to link BUFILE to PRIXURB. One of them links product and unit in BUFILE to PRODUN in PRIXURB, while the second links product, unit, and city in BUFILE to PRODUV in PRIXURB. Since PRIXURB contains prices expressed per kilo and per liter, it is necessary to have a "converted unit" in BUFILE. When the unit is grams, the converted unit is kilograms, and when the unit is centiliters, the converted unit is liters. This allows the use of a join command (which cannot use defined variables) to link BUFILE with PRIXURB. The program UNCONMOD puts product and converted unit into the currently unused variable BCPT2. The program BUVALUM1 values gifts given and received as well as barter transactions, placing the imputed value in the variable VAL. Of the 3478 transactions in kind, all but 60 were valued using this program. The remainder will have to be valued "manually" by using prices for similar products or similar units, or by using prices from outside the ENBC.

The price file will also be used to value transactions in kind in BR and food consumption in CA. The same program, BUVALUM, can be used for the valuation of BR. Valuation of food consumption in CA will be more complicated, requiring the creation of a food product conversion file, similar to PRETDAT in the rural data set. This file will convert the product-state codes in CA to the product codes in BQ and BR. It will also have the conversion factors so that valuation of, say cooked beans, is done on the basis of dry weight, since the price file contains prices in the dry, uncooked, and unpeeled state.

4. List of programs used in the ENBC/Urban budget data processing carried out during January - February 1989

- BQTOUS Assembles all the BUFILE files together so that they can be read with TABLE FILE BUFILE. Note that this does not work with MODIFY programs, nor with SCAN operations.
- BQTMOD Runs a user-identified program seven times, once with each of the BUFILE files. It is useful for running MODIFY programs. It can also be used for TABLE programs which may exhaust disk space if run on all the BUFILE files together

CFTMOD Calculates the account number for each transaction in BUFILE according to the definitions in DEFCOMP and adds it to BUFILE under the variable BCPT. The program takes over 10 hours and should be run using BQTMOD.

DEFCVER Produces a table with the number of transactions and the total value by account number, type of transaction, and product. It is useful for checking and modifying classification system.

DEFCMOD Updates the classification system in DEFCOMP using a save file with the new product-type-category combinations.

JENQ1 Calculates the number of days with transactions for each household-passage combination. Used to verify the number of days surveyed (ones with less than 15 were looked up in the questionnaire).

PEMOD Based on verification work with questionnaires, identifies household-passage combinations that were not surveyed and eliminates them from BUFILE

URECHMD1 Creates the urban sample data file, URECH, adding the sampling unit (EU) and the names of prefectures, communes, and sectors.

URECHMD2 Calculates the spatial weighting factors and adds them to URECH.

URECHMD3 Adds household numbers and passages to URECH according to those that exist in BUFILE

URECHMD4 Counts the number of passages surveyed for each household and adds it to URECH

URECHMD5 Calculates the temporal weighting factors for BQ when used alone and for the small transactions in BQ when BQ and BR are used together

URECHMD8 Calculates temporal weighting factors for large transactions in BR when BQ and BR are used together

URECHMD9 Calculates temporal weighting factors for questionnaire CA. Will be ready to run as soon as the number of days surveyed per household-passage is determined.

PONSMOD Transfers the spatial weighting factors from URECH to BUFILE

PONTMOD Transfers the total weighting factors from URECH to BUFILE

PRIXUMD1 Calculates average consumer prices at the city-passage level and adds them to the file PRIXURB

PRIXUMD2 Calculates average consumer prices at the city-annual level and adds them to the file PRIXURB

PRIXUMD3 Calculates average consumer prices and producer prices at the national-annual level and adds them to the file PRIXURB

UNCONMOD Puts the product and converted unit in the variable BCPT2 in BUFILE to facilitate the linking of BUFILE and PRIXURB for the valuation of transactions in kind

BUVALEM1 Values gifts given and received in kind and barter transactions in BUFILE according to the prices in PRIXURB

BUSOMU1 Creates a summary file of annual values for each household and each budget account number, based on BUFILE cash transactions

BUSOMUN Adds the imputed value of transactions in kind from BUFILE to the summary file

DEPVILLE Produces a table of expenditure patterns in each of the four cities based on the summary file

- URECHMD9 Pour calculer les facteurs de pondération temporelle pour la consommation alimentaire dans le questionnaire CA. Il sera lancé dès qu'on a déterminé le nombre de jours enquêtés pour chaque ménage-passage.
- PONSMOD Pour transférer les facteurs de pondération spatiale du fichier URECH au fichier BUFILE
- PONTMOD Pour transférer les facteurs de pondération totale du fichier URECH au fichier BUFILE
- PRIXUMD1 Pour calculer les prix moyens au consommateur au niveau de la ville et du passage et les ajouter au fichier PRIXURB
- PRIXUMD2 Pour calculer les prix moyens au consommateur au niveau de la ville et de l'année et les ajouter au fichier PRIXURB
- PRIXUMD3 Pour calculer les prix moyens au consommateur et au producteur au niveau national et annuel et les ajouter au fichier PRIXURB
- UNCONMOD Pour introduire le code du produit et de l'unité convertie dans la variable BCPT2 dans le fichier BUFILE afin de faciliter l'établissement d'un lien entre BUFILE et PRIXURB pour la valorisation des transactions en nature
- BUVALEM1 Pour valoriser les cadeaux offerts et reçus et le troc dans le fichier BUFILE selon les prix enregistrés dans le fichier PRIXURB
- BUSOMU1 Pour confectionner un fichier sommaire des transactions en espèces au niveau du ménage et du numéro de compte, basé sur les transactions dans BUFILE
- BUSOMUN Pour confectionner un fichier sommaire des transactions en nature au niveau du ménage et du numéro de compte, basé sur les transactions dans BUFILE
- DEPVILLE Pour confectionner un tableau de la structure des dépenses finales selon la ville de résidence à partir du fichier sommaire
- DEPQUINT Pour confectionner un tableau de la structure des dépenses finales selon le quintile de dépenses totales à partir du fichier sommaire et un fichier externe avec la valeur des dépenses totales pour chaque ménage
- MAQ Contient un boîte de titre pour décrire l'objectif de chaque programme

ANNEXE B: LISTE DES FICHIERS DES DONNEES DE L'ENBC/MILIEU RURAL

Q1PRIME0.FOC Ce fichier contient les données du questionnaire Q1, Structure et activités des ménages, pour les 1170 ménages. Les données sont au niveau de l'individu dans le ménage (les caractéristiques et les activités de chaque membre du ménage)

BUDGET.FOC Ce fichier contient les données des questionnaires Q2 et Q4, Budget rétrospectif et Budget quotidien, pour les 270 ménages "quotidiens". Les données sont au niveau du ménage-jour-transaction (e.g. l'achat des haricots dans tel jour)

Q5FILE.FOC Ce fichier contient les données du questionnaire Q5, Consommation alimentaire pour les 270 ménages "quotidiens" en milieu rural. Les données sont au niveau du ménage-jour-composant du repas (e.g. la consommation des haricots dans tel jour)

Q6FILE.FOC Ce fichier contient les données du questionnaire Q6, Patrimoine du ménage, pour les 1170 ménages. Les données sont au niveau du ménage.

