

ECONOMIC GROWTH CENTER

YALE UNIVERSITY

Box 1987, Yale Station  
New Haven, Connecticut

CENTER DISCUSSION PAPER NO. 232

DATA NEEDS IN DEVELOPMENT ECONOMICS

Carlos F. Díaz-Alejandro

July 1975

Note: Center Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Discussion Papers should be cleared with the authors to protect the tentative character of these papers.

## DATA NEEDS IN DEVELOPMENT ECONOMICS

Carlos F. Díaz-Alejandro\*

Yale University

### I

There is much to be said for the argument that there are no facts without theory. But there is also some force in the line, that "if you cannot measure it, it does not exist." Theories of economic development and growth provide weak and uncertain guidance to empirical researchers in development economics, so that progress in this area is likely to arise from the interplay of hypotheses borrowed or adapted from all branches of economics with laborious empirical work. In this process standard economic theory itself should become richer by its greater exposure to Third World circumstances. Note how, nowadays, macroeconomists of industrialized countries try to catch up, after many years of neglect, with the inflationary experience of some Latin American countries, an experience which includes many instances of the "stagflation" which fits so uneasily into standard models.

This paper, then, supposes that there is no shortage of hypotheses, hunches and insights which could be tested against Third World data, nor a lack of analytically oriented concepts, as phrased by Simon Kuznets, to guide us (even if tentatively) in translating raw primary data into quantitative economic measures. The problem is choosing the more interesting hypotheses, establishing insightful categories, and finding or generating fresh data. Indeed, the latter aspect is so serious that one increasingly prefers to read papers which involve no elaborate hypotheses but present

new solid data about an interesting issue without more than a coherent "story" to bring the numbers together.

There has been a tendency in the area of development economics to manufacture pseudo-theories to explain pseudo-facts, often leading without much pause to persistent policy advocacy. The excuse that "data are not there" has been used to justify much nonsense. Some of the most interesting work done in the development area in recent years involves this systematic examination of empirical evidence, which ends up giving a picture of reality significantly different from that commonly assumed in both the theorizing and policy-making of development. Examples include the Kindleberger work on terms of trade, that of MacBean on export instability and, more recently, the work of several authors on the nature and extent of the problem of marginality and unemployment. We will return to this type of empirical work below.

Rather than providing an exhaustive list of all data which it would be good to have, this paper will focus on data sources and techniques of obtaining data which, in my view, promise the largest pay-offs in development research. Such a view, of course, reflects my judgement about which are the most interesting and important researchable topics in the area, as well as about how contemporary capitalistic developing economies work. Analysts with different interests in the pure research-policy spectrum, or located in other parts of the world, or with different preconceptions, could very well come out with different conclusions. Before getting into all that, one should start by reviewing the major progress made during the post-Second World War period in data availability for development research.

## II

Much has been accomplished during the last two-and-a-half decades in expanding the data base available for researchers in economic development. The starting point was meager indeed: the number of developing countries with national accounts, input-output tables and flow-of-funds tables were few in 1950. By 1975, just about every developing country appears in United Nations publications with some kind of national accounts, and an increasing number of countries boast of other components of comprehensive macro-economic statistics, including those on international transactions and in the monetary field.

With all their problems (and they have many), these masses of data have supported a narrowing of our ignorance regarding the process of growth and structural change which, by historical standards, must be rated as very impressive. The Kuznets-Chenery "laws of development," of course, could not have been produced without such data base. Those observed regularities and the discipline of available national accounts have (although perhaps less than they should have) limited fanciful and sterile debates, channeling hypothesis-generation toward more fruitful areas. Furthermore, these data are on the whole easily available in national and international publications found in most libraries all over the world.

Besides their usefulness for comparative work on growth and structural change, national accounts have in some developing countries been pushed back at least into the 1930s and 1920s, stimulating research on such topics as how well different types of developing countries fared during the Great Depression, and also providing longer time series for testing secular views on growth and structural change.

