

ON THE OPTIMUM METHOD FOR CHINA'S CROP YIELD SURVEY

Chen Guoshun, Statistical Research Institute
Hebei Provincial Statistical Bureau, Shijiazhuang, China

KEYWORDS: Crop yield survey, Crop-cutting, Farmer-recording

1. Introduction

China has tried the following methods for its crop yield survey:

- (1) Reporting from the bottom (village) all the way up to the top (Nation),
- (2) Estimating by crop-cutting sampling,
- (3) Estimating by typical farmer interviews (investigating) and others.

The first method has been less applicable since the reform of China's rural economic system, while the crop-cutting method has replaced it and become the most important method - the official data of crop yield are based on it. The typical investigating or interview method and others have not been used very often in crop yield surveys.

The crop-cutting method is known as an "objective" one since under ideal conditions it is not influenced by human behavior. Therefore, many countries and areas have used this method for crop yield surveys and more and more people get the chance to study it. Although both the theory and practice of this method have been developed over the years, experts in China as well as other countries have found that the crop-cutting method has many problems which cannot be solved easily. These include the complexity of the method, the unpredictability of bias and errors, the big workload and high cost, etc. All these have limited the use and spread of the method. These are also the reasons why we cannot use this method for crop yield surveys throughout the country at various levels, especially village, township and county levels.

The question then is whether there is any other method which is simpler, more accurate, and involves less work and less cost than the crop-cutting method. The answer is positive. The method may be called the farmer-recording sampling method. In this method, the household is taken as the basic sampling unit instead of the plot. The word "recording" means that the farmers themselves weigh and write down how much they harvest. The key of the method is to let farmers record rather than state their results. By this way we can get the real crop yield of the households (sampling units) so as to estimate the total crop yield of the area. The farmer-recording sampling

method has been used in some villages, counties and prefectures in China. Experience has proved that this method is simple, accurate and easy to implement. However, because there have been no theoretical or practical comparisons with the crop-cutting method, the farmer-recording method has not been paid enough attention and therefore has not been acknowledged.

This paper compares the farmer-recording method with the crop-cutting method and testifies that the former is the optimum method for China's crop yield survey.

2. New Cognition of the Crop-Cutting Method

The crop-cutting method, which is an objective method in theory, has many limitations in practice since, in fact, one cannot easily get the real objective results from the sampled plots. This method is currently being challenged both in theory and practice. So it is useful for us to review the "objective" nature of this method. The following are some of the shortcomings of the crop-cutting method. Some of these are well-known old problems, while others are special to China.

(1) Complexity. This method involves complex procedures such as selecting sampling plots, drawing the map of the sampling plots, and cutting, threshing, drying, and weighting, the crop, etc. Before the cutting work begins, investigators and respondents must be trained. Whenever the harvest season comes, Statistical Bureau at various levels, must send large numbers of investigators to direct the work in the villages. The main reason why this method has not been used by all the villages in China is because the sampling work cannot be done independently by respondents of the villages.

(2) Unpredictability of bias and errors. This method has two kinds of errors: Sampling error and non-sampling error. While the former can be solved by increasing the number of the sampling units, the latter cannot be controlled by that. This is the biggest problem faced by sampling theory and practice at present. Tore Dalenius points out in Elements of Survey Sampling that practice has proved that the crop-cutting sampling results might result in biased estimates because the boundaries of the sampling plots cannot be correctly outlined. Honeyer and Black (1946) found that the data obtained from sampling plots of 2 sq. ft. area each was 8% higher than those of 3 sq. ft. each.

Research in five African Countries, Benin, Central African Republic, Kenya, Niger and Zimbabwe, also found that the crop-cutting method gave over-estimates of between 14% to 38%. Because of the unpredictability of the bias and errors, it is hard to know the degree of the precision of the method.

(3) Difficulty Deciding Plot Size and Shape. In order to control the bias of the crop-cutting sampling, the plot size needs to be appropriate. According to the experiences of experts from India, the plot size of 100.9 sq. ft. is the most economical size leading to unbiased estimates. But this will need a big workload. Moreover, the bias from different shapes of plots are also not the same, the bias from a row sample being the biggest. (See Estimation of Crop Areas and Yields in Agricultural Statistics, pp 82-86)

(4) Lack of cooperation. The time when the sampling plots should be cut is the harvest season. Peasants are both busy and tired at that time and they do not like to cooperate with the sampling work. At the same time, peasants also do not like to let the sampling plots be located in their fields. So the sampling work cannot be done smoothly. Besides, many investigators at the county level have families in the countryside and they have to help their wives. Therefore, the results of the sampling cannot be guaranteed.

(5) Simulation of real crop losses. The method of crop-cutting within a sampled plot is different from harvesting big areas in that it cannot reflect the degree of losses in cutting, transporting, threshing and drying when large areas are harvested. The error due to this fact cannot be easily estimated.

(6) Risk taking. The harvest season in autumn in Northern China lasts two months. Investigators cannot watch the sample plots all the time. Thus, it often happens that the sampled plots are cut by peasants without informing the investigators. In fact, complete yield results of sampled plots are rarely obtained. Also, the sample cut might be eaten by chicken or pigs, and cannot be replaced. Therefore, it is hard to get objective results even by the objective method of crop-cutting sampling.

(7) Big workload. When the crop cutting time comes, the Statistical Bureau at the county level must send almost all its staff to help with the work because of the workload. Thus other work is often neglected.

(8) High cost. Every year the nation has to provide a lot of money for the fees of investigation, training, subsidy for the investigators and compensation to farmers for loss of the sample yield. Some counties cannot use crop-cutting

sampling for the crop yield survey due to the high cost of the method.

The above are some of the main shortcomings of the crop-cutting sampling method. These have already limited the spread of the method. The most important thing is that many villages, townships and counties cannot use this method for the crop yield survey and their needs have to be met by other methods. Fortunately, in practice, a more suitable method for the crop yield survey has been found: the farmer-recording sampling method.

3. New Considerations of the Farmer-recording Sampling Method

The farmer-recording sampling method - which is the method of letting farmers weigh and write down their results - has been thought of as a subjective method and so its use has been ignored. The main reason for calling it "subjective" is that farmers may not give the right figure of their harvest and the results might be influenced by human behavior. However, this is not the case in China today. The crop yield of farmers is no longer a secret because ever since the new rural economic system came to the village farmers have enough to eat. So, the very lucky thing is, if the data of the sample units - farmer's crop yield can be got correctly, the farmer-recording sampling method will be objective.

Considering China's own condition, it is feasible to adopt the farmer-recording method at all levels of villages, townships and counties and thus meet their needs. The following features of this method show that it has many advantages as compared to the shortcomings of the crop-cutting method.

(1) Simple. This method is so simple that almost everyone can do the work, whether it is simple random sampling or systematic sampling. The techniques used by the method such as selecting families, weighing and writing down the results as well as estimating (using only simple arithmetic averages) can be learnt by the respondents of the villages in less than an hour. They can do the work completely without the help of the investigators of the Statistical Bureau of the County. So this method can be used by all villages of China without difficulty.

(2) Controllable bias and errors. This method also has two kinds of errors: Sampling and non-sampling errors. The sampling error can be decreased by increasing the number of sample units. And the non-sampling error - the deviation of the farmer's yield recording from the real results - can be solved if the grain can be weighed accurately.

This is achieved by providing respondents with weighing equipments which is no problem.

(3) Suitable for any size and shape of the sample units. No matter what size the sample units are (the households may be big or small) and no matter what shapes the fields of the household are (round or rectangular), we can get unbiased results of their yields by weighing all the grain in the farm.

(4) Easy to cooperate. This method does not involve crop cutting on sample plots in the fields of the farmers which is not welcomed by them. They only need to cooperate by agreeing to weigh the grain in the farm. If we do the propaganda properly and give farmers appropriate subsidies for their work they would like to cooperate.

(5) Crop losses. The cutting, transporting, threshing and drying are all done by farmers themselves so by this method we can get the real yield of the crop instead of an artificial estimate from sampled plots.

(6) Safeguarding. This method does not involve any risk of losing the sample because the farmers will harvest when they have to and get what they should.

(7) Small workload. The sample work can be done by respondents of the village independently and the weighing and recording can be done by farmers themselves. The county Statistical Bureau does not need to send a lot of investigators to help direct the survey, so the work is lessened.

(8) Low cost. The sample cost is reduced because this method needs less fees for investigators, less compensation to farmers for the loss of the sample and less workload.

The aforesaid statements are the advantages of the farmer-recording method for crop yield surveys and more may be listed. The key to the usefulness of this method lies in the answer to the question 'DO FARMERS GIVE REAL RESULTS?' If the answer is yes, the method is objective. If not, the method is subjective.

Let us first look at the condition of China today and study the case of another sample survey - the rural household budget survey. In this method the household is also taken as the basic sample unit to obtain income and expenditure estimates at various levels. The farmers record the incomes and expenditures of their families every day and to the last cent that they get and spend, even the money they get or lose from gambling, because the data they provide is kept confidential. So there will be no reluctance on the part of the farmers to give the real results of their crop yield.

Do farmers have the ability to record the

yields of their harvests correctly? China is a country with a glorious history in culture and people esteem knowledge very much. Usually at least one person in a peasant family has a good education so the data of the crop yield can be written down rightly without any questions.