BUSOM.FOC Ce fichier contient une sommeire des données budgétaires provenant de Q2 et Q4 et l'autoconsommation provenant de Q5. Il comprend les 270 ménages "quotidien" et les données sont au niveau de l'année-ménage-catégorie budgétaire (e.g. les dépenses annuelles pour les services de taxi).

CASOM.FOC Ce fichier contient une sommeire des données de Q5, comprenant les 270 ménages "quotidiens". Les données sont au niveau de l'année-ménage-catégorie d'aliment.

PRODDAT.FOC Ce fichier contient les coefficients de transformation pour chaque produit selon la nomenclature des questionnaires Q2 et Q4. Les données sont au niveau du produit (e.g. patate douce)

PRETDAT.FOC Ce fichier contient les coefficients de transformation et les coefficients nutritionnels pour chaque produit-état selon la nomenclature du questionnaire Q5. Les données sont au niveau du produit-état (e.g. patate douce épluchée).

MENDAT.FOC Ce fichier contient les caractéristiques sociales, économiques, et démographiques de chacun des 270 ménages "quotidiens". Les données sont au niveau du ménage.

ANNEXE C: DESCRIPTION DES FICHIERS BUDGETAIRES EN MILIEU RURAL

BUDGET.FOC

SEGMENT AU NIVEAU DU MENAGE (segment lié au fichier MENDAT)

MENID A9 Identifiant du ménage

CLASSE1 A1 Variable non-utilisée

CLASSE2 A1 Variable non-utilisée

SEGMENT AU NIVEAU DU MENAGE-QUESTIONNAIRE-PASSAGE

QUESTPER A2 Questionnaire (2 ou 4) et passage (1 à 5)

PONS I4 Pondération spatiale

PONTP I6 Pondération totale pour les transactions < 200 FRw

PONTG I6 Pondération totale pour les transactions >= 200 FRw

SEGMENT AU NIVEAU DU MENAGE-QUESTIONNAIRE-PASSAGE-JOUR

JOUR A3 Numéro du jour

SEGMENT AU NIVEAU DU MENAGE-QUESTIONNAIRE-PASSAGE-JOUR-TRANSACTION
(segment lié au fichier PRODDAT)

TABLEAU A1 Tableau (voir TYPESNOMS.DAT)

TRANS A4 Numéro de la transaction

COMPTE A4 Catégorie budgétaire (voir COMPNOMS.DAT)

MONET A1 Non-utilisé

TYPE A2 Type de transaction (voir TYPENOMS.DAT)

PROD A3 Produit ou service (voir PRODDAT.FOC)

UNITE A1 Unite (voir UNITENOMS.DAT)

QUANT I6 Quantité

VALEUR I7 Valeur versée

VALORIZ I7 Valeur imputée d'une transaction en nature

TYPEVAL A1 Type de valorisation

PQUI A2 Personne réalisant la transaction (voir PQUINOMS.DAT)

SEGMENT POUR LES ACHATS ET LES VENTES

PROD2 A3 Produit-service échangé en contrepartie

UNITE2 A1 Unité du contrepartie (voir UNITENOMS.DAT)

QUANT2 I6 Quantité du contrepartie

MODE A1 Mode de la transaction (voir MODENOMS.DAT)

DEJA12 I7 Montant déjà versé

LIEU A2 Lieu de la transaction (voir LIEUNOMS.DAT)

UTIL12 A2 Utilisation du bien (voir UTILNOMS.DAT)

RESS12 A2 Ressource pour l'achat (voir RESSNOMS.DAT)

NFOIS12 I2 Nombre de fois réalisant la transaction (Q2)

SEGMENT POUR LES CADEAUX

ADEQUI A2 A/de qui donné/reçu (AQUINOMS.DAT)

MOTIF A2 Motif pour le cadeau (MOTIFNOMS.DAT)

SEGMENT POUR LES OPERATIONS FINANCIERES

DEJA46 I7 Montant déjà versé

AFAIRE I7 Montant à faire

TOTAL I7 Montant total

AVECQUI A2 Avec qui on l'a réalisé (voir AQUINOMS.DAT)

UTIL46 A2 Utilisation (voir UTILNOMS.DAT)

RESS46 A2 Ressource (voir RESSNOMS.DAT)

NFOIS46 I2 Nombre de fois réalisé (Q2 seulement)

DUREE I3 Durée de l'emprunt ou le prêt

BUSOM.FOC

SEGMENT AU NIVEAU DU MENAGE (segment lié au fichier MENDAT)
 MENID A9 Identifiant du ménage
 PONS I4 Fondération spatiale
 SEGMENT AU NIVEAU DU MENAGE-CATEGORIE BUDGETAIRE
 COMPTE A4 Catégorie budgétaire (voir COMPNOMS.DAT)
 MONET A1 E=Espèces N=Nature A=Autoconsommation T=Transfert
 VALEUR I7 Valeur en FRw/ménage/an
 UNITE A1 Unite (pour les produits agricoles)
 QUANT D7.1 Quantité en kg/ménage/an

PRODDAT.FOC

SEGMENT AU NIVEAU DU PRODUIT
 PROD A3 Code du produit ou service provenant de Q2-Q4
 PRODNOM A40 Nom du produit ou service
 DENSITE I6 Densité en grammes/unité ou gramme/litre
 COMPTED A4 Catégorie de consommation si reçu comme transfert
 COMPTEP A4 Catégorie de production si offert comme transfert
 PRODAL A3 Code de la catégorie alimentaire
 PRODUIT A20 Nom de la catégorie alimentaire
 CONQST D4.2 Taux de transformation à la catégorie

PRETDAT.FOC

SEGMENT AU NIVEAU DU PRODUIT-ETAT
 PRET A4 Code du produit-état provenant de Q5
 PRETNOM A40 Noms du produit-état
 PROD A3 Produit qui correspond
 DENSITE I4 Densité en gramme/unité ou gramme/litre
 ALIM3 A3 Catégorie de trois chiffres (sorgho, maïs)
 ALIM30 A3 Variable non-utilisée
 ALIMPR I2
 ALIM2 A2 Catégorie de deux chiffres (céréales)
 ALIM1 A1 Catégorie d'un chiffre (1=végétal, 2=animal)
 CONQST D4.2 Taux de transformation à la catégorie
 CONQCOM D4.2 Partie comestible
 CALORIE I3 Calorie par 100 gm de partie comestible
 PROTEINE D4.1 Grammes de protéine par 100 gm de partie comestible
 LIPIDE D5.1 Grammes de lipide par 100 gm de partie comestible
 CALCIUM D7.2 Calcium par 100 gm de partie comestible
 FER D4.1 Calcium par 100 gm de partie comestible
 RETINOL I3 Vide
 CAROTENE I5 Vide
 THIAMINE D4.2 Vide
 RIBOFLAV D4.2 Vide
 NIACINE D4.1 Vide
 VITAC I3 Vide

