SUMMARY OF THE CONFERENCE

RICHARD RUGGLES
Yale University and National Bureau of Economic Research

I think I might start by saying just a word about the purpose of this conference. It was meant to be exploratory in nature. What we were interested in exploring was the impact of the computer on economic and social research methods. The conference was purposely set up to be fairly wide-ranging. We chose as topics those areas in which the computer appears to be having a major effect. In order to achieve the objective, broad coverage was needed, and a number of examples within each of the various areas, to make sure that the range of work being done was fairly represented. The result of this approach was a large mass of papers and a diversity of subjects and participants, and we suffered through long sessions and long days. In this connection, I would like to thank not only the participants for their patience in enduring these long sessions, but also the translators, who had to listen to it all and to reproduce it when at times it didn't seem to make sense.

My presentation this morning is going to be a discussion of what this fishing expedition has netted—essentially a resume, or summary, of what I consider to be the high points. Now obviously this is my personal interpretation.

The conference started off with the session on government statistical systems. The purpose of that session was to ask where governments now stand with respect to computers. We would have liked to have had a wider representation of governments from Latin America, but this would have just lengthened the conference. The constraints of time meant that we had to start with just a few.

I think the history of the Brazilian experience as described by Kerstenetzky and discussed by Janes de Souza provided a very suitable point of departure for this conference. They explained how the new technology and the advanced hardware, superimposed on the old systems of procedure, inadequate staffing, and so on, resulted in the disaster of 1960, with no census forthcoming at all. But out of the ashes arose a new organization that was more centralized and that was based upon a computer bank of microdata. The new organization was set up to process the 1970 census, and we are told that within three to four months after collection of the data, tabulations were available. This, indeed, is a happy ending to what was originally a disastrous computer experience. I suspect that this experience has been repeated more often in both Latin America and the United States than we are willing to admit. Normally, these things are buried, but the Brazilian experience is so well known that there is no use burying it; it might as well serve as an object lesson.

The Mexican story as reported by Gleason was quite different. They started serious work on computerization for the 1970 census around 1968. They trained personnel, and they took over software already developed by the U.S. Census Bureau, gradually putting into operation techniques that were fairly well advanced but tried and true. The result was quite successful—a substantial improvement over the 1960 census. As Sourrouille pointed out, much of the success of this operation was due to the technology transfer in international cooperation.

With respect to the U.S. Census, Brown gave us a glimpse of, perhaps not 1984, but at least 1980. He pointed out that the computer was now trained to talk: it could hold press conferences on business and population, it could provide tract analysis for planners and local organizations, and it could even give congressmen profiles of their districts. There are really two implications in this: one is the ability of the computer to take statistical material and provide text, highlight important elements, do comparisons over time, or perhaps even provide comparisons between regions, all in nice words rather than in tables. But, and this is really much more important, an ability has been developed to define specific geographic areas and report on them, so that, essentially, the programs were adapted to yield data on regions defined in any way that might be desired. Ferran's point was that insofar as this procedure just saves time or improves public relations, it is not relevant to our present concerns; what we really need is better information for professionals.

Now, I think that the Census Bureau has actually sold itself short in this presentation of its work on geographic coding, because there are in it considerable implications for research. In fact, the Census Bureau has built for the United States a system called DIME, which stands for Dual Independent Map Encoding System. In this system, addresses are geographically coded so that they can be located by block face within a community (a block face is the area on one side of one street between two intersections.) Once given this coding system, data contained in other files of the city—for example, education statistics, health, police, fire—all can be merged and mapped onto block faces. This gives local planners and officials in local government the ability to analyze their own data about their city in relation to the demographic and social characteristics provided in the census data. Some two hundred and eighty cities now have the DIME system of geo-encoding. The Census Bureau provides to the cities all the necessary computer programs to take their local records and match them with the census records.

In all fairness, too, quite apart from the geo-encoding work, the U.S. Census is providing researchers with masses of information for social and economic research. The 1970 census will provide more than three-thousand tapes for research use. The Bureau is also creating a set of Public Use Samples, each of which is a 1-in-100 sample of the population, approximately two million individual households. There are six different samples containing different information in the set.

