OCCASION

This publication has been made available to the public on the occasion of the 50th anniversary of the United Nations Industrial Development Organisation.

DISCLAIMER

FAIR USE POLICY

CONTACT

Please contact publications@unido.org for further information concerning UNIDO publications.

For more information about UNIDO, please visit us at www.unido.org

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
Vienna International Centre, P.O. Box 300, 1400 Vienna, Austria
Tel: (+43-1) 26026-0 · www.unido.org · unido@unido.org

16009

A REPORT FOR UNIDO

SIZE DISTRIBUTION STATISTICS IN MANUFACTURING FOR DEVELOPING COUNTRIES

John Weeks
The American University
Washington, D. C.
November, 1986

SIZE DISTRIBUTION STATISTICS FOR MANUFACTURING IN DEVELOPING COUNTRIES

TABLE OF CONTENTS

	page
PART I: SIGNIFICANCE OF THE STUDY	
I.1 INTRODUCTION	1
I.2. USES OF SIZE DISTRIBTUION DATA	2
PART II: COMPILING THE DATA AND GENERATING STATISTICS	
II.1 COUNTRY COVERAGE	6
II.2 PREPARATION OF DATA	
II.2.1 General Comments	12
II.2.2 Presentation by ISIC Sectors	13
II.2.3 Presentation by Size Ranges	14
II.2.4 Presentation of Indicators	15
II.2.5 The Standard Table	16
II.3 PRESENTATION OF STATISTICS	
II.3.1 Publication Format	17
II.3.2 Providing Analytical Tables	20
PART III. SUMMARY	21
APPENDIX: Guide to Country Data (separate table of contents)	24

PART I: SIGNIFICANCE OF THE STUDY

I.1 INTRODUCTION

This study has two purposes. First, it provides a survey of available data on the size distribution of establishments in manufacturing in developing countries,[1] and suggests guidelines for how these statistics might be presented in a useful manner by UNIDO to a wide audience of researchers. Second, it brings together data of this type as an addition to the UNIDO reference collection. No centralized bank of size distribution data exists,[2] and this report takes an important step towards creating such a data bank.

Given that institutional funds for elaborating this project might be limited, one strong point of this report is that it recommends a process of data collation and statistics generation which can reasonably be achieved within the resources currently available to UNIDO. The basic data for beginning a regular publication size distribution statistics is provided in the appendices to this report or currently exists in UNIDO. Converting these data into useful statistics is for the most part a simple exercise requiring little specialized expertise.

---

[1] Throughout this report, "size distriubtion of establishments" will refer to the disaggregation of firms in a sector or industry by ranges of numbers employed -- e.g., establishments hiring 5-9, 10-19, etc.

[2] In 1979-1981 the World Bank financed a project which included the collection of size distribution data. However, this project was subsequently abandoned. See John Weeks, "Methodology for Production of Consistent Industrial Statistics for Developing Countries," (Washington: World Bank manuscript working paper, September 29, 1979), Economic Projections and Policy Department.

This report will be self-consciously practical, organized to provide a step-by-step description of how this new and unique data bank can be developed at minimum cost, yet yield a wealth of information for researchers. As argued in the next section, size distriubtion data are an immeasurably rich source of information about the nature of manufacturing activity, allowing one to test as hypotheses issues which usually are treated as a priori assumptions. UNIDO is here breaking new ground, providing a statistical source equivalent in importance to the first UN compulation of national income aggregates or trade data across countries.

I.2. USES OF SIZE DISTRIBTUION DATA

Size distribution data are a potentially rich source for the testing of a number of important hypotheses relevant to development policy. Particularly in recent years, emphasis has shifted from equity and distributional aspects of economic development to considerations of efficiency. So-called "structural adjustment" packages are one manifestation of this emphasis upon efficiency in resource use. The industrial sectors in developing countries have particularly come under scrutiny, with allegations that much of industry exists largely due to protectionist measures such as tariffs, quotas, and subsidies of various sorts.

This stress upon the possible need to restructure manufacturing sectors, and the fact that economic decline in many countries is bringing about a de facto restructuring, raises a

number of questions which distributional data can be used to investigate. First, one can ask whether the usual assumption of the "representative firm" is a valid abstraction. It is generally recognized that economic characteristics vary significantly across industries, with regard to skill mix of the work force, wage-setting, and product market conditions. Economic characteristics may also vary considerably by size of firm. When setting industrial policy, it needs to be recognized that many policy measures and instruments are not neutral with regard to size of firm. A policy of tight credit, for example, could well harm small firms more than large ones, particularly if the latter have differential access to foreign sources of finance. Thus, knowing the size distribution of of firms is a valuable input for macro and sectoral policy making.

A second broad use of size distribution data is diagnostic. A frequent allegation is that many sub-sectors of manufacturing in developing countries, particularly those not involved in export, are characterized by a low intensity of competition and the allocative inefficiencies associated with a lack of competition. This argument is usually supported by a priori reasoning or antedotal evidence. Empirical studies of industrial concentration for developing countries are few indeed,[3] primarily because the necessary data have not been available to researchers. Size distriubtion data provide the information to

---

[3] See John Weeks, The Limits to Capitalist Development: The Industrialization of Peru, 1950-1980 (Boulder, Colorado: Westview Press, 1985), where sectoral concentration in manufacturing is compared for Peru and the United States, and other studies are referred to.

empricially test the relationship among degree of concentration, rates of industry growth, price behavior, and a number of other variables. When considering changes in policy toward manufacturing, profiles of industries by size of establishment would provide an important guide to policy impact.

