Volume Title: New Directions in Economic Research

Volume Author/Editor: NBER

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/unkn71-6

Publication Date: 1971

Chapter Title: Papers by Staff Members on Research Priorities

Chapter Author: Various

Chapter URL: http://www.nber.org/chapters/c4174

Chapter pages in book: (p. 1 - 70)

PART 1

Papers by Staff Members on Future Research Priorities

The following papers by members of the National Bureau staff were presented at joint sessions of the American Economic Association and the American Statistical Association in Detroit in December 1970. The authors examine future priorities for the National Bureau in a number of areas in which it has carried on research in the past and would hope to continue making a useful contribution.

These papers represent the conclusions and judgments of the authors and have not been reviewed by the Directors of the National Bureau. Thus, they do not represent National Bureau studies in the usual sense, although they do represent the thinking of individuals familiar with its research objectives and strengths.

The main rationale for the application of economics to law is simply that justice is not a free good. There are competing systems of administering justice, and the implementation of any system uses scarce resources. We must decide, for example, on which activities to restrict, ways to detect and prosecute violations, types of penalties, and options available to those accused. Since the alternatives are numerous and resources are scarce, choices must be made. These points are obvious to an economist. Yet in law, particularly in the field of criminal justice, principles are expounded and policies are advocated that give little attention to the scarcity of resources. Moreover, the neglect of scarcity can be the source of much confusion when one attempts to develop policies consistent with widely held legal principles. Let me cite a few examples which also point up the usefulness of an economic approach.

1. Innocent persons should never be convicted of crimes. If this were society's only goal, the solution would be clear—convict no one. Although this has appeal to anarchists, most of us accept the view that a person should be convicted if guilt is established beyond a reasonable doubt. This generalization is not very helpful since it begs the question of defining "reasonable doubt." The wider one sets the range for reasonable doubt, the stronger the evidence necessary for conviction, and the greater the resources devoted to the judicial process. We can make this choice explicit with the following question: What quantity of resources should be allocated to reducing the probability that innocent persons are convicted of crimes? Unfortunately, not only does the response to this question tend to the absolute, but the question itself arouses objection. Such a question is labelled unprincipled or immoral because it attempts to put a price on the fundamental right of an individual not to be convicted unjustly. Nevertheless, the question is the correct one to ask. The advantage of making the choice of resources explicit is to force one to evaluate the gains and costs of alternative procedures; for example, jury versus nonjury trials or settlements versus trials.

2. "Equal justice for all" is widely accepted as a desirable goal. However, its implementation may be exceedingly costly. The quality and quantity of lawyers, advisors, etc., depend in part on one's wealth. Other things being constant, wealthy individuals would fare better than poor ones in the judicial process.[1] To reduce this disparity between rich and poor, we could subsidize legal counsel for the poor (see *Gideon v. Wainwright*, 372 U.S. 335 [1963]). Alternatively, we could limit the resource investment by the rich, although the difficulties in enforcing such a limit are likely to be very great. Even with these subsidies or restrictions, there are many factors which could lead to different outcomes or "justice" in otherwise similar situations. Presumably, better-educated persons will have a comparative advantage in understanding the complexities of the law. Persons who administer the law differ in their talents, qualifications, outlook, ethics, and other characteristics. It is well-known that, holding constant a number of background factors, there is considerable variation in penalties for individuals convicted of similar offenses [13]. Although these differences may be largely random, they still promote unequal justice. The degree of inequality can be partially remedied by better-trained judges and by wider access to appeals courts. All such remedies involve calculations about the gains and losses from scarce resources. The main decision should concern the optimal amount of unequal justice, and any movement toward more equal justice need not be desirable once the costs are taken into account.

NOTE: The author wishes to express his thanks to Hal Lary and Elisabeth Landes for many helpful comments.

[1] This proposition is not as self-evident as it may seem. Under certain conditions the relationship between wealth and the probability of conviction depends upon the form of punishment. The probability of conviction tends to fall with wealth when penalties are jail sentences and rise with wealth when penalties are fines. Empirical analysis of convictions in the U.S. Courts provides some support for this hypothesis.

3. The failure to view time as a scarce resource has led to policies designed to promote "equal justice" that may have had the opposite effect. Consider the practice of charging a zero or nominal price for the courts. This is advocated as a means of not denying the poor the use of the courts; however, its effect is often the opposite. A zero price will generally lead to a queue awaiting trial as a means of rationing supply. Thus, the costs of going to trial become the opportunity cost of waiting. These costs will be greater, and hence the use of trials lower, for defendants detained in jail prior to trial than for those released. Since the bail system tends to detain low-income defendants, a zero court price will result in more pretrial settlements, and hence fewer trials, for the latter group. In my paper on the courts [8] I show that an appropriate money price can not only reduce trial delay (which is argued to be an evil itself—"justice delayed is justice denied") but can increase the use of trials by low-income defendants. Thus, a zero price, taken together with the bail system, promotes both court delay and a distribution of trial services at variance with equal justice.[2]

Another example is the mandatory prison sentence rather than giving convicted persons the option of paying a money fine. The latter is argued to be inconsistent with the principle of equal justice because fines would permit the rich to buy their way out while the poor would go to jail. Although the rich would be more likely to pay fines and be less likely to go to prison, this need not imply unequal justice once we realize that time has an implicit price. Jail sentences for all would imply larger losses from a given sentence for persons with greater values for their time. Substituting fines may move us closer to "equal justice" in the sense of equal penalties, but this of course will depend on the rate at which fines are exchanged into sentences.[3] Moreover, if one's conception of equal justice includes compensation to victims, a system of fines can be more advantageous as it provides resources for compensation.

ARGUMENTS AGAINST APPLYING ECONOMICS TO LAW

One can accept the notion that economic principles are relevant to law and yet argue that economic analysis is not appropriate for studying such problems. Economists have generally restricted their study to economic problems involving choices in the market sector. Aside from tradition, we can give two reasons why economists have avoided, or even objected to, an economic analysis of law. Neither one is especially convincing.

The first is the belief that persons engaged in illegal activities are "out of touch" or "irrational." These individuals are not motivated by the usual pecuniary costs and gains considerations. Instead they behave compulsively, and explanations for their behavior are to be found in certain distinct sociological, psychological, or possibly physical characteristics. Criminals are viewed as a type of noncompeting group, and hence the principles that explain the behavior of persons in legal activities would not be relevant in explaining entrance into illegal activities.[4] This rationale for rejecting an economic approach is not convincing because (1) researchers have had little success in discovering distinct criminal characteristics that make criminals unresponsive to cost and gain considerations, and (2) one has difficulty in defining intrinsic differences between criminal and legitimate activities. Therefore, I see no a priori reason for ruling out an economic approach. This approach should stand or fall on its ability to predict behavior, and for this we

[2] I should point out that the real culprit here is the bail system that detains low-income defendants. If the bail system as it stands were modified to permit the release of more low-income defendants, "unequal justice" between rich and poor need not be increased by a zero trial charge. However, a zero charge would still discriminate against those not making bail.

[3] For a full discussion of the case for fines see Gary Becker [1, pp. 193-198].

[4] Packer [9] implicitly accepts this view of criminals as a noncompeting group by regarding expected penalties as a type of tariff that protects criminals from the competition of ordinary businessmen.

have to wait for a full development and testing of economic models of crime.

A second and more subtle objection relates to the nonmarket character of the law. Economists usually confine their research to the market sector where prices exist. However, in law not only are there no explicit prices for principles such as justice, but when prices do exist they are often in terms of time and not money (e.g., jail sentences and court queues). Since the main theorems of economics concern the effects of changes in relative money prices, one could argue that the absence of money prices would severely restrict the economic approach. However, as long as there is choice, prices must exist. If not money prices, then we can derive shadow prices that serve the same function as money prices in the theory of choice.

To illustrate this point, consider the problem of allocating resources to reducing the likelihood of an incorrect judicial decision (i.e., either convicting an innocent person or releasing a guilty one). Let us assume the following: (1) the only scarce resource is time in the judicial process, which is fixed in the aggregate at T_o; (2) an increased allocation of time to the i^{th} case will reduce the probability of an incorrect decision, P_i; (3) incorrect decisions are weighted by S_i, the sentence an individual would receive if convicted, so that the community's losses are symmetrical with respect to releasing a guilty person or convicting an innocent one when both are faced with the same sentence; and (4) the judiciary's decision rule is to minimize the sum of its losses across all n individuals being prosecuted. Minimizing

$$L = \sum_{i=1}^{n} P_i S_i + \lambda [T_o - \Sigma T_i] \tag{1}$$

yields the first-order conditions:

$$\frac{\partial P_i}{\partial T_i} \cdot S_i - \lambda = 0 \, (i = 1, \ldots, n), \tag{2}$$

where $\partial P_i / \partial T_i \leq 0$ and $\partial^2 P_i / \partial T_i^2 > 0$. Therefore, we allocate more resources to cases involving bigger penalties and where the probability of making an error is more likely to be reduced by inputs of time. The shadow price of time in the i^{th} case is simply λ / S_i. An increase in the latter relative to the shadow price in the j^{th} case due to a decline in S_i / S_j will lead to a shift in resources from i to j. Thus, our shadow prices have the same rationing effect as money prices.

CRIME RESEARCH AT THE NBER

In this section several research projects in law and economics are set forth. The approach taken in these studies owes a large debt to the major article on crime by Gary Becker [1]. Work on these projects has already begun, and results to date suggest that they will yield useful insights in explaining observed behavior and in formulating optimal law enforcement policy.

Isaac Ehrlich makes use of the state-of-the-world approach to uncertainty to develop a model of entry into illegal activities. These activities are assumed to have uncertain outcomes due to possible punishment, in contrast to legal activities whose outcomes are known with certainty. Individuals allocate their time between illegal and legal activities to maximize their expected utility, where increases in punishments and probabilities of conviction, other things constant, lower the return and hence supply of illegal activities. Ehrlich's model enables him to derive theorems on the determinants of specialization in "crime" as against participation in both legal and illegal activities, and on the amount of self-protection (e.g., reducing the probability of apprehension and conviction) that offenders undertake to reduce their costs in illegal activities.

The main contribution of Ehrlich's study is undoubtedly an empirical analysis of deterrence. The continuing debate over whether or not punishments and conviction probabilities deter has been conducted with little evidence presented by either side. Using data from the 1940, 1950, and 1960 Uniform Crime Reports, and employing single-equation and simultaneous-equation techniques, Ehrlich is able to measure both across states and over time the responses of specific felonies to changes in variables reflecting deterrents and gains to crime. Despite numerous deficiencies in the data, the results support the basic hypotheses of the economic model. Specific crime rates vary inversely with estimates of penalties and probabilities of conviction; deterrent effects tend to be greater for crimes against property than against persons; crimes against property vary positively with both income inequality and the level of family income. Further analysis is being done on the time trend of specific crimes in the U.S. to uncover the factors underlying the observed growth of crime since World War II. Finally, a comparative analysis of crime variations in other countries is planned.

My studies of the court system and bail system are described in detail in Part II, section 3, of this Report. A brief summary is presented here. In the study of the court system, I first develop a model to identify the variables relevant to the choice between a trial and a settlement, which include estimates of the probability of conviction by trial, the severity of the crime, the availability and productivity of the prosecutor's and defendant's resources, trial versus settlement costs, and attitudes toward risk. The model is then used to analyze the existing bail system and court delay, and to predict the effects of a variety of proposals designed to improve the bail system and reduce court delay. Finally, the model is applied, with some modifications, to civil cases. Multiple regression techniques are used to test a number of hypotheses derived from the model. Considerable evidence is found to support the hypothesis that an increase in the cost differential between a trial and settlement reduces the demand for the former and increases the demand for the latter. An empirical analysis of conviction rates is also undertaken.

In the bail study I develop a model of an optimal bail system. My approach is to derive a social benefit function that incorporates both the gains to defendants from being released on bail and the costs and gains to the rest of the community from their release. I then determine the level of resource expenditures on the bail system and the number of defendants to release that are consistent with the maximization of the social benefit. An important feature of the analysis is that it permits a consideration of alternative methods to select defendants for release or, in effect, alternative bail systems. Two systems are analyzed: one, which corresponds to most existing bail systems in the United States and other countries, where defendants must pay for their release, and the other, where defendants are compensated for their detention. Although the optimality conditions are largely unaffected by whether the defendant had to pay or was paid, there are several advantages to a system in which defendants are paid. The major advantage is eliminating the punitive aspect of the bail system, since those detained are fully compensated for their losses from detention. Other advantages include reducing discrimination against low-income defendants and providing incentives for the state to improve pretrial detention facilities.

THE EFFECTS OF LEGISLATION

Legislation outlawing or restricting different kinds of economic and social behavior has grown rapidly in this century, but we have little knowledge of the actual as distinguished from the intended effects of existing legislation. Empirical studies on the effects of legislation have been much neglected. Frequently, the existence of a law is looked on as evidence of its success. Sometimes attempts to evaluate the efficacy of a law evolve into an enumeration of its enforcement provisions. Explicit consideration of the basic question is usually missing: Does the legislation have an effect on the behavior it seeks to regulate?

The methods of economics are well-suited to answering this question. Theory enables us to make qualitative predictions on the variables likely to be affected by the law, while econometric techniques enable us to disentangle the effects of the law from other factors. The great legal scholar, Oliver Wendell Holmes, noted this when he wrote:

> For the rational study of the law the black-letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics. [6, p. 83].

There are basically two complementary approaches in evaluating legislation. Both start from the presumption that legislation affects behavior primarily through the penalties handed out for violations and the probabilities of convicting violators. Other things held constant, the greater the severity of punishments and the higher the conviction probabilities, the more effective the legislation. One caveat should be noted: "effective" is not meant to denote more desirable legislation but rather that the law will have a greater impact on behavior.

The first approach attempts to measure punishments and probabilities of conviction and relate them to gains from violation of the law. Unfortunately, it is usually impossible to fully satisfy the requirements of the direct approach because data on the number of violations are not available. Hence, estimates of the probability of conviction cannot be made. (For example, we do not know how many violations of antitrust laws or fair employment laws are committed each year.) Nevertheless, this approach is still promising since certain components are available. Statistics can usually be compiled on punishments given to convicted offenders. There may be records of persons or firms accused, and the length of time taken to decide cases. Posner [10] has compiled a history of punishments meted out to convicted violators of antitrust laws. He found that those convicted and punished were a small fraction of those apprehended (the latter are a fraction of those who actually commit violations), and the punishments were not large—at least until the expansion of triple-damage suits. These results suggest that the expected costs of violating antitrust laws are small and, therefore, that these laws have not significantly deterred collusion.

The second approach is to estimate the effect of the law on observed behavior without explicitly using measures of penalties or conviction probabilities. We first identify the variables expected to change in response to the law. Let us consider a few examples. In my study of fair employment laws [7] I showed that these laws should lead to an increase in the ratio of nonwhite to white wages and to a widening in unemployment differentials between nonwhites and whites. Fuchs and Levinson [4] used death rates from automobile accidents to calculate the effects of compulsory automobile inspection in states. Geisel *et al.* [5] used homicide rates, suicide rates, and accidental deaths from firearms to measure the effects of state and local gun control legislation. Stigler estimated the effect of the Securities Exchange Commission by analyzing the performance of new issues [12], and the effect of electric utility regulation by examining the level and structure of rates [11].

To test the law's effect we take a set of observations that are distinguished by the fact that for some observations, say states, the law exists, while for others it does not. If all states were subject to the law, distinctions would be made by other characteristics, such as the number of years since passage of the law, the extent of its provisions and coverage, penalties, and expenditures used for enforcement (which serve as proxies for differences in the probabilities of conviction). Multiple regressions of the following general form could then be estimated:

$$Y = AX + BL + u, \tag{3}$$

where Y is a vector of dependent variables, X is a vector of exogeneous variables predicted to affect Y, and L is a vector of law variables. The regression coefficient on L will then measure the law's effect. In the simplest case, L would consist of a dummy variable indicating the presence or absence of the law. If data permit, L would be expanded to include other measures suggested above. Using these techniques, I found that fair employment laws raised nonwhite–white male wages and widened unemployment differentials. These variables were not affected by differences among states in enforcement expenditures or in years that fair employment laws had been in existence. Fuchs and Levinson found that compulsory inspection of automobiles reduced death rates from auto accidents across states, and that states having more than one annual inspection tended to have a greater impact than states with only one. Geisel *et al.* found that gun control legislation tended to be more effective across states and cities in reducing suicide and accidental deaths from firearms than in reducing homicide rates.

The major difficulty with the multivariate approach is that conclusions of effectiveness (e.g., Fuchs, Geisel) tend to be based on results from a cross-section analysis of one time period. However, the possibility exists that differences observed in one cross section existed prior to the legislation. In fact, a plausible interpretation of a significant coefficient on the law variable is that differences in the dependent variable led to the passage of the legislation rather than the reverse. For example, higher nonwhite to white earnings in fair employment states were indicative of a relatively strong position for nonwhites, which in turn led to the passage of fair employment legislation. States with lower automobile death rates were more safety-minded to begin with, and this led to the passage of compulsory inspection. In states with fewer guns, and hence fewer deaths from guns, there was less opposition to the passage of gun control legislation. One way to overcome this difficulty is to compute a regression similar to (3) for a time period before the passage of legislation in any of the states, using the law variable from the original cross section. Then a nonsignificant coefficient on the law variable in the earlier time period would indicate that, prior to legislation, there were no significant differences between states that enacted laws and states that did not. Clearly, such a finding would buttress the results from a single cross section. This is precisely what I did in my paper on fair employment laws. The problem is largely avoided when time-series rather than cross-section data are employed. Stigler has relied on time-series data in his research cited above. Concentration on cross-section data has been due to the lack of data for earlier years. However, a thorough study requires some time-series analysis even from fragmentary data. The acid test of effectiveness is to show that the law changed observed behavior and this will only show up in an analysis over time.