The paper by Fellegi and Goldberg describes the decision of Statistics Canada to keep their basic statistical information in the form of microdata. They are doing this because they have found that the ability to reprocess the microdata as needed gives them much more flexibility, making it possible to produce any required tabulations on demand. Also, they are finding that different kinds of economic analysis call for combining information in different ways, and this is only possible if the microdata are available.

The heart of the problem of preserving data in micro form, as Fellegi pointed out, and as Howe also recognized, is the editing that goes on in creating microdata files that are usable for a wide variety of purposes. Howe raised the question as to whether the editing did not in fact homogenize the data to a considerable degree. We will come back to this point because it was discussed a great deal during the conference, but part of the answer may be that the outliers may, in the aggregate,

be of limited significance, and their correction may improve the usefulness of the data for microanalysis by making the data sets more consistent.

Fellegi also pointed out that Canada is involved in developing a central register of all companies and establishments, giving their industry totals and so on. I believe that in Canada at this moment, this register is held within the government, but it does help provide the consistency that is needed for classification of information from different sources. I understand that in the United States, the President's Commission on Federal Statistics is going to recommend such a system. Such a register would be a great help, not only to the various government agencies but to private outside research users.

Fellegi also stressed the tremendous need for generalized programming. I think that everyone who has been involved in computerization has been depressed at the rate of obsolescence of programming when different machines come along, or when different uses are made of data, or when different data sets are introduced. There is a great need for generalized programs to handle a wide variety of information. Statistics Canada has gone a great distance in this field.

Finally, Fellegi worried about the confidentiality problem and indicated that they were very careful in their custom processing and tabulation not to allow the use of the microdata to result in disclosure. They do give access to their microdata by allowing people to run programs on it so long as this does not result in disclosure. This same technique, incidentally, is being used in the United States by the Social Security Administration, by the Census, and by some other federal agencies.

Carlos Petit questioned whether a microdata basis for stastical systems was applicable to Latin America. Fellegi's reply was that he felt that it was. If the data is fragmentary, it is much easier to squeeze meaningful information out of a microdata set than it is to make a reasonable macrotabulation that could serve a wide variety of purposes.

This session, in general, shows that computer technology in governmental statistical organizations, both in Latin America and in North America, is moving very fast indeed. People are moving toward retaining information in its original micro form and then manipulating it as needed for specific statistical purposes. Governments, instead of publishing large tomes of tabulations such as census bureaus did for many years, are beginning to make machine-readable data available on tape. This opens up different kinds of research techniques to users. After all, I can remember keypunching some of the tabulations that used to come out of the census, and having the published tabulations available on tape is a far step forward in that it not only reduces the cost of using information but it certainly improves the quality as well.

The second session, on data banks and computer centers, was essentially an attempt to find out what their current status was. It is indeed unfortunate that we had no report from CLACSO [Latin American Council of Social Sciences], but we did have a very interesting report from CELADE [Latin American Demographic Center] that indicated that they hold a substantial sample of Latin American censuses. Now these samples are mainly demographic and were aimed at the study of fertility and migration; but, I think, one of the most interesting things about this data bank is the avenues which it opens up for comparative analysis of different countries in the area of demography. I would hope that at some later time they will

add to their bank some of the other information aside from the purely demographic —such things as education and consumer budget studies—that we have heard of here. The interrelations among the demographic, social, and economic characteristics are important, and only by bringing this data together into a single data bank can its maximum potential be realized.

Domingo reported, with respect to computer centers in Venezuela, the existence of numerous computers, more or less modern. The major problem was a shortage of qualified personnel to use them. Rojas also agreed in this session that this was a major problem—that the large centers were more efficient, but the small ones were easier to start and control. I found these remarks particularly interesting because, not so many years ago, we, too, were starved for computer programmers, while now they're unemployed. This is not purely a function of economic recession; it is a function of the fact that the technology and training spread very fast. All of our graduate students practice programming; the undergraduates earn their money by programming—it is the sort of thing that people are picking up very rapidly. Part of the reason is that programming languages are becoming simpler, computers are becoming more powerful, and the problems at the applications level are not nearly as horrendous as they were a decade ago.

The question of the relation of large and small computers is also very interesting. I have been impressed at the small cost of hardware relative to other costs in running a computer center. The nonhardware costs of running a small computer center are almost as much as those of running a large computer center—you need your operators, your programmers on duty, your assistant programmers, and so on, in any operation. At Yale, we have had two or three computers and two or three staffs. It is not quite as bad as it might have been, because, theoretically, there is one head over them all—but even then, the waste of resources in essentially duplicate facilities of lower power is very marked indeed. In countries where there are shortages of both trained manpower and hardware, the sharing of facilities does seem to me to be a sensible approach, especially in the short run.