A third issue of considerable policy significance is the issue of most efficient plant size and changes in most efficient plant size over time. Mainstream economic theory usually assumes "U-shaped" cost curves for the representative firm in order to be consistent with prefect competition/general equilibrium parables. Much economic reasoning, particularly the alleged efficiency gains from freer trade, adopt this abstraction. Considerable empirical evidence as well as alternative theories of the firm suggest otherwise -- that increasing returns or constant returns may be the rule. Given the well-known difficulties with estimating production functions,[4] this issue is difficult to test directly, even with size distribution data. However, the question can be approached indirectly by investigating whether or not the size distriubtion of firms in a sub-sector remains stable over time. Cross-country comparisons would be particularly indicative here; for example, if one found a tendency for average firm size to rise in a particular sector in several countries, the increasing returns hypothesis would be strengthened (and vice-versa).

---

[4] Empirical tests flounder upon the impossibly of a specification of output/input relations which can distinguish between technical change and returns to scale. In general, evidence of non-constant returns to scale can always be rationalized as evidence of technical change or the presence of absolutely scarce factors of production.

In trade theory the technological characteristics of firms plays a pivotal role. Most theorizing and policy recommendations about the efficiency gains from trade are based upon the Heckscher-Olin-Samuelson model, which assumes that all trading partners operate on the same production functions.[5] While the concept of a production function does not easily lend itself to empirical testing, size distribution data would allow for indirect tests of this crucial assumption.

Finally, size distribution data provide an aid to sectoral-specific policies. Sectors characterized by many small firms require different institutional implementation than sectors with a few large firms. Specific programs are relevant in one case but not in the other, since large firms can usually provide their own training, research and development, market studies, etc., while medium sized and small firms cannot. Within UNIDO and beyond, size distriubtion data would be extremely valuable for developing action programs.

In summary, size distriubtion data for manufacturing sectors represent a contribution to the analysis of industry much as size distribution of income data contribute to an analysis of economic welfare of the population. In every manufacturing sector in each country of the world production is the result of the output of many establishments, operating under quite different conditions even in the same product line. It is likely that the conditions vary even more by size of firm in developing countries than in

---

[5] One must make this assumption (along with others) in order to conclude that the pattern of trade reflects a country's relative factor endowments.

developed countries. Lack of data has resulted in disregarding these differences. It is no longer necessary to do so.

PART II: COMPILING THE DATA AND GENERATING STATISTICS

II.1 COUNTRY COVERAGE

The goal of the project is to generate statistics at the 3-digit level. I suggest that the information be published in two forms. On a regular basis in UNIDO's industrial surveys, each 3-digit sector would be presented in three size categories (discussed below). In a publication of more limited circulation, perhaps to be sold at a price to recover the costs of the project, the full range of information for each country would be included. At the outset, it is useful to identify the possible universe of countries that might be covered at the desired level of disaggregation. The United Nations has over 130 member states which report information on manufacturing at least to some degree.[6] However, the relevant universe for this project is considerably smaller. First, the project refers to developing countries, so the advanced market economies and most of the centrally planned economies are excluded. This reduces potential coverage to one hundred.

A brief discussion helps to clarify the relevant refence

---

[6] Handbook of Industrial Statistics, pp. 46-47.

group of countries. With regard to the study of economic characteristics of manufacturing by establishment size, one can identify two levels of disaggregation. One is the familiar micro study of individual establishments. Another level, that for which this project is designed to provide data, are studies in which establishments are grouped together by size categories. In this second case, the underlying hypothesis is that establishments of similar size have certain common characteristics which distinguish them from establishments of larger or smaller size, and that these characteristics can be identified on the basis of a composite, "typical" establishment aggregated over a size range. If a particular country has an extremely small manufacturing sector, by the time one has diaggregated to 3-digits and a given size range one has in effect embark upon a micro study, not a "typical" establishment study. In addition to this general consideration, there is a purely practical one: in most countries manufacturing surveys are conducted subject to a "confidentially" clause. Typically, data are not presented if a size category contains only one establishment. In consequence, size distriubtion data from countries with small manufacturing sectors shwo many cells in which data have been "suppressed". Alternatively, size ranges may be reported combined together, so in practice one ends up with very little distributional data indeed.

In practice the manufacturing sector in many countries is simply too small to make size distribution data useful to collect. First, there are those countries which are so tiny in population that no manufacturing sector beyond a handful of firms

exists or is likely to exist in the foreseeable future. Examples are Reunion and Cape Verde. Second, there is a large number of countries which are so underdeveloped that their industrial sectors are tiny, even though the national populations may be in the millions. Many of the Sub-Saharan African countries fall into this categroy, as well as most of the West Asian oil states (and some of these also have tiny populations). When these virtually non-industrial countries are excluded, the potential universe drops to 70 countries at most. In my judgement a more reasonable target for collecting size distriubtion statistics is 50. That is, there are approximately 50 developing countries whose manufacturing sector is sufficiently large that size distriubtion data at the 3-digit level might be worth providing to researchers. Were such data available for the other 50 developing countries (those falling into the two exclusion categories discussed above), it would probably be a waste of resources for UNIDO to publish them. I am not suggesting that study of the manufacturing sector of Barbados, for example, is not interesting and extremely important. My point is that data by size of establishment in this and similar cases would in effect be micro-data, even individual firm data for some 3-digit sectors. Such data falls into a different category of information than that covered by this project, and in any case governments would be unlikely to make it available for publication.