MENDAT.FOC

SEGMENT AU NIVEAU DU MENAGE

MENID	A9	Identifiant du ménage
PONS	I4	Pondération spatiale
PREF	A2	Code de préfecture
PREFECTURE	A10	Noms de préfecture
ZN	A1	Code de zone
ZONE	A11	Noms de zone
REGION	A2	Région agroclimatic (voir Volume 2)
ALTITUDE	I4	Altitude en metres
PLUVIO	I4	Pluviometrie en mm par an
SEXECHIEF	A1	Sexe du chef de ménage (1=homme, 2=femme)
AGECHIEF	I2	Age du chef de ménage
SCMARI	A1	Vide
SCFENME	A1	Vide
OCCUP	A1	Occupation principale (voir Volume 3)
AGNONAG	A1	1=Agricole, 2=Non-agricole
NPER	D4.1	Nombre de personnes
NADULTES	D4.1	Nombres d'adultes
NENFANTS	D4.1	Nombres d'enfants
NADEQ	D5.2	Nombres d'adultes-équivalent (AE)
QNADEQ	I1	Quintile d'adultes-équivalent
CALAD	D5	Consommation calorique/jour/AE
QCALAD	I1	Quintile de consommation calorique
QSUPAD	I1	Quintile de superficie par AE
SUPAD	I4	Superficie par AE
COALAD	I6	Valeur de la cons. alimentaire par AE
CONPC	I6	Valeur de la consommation par tête
QUINPC	A1	Quintile de CONPC
CONAD	I6	Valeur de la consommation par AE
QUINAD	I1	Quintile de CONAD
PCTAG	I4	Agriculture comme % de la valeur ajoutée
ASGEN	I2	Autoconsommation comme % de la consommation
ASALIM	I2	Autoconsommation comme % de la cons. alim.
SUPERF	I6	Superficie de l'exploitation en ares
AGRI	D7	Valeur ajoutée agricole (FRW/an)

Q3PER.FOC

SEGMENT AU NIVEAU DU SECTEUR (lié au fichier ECHANT)
 SECTID A7 Identifiant du secteur

SEGMENT AU NIVEAU DU MENAGE

MENAGE A2 Numéro du ménage

SEGMENT AU NIVEAU DU MENAGE-PASSAGE

QUESTPER A2 Questionnaire (3) et passage (1 à 4)

JOURJ A3 Jour enquêté

PONS I5 Pondération spatiale

PONTP I6 Pondération temporelle

SEGMENT AU NIVEAU DE PERSONNE-PASSAGE

ORD A2 Numéro de la personne

AGE I2 Age

SEXE A1 Sexe

POIDS I2 Poids en kilogramme

TAIL I3 Taille en centimetres

TBRA I3 Taille du bras en centimetres (enfants)

FCN A1

LIENP A2 Lien de parentage

PRIXACH.FOC

SEGMENT AU NIVEAU DU PRODUIT-ETAT (lié au fichier PRODDAT)

PRODUN,	A4,	Combinaison de produit-unité
PROD,	A3,	Produit
PRIXNA,	D7.1	Prix d'achat au niveau national annuel
CNTNA,	I4	Nombre d'observations
PRIXVNA,	D7.1	Prix de vente au niveau national annuel
CNTVNA,	I4	Nombre d'observations
PRIXMAX,	D7.1	Prix maximum
PRIXMIN,	D7.1	Prix minimum

SEGMENT AU NIVEAU DU PRODUIT-UNITE-ZONE

PRODUZ,	A5,	Combinaison de produit-unité-zone
PRIXZA,	D7.1	Prix d'achat au niveau zone-annuel
CNTZA,	I4	Nombre d'observations
PRIXZP1,	D7.1	Prix d'achat au niveau zone-passage 1
CNT1,	I3	Nombre d'observations
PRIXZP2,	D7.1	Prix d'achat au niveau zone-passage 2
CNT2,	I3	Nombre d'observations
PRIXZP3,	D7.1	Prix d'achat au niveau zone-passage 3
CNT3,	I3	Nombre d'observations
PRIXZP4,	D7.1	Prix d'achat au niveau zone-passage 4
CNT4,	I3	Nombre d'observations

ECH.FOC

SEGMENT AU NIVEAU DU SECTEUR

SECTID	A7,	Identifiant du secteur
ZONE	A2	Zone géographique
REGION	A2	Région agro-climatique
PREFECTURE	A10	Nom de préfecture
COMMUNE	A15	Nom de commune
SECTEUR	A15	Nom de secteur
MENAGED	I3	
PROBSEL	D5.3	
SECTC	I2	Nombre de secteurs dans la commune
COMZ	D5.2	Nombre de communes dans la zone
COMTIRZ	D5.2	Nombre de communes tirées dans la zone
PONS	I4	Pondération spatiale

SEGMENT AU NIVEAU DU MENAGE

MENAGE	A2	Numéro du ménage
--------	----	------------------

SEGMENT AU NIVEAU DU MENAGE-QUESTIONNAIRE-PASSAGE

QUESTPER	A2	Questionnaire (2,4,5) et passage (1 à 5)
PENQ	I1	Passages enquêtés
JENQ	I3	Jours enquêtés
PERJRS	I4	
PONTG	D7.4	Pondération temporelle pour > 200 FRW
PONTP	D7.4	Pondération temporelle pour < 200 FRW

**ANNEXE D: LISTE DES PROGRAMMES PRINCIPAUX DE TRAITEMENT POUR
L'ENBC EN MILIEU RURAL**

- PRIXMOD1 Programme pour calculer les prix moyens au niveau du passage et de la zone géographique et les ajouter au fichier PRIXACH.FOC
- PRIXMOD2 Programme pour calculer les prix moyens au niveau annuel et de la zone géographique et les ajouter au fichier PRIXACH.FOC
- PRIXMOD3 Programme pour calculer les prix moyens au niveau annuel et national et les ajouter au fichier PRIXACH.FOC
- PRIXMOD4 Programme pour dupliquer les prix de quelques produits pour d'autres produits semblables
- BUVALM1 Programme pour valoriser les transactions en nature (sauf les ventes) dans BUDGET.FOC selon les prix dans le fichier PRIXACH.FOC
- BUVALM2 Programmes pour valoriser les ventes en nature dans le fichier BUDGET.FOC selon les prix dans le fichier PRIXACH.FOC
- BUVALM3 Programmes pour valoriser manuellement les transactions en nature qui ne sont pas valorisées par BUVALM1 et BUVALM2.
- Q5VALM Programme pour valoriser la consommation alimentaire dans le fichier Q5FILE.FOC selon les prix dans le fichier PRIXACH.FOC
- BUSOM1 Programme pour créer un fichier sommaire (BUSOM.FOC) à partir des données budgétaires de BUDGET.FOC
- BUSOM2 Programme pour ajouter la valeur des transferts au fichier sommaire BUSOM.FOC
- BUSOM3 Programme pour ajouter la valeur de l'autoconsommation à partir de Q5FILE.FOC au fichier sommaire BUSOM.FOC comme consommation
- BUSOM4 Programme pour ajouter la valeur de l'autoconsommation à partir de Q5FILE.FOC au fichier sommaire BUSOM.FOC comme production agricole et artisanale
- BUSOM5 Programme pour calculer les totaux et les sous-totaux pour le fichier sommaire BUSOM.FOC