Perhaps paradoxically, in the longer run you can afford to have smaller computer centers rather than larger ones. The paper by Boschan on the NBER data bank gave us, I think, another view of things to come. The National Bureau data bank covers two-thousand time series on national accounts, financial data, and other economic data about the economy. A number of people have mentioned to me that they would just love to be able to sit down at those terminals and use that kind of information. I am always rather paralyzed by such an opportunity: there are such a variety of things to do, and everything I can think of seems so naive that I am somewhat stuck. On the other hand, the potentialities are tremendous. The NBER bank is in operating order; it is updated currently and there are some fifty commercial users, as well as five universities. The method of use is through interactive terminals and on-line storage of the data. At the present time, four computers are involved—two GE computers, a PDP-10, and a 360/67. There are different software packages on the different computers. One of the most sophisticated ones that is now available is the TROLL simulation system, which is up on the 360/67 computers at MIT and Yale. This is a rather interesting contrast with the Domingo and Rojas discussions of computer centers. Here you have a data bank without a computer center. As a matter of fact, the same terminal at

the same place can reach different computer centers merely by using different telephone numbers. So it may well be that with interactive terminals, the actual location of the computer itself and even that of the computer center may become less relevant. Another problem that is bound to arise is that of cost. There are economies of scale; and once you have created a thing of this sort, in which the marginal cost of using the data (including the marginal computer cost) is so low, all red-blooded economists feel very unhappy if it is not made available as a free good. It is the old whether-the-bridge-should-have-a-toll problem.

Watts added another dimension to the discussion of the microeconomic data bank. He recognized the tremendous value of creating microdata sets, but he offered very sobering advice: he said to start small and develop competence with one body of data, to avoid the latest technology, to use a person-plus-machine approach, and to focus the data bank on specific needs and specific analyses. I am sure that if these rules were followed, one would not go wrong; and that had the Brazilians in their early stages thought in these terms, the disaster would not have occurred. But there is a kind of pessimistic flavor in all of this, which contrasts sharply with the achievements with the Survey of Economic Opportunity (SEO) Files, which Watts himself created. That SEO File is not designed for specific needs and specific analyses; it didn't start all that small; and it didn't develop competence necessarily with one body of data. The person-plus-machine approach didn't really work out there either. Let me give you just a brief history. The file was first turned over to Brookings, which engaged in extensive cleaning. It was then given to a firm called ASSIST for distribution, but this firm went bankrupt. I got my copy of the file from ASSIST, and it was never readable. From ASSIST, the file went to Wisconsin. In the meantime, some bootleg versions that were uncorrected had spread to the University of Georgia, Cornell University, and a number of other places. In order to get help, I went down to Georgia to get copies of their tapes, and we called Brookings up and got a little more help; it was chaos, but the point is that it worked. We have now bred, I think, four or five descendent generations which have gone to other users. I don't think that Harold Watts realizes the monster that he really created, or he would be more optimistic. At the present time, the SEO tapes are used for a wide variety of projects at the Urban Institute, Wisconsin, Yale, and other places. We have already had a number of theses started, using them in such areas as poverty, income distribution, unemployment, education, demography, and time-budget studies; and, of course, we have had a presentation here on their use in creating the MERGE File for tax analysis.

This reminds me that at a workshop on microdata held in Washington, D.C., a year ago, Okner presented some of the early results of his work on the MERGE File, and he had very much the Watts point of view—he was very pessimistic. At that time, he said, "I'd never do it again; you couldn't get me to do it." When I pointed out to him at this meeting that his paper had quite a different flavor, he said, "I got results." Perhaps, when Watts is closer to the end of his project, he may be happier.