For the majority of the 50 countries relevant to this study, no size distribution data are reported by the governments in

question. However, in countries that do not report such data, there are usually annual surveys of manufacturing, based upon reports by individual firms. Therefore, the raw data are present to provide distributional data. UNIDO could provide the expertise to have such distributional data generated as a matter of course. Most of the 50 countries compile their survey data with computer facilities, and UNIDO could encourage reporting of distributional data by developing programs to process the raw information.

Table 1 provides a list of the potential candidates for size distriubtion data. It is not intended to be definitive, but I would argue that on the basis of the criteria set out, the relevant reference group of developing countries would not be much larger than this. Alongside each country is reported the availability of size distribution data. The project seeks to cover five indicators by size range -- number of establishments, number employed, wages and salaries, value added and gross output. When a country reports all five of these, its data is identified as "complete". The appendix to this report discusses available data in more detail, in most cases providing the data itself.

The availability of data is quite "spotty", with only a few countries providing annual reporting. The picture is not a very tidy one. In the next section I give in detail how each country's data might be presented. In doing this, I postpone to a later point issues of adjustment of data and reconciling differences in national reporting.

Table 1

Developing Countries for Which Distributional Data Would be Relevant (By Region)

Country	Notes on Data Availability
Africa	
Algeria	--
Egypt	--
Ethiopia	--
Ghana	--
Ivory Coast	--
Kenya	annual, est. and emp. only
Morocco	annual, comp., 1976-
Nigeria	late 1960s and early 1970s, comp.
Tanzania	early 1970s, complete
Tunisia	--
Uganda	--
Zaire	--
Zambia	--
Zimbabwe	--
Asia	
Bangladesh	--
Hong Kong	annual comp., 1976-
India	--
Indonesia	comp. census, 1974
Iran	--
Iraq	--
Korea	annual comp., 1966-
Malaysia*	regular, 1970-
Pakistan	annual comp., 1965-
Philippines	--
Singapore	--
Sri Lanka	--
Thailand	--
Turkey	--

Table 1 (con't)

Country	Notes on Data Availability
Latin America	
Argentina	--
Bolivia	complete, 1979
Brazil*	complete, 1966-68, then various
Chile	--
Colombia	annual, complete, 1971-
Costa Rica	complete, 1966
Dominican Republic	--
Ecuador	-- (2-digit available)
El Salvador	annual, complete, 1971-
Honduras	--
Jamaica	--
Mexico	census, 1971, 1976
Panama	--
Paraguay	--
Peru	annual, complete, 1963-1975
Uruguay	--
Venezuela	complete, 1970, 1974
Puerto Rico	--

NOTES:
* Regular but not annual.
-- indicates "not available".
comp. - indicates indicators are complete

II.2 PREPARATION OF DATA

II.2.1 General Comments

Before considering the specifics of data preparation, a general comment about size distriubtion data is necessary. As will be shown below, the presentation of size distribution statistics involves organizing and summarizing a great deal of data. One might argue that this is hardly a particular characteristic of the present type of data, for national income data, UNIDO's own new compendium of manufaturing statistics, trade statistics, etc., all involve a large amount of data. Size distribution statistics are different from the other data sets mentioned in an important respect. National income accounts, manufacturing statistics, trade data, and most other data sets published by the United Nations can be reduced to two dimensions -- the statistical measure and time. This greatly simplifies both presentation and analysis, for the relevant data in principle and practice can be summarized in a single table. Size distribution data is by its nature three dimensional and cannot be summarized in a single table, no matter how sparce or rich the data. One is always dealing with 1) a set of indicators, 2) those indicators over time, and 3) those indicators divided into categories of firm size.

This three dimensional nature of the data treated in this study is the source of a number of complications. A general problem it creates is a tendency for one to underestimate the richness of available data. As indicated, each data cell has three dimensions. Expanding each cell in all three dimensions

involves considerable information. The dimension for which infomation is most limited is that of time -- for example, to provide for a series of consecutive years employment data for a given size range. When assessing the adequacy of information, one must keep in mind that any distriubtional data is a considerable step forward for research purposes. At this stage of data collection, it is unrealistic to expect to obtain data in all three dimensions for many countries. The informational break-though provided by this project is not primarily in the time dimension, but in the size-category of establishment dimension. In other words, the unique contribution of this project is to collect manufacturing data by size of establishment. Luckily, it is possible to expand this contribution into the time dimension for some countries.