Can the farmer weigh the grain of the harvest easily? The land is distributed almost equally between peasant families based on the number of people in the family. The family sizes of the farmers do not vary very largely. So there are few real big peasant farms in China. The average size of a farm land was 1.1 acres and the average grain production was 1.9 tons for one household in 1989. The total grain production is indeed not a big figure for a household. If the weighing equipments are provided, the weighing work can be done without any difficulty.

As for the data of big farms, crop-cutting sampling methods can be used to estimate the crop yield. Thus it is quite practicable to use the farmer-recording method for the crop yield survey in China.

4. Farmer-recording Method in Practice

The preceding sections have discussed the advantages of the farmer-recording method for crop yield survey in theory. This method is very promising because it has so many advantages over the crop-cutting method. The most important thing is that this method can be used by all the villages, townships and counties in China.

How does the farmer-recording method work in practice? Is it so subjective that the crop yield data got from it are not acceptable? Facts speak louder than words. Let us look at the results of the crop yields of Hebei Province estimated by both crop-cutting and farmer-recording methods.

Table 1. Crop Yields of Hebei Province
(Unit: kg/mu)

Years	a	b	c=b-a	(c/a)%
1985	199.7	197.7	-2.0	-1.0
1986	191.6	184.6	-7.0	-3.7
1987	183.2	182.2	-1.0	-0.5
1988	202.5	181.3	-21.2	-10.5
1989	204.0	195.8	-8.2	-4.0

a - crop yield estimated by crop-cutting method

b - crop yield estimated by farmer-recording method

Table 1 shows that the data of crop yields of Hebei Province estimated by the two methods are quite close, except in 1988. Given that the standard error of the estimates, $\sigma = 10$, even the biggest differences of the crop yields got from the two methods in 1988 is within 3σ . That is to say, the crop yields estimated by these two methods are statistically the same, i.e. if the crop-cutting data are acceptable, the farmer-recording data are acceptable too.

Now let us see the results for a village. If the crop yield data of a village can be got easily with a low cost and satisfactory accuracy by the farmer-recording method, we can add them up to meet the needs of townships, counties as well as provinces and the nation. Also, we can use the village data to estimate the crop yields at all levels. The following table is based on the results of Xiangyan village, Jing Xian County, Hebei Province, where a and b have the same meaning as in table 1.

Table 2. Crop Yields of Xiangyan Village
(Unit: kg/mu)

Years	a	b	c=b-a	(c/a)%
1985	172.9	151.6	-21.3	-12.3
1986	141.9	140.8	-1.1	-0.8
1987	115.7	99.1	-16.6	-14.3
1988	179.7	176.1	-3.6	-2.0
1989	176.4	164.1	-12.3	-7.0

Table 2 tells us that the crop yields got from the farmer-recording method are lower than the crop-cutting data. That is reasonable and suits the reality because of the losses of cutting, transporting and threshing the crop, etc. From the data of 1986 and 1988 we can certainly say that the farmers gave the right results of their crop yields otherwise the data of these two years would not be so close. This proves that the data given by farmers are correct and the results got from the farmer-recording method should be objective.

As for the differences, we statisticians know that that is the effect of two different methods, which is not at all surprising. Even with the same crop-cutting sampling method, differences exist within different sets of sample units. Investigators of Xingtang County, Hebei Province have drawn two sets of sampling plots to compare the results of crop-cutting method with different sample units and

found that big differences might happen. For example, the differences between the wheat yields of five villages selected from the county in 1989 as obtained from two sets of sample plots for each village were 36,3,1,22 and 9 kgs per mu respectively. Yet there is no reason for us to suspect the sampling results. Moreover, we have favourable conditions to use the farmer-recording method for the crop yield survey because the lands of families for a village are also distributed according to the grade of land and the deviation of the crop yield within a village is usually very small. Taking Dongcha Village of Xingtang County as an example, for this big village with 440 families the standard deviation of crop yields in 1989 was only 4.6 kg per mu and the relative standard error was only 1.65%.

All these have created many good conditions for us to use the farmer-recording method for the crop yield survey in China. Thus, by using the farmer-recording sampling method we obtain crop yield data that can be as reliable and accurate as the data from the crop-cutting method but with less cost, less work and less time.

5. Farmer-recording Sampling -- the Optimum Method

We have compared the farmer-recording and crop-cutting sampling methods both in theory and practice. It is very clear that the former has many advantages over the latter. Yet, in order to prove that the former is the optimum method for crop yield survey in China we should compare these two methods with certain criteria. According to Tore Dalenius in Elements of Survey Sampling, three criteria for a good sample design are (1) the measurability criterion, (2) the efficiency criterion, and (3) the simplicity criterion.

First of all, just the same as for the crop-cutting method, the farmer-recording method is measurable in that the sample units, households, are selected with known probabilities, whether simple random sampling or systematic sampling is used. Second, the farmer-recording method is more efficient than the crop-cutting one because with a lower cost it can obtain the crop yield data with the required accuracy. Third, the farmer-recording method is more simple, which is the fact well known by all. So the farmer-recording method suits the three criteria well and is a very good sample design for the crop yield survey.

From the above studies we conclude that the farmer-recording sample method is the optimum method for China's crop yield survey. We can use this method in villages, townships, counties and up to provinces and the nation. By spreading this

method and using the crop-cutting method as a supplementary method for large farms, we can meet the needs of various levels for data on crop production.

BIBLIOGRAPHY

Cochran, W.G. (1977), Sampling Techniques (3rd Edition).

John Wiley & Sons, New York, U.S.A.

Dalenius, Tore (1985), Elements of Survey Sampling.

Notes Prepared for the Swedish Agency for Research Cooperation with Developing Countries (SAREC).

Food and Agricultural Organization of the United Nations (1982).

Estimation of Crop Areas and Yields in Agricultural Statistics. FAO, ROME, ITALY.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953),

Sample Survey Methods and Theory, vol. 1. John Wiley & Sons, New York, U.S.A.

U.S. Bureau of the Census (1968), Supplemental Courses for Case Studies in Surveys and Census, Sampling Lectures, ISP Supplemental Course Series, No. 1, Washington, D.C., U.S.A.

Verma, V., T. Marchant and C. Scott (1988), Evaluation of Crop-cut Methods and Farmer Reports for Estimating Crop Production Results of a Methodological Study in Five African Countries. Longacre Agricultural Development Centre Limited, London.

THE ROLE OF NATIONAL ACCOUNTS AND ITS IMPACT ON SURVEY DESIGNS: THE ZAMBIAN PERSPECTIVE

Ackim T. Jere, Central Statistical Office,
P.O. Box 31908, Lusaka, Zambia

1. Introduction

It is generally accepted that National Accounts provide a framework through which the performance of an economy can be monitored and evaluated. National Accounting is simply a systematic way of classifying the multitude of economic activities that take place in the economy, in different groups or classes, that are regarded as being important for understanding how the economy works. Construction of National Accounts is, however, a very intricate process which requires special skills since sources of data needed to construct reliable and useable National Accounts estimates are generally derived from numerous sources.

The level of statistical development associated with a country will in general be reflected in the diversity and perhaps more importantly in the quality of statistics produced. It is argued in this paper that the quality of National Accounts estimates is affected by the designs of the different surveys whose results are an input into the construction of the accounts. Moreover, developing countries generally do not have adequate resources to mount statistical systems which will ensure that data collection is fully and effectively accomplished. Consequently, significant gaps may exist in the statistics produced. Zambia may not be a model of developing countries but it certainly has features which are quite common among developing countries, particularly in Africa.

It is also a recognised fact that the preparation of National Accounts usually demands a rather systematic way of data collection to ensure the production of useable accounts. In doing so, Zambia adheres to the recommendations of the United Nations system of National Accounts.

This paper essentially discusses National Accounting practices in Zambia and the relationship of the National Accounts with establishment surveys. A brief historical development of the statistical system in Zambia is also discussed with a focus on the types of establishment based surveys currently being conducted in Zambia.

2. Development of the Statistical System in Zambia

Production of statistics in Zambia predates

independence in October 1964. Northern Rhodesia, now Zambia, was a colony of the British Government. At that time the Federal statistical office which was based in Salisbury (now Harare), Southern Rhodesia (now Zimbabwe) granted authority for establishing a statistical office in Lusaka, the Capital City of Zambia. The setting up of the statistical office in Lusaka was accomplished soon after the attainment of political independence in 1964.

The young and small office was initially run with the assistance of technical co-operation from the United Nations Statistical office. Some of the important and salient statistics produced then included estimates of national income. The Census of Europeans or Non-Africans was also conducted.

Over the years the Central Statistical Office (CSO) has expanded and it now undertakes a lot of statistical activities. For instance, the office has conducted censuses of Population and Housing in 1969, 1974, 1980 and most recently in 1990. The experiences gathered over these years have rendered a lot of improvements not only in survey designs and methodologies, but also in the designs and construction of survey instruments. The CSO now produces a host of statistics, which include Agricultural statistics, Economic statistics such as the National income accounts, Industrial production, Prices, Employment and Earnings, Government Financial statistics, Balance of Payments and External Trade, as well as Health, Education and Migration statistics.

The CSO has been keeping abreast of statistical developments and efforts have been made in producing statistics in relatively "new" areas such as gender, environmental and the informal sector statistics. These are new areas for Zambia's CSO and it is expected that improvements will be instituted in the scope and quality of these statistics over time.

