PRIORITIES FOR STATISTICAL DEVELOPMENT

An improved statistical system should be an aim of all countries seeking economic progress. But there are difficulties and pitfalls in the way of establishing such a system. The author has some suggestions to help the developing countries.

George J. Novak

Many developing countries are prepared to spend more on developing their statistical systems if they can be reasonably sure that the additional expenditure will help to promote economic development. Would a more advanced statistical service be a good investment for government funds? Or might some alternative be better? Governments have little relevant information: although it is often pointed out that better statistical information is needed for making short-term policy and long-term investment decisions, there has been little effort to quantify the trade-offs between the cost of improving statistics and the alternative uses of resources.
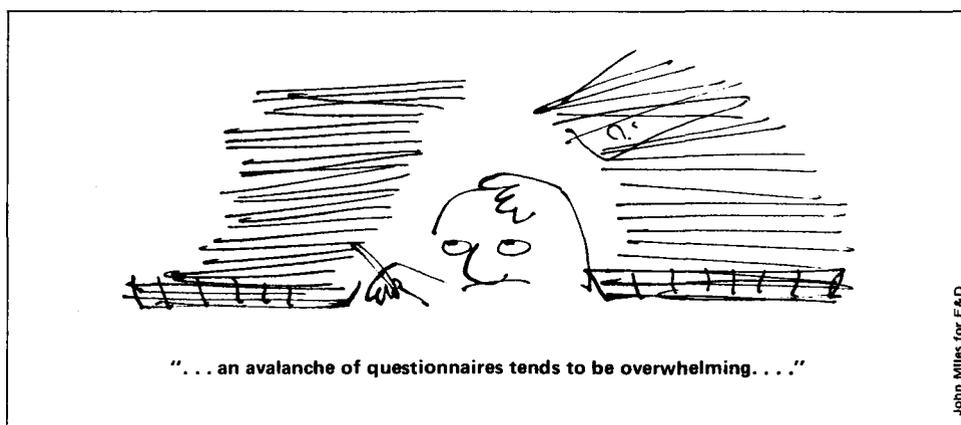
It takes years to develop a statistical system and to accumulate the time series required for economic analysis and planning. Therefore, instead of waiting idly for data, planners often resort to shortcuts which range from assumptions and analogies drawn from other countries' experiences to rough-and-ready estimates based on fragmentary information. At each stage of planning, various data inadequacies become apparent, and they tend to guide the planners in determining statistical priorities. However, this process often results in wrong and inconsistent priorities. Some countries attempt to construct complex secondary data systems of national and regional accounts or input-output tables without having first developed the necessary basic statistics. Such systems based on numerous assumptions and guesstimates may just as much mislead as they may guide both the planners and the statisticians who try to derive their statistical priorities from some of the overambitious planning schemes.

Priorities for statistical organization

An autonomous Central Statistical Office (CSO) is a cornerstone in the development of an efficient statistical system. However, most developing countries lack the skilled manpower and technical facilities required

"... an avalanche of questionnaires tends to be overwhelming. ..."

John Miles for F&D

for equipping a national CSO, or for providing multiple facilities for several government agencies and departments. Dispersing personnel with scarce skills, establishing multiple facilities for data processing, and maintaining extensive field staffs for collecting data usually result in widespread duplications, overlapping of functions, and overburdening of respondents. The latter is particularly important where the level of literacy is low and small businessmen lack the clerical staff to complete survey questionnaires requested by various agencies. Even with a centralized system, some developing countries have required their enterprises to fill out regu-