It is not a difficult task to point out that, in spite of the great deal accomplished in the area of national accounts, many gaps and imperfections remain. In many developing countries, the raw data on which the accounts are based leave much to be desired; furthermore, it is not clear that these matters are improving with time. In some countries, after an initial burst of enthusiasm setting up the national accounts system, the collection and processing of raw data stagnates and is surpassed by the changing economic reality of the country. Increasingly misleading national accounts is the frequent result. In some cases, the preparation of national accounts has drawn resources away from government statistical offices generating raw primary data, leading to a decline in their output.

Most developing countries have not moved much beyond minimum national accounts. Input-output tables, flow-of-funds and particularly aggregate income distribution estimates remain relatively rare. Where they exist, it is even rarer to find systematic efforts for periodic actualizations of those tables. Even the core national accounts are seldom available at intervals more frequent than every year; quarterly series remain a rarity in developing countries. Indeed, prompt short-term economic indicators in developing countries are extremely scarce. Cyclical macroeconomic management in developing countries, an area neglected in the literature, typically evolves amidst great uncertainty as to what is really going on in the economy.

In the area of cross-country comparisons, the data base has not allowed an adequate separation of "true" structural differences from those which show up in national accounts data simply due to differences in relative prices from country to country. One example which I have frequently used involves the substantial differences found in the relative prices of capital

goods from country to country, differences which cast serious doubts about the meaningfulness of frequently-seen comparisons of rates of capital formation among countries. Comparisons across input-output tables are also hampered by differences in relative price structures.

But it is not my intention to build up to an across-the-board plea for more and better national accounts in developing countries.<sup>1</sup> At zero opportunity cost, that of course would be splendid, and would stimulate some good research. But given high opportunity costs and my preferences regarding research priorities, I will proceed to more specific pleading.

### III

Aggregation into national macroeconomic magnitudes has been recognized for a long time as a less defensible procedure for developing than for developed countries, given the greater structural and regional heterogeneity of the former. If the national accounts were put together from regional and sectoral bits and pieces in such a way that disaggregation were only a matter of running the adding-up machinery backwards, much of the following discussion would be unnecessary. In fact, aggregate investment is not typically obtained from adding up estimates of all firms' investments, but from imports of machinery and equipment plus construction licenses issued; national savings are derived as a residual from other magnitudes; and so on. In most countries, national accounts emerge as a precious package and attempts to decompose it run the risk of leaving the researcher empty-handed.

Unless the national accounts of developing countries are substantially improved, an unlikely prospect, the juice which can be further extracted from them, either in time series or cross-section studies, appears limited. A more

promising path toward greater understanding of the structure and workings of the modern sector of market-oriented developing countries would be to focus on, say, the characteristics of the largest 200 private corporations in each country. One can conjecture that such a group accounts for nearly all of the output generated by "dynamic" industrial branches, public utilities and large-scale mining. Their share of private capital formation, outside housing and consumer durables, is also likely to be very large. The same group will also account for a dominant share of foreign trade, and international capital flows. If, to that group, one adds the 30 or so largest public and banking enterprises, one would have the core of the modern sector of most developing countries.

Data on such a group of economic units is plentiful, but buried. Even developing countries with weak administrative machinery receive a steady flow of economic information from this group. There are income and sales taxes to pay or seek exemptions from; labor legislation requiring the reporting of employment and social security taxes; import licensing demanding evidence of need; price control regulations demanding justification of price increases; banking regulations, and so on. These data sources, suitably processed, could, in most countries, provide a running census of the "commanding heights" of the economy.

I became conscious of this rather obvious way of tackling data needs when studying the Colombian import control system. The officials in charge of that system kindly allowed me to examine a large number of import license requests, which, I quickly discovered, were a golden data mine. Each company provided data for current and past years on such variables as employment, sales, taxes of various kinds paid, installed capacity, wages, inventories, etc.

Once a decision had been made on the request, alas, these forms were destroyed. From the sample of request I was able to study, it was clear that a good deal of concentration existed. For example, just 80 industrial companies captured in the sample accounted for 30 percent of all 1970 Colombian imports; these same companies accounted for 21 percent of all income and sales taxes paid during 1970 in Colombia, and employed 19 percent of all those engaged in manufacturing in the same year [3, 1976].