Research on the effects of legislation has been done on a limited scale by economists. It remains a fruitful area for further research and one that should contribute to public policy issues. At the National Bureau we have in mind several types of legislation for analysis. They include antitrust, accident liability, minimum schooling, and divorce legislation.

CONCLUDING REMARKS

We have advocated empirical analysis of the consequences of legislation without a theory of legislation. It would be preferable to treat legislation as an endogenous variable and simultaneously estimate the effects of legislation and the forces leading to passage. This is especially desirable in view of the difficulty cited above in discerning whether the effect we attribute to the legislation was not instead a factor that gave rise to the legislation. However, there is at present no widely accepted positive theory of legislation. What economists have put forth is a body of theory on divergencies between private and social cost which has been used to justify various types of government intervention and legislation. We lack a theory that would offer predictions as to the types of behavior to be restricted by legislation, the extent of restrictions, the amount of resources allocated to enforcement and the penalties for violation. Thus the difficulty of treating legislation as an endogenous variable is apparent. We do not know the structural equation that has the law as its dependent variable.[5]

In summary, I have tried to establish the contribution that economic analysis can make to two specific areas of law: crime and the effects of legislation. In both areas, administrators and legislators are continually making decisions with little empirical evidence to guide their choices. We hope that economic research can begin to provide the evidence needed for careful policy formulation and evaluation.

REFERENCES

1. Becker, G. S., "Crime and Punishment: An Economic Approach." *Journal of Political Economy,* March/April 1968.
2. Block, W., "An Economic Analysis of Rent Control," mimeographed, 1970.
3. Ehrlich, I., "Participation in Illegitimate Activities," mimeographed, 1970.
4. Fuchs, V. R., and I. Leveson, "Motor Accident Mortality and Compulsory Inspection of Vehicles." *Journal of the American Medical Association,* August 28, 1967.
5. Geisel, M. S., R. Roll, and R. S. Wettrick, Jr., "The Effectiveness of State and Local Regulation of Handguns: A Statistical Analysis." *Duke Law Journal,* No. 4, 1969.
6. Holmes, O. W., "The Path of Law," reprinted in *The Mind and Faith of Justice Holmes,* edited by M. Lerner, Little, Brown & Co., 1943.
7. Landes, W. M., "The Economics of Fair Employment Laws." *Journal of Political Economy,* July/August 1968.
8. ———, "An Economic Analysis of the Courts." *Journal of Law & Economics,* Spring 1971.
9. Packer, H. S., "The Crime Tariff." *The American Scholar,* Autumn 1964.
10. Posner, R., "A Statistical Study of Anti-Trust Enforcement." *Journal of Law & Economics,* October 1970.
11. Stigler, G. J., "What Can Regulators Regulate? The Case of Electricity." *Journal of Law & Economics,* October 1962.
12. ———, "The Economic Effects of the Antitrust Laws." *Journal of Law & Economics,* October 1966.
13. Zeisel, H., "Methodological Problems in Studies of Sentencing." *Law and Society Review,* May 1969.

[5] Without a theory, interesting work can nevertheless be done on estimating forces related to the passage of laws. For example, Block [2] in a thesis being completed at Columbia on the effects of rent control has found that cities decontrolling earlier (after World War II) were those with lower population densities and fewer tenants relative to home owners. The former suggests a more elastic supply response for new housing and the latter may be related to the political appeal of rent control. In my own work on fair employment laws I touched briefly on the forces leading to passage. One important negative finding was that the passage of these laws in states was not related to either the level or rate of change in the proportion of nonwhites.

THE NBER APPROACH TO HUMAN RESOURCE PROBLEMS—Finis R. Welch

In discussing future directions for Bureau research vis-à-vis problems of human resources, I have neither a crystal ball nor an explicit "theory" of markets for research. So in the best traditions, I will measure and extrapolate without the internal consistency constraints of either a crystal ball or a theory. The objective is limited: a discussion of recent Bureau-sponsored research in the area of human resources and suggestions of areas that I see as most promising for future research.

The Bureau has been at the center of developments in the "human capital approach" largely through its association with Gary Becker and Jacob Mincer, and much of the discussion is directed to this approach. Certainly, there is much more to human resources than capital, but the human capital approach is and has been an important tool for analyses of labor productivity. Schultz [32] draws a highly relevant distinction between human resources and human capital.

> Thus my interpretation of the term "human resources" is that it encompasses all of the many attributes of a people—physical, biological, psychological, and cultural—that account both for the social values that determine preferences and the economic value of the producer and consumer services that a people render whether they come to them as earnings or as personal satisfactions directly. (p. 11)

> Human capital is strictly an economic concept. Although it pertains to particular attributes of man, it is not intended to serve those who are engaged in analyzing psychological, social or cultural behavior. It is a form of *capital* because it is the source of future earnings, or future satisfactions or both of them. It is *human* because it is an integral part of man. (p. 6)

While there are other forms of human capital, health and educational capital have received the most attention, the original emphasis falling largely upon education. No one was surprised when those first interested in human capital asserted that education is an investment that conditions labor earnings. But when Schultz [30] and Dennison [10] suggested that between 20 and 40 per cent of U.S. economic growth this century may be attributable to increased per capita schooling; when Mincer [25, 26, 28] suggested that a larger fraction of personal income inequality is explained by differences in individual schooling and investments in on-the-job training than by differences in the ownership of physical capital; and when Becker [2] and Hansen [18] calculated the very considerable internal rates of return to schooling, the economic world took note. The arrival of human capital as a fully legitimate area of research was signalled by the Exploratory Conference on Capital Investment in Human Beings, in December 1961, which was sponsored by the Universities-National Bureau Committee for Economic Research [31]. At the conference, Becker presented his paper "Investment in Human Capital: A Theoretical Analysis," which remains as the cornerstone of the theoretical foundation for the human capital approach. Mincer presented "On-the-Job-Training: Cost, Returns, and Some Implications," which led us to think of postschool age–income profiles as something other than a purely physical maturation and aging process. Stigler presented "Information in the Labor Market," which taught us that it can be smart to be ignorant. Also at this conference, Sjaastad and Mushkin noted that migration and health are legitimate aspects of human capital and Weisbrod suggested that while sunk costs are sunk, there is real value in not sinking your costs too soon (i.e., the concept of the "option" value of staying in school).

NOTE: I owe much to my reading of T. W. Schultz', "Human Capital: Policy Issues and Research Opportunities," a paper prepared for a series of NBER colloquia in honor of the fiftieth anniversary year. His assignment was to discuss many of the same kinds of things discussed here, and I have drawn freely from his comments.

It is noteworthy that the human capital approach headed Harry Johnson's [22] list of four new approaches having high potential in analyzing social problems, which also included "time as a fundamental unit of cost," a direct descendant of the human capital approach. The list also includes Stigler's recognition that information may not be a free good, which as I have noted is itself a sibling of the human capital approach.

Since the original Bureau conference, the human capital literature has flourished but it is now clear that the main theme had been set. Much of the Bureau research now in progress is a direct descendant. In the following sections, I consider a few areas that have absorbed, and presumably will continue to absorb, much of the Bureau's "human resources" research effort. The concluding section contains a discussion of omitted areas that I consider worthwhile.

EDUCATION AND EARNINGS

It is likely that income–education relationships have absorbed a larger fraction of research resources devoted to human capital than all other aspects of the problem. Originally, we were fascinated with calculating and refining estimates of internal rates of return. That investments in man could be "more productive" than investments in machines was an intriguing idea. But these days few resources are devoted to estimating rates of return. The ballpark estimates for the U.S. came quickly and easily and improvements await better data. We did, I think, learn something from these estimates. First, ignoring all aspects of schooling other than sheer income-generating capacity, U.S. investments in schooling have been profitable. The highest returns may have accompanied a push toward universal elementary education, but the returns to high school and college appear to have been competitive.[1] We have also learned that "schooling" is a reasonably well-behaved factor in the sense that its monetary returns are affected by factor ratios, and that because of this the profitability of educational investments is subject to change as skill distributions and factor endowments change.[2]

But research interest these days appears to have shifted to more fundamental questions that hopefully will not only improve our estimates of rates of return but will be valuable for other purposes. At the micro level, the question of economic determinants of the "shape" of the age–income profile continues to absorb a good deal of attention as does the closely related question of the nature of the ability–schooling–age interrelations. At the macro level, we are focusing more on the determinants of income distributions and the determinants of the "payoff" to schooling.

Building upon the Mincer [27] analysis as refined by Ben-Porath [6 and 7], Sherwin Rosen [29] is working on optimal life-cycle income profiles with the objective of estimating the depreciation of human capital over the life cycle.[3] In these models, on-the-job training is the vehicle for transforming current income foregone into future earnings. The schooling period is viewed as a corner solution in which people "specialize" in learning, and this period is characteristically followed by one in which individuals simultaneously learn and earn. Over the course of this second period the learning–earning mix shifts steadily in favor of earning, as a result of the shortening future working span over which investments in learning are recouped. To avoid the corner solution of specializing in either earning or learning it is necessary (at a point in time) that there be increasing marginal difficulty in transforming present earnings foregone into discounted

[1] Schultz [32] offers an extensive summary of the estimates available for the U.S.

[2] Griliches [16] provides real insight into this problem in his discussion of education in production functions. Relations between relative wages and rates of return are explored by Welch [36], who provides cross-sectional evidence for the U.S. Bowles [8] adds an international cross section.

[3] For a different approach to estimating these depreciation rates, see T. Johnson [23].

future earnings. In the Ben-Porath model this is accomplished by intraperiod increasing marginal investment cost, pure and simple. But, in the Rosen model, learning is assumed to be a joint product of an occupation such that the learning rate is specified by the occupation and cannot be varied. Jobs vary in learning (i.e., future earning) intensity, and workers bid for jobs by agreeing to accept lower wages for more learning-intensive occupations. The "price" of learning is imputed from wage differentials and the necessary condition for an interior solution is that the implicit price of the learning associated with jobs must rise more than proportionately with learning intensity. That is, the unit cost of learning must be an increasing function of the rate at which learning occurs, as is true of the Ben-Porath model. The difference is the linkage to occupation, for, if the world is as Rosen hypothesizes, then his model implies an optimal pattern of occupational mobility over a person's career.

The Rosen model can be criticized on the grounds that the occupations he envisions probably correspond only slightly to occupations as they are defined in the available data. But this problem is not insurmountable. Instead of defining an occupation by a learning rate that is constant, let the rate of learning be, say, a declining function of experience in the occupation. This is not the important feature of the Rosen model. It is rather that the effects of differences in initial conditions, e.g., the availability of funds for investing in on-the-job training or the quality of schools attended, can be traced throughout a person's career. In this society, initial conditions are perhaps more affected by a person's progenitors than by himself (that is why they are called initial), so that the optimal life-cycle models of Ben-Porath and Rosen can be tailored to develop models of intergenerational mobility.

This leads to my first suggestions for future research. In my opinion economists have too long ignored the intergenerational aspects of poverty and the sociologists who have addressed the problem have worked only with descriptive intergenerational mobility matrices, leaving no room for individual choice. If reducing poverty is a legitimate social objective, then it seems that it is even more legitimate to reduce inherited poverty. It is sad that we continue to be as ignorant as we are of the factors affecting motivation and alienation. Theoretical tools for addressing these problems are, or will soon be, available and the data base is not empty.

The Bureau is now sponsoring a number of projects that attempt to refine the nature of income–schooling relationships by examining interrelations with measured ability. Taubman and Wales [35] have just completed a manuscript that explores several aspects of the problem. Among their more interesting observations they conclude that although the proportion of the student population entering college has continued to rise, their average ability levels have not fallen. Instead, colleges have admitted higher fractions of more able students from lower income families and they have become more adept at "screening" (i.e., identifying, or at least admitting) more able students. Correspondingly, the average ability of students completing high school but not going on to college has fallen. This raises an interesting possibility. The presence of "scalloped" age-adjusted income–schooling profiles has interested many persons who worked with the 1960 Census data. Briefly, the observation is that persons who attended, but did not complete, either high school or college earned less than would have been predicted by interpolation of income levels between elementary and high school completion and between high school and college completion. This effect has been alternatively dubbed the "flunk-out" or "completion norm" effect. Presumably there was positive informational content to the observation that a person had completed what he had begun. But, if the average ability of the high school graduate who does not attend college continues to fall, the "diploma" effect may become negative.

Other conclusions developed by Taubman and Wales from the data they examined are that, while omitting ability from estimates of the return to schooling does result in a systematic overestimate of the return to schooling (via positive ability-schooling correlations), the magnitude of the effect is small and that omitting ability has probably resulted in only trivial overestimates of rates of return to schooling. This, of course, agrees with the conjecture of Zvi Griliches [16]. Another tentative conclusion is that, in adjusting for education and ability, if the optimal occupational distribution as defined by income is compared with the existing distribution, then the U.S. is much more efficient in distributing its college graduates than its less-schooled citizens. Taubman and Wales attribute this result to occupational screening, but there is another possibility that could be investigated. Except for college graduates, the income associated with a given amount of schooling varies greatly with location. Persons with less schooling earn less in the South than in other regions. This suggests that there may be high returns to migration for the less schooled and the "screening" may be more geographic than discriminatory in its origin.

If the human capital approach has taught us anything, it is that income comparisons should be made in terms of something like net discounted values of lifetime earnings instead of comparing incomes at a particular point on the profile. Along these lines, John Hause is exploring two sets of data in which interrelations between income, schooling, and ability can be examined.[4] The analysis is still preliminary so the results can only imperfectly be foreseen, but it appears that in at least one body of data, ability effects may be more economically significant than the Griliches or Taubman-Wales findings indicate. Hause has a collection of Swedish data giving I.Q. scores at two points in a person's life and annual earnings data for very considerable periods of the working lifespan. Although the analysis is in its infancy, these data should be instrumental in providing information both about the ability question and life-cycle productivity. In addition, the NBER-Thorndike sample is now ready for analysis and is potentially very rich. For a fairly select population it provides earnings at a few points in a person's career together with schooling data and scores on a variety of tests administered in 1943 as the sample population entered the U.S. Army Air Force. Thus, Bureau research into the nature of the income–schooling relationship and the effects of ability in the near future is easily predicted. Analysis of the available data will continue.

Two projected analyses of the NBER-Thorndike sample are especially interesting. In one, Lewis Solmon and Paul Wachtel will merge information about university attendance together with some characteristics of state school systems for elementary and secondary schools in an attempt to identify some aspects of schooling quality vis-à-vis future earning capacity. This effort closely parallels the work by Johnson and Stafford [21] in respect to the precollege analysis, and a comparison of the results will be illuminating.[5] With regard to the analysis of quality of college, it is hard to think of a more perfectly designed sample, and the results should be of fundamental importance.

That college builds upon a very considerable educational history of the student implies that it should at any empirical level be extremely difficult to "partial out" its effects. But, the battery of scores for tests administered between high school graduation and college enrollment offers an excellent opportunity to isolate the college effects. Further, since the actual university attended is specified, schooling data should be more accurately matched than is ordinarily possible.

Tom Juster, in a continuing analysis of the contributions of measured ability to earnings, plans to focus upon interactions between occupation and implicit ability prices. This is very important. First, it can address the question of the nature of the measures we dub ability. If, for example, we measure something and call it cognitive ability and then in regression analysis find that its implicit value is the same in highly routinized manual occupations as in occupations that permit more

[4] Some of the early results are reported in [20].

[5] The main difference is that Johnson-Stafford have data for a point in time for people of varying age, while the NBER-Thorndike sample refers to people of about the same age but earnings are for several points in the career.

discretion in day-to-day functions, I for one would be very suspicious about cognitive ability. Further, if measures of physical dexterity have value that is independent of occupation, then I doubt that we are measuring what we think we are. (There is room for optimism here, for Juster assures me privately that implicit returns to digital dexterity is greater for doctors than lawyers.)

If these implicit tests of the tests pay off such that the instruments warrant a degree of confidence, then an interesting extension, similar to that of Taubman-Wales, is possible. Given an individual's ability vector, there exists a lifetime occupational mix that maximizes his income. How closely do the optimal occupational distributions agree with observed distributions and to what is the disagreement attributable?

With respect to the empirical effects of ability, I must admit a degree of skepticism. I am afraid that we simply understand too little about the nature of skills that are economically productive. The available measures of ability are often designed not to answer questions about economic productivity, but about an ability to succeed in school, which is not necessarily the same thing. I trust that future research will tackle the thorny questions of what ability and schooling mean. Presumably, it is skills that determine productivity, both economic and otherwise. Some of these skills are produced in school and others are home produced or innate. By using simple summary measures like I.Q. we may conclude that analyses of returns to schooling that omit ability give results that are similar to those that adjust for ability. From this it might be argued that little is gained by trying to assess the effects of ability. But this conclusion may say more about the value of current measures than about the nature of ability per se. Obviously, it is important to determine whether, given current measures, the returns to schooling are sensitive to ability adjustments. Current approaches will be informative on this count. But the larger question remains: What attributes of man affect his behavior and economic productivity and where and how are these attributes produced? Schooling remains a "black box" that appears to increase earning capacity. Why this is true (i.e., exactly what is produced and how this is conditioned by student characteristics determined outside the schools) remains unknown. And, of more concern to me, we seem not to even be asking the relevant questions.[6]

Herbert Gintis [14] has observed that in the available empirical studies that correct for ability a common result emerges: There is little or no downward adjustment in the return to schooling when ability is held constant. In some cases measured ability is in fact a measure of cognitive achievement *after* the student is no longer attending school. But, if the school produces cognitive abilities and it is these abilities that lend value to schooling, then why is the residual (achievement constant) value of schooling so large? Gintis argues that (1) schools are not primarily interested in producing cognitive achievement and (2) the productive value of school attendance is owing largely to the "socializing" influence. Unfortunately, socialization is not measured so that the hypothesis cannot be tested directly.