larly up to two dozen different questionnaires a month. With a decentralized statistical system, the avalanche of survey questionnaires tends to be even more overwhelming.
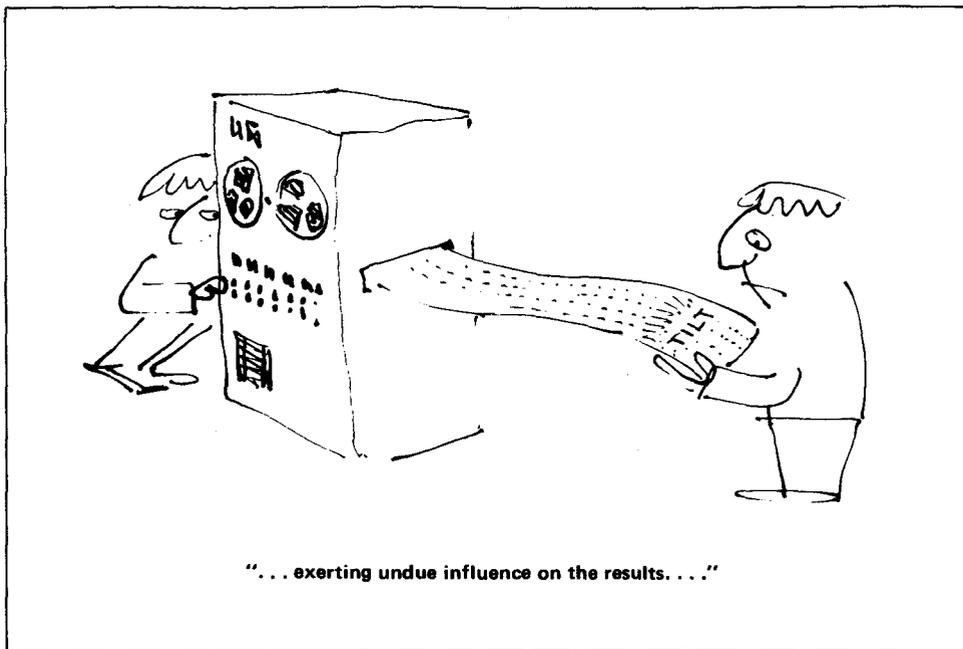
Yet it is normal for every agency to attempt to establish its own statistical services. However, they often produce shoddy statistics suffering from high nonresponse, extensive response errors and biases, and faulty sample and questionnaire designs. The agencies usually lack a trained field staff and utilize antiquated data processing facilities. The concentration of censuses and sample surveys within the

CSO, and its organization along functional lines should have top priority in the statistical organization of developing countries, even if the decentralized system—with its multiple facilities and duplication of effort—is functioning reasonably well.

In spite of the great advantages of the centralized system, the collection of administrative and specialized statistics should be left outside the CSO. Thus, public sector accounts may be compiled and perhaps even consolidated by the controller general, the ministry of finance, or the central bank. The customs collect and sometimes even process foreign trade data based on customs declarations. The central bank

9

A statistical council

Excessive independence of the CSO may, however, also be undesirable. The CSO may become less responsive to the needs of the statistical users and embark on ambitious projects which would produce data of little value and relevance to policy-makers and planners. Given a free hand in preparing their work program, statistical agencies tend to favor censuses over sample surveys, although the latter provide more timely and relevant data for the users. To strike and maintain the necessary balance requires a continuous process of supervision and negotiation. This authority should ideally be vested in an independent interdepartmental body—a statistical council—capable of remaining impartial in observing the objectivity of data and the interests of statistical users. Acting as an arbiter between the CSO and the data users, the statistical council would determine statistical priorities and formulate the general principles of statistical policy. Although each country may place a different emphasis on its priorities, the council could be expected to be responsive to the need for observing international statistical standards, recommendations, and priorities.

