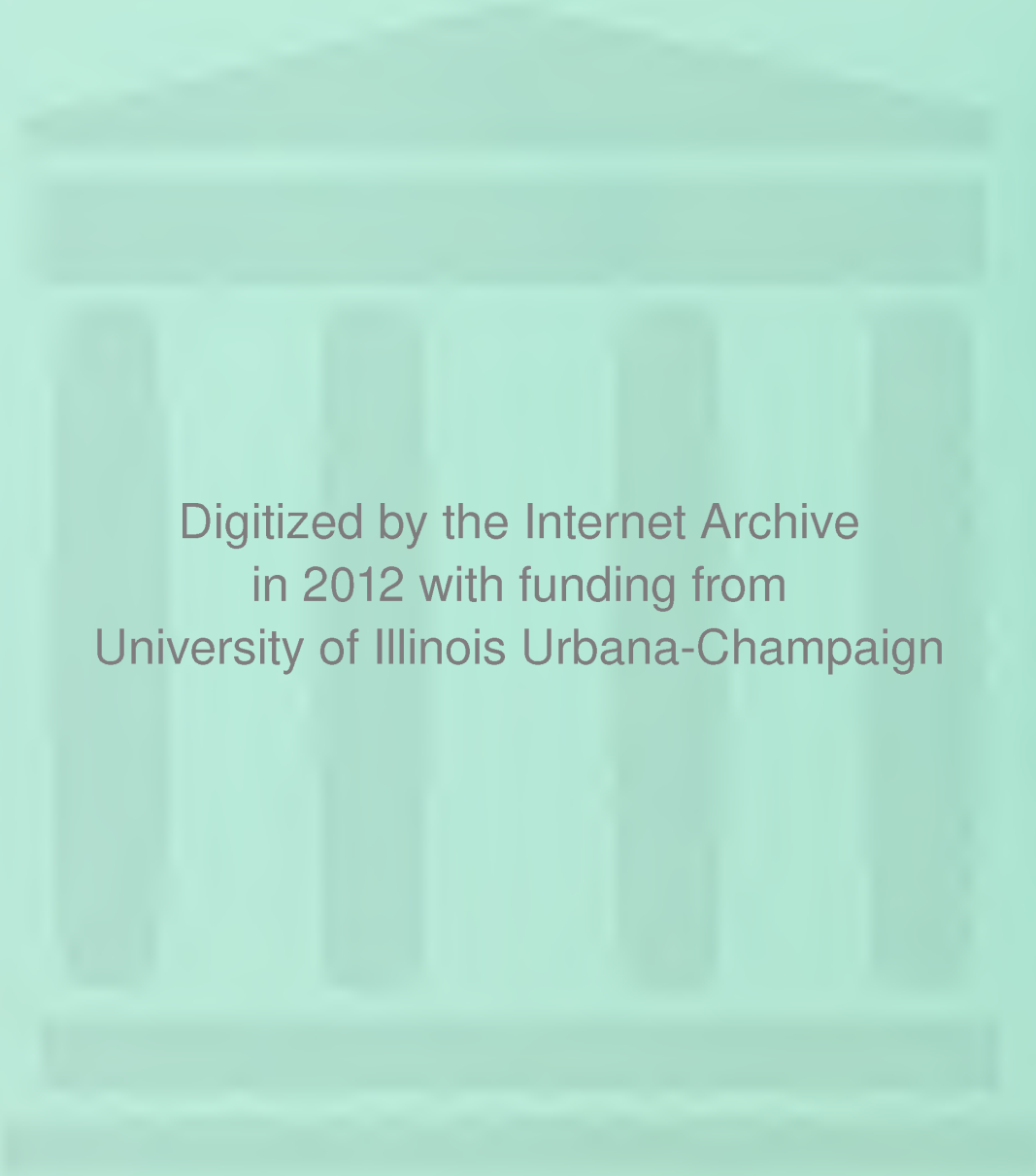




UNIVERSITY OF
ILLINOIS LIBRARY
AT URBANA-CHAMPAIGN
BOOKSTACKS



Digitized by the Internet Archive
in 2012 with funding from
University of Illinois Urbana-Champaign

<http://www.archive.org/details/problemsincollec137ferb>

Faculty Working Papers

PROBLEMS IN THE COLLECTION OF MICRO DATA FOR
INTERNATIONAL COMPARISONS

Robert Ferber
University of Illinois

Janes Angelo de Souza
Vargas Foundation

#137

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

December 17, 1973

PROBLEMS IN THE COLLECTION OF MICRO DATA FOR
INTERNATIONAL COMPARISONS

Robert Ferber
University of Illinois

Janes Angelo de Souza
Vargas Foundation

#137

PROBLEMS IN THE COLLECTION OF MICRO DATA FOR
INTERNATIONAL COMPARISONS

(Prepared for ECIEL International Conference on
Consumption, Income and Prices, Hamburg, West
Germany, October 1973)

Robert Ferber
University of Illinois

Janes Angelo de Souza
Vargas Foundation

September 1973

This paper discusses various facets of the collection of microeconomic data in less developed countries. The focus is on Latin America, the area where both authors have had experience in the collection of such data.

The paper begins with a discussion of the nature of micro data and the role they can play, and are playing, in planning and decision-making for economic development. Means of collecting micro data are covered in the second section, with attention to the pros and cons of the growing resort to administrative records as a source for such data. To illustrate the sort of problems that may be encountered in micro data collection, case studies are presented in the third section of such efforts involving data collection procedures in three Latin American countries--Brazil, Colombia and Paraguay.

Based on the foregoing material, a number of problem areas relating to international comparisons of micro data are brought out in the fourth and final section, which attempts to synthesize the results of the earlier sections and considers the implications of these findings for future data collection and data analysis.

1. Why Micro Data?

Micro data relate to the attitudes and behavior of the units of an economy, principally households and business firms. The specific unit will vary with the purpose of the study. For the consumer sector it may be the household, the family, the individual or even something else. For the business sector, it is usually the firm or the individual plant. The concept of micro units may also apply to the government sector, such as the study of individual agencies of a provincial or federal government, though here the distinction between micro and macro data begins to blur.

The role of micro data in economic analysis is twofold: To explain trends or variations in corresponding macro data, and to obtain new series and insights into economic behavior that can be obtained in no other way. Thus, by its very nature a macro series is made up of diverse components affected in many different ways by economic and political events.

The only way of understanding reasons for change is to look at these components individually and examine their behavior, which in turn often involves interviewing individual economic units. Thus, an increase in aggregate consumer expenditures may be accounted for by a sharp increase in purchase of particular durable goods. In turn, however, interviews with individual households may reveal that this sharp increase was brought about by the expectations of rising prices for durables in the near future. From the point of view of government policy, information of the latter type is of crucial value in forecasting expenditure trends and for any decisions that may be desired with regard to government policy for dealing with this problem.

That micro data can be used to obtain new insights into economic behavior, and only in this manner, is illustrated by the data that have been collected for some years in a number of countries on business plans for capital expenditures and for inventory investment. A very good example is the data on this subject collected regularly by the IFO Institute in Munich, which has been shown to provide vital information both for econometric forecasting and for understanding economic trends in West Germany. Such data are

compiled from samples of business firms and are most useful when analyzed by individual industries and even by sub-industries.