Much of the Bureau's research effort concerning the income distributional effects of human capital is summarized in a forthcoming book by Becker, Chiswick, and Mincer [5]. Their accomplishments are several and I will indicate only a few. First, Becker's Woytinsky lecture [4] is included, in which he cogently distinguishes between the effects of opportunity and ability and, in passing, provides perhaps the best available operational definitions of these two complex concepts. Opportunity is nothing more than the supply cost of funds for investing in human capital. Ability is the *ceteris paribus* rate of return to schooling. Thus, ability is comparative advantage in school attendance defined by marginal efficiencies of

[6] There are, of course, exceptions. How other characteristics affect school performance is addressed by those pursuing "educational production functions." See especially the work of Hanushek [19] on this score.

investment. Mincer's original contributions of the human capital approach for analyzing income distributions has been mentioned as being instrumental to our thinking, and the contributions of the latest essay are considerable. Aside from aptly demonstrating the sheer magnitude of the distribution of human capital in explaining the income distribution, it brings an empirical pragmatism that is valuable in helping us sort through the ever increasing stock of unexplored data. In particular, the concept of the "overtaking age" provides a convenient way of quickly deducing rates of return and of comparing life-cycle earnings. Chiswick extends the analysis, showing the value of this approach in explaining differences interregionally and internationally in both the dispersion and skewness of income distributions.

This approach to income distribution is concentrated upon what fraction of measured income inequality is associated with differences in human capital. Indeed, Mincer argues that perhaps two-thirds to three-fourths of the inequality is explained by the human capital approach. Note that in most cases the return to schooling is overestimated because of a positive covariance with ability and quality of schooling.[7] Since on-the-job training and schooling are positively related, the ability and background effects are magnified in the full life-cycle comparison. This, then, shows the necessity of relating the Mincer model to models like that of Becker [4], which attempt to explain why differences in schooling exist in the first place. If capital markets were perfect, if everyone maximized the present value of lifetime earnings, and if everyone had the same ability and opportunity, then the present worth of lifetime earnings would be independent of investments in schooling and on-the-job training. In this limiting case, the human capital approach would be instrumental to explaining the inequality in the distribution of income at a point in time, but only luck would explain differences over a lifetime. The Becker-Chiswick-Mincer approach to understanding income distributions spotlights the need to "break into the box" and try to allocate the results of the human capital between opportunity, ability, and luck.

Other Bureau research into the nature of education–income relationships includes my own work, which currently is a rather rudimentary analysis of what makes schooled people more productive. I hope that there is a place in the Bureau's research future for this type of analysis. We now have some evidence that the value of schooling is conditioned not only by the content of the schools attended and the characteristics of the student's background, but is affected by the environment in which schooling is employed. The value of schooling is related to the skill distribution, to the availability of other productive factors, to the comparative advantage of "on-site" training as opposed to the relative isolation of the formal school, and is related to the pace at which first-order empirical rules of thumb lose value relative to more basic rules, which in turn is related to the complexity and pace of change in the economic environment. An understanding of these factors is especially important for projecting the effects of schooling at a macro level and it therefore is important in the organization of an economy for growth.

HEALTH ECONOMICS

The major Bureau thrust in the economics of health began three or four years ago, signaled by Victor Fuchs' paper "Some Economic Aspects of Mortality." Following the Fuchs distinction between health and medical care, Auster, Leveson, and Sarachek [1] estimated a model in which medical care enters a health production function. This distinction is also a prime feature in the work of Grossman [17], whose theoretical model has health entering the utility function in two ways.

[7] A failure to correct for family background suggests that persons who are more schooled may also have attended better schools.

First, there is a direct effect of healthier days being happier days. Second, there is an indirect effect of health on earning capacity. Medical care is seen as one input into the production of health so that the demand for medical care is derived from the demand for health. His empirical results suggest that this frame of reference can be useful. Several related projects build upon this distinction of health and medical care. Fuchs and Grossman [11] have a paper on correlations between health and schooling, stressing simultaneous determination and implications for efficiency using the household production function concept. Also, in a project just initiated by Grossman, relations between health and education will be explored using the NBER-Thorndike sample.

Other recent projects include the work of Ro, showing that hospitals are sensitive to economic factors in utilizing their facilities. Silver [33] is involved in analyzing differential mortality rates between race and sex, and Kramer and Fuchs [12] are analyzing determinants of expenditures on physicians.

In health care, as throughout the service sector, output is especially hard to measure independently of price. This severely limits analyses of efficiency in resource allocation and analyses of the complex of factors determining productivity. In an approach that may have very general adaptability, Fuchs and Hughes are examining the distribution of surgical manpower among various tasks, where tasks are measured in common denominators of, say, hernia or appendectomy units. It is uncertain whether this kind of approach will pan out, for there are problems analytically similar to that of measuring the quantity of a product as being proportional to the quantity of one particular input. This assumption is not so bad when production is subject to fixed proportions or when relative input prices are constant. Their approach is a start and we must await results.

EDUCATION AND NONMARKET ACTIVITIES

As Becker presented his theoretical base for aspects of the human capital approach he relied heavily upon the opportunity cost of a student's time. Indeed, Schultz in originally accounting for growth suggested that the major share of investment in schooling was not in direct outlays but in the opportunity cost of not being at work. In Mincer's measures of investments in on-the-job training, *all* investments are foregone wages. Time is an input of major importance in building human capital. But, if it is important to recognize the value of time as an input into learning, is it not equally important in all aspects of consumer decisions? In his seminal paper, "A Theory of the Allocation of Time," Becker [3] developed these ideas. As work takes time, so does consumption. Things have both a direct cost and a time cost. One immediate result of this frame of reference is the concept of the household production function in which market goods, consumer time, and other family characteristics combine to produce more fundamental arguments for utility functions. A possible corollary is that since education appears to make *workers* more efficient users of time because they produce more per unit of time, might it not also make *consumers* more efficient users of time?

This led Bob Michael to a rather remarkable approach to viewing the determination of the composition of a family's consumption bundle. In his forthcoming book [24], he demonstrates that when the head of a family has more education, the composition is *as though* the family has more income than it does in fact. Its consumption bundle is skewed in favor of more income-elastic goods. Thus we have learned something. Not only does family education affect the composition of consumption but it systematically shifts consumption in favor of luxury goods. This is consistent with Becker's contention of greater household efficiency, i.e., more income per dollar.

In a somewhat related piece of research Lewis Solmon [34] concentrated upon identifying relations between life-cycle saving behavior and education. He does find a systemic relation: Persons with more education save a disproportionate share of income, the discrepancy being largest for younger persons and declining over the working lifespan. The data are not adequate to determine whether this effect stems more from the fact that persons with more education have higher incomes so that the contributions of social security and medicare are proportionately less important in retirement, or whether increased education in fact signals larger bequests.

Another aspect of the Becker theory of the allocation of time is that when wages are high, the cost of market- as opposed to home-produced goods is low, and when wages are rising, the consumption cost for time-intensive goods is also rising. Pursuing this logic, Gilbert Ghez [13] analyzed the life-cycle determinants of consumption in a model that, holding wealth constant, gives a lifespan profile to consumption. His empirical evidence reinforces the idea that consumers are aware of the value of their time.

Other current Bureau research that draws heavily upon the concept of the household production function and an interplay between education and the value of time includes the analyses of fertility and the use of contraceptives by Michael, Ross, Sanderson, and Willis, which is reported in a separate paper, and the work of Michael Grossman, who is focusing upon the economics of health.

FUTURE DIRECTIONS

In case all of my prejudices are not revealed in the earlier discussion, I will add some comments on research possibilities. These suggestions are not intended to be Bureau-specific, for a particular organization should be more specialized in its effort. For example, Bureau-sponsored research has focused more narrowly on the human capital approach than is necessarily optimal from a broader social view of the wide range of problems related to all aspects of human resources. That is, the proximity to people like Gary Becker and Jacob Mincer has given the Bureau a comparative advantage in using the "language" of human capital.

To my own thinking, education-oriented research has too long ignored the impact of the schooling process upon the formation of values. Given tastes, we have been too "classical" in our attempts to maximize. Clearly, a schooling process that spans the major formative period of a student's life will affect his values. The work of Gintis is beginning to give us some insight on these points. His approach, which addresses the question of the reward (grading?) system within schools, can reveal something about the kind of student responses schools encourage. Certainly, the growing emphasis on neighborhood schools and tuition vouchers, as examples, indicates that the public is not completely satisfied by the way objectives are mirrored in larger political decision units.

With respect to schooling, there is an obvious area for further research: schooling production functions. It is a social crime that we remain as ignorant as we are of the technology of learning. The type of approach to which Coleman [9] attracted so much attention appears to have real potential. By ordinary regressions of measures of school output (test scores) on school and family characteristics, systematic relations appear discernible. In fact, the results of Hanushek [19], using observations of individual students through time, are very promising.

But, if this nut is to be cracked, we must have extensive data, observing students over their schooling cycle. Questions of which teacher characteristics affect learning propensities and, in turn, how these characteristics are produced are obviously important. Differences in the duration of impact for experiences at various points in the learning career should be known in order to time learning experiences efficiently. Perhaps students should be exposed to the "best" teachers as early in their learning career as possible, because early experiences are longer lived and a better prepared student is a better learner. Alternatively, students may not be capable of capitalizing upon excellent teachers until they are more mature, so that it would be efficient to delay exposure to the "better" teacher. And, as a third possibility, the skills required for teaching young persons are distinct from those required for instruction of the more mature, so that it is really meaningless to speak of optimal timing of "good" teachers. We can speculate so freely because we know so little. Yet it is important that we learn. Coleman's conclusion that characteristics of the home environment are very important in determining the rate at which a student learns, a conclusion that to my knowledge is reinforced by all subsequent related research, raises serious questions regarding equity. At the least, we should attempt to determine which attributes of the home are important and how these attributes condition the productivity of schools.

It is precisely in areas like education, which are largely publicly funded and in which the reflection of consumer desires are therefore more distorted, that economic analysis can have its greatest value. And it is problems of this type that demand attention. The cost of learning about educational technology will be high because the necessary microdata will be expensive, but the returns might also be high. It is possible that because this sort of analysis requires an extensive data base, collected over considerable periods during individuals' learning careers, research organizations like the Bureau have an advantage because they can guarantee more continuity of interest than can the individual working at a university. Thus, I have a related suggestion, that the major research organizations become data banks for the kind of time-series microdata that will obviously be so important for many aspects of social research.

In an earlier section, I suggested that more research be directed toward learning about which teaching skills have value and whether these skills can be produced. Most analyses of the economic value of education show that schooling increases earnings where schooling is measured only by approximate years of attendance. For many problems, this begs the question because the sword, school years, is too dull. For example, in any of the problems associated with educational technology or "quality" of schooling, the question arises: Could the productive attributes acquired in, say, the ninth year of school have been acquired prior to that time? Unless we can find a reasonably comprehensive way of identifying the essential characteristics of the ninth year, this sort of comparison is not possible.

If areas of concentration were measured by the cost of resources devoted to researching them, I would guess that the question of life-cycle profiles of income, skill assimilation, and occupational mobility will be among the two or three most important in the next decade. Mincer has provided the benchmark, but many questions have awaited the kind of data now emerging. With respect to human capital, there is an obviously important question of how ability interacts with lifetime earnings. Is "success" cumulative in the sense that those who have high incomes early in their working career build on this advantage and broaden the gap? Does Mincer's investment view dominate, so that those who accept initially low incomes eventually overtake those who opt for high incomes early in the career? Are current measures of ability good enough to permit distinctions between returns to ability and investments in on-the-job training? To answer these questions, we must turn to life-cycle income histories. The cross-sectional data are not adequate. With these same life-cycle data the question of intergenerational linkages can be addressed. How economically mobile is our society and is mobility increasing or decreasing through time?

This leads to a host of questions related to the economics of poverty. In work now in progress at the Bureau, David Gordon [15] analyzes factors affecting earnings for ghetto residents. His conjecture is that a secondary labor force has emerged in which members of this work force expect little upward mobility and may be becoming disillusioned, so that their allegiance to the traditional working career is weakening. He points out that traditionally those at the lower rungs of the economic ladder have been immigrants from abroad or blacks who migrated from the rural south to the urban ghetto and who expected, often for good reason, their inferior economic status to be transitory. Now, many of those with the lowest incomes were born to this status and have little reason to expect upward mobility. To date, empirical evidence vis-à-vis responses to traditional incentives is sparse, but it is clear that this type of information will be fundamental in formulating social policy.

The 1960 Census carried the information that the returns to schooling for blacks were much less than those for whites. Since then, the relative income position of nonwhite to white families has improved, but we have little information about what portion of this change reflected "tightening" in the labor market as opposed to secular growth. We do not know why returns to schooling were so much less for blacks or whether they still are. If it was the effects of inferior segregated schooling, we cannot be certain of the effects of compensation in terms of similar expenditures alongside of de facto segregation. Perhaps the 1970 Census will signal rapid improvement, but if not, the question of mobility alternatives available to those who are economically depressed will be even more pressing.

There are numerous researchable questions related to the distribution of opportunities for mobility when opportunities are defined not so much as supply costs but in terms of responses or economic returns to given efforts or investments. In addition to the distribution of opportunities among various socioeconomic groups there is the question of the level of this distribution. In regard to the returns to schooling, what factors determine the population mean? Obviously the organization of an economy for growth should be related to the profitability of alternative investments. But what determines the profitability of educational investments? There is recent evidence pertaining to elasticities of substitution between laborers classified by school-related skill levels. But, there is only scant information pertaining to substitution relations between labor-skill classes and various forms of capital.

It is said that the division of labor is related to the extent of the market. Let me offer a corollary: The *complexity* of a market as measured by the degree of product and input differentiation is also limited by the extent of the market. That is, there are informational scale diseconomies in market "size." *Ceteris paribus*, the larger and the more developed the market *and* the more rapidly the market is developing, the greater is the information necessary to function efficiently in the ordinary business of life. If this is true, and if schooling is in part a ground for developing ways of efficiently processing information, then the return to schooling is derived from these characteristics of the market. The question of factors affecting the payoff to schooling is central to analyses of economic development. In my opinion, we have disproportionately concentrated on ways in which education "pays" and how this payoff can be measured, as though payoff schedules are not themselves endogenous to an economic system.

There are obvious omissions to my suggestions for further research, but much of what we will learn can, I think, be construed as falling in one of the classes of problems discussed above. Yet, any discussion of future research directions would be incomplete if it did not recognize a strong drift in social research toward policy-specific analyses. For example, at least three Bureau-sponsored projects now in progress or recently completed address the question of the effects of a minimum wage. Also, a recent empirical analysis addressed the question of income and substitution effects in labor supply functions for the working poor, which is designed to provide parameter estimates that in turn will permit estimates of the so-called disincentive effects of negative income tax schemes. At the present time several staff members are involved in a project to assess possible future avenues for research into aspects of poverty labor markets. Presumably, the Bureau's future directions for research into poverty will be affected by this analysis.

REFERENCES

1. Auster, Richard, Irving Leveson, and Deborah Saracheck, "The Production of Health: An Exploratory Study." *Journal of Human Resources*, Fall 1969.

2. Becker, Gary S., "Underinvestment in College Education?" *American Economic Review,* May 1960.
3. ———, "A Theory of the Allocation of Time." *Economic Journal,* September 1965.
4. ———, *Human Capital and the Personal Distribution of Income: An Analytical Approach.* Woytinsky Lecture No. 1, Institute of Public Administration, Ann Arbor, University of Michigan Press, 1967.
5. ———, Barry R. Chiswick, and Jacob Mincer, "Three Essays on the Effects of Human Capital on the Personal Distribution of Income." (Forthcoming 1971.)
6. Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy,* August 1967.
7. ———, "The Production of Human Capital Over Time." In *Education, Income, and Human Capital,* edited by W. Lee Hansen, New York, NBER, 1970.
8. Bowles, Samuel, "Aggregation of Labor Inputs in the Economics of Growth and Planning: Experiments with a Two-Level CES Function." *Journal of Political Economy,* February 1970.
9. Coleman, James S., *et al., Equality of Educational Opportunity.* U.S. Office of Education, 1966.
10. Denison, Edward F., *The Sources of Growth in the United States and the Alternatives Before Us.* Committee for Economic Development, 1962.
11. Fuchs, Victor R., and Michael Grossman, "The Correlation Between Health and Schooling." Unpublished manuscript, NBER, September 1970.
12. ———, and Marcia J. Kramer, "Expenditures for Physicians' Services in the United States, 1948-68." Unpublished manuscript, NBER, January 1971.
13. Ghez, Gilbert, "A Theory of Life Cycle Consumption." Unpublished Ph.D. dissertation, Columbia University, 1970.
14. Gintis, Herbert, "Education, Technology, and the Characteristics of Worker Productivity." *American Economic Review,* May 1971.
15. Gordon, David M., "Economic Theories of Poverty and Underemployment." Unpublished manuscript, NBER, January 1971.
16. Griliches, Zvi, "Notes on the Role of Education in Production Functions and Growth Accounting." In *Education, Income, and Human Capital,* edited by W. Lee Hansen, New York, NBER, 1970.
17. Grossman, Michael, "The Demand for Health: A Theoretical and Empirical Investigation." Unpublished Ph.D. dissertation, Columbia University, 1970.
18. Hansen, W. Lee, "Total and Private Rates of Return to Investment in Schooling." *Journal of Political Economy,* April 1963.
19. Hanushek, Eric, "Teacher Characteristics and Gains in Student Achievement Estimation Using Micro-data." *American Economic Review,* May 1971.
20. Hause, John C., "Ability and Schooling as Determinants of Lifetime Earnings: or, If You're So Smart Why Aren't You Rich?" *American Economic Review,* May 1971.
21. Johnson, George E., and Frank P. Stafford, "Social Returns to Quantity and Quality of Schooling." Unpublished paper, Ann Arbor, University of Michigan, December 1970.
22. Johnson, Harry G., "The Economic Approach to Social Questions." *Economica,* February 1968.
23. Johnson, Thomas, "Returns from Investment in Human Capital." *American Economic Review,* September 1970.
24. Michael, Robert T., "Effects of Education on Efficiency in Consumption." New York, NBER Occasional Paper (forthcoming).
25. Mincer, Jacob, "A Study of Personal Income Distribution." Unpublished Ph.D. dissertation, Columbia University, 1957.
26. ———, "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy,* August 1958.
27. ———, "On-the-Job Training: Costs, Returns and Some Implications." *Journal of Political Economy,* Supplement, October 1962.
28. ———, "The Distribution of Labor Incomes: A Survey With Special Reference to the Human Capital Approach." *Journal of Economic Literature,* March 1970.
29. Rosen, Sherwin, "Knowledge, Obsolescence and Income." Unpublished manuscript, University of Rochester, November 1970.
30. Schultz, Theodore W., "Education and Economic Growth." In *Social Forces Influencing American Education,* edited by Nelson B. Henry, University of Chicago Press, 1961.
31. ———, "Education and Investment in Human Capital." *Journal of Political Economy,* Supplement, October 1962.
32. ———, "Human Capital: Policy Issues and Research Opportunities." A paper prepared for a series of NBER colloquia in honor of the Bureau's Fiftieth Anniversary, November 1970.