Lebergott took a somewhat lighter view of this whole problem, indicating that, in the end, virtue would triumph. He also added another dimension that we have not discussed at all in this conference, but which I feel is probably one of the

most important elements—namely, that the survey technique is not the only source of information. Administrative records—such as invoices, bills of sale, and things that result from the current operation of economic systems—provide very valuable bodies of data, which we should utilize; if these are gathered and linked and integrated, we will have types of data that are absolutely impossible to get from survey information. I have been particularly interested in the area of alternative sources of information on consumer expenditures, because I feel that there is a very real limitation as to what you can find out from people about their expenditures: they just don't know. They certainly don't know with respect, say, to brands and sizes. Not too long ago, a student of mine did an undergraduate thesis analyzing the load sheets of supermarkets for this purpose. These load sheets indicate wholesale price and retail price and the amount sold by each supermarket for some ten-thousand items, twice a week. The stores represented were located in both poverty areas and non-poverty areas, so that he could examine patterns of consumption and substitution from one item to another when prices changed in both poverty and non-poverty areas. Thus, I think that we are overlooking very valuable sets of information if we just think in terms of censuses and surveys.

After considering data banks, the conference shifted its focus from the processing, storage, and retrieval of information to the analysis of information. The Vargas paper pointed out that the model which they were developing was providing a framework for examination of data consistency and gaps. It is certainly too early to appraise this work because it is in midstream, but this paper emphasized the importance of the feedback from the model to the data, and from the data to the model. This is a very important point indeed. We will be coming back to it; and it will, I think, have major implications for the conclusions of the conference. In a sense, it reinforces Manne's point when he said that a model is a prisoner of the data. Well, it is only a prisoner if there is no feedback; but if there is sufficient feedback, and if the modelmakers either go out and get their data or manage to get other people to do it, they're not prisoners of their data.

The Okner and Pechman paper illustrated a completely different approach from that of a macromodel. What they were considering was the merging of two very different data sets that had some information in common. The object of the game was to map onto the one data set information from the other. The two files they used were the SEO File, which gives extensive demographic and social information about households, and the tax file, which is a sample of individual tax returns provided by the Internal Revenue Service. The idea was to produce a union of the two files which would have the desirable statistical characteristic of eliminating the bias in the selection of individuals which is inherent in the tax files, and which would provide a larger set of information than is available in that file. What is being done here reflects a very interesting and, in a sense, revolutionary idea: what we are trying to do is to create synthetic data sets of make-believe, but statistically valid, people or—if you want to put it another way—we impute to certain households information (perhaps on a stochastic basis) that seems reasonable, given our knowledge of that household. Now, this has two implications: first, it makes diverse samples very much more valuable, since it provides a method of merging the information in the samples; and second, it overcomes the problem of privacy because the people in this new synthetic sample do not really exist—it

is just that the sample has the statistical characteristics of the data sets on which it is based. The validity of this procedure depends a great deal upon the overlap that is available in the two data sets.

As a national accountant, I can appreciate these problems very much, because the national accountants have been doing the same thing with macrodata all along. They essentially go to one source of data to get one type of information and to another source for another type. The various sources are seldom consistent, and the national accountants must make up their minds how to juggle them to bring them into consistency. As a matter of fact, Morris Copeland, in a recent communication, pointed out that we really should have statistical discrepancies in every account rather than just one, and there is much to be said for this.

There are biases in almost every set of data, and any imputation requires a theory of what is related to what. If you do not have a good understanding, you are going to mismatch data; and the intercorrelations among your variables will be very bad indeed. However this is not as difficult as it might appear, because there are several ways to find out about the intercorrelations among variables. For example, there is in the United States an unfinished study called LINK (not to be confused with the international project called LINK). This is a cooperative effort of the Census, the Internal Revenue Service, and the Social Security Administration, bringing together their records for identical individuals. Now, if this is done for just a sample, the principles of linkage can then be derived from this small sample and applied to other samples. Similarly, where difficulties arise in linkages, special samples can be collected which give more information about individuals, and this information can be used to develop theories of how linkages should be made. The payoff in the Okner-Pechman approach is that it yields income distributions with much richer social and demographic and income data than we have ever had before. The national accounts people in the Commerce Department have also realized this—this is not just an isolated gleam in Okner and Pechman's eyes. In the Bureau of Economic Analysis, Edward Budd is making a synthetic microdata set based on the Current Population Survey and the tax file, aligning it with national accounts. The income-distribution statistics in the future in the United States will be based upon such a microdata set. In the Brookings case, the use of the tax file for analysis of tax policy has been outstandingly successful. It has been so successful that the Internal Revenue Service has taken it over; they don't stir an inch without consulting the tax model to explore the implications of changes in the tax law upon receipts, distribution, who gains, who loses, and so forth.