Processing of statistical data had for a long time been done manually even as recently as in 1980. Since then, however, computer data processing has dominated in virtually all fields, and all survey/census data collected by the CSO are now processed using computers. In recent years micro-computers have been predominant in processing survey data and various software such as Dbase III

& IV, USP, SPSS, SAS, spreadsheet Lotus 123, etc. are in current use. The main objective is to computerise the statistical system and reduce the manual processing of data as much as possible.

3. Central Statistical Office Organisational Structure

Zambia has a centralised statistical system in which major statistical activities are undertaken by the Government's Central Statistical Office (CSO) whose head is the Director of Census and Statistics. The CSO has been and continues to be the major source of socio-economic statistical information in the country. It is empowered to carry out the statistical functions under an Act of Parliament Chapter 425 of the Laws of Zambia. It should be emphasized that, though some other organisations may be data producers of one form or the other, the CSO remains the major producer of statistical information in the country.

Since its inception in 1964, CSO has grown from a little statistical office into a large office with a wide range of statistics being produced in response to the increased demand for various socio-economic statistical information in the country.

Functionally the CSO is currently organised in three subject-matter Branches. Each Branch is headed by an Assistant Director. Branches are further sub-divided into Divisions and then Sections. The Branches are:

- (i) Agricultural and Environmental Statistics
- (ii) Economic and Financial Statistics
- (iii) Social and Cultural Statistics

In addition there are service units which service the subject-matter Branches, namely, Cartography, Information and Research, Data Processing and Administration. Furthermore, there are Provincial Statistical Offices, one for each of the nine provinces in the country. These offices are at present only data collection centres for the various CSO surveys and censuses. Each office is headed by a Provincial Statistical Officer (PSO) (who is usually a holder of a statistical diploma obtained after grade 12 level of academic education) and is manned by his Deputy, about 4 supervisors and not more than 30 enumerators. PSOs are answerable to the Director of Census and Statistics.

Moreover, there exist statistical units in some line ministries of the country which primarily service planning units in those ministries: Agriculture, Food and Fisheries; Education; Health; Development and Community Services; Mines; and Labour and Social Security.

The amalgamation of the Central Statistical

Offices Headquarters with all the nine provincial statistical offices and statistical units in the ministries is referred to as the Unified Statistical System (USS). The Director of Census and Statistics who is the de facto Head of the USS is also the advisor to the Government on matters pertaining to statistical improvements and developments in the country.

4. Economic Statistics Branch

The Economic and Financial Statistics Branch comprises mainly of the following Economic Statistics Divisions: Balance of Payments and External Trade, Public Finance, Industrial Production, Prices and National Accounts. Each of these divisions, normally headed by a Senior Statistician, periodically collect and compile data on different aspects of the economy which are primarily used in monitoring and evaluating economic development trends in the country and for planning purposes.

For Zambia the above mentioned subject areas are the ones in which establishment-based surveys dominate with the sampling and/or reporting units being establishments.

Data for these economic surveys are usually collected by means of mail-out questionnaires or schedules/forms. Examples of these surveys include the annual Census of Industrial Production; a quarterly enquiry of construction output in the private sector; a monthly enquiry among selected establishments in the Mining, Manufacturing and Electricity sectors to establish production trends; the annual enquiry of the services sector (wholesale, retail and related services trade); and a monthly enquiry on prices among selected establishments and market outlets. An establishment register which constitutes the Master Sampling frame of virtually most of the economic surveys conducted by the CSO, is maintained at the CSO Headquarters in Lusaka. The current practice is that questionnaires are dispatched by postal service from Lusaka.

The above mentioned surveys provide a vast body of data from which numerous statistics are produced. These include the wholesale and consumer price indices; the index of industrial production; balance of payments; Gross Domestic Product by kind of economic activity and by type of expenditure; and external trade statistics.

Invariably, the surveys for the different areas in the Economic and Financial Statistics Branch are designed, organised and conducted in a manner intended to serve primary data users in the specific areas concerned. However since these data also

constitute the building blocks for the construction of National Accounts, some consideration is often given to satisfy National Accounts requirements also. This is done by designing appropriate samples and survey instruments to provide for the required detail of data to meet the needs of both primary data users in the specific areas as well as the basic needs of National Accounts.

In trying to accomplish the above it becomes obvious that designing surveys will always be affected in order to ensure that National Accounts requirements are met, even though partially.

5. National Accounting in Zambia

Zambia adheres to the guidelines and recommendations set out by the United Nations System of National Accounts (SNA) in its compilation of the National Accounts. The revision of the 1968 version of the UN SNA is currently underway and the 1993 SNA draft has been published. Zambia, like most of the other countries using the UN SNA, has been following with keen interest the revision and has been participating in the debates on the revision through the UN Economic Commission for Africa.

5.1 System of National Accounts

National Income is a double flow, one physical and the other monetary representing the flow of goods and services on the one hand and the flow of money incomes or expenditures on the other, from producing to consuming units and vice-versa. Any transaction requires 'two' transactors, the one who pays and the one who receives. Indeed in the process of economic activity, producers and consumers enter into various transactions. The influence exerted by one group on the other and the way one reacts to change in the other depends on the extent of their inter-relationships.

Basically, a quantitative description of such inter-relationships is sought to be achieved through a system of interlocking accounts. National income accounting is therefore achieved through a double entry accounting system in which one transaction appears as a receipt in one account and again as a payment in another account. To facilitate this the economy is generally split into various components called sectors so that the accounts for the different sectors taken together will reflect the economic activity of the system as a whole. Five basic sector accounts identified for this, namely:

- (i) Enterprises/establishments or business sector-producing sector account
- (ii) Household sector-Consuming sector account

- (iii) The Government sector account
- (iv) The Capital Accounts
- (v) An account for transactions with the rest of the world.

The fundamental point to note is that construction of these five accounts demands a lot of data from numerous sources, and their collection and compilation should be systematic and consistent in order to produce useable National Accounts. It also requires the design of surveys which are comprehensive in scope and coverage. Since it is the enterprise/establishment sectors account which is of particular relevance to this conference, the following focuses on that account.

Enterprises or establishments are further classified into economic activities in accordance with the UN International Standards of Industrial Classification (ISIC) of economic activities. The following approach is used for the survey designs.

In each economic sector establishments are broadly classified as large and small ones. In general, all establishments which employ less than 10 employees or with capital valued at less than K250,000 in 1980 prices are classified as small. The rest of the establishments are referred to as large and these include all parastatal establishments irrespective of whether or not the value of capital owned is less than K250,000 in 1980 prices. Large establishments are surveyed on a 100% basis whereas small establishments are surveyed on a sample basis.

5.2 Sources of Data

It has been pointed out elsewhere that the data inputs used in the estimation of Zambia's National Accounts are generally obtained from a variety of sources. Divisions of CSO responsible for specific subjects provide most of the requisite data as indicated below.

- (a) The Agriculture and environmental statistics Branch provides production figures on Crops, Livestock, Fishing, etc. in respect of commercial and traditional farmers, etc.
- (b) The Census of Industrial Production (CIP) Division provides data relating to Mining, Manufacturing, Electricity and Construction industries. It also prepares various indicators such as index of industrial production.
- (c) Data for the services sectors of the economy is obtained through the annual National Income enquiry (see Section 5.4).
- (d) The Prices and Consumption Statistics Division provides a wide variety of statistical data on prices. Notable among these are the price

indices. Periodically it works out estimates of household expenditure surveys.

- (e) Balance of Payments and External Trade Statistics Division provides imports and exports data and any other transactions with the rest of the world. This data is mostly used in estimates of Gross Domestic Product by type of expenditure.
- (f) Public Finance Division supplies data on government revenue and expenditure and other information related to government activities.
- (g) Labour Statistics Division supplies data on employment and earnings of the various sectors of the economy.

It must be stressed, however, that some of these data are not in the format which is suitable for national accounting purposes and adjustments are made before incorporating them into the National Accounts.

In addition, use is made of data gathered through adhoc surveys such as Zambia's 1986 Labour Force Survey, Census data, etc. These will, in most cases, provide benchmark data. National Accounts statisticians have to make the best use of the available data to construct the components that make up the present National Accounts estimates for the country. The statistics compiled for National Accounts include the Gross Domestic Product by industry and expenditure which constitute the most important macro-economic aggregates for the evaluation of the performance of the economy. Furthermore input-output accounts have been constructed for the country for 1973, 1975, 1980 and 1985 calendar years.

Informal Sector

Though the informal sector activities have been recognised as an integral part of the Zambian economy, they are not adequately accounted for in the National Accounts. As is the case in many African economies, the nature of this sector and the lack of sufficient statistical capabilities renders the coverage of the informal sector problematic.

In Zambia, however, some economic activities in the informal sector, such as the traditional subsistence farming sub-sector of the Agriculture sector are covered on a regular basis using household surveys.

Informal sector activities are now receiving a lot of attention and specialised surveys have been identified as possible sources of data. Plans are underway to conduct a comprehensive survey of the informal sector in Zambia by next year.

5.3 Compilation of National Accounts

Gross Domestic Product (GDP) is the macro-economic measure commonly used in the evaluation of the performance of an economy. It is one of the main economic aggregates compiled in the National Accounts.

Generally, GDP is defined as the market value of goods and services produced within a given country in a given period, usually one year. Technically the GDP can be computed through three approaches, namely the Production (sometimes called output), Income and Expenditure approaches.

In the Production approach, the value of all final goods and services produced in the country during the year is measured. For each sector the production accounts are prepared to derive the value added of each sector. In the Income approach, the incomes that accrue from the factors of production provided by the residents of the country are summed up. The Expenditure approach counts up the total expenditures of the people, firms, government and local authorities. Some of this is expenditure on consumption goods, but some of it is investments in fixed assets, buildings, motor vehicles, plant and machinery, etc.

