New Approaches in Economic Analysis

Wassily Leontief, Faye Duchin, Daniel B. Szyld

Summary. The level of detail of economic models has until recently been limited by the availability of computational capability. At present, the constraint is the systematic compilation of detailed and comprehensive economic and technical data and the ability to manipulate them. The effective use of enhanced data processing capabilities will have to proceed hand in hand with a concerted effort to develop the economic database and with the shift from analytic approaches based on aggregative data to those that can take advantage of detailed information.

Powerful computers are playing an important role in transforming economics into a truly empirical science. The significance of that role is enhanced by the increasing emphasis on measuring and counting as traditional methods of production and consumption are replaced by science-based technologies. With the rapid spread of computerization, detailed factual information about every kind of activity performed in the many different sectors of a complex modern economy has become available. No attempt, however, is being made to develop an overall strategy for systematically organizing a comprehensive, fully integrated information system. The coverage in many instances reflects the mission-oriented requirements of different agencies rather than the need of attaining a better, detailed understanding of the changing structure of the entire economy. For example, no information is compiled about alternative technologies used in the same industry or about the details of monetary transactions carried out between sectors.

This lack of a complete picture results in part from the absence of a fundamental organizing principle. The United States is the only advanced country in the world that does not have a central statistical office. Each department of the government collects statistics in the area of its own particular interests. Users of such data spend much of their time trying to reconcile and align information coming from these different sources. But even if more comprehensive databases did exist, analytic approaches that have dominated economic theory have not provided a foundation for making effective and systematic use of them.

Wassily Leontief is professor of economics and director of the Institute for Economic Analysis (IEA) at New York University. Faye Duchin is associate professor of public administration and associate director of IEA. Daniel Szyld is associate research scientist at IEA, 269 Mercer Street, New York 10003.

Aggregation in Economic Models

Adoption of mathematical language has permitted economic theorists to manipulate abstract mathematical models of any size. However, the lack of computational capabilities has imposed an uncomfortably low upper limit on the number of variables that can be contained in a system intended for empirical implementation. A good example of this is Keynesian theory, which was originally conceived as a model to describe the national economy in terms of four or five, but certainly not more than six or seven, variables and equations. The simplification is made possible by the systematic use of such proxy variables as the total level of output (gross national product), average price level, total employment, and total real investments.

Even with a more detailed breakdown, such as subdividing total output into outputs of capital goods and consumer goods, neither these aggregative variables nor the structural parameters appearing in the equations that describe the relations among them can be observed and measured directly. Thus after the basic data have been assembled, the empirical implementation of a typical aggregative model requires two preliminary steps. First, the magnitude of the aggregative proxy variable must be computed. This requires summation of such heterogeneous magnitudes as the outputs of steel and of cotton fabrics (to measure the total output of manufactured goods) or the averaging of the prices of oranges and of television sets (to measure the price level of consumer goods). Second, the values of parameters to be entered in the equation of the particular aggregative model must be determined. Because such parameters describe the relations among unobservable aggregative entities referred to above, they themselves cannot be observed and measured directly but must be derived by means of various methods of indirect statistical inference. These methods, as time goes on, become more and more sophisticated, but not necessarily more reliable. This is because the properties of the distributions of the aggregative variables from which they are derived are usually unknown and can hardly ever be known.

One of the fundamental tenets of modern, so-called neoclassical economic theory is the principle of optimization. Its use in practical numerical computations has up to now been limited to the solution of special technical problems in operations research. In the more central field of economic analysis, the vagueness of aggregative formulations has militated against the effective application of this highly sensitive criterion.

Aggregation has achieved simplicity of formulation at the cost of realism. The bulk of computing performed today in connection with conventional econometric analysis consists of carrying out multivariate regression calculations. The availability of more powerful machines permits experimentation with a greater variety of curve-fitting formulas and a larger number of different combinations and permutations of the same aggregative variables that were used before or new aggregative variables defined in a different way. It is as if stronger and stronger motors were installed in fishing boats to move larger and larger oars instead of to turn propellers better suited for efficient use of the new source of power.