Consumer attitudes, based on surveys of individual consumer units, are an analog in the consumer sector, and they too have been shown in a number of countries to have analytical and predictive value in explaining economic developments.* Moreover, information obtained at the same time on reasons for particular types of attitudes and behavior can be invaluable in providing a better understanding of how people react to different economic policies and the likely effects of such policies on future behavior.

A key set of micro data of crucial importance in the study of economic development is the consumption and income patterns of private households. Such data may serve on the one hand to provide more reliable estimates of the national accounts of a country and, on the other hand, to indicate how consumption patterns differ by income levels and by other key socioeconomic factors. The estimates of income elasticities and other parameters thereby obtained are key inputs into the preparation of any econometric forecasting model or for planning economic development. The income data and savings data that may be obtained from these surveys at the same time provide basic information on the income distribution of the country, on sources of income

*See for example, Mueller, Eva, "Ten Years of Consumer Attitude Surveys: Their Forecasting Record," Journal of the American Statistical Association, 58, December 1963, 899-917; Poser, G. and Heckeltjen, P., "The Use of Anticipatory Data in a Quarterly Econometric Model of the Federal Republic of Germany's Economy," Darmstadt, West Germany: Technische Hochschule, unpublished manuscript, 1973; Shapiro, H.T., and Angevine, G.E., "Consumer Attitudes and Buying Expenditures -- An Analysis of the Canadian Data," Journal of Canadian Economics, 2, May 1969, 230-49.

by types of consumers, and on the availability of consumer resources for saving and investment.

For analytical purposes, micro data possess a degree of richness rarely possible with macro data. Because they relate to individual units of economic activity and because they are frequently available in the form of components, e.g., average expenditures per family for different types of families by type of expenditure, the data can be rearranged in many different ways to suit the needs of the particular study. Analyses can be carried out on different levels, since micro data can be combined to form aggregates, while the reverse is not true. By this means it is easily ascertained, for example, whether an increase in the price of beef will reduce demand for beef more or less uniformly in all groups of the population or only among the lower income groups.

The greater flexibility of micro data also makes international comparisons more feasible since concepts on the same topic from different countries are more easily adjusted to each other. Even so, international comparisons are by no means easy and involve a number of pitfalls, as discussed in Section 4 of this paper.

The big problem in using micro data is to collect them in the first place, and for this reason the major part of this paper focuses on this topic. If the data can be collected with adequate care and reliability, and at reasonable cost, they will not only be more readily available but their analysis and interpretation will be greatly facilitated. Hence, we proceed in the next section to discuss various means of collecting such data, to be supplemented by some case histories in the following section.

2. Ways of Collecting Micro Data

Until World War II, surveying of statistical data in less developed countries was restricted to carrying out periodically censuses and to estimating, almost always through subjective methods, global statistical data such as the value of agricultural and industrial production and volume of the output of certain commodities. Managers, statisticians and socio-economic analysts were content with having global statistical data concerning the country and, at most, with global information concerning certain areas or regions.* General disinterest characterized this work, since government participation in economic activity was minimal and uses of such data for private purposes were virtually nonexistent. The quality of the collected data was also undependable, since there was no training of qualified personnel and the statistician, in general, was "merely an unqualified public functionary charged with the task of compiling information on the activities carried out by the State and arranging such information in tabular form."**

In the last 30 years the situation has changed dramatically. Government intervention in economic activity has been growing, and the last twenty years the great majority of less developed countries has been actively engaged in preparing Plans of Economic Development to expand production and raise the standard of living as rapidly as possible.

In the light of this new situation, general global data which had been hitherto collected and made public were not enough for the researchers, who increasingly were seeking more precise and more specific data. This was the inception of the great development in processes and techniques for

*A.J. Jaffe, "Notes on Developing Countries and their Statistics," Review of Income and Wealth, September 1972, pp. 313-326.

**Lourival Camara, as quoted in "Governo da Paraíba, Estatística e Desenvolvimento", Paper presented at Ia. Conferência Nacional de Estatística, Rio de Janeiro, 1968.

collecting micro data which has taken place in recent years.

Micro data collection may be carried out by means of the so-called "cross-section" method, i.e., obtaining information over decision micro units at a certain point in time, or it may be by means of continuous collection over successive points of time. It may be in the form of a census covering the whole universe of micro units; it may be obtained through sampling; or it may refer to case studies. Data can be collected directly for specific statistical purposes, or be derived from information supplied with other purposes in view, such as tax files and administrative records.

The principal forms of micro-data collection now in use in Latin America are the following:

1) Direct collection from micro units

a) Related to individual or families

Demographic Censuses

National research projects by samplings of households

Research projects on family budgets

Special research projects of qualitative character such as on consumer intentions and attitudes

b) Related to firms or individual plants

Economic Censuses

Periodic research projects through sampling on production, employment, utilized capital, etc.

Periodic research on costs, technological levels, production processes, etc.

Business tests on entrepreneurial expectations

*H.W. Watts, "Microdata: Lessons from the SEO and the graduated work incentive Experiment", Annals of Economic and Social Measurement, April 1972, pp. 183-191.

- 2) Data collected for administrative purposes
 - a) Related to individuals or families
 - Data from income tax return forms
 - Data from social security records
 - Data from migration control services
 - Data from military enlistment and voting records
 - Data from birth registration records
 - Data from school rosters
 - b) Related to enterprises or firms or industrial plants
 - Data from income tax return forms
 - Data from industrial production tax, sales tax, etc.
 - Data from corporation balance-sheets
 - Data extracted from import and export licenses

For a better understanding of these types of data collection, each is now considered separately with some attention to historical aspects of their evolution. The principal such types are censuses, surveys and use of administrative records.

A. Censuses

Contrary to other statistical surveys at the micro level, Demographic Censuses have been carried out since the earliest times of civilization.* The first censuses, however, consisted merely in head-counting the population with limited objectives in view, usually collection of taxes and establishment of voting lists.

The first relatively complex census seems to have been carried out in Babylon, in 3000 B.C., consisting of a rather complete survey of farms and farming activities, for fiscal purposes.**

*Isaac Kerstenetzky, Tudo sobre o Recenseamento, Bloch Editores S.A., Rio de Janeiro, GB, Brasil, 1972.

**Kingsley Davis in Encyclopaedia Britannica, 1966, vol. 5.

Two censuses for military purposes are recorded in the Old Testament: one during the Exodus from Egypt, and another one in King David's time; the New Testament records a census of the population at about the time of the birth of Christ.

In Rome, starting in 44 B.C., a census was taken every five years of citizens and real property for fiscal purposes and for assembly voting. This practice was extended to all the Empire in 5 B.C., and was carried on until the downfall of Rome.

During the Middle Ages the most important census was one carried out in 1086, in England under William the Conqueror, to supply information on land tenure in his new domains.

Thus, it is evident that early statistical surveys which were carried out in the world had purely administrative purposes, and the first micro-data ever recorded were derived from administrative records.

The development of the modern idea of a population census as a complete enumeration of all persons and their main socio-economic characteristics occurred during the XVI and XVII centuries. Canada, between 1665 and 1754, made the first attempt at counting the population in a more extensive area than a city and at regular intervals.

The first national censuses were carried out in the United States, in 1790, and in England, in 1801. In Latin America the first censuses were implemented after the second half of the XIX century. Nevertheless, until World War I such censuses comprised only certain aspects related to persons. Starting in 1920 the scope for research began to expand, to comprise economic activities (Economic Censuses).