Regular publication of relevant and timely statistics is the ultimate objective of an efficient statistical system. The periodicity of publication deadlines imposes a stringent control on the performance of the system and provides the users with continuous time series. The users become accustomed to the regularly available statistics and, as the statistics are more widely used, they become more relevant to users' needs. Their usefulness is further enhanced by the publication of data not in isolation but as time series and preferably within the context of other relevant data, larger aggregates, and supporting details.

may have alternative foreign trade data based on import and export permits and foreign exchange transactions; it also collects financial and banking statistics, and it is usually in a better position than the CSO to prepare the balance of payments. The ministries of health, education, and justice usually maintain extensive administrative reporting systems which produce statistical data. Finally most agencies collect specialized information in the areas of their direct responsibility. Nevertheless, CSO experts on national accounts and sampling could be called upon to provide valuable advice for improving the relevancy and the reliability of decentralized statistics.

An autonomous CSO has the particular advantage of being able to produce more impartial and objective statistical data. In a decentralized system, each department exercises direct control over its statistical services; this makes it easier to exert undue influence on the results which favor the administrative or operational performance of the agency. Even with a centralized system, the planning office may exert undue pressure on the CSO in the choice of basic data sources, methodology, or in the interpretation of the final results which may be inconsistent with general planning objectives or specific plan targets. Supervising the CSO at arm's length, the planning office may—if the data show that planning targets are unlikely to be attained—delay the completion of data processing and the publication of the results by limiting the CSO's budgets. A close integration of the CSO within the administrative structure of the planning department has often resulted in extensive delays rather than in a more timely preparation of basic statistics. In

some countries, the release of monthly data has been delayed for two years and the processing of annual data stretched out for up to five years. The widening gap between the historical data and the planned targets provide the planners with an opportunity to construct politically more appealing but less realistic plans for economic development. The statistical office, having sacrificed most of its independence and objectivity, is then given more elbow room for bridging the gap between the ambitious targets and disappointing results. This process of adjusting statistical data to political objectives is sometimes extended to international relations—the negotiation of the level of gross national product between the Japanese and American authorities in postwar Japan is merely one example of such a statistical bargaining process.

George J. Novak

a citizen of the United States, graduated in business administration from Frankfurt University and obtained his Ph.D. degree in economics from Columbia University. Before joining the Bank staff in April 1967, he worked with the United Nations, the National Bureau of Economic Research, the University of North Carolina, and the National Planning Association. He is now an economist in the Economic Analysis and Projections Department.

Edwin G. Huffman for F&D

10

Priorities in basic statistics

Population

Population censuses have been conducted throughout history for the purpose of improving taxation policies, defense capabilities, and public administration. Nevertheless, demographic statistics may be of secondary importance for some sparsely populated countries and economies whose national income is largely derived from exports. While for these countries an approximate head count may suffice for most purposes, other countries require more detailed characteristics of their populations for educational and industrialization policies. Demographic statistics provide a framework for the formulation of social, economic, and political decisions. Thus, migration and vital statistics provide a basis for population projections by age and sex, making it possible to estimate the size of the labor force 10–15 years in advance. The number of school-age children and their regional distribution serves as a guide for the construction of new schools; rural-urban migration statistics provide a basis for the planning of new housing construction and the supply of unskilled labor; the occupational and industrial composition of the labor force derived from household surveys gives at least a crude picture of the total activity in various economic branches, with some indication of employment and unemployment in specific industries.

Foreign trade

Foreign trade statistics enjoy a high priority because they can be easily obtained from customs documents or foreign trade permits and are used for deriving the foreign balance of payments and some components of the national accounts. Basic data on central government revenue and expenditure are also relatively easily available. They are needed for the national accounts, the budgetary process, cash management, and for appraising the impact of fiscal policy decisions on the private sector of the economy. Despite their high priority, however, public sector accounts often lack economic and functional classification of expenditure, making the data less useful for analytical policy decisions. Statistical data for provincial, municipal, and other local governments are usually of poor quality, outdated, and sometimes not available at all. Decentralized agencies, government enterprises, and public corporations compile data for their own use but they are frequently not consolidated by sector or industry.