For Zambia, the Production and the Expenditure approaches are the ones normally used to estimate National product or the Gross Domestic product. The basis of Zambia's Gross Domestic Product figures is the Production approach because the data for this are much more easily available through the establishment surveys alluded to earlier.

As regards the Expenditure approach, all the components of GDP are normally estimated separately except for private final consumption expenditure which is obtained as a residual on the basis of GDP derived through the Production approach.

A firm basis for private consumption expenditures is usually a Household Budget survey (HBS). Unfortunately for Zambia the last comprehensive HBS was conducted in 1974/75. These Surveys are very costly ventures and the Zambian government has not been in a position to allocate sufficient funds for the conduct of HBS for a long time. Fortunately this year, a HBS will be conducted with funds provided by the Zambian government supplemented with donor funds and HBS survey instruments have been developed and are being tested. Data collection for HBS is scheduled to commence in July 1993 and will run for 12 months thereafter.

In Zambia, every effort is made to improve not

only to the quality but also the quantity of the statistics currently being produced. The guiding principle for a National Accounts Statistician is to strive to produce the best statistics with the data available.

5.4 National Income Inquiry

Insofar as data collection for National Accounts purposes in Zambia is concerned, mention should be made of the National Income inquiry. This inquiry is an annual survey whose coverage is confined to the non-government Service sectors of the economy. These include Wholesale and Retail Trade, Hotels and Restaurants, Transport, financial institutions, real estate and Business Services and Personal and Community services.

In Zambia this is the only survey wholly designed and implemented by the National Accounts Division in a manner which meets the National Accounts requirements, such as collecting data on components of gross output and intermediate consumption in the desired format. This facilitates computation of value added at establishment level and the subsequent summation at industry level.

The total number of establishments and/or enterprises covered under the inquiry has been around 2,400 over the last couple of years. Data collection is accomplished by mailing questionnaires to establishments/enterprises. The filled-in questionnaires are sent back to CSO, normally within 30 days of their receipt. Although the basic format of the questionnaires is similar, each questionnaire is intended to capture the characteristic economic activity (product) of each sector.

In recent years, the response rate for the inquiry has been generally on the decline. Follow-up exercises have been undertaken to raise the response rate to acceptable levels before summary totals per sector can be grossed up. A response rate of 60% and above is generally considered as an acceptable level.

Grossing up of the data to compensate for non-responding establishments and/or enterprises is normally achieved on the basis of employment data supplied by the Labour Division of CSO through the quarterly employment survey. It should be noted that grossing is confined sector-wise to the category of small establishments. Summary data totals of large establishments are not grossed because coverage for the large establishments/enterprises is on a 100% basis, and no effort is spared to ensure that data from all of them is

collected. Aggregated data of large establishments or enterprises is summed up with that of small establishments, (generally grossed-up) prior to undertaking analysis.

6. National Accounts and survey designs

In the Zambian context, in order to assess the basic relationship between the construction of National Accounts, and the various establishment surveys which provide data inputs for the Accounts, it is essential to initially examine the data sources referred to earlier.

It ought to be recognised that some of the data sources have other primary objectives and National Accounting objectives are actually secondary. Thus, some adjustments to the data are necessary before utilising them in the construction of National Accounts.

Due to the importance attached to the National Accounts in Zambia for evaluating the performance of the economy and for planning purposes, all the statistics currently produced within the Unified Statistical System almost always take into account National Accounts requirements to a large extent. In fact, the Director of Census and Statistics has a responsibility to ensure that National Accounts data requirements are satisfied as much as possible within the USS. This, however, does not necessarily imply overriding the major objectives for collecting and compiling the statistics concerned. In view of this, the design of surveys ensure that sufficiently detailed data are collected which are subsequently used to compute, for example, gross output, intermediate consumption and consequently value added accruing to the different sectors of the economy. Invariably this implies ensuring that sample sizes and appropriate survey designs are constructed to compute estimates which will have a national representation.

In most cases, adjustments will have to be made to some data to conform to National Accounting specifications. A case in point is with regard to industrial production. From the Industrial production point of view what is of paramount importance is the computation of Industrial Production Census value added. This concept is different from the National Accounts value added in that it excludes services transactions. Data on services transactions is collected by the Industrial Production Division though it is not directly used for evaluating and assessing the industrial production trends in the country. In the above example it can be clearly observed that in the design of the Census of Industrial Production for Zambia

both the Industrial Production and National Accounts requirements are considered. There are situations in which only basic requirements for National Accounts are met. In such cases other supplementary sources of data may be sought.

7. Establishment Based Surveys

An establishment is defined as the combination of activities and resources directed by a single owning or controlling entity towards the production of the most homogeneous collection of goods and services, usually in a single location but sometimes over a wider geographical area, for which separate data can be compiled in respect of production and all the intermediate inputs of labour and capital resources employed for a purpose, directly and in support of, or ancillary to, the production.

The history of conducting establishment based surveys in Zambia can be traced as far back as 1964 when the CSO in Zambia was established. These surveys constitute the core of statistical surveys which make up the data base for economic statistics including the Index of Industrial Production, Balance of Payments, Prices, National Accounts, and Labour and Earnings.

An establishment register exists at CSO Headquarters which has a listing of establishments classified by their main economic activity, in conformity with the UN's ISIC. As pointed out elsewhere this register is the master sampling frame from which different samples are drawn for the economic surveys. It should be stressed that at the national level it is only the CSO which has the capacity to carry out large scale sample surveys.

There are a number of problems associated with the establishment register. Notable among them is non-updating of the register on a regular basis. Zambia being a developing country does not have a lot of resources for frequent and regular updating of the register. However, efforts are generally made at least once every year to review the status of the register vis-à-vis closing down and opening of businesses during the year. Non-response is a major problem encountered in these establishment surveys. Factors which have been identified as major causes for non-response include the defective sampling frame, wrong postal addresses, unreliable postal systems and at times non-co-operating respondents. Measures instituted to reduce non-response include physical follow-ups, phoning non-respondents and carrying out educational campaigns through various media.

There are statistical units in some organisations which also produce statistics on establishments,

essentially to serve internal interests of those organisations. Notable statistical units are in the Central Bank of Zambia and Zambia Industrial and Mining Co-operation. In many instances these statistics are derived out of the administrative records kept by the organisations. As a matter of policy, line ministries in the country are being encouraged to establish statistical units for the purpose of processing and analysing the administrative data records under their custody.

8. Conclusion

National Accounts are among the most important statistical data or information which are in high demand by various data users in Zambia. It is imperative that requisite data inputs for the compilation of National Accounts are systematically and efficiently collected.

Establishment surveys form the core from which National Accounts building blocks are constructed. Consequently, improvements to the quality and coverage of establishment based surveys has been recognised to be a necessary continuous process. The CSO in Zambia will pursue this task diligently and conscientiously in order to enhance the usefulness of the National Accounts and other statistics, particularly those derived through establishment surveys. However, the major constraint to this effort is the limited financial and material resources.

References

1. Jere, A.T. "Basic economic statistics required for planning purposes Unpublished paper prepared for the economic commission for Africa (ECA) working group on improving the basic economic statistics required for planning purposes", Addis Ababa, Ethiopia, 19-23 September, 1988.
2. Mulenga, M.S.C. (editor). Occasional papers No. 2. A Workshop report on National Accounts of Zambia. Central Statistical Office, Lusaka, October 1980.
3. United Nations, "Manual on National Accounts at Constant Prices", Statistical papers, series M, No. 64, rev., 3, New York, 1968.
4. United Nations, indexes to the International Standard Industrial Classification of all economic activities. Series M, No. 4, rev. 2, New York.

A WORKABLE SYSTEM FOR UPDATING INDONESIA'S MANUFACTURING DIRECTORY

Alex Korns, Development Studies Project and Central Bureau of Statistics, Indonesia
Jl. Teuku Cik Ditiro 29A, Jakarta 10310, Indonesia

I. Background

Manufacturing -- the fastest-growing sector in Indonesia -- is estimated in 1991 to account for 21% of GDP, exceeding agriculture's share for the first time. Unique or nearly so among developing countries, Indonesia's Central Bureau of Statistics (CBS) each year conducts a complete census (using a 13-page form) of factories with 20 or more workers. A major problem with this data set has been how to update a Directory of 16,500 active establishments to keep pace with the appearance of an estimated 2,000 new ones per year.

In developing its manufacturing Directory, Indonesia, in common with many other rapidly industrializing countries, has wrestled with problems of under-coverage, double listing, and nonresponse adjustment. This report is presented in the hope that other countries may be able to benefit from the Indonesian experience. After an initial evaluation phase, CBS began to address the Directory problems in 1988 through piecemeal improvements, with mixed results. Faced with a need to monitor more closely the country's rapid manufacturing growth, CBS then opted in 1990 for development of an integrated Directory updating system designed simultaneously to address all the major problems. The system, which is being implemented in stages, has given satisfying results so far, with some "bugs" still remaining. The story of the development of the system is here told chronologically to help readers understand the problems encountered along the way, but readers impatient to know about the new system can jump immediately to section III.