Even though a census supplies ample data regarding people, family units, business enterprises, firms and political units, as well as a general picture of the economic and social structure of a country, it presents two basic problems. First, because of the high cost involved, a census can only be carried out between long intervals of time (generally ten years); and second, in view of the volume and diversity of the information to be collected the census has to be rather general and must omit details which are often essential for the economic analyst.

B. Surveys

Inasmuch as the information supplied by a Demographic Census only becomes available every 10 years, and in view of the extensive and complex data sought, it may come out with a minimum two to three year lag. As a result, there came into being sometime in the 60s the so-called national sampling home surveys.

This type of investigation, which may be repeated quarterly or half-yearly, can supply rapid information on basic characteristics of the population and living conditions, such as birth rate, death rate, fertility, internal migration, manpower changes, employment, unemployment, education and health levels, standard of living and similar parameters*. These statistics have become of such importance that collecting such data now constitutes a project in which all Latin American countries take part -- the "Atlantida Project", set up on the basis of a methodological study prepared by experts of the U.S. Bureau of the Census in 1967.**

Surveys of this type have been carried out in the following countries in or near Latin America:***Argentina (1963), Barbados (1965), Brazil (1967), Chile (1966), Colombia (1970), Costa Rica (1966), Ecuador (1968), Jamaica

*J.W. Brackett and R. H. Reed, "Uso Analitico de Datos Obtenidos em Encuestas de Hogares", Paper presented at Second Symposium on Household Sample Surveys in Latin America, Rio de Janeiro, 1970.

**U.S. Bureau of the Census, Atlantida: A Case Study of Household Sample Surveys, Washington, D.C. 1967.

***Fundacao Instituto Brasileiro de Geografia e Estatistica, Provisional Report on the "Second Symposium on Household Surveys in Latin America, Rio de Janeiro, 1970.

(1968), Panama (1963), Peru (1968), Dominican Republic (1969), Trinidad and Tobago (1963), Uruguay (1967), Venezuela (1966).

Another way of collecting micro data which has been assuming increasing importance is through surveys on family budgets. Early investigations of family budgets were started at the turn of the century, aimed at supplying data to determine the amount of expenditure required by families of various size to survive and maintain working efficiency.* Among the earlier investigations of family budgets in modern times special attention must be given to the surveys of family budgets of urban workers in England in 1904, to the surveys on food consumption of the working classes in the United States in 1901, and to the broad survey on family budgets of salaried employees carried out in 1918-1919 in the United States.

With the advent of inflationary problems in several countries after World War I, the objective of surveys on family budgets developed mainly toward collection of basic data for constructing weights for cost of living indexes. However, with the development of national economic analysis in the last 30 years, the demand for data resulting from investigations into family budgets increased considerably, not only in regard to cost of living problems but also in order to obtain essential data for studying such socio-economic problems as:

- Structure of individual and family expenditures and consumption
- Distribution of income
- Measure and make-up of private consumption (which represents over 70% of gross internal expenditure in most countries)
- Nature and distribution of private savings

*Andrew Donald Roy in Encyclopaedia Britannica, 1966, Vol. 6.

- Estimates of demand functions and projections thereof
- Evaluation of nutrition and identification of population sectors presenting nutritional deficits
- International comparisons of income levels, purchasing power and real exchange rates

There are three methods which are usually employed in collecting micro-data on family budgets, all three of them having recourse to sampling.

a. The memory method, which entails having an interviewer fill out a questionnaire on family expenditures on the basis of recall over periods varying from a week to a year. It has the advantage of greater ease in obtaining answers to queries and lower cost, and permits much wider application than the other methods. However, the results obtained may be subject to large margins of error principally on account of the memory factor.

b. Filling out family diaries, carried out by the family unit itself. It is a difficult method to apply in less developed countries since it requires a minimum level of education of the participating families. Also, sample mortality tends to be high if carried on for some period of time.

c. Recording of expenditures daily and direct weighing of foodstuffs by a specially trained interviewer. Results obtained are likely to be more accurate but the method is very costly and difficult to implement, which circumscribes its application to a small number of families.

Surveys on family budgets have been carried out in Latin America for a number of years. For example, in Brazil studies of this type have been carried out by different agencies since 1936. However, only in the second half of the decade of the 60s were they systematically set up and

made comparable in all member countries of LAFTA (South America and Mexico) and in some Central American countries, under the sponsorship of the ECIEL Program, coordinated by the Brookings Institution, of Washington, D.C.

In regard to qualitative surveys on consumer behavior, even though such surveys are well developed in certain countries such as the United States (Survey Research Center, University of Michigan) and West Germany (IFO Institut), this type of survey is not yet current in Latin America.

Regarding the entrepreneurial sector, the collection of micro data, except for Census data (as mentioned earlier), has been carried out for a longer time and with a larger wealth of detail than family surveys. Surveys of industrial and business enterprises have been conducted annually and sometimes monthly in several Latin American countries since the 40s covering such data as type of activity, number of employees, volume and value of production, salaries paid, power consumed (fuel, electricity, and others). Only in the agricultural sector have less data been collected directly from producing units, inasmuch as researchers in Latin America had previously concentrated attention on industrial development which for many years was thought to be the only sector able to accelerate the developing process of the Latin American countries. On the other hand, direct surveys of micro units in the agricultural sector are much more difficult in view of the vast distances between units.

The collection of qualitative micro data to forecast entrepreneurial behavior (business tests) has been developing for about eight years in various Latin American countries, chiefly Argentina and Brazil. This type of investigation is becoming more frequent inasmuch as it is simple to implement and tabulation of results can be rapidly performed.

C. Administrative Records

Just as micro data may be collected directly from micro units for specific statistical purposes, they can also be withdrawn from administrative records, i.e., from the registers which are kept by public and sometimes by private agencies engaged in activities such as listing persons and firms for a determined objective (social security, polls, military enlistment, birth and death records, school rosters and others), or tax returns (income tax, sales tax, as well as others).

As indicated previously, the world's early collections of micro data were much more from administrative records than from statistical surveys. Even in recent times, during the period which antedated the ideas of economic development and planning after World War II, researchers in several Latin American countries had to have recourse to the information contained in the administrative records existing in several Government departments, since such countries did not have organized statistical services. For a number of years that information was relegated to a secondary role, but in recent times it has come up as a most interesting source of data for measuring important socioeconomic aspects. As a result, there is the problem of having to choose between collecting micro data by direct surveys of the relevant populations or by administrative records. Both systems present advantages and disadvantages.

The main advantages of the direct surveys of micro units is that data collected for specific purposes directly, supply the kind of information which is as precise and as broad as the researcher wants. Nevertheless, three types of disadvantages are attached to this approach. First, it

entails high costs, and very often - particularly in low income countries - it becomes difficult to obtain the required funds. Governments with limited resources or with resources in short supply are generally more interested in investments aimed at improving their tax collections than in spending part of their resources in statistical surveys which may throw more light on their country's economy.

The second disadvantage is related to the time for collecting and tabulating the data. Despite modern electronic systems for data processing, generally a long period of time must be spent in the collecting process and in the interpretation and processing of large volumes of micro data collected through direct research, so that when the information finally becomes available it may be out of date.