33. Silver, Morris, "An Econometric Analysis of Spatial Variations in Mortality Rates by Race and Sex." In "Studies in the Economics of Health and Medical Care," edited by Victor Fuchs, Unpublished manuscript, NBER, 1971.
34. Solmon, Lewis C., "The Relationship Between Schooling and Saving Behavior: An Example of the Indirect Effects of Education." Unpublished manuscript, NBER, February 1971.
35. Taubman, Paul, and Terence Wales, "Higher Education: An Investment and a Screening Device." Unpublished manuscript, NBER, 1970.
36. Welch, Finis, "Education in Production." *Journal of Political Economy*, February 1970.

THE NBER URBAN SIMULATION MODEL AND URBAN ECONOMICS IN THE 1970's—John F. Kain

INTRODUCTION

Many economists and others have assumed that urban economics is nothing more than economists' observations about a variety of policy questions, such as poverty, racial discrimination, low-income housing, transportation, and education. This view is mistaken.

Urban economics has, or should have, a number of distinct analytical and theoretical foci. These are the determinants of the spatial distribution and location of activities within urbanized areas, especially the individual and aggregate location decisions of households and firms. The potential value of urban economics to public policy arises from the special insights it can provide about the ways in which public policy either influences or is influenced by the location decisions of firms, households, and other decision-making units within urbanized areas. To provide these special insights, urban economics must develop a theory which describes the determinants of urban spatial structure and the processes of urban development.

Existing techniques of economic analysis are inadequate for the development of such a theory. The weaknesses of existing theories of location and urban structure arise from their need to resort to unrealistic assumptions about the distribution of employment in urban regions, the characteristics of housing and of housing markets, and the interrelationships of workplaces and residences. These unrealistic assumptions are needed to make the theory analytically tractable. The calculus, the primary tool of microeconomics, is simply incapable of handling the complexity of urban economic phenomena. As long as urban economics theory continues to rely on the calculus, its progress will be painfully slow; and it will never be able to deal adequately with a number of the central issues of urban economics.

The apparent inadequacy of existing urban models and their inability to handle important aspects of urban phenomena has led numerous researchers to consider whether large computer simulation models might provide a means of developing more satisfactory explanations of urban growth and development. The appeal of large computer simulation models lies in their ability to handle far more complex systems and to employ more elaborate and realistic assumptions. It may be possible by iterative techniques to solve problems that could not be solved by existing analytical methods.

This possibility led John Meyer and me to propose the design of a large scale urban simulation model over ten years ago as part of a RAND Corporation study of urban transportation sponsored by the Ford Foundation.[1] Our sketch of a computer simulation model of urban growth and development was one of several roughly simultaneous efforts to formulate such a model. Subsequently various efforts were made to construct computer simulation models in a number of metropolitan areas, generally as part of comprehensive urban transportation studies.

NOTE: Building a computer simulation model of the kind represented by the NBER Urban Simulation Model is a complex undertaking requiring a diverse set of skills and the efforts of a large number of individuals. All members of the NBER Urban Studies Group participated. However, Gregory Ingram, Royce Ginn, H. James Brown, and I were primarily responsible for the design, programming, and implementation of the model. Stephen Dresch played a major role in estimating the critical submarket demand equations for the model. The contributions of John R. Meyer, as usual, are both major and difficult to describe. Mahlon Straszheim, Stephen Mayo, Daniel Fredland, Elizabeth Pinkston, Robert Goldberg, John Quigley, Joseph Persky, and Irving Silver also made valuable contributions during the two and one-half years the model was under development.

[1] John F. Kain and John R. Meyer, "A First Approximation to a RAND Model for Study of Urban Transportation," The RAND Corporation Memorandum RM-2878-FF, October 1961.

These models never lived up to the promises of their most enthusiastic proponents, and the sponsors were often disappointed by the results. As a result, urban simulation models are somewhat out of fashion.

The disillusionment with urban simulation models resulted in large measure from unrealistic expectations about what could be *quickly* learned from them, serious underestimates of the difficulties of constructing truly useful models, and the lack of adequate, long-term financial support for their development. It has long been evident to researchers that to build a truly useful urban simulation model would be exceedingly difficult, expensive, and time-consuming.

Inadequate theory and empirical knowledge about urban phenomena, limitations of computer technology, and lack of financial support for the modeling project prevented us from completing our proposed urban simulation model at RAND. Our interest continued, however, and in the ensuing years research on urban phenomena increased the base of theory and empirical knowledge, and computers increased in size and speed. A grant from the Department of Housing and Urban Development to the National Bureau of Economic Research in the spring of 1968 eased the budgetary constraint, and we began again to develop an urban simulation model.

Last fall we completed programming, preliminary calibration, and initial runs of the NBER Urban Simulation Model. Our experience in constructing the model, the growing knowledge of firm and household behavior, and the rapid growth in computer technology make me exceedingly optimistic about the contribution that large computer simulation models can make to economics and particularly to urban economics. Such models have the potential to revolutionize urban economics in the decade of the 1970's. The discussion of the NBER model that follows and the manner in which it corrects a number of the most serious deficiencies of existing theories of location and urban structure should make evident the reasons for my optimism.

A BRIEF DESCRIPTION OF THE NBER MODEL[2]

The NBER Urban Simulation Model is a generalized, policy-impact model based on empirical research for a large number of cities; it represents those aspects of firm, household, and market behavior common to all cities. The current version of the model makes extensive use of data on Detroit during 1960-63. However, the model city is not Detroit. We needed a reasonably consistent description of a city in terms of employment, population, housing, housing prices, and transit and highway networks to begin the simulations. Detroit was a convenient source of data for this purpose, although not entirely satisfactory.

The hypothetical metropolitan region included in the current version of the NBER model is divided into nineteen workplace and forty-four residence zones. The model city contains only employed households, and each household has only a single worker employed at one of the nineteen workplaces. Each household in the model city belongs to one of seventy-two household classes, defined in terms of family size, family income, and the education and age of the head. Each household lives in one of twenty-seven distinct types of housing, defined in terms of structure type, number of rooms, quality, and lot size. The worker travels to and from his workplace by one of two modes of travel. The modes are depicted in terms of interzonal travel time and cost.

The model is designed to simulate major changes in urban spatial structure that occur over periods ranging from ten to fifty years. Our principal theoretical inter-

[2] A more detailed description of the model is contained in: Gregory Ingram, J. Royce Ginn, H. James Brown, and John F. Kain, *The NBER Urban Simulation Model: Vol. I:* "The Model Description," Report to the Department of Housing and Urban Development, Office of Urban Transportation Development and Liaison, Division of Systems Research and Development, December 1970.

est is with the effects on the spatial structure of urban areas of long-term trends in the level and spatial distribution of employment, of changes in transportation technology, of increases in income, and of the growth in employment and population. Our principal policy concern is with the indirect and relatively long-term impacts of various public policies on urban spatial structure, on investments in residential and nonresidential capital, and on changes in the characteristics of neighborhoods. Other important policy concerns, such as racial discrimination in housing markets, could not be incorporated in this version of the model for reasons discussed later in this paper.

The model begins with a description of the spatial structure of the model city at a point in time and modifies this structure over a period of years by simulating the location and investment decisions of firms, households, and housing suppliers.

The number of years covered by the simulation and the period of time represented by each iteration depend on the objectives of the simulation. For both technical and budgetary reasons, the period of time represented by each iteration will usually be greater for longer simulated periods. If we are interested in the effects of particular policies within a ten-year period, each iteration of the model would probably represent a year. However, if we are interested in the effects of a policy or major public investment over a period of twenty or thirty years, each iteration might represent two years or more.

The NBER model differs from previous urban simulation models in several respects. The most important is its explicit representation of the structure and behavior of the housing market. Previous models have represented household location decisions and changes in urban spatial structure by elaborate statistical descriptions, usually with little or no theoretical justification. Market behavior may be implicit in these empirical regularities; however, concepts such as supply and demand, and prices are rarely included in the models. During each iteration period, the NBER model simulates most of the important types of behavior which influence urban spatial structure. These include: (1) household decisions to move; (2) determination of housing prices by housing type and location; (3) the types and location of housing selected by new and moving households; (4) filtering of the housing stock from one quality stratum to another; (5) renovation and modification of the housing stock; (6) the construction of new housing; (7) changes in the pattern of interzonal travel to and from work.

The current version of the model obtains market prices for each of the twenty-seven house types in the forty-four residence zones during each iteration, and uses this price information in determining the demand by housing consumers for each kind of housing in each residence area and the response of housing suppliers to this demand. Housing prices and access costs influence the choice of both housing type and location by new households, immigrants, and intrametropolitan movers. The rate at which the housing stock in a particular neighborhood is improved or allowed to decline in quality depends on the maintenance policies of both landlords and resident owners. In the NBER model the choice of maintenance policies depends on their relative profitability. Relative profitability, in turn, depends on the relative prices of housing of each quality stratum in each neighborhood.

More substantial physical modifications of the existing housing stock, such as the merging or conversion of existing structures and new construction, are similarly based on the profit which housing suppliers can expect to obtain from engaging in each kind of supply activity in each residence zone or neighborhood. Prices determine how much a housing supplier can expect to obtain from the sale of each kind of housing in each neighborhood.

During each iteration, the model seeks to satisfy a target demand for each type of housing, including a normal vacancy rate. Vacancies can fall below this normal level either because the provision of some housing types is not profitable enough or because large numbers of a particular kind of unit are transformed into other kinds. The result will be an increase in the price of these units in the next period.

The model is designed to assess the spatial implications of exogenously specified employment distributions. It is not designed to assess the effects on industry location, of changes in the distribution of households, or even to evaluate the effects on industry location of changes in technology, income, and tastes.

It is likely that later generations of the model will handle population-serving employment endogenously. Population-serving employment is that one-third of total metropolitan employment that serves the resident population, e.g., employment in convenience retailing, public schools, and personal services. The location of these activities bears a predictable relation to the population of each residence zone. As a result it is fairly easy to add this feature to the model.

By contrast, we have no plans to make the location of basic employment—e.g., manufacturing, wholesaling, and other exporting activities—endogenous in the foreseeable future. Furthermore, we do not see how it could be done with present knowledge about the determinants of industry location. Our assessment of the available evidence suggests that, while the location of basic employment has a strong influence on the residential location decisions of individual households, as well as on the prices of land and housing, the density of urban development, and on other aspects of the spatial structure of urban areas, it is not in turn strongly dependent on the spatial distribution of the residential population. In short, we do not believe that the location decisions of basic industry are influenced by residential location decisions and housing market behavior to an extent that would justify making the location of nonpopulation-serving industry endogenous. And if a significant dependence exists, our understanding of it is too limited to dignify with an explicit representation in the model. Still, we regard an understanding of industry location as a fundamental building block for a complete theory of urban growth and development, and we are therefore analyzing changes in the location of manufacturing employment in several cities.[3] As this research increases our knowledge of the determinants of industry location, we will be able to use the NBER model to examine the consequences of employment changes for the location decisions and commuting patterns of urban households, for the behavior of the urban housing market, and for the spatial structure of metropolitan areas.

RELATION TO EARLIER STUDIES

The NBER model is perhaps best viewed as a hybrid of the empirically based computer simulation models used during the past ten years in land-use–transportation studies and the economic theories of location and urban spatial structure developed by urban economists during roughly the same period.[4] Both kinds of models have improved our understanding of the processes of urban development, but both are seriously deficient in a number of important respects. One reason is that there has been relatively little cross fertilization between these two model-building traditions.

Engineers and planners, the principal architects of most previous urban simulation models, have based their models on empirical regularities obtained from the analysis of large surveys, with little or no consideration of the theoretical problems emphasized in economic theories of location and urban structure. In the same way urban economists have paid scant attention to the descriptions of empirical reality constructed by model builders for urban transportation studies.

[3] Franklin J. James and Raymond J. Struyk, "Recent Trends in Industrial Composition and Location," in John F. Kain, editor, *The NBER Urban Simulation Model: Vol. II:* "Supporting Empirical Studies," Report to the Department of Housing and Urban Development, Office of Urban Transportation Development and Liaison, Division of Systems Research and Development, December 1970.

[4] A survey and critique of the land-use models developed for five of these studies is presented in Brown, *et al.*, "A Survey of Empirical Models of Urban Land Use: Suggestions on Research Objectives and Organization," National Bureau of Economic Research, forthcoming.

An important exception was the Herbert-Stevens model.[5] John Herbert and Benjamin Stevens proposed using a linear programming algorithm as the basis of a residential location model for the Penn-Jersey Transportation Study. They conceived of their linear programming model as a direct analog to the utility maximizing behavior assumed in economic theories of location and urban structure. Britton Harris, Vladimir Almendinger, and others on the Penn-Jersey study made strenuous efforts to implement the Herbert-Stevens model,[6] but their efforts revealed a number of unsolved theoretical problems. These difficulties, combined with the immense data and computational requirements of the model, forced the Penn-Jersey study to abandon the linear programming framework in favor of a more conventional empirical statistical land-use model. Britton Harris has continued to refine the approach proposed by Herbert and Stevens, but it is our impression that several problems remain unresolved.

ECONOMIC THEORIES OF LOCATION AND URBAN SPATIAL STRUCTURE

The NBER model differs from the recent theories of location and urban spatial structure principally in terms of the greater realism of its assumptions and its far more detailed representation of urban structure. For these characteristics it owes much to the urban simulation models developed by transportation planners.

All economic models of residential location and of urban spatial structure depict the location decisions of urban households as resulting from utility-maximizing behavior.[7] Specifically, those models assume that a household chooses that residential location which maximizes its real income. This behavioral assumption, as well as the assumptions that all employment is located in the center and that monthly travel outlays increase as the distance between work and home increases, permits the authors of these theories to generalize about the locational patterns of different income groups and the spatial configuration of housing prices and density. Although these theories pay less attention to the determinants of industry location, firms are also depicted as choosing that location within the metropolitan region that maximizes their profits.

The several models differ in their precise formulation of the problem. In every case, however, competition for sites more accessible to the center produces a systematic decline in the price of urban land with distance from the core. These rent gradients and the greater travel costs required to reach more distant residence locations interact to influence both the location of particular households and the intensity of residential development. In the typical theoretical treatment, households consuming larger amounts of urban land find it advantageous to commute long distances from the single employment center to outlying residences. Households consuming smaller amounts of land or having unusually high transportation costs find that the savings from cheaper peripheral land are too small to justify the long trips required. Higher land prices cause households and firms to use land more intensively. Land prices decline with distance from the center; therefore, densities also decline.

These models have widespread acceptance. Many persons believe that they pro-

[5] John Herbert and Benjamin Stevens, "A Model for the Distribution of Residential Activity in Urban Areas," *Journal of Regional Science*, Vol. II, No. 2, Fall 1960, pp. 21-36.

[6] V. V. Almendinger, "Topics in the Regional Growth: I" (Penn-Jersey Transportation Study, PJ Paper No. 4, Philadelphia, 1961); Britton Harris, "Linear Programming and the Projection of Land Use" (PJ Paper No. 20, 1963).

[7] For example: William Alonso, *Location and Land Use* (Harvard University Press, Cambridge, 1964); John F. Kain, "The Journey-to-Work as a Determinant of Residential Location" (Papers and Proceedings of the Regional Science Association, Vol. IX, 1962, pp. 137-161); Edwin S. Mills, "An Aggregative Model of Resource Allocation in a Metropolitan Area" (*American Economic Review*, May 1967, pp. 197-211); Richard Muth, *Cities and Housing* (University of Chicago Press, Chicago, 1969); Lowdon Wingo, Jr., *Transportation and Urban Land* (Resources for the Future, Inc., Washington, D. C., 1961).

vide satisfactory explanations of the geographic stratification of different income groups and of density gradients at any moment in time as well as changes in these distributions over time. For example, the tendency for low-income households to locate in the central parts of cities is attributed to the lesser quantity of housing they consume. The existence of slums is explained as the natural result of market forces and of the tendency of centrally employed low-income households to spend too little on housing maintenance. Historical declines in the level and density of population in central areas and increases in the level and density of suburban areas are alleged to be the result of secular increases in incomes and of declines in the cost of passenger transportation.

These models have provided useful insights about the determinants of residential location and the behavior of urban housing markets, and there is undoubtedly a good deal of truth in their conclusions about the determinants of central city declines in population, the suburbanization of urban households, and the creation of slums. However, closer examination of these models, and especially the comparison of their asumptions with empirical reality, raises serious doubts about their completeness.

Any theory must abstract from and simplify reality in order to make the world understandable. Indeed, this is the essence of good theory. However, the admission that simplifying assumptions are both desirable and necessary does not mean that the realism of a model's assumptions should be ignored.

An appropriate test of an economic theory of location and urban spatial structure is its ability to explain historical patterns of urban development. There have been a number of ingenious attempts to test these theories empirically, and many persons believe these tests substantiate the theories.[8] Unfortunately, the tests are too primitive and do not provide a measure of the extent of correspondence between theory and reality.