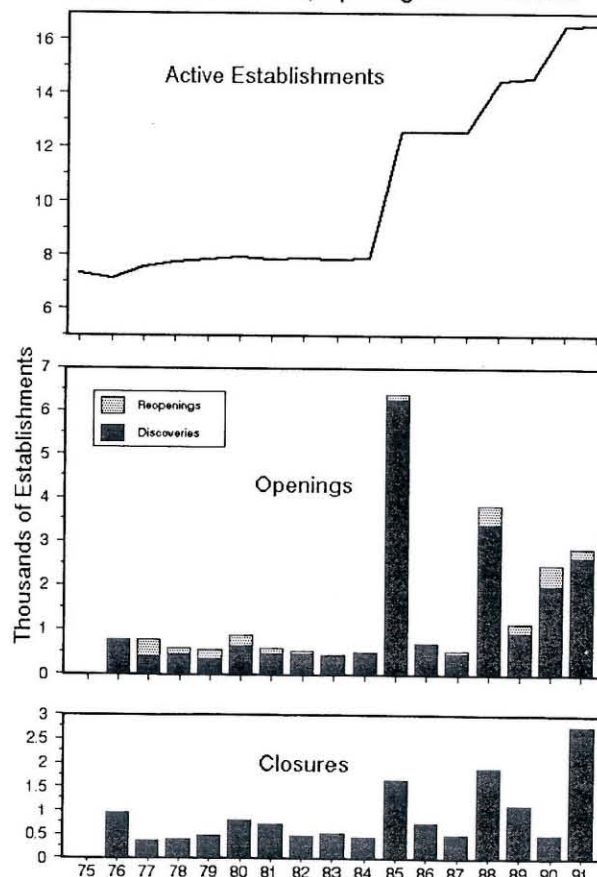
Diagram 1 shows Directory size and gross changes (discoveries, reopenings, closures) during 1975-91. On the face of it, the fluctuations appear erratic and unlikely to reflect fluctuations in economic activity. On examination of updating procedures, pre-1990 fluctuations can be divided into two kinds of periods.

* Years of perfunctory updating (1976-84, 1986-87, and 1989). Directory size remained flat, with openings roughly equalling closures, equal on average to about 6 percent of active establishments. During these years, CBS sent local offices a Directory of establishments in the previous year's survey and instructed field staff to deliver questionnaires to listed establishments as well as to all new/missed ones.

* Years of special efforts (1985 and 1988). For survey year 1985, as part of the 1985-86 Economic Census, door-to-door enumerators found over 6000 new establishments. For survey year 1988, enumerators made a special effort discussed below, finding 3300 establishments. During both years, reported closures also were unusually large (13 % of active establishments) but remained much smaller than discoveries.

In sum, pre-1990 fluctuations reflect variations in activity of the statistical agency, not in manufacturing activity. Even 1990-91 results, while reflecting progress toward a new Directory system, do not yet provide the desired measure of growth in establishment numbers. As the system gradually bears fruit, it is hoped that 1992-93 data may at last begin to measure growth in the number of establishments.

DIAGRAM 1: Manufacturing Directory 1975-91:
Active Establishments, Openings and Closures



II. Directory updating, 1975-89

A. Initial findings. A CBS team began in 1988 to study Directory problem, with help from the USAID-funded Development Studies Project. The study was triggered by a realization that the 1985 coverage jump had broken the time series. Initial field observations in areas of industrial concentration, showed that the most significant establishments (in terms of employment) missed before 1985 had been quite visible. Enumerators often knew where they were. Nevertheless, undercoverage was widespread, for several reasons.

- * Although enumerators were exhorted to deliver Annual Survey questionnaires to all new/missed establishments, management lacked tools for checking enumerator compliance.
- * The only procedure for reporting a new establishment was to have it fill out an Annual Survey questionnaire. There was no channel for reporting the existence of new nonrespondents (although a procedure existed for estimating nonrespondents already in the Directory).
- * Enumerators were evaluated on the percent of their "target" (calculated as their share of the existing Directory) for which they got completed Annual Survey questionnaires, not (for lack of a standard) whether they reported new establishments. They were unenthusiastic about reporting new establishments, which were often uncooperative, because they foresaw having to return over and over at their own expense to obtain a completed questionnaire.
- * Enumerators received no payment for searching for new establishments, although for other field work they usually got piece-rates that reimbursed for both time and transport costs.

B. Conceptual issues. The finding that the survey suffered from undercoverage raised conceptual issues. Directory updating rules in force before 1988 had been formulated in 1975, when the share of industry in GNP was only 9 percent. The rules were intended to free enumerators from having to annually recheck employment at thousands of small factories that might have crossed the 20-worker threshold. The rules were: Once entering the Directory, establishments remained in it until the next Census even if employment dipped below 20, while establishments outside it that once had fewer than 20 workers were not in scope (even if they rose above the 20-worker threshold) until the next Census. In light of the rules, CBS suspected the 1985 flood of new establishments consisted mainly of ones that had been out-of-scope before the 1985/86 Economic Census.

Casual field inquiries showed, however, that enumerators often ignored these rules in reporting new/missed establishments.¹ Recognizing that users wanted data for all establishments in any year regardless of their previous employment history, CBS decided in 1988 to replace the old rules for Directory membership with new ones based on employment in the current survey year only. As a result, enumerators could no longer excuse failure to report a new establishment on the ground that it may just possibly have started with fewer than 20 workers. The new rule was adopted in recognition that it would be inefficient annually to recheck all borderline establishments; hence, enumerators might sometimes fail to notice an establishment crossing the 20-worker threshold.

C. Jabotabek pilot. A pilot study was conducted in seven industrial districts in and near Jakarta to test alternative techniques for identifying new establishments. BPS staff at provincial and kabupaten levels matched the BPS Directory with establishments at other agencies, including the Department of Industry. About 1900 "candidates" for addition to the Directory were obtained through matching and field checked; about 900 qualified for addition to the Directory.

The study showed that other agency lists were very useful for updating the Directory. However, inasmuch as none of them were either complete or accurate, they could not be taken at face value. Field checking showed that some candidates did not qualify for the directory because they had once existed but were already closed, were under construction, had fewer than 20 workers, or existed only on paper and were never realized at all. Accordingly, other agency lists in Indonesia, which are based on permits that reflect an intention to undertake activity, cannot serve the same benchmark function as tax rolls in the United States and some other countries, which reflect ongoing activity.

D. Coverage of new establishments. In an attempt to reduce undercoverage in the 1988 Annual Survey CBS implemented five nationwide changes intended to improve coverage, beside the Jabotabek pilot:

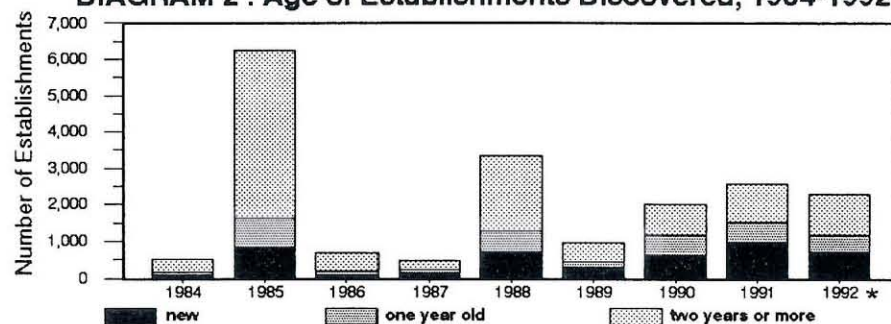
- * A small "finder's fee" paid to enumerators nationwide for reporting new establishments.
- * A sharp increase in the piece-rate for obtaining a completed Annual Survey questionnaire, so that enumerators would feel less burdened by the addition of new establishments to the survey.
- * The announcement of the new, simplified, rule that made coverage dependent only on having 20 or more workers in the survey year, as

mentioned above.

- * A new procedure for listing establishments even if the enumerator is unable to obtain a completed Annual Survey questionnaire.
- * A new updating manual recommending matching with other agency lists.

The big coverage gain for 1988 (3300 establishments, of which 1000 were in the Jabotabek pilot area) proved there had been widespread undercoverage and that field staff could find many new/missed establishments without a costly door-to-door canvass. The gain raised hope that the coverage problem had been solved -- a hope tested by the 1989 survey, which kept all five nationwide innovations introduced in 1988. Disappointingly, only 967 establishments were discovered in 1989, only 297 of which began production in 1989. The sharp decline of discoveries from 1988 appeared in part to reflect a drop in field staff enthusiasm for updating due to preoccupation with preparations for the 1990 Population Census.

DIAGRAM 2 : Age of Establishments Discovered, 1984-1992



* 1992 data are preliminary and incomplete

When 1984-89 discoveries were analyzed by cohort (year starting commercial production), discoveries of truly new establishments were seen to have fluctuated sharply (dark block in diagram 2). In years of special efforts the number reached 800. In years of routine efforts before 1988, the number never topped 135; many establishments discovered had begun commercial production some years before discovery. From the cohort analysis and analysis of the 1988-89 fluctuation in discoveries, CBS concluded that the key to solving the coverage problem was to devise a procedure satisfying two requirements:

- * The procedure should discover most new establishments in the year they begin production, or at most one year afterwards.
- * The procedure should operate year-in year-out with equal effectiveness regardless of staff enthusiasm.

E. Closures and doubles. The unusually large number of closures and reopenings in the 1988 results drew attention to two other problems.