The third disadvantage is the possible low quality of the collected information. We are not aware of legislation in any Latin American country imposing any legal penalty for supplying erroneous or inaccurate information sought for purely statistical purposes, and since the average cultural level of the population of these countries is low and the informers are unaware of how useful the information can be, quite frequently they make no effort to supply trustworthy data. For example, in some tests carried out by one of the authors, on data from industrial and agricultural enterprises on sales value, salaries paid, pending orders, investment volume, etc., in few cases was he able to convince the reporting firms to allow consultation of their books and records. Their main concern was to spend as little time as possible in the interview, and to supply the information from memory. A later check of some of this information showed it to be totally erroneous.

Micro data from administrative records also presents advantages and disadvantages. The first advantage is its low cost and the speed with which the data can be obtained, having already been regularly collected and processed for other purposes. The additional cost involved usually represents only the cost of preparing special tables to better serve the objectives of the research work. A second advantage resides in the relative assurance of such original data (at least in some cases), since legal sanctions which sometimes can be rather drastic may be imposed for false or mistaken reporting, and the informants are much more careful in filling out the return forms.

The disadvantages, however, are of a different nature. The first is that data collected for other purposes can hardly serve to supply sufficient information for a particular socio-economic analysis. For example, studies on income distribution based on data in income tax files would have to leave out that part of the population with low incomes which is therefore tax exempt. At the other extreme, the data available refer to "taxable income" and not to "total income", thus omitting a large portion of capital income not subject to tax, and non-monetary income. As another example, data obtained from social security records omit workers who are not contributors or who receive benefits from Social Security, and in the underdeveloped countries the number of such workers makes up almost the total number of unemployed, self-employed, and farm workers.

Another disadvantage that may arise is intentionally distorted data. Since many such surveys are directly or indirectly related to tax collection, there may be a deliberate bias in supplying the information with the intent of diminishing the amount of tax to be paid.

Thus, in order to make them usable, data from administrative records must be "adapted" to meet the requirements of the research project for which it is needed. The simplest form consists in combining the primary data in a different fashion than that followed for nonstatistical purposes. A second form would be to supplement the information by adding data from other sources. A third form would be to convince those responsible for collecting the data to adapt their questionnaires, records and files in such a manner as to take care of both administrative and statistical objectives. However, this is not an easy task because it is often difficult to convince, say, a tax collector to worry over (and to spend money and effort on) problems such as the distribution of income or the propensity of the population to save.

Any means of collecting micro data presents advantages and disadvantages. No simple body of data, even though obtained from direct investigation, is usually capable of furnishing all the data required by the researcher, except for restricted projects and very limited objectives. For this reason, one need not consider collection of micro data as a conglomerate obtained by one single investigation following one single method.

Obviously, when the objectives which the analyst has set forth may be attained with a single body of data, this facilitates enormously the task of collecting and processing the data. Nevertheless, when the objective is a thorough exploration of variations in aggregates of economic activity by identifying changes at the micro level, the only way is to correlate fragments of information from different sources concerning economic units observed at a certain point in time and to continue these observations at

different periods of time.*

But how can these data be correlated, if the methods for linking the different files of micro data are still primary and rudimentary? According to Benjamin Okner** who voices the general opinion of the experts, few processes are generally accepted for determining the validity of the connecting processes. There is no objective method for deciding if any method is "unsatisfactory", "good" or "best". The only valid way to pass judgment on the appropriateness of a synthetic file of micro data, derived from several sources, consists in appraising if it meets the requirements of the research for which it was set up.

3. Case Studies for Individual Countries

To gain a better insight on how micro data have been collected in the less developed countries and the sort of problems encountered, it seems instructive to review some actual studies of this type.

Three types of case studies are presented in this section. Two relate to data collection by surveys of human populations. One of these is a survey of consumption expenditures and incomes, where both the planned and actual procedures differed substantially from one country to another even though international comparisons were the primary objective. To illustrate the nature of these differences, the manner in which this study was carried out in three different countries is described--Brazil, Colombia and Paraguay. The second is a price survey where identical plans and procedures could be employed in all countries. For these reasons, a general description of this study is provided which applies with minor variations to all the countries involved.

*Isaac Kerstenetzky, op. cit.

**Benjamin A. Okner, "Reply and Comments" on "Constructing a New Data Base From Existing Micro-Data Sets: The 1966 Merge File", Annals of Economic and Social Measurement, July 1972, pp. 359-362.

Some specific problems encountered in this study are noted with reference to Brazil.

The third type of case study refers to a much more recent form of data collection, namely, the use of administrative records. This form of data collection is still in its infancy, both in developed as well as less developed countries, and not much experience has accumulated with it. One such experience is recounted here, with reference to Brazil.*

In each case a more or less standard format is followed. After a brief description of the background and purpose of the particular study, the presentation focuses on the ex ante design of the data collection phase, how the design in practice conformed with the ex ante plan, reasons for divergence, and an evaluation of the efficiency of the operation and, in a very general sense, of the reliability of the data. More general implications, both for data collection and for analysis, are left for the following sections.

A. The Consumption Study

Under the sponsorship of ECIEL, plans were made in 1966 and 1967 for a survey of consumption expenditures and incomes in the major urban areas of the LAFTA countries (the ten Spanish and Portuguese speaking countries of South America, plus Mexico). To be covered were private households in areas with a population of 500,000 or more. If a country had many such areas, as in the case of Brazil, a purposive representative sample of such areas was to be chosen.

The primary objective of the study was to make international comparisons of consumption and income patterns. At the same time, most of the countries were also interested in using the data for internal purposes, and since many of them had made similar studies of one sort or another in the past, they

*This case study is not complete and is not included in this draft.

were influenced to use the same or a similar format on this study for the sake of comparability. Partly for this reason and partly because of inherent differences in consumption standards and seasonal patterns among countries so far apart, it was not possible to use fully standardized procedures among the countries on this study. The type of differences that arose is illustrated in the following presentation on how the study was carried out in three of the countries.

(i) Brazil

The survey in Brazil was carried out by the Centro de Estatística e Econometria do Instituto Brasileiro de Economia of the Getulio Vargas Foundation. The survey was designed to permit data collection in four quarterly stages, comprising three different samples, as follows:

- a) families to be interviewed in 4 successive quarters,
- b) families to be interviewed in only two quarters,
- c) families that would supply information during only one quarter.

There were three reasons for selecting this sampling scheme: first, for financial reasons, to minimize the cost of field operations; second, for substantive reasons, to obtain data on the seasonal variations in the consumption of certain items and services throughout the year of the survey; and third, for methodological reasons, to have a means for assessing the effects of panel mortality and panel conditioning.

The size of the sample was established a priori, based on budgetary availabilities and previous experience of the Institute in 1962 in this type of research, at 1,000 families in Rio de Janeiro, 700 families in Pôrto Alegre and 700 families in Recife. These were the only three cities included in the study from Brazil, again for budgetary reasons. In each city stratification was sought by income and by size of family.

The first problem in sample selection was setting up a frame for the population component of the universe to be studied. The most recent population census had been carried out 7 years prior, in 1960, and was completely outdated. More recently, in 1965, a school census was taken by the State Secretariat of Education to evaluate requirements for expanding public primary education. No data on income were sought; the census merely included a list of domiciles, addresses and size of families. Accordingly, the information in this census had to be supplemented with data from administrative records from the State Secretariat of Finance relating to the collection of real estate taxes, and a proxy variable for income was used, namely, the amount of tax paid on the value of residences included in the frame.