A paper using a related technique but focused on different questions was presented by James Schulz. He undertook to simulate the effects of the social security system on the aged population in the future. He wanted to examine the income distribution of the aged population in the year 1980, given the present configuration of social security, private pension plans, and other savings that individuals might have. In order to do this, he created a life-process model which aged each person a year at a time, using a stochastic approach. For example, he first asked whether the person lived or died. To do this, he looked up in a mortality table the probability of death for a person of those characteristics and then drew a random number to apply that probability to the particular case. The basic data used was the Census 1-in-1000 sample for 1960, but a great deal of other information

was added to this—information on labor turnover, unemployment, retirement, work status of the wife, mortgages, and other financial information—with the result that he was able to create an income statement and balance sheet for each household, adding savings in each year. The result of this analysis not only provided him with a distribution of income for the aged in 1980, given his initial institutional assumptions; but also, like most simulations, permitted sensitivity analysis to see how the income distribution changed with various changes in policy, or in assumptions, or in exogenous factors, and so on. Recently Schulz has extended this work to the social security systems of Austria and Sweden, applying their provisions to his base population. This allows him to appraise the effect upon his given population of alternative social security systems. This, of course, is very useful for revising our own social security system, and for getting some appreciation of how social security systems compare with each other. The basic principle in this work is applying the analysis to a set of microdata for households, household by household, and then tabulating the results at the end, or as might be said, running a census at the end of the period.

The paper by the Urban Institute group was along much the same line, except that the project it reports upon is much more ambitious. Rather than devising a special-purpose tool, the Urban Institute project views this technique of microanalytic models of household behavior as a general-purpose tool which can be applied to a wide variety of studies relating to economic and social problems of the household. One of the principal segments of the project was the demographic simulation of births and deaths and marriages and divorces. This microanalytic model is to be imbedded in a macro system of accounts which will provide an environment for the microsimulation of household behavior. I think that the methodological importance of this type of research is very great indeed. It is a long-range project—probably longer range than the people in it would like it to be—but it is very ambitious, very comprehensive, and tries to take an integrated approach to the whole of economic analysis in this particular area. It provides us with a tool for studying social problems—health, education, poverty—problems which our macromodels are not suitable for.

The final paper in this session, by Gamba, described a sectoral simulation of the Argentine beef industry. This was neither macro nor micro, but somewhere in between. It demonstrates that for particular sectors of the economy, partial simulation on a disaggregated basis may be very useful. I think Kresge's point that it would be interesting to explore this with the more recent simulation packages and to see how the results compare was extremely pertinent.

Next, we came to the session on macroeconomic models—the bread and butter of economists. De Souza and Monteiro gave an overview of seven different models for Brazil. Success was not claimed for any of them—there were major difficulties about the price-level problem—but the authors came, as everybody comes in this area, to the conclusion that what they needed was larger, more complex and sophisticated models.

The Klein-Beltran del Rio paper gave us one of those larger, more complex and sophisticated models. I think there were about 140 equations, with 40 of them behavioral, and some 46 exogenous variables. This model was not taken over from the models of developed countries directly. It recognized that for a developing

country, special features are needed to take account of bottlenecks, the role of agriculture, the income distribution, internal migration, and the process of capacity creation. In the Mexican model especially, the special trade features with respect to the United States and the proximity to the United States are taken into account. The Klein model was of a long-term nature—not long term in the sense of time-period covered, but long term philosophically; it covered a four-year period, showing the output-price relationship which could be expected under deflationary policies and expansionary policies over the four years.

The Manne model was also on the Mexican economy. This paper, which is part of a much larger study, also presented alternatives—eight different alternatives —for Mexico, using a special model developed for this purpose. I was struck with the similarity of the results of these models, despite the fact that the models themselves were very different and were done by different people; this fact is indeed encouraging.

The Foxley model that was developed for Chile dealt with some 15 sectors and 115 variables. It was built with special emphasis on the copper sector, which has accounted for 70 percent of Chile's exports. The model was used to analyze the export subsidy and expansion policy.