Agriculture

Since over half of the population in most developing countries belongs to the agricultural sector, agricultural statistics rank high on the list of basic priorities. Information on harvested areas and average yields of major crops provides an estimate of production which in turn permits estimates of relative prices and foreign exchange requirements—information particularly important in case of crop failures. To be really useful, however, such information must be available at an early date. It is often important to have a rough estimate of the expected agricultural output even before the harvest time or as soon as some disaster strikes. Therefore, instead of, or in addition to, data on harvested

"... politically more appealing but less realistic plans. ..."

areas and average harvested yields, average biological yields, and sown areas under various crops could serve as a basis for the assessment of the future agricultural situation. The estimation of total areas and average yields is difficult and costly in tropical countries with continuous sowing and harvesting seasons. Some countries have millions of farm holdings with an even larger number of plots where dozens of different crops are almost continuously sown and harvested. Under these conditions, it would be a truly Herculean task to measure the total areas and average yields with conventional census methods— or even with surveys based on modern sampling techniques.

In view of these difficulties, the priorities of agricultural statistics are heavily weighted in favor of probability sampling. The crop-cutting surveys should focus on a few key crops such as rice, wheat, coffee, cocoa, or sugar, while the remaining crops could be estimated from the household expenditure surveys.

Industry

Since promotion of industrial development is of vital concern, industrial statistics rank high on the list of priorities even in countries where industry still remains in its infancy. In most countries, quarterly data on employment, sales, value of nonfactor inputs, wages and salaries, and the output of major products would probably suffice for the construction of quarterly indexes of industrial production and preliminary annual estimates. Such data may be collected quarterly from large establishments with 100 or more employees. For medium-sized establishments with 10 to 99 employees, the same information may be collected once a year on a probability sample basis; for small establishments with less than 10 employees, a sample survey once every five years would probably be sufficient in most developing countries.

Priorities of other basic statistics vary from country to country and from period to period. Large public utilities in transportation, communications, electric energy, water supply, and sanitation can usually provide adequate basic statistics on employment, finances, and output. Large financial intermediaries and insurance companies also have fairly adequate basic data in most countries. The systematic collection and processing of these data are important for public policy decisions and for the national accounts. In view of the relatively low cost of collection and processing, the CSO should make every effort to standardize the compilation of these data at the establishment level and to set up regular reporting procedures.

Low-priority statistics

Statistics on construction, domestic trade, and private services rank low on the priority list in view of the relatively high cost of obtaining them. These activities are difficult to quantify and they are dispersed among small production units. The use of probability sampling and census frames could lower the cost of occasional surveys and make the collection of data feasible for these important subject-matter areas.

Statistics on health, education, and various other social indicators are usually compiled by specialized agencies from administrative reporting systems. Therefore, the CSO tends to give them the lowest priority within its work program. The analytical usefulness of social statistics is still quite limited. In spite of the recently increased interest in social indicators, they are still

11

useful primarily for partial analyses of social phenomena. The social sciences have not produced an adequate theoretical framework within which social indicators could be integrated, and their relative importance weighted, to interpret meaningfully the interaction of social phenomena which the indicators represent.

Statistical Priorities for Major Series

(Priorities within each group are illustrative)

Group 1.
Demographic and Social Statistics
—Population
  —age and sex distribution
  —vital statistics (births, deaths, etc.)
  —internal and external migration
  —regional distribution
  —distribution by industry and occupation
  —employment and unemployment
—Health and nutrition
—Education
—Housing
—Income distribution

Group 2.
Basic Economic and Financial Statistics
—Agriculture
  —harvested areas, yields, and prices of major crops
  —production and value of minor crops
  —livestock and products
—Industry
  —mining
  —manufacturing
  —electricity, gas, and water supply
—Foreign trade and balance of payments
—Price information
  —consumer price indices
  —wholesale price indices
—Public sector accounts
—Financial and monetary statistics

Group 3.
Derived Statistical Data
—National accounts
  —expenditure on GDP
  —industrial origin (value added)
  —distributive shares
  —national accounts at constant prices
—Production indices
  —index of agricultural production
  —index of industrial production
—Regional accounts
—Input-output tables