First, many closures and reopenings were due to churning of double-listed establishments that had entered the Directory in previous years. In most cases, only one of the partners in a doubles pair was shown as active in any year while the other was shown as closed. Whereas before 1988 enumerators received lists of active establishments only, in 1988 they got lists of all ever-active establishments with instructions to check which ones were still active. For doubles, they responded by designating one "partner" as active and the other as closed; in many cases, the one designated as active happened by chance to be the one that was inactive the year before, while the one closed had been previously active. The discovery of this churning led to a year-long study that in 1991 identified about 5.5 percent of the 25000 ever-active establishments in the 1989 Directory as double-listed.

In itself, double listing did not much affect the accuracy of Annual Survey tabulations, because both entries were seldom recorded as active at the same time (one being based on real data, the other estimated as a nonrespondent). However, the doubles did clutter the Directory and disturb the CBS project for backcasting synthetic histories for establishments missed in earlier years, and were,

accordingly, purged from the Directory. In diagram 1, openings and closings are calculated on the basis of a merged time series for each pair of doubles.

Second, field checks of 1988 closures showed that a few active establishments had been reported as closed. Because closure reports were transmitted informally, leaving no paper trail, responsibility for this mistake could not be assigned. The inverse error also took place. That is, newly closed establishments were sometimes treated as active nonrespondents for a year or more after closure and inputted during Annual Survey tabulation. This error was impossible to trace because active nonrespondents were not positively designated; rather, they were a residual set.

F. Software for updating. Some problems with the old Directory updating system can be traced to software limitations. Before 1988, all Directory updates were entered into the CBS mainframe; listings were sent as

hardcopy to the provinces for distribution to districts. Because the procedure for transmitting Directory corrections from districts to provinces and thence to the head office was cumbersome and ill-defined, field staff was often surprised by discrepancies appearing in next year's Directory. On the hard copy, establishments were ranked by area and SIC; no alphabetical listings were available. These procedures sometimes made it difficult for staff to check whether an establishment was already listed or not.

In 1988, in the context of administrative decentralization, CBS provided statistical offices in each of Indonesia's 27 provinces with software (using an overlay on DBASE3) for updating their own Directory. The software opened a channel for reporting the existence of new nonrespondents. However, it lacked summary tables and frequently failed to function due to viruses, technical misunderstandings and long communication lags with CBS. Provinces updated their Directories inaccurately and tardily, because they lacked procedures for checking Directory accuracy and felt they were doing the task for the head office not for themselves. Consequently, head office staff was obliged to edit provincial diskettes to assure consistency with the Annual Survey, returning diskettes to provinces for use the following year. For example, because provinces sometimes sent questionnaires for new factories but failed to add them to their Directories, head office staff added these new names. In the course of editing, some new mistakes such as new doubles inevitably occurred. Some corrections conveyed to the head office by letter or word of mouth were ignored in editing for lack of standard procedures. A vicious cycle arose whereby the provinces complained of inaccurate CBS editing and absolved themselves of responsibility for a clean Directory.

III. The new system, 1990-92

Faced with the need more closely to monitor Indonesia's rapid industrial growth, CBS decided in October 1990 to develop a new, integrated Directory updating system. The system was implemented in stages, with complete nationwide implementation beginning in calendar 1993 for survey year 1992. Diagrams 3-4 depict its major elements.

A. Matching. A standard procedure was developed for annual Directory updating, based on nationwide matching with establishment lists of other agencies to select a list of candidates for field checking. To assure consistent updating efforts from year to year, CBS assigns an annual target number of candidates to each province. Matching is mainly carried on at two levels, provincial and district, with limited matching at

CBS. Provincial statistical offices match other agency lists with an alphabetized Directory; unmatched names are designated "candidates," sorted by district, and sent to districts. District offices similarly match other agency lists; unmatched names not present in the provincial list of candidates are designated additional "candidates." All candidates are sorted and given to enumerators for field checking.

The target number of candidates was set at 60 percent of the active Directory for areas implementing matching for the first time, and 35 percent for those implementing it for a subsequent year. If the initial number of candidates is below target, supervisors are told to add lists from other agencies. If, however, the initial number of candidates is above target, supervisors are told to discard certain low-priority groups of candidates, such as establishments below or slightly above the 20-worker threshold or unincorporated ones in subsectors usually dominated by small industry (e.g., tofu and tempe). Supervisors are told to review each year the "productivity" of each source (agency list) and give preference to more productive ones.

Field staff are encouraged spontaneously to report new/missed factories directly observed in the field but not derived from matching. Supervisors, however, are told to treat a high share of direct observation in a district as an indicator that matching may need improvement, because only good matching can provide management with the tool to guarantee consistent updating results independent of fluctuations in enumerator enthusiasm.

B. Field checking. Field checking of candidates is done with a standard form (I-B) that was pilot tested for two years. It requires the enumerator to check one of 9 findings that can be divided into three groups:

1. Establishments engaged in commercial production, with 20 or more workers during the survey year. These establishments will be added to the Directory unless found to duplicate ones already in it.
2. Three codes for establishments not yet engaged in commercial production but needing to be rechecked the next year. This includes factories in trial production, factories under construction, and vacant lots with a signboard indicating that a factory will be built there.
3. Five codes for establishments not fulfilling conditions for the Directory and not needing to be rechecked the next year. This includes small establishments, ones located in another district, non-manufacturing ones, ones that once existed but are now closed, and candidates that cannot

be located.

For candidates in the first group, enumerators collect data for main product, employment, and the year starting commercial production from either the establishment itself or (for uncooperative establishments) guards or neighbors. For all candidates, enumerators fill out a form and receive a small payment. Supervisors check whether candidates in the first group are already in the Directory. Provinces enter data for all candidates in a database using uniform CBS software, which transfers successful candidates to the Directory and prints lists of candidates that need/don't need to be rechecked the next year.

Nonactive establishments, all of which the old software had designated as "closed," are now designated with three codes: Closed, small and obsolete. Obsolete includes three groups of drop-outs: Doubles, mergers, and establishments which had been added during updating but turn out not to qualify for the Directory upon administration of the Annual Survey. Despite safeguards in the new system, a few doubles were expected to continue to appear. However, the use of form II-B was expected to automatically lead to all doubles receiving an "obsolete" code. To minimize doubles the rule for assigning identification numbers to new establishments was changed. Starting in 1991, the ID was based on the originally reported SIC,

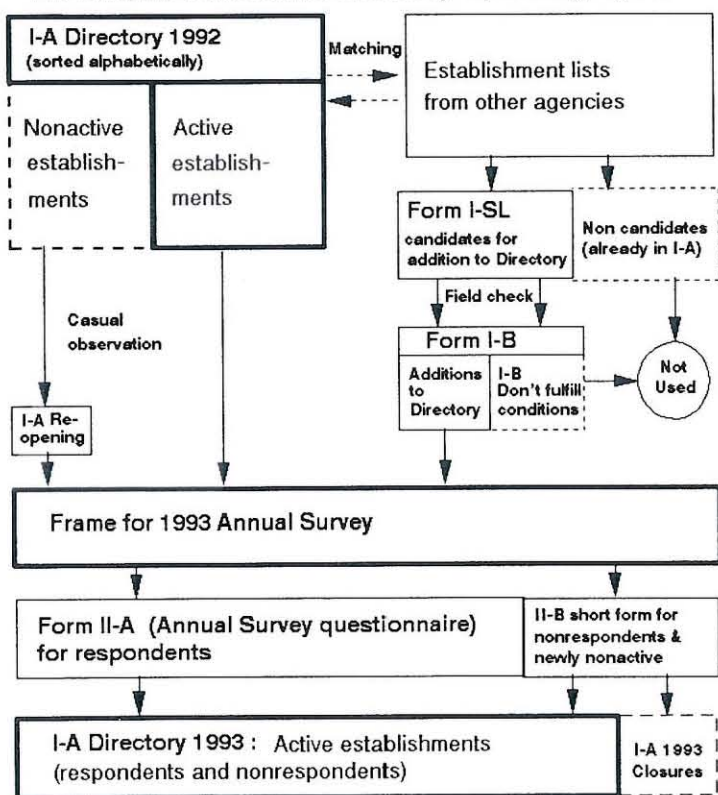
whether or not it subsequently changed. Previously it was based on the current SIC, so that a change in main product (or recoding) forced staff to "close" an establishment and "open" a new one. The new rule avoided churning ID's while preserving a mnemonic basis for most ID's (except in the case of an SIC change).

D. Software. A new applications package, distributed to provinces in September 1992, was based on Clarion software, which provided more advanced and user-friendly capabilities than DBASE3. The package was designed to serve management needs of provinces and dovetailed with an illustrated manual for field implementation of the new updating system. The software allows the Directory to be accessed in three sequences: Alphabetical, regional, and sectoral (SIC). Newly discovered establishments automatically received new ID numbers. Modules were added for provinces to monitor receipts of forms II-A and II-B and shipments to CBS. Tables were provided to monitor work flows. These administrative conveniences were intended in part as "bait" to stimulate provinces to update their own Directory quickly and ensure consistency of ID numbers between Directory and Annual Survey questionnaire.