In our view, the theories are particularly vulnerable on three counts:

1. They assume that all production takes place at a single location.

2. Because they obtain only long-run equilibrium solutions, they ignore capital stocks entirely and fail to consider the effects of heterogeneous, durable, and locationally fixed capital.

3. They make no effort to acknowledge various kinds of interdependence that appear to be important in urban housing markets. These include housing consumption and production externalities, racial segregation and discrimination, and the provision of local public goods.

The current version of the NBER model deals explicitly with the first two problems. More advanced versions of the model will deal with the third set of issues. As a consequence, it is useful to view the NBER model as an economic theory of residential location and urban spatial structure which includes a large number of spatially separate workplaces and which explicitly incorporates durable and heterogeneous stocks of residential capital.

THE MONOCENTRIC ASSUMPTION

The assumption that all production takes place at a single center is acknowledged by its authors as lacking in realism, and all of them attempt to incorporate some noncentral employment into their models. The most common assumption is to define a category of local workers, who presumably provide services to the neighborhood. The inclusion of these local workers into the model cannot, however, be regarded as a meaningful departure from the monocentric assumption, since the behavior of these local workers is considered only in the most trivial way, and their inclusion has no effect on the solutions obtained from the theories.

---

[8] The most elaborate tests are by Muth, *op. cit.* Other evidence is contained in Mills, "Urban Density Functions," *Urban Studies*, Vol. 7, No. 1, February 1970.

If the comprehensive urban transportation studies of the postwar period have done anything, they have made clear the inappropriateness of the monocentric assumption. It is rare that as much as 10 per cent of all employment is located in the core or central business district, and the central city will often contain less than half of all metropolitan employment. In 1963, 52 per cent of all manufacturing and 29 per cent of all wholesaling employment in forty of the largest metropolitan areas was found outside the central city. The fractions are still larger today.

The prevailing trend over the past half century has been a relative, often absolute, decline in central employment and a rapid growth of jobs in suburban areas. Theories that claim to explain the suburbanization of urban populations, changes in the length of the journey-to-work, and modifications of central and suburban densities, without explicit references to these changes in the distribution of employment, must be viewed with suspicion. Many of the past changes in urban structure which these theories attribute to increases in incomes and to declines in the real costs of transportation, may instead be the result of changes in employment location. Existing empirical tests of the theories do not distinguish between these explanations.

The NBER model, in the tradition of economic models of residential location, assumes that a household chooses that housing type and location that maximize its real income. The household bases its decision on the combined housing and transport costs it would incur in consuming each type of housing. These combined costs, or residential expenditure amounts, are the expected sum of outlays for the journey to work, including both money and time expenditures, and housing expenditures. Each household employed at each workplace estimates the expected residential expenditure required to consume each of the different kinds of housing. The prices represented by these combined costs are used in econometrically estimated housing demand functions which determine the kinds of housing members of each household class purchase.[9] These prices, which differ from one workplace to another, explain the tendency for workers employed in suburban workplaces to live at lower densities than workers with similar incomes and other characteristics employed at central workplaces.[10]

As indicated earlier, the current version of the NBER model takes the approximate distribution of employment in Detroit in 1963 as the starting point in obtaining location solutions for single worker households employed at nineteen spatially separate workplaces and produces housing prices for each of forty-four residence areas. This exceedingly large and complex problem cannot be solved analytically with existing methods. However, problems of this kind can be solved by iterative techniques using the computational power of high-speed computers. We have

[9] The demand equations used in the current version of the NBER Urban Simulation Model are based principally on the research of Stephen Dresch, using data from the Detroit TALUS Study. (Stephen P. Dresch, "The Demand Model: Theoretical and Empirical Foundations," Chapter 2 of the *NBER Urban Simulation Model: Vol. II, op. cit.*) Jim Brown and I obtained generally consistent and supportive results using the supplementary home interview study of San Francisco, collected by BATS. (H. James Brown and John F. Kain, "Submarket Demand Equations for San Francisco," *ibid.*, Chapter 3.) Mahlon Straszheim, using a somewhat different approach, obtained findings which generally support the approach used in the model and suggest a number of ways in which the current model design might be improved and modified. (Mahlon R. Straszheim, "The Demand for Residential Housing Services, Housing Markets, and Metropolitan Development," *ibid.*, Chapter 4.)

Serious data problems magnified the theoretical and estimation problems inherent in calculating the submarket demand functions. Although data from the Pittsburgh Transportation Study appeared to be the most suitable, they did not become available until June 1970—too late to be used for the current version of the model. There would probably be substantial advantages in shifting to a Pittsburgh data base for further model development.

[10] John F. Kain, "The Journey-to-Work as a Determinant of Residential Location," *op. cit.*; John R. Meyer, John F. Kain, and Martin Wohl, *The Urban Transportation Problem*, Harvard University Press, Cambridge, 1965; John F. Kain, "A Contribution to the Urban Transportation Debate: An Econometric Model of Urban Residential and Travel Behavior," *The Review of Economics and Statistics*, February 1964.

used a linear programming algorithm to help us solve the problem.[11] When we began work on the NBER model, we regarded the use of linear programming as appealing but impractical. After considering and discarding a number of alternative approaches, however, we conceived of an operationally feasible way of using linear programming, both to locate individual households and to provide location rents by housing submarket.

We made the problem computationally feasible by breaking it into two parts. Households employed at specified workplaces are first allocated to housing submarkets (housing types) using the submarket demand equations; then the households competing in each of the housing submarkets are allocated to specific residence areas by means of linear programming. All households competing in each housing submarket, employed at spatially distinct workplaces, are located in the several residence areas so that total journey-to-work costs are minimized.

The linear programming solutions provide two critical kinds of information for the model: (1) the number of worktrips between each workplace and residence zone for households choosing residences during the period; and (2) shadow prices for each type of housing in each residence zone. These shadow prices are averaged over several years and are interpreted as estimates of market prices for the current year. These current prices, which are used by housing demanders and suppliers in making their decisions, provide the links between the choice of housing types by households and the choice of residence areas, as well as the links between the demand for housing by type and location and its supply.

HOUSING STOCKS AND LONG-RUN EQUILIBRIUM

In addition to evaluating the implications of relaxing the monocentric assumption, the NBER model abandons the highly restrictive long-run equilibrium framework which characterizes all existing economic theories of location. These theories employ the method of comparative statics, which involves the analysis of the distribution of employment, population, income, and other relevant characteristics that would exist in long-run equilibrium. The long-run equilibrium state considered by these theories requires a full adjustment of the capital stock to any changes in supply or demand conditions.

Long-run equilibrium models provide no information about the process of stock adjustment or the time path of adjustment. The failure to consider explicitly these dynamic adjustment mechanisms is a general weakness of economic theory and analysis. However, because capital stocks are especially important for urban housing markets, this weakness is particularly serious in the analysis of urban development.

Stocks of nonresidential and residential capital in cities are seldom demolished and replaced by new structures. Furthermore, these stocks have a powerful effect both on the types of new investment and on its location. New construction will be concentrated on those types of housing services that are not easily or cheaply produced from the existing stock of residential capital. Except when there are compelling locational advantages, new construction will occur on vacant land—most of which is found at the periphery of the built-up area. The result is that the spatial distribution of housing capital of different types will depend on the timing of development and will differ from that which would occur if the city were built *de novo* each year.

The NBER model explicitly represents stocks of residential capital. The model does not obtain the long-run equilibrium distributions of residential capital consistent with the existing levels and distributions of employment, incomes, and other factors believed to influence the quantity and location of housing services

[11] Use of linear programming algorithms as a way of solving the multiple workplace-multiple residence problem was first proposed by Herbert and Stevens *(op. cit.)*. However, the approach used in the NBER model is completely different from the one they proposed.

consumed by urban households. Instead, it obtains estimates of the desired demand for housing by type and location during each time period, and modifies the existing stock through maintenance, renovation, repair, and new construction. Moreover, the prices that determine the desired demand in each period are not long-run equilibrium supply prices, but rather a set of market prices that reflect the composition and location of existing stocks of residential capital.

HETEROGENEITY OF THE STOCK

A major advantage of the long-run equilibrium assumption is that it permits the theorists to ignore all aspects of the heterogeneity of housing except location and price. In the long run, any kind of housing can be produced at any location within the metropolitan area at its long-run supply price. Since nonland factors of production are assumed to cost the same everywhere in the metropolitan area, housing prices vary from one part of the region to another only because land prices differ. However, once durable stocks are included in the model, this condition no longer holds; and no simple relationship exists between the price of various types of housing and their location within the region.

The NBER model permits the relative prices of different types of housing bundles to vary from one part of the region to another. At most, existing economic theories of location recognize only two attributes of housing services: accessibility to the center and the quantity of residential space consumed. However, it is apparent that the bundles of residential services consumed by urban households consist of a large number of additional attributes. Many of these attributes are difficult to modify by the actions of individual property owners and require collective action.

Economic theories of location and urban structure obtain only a single price gradient for urban land. Empirical research by Mahlon Straszheim, however, indicates that the situation is far more complex than these theories assume.[12] Straszheim's results strongly suggest that there are different price gradients for different attributes of the bundle of residential services.

The NBER model defines a series of housing submarkets in each of which housing services are assumed to be homogeneous. The present version contains twenty-seven distinct submarkets defined in terms of structure type, number of rooms, and dwelling unit/neighborhood quality.[13]

PROBLEMS OF INTERDEPENDENCE

Especially vexing problems of heterogeneity arise from consumption externalities, from the public provision of local public goods, and from other so-called neighborhood effects. As analyses by John Quigley and me indicate, attributes of the bundle of residential services that are in some sense external to the particular property or dwelling unit, (i.e., the quality of local public schools, neighborhood crime, neighborhood prestige, etc.) appear to be as important as attributes of the dwelling unit itself (i.e., size, condition, number of rooms, etc.).

Existing economic theories of location and urban structure ignore these external dimensions of the bundle of residential services. Unfortunately, the lack of adequate theory, the paucity of persuasive empirical evidence about how these attributes are produced and how they influence household behavior, and model-building priorities have prevented a satisfactory representation of them in the present

[12] Mahlon R. Straszheim, "The Demand for Residential Housing Services, Housing Markets, and Metropolitan Development," *NBER Urban Simulation Model, Vol. II,* Chapter 4.

[13] The housing submarket definitions used in the present version of the NBER model must be regarded as highly tentative. We suspect the model's performance would be greatly improved by better submarket definitions. The best way to accomplish this difficult and complex task is still unclear, but we are actively researching this area.

version of the NBER model. However, research on household demand for these attributes and on the determinants of their supply has been a prominent part of all our econometric studies of the housing market. Moreover, the NBER model has great potential for testing various hypotheses about the relationship between individual and aggregate location decisions. For example, when an individual household makes location decisions, it is influenced by its perception of the present and anticipated quality of public goods in various jurisdictions. But it is just as evident that the quality and types of public goods in a jurisdiction depend on the characteristics of past and present residents of that community. Heretofore, no suitable models have existed for considering these kinds of interdependencies. The NBER model seems ideal to study the interpendence between the demand for and provision of local public goods.

Racial discrimination in urban housing markets is an obvious and particularly important type of interdependence. The fact that it is not included in the present version of the NBER model results from model-building priorities rather than our evaluation of what is most important. Before introducing any additional complexities, even one as important as housing market discrimination, it was first necessary to have a basic market model. We have, however, several ideas about how we might represent housing market discrimination in the model, and we regard this extension of the model of highest priority. If these ideas prove fruitful, the use of the NBER model to evaluate the consequences of discrimination could be its most valuable contribution.

CONCLUSION

The NBER Urban Simulation Model should provide valuable insights about the probable effects of a wide variety of proposed public policies. For example, we should be able to evaluate the effects of alternative transport investments on the location decisions of urban households, on the kinds of housing they consume, and on the density and structure of urban development. Similarly, the model should be useful for evaluating a wide variety of housing programs. Among the most important of these are programs, such as rent supplements, that seek to improve housing conditions by increasing the purchasing power of low-income households. A central concern about these proposals is that the subsidies may not increase the supply of housing; they may simply increase prices and enrich existing property owners.

A variety of other programs—most notably the urban renewal and model cities programs—are concerned less with improving housing standards generally than with improving the quality of particular communities or neighborhoods. The NBER model, with its emphasis on the spatial dimensions of the housing market, is ideally suited for evaluating both the direct and indirect consequences of such programs.

It would be misleading, however, to suggest that the current version of the NBER model can provide satisfactory answers to these and similar questions. It is still a crude and incomplete prototype. Moreover, the empirical estimates of some of the functions included in the model are unsatisfactory.

The current version of the NBER model has been programmed for the IBM 360 computer and has been run on both IBM 360-67 and IBM 360-91 machines. It is capable of simulating the behavior of households and of the housing market in the way described; however, the results of the simulations are inadequate representations of Detroit or of any other city. The problem is not the operation of the model or its theoretical structure. The major source of difficulty is the lack of some critical data for Detroit, the city we have relied most heavily on in calibrating the model. These data problems can be circumvented to some extent. To obtain satisfactory results, however, we may have to recalibrate the model for Pittsburgh, San Francisco, or some other city where data problems are less severe.

Perhaps the most important lesson to be gleaned from our efforts is that the time of computer simulation models is here. When John Meyer and I proposed the development of an urban simulation model at RAND, ten years ago, the largest computer generally available was the IBM 7090. We estimate that to run a single year's simulation of the NBER Urban Simulation Model on the IBM 7090 would have required at least four hours; to run the model on the 7094, the largest computer in widespread use until a year or two ago, would have taken from forty-five minutes to an hour of computer time and probably another 20 minutes for tape storage. This assumes, of course, that enough core storage existed on the 7090 and 7094 computers to run a model of this size, and that is not true by a factor of two. In the late 1950's, a year's simulation with the NBER model would have cost at least $7,200, assuming again that the model could have been run at all. Today it costs less than $90 per simulated year on the IBM 360-67 and $50 per simulated year on the 360-91.

Even a conservative projection of computer technology indicates that the only barriers to the development of truly useful urban simulation models are our ingenuity and technical know-how. The projected growth in computer capability, more than any other factor, is the basis for my contention that computer simulation models will revolutionalize urban economics in the next decade. If they do not, we have failed to respond to the opportunity available to us.

ECONOMIC MODELS OF FERTILITY: SOME EXAMPLES AND IMPLICATIONS — Warren Sanderson and Robert J. Willis

Until recently, it has been difficult to ask many questions of economic theory concerning household behavior because of the rigid characterization of the household in standard theory. In this theory, a household is characterized as an entity which attempts to maximize its utility, subject to the constraint that its expenditures on market goods not exceed its fixed income. It is obvious that an attempt to apply economic theory to a phenomenon such as fertility will have to be made largely outside the framework provided by this model. Outside of the model, to name but a few things, are factors influencing family formation and composition, the division of labor within the household, supplies of labor by family members to the market, educational and job choices by family members, child rearing practices, methods of fertility control chosen and efficacy with which they are used. Inside the model is mainly the postulate of rational economic behavior, the principle of constrained maximization.

Recently, the capacity of economic theory to accommodate a more richly specified structure of household tastes and constraints has been greatly advanced by the work of Gary Becker [1] and, independently, Kelvin Lancaster [3]. This work has made it possible to apply the postulate of rational, maximizing behavior rigorously and flexibly to a wide range of problems, including fertility behavior.

Households, following Becker's development of the argument, cannot directly satisfy their wants by the consumption of market goods, but instead use these goods in combination with the time of household members to produce psychic "commodities," which are the true object of consumer wants in the sense that their "quantities" enter as arguments into the household utility function while the quantities of market goods or members' time do not. The properties of household production functions by which inputs of goods and time are transformed into commodities are governed by the state of consumer technology in exactly the same sense that the properties of manufacturing production functions are said to be governed by the state of manufacturing technology. The household obtains its supplies of market goods either by spending nonlabor income or by withdrawing the time of household members from home production, "exporting" it to the labor market and, with the income earned thereby, "importing" market goods. In the short run, the market wage rate obtained by each household member may be regarded as parametric, being determined by the set of market wage rates for persons of given characteristics. In the long run, however, a person may to some extent choose some of his characteristics by building up or running down his stock of human capital through investment in education, training, health, etc., which are themselves produced in the household with inputs of time and goods. The relationship between a person's market productivity and his stocks of various types of human capital may be termed his "earnings function."

The structure of the household, in Becker's theory, thus consists of (1) the household utility function whose arguments are quantities of psychic commodities, (2) the household production functions by which inputs of members' time and market goods are transformed into commodities and additions to stocks of human capital, and (3) the earnings functions of each household member which relate his stocks of human capital to his market earnings, given the set of market wage rates. The variables subject to household control (i.e., its control variables) are (1) the quantities of commodities, (2) the quantities of time and goods inputs to the production of each commodity and type of human capital, (3) the amounts and distribution among members of each type of human capital, and (4) the quantity of time each household member allocates to market work and to home work. Finally, the "state" variables, which the household cannot control, are in two groups: (1) variables whose values were determined by a past decision, such as the amount of human or physical capital each member has during a particular period and (2) variables which are outside the control of the household throughout its life-cycle. Principally, these are each member's endowment of time and the set of market prices for goods and labor.

The method of constrained maximization now provides the economist with a very powerful tool with which to determine the implications of the assumptions he has made about the household's structure. Basically, this method asks how the economic actor should choose the values of the variables over which he has control so as to maximize his utility, subject to the condition that he not violate the set of constraint functions whose values, in turn, are determined by the set of state variables not subject to his control. The answer to this question is obtained by solving the maximization problem for the optimal value of each control variable as a function of the values of each of the state variables. Each of these is a function of household members' endowments of time and of the market prices for goods and labor that the household will face during its lifetime. The nature of these functions will depend on the assumed properties of the structural equations of the model.

As any general-equilibrium theorist knows, the cost of an analytic framework which incorporates every phenomenon is that it implies nothing about any particular phenomenon. Restrictions on the elements of the model and on the model as a whole must be made, if it is to yield any implications. When such restrictions are made, the power of this approach to household behavior becomes very great indeed. The analytic methods by which this framework may be applied to actual phenomena are familiar to every economist. The model may be general or partial equilibrium, dynamic or static, disaggregated or aggregated. The appropriate technique to choose depends on interest.