Priorities for derived data systems

The main derived data systems include the balance of payments, government and national accounts, monetary and banking statistics, and input-output tables. They establish conceptual frameworks for determining the economic relations of the country with the rest of the world, of the government and the banking system with the rest of the economy, and for identifying structural relationships and inter-industry flows. Such derived data systems require consistent and comprehensive basic data and this requirement tends to establish the relative importance of the relevant basic statistics. The balance of payments, government and national accounts, and monetary statistics rank considerably higher on the priority list than do input-output tables. The latter impose a heavy burden on statistical resources which often exceeds the analytical benefits derived from the use of input-output coefficients.

Price and production indices are essential for the deflation of national accounts components and other macroeconomic aggregates. Current price information is also vital for identifying short-term consumer price movements. The derivation of income distribution statistics and systems of social indicators is important for the formulation of social policy objectives. All of these derived data systems enjoy a high priority, particularly if their derivation involves relatively little effort.

Priorities for data reliability

Most developing countries would benefit from greater reliability of their economic and social statistics, although the improvement process is costly and benefits must be carefully measured against costs.

Plan targets and policy objectives relate to projected future conditions whose imperfect linkages with the past impose the first strain on reliability. The independently projected targets are subjected to an optimization procedure which—no matter how rigorous and sophisticated—only crudely approximates reality, even when all the explicit and implicit targets are iterated into a consistent system of relationships. Plans and policies constitute oversimplified logical abstractions of complex social and economic processes. We may count the number of people, but we may never be able to aggregate with the same degree of reliability their detailed characteristics—such as incomes—to say nothing of the satisfactions which people derive from their relative well-being. And yet it may be the national welfare rather than the mere number of people which is the important indicator needed for the analysis.

Several elements are involved in determining the benefits and costs of priorities for data reliability. In assessing the benefits to be derived from improved data, the following sources of possible errors must be considered: (1) adjustment of historical data to the concepts, definitions, and classifications to be used for making decisions; (2) projection of independent variables; (3) optimization of planned targets and their reconciliation with policy objectives; (4) validity of hypotheses and theories used in the formulation of policy objectives; and (5) policy priorities determined by the political process.

The sum total of incremental benefits must be balanced against the necessary additional cost of improving the data at each stage of collection and processing. Additional costs arise in (1) reducing non-response by improving statistical survey frames, repeated interview calls, and indirect estimates; (2) improving response by a better training of interviewers, use of alternative questions and their improved sequencing, consistency checks, and introduction of stricter survey controls; (3) reducing sampling errors by greater stratification and larger sample size; (4) reducing the processing errors by more extensive verification, consistency checks, editing, and the introduction of modern data processing techniques and equipment.

The cost-benefit studies of priorities for data reliability must also consider the timing of data availability. If the data are not available at the time when they are needed, then the benefits will be considerably reduced. In the absence of timely data, preliminary estimates must be used, and there may be little need for further refinements at some later date. Therefore, good preliminary estimates often enjoy a higher priority than more reliable but belated statistics.

Methodological priorities

Probability sampling based on reliable statistical frames has the highest priority among the alternative statistical approaches to data collection in developing countries. Contrary to widespread views, modern and properly designed sample surveys produce better results at considerably lower cost than a complete enumeration attempted by censuses. High illiteracy rates make self-enumeration impractical. Censuses use large numbers of temporary interviewers who are poorly trained and cannot secure adequate response results. The checking, editing, and further processing of millions of questionnaires usually results in many additional inaccuracies, even with the use of modern data processing equipment.

These difficulties are greatly reduced in well-designed surveys. A smaller number of permanently employed interviewers can be better selected, trained, and provided with necessary experience in survey work; a smaller number of questionnaires can be checked and edited more thoroughly; data processing takes less time; and the results become available at an earlier date than those of censuses. Since sampling errors are usually smaller than response errors and biases, survey data tend to be at least

12

as good or even better than the census results. Moreover, since surveys are cheaper, they can be conducted frequently to match the latest priorities of policymakers.