It is not surprising that the first practical application of the optimization principle, carried out by George Danzig in the mid-1940's when he faced the problem of maximizing the output or minimizing the cost of production of a system consisting of a large number of interdependent industries, was based on a highly disaggregated input-output matrix. Danzig solved the system by inventing the well-known simplex method of linear programming. Development of various methods of nonlinear programming and more efficient computational procedures (some of which are described below), along with the creation of more comprehensive databases and the introduction of more and more powerful computers, will open up the possibility of applying the optimization principle to larger and larger economic systems represented by increasingly disaggregated models.

Development of the Input-Output Model

The first economic model formulated explicitly to take advantage of large amounts of economic data without losing the meaning of the information was the input-output model, whose database in its simplest form consists of a rectangular input-output matrix. The *i*th column of the matrix registers the purchases of one sector from every other sector of the economy including primary inputs, and the *i*th row shows that sector's deliveries to every other sector including so-called final demand sectors, such as households, and investment (Fig. 1). Most of the information used for compilation of national input-output tables comes from the population and economic censuses.

The first use of a large automatic computing machine for analytic purposes in economics was made in 1935 (*1*). The computer used was the Mechanical Simultaneous Equation Solver, invented and constructed at the Massachusetts Institute of Technology by John Wilbur, and the database comprised the first true input-output tables of the U.S. economy. The computation yielded a general solution by inversion of the matrix of an unusually large system of linear equations describing the interrelations among the levels of output as well as among the prices of goods and services produced and absorbed by 46 different sectors of the U.S. economy in the years 1919 and 1929. Because of the limited power of that early computer, these 46 sectors had to be combined into 19 sectors. The principal purpose of these calculations was the quantitative assessment of the effects of technological changes between the years 1919 and 1929 on the level of output and on corresponding employment and relative prices of products in different sectors of the U.S. economy.

Structure of the Economic Database

By the mid-1960's, a large volume of economic and social data was being periodically collected by different government agencies, and various specialized surveys were being conducted on all aspects of the economy. Creation of a centralized Federal Data Center was recommended, to be charged with reducing duplication, integrating different data sets, and providing easy but controlled access to what began to be called the economist's database (*2*). Little if any progress has yet been made in this direction.

These data were often published in printed reports, and in the 1970's they began to be available systematically in a form that could be interpreted by computers. However, most economists' data needs were satisfied by relatively small time series, which are simple data structures (for example, a sequence of independent observations each containing several variables). Most economists' computational requirements were satisfied by generalized, prepackaged software for estimating statistical parameters and projecting variables on the basis of regression equations and other techniques of statistical inference.

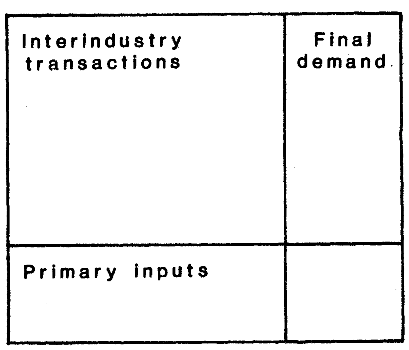

Fig. 1. Input-output table showing final demand, primary inputs, and interindustry (or input-output) transactions.

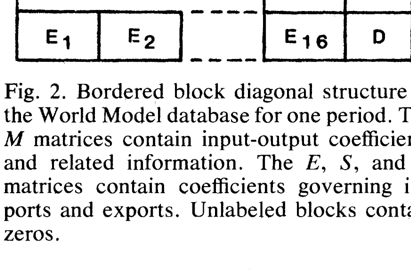

Fig. 2. Bordered block diagonal structure of the World Model database for one period. The *M* matrices contain input-output coefficients and related information. The *E*, *S*, and *D* matrices contain coefficients governing imports and exports. Unlabeled blocks contain zeros.

