

PN - AAP - 710

ISN 34179

The Dominican Republic Central Bank
Household Budget Survey, 1976-77:
Data Quality, Analytical Uses
and Resource Needs

A Report to the USDA/OICD Nutrition
Economics Group, under Contract
No. 53-319-R-9-164
Effective 5 July 1979

Philip Musgrove
3709 Northampton St., N.W.
Washington, D.C. 20015

27 July 1979

RSSA 3-77

Economic Analysis of Agricultural Policies

1. This report summarizes the findings of a two-week trip to the Dominican Republic, 8-21 July 1979, to evaluate the Central Bank's 1976-77 nationwide household budget survey and its potential use in studies of the Consumption Effects of Agricultural Policies (CEAP), in conjunction with the Ministry of Agricultural (SEA). Laurent Ross reports separately on computational aspects.

2. Most of our time was spent with the Bank's Division of Economic Studies, with Gumersindo Del Rosario, Antonio Rodríguez, and Gladys Santona. We also briefly visited Rubén Núñez of SEA's Department of Information, Statistics and Computing, and twice visited the USAID mission, once for briefing by Felipe Manteiga and once for de-briefing with him, Cam Wickham, Henry Wellhouse and John Clary. At the Bank's invitation I gave a seminar on the 18th describing our work to an audience from the Bank, SEA, the National Offices of Statistics (ONE) and Planning (ONAPLAN) and other agencies.

3. Being "captured" by the Bank meant I worked for them on some questions, particularly about price indexes, only distantly related to CEAP. This possibly lost time was more than compensated by the Bank's increased confidence in us and willingness to discuss difficulties. Since the Bank has in the past received visits from several "experts" who were apparently incompetent and did not gain that confidence, this is an important consideration for future technical assistance. It helped that we answered as many questions as we asked; some of those answers form a series of documents attached as an Appendix to this report.

4. The principal conclusions are quickly summarized:

(1) The data are of reasonable quality and should in principle serve for the kind of analyses needed for CEAP studies: there are no serious limitations in sample design or scope of data.

(2) Considerable data-clearing is desirable before analysis, since manual checking left numerous errors. The first stage of machine-cleaning is underway, and others should be defined.

(3) Documentation is good up to the receipt of questionnaires, and poor thereafter. It is essential to document subsequent stages while memories and notes make this still possible.

(4) Bank staff have tried some preliminary analyses, with sometimes acceptable results. Where the results are implausible, the fault lies less in data deficiencies than in the lack of economic and statistical knowledge. This analytical limitation is where assistance is most needed.

(5) There seem to be no institutional obstacles to Bank-SEA collaboration, which both agencies profess to want. It is premature to transfer the data, but they can be used at the Bank with SEA's help until a clean master file is ready. Details remain to work out.

(6) The situation is better than we anticipated: the problems are mostly of analytic capacity rather than of data inadequacy or institutional resistance. The needed capacity can only be acquired slowly by Bank staff, but they want to learn; and they can use help from within the country (SEA) or outside.

5. The sample design modifies the Census Bureau's Atlantida scheme slightly, so as to reduce interviewing costs in rural areas while capturing urban heterogeneity. Proportionality (probability of 0.005 for all families) is maintained in the design by month, region (municipio) and zone (urban or rural): now weights need to be calculated and published to adjust for accidental undersampling (doc. 2, Appendix). It may then appear that some sample cells lost too many observations to be used for detailed analysis, although overall, the sample is large enough. The sample design is well documented; data users should be able to understand it and draw subsamples with only moderate effort. Documentation is needed for the sample actually obtained, however (doc. 1, Appendix).

6. Interviewers appear to have been well trained and supervised and provided with good written instructions. There seem to have been conscientious efforts to prevent cheating and to check the questionnaires for errors early enough that the family could be re-interviewed. The original questionnaires are still available to help in machine-checking. It is less easy to evaluate the coding, although four observations can be made: (1) there are obviously still errors in the data, some of which were probably introduced in coding and keypunching. These have turned up in machine consistency tests and early tabulations (see Laurent Ross' report). (2) The coding in some cases aggregates information slightly, reducing the number of possible values for a variable -- e.g., age is reported in months for infants, but coded only to the nearest year. No significant detail seems to have been lost in this process. (3) The data on brand or model of goods purchased has not been coded at all. For nutritional analysis this represents no loss since individual foods are well-distinguished: for some kinds of marketing

analysis, it might be worthwhile to code these variables and add them to the file. (4) Information on direct saving, and on changes in debts, has not been coded yet and is feared to be of poor quality. A complete saving balance is impossible even in principle since some variables are missing, but it is advisable to code what exists and make some tests of its reliability. This is perhaps important even for nutritional analysis, if families' access to credit and debt burden affect their consumption.