II.2.2 Presentation by ISIC Sectors

There are great variations in data availability, as Table 1 shows. In order to keep my discussion as tidy as possible, I shall first deal with the manner in which the raw data might be adjusted, then consider the form in which they might be published. First, the data would be presented at the standard UN 3-digit ISIC level. While some of the reporting countries still use the "old" UN ISIC categories and some employ national categories particular to a country, the data for each reporting country can be converted into the present UN system. In some case, however, the conversion involves loss of detail. For some countries certain sectors could only be presented at the present

2-digit level. For example, El Salvador disaggregates sector 32 into 321, 322, 323, and 324, while Brazil does not, reporting only at the 2-digit level (see the discussion of each country in the Appendix). Therefore, when publishing the adjusted data, I recommend that for each country both the 2-digit and the 3-digit totals be included. This would allow for direct comparisons across countries. Further, the aggregate manufacturing totals (sector 3), should be provided.

In some cases, a perfect fit between the present ISIC system and other systems is not possible. For example, under the old UN system, plastics was included in "other manufacturing" (old sector 39), and is not separately identified (Pakistan is a case in point). In such cases, I recommend that no attempt be made to reallocate indicators, but the anomaly be indicated by footnotes. Obviously, this limits some types of cross-country comparisons. It would seem preferable to making adjustments of spurious precision.

II.2.3 Presentation by Size Ranges

The degree to which countries disaggregate their data by size ranges varies tremendously, from the case of Bolivia with thirteen ranges by number employed to Venezuela which uses only four. Fortunately, the data for all countries can be converted into the three standard UN categories of 10-19 ("small"), 20-99 ("medium"), and 100 and over ("large").[7] All size ranges are to

[7] United Nations, International Recomendations for Industrial Statistics, Series M, No. 48 (New York: United Nations, 1983).

some extent arbitrary, so the criteria of selection can justifiably be based upon convenience and the usefulness of the resultant statistics. These three categories qualify on bot grounds. First, they represent the ranges which the UN recommends for all its member countries. Second, they provide at least a reasonable quantification of terms used commonly in the literature on industrial organization. Most reporting countries provide data considerably more detailed than these three categories. How this more detailed data might be made available to researchers is discussed below.

II.2.4 Presentation of Indicators

As said before, I recommend that five indicators be presented: number of establishments, number employed, wages and salaries, value added and gross output. There is some variation among countries in the definition of all but the first of these. With regard to establishments, all reporting countries covered in this report count the number of separate production units; i.e., not the number of ownship entities ("firms"). Researchers must be warned of this in notes, for it affects calculations of industrial concentration when one intends by those statistics to measure the degree of market control.

The other indicators vary in ways familiar to UNIDO work on manufacturing statistics. The "number employed" may or may not include proprietors, unpaid workers (e.g., apprentices), and part-time workers. "Wages and salaries" may refer to labor costs inclusive of various types of non-salary benefits, or only to

direct payments to workers. Value added and gross output may or may not include certain categories of taxes. Attempting to render these four indicators consistent across countries would be an extremely time-consuming and costly task of dubious utility. How important these variations are is determined by the use to which the statistics are to be put. Merely providing the statistics based on national definitions is at this stage a tremendous service offered by UNIDO. Fine adjustment can be reasonably left to individual researchers. One might in footnotes indicate differences in defintion, but a general note of warning would be sufficient.

II.2.5 The Standard Table

From this point on in the discussion, I shall use the term "standard table", which refers to a five-by-three table with the five indicators listed horizontally and the three size categories vertically, at the 2-digit or 3-digit UN ISIC level for a given year for a given country. As mentioned above, this is a three dimensional table (indicators, size category, and year, with a fourth dimension not previously discussed, industrial sector) An example of the standard table is given below. In this example, value added is not reported in the country source. An alternative way of presenting the information would be to organize standard tables by indicator, with years listed horizontally. The manner I have chosen has the advantage of being appropriate for countries which have to date reported only one year's size distriubtion data.

COUNTRY: El Salvador
SECTOR: 324 (Shoes and other Footwear)
YEAR: 1974

(Number or National Currency units in thousands)

Size Range	Estab.	Emp.	Wages& Salaries	Gross Output	Value Added
10-19	0	0	0	0	--
20-99	2	57	130	445	--
100+	1	1308	4829	42773	--
Totals	3	1363	4959	43218	--

II.3 PRESENTATION OF STATISTICS

II.3.1 Publication Format

The most important decision to be made in this project is how to present the statistics. Problems arise from the great variation in the availability of data over time. First let us consider those countries for which there are more-or-less continuous annual data. These countries are listed in Table 2. It is to be noted that the year of first appearance of the data varies among countries, though further investigation of national sources may produce data for earlier years for some of the countries.[8] Further, some countries report their distributional statistics later than other, so even for this set of countries one cannot presume that the data for a particular year would be available to UNIDO at the same time from several countries. In

[8] The data availability reported represents a survey of sources in the UNIDO collection, the IMF/World Bank Library, and the Library of Congress.

Table 3 are listed the countries which have data for only one may have supplied data for several years but discontinued the practice.

I recommend that UNIDO publish all available distributional data at the 3-digit level. Thiscould be done in two forms, in a reference work and in the Industrial Development Surveys, where the latter would involve up-dating of data in the former. Following the practice of other United Nations publications, the

Table 2

Countries with Continuous Annual Data

Country	Time Period
Colombia	1971-1983
El Salvador	1970-1979
Hong Kong	1976-1978
Korea	1966-1982
Morocco	1976-1980
Pakistan	1964-1980
Brazil	1960's,1970's,1980's
Malaysia	1970's
Malta	1960's & 1970's
Peru	1963, 1965-1975 (not 1968)

NOTES:

Brazil -- 1966-1968, 1970, 1973, 1982, reported in the Anuario Estadistico, regularity of data collection not clear

Malaysia -- 1970, 1972-1974, 1978, apparently regular publication.