The sample was designed to be self-weighting. The probability of selection for each family was to be equal, and the selection was carried out in two stages. In the first stage a certain number of Census Sectors* were chosen, with probability proportional to their size, and within each sector a fixed number of families was selected. These families constituted the "initial sample", and were classified into strata according to size and amount of tax paid on homes. In the second stage, the homes composing the

*For Census purposes, cities are divided into Districts, and the latter are subdivided into smaller areas which are designated as Census Sectors.

"final sample" were selected from those which composed the "initial sample".

The questionnaire used in the survey was based upon the experience and model utilized in the previous 1961-63 investigation. However, care was taken to adapt the questionnaire for international comparisons, by providing for detailed schedules of family expenses and income. All data were sought by memory, resulting in a long questionnaire which required, on the average, over two hours to respond. The questionnaire had six different parts. The first part deals with individual characteristics of the members of the family; the second requested information concerning the home, an inventory of major durable and semi-durable assets, and expenditures and purchases of home related items; the third part sought detailed information on consumption of very diverse goods and services such as foodstuffs and beverages, textiles, clothing items, personal care, recreation, reading and education, public transportation, trips, health and miscellaneous expenses. The fourth part of the questionnaire asked for family income, specified according to class and origin as well as changes in debts and capital; the fifth part recorded ownership, purchase and sale of real estate and vehicles, donations and presents given and received; and the last part contained general questions to be answered by the interviewer on the reliability of the data, which members of the family supplied the information, and a summary balance sheet to bring together an income and expense balance of the family.

In view of the nature of the data which was to be sought and the experience of the previous investigation, a female member of the family, preferably the lady of the house, was the preferred respondent. However, since some interviews had to be conducted in low income areas or difficult to reach, a limited number of male interviewers were employed. In Rio de Janeiro the

majority of interviewers were primary school teachers and homemakers, with a smattering of education and sociology students. In Recife the interviewers consisted of social sciences students; and in Pôrto Alegre, of students of economics and social sciences.

Organization of the field team was a serious problem, because many of the interviewers who had been selected and trained for the job withdrew and had to be replaced after the first interviews, in view of the complexity of the survey and their unfitness for this type of work. This problem persisted during the whole duration of the survey. Taking Guanabara (Rio de Janeiro area) as an example, of 100 initially selected and trained interviewers, 23 lasted through the first quarter; the number went down to 16 during the second quarter; to 15 during the third, and to 8 during the fourth quarter.

The lack of a permanent and continuous team of interviewers during the investigative period contributed greatly to extending excessively the duration of the field work and must have decreased the reliability of the data. Nevertheless, receptivity was good and the rate of refusal was quite small: 13% in Rio de Janeiro, 11% in Pôrto Alegre and 7% in Recife. The rate of refusal was not proportionately distributed among the various strata, but was considerably larger for higher income families.

Before accepting any questionnaires, they were put through a series of checks to test the consistency and quality of the information collected by the interviewers. Parallel means were used to confirm if families had been in fact interviewed at the addresses chosen. Ten percent of the questionnaires, chosen at random after having been revised and submitted to criticism, were turned over to the supervisors for a rapid second visit in which they would repeat the queries on the composition of the family, food consumption habits and income. The amount of unconfirmed data was small, and not enough to

adversely affect the result of the survey, though in some cases, there were grounds for discharging the interviewer. In addition, a number of questionnaires were returned to the interviewer for clarification. In most such cases it became necessary to revisit the family to seek more complete data. Only after the supervisors had decided they were satisfactory were the questionnaires forwarded to be coded.

Final evaluation of the reliability of the information and errors of sampling can only be made after processing of the data, being performed at the Brookings Institution. Nevertheless, some preliminary tabulations do indicate discrepancies between ex-ante information within the frame and the actual results. Table 1 presents, as an example, results for Guanabara for the first quarter. These figures clearly indicate that ex-ante data derived from the tax files frame show a concentration in the first strata, which if true would mean an excessively flat income-distribution pyramid, with almost one third of the population at the base. However, the ex-post information from the survey data show heavy concentration in the middle-income groups, which strongly contradicts conclusions that might be derived from the ex-ante data.

This finding highlights one of the problems pointed out in the previous section on the use of administrative records for statistical purposes. Inasmuch as ex-ante stratification was carried out from the assessed tax value of households, and granting that the correlation between income and quality of dwelling may not be perfect, this would not be sufficient to explain the large divergence disclosed. This divergence may be explained by the fact that, since the objective of the assessment was to collect the ad-valorem tax, such assessments were established considerably below the rent values.

Table 1

Ex-Ante Design of the Sample and Ex-Post
Data, by Income Levels and Family Size, Guanabara
 (First quarter)

<u>Tax assess-</u> <u>ment class</u>	<u>Family Size</u>								<u>Total</u>	
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7-9</u>	<u>10+</u>	<u>Number</u>	<u>Pct</u>
<u>Ex-Ante Design</u>										
A	8	17	21	17	16	14	18	6	117	27.9
B	2	7	7	10	7	3	3	3	42	10.0
C	3	8	12	9	4	3	6	1	46	11.0
D	4	5	6	6	9	5	4	1	40	9.6
E	3	11	12	11	6	6	5	1	55	13.1
F	2	8	10	10	9	4	4	1	48	11.5
G	5	7	7	5	6	4	3	-	37	8.8
H	1	4	3	4	4	2	5	-	23	5.5
I	-	2	3	3	2	-	1	-	11	2.6
Total	28	69	81	75	63	41	49	13	419	--
Pct	6.7	16.5	19.3	17.9	15.0	9.8	11.7	3.1	-	100.0

Ex-Post Data

<u>Income class</u>										
1	10	4	3	3	-	2	2	-	24	5.9
2	2	6	5	2	4	1	3	-	23	5.6
3	2	15	7	11	3	4	5	1	48	11.7
4	9	16	20	19	16	11	8	1	100	24.5
5	6	10	14	14	13	6	10	4	77	18.9
6	2	12	15	14	12	7	6	5	73	17.8
7	3	5	6	7	4	3	1	1	30	7.3
8	1	2	3	2	4	4	4	-	21	5.1
9	-	-	2	6	2	2	1	-	13	3.2
Total	35	70	75	79	58	40	40	12	409	--
Pct	8.6	17.1	18.3	19.3	14.2	9.8	9.8	2.9	-	100.0

(ii) Colombia

As part of the ECIEL study of consumption and income patterns in Latin America, a survey on this subject was carried in four of the major cities in Colombia in 1967-68--Bogata, Baranquillo, Cali and Medellin. These four cities constituted 22 percent of the total population of the country at that time and nearly 40 percent of the urban population. The objective of the study was to compare household consumption expenditures and income in the major urban areas of Latin America. To this end, surveys were planned among private households in selected urban areas of 500,000 or more in all the Spanish and Portuguese speaking countries in South America and in Mexico. The four cities selected in Colombia were felt to provide very good representation in the light of these objectives.

Planning. The plans for the study were carried out mostly during 1966, including a pretest. The plans called for a panel design with some housing units interviewed each of four quarters over the period of a year, some families interviewed only twice and others only once, in the manner illustrated by the following diagram; the figures are initial sample sizes after allowance for noncontacts and refusals on the first wave:

<u>Subsample</u>	<u>Quarter</u>				<u>Total</u>
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	
1	260	225	135	160	830
2	260	225			485
3	260		205		465
4	260			170	430
5	170				170
6		170			170
7			170		170
8				170	170
Total	1,210	620	560	500	2,890

It was estimated that 1,720 different housing units would be interviewed to obtain 2,890 completed interviews and that, with allowance for initial refusals and noncontacts, about 2,015 housing units would have to be contacted.