Because of the dominant position of time-series analysis in present econometrics, much emphasis is placed on the length of each time series and the definitional comparability of its members measured at different times. In the case of a time series describing, for example, the production of capital goods over a period of several decades, these two objectives are irreconcilable because of fundamental changes in the nature of production. Furthermore, insufficient attention has been paid to describing the interrelations (rather than making comparisons) between different variables, such as the output of aluminum and the input of electricity used to produce it.

In corporate information systems, essentially the same problem is being tackled by shifting from file management to database management, a transition requiring the integration of separate files of data. In business applications this integration is often based on paths of access. Two variables are related if, for example, they both appear frequently in the same report. In research, this integration must of course be provided by theory, which imposes structure on the database.

Input-Output Tables and the National Economic Database

Input-output tables are sometimes viewed as an extension of the System of National Income and Product Accounts. In this system, gross U.S. expenditures for personal consumption, private investment, and other items (also called final demand) are tabulated for each year, as well as gross income in wages and salaries, corporate profits, and the like (also called value-added or primary inputs) (Fig. 1). Efforts are made to maintain comparability of the aggregated annual figures over long periods of time. Thus the Accounts have the same limitations as all time-series data.

Originating from double-entry business bookkeeping, the System of National Accounts depicts economic transactions as value flows and measures them in dollars. The corresponding physical amounts are described indirectly by mostly aggregative index numbers, in which the flows originally accounted for in current dollars are revalued in terms of the prices of a somewhat arbitrarily chosen base year. In construction of input-output tables, this conventional device is being slowly but systematically superseded by the introduction of real, physical units.

The transition to physical units enables a detailed input-output model to depict the relation between an economic activity and its natural environment. Some recent input-output tables show, for instance, the number of tons of sulfur dioxide produced by different sectors and released into the air and the millions of cubic feet of water, among other inputs, absorbed by different industries. Most important, the use of physical units makes it possible to use direct engineering information in input-output tables.

An Example: The World Model

A good example of the direction in which analytical use of the input-output approach might advance is provided by the World Input-Output Model (*3*). Sixteen regions of the world are represented by about 55 producing sectors and are linked together through the trade in individual goods and services at 10-year intervals for seven decades (1970 through 2030). The emission and abatement of different pollutants are represented, and the cumulative use of individual raw materials is monitored. The model integrates input-output tables of many countries for different years into a highly structured database.

For each period, the World Model can be regarded as a linear system of about 3300 algebraic linear equations. The associated matrix has what is called a bordered block diagonal structure (see Fig. 2). Because the inversion of a matrix usually requires a number of operations on the order of the cube of the number of its columns, in this case it would take more than 35 billion operations to compute the inverse matrix and more than 10 million memory locations to store it. On a powerful sequential computer that operates at a rate of about 1 megaflop (1 million floating point operations per second), this formidable task would take about 70 hours of computing or 3 days for the full seven-decade calculation.

The first method for the solution of the World Model developed in the mid-1970's reduced the problem to a series of linear systems whose matrices of coefficients M_i are the 16 diagonal blocks in Fig. 2. In this way, the inversion of the whole system was avoided. Nevertheless, the matrices M_i were inverted, and the inverses were stored (*4*).

The analytic use of the World Model and similar models resides in the possibility of comparing the results of alternative sets of assumptions (scenarios). Several computations are performed and their results saved for later display, analysis, and comparisons. The software originally developed made it possible to change selectively the figures in specific rows and columns as the basis for a new scenario—for example, a modernized technology for making steel in India in 1990—and to tabulate selectively the results of a computation. One solution involving the full description of a scenario and matrix inversions took several hours of computer time, could be run only once a night, and cost several thousand dollars. For these reasons, scenarios were simple and few in number, and approximation techniques were developed to minimize the number of full-matrix inversions.

The Use of Sparse Matrix Techniques

As in most highly detailed economic models, the World Model matrices are very sparse; that is, few entries are nonzero because each variable is directly related only to a relatively small number of other variables. Techniques have been developed since the early 1970's to solve sparse linear systems. Sparse storage techniques record only the nonzero entries so that testing for the existence of zeros is not required.