7. The questionnaires were consistency-checked, which served to catch some errors, chiefly among sociodemographic variables (age, education, etc.) or between foods purchased and foods reported eaten. The Bank initially believed no significant errors remained, but the first tabulations, by Dato-Centro, turned up many such errors still. A series of machine checks is therefore underway; these include tests for invalid values of single variables as well as two-variable consistencies. Because this need was seen already, and because Ross concentrated on preparing a codebook and other documentation, we did not run any checks ourselves. Priority should be given to reviewing the machine-checking, with subsequent definition of tests which are statistical rather than logical -- that is, tests for extreme values of financial variables. These can be a by-product of initial econometric analyses, identifying extremes by large residuals (see my paper on multi-variate tests, Appendix).

8. The preliminary judgment of data adequacy rests on what is known so far, and on the expectation that the errors found in testing either will be infrequent, or will have been introduced after interviewing, so they

can be corrected from the questionnaires. There is another kind of "error" which is easy and imperative to repair, and that is that food quantities are in two or more different units for a given item. A matrix of conversion factors has to be defined and applied to the data for standardization. (It is already known, from price index tabulations, which units were reported for each good, so there is a preliminary check of average unit prices which gives reasonable results.) I doubt that the implicit unit prices will serve to estimate price elasticities, although there may be large and systematic enough urban-rural or regional differences to permit such analyses. Estimates of price effects will probably have to come from exogenous price data or from constrained demand functions.

9. Given correction of errors, standardization of quantities, and an exogenous matrix of nutrient contents of foods, the data should serve for:

(1) Descriptive statistics of budget allocation, physical food consumption, and nutrient intake, with households classified by size and composition and by income (or related measures, or income determinants such as education and age) as well as by region and zone.

(2) Econometric estimates of expenditure, food consumption or nutrient intake as functions of the same classifying variables, either as single equations or in systems subject to constraints. Elasticities can be derived from such functions for simple calculations, or the entire functions can be used for projections.

It is premature to specify the statistics or functions to estimate, and in any case several different ones should be tried; there are no "right"

answers, and experimentation is needed. It is also too early to say which variables will be the most reliable or useful. Priority now has to go to getting the data ready.

10. Bank staff have tried three kinds of calculations so far. They are premature, given the state of the data, but they have been useful in revealing the Bank's interests and capacities. These are:

(1) Price index calculations: mean expenditures by item, mean quantities and unit prices by unit of measurement, and preliminary indexes for the country, all urban areas, and Santo Domingo alone. These calculations have highest Bank priority; they ought to be reviewed on clearer data, but would probably change very little. The price index methodology is under revision, offering the chance to clarify several issues (docs. 3 and 4, Appendix), particularly the possibility of making true indexes of spatial price differences (only spatial differences in inflation rates are currently considered).

(2) Consumption functions for 22 foodstuffs, estimated with a curve-fitting program providing little statistical information. We showed the Bank how to use their regression program, modified it to yield more statistics, and evaluated the consumption functions at length (doc. 5, Appendix). This experience illustrates dramatically that the Bank (i) wants to undertake analyses related to CEAP studies but (ii) lacks the analytical capacity to set up and evaluate econometric estimates. The deficiencies are economic, mathematical and statistical, so that a variety of assistance is required.

(3) Estimates of poverty lines, using a simple World Bank procedure (doc. 6, Appendix). I did not evaluate these estimates, but note

that they also demonstrate Bank interest in studies related to CEAP. The analytical difficulties are much smaller.

11. The four principal steps to be taken now are:

(1) Completion of data cleaning. The purely logical tests should be completed before any analysis, but the statistical cleaning may overlap with it. This phase will take at best several weeks and might require a few months, depending on how detailed the testing becomes and on the frequency and severity of errors found. The chief sort of assistance required is of two sorts -- installing some easier programs, offered by SEA, than those now in use; and evaluating the statistical tests. A Programmer's help is more important than an analysts' here.

(2) Creation of a single, standardized file of uniform length and definition (see Ross' report). Until this is done, it will be very difficult to undertake work with the individual observations rather than aggregates. Except for an analysts' advice on the transformed variables (totals, averages, binary variables, etc.) to include in the file, all the help needed here is a programmer's. (The Bank can already, it should be noted, manipulate the different files in which the data are stored, but this is cumbersome and not well documented.) This is a job of a week or so.