Data processing priorities

Data processing should be set up so that a series of preliminary estimates could be produced as quickly as possible with only partial data returns. While it may be easier to come up with one single estimate after a thorough checking and editing of all the responses, the release of preliminary estimates at an early date is essential for the more effective use of data. Moreover, the

to particularly urgent data processing jobs, provided that the disruption of the established routine would not significantly affect the data processing.

The centralized organization of the statistical system greatly facilitates the use of modern electronic data processing equipment. Many developing countries have imported second- and third-generation computers under their economic development programs. Unfortunately, various government agencies have frequently tried to install incompatible computer facilities of their own, sometimes as a matter of prestige rather than of real need. It is thus impossible to hook up the existing com-

Helping improve statistics

Whatever the difficulties and pitfalls, an improvement of the statistical system constitutes an integral part of economic and social development. While the developing countries have the willingness and sometimes even the resources to improve their statistical systems, they often lack the necessary skills and the sense of priorities. Weak statistical systems are overburdened with censuses which drain limited resources to the detriment of the immediate need to produce current economic and social statistics. In the absence of current data, governments have attempted to make heroic assumptions about the trends and levels of economic activity. Consequently, they sometimes embark on policies which give misplaced emphasis or even proceed in the wrong direction before more reliable data become available. The lack of statistical priorities is then reflected in misplaced and only vaguely defined national priorities.

The United Nations, other international agencies, and member nations have provided technical assistance in statistical methodology to many developing countries. The Fund has maintained a systematic service of technical assistance on balance of payments data, monetary and general statistics, and most recently on government finance statistics, while the Bank has specialized in setting up a reporting system of external debt statistics for its member countries. Since 1969, the Fund has provided technical assistance in statistics to 69 countries under a program for the establishment and improvement of central bank bulletins. As a major user of economic and social statistics, the Bank has recently reviewed the statistical systems of, among others, Pakistan, Indonesia, and Iran. These comprehensive reviews identified shortcomings, established priorities, and made recommendations for the improvement of statistical organization, basic statistics, national and government accounts, field operations, and data processing. Statistical mission reports have helped the countries to raise the standards and improve performance of their statistical systems.

Once priorities have been determined, gaps in technical capabilities can be identified and, if necessary, foreign experts brought in to strengthen the vital elements of the statistical systems in these countries. Close cooperation between the users and producers of statistics is needed at both the national and international level to ensure proper standards and priorities in improving social and economic statistics in developing countries.

"... inadequate response results...."

comments of users on the preliminary results help to avoid gross errors in the final estimates. The preliminary data could be derived by a probability sub-sample of returns, by partial projection of missing elements, and by other estimating techniques for closing information gaps. With electronic processing of frequent industrial surveys, data can be punched on cards as the questionnaires come in and the results are aggregated. Tentative estimates for the belated response can be derived from previous surveys in order to fill the gaps in the preliminary estimates.

Small and frequent surveys should generally have a higher priority than larger and less frequent surveys and censuses. Also, weekly and monthly surveys should have precedence over quarterly and annual reports. These general priorities may be modified by giving a greater preference

puter facilities or to shift the overflows in peak demand periods to other government agencies. Consequently, excessive computer capacity tends to build up in some developing countries.

Most of the computers are operated in single shifts during the regular office hours. At night, on weekends, and during frequent power failures, the air-conditioning is turned off and the equipment is subjected to excessive heat and humidity which contributes to frequent technical malfunctions and complete breakdowns. Scarcity of proper spare parts, adequate service, software, and experienced programmers shorten the useful life of electronic computers and reduce their utilization. In addition, the introduction of electronic computers tends to replace many more low and medium-skilled jobs than the highly skilled ones it creates.

13