There are a large number of issues which could be fruitfully studied in the context of ample data for the core economic units. All of the standard development questions on such matters as efficiency of trade policy, choice of techniques, productivity changes, etc., can be asked. Furthermore, the questions raised are likely to become more pointed when put in the context of the structural reality of the modern sector. Rather than imagining many atomistic firms behind some macro variable like total manufacturing output, one will have to face obvious departures from simplistic theories of the firm. Both empirical work and theorizing should benefit from the more realistic starting point; one is likely to understand the play better after having identified the major actors.

Let me again illustrate what I have in mind by reference to my work on Colombian foreign trade. From much of the debate on export promotion vs. import substitution, one could have imagined that the economic actors on both sides of that dichotomy were substantially different types: one type of unit labor-intensive, competitive and small, while the other was pictured as capital-intensive, large and heavily dependent on government favors. Yet it turns out that in Colombia, for the period I examined, many of the



firms engaged in new manufactured exports had been, or were still, engaged in much import substitution.<sup>2</sup>

While pure trade theory features individual industries having identical linear homogeneous production functions within and across countries as units of analysis ("wine and cloth"), often one finds large firms producing jointly both "wine and cloth." Rather than attempt to separate the various activities of a given firm so that they can fit into standard industrial classifications, a process typically full of doubtful procedures, in many instances it may be better to focus on the large multi-product firm as the unit of analysis.

Focussing on the core economic units should advance research in a number of other lines. It obviously facilitates measurement of the degree of foreign ownership in developing economies, and provides a framework for comparing the behavior and characteristics of domestic and foreign-owned firms in such things as exporting, research and development and choice of technique. The history of a country's industrialization may best be written as the history of the rise and expansion of the core economic units, of conglomerates and economic groups embracing several of those units, and of their interaction with government policy which stimulated such industrialization. It may, however, be much more difficult to obtain data on the history of the core economic units and of economic groups than on their current operations.

The difficulty to which reference has been made is not the political, man-made variety. In addition, many may fear that focussing on core economic units for data gathering is but a first step toward their eventual socialization. Somewhat spurious pleas for protecting corporate "privacy" will be heard also. Be that as it may, large companies and powerful economic groups

are likely to feel more comfortable with their numbers diluted in macro-economic aggregates than with schemes which highlight their largeness. Individual researchers may still obtain their cooperation for special projects; this is likely to remain the most realistic path toward obtaining a better understanding of the role and behavior of core economic units and groups in a number of development topics. Individual researchers may find it difficult to cover all 200 top firms, and may have to remain satisfied with sample surveys. But that takes us to the next topic.

#### IV

Exclusive emphasis on large companies and groups would neglect some very important research areas, such as income distribution, unemployment/underemployment and extreme poverty, as well as sectors dominated by small firms and farms. The fresh data needed to get a firmer understanding and better measurement of these matters and sectors is unlikely to come from the further refinement of national accounts. Better and more frequent censuses may not be the answer either. The most efficient, although far from cheap, instrument here seems to be the greater use of sample surveys.

In a lively Presidential address to the Eastern Economic Association, Barbara R. Bergmann chided U.S. economists for not following the example of scientists in generating their own data base [1, 1974]. She also advocated a much greater use of sample surveys to find out "first hand" about economic reality, rather than waiting for a government agency to produce numbers which are seldom exactly what we wanted. Her remarks, which make sense even in a country where government offices turn out massive amounts of high quality data, apply a fortiori to developing countries. Similar points have been

made by G. H. Orcutt [8, 1970].

There is already evidence supporting enthusiasm for sample surveys in development research. During the late 1960s and early 1970s, the conjecture of large and growing unemployment in developing countries became so publicized that it soon was taken as fact. Dramatic policy proposals for how to deal with this problem followed. Only somewhat later came careful efforts to ascertain the extent and characteristics of open unemployment. Such efforts naturally involved increasingly sophisticated sample surveys.