The purpose of this design was to be able to evaluate conditioning effects and mortality as a result of the panel and to adjust the data for any such biases that might be uncovered (recognizing that such adjustments are not easily made). Accordingly, each subsample was a systematic random selection from the main sample frame.

To assure representation from different income levels, three such strata were used--high, medium and low--the basis for stratification being data provided by the municipal planning offices to classify each neighborhood (barrio) by these levels. The stratification was on an area basis, with reference to the general condition of the barrio.*

For sampling purposes the "family" was treated as the unit, defined as consisting of all those in the same dwelling unit and sharing cooking facilities, excluding boarding units (defined as dwelling units with four boarders or more). For analysis, the family was divided into three types of consumer units--primary units (family members sharing at least food expenditures with the main wage earner), secondary units (family members sharing at least food expenditures with each other but not with the head), and supplementary members (individual family members sharing at least food expenditures with other family members but not other types of expenditures, e.g., servants, grown children with own income). A family therefore would consist of one primary unit, additional secondary units and supplementary units (each numbering from zero onward).

*In terms of population, about 65 percent of the barrios were classified as low, 25 percent medium, and 10 percent low.

The data were to be obtained by recall, based on a questionnaire administered to the head of the consumer unit and/or his wife. Three different questionnaire forms were to be used, one applicable to primary and secondary units, one for supplementary units, and one for interviewers to record the circumstances and quality of the interview. A pretest of these questionnaire forms versus diary forms indicated that they yielded about the same level of expenditure data though the diary forms provided more details; these forms were therefore used in the main study.

Execution. The study in Colombia was planned and carried out by staff of the Center for Economic Development (CEDE) of the University of the Andes, in Bogata. The questionnaires and instruction manual were developed by this Center using prototypes developed by the coordinating staff at Brookings. Interviewers, hired and supervised by CEDE, were students at the University who were given intensive training in using this type of questionnaire. In particular, they were trained before the first wave in listing procedures (as a basis for probability selection of households), contacting and approaching sample members, explaining the objectives and importance of the study, use of the various questionnaire forms, and means of securing future cooperation. Following the training session each interviewer was required to complete two practice interviews and attend a final review session. Training sessions at later waves were much shorter, and were mainly for review purposes (except for any new interviewers that might be hired) since the same procedures and questionnaires were used.

The field work conformed well with the original schedules, the four successive waves being conducted in the four cities in May, August, and November, 1967, and May 1968, respectively. Extensive quality checks were incorporated, including verification of 10-20 percent of each interviewer's questionnaires relating to various parts of the interview as well as to correct

selection of the sample household. Questionnaires were also checked in the office for completeness of information and for arithmetic and other inconsistencies. If major errors or omissions were found, the questionnaire was returned to the interviewer for further work.

The data were coded in Bogota in accordance with a coding manual developed for all of these studies by the coordinating staff at Brookings. After keypunching and machine verification in Bogota, tapes were prepared and sent to Brookings where the data were put through an elaborate set of machine cleaning programs including corrections for extreme values by regression techniques.

A final stage of data purification consisted of analyzing the data for evidences of bias due to sample mortality or panel conditioning, and corrections for any such biases as well as insertion of missing values. Checks of sample representativeness showed no evidence of bias in the initial wave, for the total sample or by subsamples. However, panel mortality over time was detected, with more than proportional losses among small households, those with young heads, those with unemployed heads and households at very low or very high income levels. On the other hand, panel conditioning effects were slight.* Hence, a series of adjustments were applied to impute missing values where needed, to make quarterly estimates of income and expenditures by category for each family unit, to combine them into annual estimates where the data permitted and to weight the quarterly and annual records to correct for panel mortality.**

*A complete analysis of sample biases in these data is contained in Musgrove, Philip, The Collection and Interpretation of Household Income and Expenditure Information. Washington, D.C.: Brookings Institution mimeographed report, 1972.

**Ibid.

In retrospect, the principal problem with this study seems to have arisen in the field operation, especially with the panel subsample. The extent of panel mortality was far higher than anticipated, due in part no doubt to the prior lack of experience of the field staff, as can be seen from the following comparison of planned interviews, initial sample only, and initial sample with replacement interviews:

	<u>t₁</u>	<u>t₂</u>	<u>t₃</u>	<u>t₄</u>	<u>Total</u>
Planned	260	225	185	160	830
Initial and sub- sample only	183	125	86	66	460
Initial interviews with replacements	215	241	245	228	929

The starting sample size with the panel was much lower than expected and, in addition, as evident from the second line of this tabulation, panel mortality took a heavy toll. Only about a third of the original panel members were interviewed all four waves as compared with an expected 60 percent. The quick and substantial replacement of panel dropouts more than offset the panel losses, as seen in the third line of the tabulation, but raised some difficult measurement problems that undoubtedly affected the reliability of the data to some extent.

In fact, the overall number of interviews was somewhat higher than expected (2,949 vs. 2,890), but the general experience was for fewer than anticipated originally interviewed units to be interviewed over again in the panel subsamples n_1 to n_4 , and for the gaps to be filled by replacement interviews. By hindsight it would have been simpler and more efficient to have deleted the partial panel subsamples, n_2 to n_4 , and to have increased the size of the panel subsample, n_1 , and to have put more emphasis on maintaining the cooperation of those sample members. Even so, a great deal of very useful data were obtained which has served, and will serve, as a basis

for many studies of income and consumption patterns in the principal urban areas of Colombia.*

(iii) Paraguay

The consumption study in Paraguay, like the one in Colombia, was also carried out under the ECIEL program, with the same objectives and under the same criteria. However, Paraguay is a much smaller country, with about two million people, and with only one major urban center, Asuncion, having approximately 300,000 people. Hence, the study was restricted to that area.

Planning. The plans for the study were made in 1969 and early 1970, later than the study in most of the other countries. The study plans were prepared by the Centro Paraguayo de Estudios de Desarrollo Economico y Social in Asuncion, with the collaboration of the coordinating staff at Brookings. The plans called for an initial sample of 300 housing units (the sample size limited by available funds) and for a modified panel design of three equal subsamples, one a panel interviewed twice six-months apart; seasonal variations were felt to be of much less consequence for Asuncion and six-month recall of major durables seemed feasible on the basis of a pretest.

The same units of analysis were employed as in Colombia but only one questionnaire form was used for all consumer units, based on the common questionnaire designed by ECIEL several years earlier. The pretest supported this type of approach, the main difficulty emerging from the pretest being that of poor maps and inadequate instructions for locating housing units. As in Colombia, sample selection was on an area probability basis with stratification by income level of neighborhood.

Execution. Interviewer selection, training and supervision was similar to that in the consumption study in Colombia. The interviewers were

*For example, Prieto, Rafael, Encuesta de Presupuestos Familiares Barranquilla, Bogata, Cali, Medellin, Bogata: Centro de Estudios sobre Desarrollo Economico Universidad de los Andes, 1970.

either university social science students or social workers. They were put through a similar training program and were subject to the same sort of supervision and verification as in Colombia. The field work was carried out in September-October 1970 for Wave 1 and in March-April 1971 for Wave 2. The sample results were remarkably good, as is evident from the following data:

	<u>Wave 1</u>	<u>Wave 2</u>
Housing units selected	303	303
Not housing units	30	3
Refusal	4	1
Interviewed	269	299

All of the 101 families in the panel subsample interviewed on the first wave were also interviewed on the second wave.