(3) Producing and publishing a set of basic tabulations and descriptive statistics. One group of such results, a set of frequency tables, is already defined; it remains to define budget structures and other sorts of tables. This phase should take only a few weeks and require little if any outside assistance, although it would help to have an economist review any analysis based on the tables.

(4) Analysis, including econometric estimates, of consumption functions and other results relevant to CEAP (along, of course, with other studies of interest to the Bank or other users). This phase is open-ended, and should wait on a first review of the summary and descriptive tables. An analysts' help will be essential, perhaps full-time over some months.

At none of these stages do we anticipate serious difficulties of computer time or capacity: the scarce resource is always human, although the required skills differ.

12. The Bank, SEA and USAID mission all express interest in this work, willingness to devote resources to it, and willingness to collaborate; all say that the atmosphere for cooperation is much better than it was six months or more ago. The Bank proclaims an "open door" for data use, already supposedly in effect for some kinds of data needed by ONE. To turn these expressions into results, one must take account of the following specific resources, interests and limitations:

(1) Del Rosario will want to participate in everything done with the data, and should do so for some time to come. Given his and his colleagues' limitations, he needs technical help which the Bank does not seem likely to provide through more staff; their institutional interest is concentrated on the price indexes, not, for the moment, the more sophisticated analyses. However, the Bank is very receptive to getting such help.

(2) Núñez of SEA has specifically offered, once the data are ready, to try to hire an analyst who would work with (and perhaps at) the Bank, all analyses to be considered joint products of the two agencies. He himself

could do the job, but is unlikely to take the required time away from his duties running the SEA computer center, although he would presumably review the analyst's work. How much outside help will be needed depends greatly on whether Núñez can hire such a person and dedicate him full-time to work with the Bank. To avoid personal or institutional friction, the person also needs a deferential, diplomatic manner that may seem incogruous with the high analytic skills needed. Our position as outsiders, and the mission's selling of our help, gave us advantages a person from SEA or another agency will have to duplicate for himself.

(3) The mission's contribution is unlikely to be in the form of a person's time, despite Manteiga's great interest; they are all too busy. The mission can however help format collaboration, continue to lend moral support, and of course help any future technical assistance visit. It is also worth remarking a skepticism expressed by Clary, of the value of paying for any more survey work when most data are never adequately analyzed: here is a survey that can be analyzed at modest cost, offering a surer return on AID money and effort. This does not mean weakening relations with SEA, or trying to build the same relation with the Bank (which would probably fail), but it does suggest a shift of interest from gathering numbers to studying them.

13. If SEA provides a competent analyst to work full-time for a year or so with the Bank, there will probably be no need for a resident advisor to the project. It would still be desirable for mission staff to know more about the project and to recruit visitors (TDY) to help with particular problems. What these will be will depend on the skills available locally.

I suspect outside review would be especially valuable for (i) evaluating the master data file after cleaning and consolidating and documenting its virtues and weaknesses, and (ii) setting up and evaluating groups of consumption functions or other econometric estimates. Both of these are fairly discrete interventions that could be accomplished in two or three weeks. What should definitely be avoided are very quick visits that leave no time for questions to emerge; it is also desirable not to use many different people, so as to maintain good working associations and minimize learning costs.

14. If an adequate person cannot be found to be hired by SEA (or possibly the Bank), then a resident advisor makes sense just because there is so much work to do over about the next year. This will be a good solution, however, only if the person can cover all the different skills required; if instead he would have to rely on many TDY visits by experts in specific matters, it would be simpler to have Manteiga coordinate their visits and oversee the project in between. Using TDY help only is feasible, but may require many visits. Discussion of the costs and benefits of different approaches with the mission seems essential. One approach mentioned by Manteiga is to sponsor a series of workshops or seminars to discuss problems and present results; these could coincide with TDY visits, and they would be a good way to draw in other data users as well as a useful discipline for people involved in the project.

15. The horizon for this work should be about one year off -- mid- to late 1980. That means technical assistance should be planned over that period, and long-term Bank-SEA collaboration promoted. Some results can be

available much earlier, but econometric work useful to CEAP may wait until late this year or early 1980. Considering how long projects of this sort usually take, and the great waste of resources if data are gathered but not adequately analyzed, that is really an optimistic prognosis.