**ANNEXE B: LISTE DES PROGRAMMES POUR CONFECTIONNER LES
TABLEAUX DU VOLUME 3 DE L'ENBC/MILIEU RURAL**

TCR00	2.01	Résumé de la consommation finale des ménages ruraux
TCR20	2.02	Origine de la consommation des ménages ruraux (FRw/ménage/an)
TCR21	2.03	Origine de la consommation des ménages ruraux (% en ligne)
TCR22	2.04	Origine de la consommation des ménages ruraux (% en colonne)
TCR10A	2.05	Pourcentage des ménages ruraux consommant chaque type de bien et service
TBM10A	2.06	Pourcentage des ménages ruraux réalisant chaque type de dépenses finales
TCR50	2.07	Distribution des ménages ruraux selon la valeur de la consommation par ménage
TCR60	2.08	Distribution des ménages ruraux selon la valeur de la consommation par tête
TCR70	2.09	Distribution des ménages ruraux selon la valeur de la consommation par adulte-équivalent
TBM20	2.10	Distribution des ménages ruraux selon les dépenses finales par ménage
TCR01	2.11	Structure de la consommation alimentaire
TCR30	2.12	Origine de la consommation alimentaire (FRw/ménage/an)
TCR31	2.13	Origine de la consommation alimentaire (% en ligne)
TCR32	2.14	Origine de la consommation alimentaire (% en colonne)
TCR10B	2.15	Pourcentage des ménages ruraux consommant chaque type de produit alimentaire
TBM11	2.16	Pourcentage des ménages ruraux réalisant chaque type de dépenses alimentaires
TCR80	2.17	Distribution des ménages selon la valeur de la consommation alimentaire par tête
TCR81	2.18	Distribution des ménages ruraux selon la part de la consommation allouée à l'alimentation
TCR02	2.19	Structure de la consommation non-alimentaire
TCR10C	2.20	Pourcentage des ménages ruraux consommant chaque type de bien et service non-alimentaire
TBM12	2.21	Pourcentage des ménages ruraux réalisant chaque type de dépenses non-alimentaires
TCR90	2.22	Distribution des ménages ruraux selon la valeur de la consommation non-alimentaire par tête
TPR05	3.01	Structure de la valeur brute de la production et la valeur ajoutée selon le secteur économique
TBM07	3.02	Structure du revenu monétaire brut et net selon le secteur économique
TPR20	3.03	Utilisation de la production des biens et des services (FRw/ménage/an)
TPR21	3.04	Utilisation de la production des biens et des services (% en ligne)
TPR22	3.05	Utilisation de la production des biens et des services (% en colonne)

TPR10A	3.06	Pourcentage des ménages ruraux produisant chaque type des biens et des services
TBM10B	3.07	Pourcentage des ménages ruraux percevant chaque type de revenu monétaire
TPR50	3.08	Distribution des ménages ruraux selon la valeur ajoutée par ménage
TPR60	3.09	Distribution des ménages ruraux selon la valeur ajoutée par tête
TBM30	3.10	Distribution des ménages ruraux selon le revenu monétaire net par ménage
TPR01	3.11	Structure de la valeur ajoutée agricole
TBM03	3.12	Structure du revenu monétaire agricole
TPR30	3.13	Utilisation de la production agricole (FRW/ménage/an)
TPR31	3.14	Utilisation de la production agricole (% en ligne)
TPR32	3.15	Utilisation de la production agricole (% en colonne)
TPR10B	3.16	Pourcentage des ménages ruraux produisant chaque type de denrée agricole
TBM13	3.17	Pourcentage des ménages ruraux percevant chaque type de revenu agricole
TSR10	3.18	Distribution des ménages ruraux selon la valeur ajoutée agricole par ménage
TSR50	3.19	Distribution des ménages ruraux selon l'importance de l'agriculture
TPR02	3.20	Structure de la valeur ajoutée artisanale
TBM04	3.21	Structure du revenu monétaire artisanal
TPR10C	3.22	Pourcentage des ménages ruraux produisant chaque type de bien et service artisanaux
TBM14	3.23	Pourcentage des ménages ruraux percevant chaque type de revenu artisanal
TSR20	3.24	Distribution des ménages ruraux selon la valeur ajoutée artisanale par ménage
TSR60	3.25	Distribution des ménages ruraux selon l'importance de l'artisanat
TPR03	3.26	Structure de la valeur ajoutée commerciale
TPR10D	3.27	Pourcentage des ménages réalisant chaque type de commerce
TSR30	3.28	Distribution des ménages ruraux selon la valeur ajoutée commerciale par ménage
TSR70	3.29	Distribution des ménages ruraux selon l'importance du commerce
TPR04	3.30	Structure de la valeur ajoutée salariale
TPR10E	3.31	Pourcentage des ménages percevant chaque type de salaire
TSR40	3.32	Distribution des ménages ruraux selon le niveau des salaires par ménage
TSR80	3.33	Distribution des ménages ruraux selon l'importance des salaires
TDM10	4.01	Caractéristiques des ménages en milieu rural selon la valeur de la consommation
TDM18	4.02	Structure de la consommation finale en milieu rural selon la valeur de la consommation
TDM19	4.03	Structure de la valeur ajoutée en milieu rural selon la valeur de la consommation
TDM20	4.04	Caractéristiques des ménages en milieu rural selon l'occupation principale

TDM28	4.05	Structure de la consommation finale en milieu rural selon l'occupation principale
TDM29	4.06	Structure de la valeur ajoutée en milieu rural selon l'occupation principale
TDM30	4.07	Caractéristiques des ménages en milieu rural selon le taux d'autoconsommation alimentaire
TDM38	4.08	Structure de la consommation finale en milieu rural selon le taux d'autoconsommation alimentaire
TDM39	4.09	Structure de la valeur ajoutée en milieu rural selon le taux d'autoconsommation alimentaire
TDM90	4.10	Caractéristiques des ménages en milieu rural selon la consommation calorique
TDM98	4.11	Structure de la consommation finale en milieu rural selon la consommation calorique
TDM99	4.12	Structure de la valeur ajoutée en milieu rural selon la consommation calorique
TDM40	4.13	Caractéristiques des ménages en milieu rural selon la superficie de l'exploitation
TDM48	4.14	Structure de la consommation finale en milieu rural selon la superficie de l'exploitation
TDM49	4.15	Structure de la valeur ajoutée en milieu rural selon la superficie de l'exploitation
TDM50	4.16	Caractéristiques des ménages en milieu rural selon la zone géographique
TDM58	4.17	Structure de la consommation finale en milieu rural selon la zone géographique
TDM59	4.18	Structure de la valeur ajoutée en milieu rural selon la zone géographique
TDM60	4.19	Caractéristiques des ménages en milieu rural selon la taille du ménage
TDM68	4.20	Structure de la consommation finale en milieu rural selon la taille du ménage
TDM69	4.21	Structure de la valeur ajoutée en milieu rural selon la taille du ménage
TDM70	4.22	Caractéristiques des ménages en milieu rural selon le sexe du chef de ménage
TDM78	4.23	Structure de la consommation finale en milieu rural selon le sexe du chef de ménage
TDM79	4.24	Structure de la valeur ajoutée en milieu rural selon le sexe du chef de ménage
TDM80	4.25	Caractéristiques des ménages en milieu rural selon l'âge du chef de ménage
TDM88	4.26	Structure de la consommation finale en milieu rural selon l'âge du chef de ménage
TDM89	4.27	Structure de la valeur ajoutée en milieu rural selon l'âge du chef de ménage

ANNEX F: SUMMARY OF COMMANDS IN FOCUS

I. FOCUS COMMANDS - Used with the double arrow prompt. Some are used to enter a given mode (DEFINE, TABLE, MODIFY, TED, SCAN, DOS), some perform a specific operation (EX, CHECK, ?), while the others establish setting which remain active until you leave FOCUS or change them with another command (JOIN, ONLINE, OFFLINE, FILEDEF, SET).