Separate modules were provided for backing up and regenerating data:

- * The backup module was expected to stimulate provinces to backup data often -- the key to data security in a virus-rich environment.
- * The regeneration module enabled provinces independently to make the transition from one survey year to the next. It converted "new" codes to "old" ones and erased receipts/

DIAGRAM 3 : Annual Directory Updating Cycle



Notes: Form I-A is the Directory
Form I-SL is for listing candidates for the Directory
Form I-B is for field checking candidates for the Directory
Form II-A is the Annual Survey questionnaire
Form II-B is for reporting nonresponse and newly nonactive

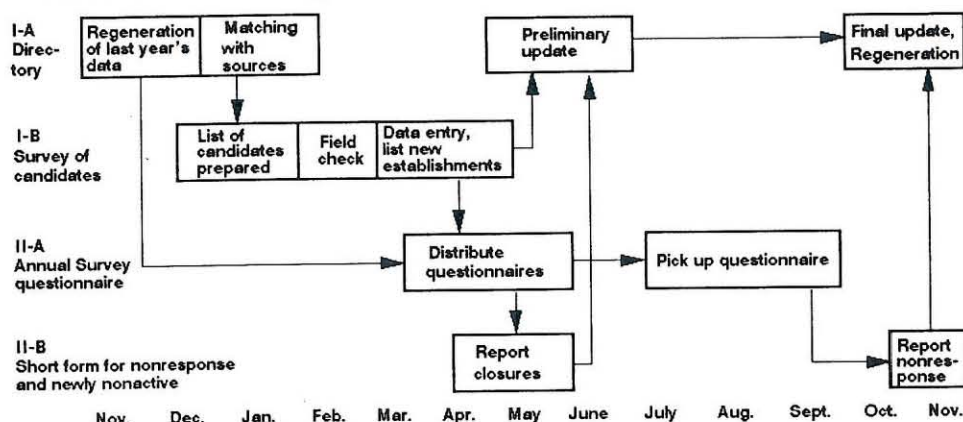
C. Reporting of closures, small, doubles. To prevent closures and nonresponse errors, a second form (II-B) was introduced for enumerators to report all active establishments that fail to submit an Annual Survey questionnaire (II-A). As shown in diagram 3, the form closes the set of active establishments, thereby providing positive designation of which nonresponding establishments are still active and which have closed.

shipments data.

Simultaneously, provinces were given full responsibility for Directory updating. CBS pledged it would never again edit provincial Directories, although it would check them for quality and provide hard copy feedback.

were new and 19 percent were one year old. For 1991, when all Java experienced systematic updating, regional differences in the age composition of discoveries may in part be due to real differences in growth patterns. For example, most Botabek growth is due to new (green-field) factories, whereas much growth in the older Jakarta area reflects incremental growth at smaller, older factories.

DIAGRAM 4 : Timing and Interrelationship of Main Field Activities



The number of discoveries of newly-commercial establishments in Java in 1991 (905) is unprecedentedly large. It is equal to 6.7 percent of the number of active establishments for Java in 1990 (13,500), which compares reasonably well with our presumption, based on fragmentary evidence, that the true annual number of

IV. Pilot results, 1990-91

A. Discoveries. The system for discovering new establishments was tested in 10 industrial districts of Jakarta and West Java in survey year 1990 and throughout the island of Java (with 75 percent of manufacturing) in 1991. Results indicate that the system reliably turns up many new or one-year-old establishments. In its first year of operation in any area, matching can be expected to turn up many establishments that were previously missed; such delayed discoveries will presumably decline in subsequent years. As seen in diagram 5, matching in its first year in all areas has shown a jump in the number of discoveries and, usually, in the ratio of new establishments to total discoveries. Moreover, for areas subjected to matching over several years, there is some tendency for the latter ratio to rise -- an indicator of shrinkage the pool of missed establishments. Jakarta, which experienced a pretest of a matching system in 1988, showed improvement in 1990 and further improvement (in the ratio) in 1991, when 28 percent of discoveries were new establishments and another 28 percent were only one year old. The fast-growing Botabek suburbs, which also joined the 1988 pretest, showed sharp improvement in 1990 and again in 1991, when 73 percent of discoveries were new. The city and district of Bandung, which first matched in 1990, showed a big improvement in 1991, when 50 percent of discoveries were new. For the rest of Java, 1991 was the first year of systematic updating: 36 percent of discoveries

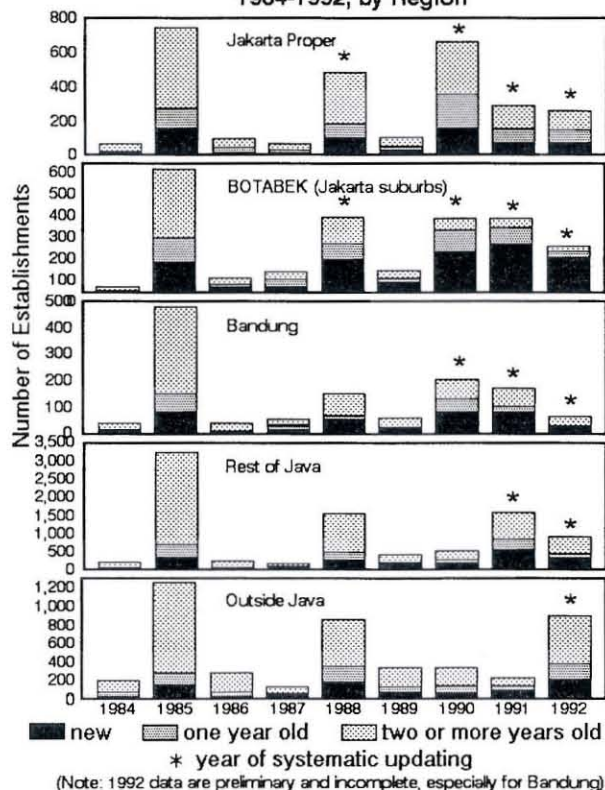
new establishments may be about 10 percent of the number of active establishments. In fact, the 1991 measure may even be more complete than appears at first glance, if allowance is made for establishments starting production in 1991 but not yet qualifying for the Directory because their employment is still less than 20. Most 1991 I-B discoveries were found through matching. The 22 percent found by direct observation constitute a reasonable share, although too high in a few districts. Enumerators have gained a new awareness that CBS expects zero undercoverage and show a new enthusiasm about reporting new or missed establishments.

The average 1991 discovery had 107 workers, much smaller than the 161 workers at the average establishment in the 1990 Directory. Establishments discovered in their first year of commercial production tended to be about a third larger than ones discovered after two or more years of commercial production.

B. Process of discovery. Tables in the software sent to each province enable them to evaluate their own updating efforts and provide insights into the discovery process. Total discoveries in Java (purged of false ones, discussed below), 2400, were equal to 39 percent of the 5950 candidates checked. Of the 61 percent that failed to qualify, 690, or 11 percent of candidates, were slated for rechecking the next year because they were found to be in trial production, under construction, or holding a vacant lot and signboard.

The success rate (discoveries as a percent of candidates) varied from 15 to 60 percent in industrial districts and even more widely in nonindustrial ones. The variation reflects district variations in the skills of CBS officials in selecting candidates, the quality of other agency lists, and underlying (and unknowable) real factors such as true growth rates, the backlog of missed establishments, and (inversely) the ratio of the target to the number of factories awaiting discovery.

DIAGRAM 5: Age of Establishments Discovered 1984-1992, by Region



The findings indicate that the target number of candidates everywhere needs considerably to exceed the expected number of discoveries, to allow for error in both our expectations and in the selection of sources and candidates. Beyond a certain minimum target, the likely effect of further raising the target is probably to increase discoveries of borderline establishments with 20-30 workers. Even then, the marginal yield in discoveries probably declines sharply as the number of candidates gets very large. Surabaya, a central city with few new factories, discovered only 134 establishments after checking 886 candidates (equal to an extraordinary 1.44 times the number of active establishments in its Directory).

The major weakness of the updating system in 1991 was its tendency to draw into the Directory many false discoveries -- about 8 percent -- that were deactivated

later in the year because they turned out to be double or small. (False discoveries are not included in the data underlying diagrams 1, 2 and 5). Supervisors have been asked to tackle this problem in two ways:

- * Many doubles go by the same name and can rather easily be avoided if provincial and district supervisors will merely insist on careful cross-checking with the existing Directory. Some doubles go by different names or addresses and require careful checking at the enumerator level.
- * Borderline small factories in seasonal industries (e.g., garments) may report 20 or more workers on the I-B then report fewer than 20 on the II-A long form. For such industries, discrepancies can be avoided by asking I-B respondents to fill out a monthly employment worksheet like the one in the II-A.

C. Software. The provincial statistical offices welcomed the new software received in September 1992, using it to enter I-B results and update Directories promptly. Many offices used the module for recording Annual Survey questionnaire receipts but few used the one for recording shipments to CBS. Rapid acceptance of the new software has been supported by three factors:

- * CBS staff keeps in close contact with provincial offices to resolve problems caused by viruses, incompatible DOS, misunderstandings, etc.
- * Provinces are asked to bring their diskettes to interprovincial meetings and send them to CBS at certain deadlines. Most provinces sent their 1991 diskettes twice to CBS, in October 1992 and March 1993.
- * Provinces have been trained in the use of various printouts -- e.g., how to administer form II-B by sending to each district a printout of establishments for which documents have not yet been received.

Major industrial provinces each dedicated a PC for the Directory. Viruses have caused program failures in a few provinces; these were quickly repaired except in small, remote provinces. No province has yet irretrievably lost a substantial amount of data.

D. Receiving at CBS. CBS improvised a system for recording receipts of Annual Survey questionnaires from provinces using the software designed for provincial use. The receiving officer was instructed never to correct the Directory, but merely to compare each incoming questionnaire with it, as follows.