The data reduction and purification processes were the same as with the Colombian consumption data. Needless to say, panel mortality was of no consequence, nor in fact was panel conditioning. Construction of annual records was much simpler on account of the simpler panel design, as was estimation of missing values.

B. Price Study

This was another of the survey studies in Latin America coordinated by ECIEL. Unlike the consumption study which had both national and international objectives, the objectives of the price study were almost entirely international, to enable price comparisons to be made across the LAFTA countries. This made the development and execution of a standard approach much more feasible.

Planning. Prices on a wide variety of goods and services were to be obtained to yield a representative picture of price conditions. Four classes

of products were to be covered: consumer goods and services, investment goods, prices of the inputs (mostly wages and salaries) into public goods, and prices of exports and imports. The prices were to be collected in the capital city of each country except for Ecuador, where Guayaquil was included in addition to Quito. By far the largest segment was the first, private consumer goods and services, the sample for which consisted of 416 separate items. Three qualities were priced for each item, with about seven observations for each quality, the actual number varying with the item.

The sample of items was standardized for all countries, the selection based in part on a similar study in 1961 by the Economic Commission for Latin America (ECLA) and partly on the experience with the consumption study. A uniform set of questionnaire forms was prepared by the coordinating staff at Brookings, and all data were to be collected during the same period, in May 1968. Outlets for collecting price data were selected on a stratified probability basis, using for consumer prices areas included in the consumption study. Altogether, about 500 pricing outlets were used per city. In addition, some prices, such as housing, were taken directly from the questionnaires on the consumption study. All data reduction and data processing was to be done at Brookings using standardized procedures.

Execution. While the uniform procedures and standardized questionnaires helped to simplify the planning and later data reduction and processing stages, it was not as easy to implement at the data collection stage, despite the use of coordinating staff and others to visit each country to ensure comparable selection of goods and outlets and uniform implementation of the data collection. While these visits definitely helped to promote standardization, especially since these visitors were used to help standardize interviewer training as well, the attempt to apply standardized procedures to 11 countries with highly diverse economies and products was bound to cause difficulties.

One problem was that despite the training interviewers showed a tendency in all countries to select more modern and more easily available outlets, which might have introduced a bias for some commodities toward higher prices. Another problem was the inability at times to obtain price quotations on the same, or even similar, products in all countries. Some products, or product qualities, were not available; in other cases price controls and suspicion on the part of manufacturers made it difficult or impossible to obtain realistic prices; and in still other cases manufacturers did not want to supply price quotations according to the preset specifications (this was especially true of machinery and construction goods).

As a result, instead of the 8,700 price quotations planned to be collected per country, the actual number was in the range of 4,000 to 5,000. Had the data collection period been extended beyond May 1968, more prices would undoubtedly have been collected but it was felt important that all prices refer to the same month insofar as possible.

In accordance with the original plans, the data were coded by the individual institutes and these sent with the original questionnaires to Brookings. There the coding was checked, errors corrected by contacting the particular institute, and then the data were keypunched and put through an elaborate cleaning and editing process.* In view of the fact that the same process was used for all countries and in view of the standardization of the sampling and the data collection, the comparability of the results across countries

*The details are given in Salazar-Carillo, Jorge, "The Use of the Computer in Handling Large Price Files," paper given at Conference on the Role of the Computer in American and Social Research in Latin America, Coernavaca, Mexico, October 1971; to be published by National Bureau of Economic Research.

was about as satisfactory as any such study could be, though in a more general sense one might question the representativeness of a basket of goods that is the same, essentially, in all countries.

The Case of Brazil

The price study in Brazil covered, as in the other countries, a large number of goods and services, as well as civil service salaries.

The choice of goods, as well as detailed specifications, were made by the coordinators and forwarded to the participating institutes. Specifications were very detailed, including brand name if possible.

In carrying out the survey, the Instituto Brasileiro de Economia benefited from twenty years of experience in price surveys and calculation of price indexes. The team that did the work was basically the same group that has been engaged for many years in collecting prices, and the majority of the firms which were visited were customary informants of the Institute. Nevertheless, in view of the complexity of the work and the urgency required in carrying it out, only experienced interviewers were used who belonged to rather high social strata and for that reason could have access to corporation executives. Hence, no difficulty was encountered in contacting executives and in obtaining the data, whenever available.

For consumer goods the selection of business firms was done to include the most representative types in the marketing sector of the city of Rio de Janeiro with regard to the category of the firm and the socio-economic level of its clientele. Data collection on foodstuffs was carried out in 9 retail firms, which included three self-service chain supermarkets, three large chain grocery stores, and three small corner grocery stores. Of the three stores of each type visited, one was located in a mainly high income zone, another in a middle-class borough, and the other in a low income section.

For clothing, appliances and similar goods, 6 stores were selected, which included two large chain department stores, two stores in a preponderantly higher middle-class and high-class zone, and two stores in a mostly lower middle-class and lower class sector.

The main hurdle in collecting consumer good prices was to identify the products and to establish the linkage between the specifications of the coordinators, mostly based on the production and marketing structures of the United States, and the types of products currently found in the Brazilian market. A number of difficulties was encountered, such as the following:

a) Specifications of the goods were so detailed that salesmen were not always able to identify the article; for example, men's suits made of precisely specified combined textiles, with percentages of natural and synthetic fibers making up the combination.

b) Unavailability, in the Brazilian market, and particularly in Rio de Janeiro, of items only used in cold climates, for example, woolen undershirts, winter gloves.

c) Specification of items no longer consumed, e.g., ladies' Ban-Lon ensembles, men's cotton socks.

d) Specification of items which in Brazil can only be purchased in luxury stores, for example, specified types of lingerie which could only be found in some of the most sophisticated "boutiques".

e) Sizes and unit measurements different from those generally found in Brazil, for example, bedsheets larger than can be purchased in Brazilian stores, different width of textiles. Also, in Brazil fruits are generally sold by the dozen, and the price requested was by Kg.; and the size and weight of canned foods did not coincide with the Brazilian products. Also, some

foods do not exist in Brazil, and meat-cuts were different as well.

The system of weights for the basket of consumer goods was based on the family budget survey mentioned previously with only two small adjustments. Estimates were made of rent imputed to owner-residences, and data for families in the two upper income deciles were excluded, as in other countries.

Price collection for machinery and equipment and durable consumer goods are discussed jointly, since procedures were similar. As for consumer goods, the list and specifications were supplied by the coordinating staff. Prices of these goods were obtained from trade representatives of major manufacturers and main importing firms. In general, five firms were selected for each product (although some times the number of existing manufacturers or importers was less) that supplied prices for the items contained in their total production line. Thus, for electrical alternate current motors, 4 manufacturers were chosen and they supplied prices for 46 different specifications.

Problems of adapting specifications of these products were more serious than for consumer goods. Although Brazil still imports a considerable number of machines and equipments, the major portion of these items used in Brazil are manufactured domestically. In addition, Brazilian laws render it practically impossible to import items also produced in the country. About 80% of the items specified in the basket of machinery and equipment and durable consumption goods, were found to be domestically produced, with brands and other specifications different from those in the list supplied by the Coordinator. Hence, we were obliged to make a rather exhaustive survey of prices of Brazilian products with characteristics more or less approximate to those in the basic list, to allow the final prices that were utilized to be inspected by regression methods.