The main point is that this model gives the economist a framework in which to organize his knowledge about the structure of the household in such a way that the implications of this knowledge for observed behavior can be derived rigorously rather than heuristically. We hope to show by example that the use of rigorous economic models can provide realistic implications for fertility behavior which could not be obtained without an explicit model.

TIME INTENSITY AND THE EFFECT OF INCOME ON FERTILITY

In [6], Willis proposed a model of completed fertility within the framework provided by Becker's theory. The model assumes that a married couple, living in a world of perfect fertility control and perfect foresight, chooses the number and "quality" of its children and their own standard of living so as to maximize their lifetime utility. The services of children and the parents' standard of living are assumed to be psychic commodities produced within the home, with inputs of the wife's time and market goods according to household production functions whose properties are determined by the state of consumer technology.

Given the state of consumer technology, the household's consumption and production possibilities are limited by the amount of the wife's time and of market goods available to produce child services and the parents' standard of living. The supply of the wife's time to home production is equal to her lifespan after marriage minus the number of years she devotes to market work. The supply of market goods is limited by the husband's lifetime income (he is assumed, for simplicity, to work full time and his time at home is assumed to be unproductive) plus the earnings of the wife equal to her market wage rate multiplied by her years of market work. The composition of resources available to the household is, therefore, subject to household control via the wife's supply of labor.

The wife's labor supply decision depends on a comparison between the productivity of her time in the household and the productivity of the market goods she could earn by withdrawing time from home production. To produce commodities efficiently, the wife will allocate her home work and market goods between children and the standard of living so as to equate the ratios of the marginal products of time and goods.

These ratios, which determine the shadow price of the wife's time in the home, are called the wife's home wage W_h. If W, her market wage, exceeds W_h, the wife will supply labor to the market until W_h rises to equality with W. The home wage rises because the ratio of goods to time in home production rises, causing the marginal product of time to rise and the marginal product of goods to fall. Thus, the working wife will adjust her supply of labor so as to keep her home wage equal to her market wage. Assuming the market wage to be parametric, it follows that the price of the working wife's time in home production is fixed and equal to her market wage. In contrast, if the wife is supplying no labor to the market and finds that W_h exceeds W, there is no way for her to increase her supply of home time and reduce the supply of market goods in order to reduce W_h to equality with W. Consequently, the price of the nonworking wife's time is endogenous to the model and will, in general, be an increasing function of the amount of market goods purchased with the husband's income, in the same way that the real wage of labor tends to be an increasing function of capital-labor ratios in standard economic models.

As it stands, this model has no particular implication for the fertility demand function, the relationship between the optimal number of children and the exogenous variables of the model, the husband's lifetime income and the wife's market wage rate, because the structure of the model has not been sufficiently restricted. In [6], the major implications of the model followed from the assumption that the household production functions for children and the standard of living were linear and homogeneous and, most importantly, that children are always produced more time-intensively than the standard of living. Given these two assumptions, it was proved that there is a positive one-to-one correspondence between the home wage of the wife and the opportunity cost of children in terms of the parents' standard of living. The heuristic argument is quite simple. Given constant-returns production functions, the opportunity cost of children in terms of the standard of living is given by the unit costs of the time and goods devoted to children divided by the unit costs of time and goods devoted to the standard of living, where the value of time and goods are evaluated at equilibrium shadow prices. Now let the price of the wife's time rise relative to the price of goods. Since children are more time-intensive, their unit costs rise relative to the cost of the standard of living.

The implications of this time-intensity hypothesis for household fertility behavior can be outlined as follows. Assume that, if household resources increase (holding the cost of children constant), the household will tend to increase its fertility as well as its level of child quality and standard of living. Holding utility constant, on the other hand, an increase in the opportunity cost of children will tend to reduce fertility. It does not necessarily follow, however, that an increase in the wife's market wage tends to reduce fertility and an increase in the husband's income tends to increase fertility.

Holding the wife's wage constant, let her husband's lifetime income (H) increase. Suppose, too, that at the initial low level of H the wife was working. As H increases, the cost of children remains constant and the wife withdraws labor from the market to keep W_h from rising above W, until she ceases work altogether. In this phase, the husband's income has a positive effect on fertility. As H continues to increase after the wife has ceased working, however, the increase in her home wage cannot be prevented by further additions of the wife's time to home production. As W_h rises, the opportunity cost of children rises, causing substitution effects against fertility which may outweigh the income effects in favor of fertility. Conversely, holding H constant at a high level, the market wage of the nonworking wife may rise without affecting the home wage, until it rises high enough to induce the wife back into the labor force. After that point, further increases in W will tend to raise the cost of children and reduce fertility.

Perhaps the most intriguing implication of the model for fertility is its suggestion that the form of the relationship between the optimal number of children, the husband's income, and the wife's potential market wage depends on whether the value of the home wage is governed by the wife's market wage, as it is for working wives, or whether it is governed by the husband's income, as it is for nonworking wives. In the former case, fertility will tend to be negatively related to W and positively related to H while, in the latter case, the desired number of children will tend to be unrelated to W and less positively or negatively related to H depending on the relative strengths of the income and substitution effects.

A provisional attempt has been made to verify these implications. If our model is correct, the effect of husband's income on fertility should depend on the level of the wife's market wage, while the effect of the wife's wage should depend on the level of the husband's income. The appropriate regression model to detect such an effect, if it is present, contains an interaction term. We, therefore, estimated a regression model of the following form:

$$N = a + bH + cW + d\,(H \times W),$$

where $\partial N/\partial H = b + dW$, and $\partial N/\partial W = c + dH$.

If the level of W is low, a wife will be less likely to work and the effect of H on fertility is most likely to be negative because the positive correspondence between H and W_h among nonworking wives results in a positive relationship between H and P, the opportunity cost of children. Accordingly, we expect b possibly to be negative while d must certainly be positive if the sign of the "true" income effect on fertility is positive. An increase in the wife's potential market wage will raise the opportunity cost of children and cause a substitution effect against fertility, if the wife works. It is more likely that the wife will work, the lower is her husband's lifetime income. Thus, we expect c to be negative.

This regression was run using data on the number of children ever born to white women over thirty-five, married once with husband present, drawn from the 1960 Census one-in-one-thousand sample and from a published tabulation of a 5 per cent sample of North Central, urban, native-born couples, compiled by Richard and Nancy Ruggles [5] from the 1940 Census. The parameters of the "interaction" model were estimated from seven independent subsamples consisting of three age groups of women from the 1960 Census (35-44, 45-54, and 55-64) and four from the 1940 Census (40-44, 45-49, 50-54, and 55-59).

The results of these regressions can be used to test the proposition that the regression coefficients within each age group are of the theoretically expected sign. They may also be used to see if the relationship between fertility and household economic status has remained stable over time.

The independent variables of the theoretical model are the husband's lifetime income and the wife's permanent market wage. Since neither of these variables can be observed in data, the husband's census-year income is used as a proxy for H and the wife's years of education is used as a proxy for W. Each proxy may be considered to measure the corresponding "true" variable with a random error, with the result that the estimated regression coefficients will tend to be biased. One way to reduce bias from "errors in the variable" is to use cell means from grouped data instead of data on individuals when the grouping scheme is chosen so as to be correlated with true variable(s), but uncorrelated with the error(s).[1]

[1] See Malinvaud [4, pp. 359-362]. An empirical test of the magnitude and direction of this bias was carried out by comparing the presumably unbiased coefficients estimated from grouped data with the presumably biased coefficients estimated from individual observations on members of the three age groups from the 1960 one-in-one-thousand sample. Generally the coefficients from the ungrouped regressions were smaller in absolute value than their counterparts from the grouped regressions, indicating that the direction of bias was toward zero. The magnitude of the bias over the three groups averaged about 50 per cent, 15 per cent, and 20 per cent for the coefficients of H, W, and $H \times W$, respectively.

Following this strategy, the regressions with the 1960 data were run using mean values of the husband's 1959 income for H and the wife's years of education W. The data were cross-classified by seven husband's occupation classes (farmers omitted), three husband's education classes, and five levels of wife's education. The 1940 data were cross-classified by the wife's education (five classes) and the husband's 1939 income (seven classes). The independent variables H and W were assigned values equal to the midpoint of each category and the husband's 1939 income was multiplied by two in order to measure it in dollars of 1959 purchasing power.

In addition to the interaction model with grouped data, we have run a noninteraction model from the 1960 and 1940 Censuses by omitting the interaction term $H \times W$, to see if the resulting regression coefficients tend to be unstable across successive cohorts. Such instability in the effects of H and W on fertility might be expected on the basis of our theoretical model, given substantial intercohort growth in average husband's income and wife's education.

The results of the noninteraction model are reported in Table I-1. The coefficient of W, the wife's years of education, is always negative and statistically significant, a finding that supports our hypothesis that fertility declines as the opportunity cost of children rises. The effect of income on fertility, according to the results of these regressions appears to have a trend from negative to positive across successive cohorts. The coefficient of husband's income (H) is negative and statistically significant for the earliest cohort (1881-85), remains negative and significant but declines in magnitude progressively for the cohorts (1886-90 to 1896-1900) represented in the 1940 Census, and becomes positive but insignificant for the last two cohorts (1906-15 and 1916-25) from the 1960 Census. Alternative interpretations of this trend in the magnitude of the "income effect" are available. One is that the relationship between income and fertility is an unstable one that basically reflects the effects on fertility of omitted variables such as birth control or tastes, which are correlated with income in the cross section but change independently of income over time. Another is that the noninteraction model misspecifies the way in which economic variables affect fertility. We may examine the case for the second alternative by looking at the results of the interaction model suggested by our economic theory of fertility.

The results of the regressions using the interaction model with grouped data are also shown in the table.[2] Each variable in every cohort has a regression coefficient of the expected sign and all are statistically significant. Thus, the sign of the coefficient of the interaction term $H \times W$ is positive, the coefficient of H is negative, and the coefficient of W is also negative. However, the numerical magnitudes of the coefficients b, c, and d are remarkably stable from cohort to cohort.

The empirical implications of the interaction model for the effects of economic growth on fertility trends and differentials depend on (1) the stability of the parameters of the fertility demand function across cohorts and (2) the direction and magnitude of the effects of H and W on the fertility of various population groups. Both of these questions may be examined in Chart I-1A and Chart I-1B in which, respectively, are plotted $\partial N/\partial W = c + dH$ and $\partial N/\partial H = b + dW$, using the values reported in the table for the regression coefficients b, c, and d for the seven age groups from the 1940 and 1960 Census samples. These plots vividly portray the high degree of stability of these parameters across the birth cohorts 1881-85 to 1916-25. This stability in the face of the substantial growth in average levels of husbands' incomes and wives' years of education and the substantial variation in other conditions faced by these cohorts suggests that it may be possible to apply our model to the explanation of trends in cohort fertility as well as to the explanation of cross-section fertility differentials within cohorts.

---

[2] Each cell was weighted by the square root of the number of women in the cell in order to have a homoscedastic error term. See Malinvaud [4, pp. 242-245].

TABLE I-1

Interaction and Noninteraction Regression Results
(t values in parentheses)

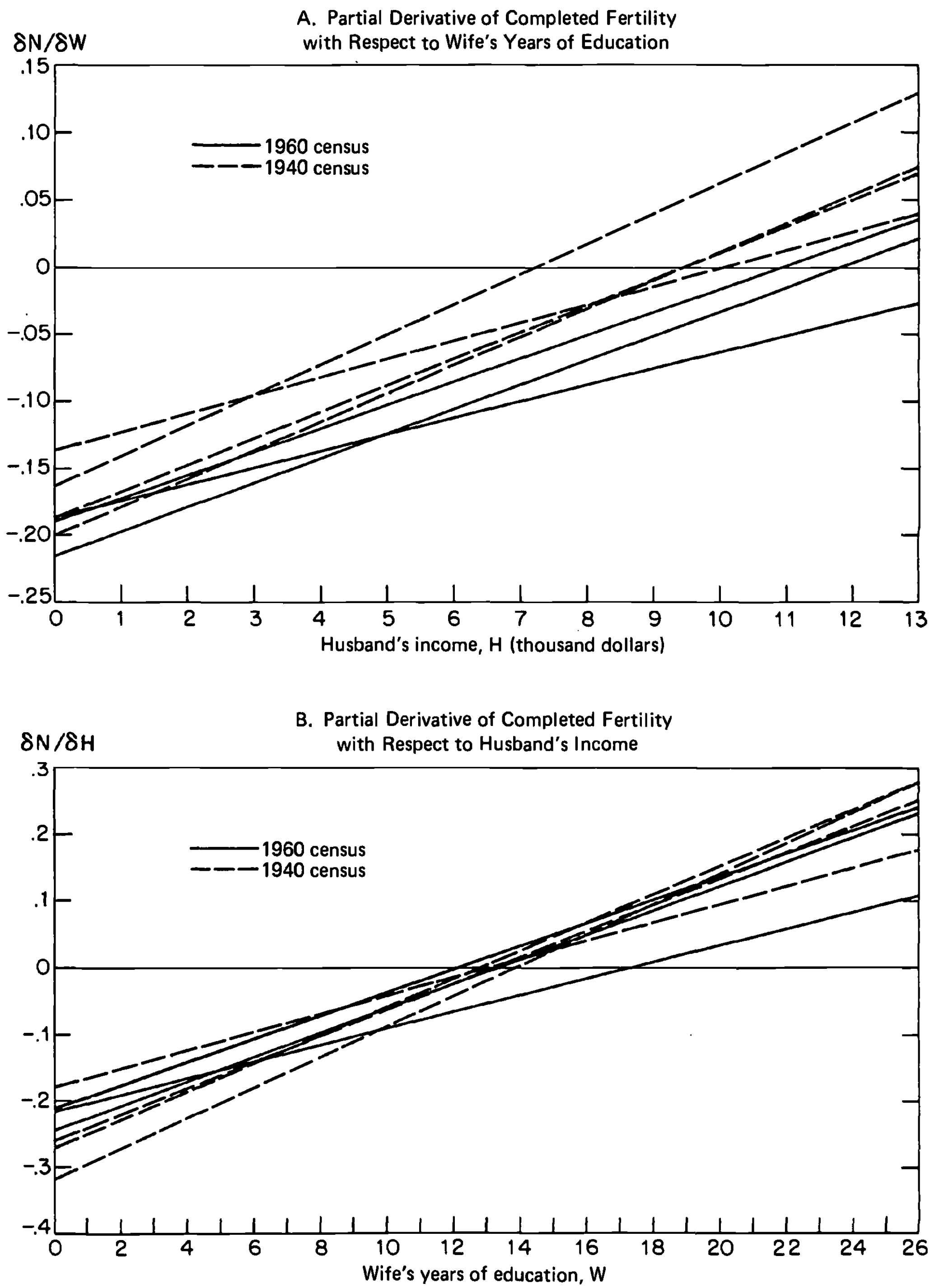
Birth Cohort	Age at Census	Husband's Income (in $10,000)	Wife's Years of Education	$H \times W$, Interaction Term	Constant	R^2
		1960 Census				
1916-1925	35-44					
Noninteraction model		0.2177	−0.0892	—	3.397	0.327
		(1.46)	(6.07)			
Interaction model		−2.1135	−0.1892	0.1723	4.724	0.436
		(3.93)	(7.25)	(4.48)		
1906-1915	45-54					
Noninteraction model		0.0400	−0.1181	—	3.466	0.475
		(0.26)	(7.27)			
Interaction model		−2.4622	−0.2153	0.1821	4.776	0.586
		(4.91)	(9.09)	(5.18)		
1896-1905	55-64					
Noninteraction model		−0.7072	−0.1212	—	3.940	0.672
		(4.41)	(7.90)			
Interaction model		−2.1509	−0.1867	0.1221	4.460	0.695
		(4.12)	(6.91)	(2.89)		
		1940 Census				
1896-1900	40-44					
Noninteraction model		−0.2224	−0.0895	—	3.138	0.643
		(1.74)	(6.36)			
Interaction model		−1.7864	−0.1386	0.1361	3.670	0.728
		(3.63)	(7.13)	(3.26)		
1891-1895	45-49					
Noninteraction model		−0.4292	−0.1148	—	3.622	0.663
		(2.70)	(6.23)			
Interaction model		−2.6246	−0.1876	0.1974	4.378	0.897
		(5.10)	(8.48)	(4.40)		
1886-1890	50-54					
Noninteraction model		−0.4626	−0.1195	—	3.734	0.561
		(2.29)	(4.85)			
Interaction model		−2.6983	−0.2000	0.2098	4.516	0.687
		(4.28)	(6.63)	(3.68)		
1881-1885	55-59					
Noninteraction model		−0.7176	−0.0916	—	3.531	0.535
		(3.38)	(3.84)			
Interaction model		−3.1876	−0.1658	0.2262	4.279	0.655
		(4.22)	(5.52)	(3.37)		

Year by year, secular economic growth has raised the market wage rates available to men and women so that the lifetime earnings capacities of husbands and wives have tended to increase. Our theoretical model and the estimates of its parameters can help us to understand the implications of secular growth in H and W for changes in cohort fertility. In Chart I-1A, we can see that growth in W will always tend to depress fertility because $\partial N/\partial W$ is negative throughout the sample range of H. In contrast, growth in H may either raise or lower fertility depending on the level of W. In Chart I-1B, we can see that $\partial N/\partial H$ is negative for values of W less than about twelve years and positive for levels of the wife's schooling above thirteen years. Thus, as the average level of W increases across cohorts, increasing proportions of married couples in the population will tend to raise their fertility in response to an increase in H, holding W constant. At the aggregate level, the sign of the partial effect of income increases on cohort fertility will change from negative to positive as the average level of W increases. This is illustrated by the change in the sign of the coefficient of H in the noninteraction regressions from negative to positive as the average level of W rose from about nine to eleven years of education across the cohorts of 1881-85 to 1916-25.