The solution of the linear system is obtained by factorization of the matrices, which makes it possible to avoid computing the inverses of the matrices. The number of operations required for the factorization of a sparse matrix and the solution of the entire system is proportional to the number of nonzeros in the factors; with appropriate numerical techniques, the number of nonzeros in the factors can be kept close to the number of nonzeros of the original matrix. In other words, the computational cost of solving a sparse linear system is one or two orders of magnitude smaller than that of computing the inverse. Because the factors remain sparse whereas the inverse of the matrix is generally full, storage requirements are also lower. The implementation of sparse matrix techniques for the World Model has made it possible to contain the on-line storage requirements and to obtain a full seven-period solution in about 3 minutes of central processing unit (CPU) time (*5*). Extremely fast methods of solution and the prospects for flexible software and database management have led to the decision not to store any factors (the equivalent of storing inverses to update them or reuse them) but to recompute them for every new scenario. Sparse factorization techniques are being introduced as a standard in large economic models (*6*).

The Use of Supercomputers

The current generation of supercomputers can achieve peak rates of 100 to 200 megaflops in certain portions of the calculations by simultaneously performing a series of tasks. Rewriting large portions of the World Model codes has resulted in full seven-decade computations in less than 1 minute on a Cyber 205 (*7*). Now, models one order of magnitude larger can be considered. To our knowledge, the World Input-Output Model represents the only instance in which a supercomputer has been employed for large-scale economic modeling (*8*). This research provides a glimpse of the kinds of computational and programming problems that can be expected to arise in the future.

Computers with thousands of parallel processors are being designed and tested in several research facilities on campuses across the country. An example is New York University's Ultracomputer Project, which is currently experimenting with an eight-processor prototype of a machine designed for 4096 processors (*9*). The effective utilization of this computational capability in the analysis of a real economy will require the increasing integration of the conceptualization of the economic model and its implementation by computer. A team research effort—a practice long established in the natural sciences but still rare in economics—is possibly the only way to proceed.

A Dynamic Model of the Economy

With the implementation in the 1980's of a dynamic input-output model describing the U.S. economy year by year from 1963 through 2000 (*10*), the structure of the input-output database has again been elaborated, this time to include a detailed description of each sector's use of plants and equipment and various types of labor. Three matrices have been added to the basic input-output transactions table. One of these matrices describes the structure of each sector's purchases of plants and equipment to replace the existing capital stock. The second describes capital requirements to expand each sector's capacity to produce through new plants and factories. The third matrix describes each sector's requirements for labor of different qualifications. A set of dynamic input-output equations specifies the relations among the four matrices in any year and also relates the production of capital goods in one year to their subsequent use. The computational approach developed for the World Model has been adapted for this model.

The Present Situation

The extension of the traditional curve-fitting methodology to its upper limits is represented by the model of the world economy used in Project LINK (*11*). This large-scale economic model links together an increasing number of large and small, annual and quarterly, and national econometric models of the Keynesian type by means of additional equations describing international trade in four aggregative categories of goods.

Today, standard econometric procedures based on aggregation and statistical inference still dominate the field of economic research. However, the input-output approach, designed to make the fullest possible direct use of detailed factual information, is now being employed in most of the major areas of economic inquiry. Compilation of at least small national and regional input-output tables is being carried on in all developed and most developing countries. In Norway, for instance, and particularly in Japan, systematic compilation of such a database is considered to be one of the principal tasks of the official statistical organizations.

As the demand for more realistic modeling increases, detailed engineering and other types of technical information are beginning to supplement and, in some instances, even to replace the more conventional sources, such as census figures and accounting records, in the compilation of input-output information. The cost of constructing and maintaining the comprehensive database that would be needed to carry on an analysis of the operations of a modern economy useful for public and private decision-making is bound to be high—much higher than the amounts now spent by official statistical agencies such as the U.S. Bureau of the Census. At present, however, the collection of statistics by the government seems to have been gradually deteriorating. Partly as a result of this, corporate data gathering is on the increase. Privately collected information tends to be fragmentary and, because of its proprietary character, cannot be made available (except at prohibitive cost) for scientific use. Economists may be facing the unenviable prospect of entering the information age without sufficient information.