Malta -- 1964, 1969, 1972, 1976, 1977, apparently regular publication.

Peru -- publication suspended in 1968, available subsequently from tabulations in Ministry of Industry and Tourism

IDS could in each issue publish more than one year's standard tables. For example, the forthcoming issue of the IDS could include the first six countries in Table 1 for the years 1978-1980, with the subsequent issue covering 1981-1983, when presumably data for those years would be available for all of the six. The alternative would be to publish the most recent data for all of the countries. This would have the drawback of being less "tidy", since tables from the various countries would refer to different years -- 1983 being the most recent for Colombia, 1979 for El Salvador, etc.

In any case, publishing more than a few years at a time would be quite costly in a high-quality publication like the IDS. If the recommendation is followed to publish the 1-digit and 2-digit aggregates, then an average of thirty standard tables per country per year is involved. For only six countries this works out to 180 standard tables, and perhaps double that if a simple analytical table such as the percentage distribution of

Table 3

Countries with Occasional

Or Single Data Points

Country	Years
Bolivia	1979
Costa Rica	1967
Cyprus	1976*
Indonesia	1974
Mexico	1970, 1976
Nigeria	1968-70, 1971
Tanzania	1970, 1973-74
Venezuela	1970, 1974

indicators by establishment size is included. At the minimum twenty pages of tables are involved for one year and six countries. Including three years -- the first suggestion above -- implies sixty pages of standard tables alone.

The remaining size distribution data, past data for the countries with regularly-published information and data for countries with occasional surveys, could be published in the form of a computer-generated output. Since this would be an extremely valuable source to researchers, it could be sold on a semi-commercial basis. This publication would give standard tables for all years prior to 1978 (for example), and for all countries in Table 3 regardless of year.

The form in which the statistics will be published is an issue involving cost considerations which are best known to UNIDO. The purpose of discussing this issue here has been specify the format of data presentation for the treatment of individual country data, which follows in the appendix. That treatment is based upon the assumption that UNIDO wishes to publish distributional data on a regular basis for as many countries as possible for at least some years. The judgement reached here is that regular publication of data should be restricted to the few countries which provide a more or less continuous flow of consistent information.

II.3.2 Providing Analytical Tables

If cost and space permit, provision of some simple analytical statistics along with the standard tables would be a useful addition. The most obvious of these statistics would be a percentage distribution of indicators. Such an analytical table is provided in the statistical sheets for several countries in the Appendix (see, for example, Brazil). These sheets are from the non-defunct World Bank project on industrial statistics. Those sheets provide several analytical tables, though all but the percentage distribution of indicators are not useful for this project, for reasons explained in the introduction to the Appendix. In light of cost considerations, I recommend publishing only the standard tables in the IDS.

PART III. SUMMARY

In the Appendix I give in detail the manner of preparing each country's data for presentation. Here I summarize my recommendations and assess the output of the project.

1. First, the project provides a collection of size distribution statistics. While it cannot be claimed that all extant size distribution data are provided here (or in the UNIDO collection), coverage far exceeds any attempted before (i.e., the World Bank industrial statistics project).

2. Second, it provides a detailed guide to converting the size distriubtion data from eighteen countries into standard tables. For six of these countries, annual data can be published in the next UNIDO Industrial Development Survey. Further

investigation of national sources may allow the addition of four more countries (see Table 2).

3. It is recommended that UNIDO use the standard tables as guides to encourage more governments of developing countries to provide size distriubtion statistics.

As indicated in point two above, the initial set of countries for which UNIDO would provide regular publication of data would be a minimum of six and a maximum of ten. This may seem a rather inadequate beginning, a fraction of the fifty countries listed in Table 1. However, this offering must be placed in context. As explained above, the data associated with this handful of countries will be quite voluminous. Second, this small beginning would represent the first attempt to provide researchers with size distribution data and should be compared to the initial publication by the United Nations of national income statistics after World War II. If one looks back at those early publications, one is struck by how shockingly incomplete they are by comparison to the same publications today. However, at the time they represented a major step forward in providing internationally comparable statistics. Further, the publication of these statistics, even in limited and income form, served to encourage their collection and construction in countries where they were not available. As UNIDO, other multilateral organizations and individual researchers begin to use the initial small offering of data for policy studies, the importance and utility of such data will become obvious. One reason that so little size distribution data are available is that no one uses

them. By making what exists generally accessable, analytical use of size distribution data will become common and the demand will increase.

In light of these considerations -- recognizing that the effort is a pioneering one -- I have placed emphasis upon reliability and consistency of published data. In principle it is possible to attempt adjustment of data, to fill in blanks, and general produce a set of statistics that appear more complete. Such an exercise would largely be an illusion, however. Statistics incorporating dubious and arbitrary adjustment exercises would generate an output of questionable utility. The result might well be to discourage rather than encourage the collection and anlysis of size distribution statistics for policy purposes.