The picture of open unemployment yielded by that research, such as that of J. Ramos [9, 1974], certainly has a richness and sophistication not present in earlier discussions of the topic. Heavy participation of non-head of households and young persons among the openly unemployed raises serious questions, for example, about the connection between open unemployment, poverty and income distribution. That research also probes toward better measures of underemployment in city and country. It is striking that only after two-and-a-half decades of theorizing and mostly casual empiricism about surplus labor in development, one begins to see now systematic empirical studies on the topic.

A related area where sample surveys show great promise is the mapping out of "heterogeneous poverty" in the Third World. The logical data sources for researching this issue are of course national censuses, or stratified samples drawn from them. Impressive results have been obtained by Albert Fishlow via that route [4, 1972]. But the censuses are infrequent and not all developing countries may have the administrative infrastructure to carry out as thorough a census as those recent ones for Brazil. Furthermore, if emphasis is placed on studying the characteristics of poverty, rather than

income distribution, even partial surveys may provide more flexible and direct tools of research than censuses. They certainly are a natural complement to infrequent censuses which remain necessary for guiding the structuring and design of samples.

Have the poorest fifty percent of the population benefitted from growth in developing countries? Conflicting answers one hears to this question reflect the shabby state of empirical work in development economics. Answering it in a somewhat more scientific manner may involve not only greater reliance on tools such as sample surveys but also the refinement of welfare indices, and the taking into account of the poor's own perception as to whether they are better or worse off. A modest new rural water supply system may make villagers feel much better off, even though in the national accounts such projects would have a negligible impact on any sort of quantitative indicator. Much of the same can be said about the provision of other public services in the areas of health, education, transportation, etc. How much of the welfare gains arising from the eradication of malaria and the decline in infant mortality have been captured in the national accounts of developing countries? Just about any junk-producing new import substituting industry has had a bigger impact on those national accounts!

United Nations efforts to develop indicators of social welfare could be very helpful. Simple data on numbers of bicycles, radios and shoes in a village, and on materials used for roofing may say more about welfare than "guesstimates" on aggregate per capita consumption. My own favorite measure of welfare and its dispersion among social classes is life expectancy; there is no more brutal indicator of inequality than the higher life expectancy of the rich (and of their children) than of the poor.

On the whole, the sample survey may be the closest instrument we have to those in the experimental sciences, where the researcher typically generates his own hypotheses and his own data. Its greater use in development economics may limit the negative consequences of the impossibility of controlled experiments, as noted by Simon Kuznets: great loss in data economy and proliferation of competing hypotheses that survive for long periods of time [7, 1973, especially pp. 248-249].

Why have development economists been so slow in following the example of their sociology and political science colleagues in using sample surveys? Properly designed sample surveys require team-work, continuity of effort and organization, not to mention faith in empirical research with long gestation periods. These are stiff requirements, particularly to a profession afflicted with extreme individualism, theoretical hubris and, in developing countries, threatened by institutional instability. But the sharply diminishing returns available in further squeezing public data sources is pushing us in the right direction. One hears that in some developing countries, private consulting firms are emerging which offer to carry out sample surveys with a high degree of sophistication.

Autonomous research institutions in developing countries have a natural comparative advantage for this kind of empirical work if institutional stability can be maintained. Here is an area where economists in developing countries could take up an important research and pedagogical leadership role. Such a role would exclude both being simple gatherers of data to be processed in research centers of developed countries, and being gatherers of data which never get analyzed in a systematic fashion. Neither danger should be dismissed lightly.

V

Trade and financial links between developed and developing countries have historically generated a disproportionate amount of data, contributing to the appeal of research in the area of trade and development. The international institutions which emerged from the Second World War have contributed much to stimulate and standardize the production of data on international trade and financial flows. Nevertheless, important gaps remain.

The biggest scandal in this field remains the lack of serious indices of international prices. After many years of arguments about trends in the terms of trade of developing countries, most international agencies still rely on unit value indices for imports and exports in computing terms of trade, in spite of widespread agreement that such unit value indices are rubbish, or close to it, as shown by Kravis and Lipsey [5, 1971]. As in the case of unemployment and underemployment, the amount of theorizing and policy debate on the terms of trade relative to the data base on the phenomenon itself makes the discipline of development economics appear singularly frivolous.