Alterations in the pattern of fertility differentials across cohorts are also capable of being explained, in part at least, by the results of our regressions. Thus, the "traditional" negative correlation of fertility with the socioeconomic status of

Chart I-1

Partial Derivatives of Completed Fertility with Respect to Exogenous Variables Implied by Interaction Regressions from 1960 and 1940 Censuses



either husband or wife is consistent with a population of relatively low average levels of H and W, in which both $\partial N/\partial H$ and $\partial N/\partial W$ are negative. As the level of W rises and $\partial N/\partial H$ becomes positive for households containing well-educated wives, income growth will tend to raise the fertility of these wives relative to that of less well-educated women, as it did in the early postwar years of the baby boom. As W continues to grow, a U-shaped pattern of fertility differentials emerges, in which there is a negative relationship between socioeconomic status and fertility among lower-status groups and a positive one among higher-status groups.

The assumption that children are time-intensive compared with other household activities, combined with a simple comparative static model of lifetime household behavior, has been shown to yield implications that are consistent with a variety of observed relationships between fertility and socioeconomic variables. In the next section, we shall demonstrate that these implications largely remain intact when the time-intensity assumption is applied to a dynamic model of household behavior, and that a variety of additional features of household behavior, including the timing of births and of labor force activity by the husband and wife, become subject to analysis and empirical verification.

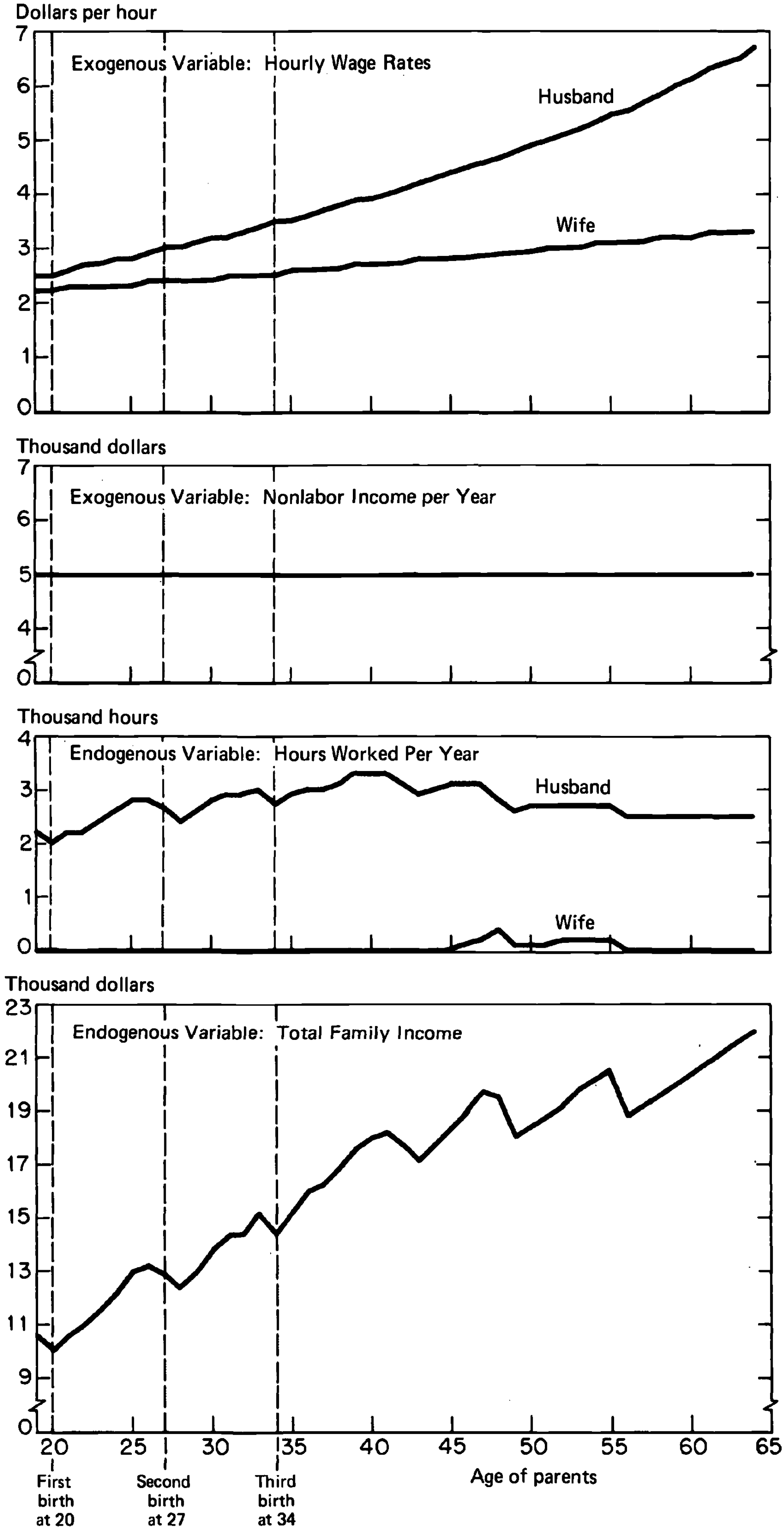
TOWARD A LIFE-CYCLE MODEL OF FERTILITY

The basic framework of the economic model of fertility discussed above may be incorporated into a life-cycle model of fertility behavior which has implications for the spacing as well as the number of children. We have been experimenting with a numerically specified dynamic model of utility maximization which is solved using a successive-approximations dynamic programming technique. While the results of our efforts with these models are still highly tentative and incomplete, our experience to date indicates that they hold great promise for modelling fertility behavior. Although the life-cycle model is based upon the comparative static model, there are some important differences that are worth noting. In the life-cycle model, the husband's time and the wife's time are inputs into all household production functions and each child is treated as if it were a stock whose value at any point in time depended on the value of the cumulated net investment his parents have made in him. Investment in each child is produced via a series of household production functions in which the goods-intensity rises and the wife's time-intensity falls as the child grows older. Given the structure of household production, the utility functional, an exogenous life-cycle hourly wage path for each parent, and an exogenous path of nonlabor income, the model determines the optimal number and timing of children, the optimal amount of time each parent spends in the labor market each year of married life, and, combining the latter with the exogenous wage rates, the optimal path of family labor income.

Chart I-2 illustrates the results of one experiment with the model. In that trial, we assumed that both parents were nineteen years old when they married, that their time horizon for childbearing decisions terminated at age sixty-four, that they faced the exogenously given wage paths shown on the uppermost panel of Chart I-2, and that they had the exogenously given path of nonlabor income shown in the next lower panel. The hourly wage-rate paths shown here roughly represent the observed cross-sectional age-specific wage rates of college educated people at the time of the 1960 Census. The lower two segments of Chart I-2 show life-cycle paths of hours of work and total family income, which are determined by the model. The vertical lines indicate when children are born.

Chart I-2 was one of three experiments designed to test the consistency of the dynamic and static models. Since nonlabor income in the life-cycle model, and the husband's permanent income in the static case represent comparable exogenous income variables, the consistency of the models was investigated by comparing the implications for fertility of changes in these variables. The results of these three experiments are summarized on p. 41. Line 3 represents the situation shown in Chart I-2.

Chart I-2
Life Cycle Paths of Selected Variables



Line No.	Annual Nonlabor Income	Present Value of Total Family Income Discounted at 7.5%	Number of Children	No. of Years Wife Participates in Labor Market
1	–$5,000	$125,006	2	34
2	0	156,254	3	22
3	5,000	191,793	3	10

These trials show that the life-cycle and static models are consistent. As predicted by the static theory, the positive effect on fertility of increasing nonlabor income diminished as the wife's labor-force participation decreased. Although the consistency for the two approaches is not surprising, it is encouraging since dynamic models need not have implications similar to those of their static analogs.

Four more experiments were conducted to investigate the implications of the life-cycle model for the effects of wage-rate changes on fertility. The hourly wage rates used in these examples roughly represent the observed cross-sectional age-specific wage rates of high school educated people at the time of the 1960 Census. Let us define (A) as this hourly wage-rate path for the husband and (B) as the hourly wage-rate path for the wife. The results of these trials are summarized below.

Line No.	Husband's Wage-Rate Path	Wife's Wage-Rate Path	Present Value of Total Family Income Discounted at 7.5%	Number of Children	No. of Years Wife Participates in Labor Market
1	10 × (A)	0.1 × (B)	$999,132	1	0
2	$2 + (A)	$2 + (B)	215,060	3	29
3	(A)	(B)	120,257	2	30
4	0.1 × (A)	0.1 × (B)	10,903	3	20

Comparing line 3 with line 4, we see that a large equiproportional upward shift in wage-rate paths for both parents is associated with a decrease in the number of offspring and an increase in the number of years the wife participates in the labor market. Thus, the growth of male and female wage rates at a constant rate may be sufficient to account for both a fertility decline and an increase in female labor-force participation. This implication may be useful in explaining observed decreases in fertility.

The wage-rate paths denoted in lines 2 and 3 differ by $2.00 an hour for each parent and each year of married life. In other words, moving from the situation in line 3 to that in line 2 represents an equal arithmetic (as opposed to proportional) shift in wage-rate paths. Since at each age the male's wage is higher than the female's in the initial wage-rate sets, (A) and (B), the wife's wage rates grow more rapidly than the husband's as we go from the situation indicated in line 3 to that indicated in line 2. Similarly, since we have characterized both the husband's and wife's wage rates as increasing monotonically with age, as we move from line 3 to line 2, both parent's wage rates grow more rapidly when they are younger than when they are older. In this particular case, these wage-rate changes lead to an increase in fertility and to a slight decrease in the number of years the wife participates in the labor market.

Line 1 represents a situation in which the husband has a relatively high market wage rate and the wife a relatively low one. This circumstance was chosen to investigate whether the dynamic and the static models were consistent in the implication that families in which the husband's income was very high would tend to have relatively few children. Line 1 does, indeed, show that this comparatively wealthy couple has only one child and that the wife never works in the labor market. Comparing lines 1 and 4, it is possible to see an example of the point we made in discussing the static model, that increasing the husband's income alone can, under certain circumstances, be associated with a decline in fertility.

In addition to the trials described above, we have varied parameters of the utility functional, altered the rate of time preference, and experimented with values of fertility control costs. Space does not allow a discussion of these tests. It is our aim to broaden the dynamic model presented here and, with parameters estimated from United States experience, to investigate whether the explicit aggregation of predicted life-cycle patterns of fertility and labor-force participation can aid in our understanding of these complex phenomena.

CONCLUSION

The economic theory discussed in this paper may be elaborated in a large variety of ways. We have already shown that the one-period, lifetime framework of the static model could be recast into a multiperiod, life-cycle framework, thereby broadening the scope of the model. The broader model includes implications for the spacing of children as well as for their numbers, and for the time paths of the husband's and wife's labor-force participation as well as for their total lifetime hours. The scope of the model, within either the static or dynamic framework, may be further broadened to include other aspects of household behavior by adding additional equations to the production structure of the model. For example, the paths of the husband's and wife's wage rates, which are treated as exogenous in this paper, may be explained endogenously by adding earnings functions in which the market wage available to a person of a given age depends on his education and experience in the labor force. Since labor-force activity is subject to choice, the market wage becomes endogenous and is chosen simultaneously with hours of work and fertility. Another example would be the addition of a fertility control function that makes the expected number of children depend on the effort devoted to reducing fertility below its "natural" level.

Given the scope of the model, alternative assumptions may be made about the form and arguments of the structural equations, which may result in alternative hypotheses about the effect of these variables on observed behavior. For example, in our regressions we have already tested the hypothesis that children are more time-intensive than other household activities against the alternative that children are equally or less time-intensive. It would also be possible to investigate the implications of assuming that household production functions shift in a particular way with the education of household members, that returns to scale have a particular pattern, or that market goods are disaggregated in some way.

The theory, then, provides a skeletal framework or structure upon which theoretical and empirical investigations of lifetime or life-cycle household behavior may be based and in light of which other work in the area may be evaluated. Given the wide variety of observed relationships between economic variables and fertility, it is easy to be skeptical about the relevance of economics to fertility research. We have shown in this paper that economic models of fertility imply connections between these variables and fertility as varied as those we observe. Despite their promise, however, the models discussed in this paper should be regarded merely as examples, drawn from research in progress, of how economics may help us to understand fertility behavior.

REFERENCES

1. Becker, Gary S., "A Theory of the Allocation of Time." *Economic Journal*, September 1965.
2. Johnston, J., *Econometric Methods*. McGraw-Hill, New York, 1963.
3. Lancaster, Kelvin, "A New Approach to Consumer Theory." *Journal of Political Economy*, 1966.
4. Malinvaud, E., *Statistical Methods of Econometrics*. Rand McNally, Chicago, 1966.
5. Ruggles, Richard and Nancy, "Differential Fertility in United States Census Data," *Demographic and Economic Change in Developed Countries*. Princeton University Press for National Bureau of Economic Research, 1960.
6. Willis, Robert J., "A New Approach to the Economic Theory of Fertility Behavior," mimeo, 1969; also "The Economic Determinants of Fertility Behavior," unpublished Ph.D. dissertation, University of Washington, 1971.

AGENDA FOR FUTURE RESEARCH: THE MEASUREMENT OF SOCIAL AND ECONOMIC PERFORMANCE—F. Thomas Juster

INTRODUCTION

Both practical men of affairs and economists would agree that income is derived from wealth, and that income is equal to the flow of current services from wealth plus or minus changes in the stock of wealth itself. Accounting systems for measuring performance reflect these relationships to different degrees, however. In one sense, the practical men of affairs do better: corporate assets generate gross income, and subtracting depreciation on assets provides a measure of profit that, while not very accurate, at least broadly reflects the return on assets. But the national income, whether measured gross or net, is clearly not designed to represent the return on the national wealth—at least, not on the national wealth as ordinarily measured. The basic reason is not difficult to see. Some two-thirds of the national income represents a payment for labor services, and the asset from which these services are obtained is not part of the national wealth as customarily defined.

The Concept of Performance

For purposes of this paper, social and economic performance is deemed to be synonymous with net social and economic output—the flow of goods and services that contribute to material welfare. This concept obviously goes well beyond the notion of output produced and sold in the market at a price, since a great many nonmarket activities that produce net output and thus economic welfare have market counterparts from which guidance may be obtained on valuation. It makes no more sense to permit the allocation of activities between the market and nonmarket sectors to influence the total of material well-being than it seemed to Kuznets many years ago to exclude residential housing from the concept of national income simply because housing services were not generally sold by business firms to consumers.

On the other hand, even a broad measure of social and economic output is not synonymous with a psychological measure of human welfare. The measurement of social and economic performance will inevitably be based on evaluation derived from activity in the market, and hence will necessarily assume that goods and services acquired or foregone are worth their market price or the equivalent.

A Framework for Measuring Performance

In general terms, economic and social output can be thought of as a flow of satisfactions or utilities generated by combining the services of various types of capital assets. A wide variety of such assets exist in the system, and they produce a number of different kinds of utilities. The assets themselves can be classified into five broad categories: tangible capital assets (equipment and structures); intangible capital assets (knowledge); human capital assets (acquired skills and basic abilities); physical environmental assets; sociopolitical environmental assets.

Tangible capital assets comprise business assets, consumer assets divided into housing and durables, and government assets.[1] Intangible assets result from the application of human capital and other resources to research and development problems. This process results in the production of socially useful knowledge, a type of asset that is analytically distinct from the skills and talents of the people

NOTE: This paper draws heavily on my essay in the *1970 Annual Report* of the National Bureau, "On the Measurement of Social and Economic Performance." The emphasis here is somewhat different, being on research priorities, but the conceptual framework is taken directly from the other paper as is the section on Environmental Assets and Returns.

[1] In the present U.S. Income and Product Accounts, only tangible business assets and housing are treated as capital assets.

who produced that knowledge. It includes, but is not limited to, patented products and processes. Human capital assets represent both innate ability and training, the latter ranging all the way from parental time spent with children through formal schooling and on to work experience designed to aid future productivity. Physical environmental assets can be thought of as comprising both natural resources as traditionally viewed, such as mineral and agricultural wealth, other natural assets like temperature, precipitation, water and air, and partly man-made assets like dams, forest preserves, and parks. The assets comprising the physical environment and the sociopolitical environment overlap to some degree. While welfare-producing assets like the amount and distribution of water resources and the quality of the atmosphere clearly belong in the physical environment category, environmental factors like population density are partly physical and partly social. The major assets in the sociopolitical category are difficult to define precisely, but are meant to cover concepts like equity, security, freedom, social and economic mobility, privacy, and so forth.

Net economic and social output can be defined as the sum of direct consumption benefits yielded by this collection of assets plus or minus net changes in the assets themselves. The major differences between the present income and product accounts and the ones implied by the concepts just outlined are: First, a much wider range of outputs would be recognized as contributing to economic and social welfare, including some that are free for some countries or regions while only obtainable through the use of scarce resources for others. Second, changes in stocks for a much wider range of assets would be explicitly taken into account, with a resulting tendency to increase or reduce measured net output depending on whether assets were being augmented or reduced as a consequence of activity in the system. For example, deterioration of the physical or sociopolitical environment because of various types of externalities—air, water, noise and waste pollution, congestion, rising crime rates—means that the flow of benefits from these assets has been reduced. Thus, where the process of economic growth deteriorates the environment, an augmented set of accounts would register the usual increases in net output resulting from growth in the market sector, but they would also record an offset consisting of the degree to which environmental assets had been depreciated with a consequent reduction in the flow of future benefits.

RESEARCH PRIORITIES

Empirical implementation of this general framework for measuring social and economic performance is the main subject of the paper. The areas that seem most amenable to quantitative investigation and of most importance for the illumination of policy choices are: (1) measurement of the returns from human capital, especially the nonmarket returns; (2) measurement of the by-product effects of economic activity on the yield from environmental assets; (3) measurement of output for economic activities that are typically valued at production costs and sold collectively, e.g., services like public education, police protection, etc.; (4) provision of a more analytically useful distinction between output that yields immediate consumption benefits and output that yields future benefits; (5) measurement of the way in which comprehensively defined social and economic output is distributed among the population.