APPENDIX:

GUIDE TO COUNTRY DATA

APPENDIX: GUIDE TO COUNTRY DATA

TABLE OF CONTENTS

	page
EXPLANATORY NOTES	26
COUNTRIES	
Bolivia	27
Brazil	28
Colombia*	29
Costa Rica	30
Cyprus	31
(Ecuador	32)
El Salvador*	33
Hong Kong*	34
Indonesia	35
Korea*	36
Malaysia	37
Malta	38
Morocco*	39
Mexico	40
Nigeria	41
Pakistan*	42
Peru	43
Tanzania	44
Venezuela	45

* Minimum countries recommended to be included in regular publication of distributional statistics. See Section II.3.1.

* Ecuador's data is reported at the UN ISIC 2-digit level. For all other countries data are at 3- or 4-digit level.

EXPLANATORY NOTES:

DATA TABLES

In the pages that follow are provided data for the countries discussed. These tables are of two types, from nationaal publications and the output of the World Bank Industrial Statistics Project (see footnote 2). The World Bank tables are all organized by 3-digit ISIC sector, with all sectors for a given year listed in numerical order. This output provides five tables for each sector, which unfortunately are neither well-identified in the output nor self-explanatory. I explain them below.

1. Title: "National Currency" -- This first table gives the raw data by size categories. The columns are, reading left to right:

size -- number given is the lower limit of the size range, by number employed;

estab -- number of establishments;

employ -- number employed;

wages -- wages and salaries paid;

GQ -- gross output (value added plus intermediate costs); and

VA -- value added.

2. Title: "Proportion of Total" -- This table provides the proportional distribution of each indicator by size category. A line is provided that verifies that the proportions add to unity. The last three lines of the table provide statistics of variation.

3. Title: "???". This table provides certain ratios by size of establishment. The line noted as "TOTAL" gives average values for the sector as a whole. The ratios are:

LKJ/EKJ -- employment per establishment within the size range;

WSKJ/LKJ -- average wage per worker in the size range;

GQKJ/LKJ -- average gross output per worker in the size range; and

VAKJ/LKJ -- average value added per worker in the size range.

4. Title: "Relative to Industry Mean". This table was prompted by a useful idea, but in practice is useless. Its intention was to provide the values in the previous table relatively to the industry mean; so, for example, one could quickly read how much hirer were wages in the largest establishment category and in the smallest. However, through an error the industry mean in each case is not the true mean, but the mean of the average for each size category. Since indicators are not distributed evenly over the size categories, the true mean usually lies far from this arbitrarily weighted mean.

5. Title: "Measures of Dispersion". This table calculates concentration indices for for all variables but number of establishments. The relevant concentration indiex is given in the last line of the table. The measure, called "the Herfindal Index", is quite simple. It seeks to summarize the entire size distribution, not just calculate the market share (say) of the largest four establishments. If one simply added the shares of all establishments, one would always yield unity as a result. The avoid this result, the index takes the square of each

establishment's share of an indicator. If an industry had only one establishment, the concentration index would be 1.00 (1 squared); if there were two establishments of equal size, the index would be .5; and for many establishments of small size, the index approaches zero. For grouped data, one must measure the share of an indicator by an establishment as the average share for the size range. In this case, the index is calculated as follows.

$$H = \sum_{i=1}^{m} N(i)\{[1/N(i)][x(i)/x]\}^2$$

N(i) = number of establishments in a size range i;

x(i) = the absolute value of an indicator in size range i;

x = the total absolute value of an indicator for all size ranges;

m = number of size ranges.

NOTES

All data referred to in this appendix can be found in the UNIDO collection or in this appendix. Location of sources is indicated by the following.

* Given in this appendix, in the form of tables from the World Bank project.

** Given in this appendix, photocopy of national publication.

[] (enclosed in brackets) Found in the UNIDO collection.

COUNTRY: Bolivia

YEARS: [1979]

MANUFACTURING CODE: UN ISIC, 4-digit level.

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[no]
Gross Output	[yes]
Value Added	[no]

SIZE RANGES: 5-9, 10-14, 15-19, 20-24, 25-49, 50-74, 75-99, 100-149, 150-199, 200-299, 300-499, 500-999, 1000+ (15)

ADJUSTMENT: Two adjustments are required:

1) compression of the fiften size categories to the standard three; and

2) aggregation from 4-digit to 3-digit level (no difficulty here, since 4-digit categories are the standard UN ISIC).

SOURCES: Estadisticas Inudstriales Manufactureras (Estrato Fabril) Resultados Preliminarias 1978-1979

COUNTRY: Brazil

YEARS: 1966**, 1967**, 1968**, 1970*, 1973*, [1982]

(1966&1968 tabulated by hand in standard UN ISIC)

MANUFACTURING CODE: National categories similar to old UN ISIC.

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500+ (7)

ADJUSTMENT: Two adjustments are necessary:

1) Compression of the seven size categories into the standard three; and

2) identifying national categories as UN ISIC cateories; this is a re-labelling exercise, with the conversion given below.

SOURCES: Anuario Estadistico do Brasil, and Censo Industrial, 1970.