DEFINE FILE file1

Marks beginning of block of definitions applicable to file1. File1 must be a FOCUS file or a HOLD file. See description of definitions in section II.

TABLE FILE file1

Marks beginning of block of commands for producing a report from file1. File1 must be a FOCUS file or a HOLD file. See description of TABLE commands in section III.

MODIFY FILE file1

Marks beginning of block of commands for modifying file1. File1 must be either a FOCUS file or a HOLD file.

SCAN FILE file1

Enters scan mode to review FOCUS files. See section IV.

TED file1

Enters text editor to create, review, and edit programs and any files except FOCUS files (i.e. those with the suffix .FOC). See section V for a more complete description.

DOS

Returns temporarily to the DOS mode for file management (copying, renaming, and deleting files and so on). The EXIT command in DOS will then cause a return to the FOCUS environment.

EX file1

Executes FOCUS program by the name of file1.

CHECK FILE file1

Check the structure of file1 for errors. Often used for newly created master files. PICTURE option at end produces a diagram of the structure of the file.

? SET

Gives all the current settings for those options which are determined by the SET command.

? nnn

Gives the full explanation for error message number nnn.

JOIN var1 IN file1 TO var2 IN file2

Links two FOCUS files together using one or more variables in common. Allows use of variables in second file while using first file. Var2 in file2 must be an indexed variable.

ONLINE

Sends output to the screen.

OFFLINE

Sends output to the printer, or to a file if a FILEDEF has been issued.

FILEDEF OFFLINE DISK file1.xxx

Used to send output to a file instead of to the printer. Functions only if OFFLINE command has been issued.

FILEDEF var1 DISK file1.xxx

Used with a DECODE definition to identify the external SAVE file where the decoded variable is found. Var1 is the same as the variable in parentheses in the DECODE definition. The external file (file1.xxx) contains two columns, the code and the decoded name.

SET WIDTH= n, CDN=ON

Sets the width of the printout (maximum is 132) and sets continental decimal notation on, so that 1.200.3 becomes 1.200,3 as is the standard notation in Europe and in Rwanda. The default is off.

SET MORE = OFF

In processing the data, status messages and error messages may be sent to the screen. When the screen is full, processing normally stops until a key is tapped. With MORE=OFF, processing continues without the need to tap a key. This is useful for programs with long run-times.

II. DEFINITIONS - Definitions are included in a block that starts with DEFINE FILE file1 and ends with END. The definitions apply only to that file (file1). The definitions take the form:

var1/format = expression ;

The most common formats are A (alphanumeric), D (decimal), and I (integer), so that A5 means alphanumeric with five characters and I3 means a 3-digit integer. The format D7.2 means a decimal with a total length of 7 characters, including 2 to the right of the

decimal point, the point itself, and four to the left of the point. There are also various date and time formats, as well as options for suppressing values of zero, bracketing negative numbers, including dollar signs, and so on.

The expression may include IF/THEN/ELSE, AND, and OR statements, and nesting of these operations is allowed. All variables in the expression must be existing variables (in the active file or in linked files) or previously defined variables. The expression may include the following functions:

EDIT (var1)

Defines a new variable which is the same as var1 except that the format is switched from alphanumeric to numeric or numeric to alphanumeric.

EDIT (var1, '999\$\$\$')

Defines a new variable which contains only some of the characters of var1. The new variable contains the characters in var1 represented by 9's, leaving out those represented by \$'s (in this example, the first three characters are selected). Var1 must be alphanumeric.

var1ùvar2

Defines a new variable which combines var1 and var2, both of which must be alphanumeric. The resulting variable is alphanumeric as well.

DECODE var1 (var2 ELSE value1) ;

or

DECODE var1 (value2 valueA value3 valueB ...
value4 valueC ELSE value1) ;

Defines a new variable according to the value of the already defined variable var1, e.g. defining names that correspond to existing codes. If no correspondence is indicated, var1 takes the value value1. In the first example, the correspondence between the old and new variables is defined in an external file containing two columns (old and new). The file is identified by a FILEDEF command containing var2. In the second example the correspondence between the old and new variables is determined by the matched pairs in parentheses. Alphanumeric expressions (value names) must be in single quotation marks.

III. TABLE COMMANDS - The commands to produce a table are included in a block that starts with TABLE FILE file1 and ends with the command END. The program can be written in TED (see section IV) or simply by issuing the command: TABLE FILE filename. The order of commands should be roughly the following:

TABLE FILE file

HEADING

" title "

WRITE/PRINT/COUNT var1 var2 ...

COMPUTE var3/format = expression ;

```

ROW-TOTAL/COLUMN-TOTAL
HOLD/SAVE
ACROSS var4 AS 'expression'
BY      var5 AS 'expression' SUB-TOTAL
BY      var6 AS 'expression' IN-GROUPS-OF value
IF condition
END

```

Some of the important phrases are described below.

HEADING

```
"      title      "
```

Places a title over the table. The title may have multiple lines, each with a pair of quotation marks.

```
PRINT var1 var2 ...
```

Lists the values of the named variables for each record in the file. The variables may be numeric or alphanumeric.

```
WRITE var1 var2 ...
```

Calculates the sum of the named variables at the level of aggregation determined by the ACROSS and BY phrases. The variables must be numeric. The main options are:

PCT.var1	Column percentage of variable
RPCT.var1	Row percentage of variable (used with ACROSS)
CNT.var1	Number of records at the level that the variable appears.
PCT.CNT.var1	Column percentage of number of records.
AVE.var1	Average value of variable.

Other prefix options include MIN (minimum), MAX (maximum), and ABS (absolute value).

```
COMPUTE var1/format = expression ;
```

Calculates and displays a new variable (var1). This command comes after the verb statement (e.g. WRITE var1 var2). The expression must be in terms of variables already listed after the verb. C1, C2, C3 ... can be used to refer to the variables already displayed. The equation may use IF/THEN/ELSE, AND, and OR conditions.

```
ACROSS var1
```

Creates a cross-tabulation in which the figures are organized horizontally according to the values of var1. ACROSS must be used with WRITE and var1 must have a relatively limited number of values, generally less than 6. Multiple ACROSS's are allowed but often create excessively wide tables.

```
BY var1
```

Creates a cross-tabulation in which the figures are organized vertically according to the value of var1, ordered according to the value of var1. Multiple BYs are possible.

```
preposition var1 IN-GROUPS-OF value1
```

Creates a cross-tabulation in which the figures are organized according to the range that var1 falls into. For example, BY VALEUR IN-GROUPS-CF 100 gives one line for VALEUR from 1 to 100, another for VALEUR from 101 to 200, and so on. Adding the expression TOP value2 defines the last group as greater than value2.

```
FOR var1
value1 TO value2 OVER
value3 TO value4 OVER
:
value5 TO value6 OVER
value7 TO value8
```

Similar to BY var1 IN-GROUPS-OF value1 except that each line is explicitly defined. For example, value1 and value2 define the lower and upper limits of var1 for the first line of the table.

```
var1 AS 'expression'
```

Gives a label to the variable var1 which is used in the table. In the absence of an AS phrase, the variable name itself appears in the table.