The weights for this basket of these goods were calculated from statistics on the value of national production and of imports.

No major problems were encountered in the survey of Civil Service salaries, since there was detailed information in the administrative records of the Civil Service Central Control Department.

For building materials, a very different kind of problem was encountered. The collection of prices for the four selected types of structures (residential homes, apartment buildings, commercial buildings, and highways) did not present any difficulty inasmuch as the specifications of the components and materials corresponded exactly to standards used in Brazil. The problem consisted in setting up a weighting system for these four types of buildings. Weights should be derived from data on investments in national accounts. However, investment estimates in Brazil are in global terms, by means of apparent input consumption, and only public and private investments are separated. In view of this fact it was necessary to make an estimate, albeit gross, of the relative importance of each type of building within the flow of national investments. To do this, data were derived from administrative sources and from balance sheet statistics of the Union, the States and Municipalities, as well as from autonomous agencies, foundations and governmental enterprises. The value of buildings according to type was estimated by the Instituto Brasileiro de Geografia e Estatística. In effect, therefore, administrative records and other sources had to be used to supplement the survey data.

4. An Overview

Based on the foregoing material as well as on more general considerations, it would seem appropriate to summarize in this concluding section of the paper some of the pitfalls involved in international comparisons using micro data and what sort of measures might deal with these pitfalls in future studies. Of the many such pitfalls, only those are covered which in the opinion of the authors deserve special consideration.

1. A pitfall that is likely to be overlooked at the initial stages is the extent to which studies designed for making international comparisons may be distorted by national interests. If the pertinent data are relevant for national policy making, and especially if similar data have been collected for this purpose in the past, the particular institute will be under strong pressure to modify the study design to obtain data that are comparable with the data it has collected in the past rather than data on a basis comparable with those to be collected in other countries. To attempt to influence a country to alter its questions and categories to fit the purposes of an international study at the expense of the comparability of the data for its own purpose is hardly likely to succeed and is likely to cause trouble at a later time. A more satisfactory solution is to develop a study design and a questionnaire format that can satisfy both needs with appropriate modifications, even if higher costs are involved.

2. Structural differences in consumption and in production make the study design very difficult when countries are involved that span many different latitudes and customs. Thus, in the case of comparisons relating to Latin America, consumption differences resulting from climatic conditions, customs and traditions are vast. Goods which are essential for the population of a particular country--winter clothes, for example--do not exist in other countries. Food habits are also very different. For example, in Brazil rice and manioc flour are more important basic foodstuffs than bread or potatoes, and poultry is considered to be more desirable than beef.

The production structure also makes comparisons difficult and sometimes even impossible. In some countries, like Argentina, Brazil and Mexico, the industrial sector is chiefly made up of large corporations in such sectors as steel, petrochemicals, automobiles, shipping. In other countries, such as

Bolivia and Ecuador, basic industrial activity consists in primary processing of agricultural commodities and family-type craftmanship. Comparisons of the industrial sector in these two type of countries may result in absurd conclusions.

3. Difficulties of standardization of units have already been illustrated in the preceding section. Units of measurement, sizes and types of packages, product quality, brands and product types may all differ from one country to another, and to such an extent that rigorous specification is not likely to be practicable. How to handle this problem is very difficult. To some extent equivalent measures may be designed. In other cases, it may be best to use looser specifications and relate the products and the categories more to their uses or type of consumption than to seeking an exact replication in each country. In still other cases statistical equivalents may be obtained by regression methods, although these methods necessarily involve assumptions of similar interrelationships among products in different countries that may not be valid.

4. In the case of administrative records the quality of a particular type of data may vary substantially from one country to another. In some instances, it is impossible to obtain the necessary data because they are kept confidential by the particular agency. In still other instances, data may be intentionally distorted for political purposes.

For these reasons, unless the data have been evaluated and validated in the past, a prudent approach is to compare data from administrative records with other sources. The comparison cited earlier in describing the consumption study in Brazil is a good example; while the income data from the survey can not be considered fully accurate, they nevertheless serve as a good yardstick for evaluating the general adequacy of the proxy variable from administrative records used for income stratification.

5. Differences in the type of sampling materials and other survey resources from one country to another may have a substantial effect on the quality of survey data obtained from different countries. If one country has a master sample of its urban areas while another country has only census materials eight to ten years old, one may expect substantial differences in the representativeness of the samples obtained from the two countries. Devoting extra resources to obtaining better sampling materials in the latter country, an obvious solution, is not always feasible either from the point of view of personnel or resources.

6. Still another pitfall is that the ex post results of a survey operation may be very different from the ex ante plans, and these differences are likely to vary substantially from one country to another. Differences in the quality of field operations are especially likely. It does not necessarily follow that the organization with the most experienced field staff will do the best job, since the field staff may not be experienced in that type of work. Also, the quality of supervision is a key variable, and this can vary greatly among organizations. To some extent, standardization can be promoted by developing uniform training procedures and materials as well as supervisory and verification practices. The fact remains, however, that every field organization tends to do some things in its own way, and it is only by constant communication that some semblance of standardization is achieved.

Coding, editing and other data reduction operations are also subject to considerable variation among countries even when common codes and standard coding conventions are provided. Ideally, data reduction standardization could be achieved by having it all done at a single location; in practice, this is usually impossible, and also may be undesirable from a field point of view. The only answer, both for data reduction and for field operations, is

to provide adequate budget for coordinating staff to visit regularly every country while the study is underway and to check on the procedures being used. Relying on mails, cables or telephone is hardly feasible in the less developed countries, particularly since one organization may be unaware that it is doing things differently from the others until it is noticed during the course of a visit.

7. The timing of a study is likely to be interrupted by numerous factors, not the least of which are political upheavals. Rarely is it possible to carry out a single study at the same time in many countries (the ECIEL price study was an exception), and rarely is it possible to stick to a preset time schedule.

8. Nowadays practically every country imposes certain restrictions on foreign trade and on exchange rates, so that any comparison of micro data in terms of "monetary value" are necessarily biased by such restrictions. Even in countries where there are no governmental restrictions on the exchange rate, it is subject to the influence of so many factors that the correlation between this value and purchasing power is small. Among these influencing factors are the flow of goods and services, components of relative prices, political stability, governmental restrictions to incoming or outgoing foreign currencies, flow of tourist travel, balance or deficit of the balance of payments, and of course government restrictions on the exchange rate. To cite one example, in February, when there is a marked increase in the number of tourists into Brazil, particularly Americans, the free (illegal) exchange rate raises the value of the cruzeiro; in July, when there is an opposite flow (from Brazil to the United States and Europe) the free rate of exchange reduces the value of the cruzeiro.

9. As a final point, it cannot be overemphasized that micro data collection for international comparisons involves constant coordination effort, an effort that cannot be conducted from a fixed location. Achieving uniformity of concepts and procedures at the ex ante stage is a prerequisite for the success of such an operation but is no guarantee that it will yield comparable data. To achieve the latter, constant checks must be carried out to assure that the survey work in the field, and on the questionnaires when they are returned from the field, are handled in a similar manner in all countries. Yet, rather paradoxically, strict standardization may be self-defeating if standardization is unrealistic with regard to the variables under study. A certain amount of flexibility must be incorporated to allow for intercountry differences in living and in production patterns.



UNIVERSITY OF ILLINOIS-URBANA



3 0112 060296826