COUNTRY: Colombia

YEARS: 1971*, 1973*, 1974*, 1975*, [1977-1983]

MANUFACTURING CODE: UN ISIC 3-digit

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 0-4, 5-9, 10-14, 15-19, 20-24, 25-49, 50-74, 100-199, 200+ (9)

ADJUSTMENT: The only adjustment required is to compress the data into the three standard size ranges.

SOURCES: Departamento Administrativo Nacional de Estadistica, Industria Manufacturera.

COUNTRY: Costa Rica

YEARS: 1967*

MANUFACTURING CODE: UN ISIC 3-digit

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 5-9, 10-19 (by 10s to 49), 50-74, 75-99, 10-199-200. 200+ (9)

ADJUSTMENT: Compression of data into standrad size ranges.

SOURCES: Eastadistica Industrial, 1967 (data sheets follow in this appendix)

COUNTRY: Cyprus

YEARS: 1976**

MANUFACTURING CODE: UN ISIC

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[no]
Value Added	[no]

SIZE RANGES: 1-4, 5-9, 10-19, 20-29, 30-49, 50-99, 100-499, 500+

ADJUSTMENT: Compression of data into standard size ranges.

SOURCES: Census of Industrial Production 1976 (UNIDO collection)

COUNTRY: Ecuador

YEARS: 1975**, [1976], [1980]

MANUFACTURING CODE: UN ISIC at 2-digit level only

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 1975 and 1976 -- <7, 7-9, 10-19, 20-49, 50-99, 100-199, 200-499, 500+ (8); 1980 -- 1-2, 3-4, 5-9, then same as previous years.

ADJUSTMENT: Compression of data into standard size ranges.

SOURCES: Encuesta de Manufactera y Mineria (UNIDO collection)

COUNTRY: El Salvador

YEARS: 1970**, 1971*, 1973**, 1974*, 1975**, 1976*, 1977*, [1978] (1970 and 1973 tabulated by hand)

MANUFACTURING CODE: UN ISIC 3-digit

INDICATORS:

Establishments	[yes]	
Employment	[yes]	
Wages&Salaries	[yes]	
Gross Output	[yes]	
Value Added	[yes]	not available for 1971 and 1973

SIZE RANGES: 1971-1977 -- 5-9, 10-14, 15-19, 20-24, 25-49, 50-99, 100-499, 500+ (8); 1978 -- 5-9, 10-19, 20-49, 100-199, 200+ (6)

ADJUSTMENT: Compression of data into standard size ranges.

SOURCES: Direccion General de Estadistica y Censos, Anuario Estadistico

COUNTRY: Hong Kong

YEARS: [1976-1978], 1981 **

MANUFACTURING CODE: UN ISIC, 4-digit

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 1-9, 10-19, 20-49, 50-99, 100-199, 200-499, 500-999, 1000+ (8)

ADJUSTMENT: Two adjustments are required;

1) compression of data into standard size ranges; and

2) aggregation from 4-digit level to 3-digit level.

SOURCES: Census of Industry

COUNTRY: Indonesia

YEARS: 1974*

MANUFACTURING CODE: UN ISIC, 3-digit

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES:

ADJUSTMENT: Compression of data into standard size ranges.

SOURCES: World Bank data sheets.

COUNTRY: Korea

YEARS: 1966**, 1967**, 1969**, 1970*, 1971**, 1972**, 1973**, 1974*, 1975*, 1976*, 1977*, 1978**, 1980**, 1981**, 1982**

MANUFACTURING CODE: 1966-1969, "old" UN ISIC system; 1970, onwards, UN ISIC 3-digit

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 5-9, 10-19, 20-49, 50-99, 100-199, 200-499, 500+

ADJUSTMENT: Two adjustments are required:

1) for all years, compression of data into standard size ranges; and

2) for 1966-1969 conversion of old ISIC sectors into standard UN ISIC.

The 1966-1969 data are reported at a level of disaggregation which allows for 2) to be done without difficulty.

SOURCES: Economic Planning Board, Report on Mining and Manufacturing

COUNTRY: Malaysia

YEARS: 1970*, [1972**, 1973*, 1974*], [1978]
**notes that 1972 is provided tabulated by hand.

MANUFACTURING CODE: UN ISIC, 4-digit

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

Note: employment and value added only for 1978.

SIZE RANGES:

ADJUSTMENT: Aggregation to 3-digit level provided in this appendix for 1970-1974. Data must be compressed into standard size ranges. Data cover peninsular Malaysia only.

SOURCES: Survey of Manufacturing Industries, Peninsular Malyaysia; 1978 data from unpublished tables summarized in World Bank, Malaysia: Development Issues and Prospects of Small Enterprises, Vol. III: Annexes and Statistical Tables (June 25, 1982).

COUNTRY: Malta

YEARS: [1964, 1969, 1972, 1976, 1977]

MANUFACTURING CODE: 1964, 1969, 1972 -- "old" ISIC at 3-digit level; 1976, 1977 -- standard ISIC, 4-digit.

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 1-5, 6-10, 11-19 (by tens to 99), 100-149, 150-199, 200-299, 300+ (ranges frequently combined when few establishments in sector)

ADJUSTMENT: Two adjustments are required:

1) Compression of data into standard size ranges; and

2) coversion of early years to current ISIC and aggregation of all years to 3-digit level.