ROW-TOTAL

Adds the row-total to the table. Can only be used with the ACROSS statement. Does not give meaningful results when the numbers are averages or column percentages.

COLUMN-TOTAL

Adds column-totals to the table. Can only be used with the BY statement. Does not give meaningful results when the numbers are averages or row-percentages.

BY var1 SUB-TOTAL

Calculates a sub-total each time var1 changes value. If there are multiple BY statements, SUB-TOTAL gives sub-totals for this BY and for all BY's above it. In contrast, SUBTOTAL gives sub-totals only at this break.

HOLD

Sends results to a HOLD file. A HOLD file can be read with TED (unlike a FOCUS file) and has a master file. Most importantly, it can be used as a data file; in other words, a table can be produced from a HOLD file using TABLE FILE HOLD and normal table commands. The file is called HOLD.FTM and the master file HOLD.MAS unless AS filename is used to designate another name.

SAVE

Sends results to a SAVE file. A SAVE file is in ASCII format, can be read with TED, but has no master file, nor can it be used as a data file to produce tables. It can be used as an external

file to establish a correspondence between two variables in conjunction with FILEDEF and DECODE commands. It is also used to create and modify FOCUS files in conjunction with a MODIFY program. The file is called SAVE.FTM unless AS filename is used to designate another name.

IF condition

Selects transactions to be processed. The condition must be a simple equality or inequality, without AND or OR operations.

IV. TEXT EDITOR (TED) COMMANDS - Used to create, review, and edit programs and any files other than FOCUS files. Entered through the command: TED filename.

1. Commands at foot of page

<tab>	Go to beginning of next line
<return>	Go to foot of page; execute deletion, copy, and move
<insert>	Change from write-over to insert mode or vice-versa
A n	Add n lines at the top of the screen
D n	Move down n lines
U n	Move up n lines
TO	Move to top of file
BO	Move to bottom of file
QU	Exit file without saving
QQU	Exit file without saving ignoring changed file
FI	Save file and exit
RU	Save file and run program it contains
E	Change to edit mode which allows line commands
NU	Change to number mode in which line numbers are shown.
TY	Change to type mode
GET file	Copy external file into current file at top of page
CASE U	Change to mode in which all letters made uppercase
CASE M	Change to mode in which letters left as is
F6	Move rest of current line onto next line
Alt F6	Join next line to the end of current line
L/a	Search for string a
C/a/b/de	Search and replace string a by string b for d lines and e times each line

2. Line commands

Ctl A	Insert a blank line below current line
Ctl D	Delete current line
M	Identifies line to move
C	Identifies line to copy
D	Identifies line to delete
MM	Identifies beginning or end of block to move
CC	Identifies beginning or end of block to copy
DD	Identifies beginning or end of block to delete
P	Copy or move text to current line
"n	Duplicate current line n times

IV. SCAN COMMANDS - Used for reviewing a FOCUS file. Scan mode started by the command: SCAN FILE filename. Editing in scan mode is possible but not recommended.

NN Display next record
SHOW x y Display the following variables
TO Go to top of file
BO Go to bottom of file
LO expr Locate records meeting following condition
F9 Move to left
F10 Move to right
QU Quit scan mode

V. TEXT EDITOR (TED) COMMANDS - Used to create, review, and edit programs and any files other than FOCUS files. Entered through the command: TED filename.

1. Commands at foot of page

<tab> Go to beginning of next line
<return> Go to foot of page; execute deletion, copy, and move
<insert> Change from write-over to insert mode or vice-versa
A n Add n lines at the top of the screen
D n Move down n lines
U n Move up n lines
TO Move to top of file
BO Move to bottom of file
QU Exit file without saving
QQU Exit file without saving ignoring changed file
FI Save file and exit
RU Save file and run program it contains
E Change to edit mode which allows line commands
NU Change to number mode in which line numbers are shown
TY Change to type mode
GET file Copy external file into current file at top of page
CASE W Change to mode in which all letters made uppercase
CASE M Change to mode in which letters left as is
F6 Move rest of current line onto next line
Alt F6 Join next line to the end of current line
L/a Search for string a
C/a/b/de Search and replace string a by string b for d lines and e times each line

2. Line commands

Ctl A Insert a blank line below current line
Ctl D Delete current line
M Identifies line to move
C Identifies line to copy
D Identifies line to delete
MM Identifies beginning or end of block to move
CC Identifies beginning or end of block to copy
DD Identifies beginning or end of block to delete
P Copy or move text to current line
"n Duplicate current line n times

ANNEX G: COMPARATIVE TESTS OF SPSS AND FOCUS

Files with identical data content were created in SPSS and FOCUS. Several similar programs were written for each package to produce reports typical of those needed for analysis of budget and consumption surveys. The two packages are compared in several activities, such as creation of a summary file and producing different kinds of tables. Three criteria were considered in evaluating the packages: the efficiency of data storage, the speed of operation, and the ease of programming.

1. Comparison of data storage efficiency

The BUDGET file contains cash and non-cash transactions made by 270 rural Rwandan households over one year. It has around 11,000 cases and 19 variables. The original FOCUS file was used to generate an ASCII file which was in turn loaded into an SPSS system file. The FOCUS file is 3 Mb in size. The uncompressed SPSS file occupies 11 Mb, while the compressed version of the same file is only 2.7 Mb. The compression of the SPSS file does reduce the speed of operations somewhat, but the large saving in disk space was deemed worth the reduced performance. Hence, subsequent tests were done on the compressed version of the SPSS file.

It should be noted that FOCUS took a full ten hours to generate an ASCII file from the 11,000 cases in BUDGET. The slowness of this procedure is apparently due to the hierarchical nature of FOCUS files. For the purpose of comparison, an ASCII file was produced from the SPSS system file, a procedure that took less than two hours. This is presumably because SPSS system files are already rectangular and more similar to ASCII files.

2. Comparison of run times

Four tests were done to compare the operation and performance of SPSS and FOCUS: a simple cross-tabulation, the creation of a summary file, an average expenditure breakdown, and a table of average expenditure as a function of a household characteristic. The run times for the SPSS programs should be considered tentative, as Chris Wolfe is working on improving the efficiency of storage and retrieval.

Test 1: Simple cross-tabulation

Test: The two software packages were compared in their ability to count records as a function of two variables. This program was run on the BUDGET file with its 10,000 transactions.

Results: FOCUS produced the table with row and column totals in 6 minutes. SPSS generated a similar cross-tabulation in 12 minutes.

Commentary: Although FOCUS was about twice as fast in this task, it should be noted that the SPSS table included row and column percentages. Although FOCUS can calculate these percentages easily, generating the row and column totals of these percentages is cumbersome.