The conclusion of this session was that the macromodel does provide a description of the economic system in a general, comprehensive way. Ranis pointed out that, obviously, different economists view these different models in different ways, and that there are widely differing views as to their suitability for handling a particular problem. Manne stressed the importance of information; he felt that were the information base different, his model would also be different. Denise Ford bemoaned the fact that the income distribution somehow got lost in macromodels, and I don't think that there was any specific suggestion as to how it might be introduced into them, although I am sure that the modelmakers are ingenious enough to do it. There was also the suggestion (and I believe a general conclusion) that the models themselves could be successfully expanded in a number of directions; that input-output could be introduced to take care of the production sector; that flow-of-funds information could be included to improve the treatment of the monetary sector; that short-term models could be related to long-term models; and that models in different countries could be linked with each other, as is being done in the LINK project.

The next session moved to demography, manpower, employment, and education. This was a much more heterogeneous session, which showed the wide variety of different uses of demographic information.

Lopes started out emphasizing the need for family data relating to the household, the family, and subfamily, and he presented some work in which he had united these concepts. It is actually very paradoxical; if you examine the microanalytic models of the Urban Institute, or of the other groups, it is so obvious that the household is of major importance that one wonders why it was never really taken into account in demographic information before. I think the reason is one of technology: prior to the computer, the head count was the most important thing, and tabulations were the major products. It is only now that we are able to examine the micro information; but certainly for microanalytic models, we need the household concept.

Grajal dealt with the short-term demand for manpower. This is a very pressing problem, not only in Venezuela but in places like the United States. We have problems of on-the-job training, locating areas of unemployment, identifying structural problems. But what is really fundamental here is not just the analysis of a few projections. There is a more fundamental problem of poverty, of the unemployable. I have the uneasy feeling that short-term unemployment is actually long-term unemployment in many of these countries, and that it cannot be cured by changing the level of economic activity of the system. We need more and better demographic models linked with macromodels to really approach the problem.

Schiefelbein studied the efficiency of the educational system in combining resources to serve educational needs and to provide the best educational payoff. He had considerable data problems in blending census material with education statistics, as happens wherever this is tried. But again I am impressed here with the potentiality of the microanalytic model in working on problems of this sort. The regional data which I mentioned in connection with the discussion of the census material is also very important here. Good regional information linked with social and demographic information would be very helpful.

Browning introduced into the conference an extremely interesting proposal for collecting life histories. The novelty of his approach was the idea of tapping the respondent's stream-of-consciousness, and in the context of a sequence of life events, to elicit as much information as possible. I feel that this is a richer area than one might think at first; however, I would like to direct attention to other bodies of data which we may not normally think of in this connection, such as social security records which link individuals over their lifetimes, showing who they work for and where they live, and are of great use for tracing the individual over a longer period of time. Over shorter periods of time, we have had references to the panel history which is being obtained at the Survey Research Center in Michigan, which interviews households continuously over a five-year period, getting a great deal of information on each household. The time dimension of households is something that we are going to need if we are going to drive the microanalytic models sensibly.

CELADE [Latin American Demographic Center] also presented a paper on the use of a computer for interrelating economic and demographic variables. Again, they emphasized the importance of household and family data. The data processing involved is not something to be taken lightly—it is a tremendous amount of material; but I think the point that computing isn't enough, that we need analysis, does deserve to be emphasized.

The conclusions that I would draw from this session are that there is a great deal of data already available in the social and demographic area, and even more problems; that this is a very fruitful ground for certain types of modeling, but that the modeling does need to be done if we are to relate the data to useful analytic procedures.

The session on international comparisons of income, consumption, and prices showed how the computer was used in this major research area. The paper by Kravis and Lipsey evaluated price data for international trade flows, using international prices, wholesale prices, and export prices. Their results showed that the price elasticities differ considerably, depending on which prices are used.

Arturo Meyer pointed out that although such results are interesting, you do not dare throw wholesale prices out because they appear to be invalid. I have long wanted to throw wholesale prices out, but no one has ever let me. From a data processing point of view, this sort of comparison would not really be possible to do except by computer. It involves many very detailed comparisons and many computations.

This is also true of the work that Ronchetti and Bertaud described with respect to the European Economic Community. Here there are some six-hundred prices with three-hundred-thousand items of data being analyzed. They used a national accounts frame, and tried to develop real-product measures. I would agree with Joel Popkin in this context that there are tremendous advantages in a data bank composed of original microdata, since it provides flexibility in the development of analytic strategies.