SOURCES: Central Office of Statistics, Census of Production

COUNTRY: Mexico

YEARS: [1970, 1975]

Note: data from 1960s should also be available from earlier industrial census, but volume not in WB/IMF library. The 1980 census should also have been published by now. A further search for census documents should provide data for 1960, 1965, and 1980.

MANUFACTURING CODE: "Old" ISIC code at 4-digit level (corresponds roughly to current 5-digit level)

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: variable from sector to sector, with frequent combinations when there are few establishments.

ADJUSTMENT: Mexico involves the most tedious adjustments. The "old" 4-digit sectors must be aggregated up to the current 3-digit sectors. In some cases this involves combining sub-sectors which have different size categories. The easiest procedure is to convert the raw data into the standard size categories at the level of disaggregation presented in the national publication. Once this is done, then the sub-sectors can be combined.

SOURCES: Secretaria de Industria y Comercio, IX Censo Industrial

COUNTRY: Morocco

YEARS: [1976-1978, 1980]

MANUFACTURING CODE: For all years, the "old" ISIC system, with 2-digit sectors number 10-27 instead of 20-39.

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 0-10, 10-19 (by tens to 99), 100-199 (by hundreds to 499), 500+

ADJUSTMENT: Two adjustments are required:

1) Compression of data into standard size ranges; and

2) coversion to current ISIC and aggregation of all years to 3-digit level.

SOURCES: Direction de L'industrie, Situation des Industries de Transformation

COUNTRY: Nigeria

YEARS: 1968*, 1969**, 1970**, 1971**, 1972*

MANUFACTURING CODE: UN ISIC 4-digit

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 1-19, 20-49, 50-99, 100-299, 500-999, 1000+

ADJUSTMENT: One assumption and two adjustments are required.

1) Assume that the 1-19 data in the national publication corresponds to data for the standard 10-19 size range. Little error is involved as a result of this assumption. When I was in Lagos in 1980, statisticians at the Federal Office of Statistics told be that few establishments below 10 employees reported.

2) Aggregate 4-digit data to 3-digit level (as in World Bank sheets); and

3) compress to standard size ranges.

SOURCES: Federal Office of Statistics, Industrial Survey, Nigeria

COUNTRY: Pakistan

YEARS: (data refers to over-lapping twelve month periods; e.g., 1964 = 1964/65) 1964**, [1965], [1966], [1969], 1970**, 1975**, [1976-1980]

MANUFACTURING CODE: "old" ISIC for all years, 2-digit

INDICATORS:

Establishments	[yes]	
Employment	[yes]	not given for early years, see below
Wages&Salaries	[yes]	
Gross Output	[yes]	
Value Added	[yes]	

SIZE RANGES: 0-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, 1000-1999, 2000-4999, 5000+ (10)

ADJUSTMENT: Three adjustments are required:

1) compress into standard size ranges;

2) convert to standard UN ISIC (some sectors must be converted to 2-digit level); and

3) estimate missing employment data in early years.

The last adjustment can be done quite simply. Elsewhere in the Census one is given total employment for each sector. If one uses the average size of establishment for each size range from the first year in which it appears, employment can be estimated for the missing years. The only complication is to constrain the estimation such that total estimated employment for a sector equals the told reported elsewhere in the year's document for that sector.

SOURCES: Federal Bureau of Statistics, Census of Manufacturing Industries

COUNTRY: Peru

YEARS: 1963*, 1965*, 1966*, 1967*, 1968*, 1970*, 1971*, 1972*, 1973*, 1974*, 1975*

MANUFACTURING CODE: data sheets provide information at UN ISIC

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 5-9, 10-14, 15-19, 20-49, 50-99, 100-199, 200-499, 500-999, 1000+ (9)

ADJUSTMENT: Compression of data into standard size ranges.

SOURCES: The information is provided in this appendix. The original sources are:

1963 Censo Economico 1963

1965-1968 Estadistico Industrial

1970-1975 unpublished data from Ministry of Industry and Tourism

COUNTRY: Tanzania

YEARS: [1970], [1973], 1974*

MANUFACTURING CODE: UN ISIC, 3-digit

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[no]
Gross Output	[yes]
Value Added	[no]

SIZE RANGES: 10-19, 20-49, 50-99, 100-499, 500+ (5)

ADJUSTMENT: Compression of data into standard size ranges.

SOURCES: Ministry of Economic Affairs and Development Planning, Bureau of Statistics, Survey of Industrial Production

COUNTRY: Venezuela

YEARS: 1970**, [1974]

MANUFACTURING CODE: UN ISIC, 3- and 4-digit

INDICATORS:

Establishments	[yes]
Employment	[yes]
Wages&Salaries	[yes]
Gross Output	[yes]
Value Added	[yes]

SIZE RANGES: 5-20, 21-50, 51-100, 100+ (4)

ADJUSTMENT: Compression of data into standard size ranges. The smallest national size range is 5-20, rather than the standard 10-19. Because establishments hiring 5-9 make such an insignificant contribution to the share of indicators in most sectors (with the exception of number of establishments), this difference in size coverage might merely be indicated in a footnote. Any adjustment procedure would be arbitrary.

SOURCES: Ministerio de Fomento, Direccion General de Estadistica y Censos Nacionales, Encuesta Industrial.