Test 2: Producing a summary file

Test: SPSS and FOCUS were compared in their ability to create a summary file from the original BUDGET file. The BUDGET file contains data at the level of the household-transaction, for example the purchase of a shirt by household X on day Y. The summary file would have data at the level of household-budget category, for example the annual expenditure on new clothing by household X. This new file would include household number, budget category code, transaction type code (indicating cash or in-kind), value in francs per year, and quantity in kilograms or liters per year. This procedure involved the application of temporal weights and standardization of quantity across various units (e.g. grams, centiliters, etc.) for agricultural products. An important aspect of this test is that it required the use of product-specific conversion factors stored in another file.

Result: FOCUS created the summary file in about an hour, including about five minutes to create the "master file" which describes the structure of the new file and 53 minutes of run time. SPSS took 6 hours and 17 minutes, although a more efficient program could have reduced considerably the time required.

Commentary: The speed of FOCUS is partly the result of the permanent link which had been established between the BUDGET file and the auxiliary file containing the conversion factors for each product. As a result of this link, the conversion factors in the auxiliary file were available from the BUDGET file. Furthermore, this look-up function does not require sorting or the creation of a file combining data from both files.

As written, the SPSS program required the following time for each step:

Sort by product code	2 hrs 30 min
Join files	35 min
Compute new variables	2 hrs 0 min
Sort by household etc.	58 min
Aggregate	14 min

In addition, the program stopped several times, having exhausted memory or disk space. This is because in order to use a variable from another file, a new file must be created combining the two files using the JOIN MATCH command. If the original file is not sorted by the key variable (in this case, the product code), the combined file is assembled in memory. However, for a large file such as this, memory capacity is quickly exhausted. On subsequent runs, the file was sorted first. Even so, the need to sort a 30,000 case file also required a large amount of disk capacity. At one point, even 17 Mb available disk space was not enough.

It is important to note that the program could have been improved to run faster and use less disk space, most importantly by eliminating unused variables before sorting. Only 10 of the 39 variables in BUDGET were needed, and Chris Wolfe estimates that the sorting time could be reduced by 75% or more by dropping the others. One implication of this is that SPSS requires the pro-

grammer to more precisely identify the file management activities which need to be done. FOCUS handles automatically some of the tasks for which SPSS depends on commands from the programmer.

Test 3: Producing a summary budget table

Test: Using the summary file created in the previous test, a breakdown of average household expenditure by budget category was produced with each software package. The summary file has about 10,000 records and five variables.

Results: When the total number of households in the sample (270) was included in the program, FOCUS took 1.5 minutes and SPSS (using the MEANS command) took 5 minutes. When this number was not included in the program, FOCUS required 3 minutes, while SPSS ran for 5.5 minutes using MEANS and 6.7 minutes using the TABLE command.

Commentary: Including the number of households reduced the times for both packages because they could simply calculate the sum of expenditures on each category and divide by 270. Without this number, an intermediary file had to be created and the averages calculated from this file. In either case, however, FOCUS ran two to three times faster. The speed of SPSS depended on which command was used: the MEANS command is faster but the format of the table is not adjustable, while the TABLE command is slower but the output can be formatted for improved presentation. The output in these FOCUS runs were formatted, making them more comparable to the SPSS runs using the TABLE command.

Test 4: Producing a budget table by sex of head of household

Test: This test was similar to the previous one except that average expenditures were cross-tabulated with the sex of the head of household, a variable found in a separate household characteristic file. Again, the summary budget file with its 10,000 records was used.

Results: FOCUS performed this task in 3.3 minutes, while SPSS required 30.5 minutes.

Commentary: The speed of FOCUS in this test was due to the existence of a permanent link between the budget file and the household characteristics file, allowing household characteristics to be used directly from the budget file. By contrast, SPSS had to create a new file combining the budget data and the sex of head of household. Establishing this link in FOCUS did, of course, require an extra step in the creation of the summary file, which was part of the five minutes creating the "master file" in Test 1. Once created, however, this link accelerates the calculation of cross-tabulations involving household characteristics.

3. Ease of programming

The ease of programming in two software packages is difficult to compare objectively and involves different aspects. This topic will be covered only briefly. For further insight, it may be useful to refer to Annex B which provides the SPSS and FOCUS programs used in the comparative tests.

Creating and modifying files is clearly simpler in SPSS than in FOCUS, because the latter involves hierarchical data and permanent links between related files. Although facilitating the production of reports, these features complicate the design, creation, and maintenance of FOCUS files. FOCUS does include a menu-driven interactive module for defining file structure (FILETALK), a built-in data entry program (without data validation), and a process for changing the structure of existing FOCUS files (REBUILD). These features alleviate but do not completely eliminate the problems associated with the complexity of FOCUS files. On the other hand, although more difficult, the data-entry facility in FOCUS allows for a wider range of data validation options, particularly regarding cross-checking among files.

With regard to the syntax for generating tables, each package has its strengths. SPSS may be easier for the creation of standard descriptive statistics for one (or a series) of variables; FOCUS requires more commands to obtain frequency distributions, standard deviations, ranges, etc. and obtaining them for a series of variables requires more steps. In addition, SPSS does simple cross-tabulation with row and column-percentages more easily. On the other hand, FOCUS seems more powerful in producing complex tables and in specifying the format of them. In addition, FOCUS does automatically some intermediate steps for which SPSS depends on programmer instructions, for example sorting and determination of column width. Finally, FOCUS commands are closer to natural English and appear to involve more flexible syntax. For example, some commands can be expressed in different ways depending on the preference of the programmer. Both the following are acceptable:

```
IF QUANT GE 1 AND QUANT LE 5 OR QUANT EQ 10
IF QUANT FROM 1 TO 5 OR QUANT IS 10
```

The identification of errors in SPSS is facilitated by the fact that a command-by-command record of operations appear on the screen and are stored for later consultation. FOCUS procedures in the course of running a program are not as well documented, although the error messages are quite specific and detailed.

4. Summary and conclusions

It is clear that FOCUS is faster than SPSS. With the exception of creating an ASCII file, FOCUS was faster than SPSS in every test performed here. For ordinary tables involving variables all from the same file, FOCUS was two to three times faster. However, the real strength of FOCUS lies in dealing with tasks involving variables from different files. Its ability to establish permanent and temporary links between files appears to be considerably faster than the JOIN MATCH operation in SPSS. This appears to be due to the fact that JOIN MATCH must create a

new file combining all the variables to be used and it must sort the file in order of the key variable.

In terms of file size, there does not appear to be that much difference between SPSS and FOCUS. An uncompressed SPSS file may be several times as large as an equivalent FOCUS file, but once compressed, the SPSS file is only 20% larger.

With regard to the ease of programming, SPSS involves fewer commands for basic descriptive statistics and simple cross-tabulation, but seems less flexible in producing complex reports and allowing a wide range of formatting options. FOCUS seems to require less specification of intermediate steps and the commands seem closer to natural language. And finally, FOCUS handles data from different files and at different levels more easily, but at a cost in terms of the ease of file creation and management. And finally, SPSS has an edge in the ease of debugging as a result of its description of procedures during the running of the program.