Jorge Salazar-Carrillo's account of the difficulties of computerization rings a familiar bell. Many of us have encountered problems similar to those which Salazar-Carrillo has so aptly described. It should be noted that even Salazar-Carrillo does not suggest that there is any practical alternative to forging ahead, in spite of the difficulties encountered. It is my own belief that the process is one of learning, and that the processing and analysis of data is not doomed to perpetual frustration, in which the individual is pitted against the computer. Kenessey's point was that probably one should take a broader view of the role of the computer in the research process. Perhaps by changing the procedures of operation, increased substitution of computers for individuals would be possible.

Howe's paper really was along the same line, telling about the preparation of the data for the ECIEL [Joint Studies on Latin American Economic Integration] budget-study project. This is again a very large volume of data—19,000 households with 800 to 1,000 variables per household, covering some 21 urban centers in 11 countries. This project encountered considerable problems of standardization at the level of the original data, but it was found that there were considerable advantages in central computer processing. Again this question of editing the data, and the elimination of outliers, was raised. Betancourt suggested, I think quite reasonably, that the outliers should be retained for two reasons: first, some analyses with the outliers included may be useful; and second, the strategy of editing may later be altered so that outliers are processed differently. I think that this point is very well taken; and, incidentally, the principle is also true in combining data sets. For example, if two data sets are combined which have different definitions of income, both should be retained. This makes it possible either to separate the data sets if desired, or to develop a different strategy of combining them, or to apply different hypotheses. This underlines the necessity for having extremely flexible computer technology, with records that are easy to read and open-ended so that information can be added.

The conclusion, I think, in this session was that price research on the scale reported on here could not be done without the computer. It is true that the early Kravis-Gilbert studies accomplished a tremendous amount without the present high-speed computers, but it is clear in these studies that computational limitations prevented them from doing things that they would have liked to have done. The same is true of the early Economic Commission for Latin America (ECLA)

studies. Braithwaite in ECLA accomplished a great deal, but again the tabulations were strictly limited. I am particularly conscious of this because we computerized the basic ECLA data at Yale University, and we were able to do types of calculations that Braithwaite just could not do. He was not, for instance, able to compute Paasche and Laspeyres indexes; he was not able to develop more sophisticated weighting systems, or alternative weighting systems. The computer is really required for the large amount of computation involved in creating price indexes. This session also emphasized the importance of computer editing. The editing process has analytic implications, and these must be taken into account.

The last session dealt with international trade patterns and commodity markets. The Balassa and Schydlowsky paper was directed toward developing indicators of protection. This was an econometric analysis dealing with nominal rates, effective rates, and the effects of tax credits and government expenditures. While the results were very interesting, the authors themselves suggested the possibility of developing a wider framework of simulation models which would be concerned with this kind of analysis, as well as policy packages that could test alternative policies.

The Gomez-Sastry paper was an assessment of the UN trade projections for developing countries. Such projections are of obvious importance for both national and international policy. The basic approach here was the two-gap approach suggested by Chenery. Gomez felt that he obtained better results with a general international model than with the domestic models, which gave too high export estimates. There is in this area a very large problem of masses of data, minicountries, lack of consistency, and interdependence. I am sure that future work will go in the direction of linking various kinds of models. But I was particularly impressed at the quality of the projections that were made.

The Perez Castillo paper, on the evaluation of foreign-trade policies and economic-development simulation, really did not belong in this session; it was on simulation systems and belonged in the session devoted to macrosimulation models, but it is a far larger and more detailed paper than any that were presented in that session. Its function was to construct a general model for studying economic and social problems. It includes such things as income distribution and savings by income class. Projections were made for the year 2000 under different strategies. This is the sort of model which really gladdens the heart of a computer salesman; I am sure that the amount of time it must gobble up is horrendous. It is methodologically important because, as the authors point out, they started with an aggregated model and kept disaggregating it, and disaggregating it, and disaggregating it. It would have been useful for this paper to have been discussed in the simulation session in conjunction with the other models.

Luft's paper, also, did not really fit into this session. While he does study the effect of alternative trade policies on the Colombian economy, his, too, is a general model. It is an offshoot of a macroeconomic transport simulator model that was constructed for Colombia. What Luft did was to eliminate the transport segment and to add more policy variables. The result is a combination of a macroeconomic model and a more disaggregated simulation model. It is not probabilistic in nature, but has calibrated values. John Meyer pointed out the relationship of this model to the data base. It brought together large bodies of heterogeneous information

from a great many sources, and was, in fact, responsible for creating a consistent set of data for the Colombian economy.