The rise of multinational corporations is a matter of hardly exclusive interest to development economists, but a greater flow of information on their activities would certainly facilitate objective research in the controversial area of their net contributions to development. The emphasis on obtaining data from core economic units, advocated earlier in this paper, should be helpful here. But given the international nature of these units, international efforts at data gathering will also be necessary. The United Nations has started such an endeavor, but the field is a difficult one, as

noted by Somavía [10, 1974]. Part of the difficulty lies in the conceptual weakness of economic ideas in areas such as that of technological transfer; measurement becomes particularly difficult when one is not sure about the nature of what one is trying to measure!

Pressures from the United Nations could be helpful in inducing the home countries of multinational corporations to cooperate in exchanges of information regarding their activities, a cooperation which has not been spontaneously forthcoming until now. These efforts may eventually lead to a commonly agreed body of "stylized facts" on which positive and normative theories about multinational corporations can be firmly based.

Finally, international agencies seem the logical bodies to improve data availability on international migration of skilled and unskilled labor, a subject which is likely to become one of increasing interest to development analysts.

## VI

From all that has been said above, it should be clear that I am skeptical regarding how successful the researcher in development economics will be in generating new insights by just relying on information already accumulated in libraries, whether located in developed or developing countries.

It is doubtful that libraries in developed countries should, or could, efficiently gather data on core economic units or on the small but growing body of sample surveys scattered throughout the Third World. International organizations may be able to provide an inventory of major bodies of such data and their location, particularly household sample surveys, a list which

can then be disseminated through libraries. But the researcher will sooner or later have to go to the source of those data to check on gathering methods, robustness, and definitions.

Indeed, my bet is that more often than not already collected data "lying around," unused and unanalyzed, in the Third World, will prove of minor use to researchers. It is an old game in our profession to dream up hypotheses which can be tested against available data, a procedure frequently accompanied by many "proxies" and "dummies," and leading to murky results. But my conjecture is that the more interesting and focussed a hypothesis in development economics, the less likely is that data "lying around" will be of help in testing it. Like the natural scientists, the pioneering development researcher may have to devise not only his or her own hypothesis, but also gather the fresh data needed for testing.



Notes

\*Comments on an earlier draft by Benjamin I. Cohen, Simon Kuznets and Hugh Patrick are gratefully acknowledged. The paper also benefitted from discussions at the 1975 Kiel seminar of the International Economic Association.

<sup>1</sup>Simon Kuznets has called for more comparative analysis of economic growth experience [6, 1972, especially pp. 80-86]. I share this view, but place that area of research more in the "growth" than in the "development" category, meaning by the latter the analysis of the early stages of growth acceleration and structural change in countries with an income per capita of, say, less than \$1,000.

<sup>2</sup>In Asian countries, B. I. Cohen has found a greater degree of specialization [2, 1975].

References

1. Bergmann, B. R., "Have Economists Failed?" Presidential Address, Eastern Economic Association, Albany, New York, October 27, 1974 (mimeographed).
2. Cohen, B. I., Multinational Firms and Asian Exports. New Haven: Yale University Press, 1975.
3. Díaz-Alejandro, C. F., Foreign Trade Regimes and Economic Development: Colombia. New York: National Bureau of Economic Research, forthcoming, 1976.
4. Fishlow, A., "Brazilian Size Distributions of Income," American Economic Review, May 1972, 62(2), pp. 391-402.
5. Kravis, I. B., and Lipsey, R. E., Price Competitiveness in World Trade. New York: National Bureau of Economic Research, 1971.
6. Kuznets, S., Quantitative Economic Research: Trends and Problems. New York: National Bureau of Economic Research, 1972.
7. \_\_\_\_\_, Population, Capital and Growth: Selected Essays. New York: Norton and Company, 1973.
8. Orcutt, G. H., "Data, Research, and Government," American Economic Review, May 1970, 60(2), pp. 132-137.
9. Ramos, J., "An Heterodoxical Interpretation of the Employment Problem In Latin America," World Development, July 1974, 2(7), pp. 47-58.
10. Somavía, J., "Transnational Corporations: Information Gathering by the United Nations," Development Dialogue, 1974, (2), pp. 11-24.