ANNEX H: INVENTORY OF IBM-COMPATIBLE COMPUTERS AT MINIPLAN

- IBM AT with 1 hard disk and 1 disk drive; provided by D.G. de Planification; used for data entry for Enquete Nationale sur l'Emploi
- IBM AT with 2 hard disks and 1 disk drive; provided by USAID; used for processing of ENBC/Urban budget data
- IBM AT with 2 hard disks and 1 disk drive; provided by D.G. de l'Informatique; used for programming for ENBC/Urban and word processing
- IBM AT with 2 hard disk and 1 disk drive; provided by the budget for the Enquete Nationale sur l'Emploi; used for word processing
- Wang AT equivalent with 1 hard disk and 1 disk drive; provided by UNICEF; used for processing of ENBC anthropometric data
- Micral AT equivalent with 2 hard disks and 1 disk drive; provided by the Cooperation Francaise; used for data entry for ENBC/Urban and word processing
- IBM PS/30 with 1 hard disk and 3.5" disk drive; provided by USAID; used for calculation of price indices using Lotus
- IBM PS/30 with 1 hard disk, 3.5" disk drive, and external 5.25" disk drive; provided by USAID; used for calculation of price indices using Lotus
- IBM XT with 2 disk drives and 2 external Bernoullie disks; provided by UN Fund for Population Activities; used for data entry of civil statistics
- IBM XT with 2 disk drives and 2 external Bernoullie disks; provided by UN Fund for Population Activities; used for data entry of ENBC/Urban data
- IBM XT with 2 disk drives and 2 (out of order) external Bernoullie disks; provided by UN Fund for Population Activities; used for data entry of civil statistics
- IBM XT with 1 hard disk and 1 disk drive; provided by the budget for the Enquete Nationale sur l'Emploi; used for data entry of National Employment Survey data

ANNEXE I: EVALUATION DES FICHIERS LAISSES PAR M. MULLER

M. Christophe Muller, assistant technique de la Coopération Française, est parti définitivement en octobre 1988 ayant terminé plus de trois ans sur l'Enquête Nationale sur le Budget et la Consommation. La Direction Générale de la Statistique m'a demandé d'évaluer les fichiers laissés par M. Muller afin de déterminer leur utilité.

M. Muller a supervisé les travaux de codification et de saisie des données et a participé à la préparation des Volume 1 (Méthodologie) et Volume 2 (résultats budgétaires provisoires) du milieu rural. En juillet 1987, la Direction Générale lui a assigné à l'analyse de la patrimoine, des prix, et des activités économiques en milieu rural. Cependant, il a continué à travailler sur l'approvisionnement en bois et en eau, à partir d'un dépouillement manuel des données. En mars 1988, il a commencé à traiter les données budgétaires, refaisant (avec des modifications) les travaux que NIYIBIZI Bonaventure et moi avions entrepris depuis 1986.

A ma connaissance M. Muller n'a laissé aucune documentation de ces travaux. Sans descriptions des fichiers ou au moins une liste des fichiers qu'il utilisait, il est difficile de les trouver et d'interpréter leur contenus. Sur l'ordinateur Micral 60, j'ai pu trouver trois sous-directoires créés par M. Muller:

C:çCOP	21 fichiers de texte
C:çENBCçPGMçESSAI	59 fichiers FOCUS
C:çCOSTWAL	0 fichier (vid)

Les 21 fichiers de texte représentent les documents préparés et distribués par M. Muller concernant l'approvisionnement en bois et en eau, les prix, la méthodologie en milieu urbain, et le résumé des résultats provisoires du Volume 2. Plusieurs fichiers ne contiennent que des morceaux de ces rapports.

Le sous-directoire ENBCçPGMçESSAI contient 59 fichiers, dont la répartition selon le type de fichier est comme suit:

16	fichiers de données en format FOCUS
33	fichiers des structures de fichiers de données
0	fichiers des programmes
3	fichiers des résultats des programmes
7	fichiers de divers types

D'abord, il vaut la peine de mentionner que le nombre total de fichiers est assez faible. Au moment d'effectuer le traitement des données budgétaires rurales, nous avons environ 300 fichiers. Le traitement des données en milieu urbain vient de commencer et on a déjà 150 fichiers.

Ensuite, la répartition des fichiers est curieux. Normalement, les programmes représentent environ 75% des fichiers. Ce sont les programmes qui effectuent l'exploration des données, la modification des fichiers, la confection des nouveaux fichiers, et la

confection des tableaux des résultats. Par conséquent, il est impossible de déterminer comment les fichiers de données ont été modifiés, quelles sont les définitions du système de classification, et quel type de traitement M. Muller effectuait. L'absence totale des programmes indique que les programmes ont été stockés ailleurs, par exemple sur les disquettes, ou que M. Muller les a amené avec lui lorsqu'il est parti.

En ce qui concerne les 16 fichiers de données, il y a sept fichiers de données brutes provenant des questionnaires et neuf fichiers qui sont des versions des fichiers confectionnés par d'autres personnes. Parmi les sept fichiers des données brutes, il y a des copies des données du questionnaire Q2 (budget rétrospectif), Q5 (consommation alimentaire), Q6 (patrimoine), et A1 (prix). Deux des sept fichiers sont des anciennes versions des fichiers n'ayant qu'un sous-échantillon des données. Le seul fichier dont je ne savais pas concerne l'approvisionnement en bois. Il est étonnant de ne pas trouver une copie des données du questionnaire Q4 (budget quotidien), le questionnaire budgétaire le plus important.

Les neuf fichiers de données qui restent sont des versions des fichiers confectionnés par Jim Otto (2 fichiers), Niyibizi Bonaventure (2 fichiers), et moi (5 fichiers). Parmi les neuf, il y a deux qui n'ont pas du tout été modifiés et un autre qui est une version obsolète d'un fichier que j'ai confectionné. Les autres ont été modifiés, mais il est difficile de déterminer de quelle façon. Il n'y a nulpart de nouvelles variables. Le seul changement visible est que M. Muller a ajouté des liens entre les fichiers d'une façon qui démontre une mauvaise connaissance de cet aspect de FOCUS (voir par exemple le fichier MENDATX).

Les fichiers des résultats indiquent l'étape de traitement. Malheureusement, il n'y a que trois fichiers de ce type. Un fichier contient un tableau confectionné avec environ sept commandes. Le deuxième tableau est presque identique. Le troisième est un tableau sommaire des données budgétaires. Il démontre un système de classification des dépenses avec 240 postes (contre 79 postes dans le système des Volumes 2, 3, et 4). Ce système de classification est tellement désaggréé qu'il comprend les postes comme "Instrument de musique", "Coiffure", "Moto", "Beignet", et "Réparation chaussure". En milieu rural, ces genres de transaction sont assez rares, et par conséquent, le nombre d'observation est très faible. Par exemple, les dépenses des motos sont basées sur une seule transaction en milieu rural. Il est aussi notable que le système de classification ne comprend pas le revenu. Enfin, le tableau indique que M. Muller n'était pas encore arrivé à l'étape d'intégrer l'autoconsommation aux budgets ruraux. Ceci veut dire que le traitement est au moins une année avant de pouvoir sortir des résultats définitifs.

En résumé, les fichiers sont totalement inutilisables pour les raisons suivantes: il n'y a aucune documentation, tous les programmes manquent, le système de classification est mal conçu, et les données sont à une étape très préliminaire dans le traitement. Le fichier sur l'approvisionnement en bois est le seul qui pourrait être utilisable.