This session, although it was nominally on international trade, underlined the same conclusions which were reached in the previous sessions. It pointed out that general models of the economy are needed to evaluate policy and to make projections under different alternatives. The specific objectives of any given use determine the variants of models, the kinds of variables, whether various areas are disaggregated, or how the model is expanded and in what directions. Finally, it was, I think, agreed in this session that we need to link various models because of the basic interdependence of all economic and social activities.

As a conclusion to the conference as a whole, I came away with the following feelings: that the computer has drastically changed the technology of data processing, storage, and retrieval; that the increasing availability of data and the ability to use the computer have had very significant effects on research methodology; that the variety of micro- and macromodels and the techniques of simulation which were presented here have wide applicability and are intimately related to the improvement of the data base; that there must be feedback between the research work and the data base and between the data base and the research work; that the computer is already essential in many well-established areas—in demography, social research, price comparisons, and international trade. However, there were evident during the conference other currents besides these major formal conclusions.

A declaration was circulated expressing the strong feelings of many people that the relationship between producers and users of data was very unsatisfactory, that the technology of the computer, rather than improving this, had in fact worsened it, and that radical changes were needed. The most important change proposed was the introduction into central statistical offices of research groups who would utilize the data as it was being produced, so that the feedback between the research process and the collection and processing process would be closer. Also recognized was the tremendous need for international cooperation and, perhaps, multinational institutions which could effectively provide for the transfer of technology and know-how. I would add to this that it is not just computer technology that is really needed; it is the technology of economic-research methodology and social-research methodology, because you cannot divorce data processing and analysis from one another.

There was a further undercurrent in the conference that I personally was gratified to see. That was a feeling that despite all of our labors over the years, GNP is not necessarily the goal; that optimizing with respect to monetary transactions of one sort or another, or a transaction network, or perhaps even other macro measures such as unemployment, is not perhaps the ultimate end of our models. There are a wide range of social problems and social ills that somehow do not seem to have been attacked by the massive amount of computer work that economists are turning out. I do not know quite what the answer is, but I suspect that we probably should try to attack specific social problems more directly, and to develop analytic procedures that are more relevant to these problems. This same matter came up at the meeting of the International Association for Research in Income and Wealth in Sweden this past year, and an Irishman got up and said

that as far as his country was concerned he did not think they could adequately measure welfare, but they could certainly measure important aspects of "illfare." Perhaps it is true that in our desire to take account of the systems effects and interdependence, and in our desire to get a measure of welfare, we have overlooked the problems which are not adequately reflected in the overall totals. Perhaps if we concentrate on what we can determine is wrong with societies, rather than trying to identify the optimal society, our approach might change.

In closing, I would like to reflect for a moment on the future. I feel that this conference has succeeded in opening the door to a whole wide range of problems concerned with the role of the computer in economic and social research. I personally am gratified at the methodological discussions that were held here, and I feel that this has encouraged more methodological divergence.

We are very grateful to IBM for their generous support for this conference. They gave it without any reservations whatsoever, and we have used it without reservations.

We are very happy, also, to announce at this time that IBM is continuing its support over the next five years. We would like to continue our cooperation with national and international research organizations in Latin America. We shall try to get some additional support from other sources so that we can hold workshops and conferences in various areas, and work together with other research organizations in this area on a cooperative basis over the next few years. I would expect that in the future we shall not have such a large or lengthy conference. This conference has been ideal for accomplishing its purpose, people have been very patient, the surroundings have indeed been lovely, and it has been most productive indeed, but I think that at the next stage a narrowing of the focus and perhaps smaller and shorter workshops might be in order. As to the particular areas which I feel show promise as topics for workshops, it is quite obvious that economic models and simulation techniques were some of the more exciting things in this conference. But I think we also need to give some direct attention to the development of microdata, macrodata, and statistical systems. We cannot depend purely on the discussion of models to cover data development, and, conversely, we cannot expect our discussion of data to cover the analytical problems. Finally, I also feel that we should recognize demographic and social research explicitly and, perhaps, try to concentrate on some of the "illfare" problems and see how they can be tackled, given the techniques of the computer and the availability of information.