- * Questionnaires with name, address and ID consistent with those in the Directory are merely logged in.
- * Questionnaires with missing or discrepant ID are logged in and ID numbers are entered or corrected.
- * Questionnaires from establishments not in the Directory are held to await updated provincial Directories.

As Provinces sent updated Directories on diskettes, the receiving officer merged them with the receipts data already logged in, and resumed receiving. He recorded receipt of 13,500 questionnaires for the 1991 survey, providing the Industry Division, for the first time ever, with an electronic checklist of which questionnaires had been received. At the end of the survey year, 38 unresolved discrepancies were found: 13 establishments that appeared active in the Directory but were in reality "nonactive: small," and 25 new establishments that had not yet been entered by provinces. Moreover, the receiving officer was obliged to personally update the Directory for one small province with little industry. These discrepancies are trivial compared to previous years, when the head office entered hundreds of new establishments and corrected activity status for hundreds more. The new system has thus succeeded fully in shifting updating responsibility to provinces and eliminating the creation of doubles at the head office.

E. Closures and nonrespondents. The system for positively reporting closures and nonrespondents with form II-B, first tested in survey year 1991, succeeded in "closing the set" of active establishments; almost all nonrespondents in Java were positively confirmed. The overall national nonresponse rate was only 17 percent of establishments. Considering the negative attitude of many establishments toward surveys, this is great testimony to the perseverance of CBS field staff assigned individually to each of 3500 sub-districts.

Field staff needs additional training in the use of form II-B, to minimize tardy and inaccurate reporting of closures. A study has been launched to "score" the accuracy with which enumerators filled in form II-B in 1991. Preliminary results show that the month of closure was not reported for a large share of closures; for many of these the year of closure reported, 1991, is merely the survey year and not the actual year of closure. Results also show that, due to a combination of misunderstanding and oversight, district-level supervisors failed to check the closure of many large establishments, as required by form II-B.

The big surprise in 1991 was the unprecedentedly

large number of newly nonactive establishments: 3120, or 18.8 percent of the previous year's Directory. Preliminary findings show:

- * The number of newly closed (1790) and newly small (620) is well above the rate for a normal year (about 1000 closures and a few hundred newly small), and, as a joint percentage (14.8 percent), exceeds the 13-percent rate in two previous peak years, 1985 and 1988. Many of the closures (690) occurred in Jakarta province, where they equalled an extraordinary 26 percent of the number of previously active establishments. Jakarta statistical officials attribute the flood of closures to form II-B, which forced field staff to report closures overlooked in previous years. This hypothesis is supported by tabulations showing that, among 1991 Jakarta closures, 69 percent did not complete Annual Survey questionnaires in 1990, and 37 percent had not done so during 1988-90.
- * The number of newly obsolete, 700, is legitimate but startlingly large and due to two factors that are being explored in a CBS study of 600 newly obsolete establishments in Java. First, enumerators identified 77 doubles missed in the 1989 doubles study, and 83 that were added to the 1990 Directory. Second, enumerators identified 339 1991 discoveries as false ones, of which 243 turned out to be double, and another 96 were discarded for other reasons. Of the 243 doubles, 133, over half, cropped up in Central Java. Preliminary indications are that over half of these doubles were not discovered in the new updating procedure, with form I-B, but were entered directly into the Directory. This happened when districts sent provinces Annual Survey questionnaires (II-A) with missing or outdated ID numbers (BPS reassigned ID's in 1991). Provinces tried to match the establishments by name but sometimes failed, as for establishments with multiple names. Stymied, provinces then assumed the II-A's were for new discoveries, instead of returning them to districts for more positive identification as they should have done.

V. Preliminary national results, 1992

Preliminary results for 1992, shown in diagrams 2 and 5, are based on August 1993 data that remain highly incomplete for some West Java districts (including the Bandung area) and 7 remote provinces outside Java. Despite this, the expected surge is seen in discoveries outside Java, where systematic matching was applied for the first time.

A remarkable feature of the preliminary 1992 data is that firms discovered in the first year of production in Java fell sharply from 905 in 1991 to 542 in 1992. The latter figure is unlikely to rise above 600 even when West Java returns are complete. It is tempting to speculate that the decline reflects a 1992 slowdown in industrial growth, related to tight money policy. A decline in new factories would be consistent with the observed sharp decline, starting in mid-1991, in monthly project approvals by the Investment Board.

However, the possibility cannot be excluded that the decline in discoveries of new establishments partly reflects as well a decline in the number of candidates that were field checked during Directory updating in Java, from 5950 in 1991 to about 4000 in 1992. The latter decline is due to three factors: the use of lower, second-year targets for East and Central Java, further adjustments to targets for budgeting reasons, and difficulty in finding enough candidates to fill the target in Jakarta and Central Java. The difficulty in interpreting the 1992 data draws attention to the need for consistency and appropriateness in target determination and for studies of the impact on discoveries of fluctuations in the number of candidates.

An interesting feature of the results is that when 1992 discoveries of establishments in Java in their second year of commercial production (263) are added to 1991 discoveries of first-year establishments (905), the total (1168) is equal to 8.6 percent of the 1990 Directory for Java -- surprisingly close to the presumed 10 percent rate of appearance of new establishments that was mentioned earlier.

Operationally, the 1992 survey year marked the first time that the Manufacturing Statistics Division assigned a subdivision to the main task of supervising Directory maintenance. Supervision focused on the 22 provinces outside Java, where the new updating system was implemented for the first time. Time and funds were found for visits to 9 provinces outside Java, in some of which serious implementation flaws (including the omission of key steps in the updating process) were discovered and corrected. By mid-August, 21 of 27 provinces had submitted the preliminary Directory diskette due July 1, some in response to telephone prodding, and 18 had used the software module for recording shipments of Annual Survey questionnaires to the head office. Analysis of the diskette data alerted the head office to implementation problems in several provinces, which triggered phone calls and letters from the head office. To spur provinces to pay more attention to quality, the head office has decided to take the unprecedented step of "scoring" the quality of responses to forms I-B and II-B.

The 1992 experience shows that the integrity and timeliness of the updating system cannot be taken for granted but can only be guaranteed through continuing attention to the details involved. This requires intensive communication between the head office, provinces and districts, by visit, circular letter, telephone and meeting. It also makes heavy demands on the time of the Manufacturing Statistics Division, and of provincial and district statistical offices.

VI. Summary

Indonesia's rapid industrial growth has obliged CBS to tackle the difficult problem of annually updating its manufacturing Directory to keep pace with the appearance of new factories. The new CBS Directory updating system was developed after piecemeal improvements were tried, leading to often frustrating, trial-and-error, learning. The new system makes integrated, mutually-supporting use of user-friendly, management-oriented software, two new field instruments (forms I-B and II-B), and systematic matching with other agency lists. By these means, the system enables most new establishments to be quickly entered into the Directory, while all closures and nonrespondents are positively identified. The field instruments and software provide many checks and balances to prevent errors and assist management to detect and correct them when they occur. Provincial statistical offices find it convenient promptly to update their Directories and share printouts with district offices. Provincial offices also find themselves saddled with some of the perplexing coordination and supervision problems that previously found their way to the head office. The head office plans to provide regular feedback to provinces to help them master the new tasks.

The new system can also be viewed in the context of statistical development in a developing country. The new system aims for a breakthrough to higher data quality by simultaneously tackling a tangle of institutional obstacles. More specifically, the new system targets six obstacles to successful Directory updating: The lack of a clear system of tasks and responsibilities, and the lack of performance indicators, of communicative software, of training, of funding, and of confidence in the system by participants. However, one remaining obstacle still handicaps the new system: a shortage of management time, day-in and day-out, at national and provincial and district levels, to coach participants in their tasks and to monitor and troubleshoot performance flaws. CBS is trying to assign additional staff to the task, but is constrained by institutional obstacles to resource reallocation.

Although quantitative results for updating conform in

large part to our expectations and show great improvement over previous performance, the results do not yet begin to provide a direct, reliable measure of growth in establishment numbers; the extraordinary number of closures reported in 1991 prevents that. As the new system continues to be implemented in a consistent manner, it is hoped that 1992-93 data may at long last begin to provide measures of growth in establishment numbers, providing in the process critical input for adjusting the quarterly manufacturing survey for the influence of growth in establishment numbers. As a follow-on project, CBS is developing a scheme for downloading provincial Directories to district statistical offices, for use in a management information system to monitor field work for three manufacturing surveys: The Annual Survey of Manufacturing, the Quarterly Survey of Manufacturing Production, and the Quarterly Survey of Manufac-

turing Wages. The MIS will provide a tool to help management accelerate document flow and check data quality. Plans are also taking shape for creating an integrated directory of non-industrial businesses.

The Indonesian Directory updating system would appear to differ from the US Census Bureau's SSEL in two respects. First, CBS cannot rely on matching alone for updating, due to widespread deficiencies in other agency lists. That is, no other agency list is complete enough to serve as a definitive benchmark; all such lists constitute nothing more than input for CBS examination. Second, CBS has abundant staff, so it can afford to field check every "candidate" as the Census Bureau cannot. Indonesia's workable system for directory updating may be suitable for use in many other developing/transitional countries.

1. Final proof that enumerators had often ignored the rules was not forthcoming until a "backcast" of establishments discovered by the Annual Survey some years after they began production revealed that over 40 percent of establishments discovered in 1985 (with over 75 percent of employment at newly discovered establishments) began production before 1985 with 20 workers or more. That is, they were in-scope before 1985 but had been missed by enumerators.