As recently as thirty years ago, effective application of the modern model-building approach to the study of large economic systems was restricted by the absence of adequate computing facilities. Further progress in the field will be limited by the lack of requisite factual information. In the long run, economics as an empirical science will be able to take full advantage of the immense data processing capabilities of modern computers only by modeling the economic system in very great detail and by creating the large, comprehensive, and at the same time detailed database required for the implementation of such large models.

References and Notes

1. C. Eames and R. Eames, *A Computer Perspective* (Harvard Univ. Press, Cambridge, 1973).
2. For a brief history through the late 1960's, see N. D. Ruggles, in *Economics*, N. D. Ruggles, Ed. (Prentice Hall, Englewood Cliffs, N.J., 1970), pp. 1–27. For more recent history, see C. G. Renfro, *Rev. Public Data Use* **8**, 121 (1980).
3. W. Leontief, A. Carter, P. Petri, *The Future of the World Economy* (Oxford Univ. Press, New York, 1977). The model has subsequently been elaborated and used in several other studies.
4. P. Petri, Report No. 3, Department of Economics, Brandeis University, August 1977.
5. For a full description of the solution algorithm, see F. Duchin and D. B. Szyld, *Economics of Planning* **15**, 142 (1979); D. B. Szyld, in *Sparse Matrices and Their Uses*, I. S. Duff, Ed. (Academic Press, London, 1981), pp. 357–365. Detailed timings can also be found in D. B. Szyld, in *Cyber 200 Applications Seminar*, J. P. Gary, Compiler (NASA Conference Publication 2295, Washington, D.C., 1984), pp. 279–286.
6. See for example K. R. Pearson and R. J. Rimmer, in *Proceedings of the Sixth Biennial Conference of Simulation Society of Australia* (Adelaide, Australia, 1984), pp. 103–108.
7. Computations on Colorado State University's Cyber 205 were made under a grant from Control Data Corporation. Current research on the World Model on that computer is being carried out under a grant from the Institute for Computational Studies at Colorado State University.
8. C. Sims, report to the National Science Foundation, 1985.
9. A. Gottlieb *et al.*, *IEEE Trans. Computers* **C32**, 12 (1983).
10. The model is described in F. Duchin and D. Szyld, Institute for Economic Analysis Report, March 1985. It was first used in W. Leontief and F. Duchin, *The Future Impact of Automation on Workers* (Oxford Univ. Press, New York, 1985).
11. V. Filatov, B. G. Hickman, L. R. Klein, in *Global International Economic Models*, B. G. Hickman, Ed. (Elsevier, Amsterdam, 1983), pp. 29–52.

Computers in Production Agriculture

Donald A. Holt

Computers are beginning to play important roles in production agriculture, just as they do now in many other industries. In general, they are being used in data capture and processing, automatic control, and as decision aids. Eventually, through a combination of communication, monitoring, analysis, simulation, expert systems, and automatic control, computers will make large, efficient farm operations even more productive and efficient than they are now. The spectrum of possible benefits is projected in the following scenario.

Scenario

It is a clear June morning somewhere in the midwest. During the night, the farm computer automatically dialed several local and national databases to obtain information on current fertilizer, seed, fuel, and pesticide supplies and prices; weather; markets; insect and disease predictions; and buyer offers. Now it turns on the radio, which gently awakens farmer Bob with music. After a few minutes, information gathered and processed by the computer during the night appears on the bedroom monitor.

Sensors in nose rings, ear tags, and implanted devices have been scanned to assess the physiological condition of the farm's animals. Confined sows and cows coming into estrus have been identified automatically by sensors monitoring mounting activity and vaginal secretions, and they have been scheduled for receipt of frozen embryos.

The automatic feed grinders and mix-

Donald A. Holt is associate dean of the University of Illinois College of Agriculture and director of the Illinois Agricultural Experiment Station, Urbana 61801.