Human Capital Returns

Economic accounting systems have fallen well behind the development of economic concepts of human capital and its deployment in activities that produce real income. Our accounts recognize that human capital produces an output equal to its rental price when it is purchased in the market. Even here, it is implicitly assumed that output consists only of the labor contribution to the firms' current production of goods or services, and thus that the stock of human capital is not altered either by the learning, aging, or obsolescence involved in its use.[2]

However, formal job market activities are only a fraction of the total income-producing activities that use inputs of human capital. Going to school is not conceptually different from going to work; raising children at home is not conceptually different from teaching nursery school; serving on an unpaid public board of directors is not conceptually different from serving on a private board for compensation; keeping house for one's family is not conceptually different from keeping house for other people's families; and rejecting an opportunity to work overtime or to moonlight in order to enjoy more leisure time sounds as if it should increase real income rather than reduce it.

Measuring the contribution of nonmarket activities to social and economic output is a problem that has become of considerable interest to social accountants. In one sense, the measurement problem is trivial: If one is willing to assume that all human activities are adjusted so as to yield equalized returns at the margin, all nonmarket activities can be valued at market wage rates and the gross output of human capital is simply the hourly wage rate times 24.

There is much to be said for this view, both theoretically and empirically. Individuals can generally allocate their time between market and nonmarket activities as they see fit, and families can do so with even greater ease than individuals because they have more degrees of freedom. Families can and do choose to substitute purchased services (housekeeping, child care, home maintenance, etc.) for inputs of their own time, and the higher the value of time as measured by market wage rates the more likely the substitution. Similarly, families can and do substitute market earnings for leisure or nonmarket activities, either by varying hours of work via overtime and multiple-job holdings or by moving into or out of the labor force as the demands on time spent in the household vary over the life cycle.

A good first approximation, therefore, is that time is allocated at the margin so as to equalize its value in all activities, given the market wage rate. It is possible that efforts at quantification can make little further headway than this. There are, however, some areas that appear to be promising candidates for additional research.

First, imputation of the marginal wage rate to all activities, while conceptually able to resolve the measurement of gross output, cannot provide an estimate of net output; the latter is clearly the more useful measure. Some amount of time must be allocated to activities that do no more than maintain the stock of human capital intact. Eating, sleeping, and perhaps leisure activities fall into this category. In principle, any allocation of time essential to the maintenance of market productivity should be considered as an intermediate rather than final product.

A thorny and perhaps insoluble question of purpose and motivation seems to be involved. If the President needs to play golf in order to continue presidenting at the same level of efficiency, then golf-time is not a final consumption good but simply a capital maintenance activity. If, however, presidential golf-time is really thought of (by the President) as final consumption, is it worth more than my golf-time simply because the President's market wage rate is higher than mine? Tentatively, I am inclined to argue that "pure" leisure is not worth more to one person than another unless there are differences in the efficiency with which "leisure" is pursued,[3] but that other nonmarket activities of a more purposeful

[2] The effects of learning (on-the-job training, to use Mincer's terminology) are heavily concentrated in the early years of labor force participation and appear to be of greater importance for those with more formal schooling. The effects of obsolescence are more heavily concentrated in the later years. Thus the aggregate net effect depends on the age (and schooling) distribution of the population, on the rate of technical change and on longevity.

[3] For example, a high-wage person would presumably get more benefits than a low-wage one from a given amount of chronological time devoted to playing golf: his golf equipment would be more efficient, he might use a time-saving golf cart or an energy-saving caddy, he might belong to a private club that enables him to play exactly when he wished, etc.

sort are appropriately valued at market wage rates and are therefore likely to be of differential value depending on who is engaged in the activity.

A second question concerns the allocation of nonmarket time between present and future benefit flows. The general principle seems clear enough. If one spends his evenings at home studying for a CPA examination, the time input represents an investment in future income. If time is spent teaching one's children how to recognize letters or how to resolve social conflicts with peers, the time input represents an investment in the child's future real income.

A third question concerns the efficiency of time use. Economists have traditionally viewed time as the ultimate allocative constraint: labor or capital can always be increased or augmented by training or technical change, and natural resources can always be discovered or manufactured in the laboratory. But all these things must be accomplished within the constraint of a twenty-four hour day. As a physical fact, the time constraint is tautologically valid. As an input into the usual household or firm production functions, the time constraint can be stretched: One can, and many do, commute to work or to a professional meeting in Detroit and at the same time enjoy the leisure-time activity of listening to the radio or the productive activity of finishing up a professional paper. One can combine eating dinner with the consumption activity of conversing with adults or the investment activity of training one's children. With a long enough extension cord, an efficient housewife can spend an hour setting up the next night's PTA meeting and also wash the dishes, prepare dinner, and write a few Christmas cards. In short, the right time units are efficiency hours and not calendar units; there are a great many people who manage to get thirty hours worth of activities out of a "standard" twenty-four hour day, and others who manage to get only twelve.

The last question is tangentially related to the third. Traditional theory argued that the labor-supply function had an upward slope because a higher real wage was needed to overcome the increased disutility from additional work. More modern theory has discarded the idea of disutility and put in its place the marginal rate of substitution between work and leisure—along with the proposition that leisure is, probably, a superior good. If by work one means turning screws on the assembly line, washing dishes in a restaurant, or picking lettuce in the San Fernando Valley, work probably does have net disutility, although conditions of work have probably lessened that disutility over time.[4] Moreover, occupations of this type have surely been declining as a proportion of total labor force activity.

Market wage rates are, of course, associated with any net disutility or utility (let us say consumption benefits) associated with work; other things equal, the higher the consumption benefits, the lower the market wage. We have no (presently) observable measure of these benefits, however, since the "other things" are plainly not equal between occupations. Not only are disagreeable and uninteresting jobs apt to require little training and minimal levels of ability, but the consumption benefits associated with any given occupation presumably vary among people. Hence free occupational choice will tend to maximize the consumption benefits associated with work in any given occupation. One suspects, nonetheless, that there are marked differentials in consumption benefits among occupations, that the amount of benefit is positively rather than negatively associated with the level of market wage rates, and that the aggregate amount of consumption benefit involved in market work has been increasing over time. It may well be the case that the community as a whole now derives net benefits from work, whereas formerly net negative benefits were the dominant characteristic of labor market activity.

---

[4] Restaurants can be air conditioned and supplied with mechanical dishwashers, steel mills use vastly more mechanization and have more work relief crews, etc. Many of these changes are measurable, at least in terms of their cost to employers.

Environmental Assets and Returns

One of the oldest questions posed by social accountants is the proper treatment of activities, mainly but not entirely governmental, designed primarily to prevent a reduction in social or economic welfare, e.g., the use of resources for national defense purposes. During the Second World War the United States devoted close to half of its total resources to military purposes: Was this net output or simply a cost of maintaining the sociopolitical environment—a "regrettable necessity," to use the prevailing terminology.

In principle, it is hard to see why the costs of maintaining any "given" environment, whether it be sociopolitical or physical, should be viewed as a part of net output. A country that spends part of its resources to maintain a military establishment for defense against actual or potential enemies is less well off than one that uses less of its resources in this way, other things equal. And a shift in the political stability of the world community that results in the need, real or imagined, for an expansion of military expenditures from 10 to 20 per cent of total output has clearly resulted in a loss of social and economic well-being for the entire community.

It is not, of course, only military outlays that fit this category. A community that needs to spend more resources on policemen, firemen, burglar alarms, safety locks, night watchmen, etc. is clearly worse off than one in which these outlays are not necessary. No one buys police or fire protection, or hires night watchmen, because these services are desired per se. If there were no crime or fires, and no risk of either, there would be no expenditure on crime or fire prevention and everyone would be better off.

And the relevant class of activities extends far beyond the national or personal security outlays discussed here. For example, resources used for medical care are largely in the same category. Few people go to hospitals because they enjoy the rest and the good food!

The analytically appropriate treatment is to view the sociopolitical environment as an asset that may yield direct consumption benefits in and of itself, and that permits productive activities to be carried on without interference. Like any asset, the sociopolitical environment can deteriorate or depreciate. Expenditures required to "maintain the asset intact" constitute gross but not net output of the system. Thus wars, crimes, and fires are specific manifestations (costs) of environmental deterioration, while resources spent to suppress these manifestations must be presumed to have enabled environmental assets to be better maintained than in their absence. Further, depreciation of the sociopolitical environment can be estimated as the *sum* of two components: first, costs imposed by the amount of deterioration that has been permitted to occur, as represented by losses resulting from crimes, fires, wars, etc.; second, costs incurred to maintain the asset at its given level, as represented by expenditures on policemen, firemen, members of the Armed Forces, etc. In the absence of these maintenance expenditures, or in the event of their reduction, it must be presumed that the asset would deteriorate further and that the costs represented by specific manifestations of deterioration would thus increase. Optimum social policy, of course, consists of equating at the margin the cost functions associated with these two activities.

The appropriate treatment of the much-discussed subject of pollution of the physical environment should now be evident. A community starts off with a stock of such assets—air and water of a certain degree of purity, roads that are free of abandoned cars, playgrounds and streets free from discarded newspapers, broken bottles, etc. As a, perhaps inevitable, part of the process of industrialization and economic growth, these environmental assets tend to deteriorate or depreciate, thus reducing the flow of benefits from enviromental assets. Expenditures designed to slow down or reduce deterioration are clearly costs associated with the maintenance of the asset rather than an output of the system. As with the sociopolitical environment discussed earlier, the full cost of deterioration is the sum of the reduced yield on the asset plus any costs incurred to prevent even greater deterioration. Alternatively, one could view industrialization and economic growth as producing a series of dis-products and dis-services—various kinds of impurities and undesired products introduced into the physical environment and left there. In the absence of expenditures designed to reduce environmental deterioration, real net output is decreased by the negative value of these dis-products and dis-services.

There is an interesting difference between types of environmental deterioration cases that superficially appear to be similar. Compare the situation in which real or imagined needs for defense cause a country to use x per cent of its resources for military purposes with one in which deterioration of the physical environment causes the country to use the same fraction of resources to control pollution. In the latter case, there is a strong presumption that deterioration of the environment is a direct consequence of the normal functioning and growth of the economy. If so, the flow of benefits from economic growth is clearly overstated unless allowance is made for the negative byproducts of growth.

In the former case, however, it is far from clear that the environmental maintenance costs represented by the need for a large defense establishment is a result of the normal functioning and growth of the system. One could conceive of circumstances in which that might be the case; e.g., an aggressor nation building up its military strength in order to overcome other countries and thence derive future economic benefits. But in general the causality is unclear.

If defense outlays are basically unrelated to the functioning of the socio-economic system and simply reflect exogenous events, should one "penalize" the system by treating such outlays as environmental maintenance costs? If the objective is to measure changes in social and economic welfare, it seems that the answer should be yes. Resources used for defense cannot be used elsewhere—and I cannot see that it matters *for purposes of measuring* whether defense welfare needs are a cause of one's own actions, are real but exogenous to one's actions, or are wholly imaginary. It does, however, make a great deal of difference *for the analysis of policy alternatives* whether or not the system has created its own defense needs. If this is the case, there is a hidden cost to a change in policy that increases the optimum size of the defense establishment, just as there is a hidden cost to a growth policy that produces deterioration in the physical environment as an inevitable byproduct.

If defense needs are unrelated to economic and social policy, however, the appropriate analogy is to phenomena like earthquakes, floods, and other natural disasters: welfare is willy-nilly reduced, and there is nothing that can be done about it. But the reduction is real and needs to be registered in the accounts.

Public Sector Performance

As with regrettable necessities, social accountants have long been concerned about the problems associated with the measurement of output in the public sector. Our social accounting conventions essentially say that public services are worth their cost. The rapid growth of resources devoted to services generally, and especially to those services for which no output measure is available, has created a renewed interest in these problems. It is not much of an exaggeration to say that, unless we can devise meaningful measures of output or performance in the public service sectors of the economy, and unless we can devise incentives to improve productivity as reflected by those performance measures, the United States may be faced with an increasingly serious drag on the rate of over-all economic growth and of social and economic welfare.

The output measurement problems in the public sector can be illustrated by the education industry. The provision of educational services is largely a public sector activity, but by no means exclusively so. Hence there is a market test for output, of the same kind as is applicable to the production of wheat or steel or automobiles. People can choose to send their children to private profit-making schools, to private not-for-profit schools, or to public institutions. However, neither public nor private institutions can really specify what parents are buying when they send their children to school, and the "managers" of the "industry" would be hard put to say what it is that they are producing, whether they are producing it as efficiently as possible, or the technical characteristics of their production functions. In contrast, the producers of wheat, steel, and automobiles certainly know which production techniques are efficient and which are not, and by how much output will change if more of one input is used and less of another. Further, what they are producing is a product with specified characteristics, and consumers are free to accept or reject the product depending on the value placed on these characteristics. Hence the problem is only partly a public sector problem and is more appropriately described as a subclass of quality change problems.

In principle, what one would like to measure is reasonably clear: schools take youngsters with a given amount of innate ability, a given family background, given motivation, etc. and expose them to several years worth of instruction in the development of basic skills, the acquisition of factual knowledge, the techniques of problem solving, and so forth. The output of the school system is presumably the total *gain* in achievement accomplished by all students. Even if we could not translate gains in schooling achievement into their ultimate output—their influence on adult behavior and performance—it would still be of considerable value for improving efficiency in the education industry if we could simply measure the gains in achievement attributable to schooling per se and not to other inputs.

The type of problem encountered in measuring output in the education industry is found in virtually all of the public service sectors and in many private sectors as well. For example, we do not measure the output of the health industry by what is achieved in reducing work-loss time via a lesser incidence of disabling illness, nor by its effect in increasing expected lifetime; we do not measure the value of police or fire protection by the number of crimes prevented or the value of lives and property saved; and so on. For many if not most of these activities, we know in principle what the proper output measure is, although in many cases the output is multifaceted and not easy to quantify. But it is much easier to measure the *costs* of providing these services than to measure the *results* of providing them, and we are generally driven to settle for the former. One of the high priority items on the research agenda for social and economic measurement during the next decade will be to start providing true output measures for all of the areas in which we now rest more or less content with input measures.

Present Versus Future Output

Our present accounts divide total output into consumption and investment components that clearly fail to coincide with an analytically useful separation of the two. Social accountants have been working on this problem for several years now, and recent studies by Kendrick and the Ruggles[5] have provided quantitative estimates for a much more comprehensive measure of investment.

One area that has not yet been examined for its investment-consumption implication was touched on earlier—the division of nonmarket activities between those geared to the provision of present benefits and those geared to the future. Some of the important questions here concern investments in learning or training in the labor market, which can be either positive or negative, and parental investments in preschool and school-age children. All these are essentially problems of esti-

---

[5] See Richard and Nancy Ruggles, *The Design of Economic Accounts*, New York, NBER, 1970. Kendrick's work is still in manuscript form.

mating investments in human capital, which take place both in and out of the labor market and both during and before formal schooling.

The Distribution of Economic Welfare

All of the problems discussed above have been concerned with the measurement of aggregate social and economic welfare. On any research agenda this is the most useful place to start, for several reasons. First, unless one can measure the aggregate one probably cannot measure anything at all. Second, policy priorities should probably be geared to the size of the aggregate gains or losses involved: to spell out the obvious, if deterioration of the physical environment results in a welfare loss of $5 billion per year, it is hard to justify spending $10 billion per year to prevent that loss.

While the aggregate measure thus has the higher priority, the distributional questions are important and warrant serious consideration. This is especially so in view of the fact that the existing evidence on income distribution is based almost entirely on income as measured by the National Income and Product Accounts. An augmented measure of income would differ drastically from the present one, and there is no presumption that the distribution of broadly defined economic welfare looks anything like the distribution of narrowly defined money income, nor that changes over time in the distribution of welfare are similar to changes over time in the distribution of money income.

A case can be made for the proposition that many of the flows that would enter a broadly defined measure of welfare, but are excluded from our present narrowly defined measure of money income, tend to be distributed less equally than money income, hence that economic welfare may be less evenly distributed than money income. And it is distinctly possible that the inequality in welfare may have been increasing over time. Some of the major differences between a broad economic welfare measure and a money income measure are discussed below, with comments about their possible distributional characteristics.

1. Although it has not been discussed above, a major exclusion from money income as measured in the Income and Product Accounts is the flow of capital gains. These certainly warrant inclusion in a measure of the way in which income is distributed over any reasonably long time period: capital gains are obviously distributed very unequally, presumably more like the wealth distribution than the income distribution.[6]

2. The distribution of nonmarket output, as reflected by the allocation of nonmarket time, might have relatively little impact on income inequality. First, hours of labor force time tend to be positively correlated with total earnings, hence people with higher earnings will tend to have less available nonmarket time. But second, differences in the efficiency of time use among the population are likely to be positively correlated with differences in the level of education and earnings: thus differences in the distribution of time measured in efficiency units are probably an offset, and perhaps a more than complete offset, to differences in the distribution of physical time.

3. The incidence of environmental costs and benefits is not obvious. For the sociopolitical environment, the costs of resources used to maintain the asset are distributed in accordance with the burden of taxation; the benefits could plausibly be distributed either in terms of wealth or in terms of income. If the benefits are distributed by wealth, income inequality is increased by the burden of maintaining these assets, while if the benefits are distributed by income there is probably not much impact.

For the physical environment, where we have so far chosen to absorb the losses

[6] Some empirical work on this subject is contained in an unpublished Ph.D. dissertation by Michael McElroy.

rather than pay the maintenance costs, the burden is probably larger on low-income urban families, and hence the welfare distribution is pushed towards less equality.

4. The inclusion in welfare of any net consumption benefits from work would probably tend to increase inequality, since there is almost certainly a strong positive correlation between money income and net consumption benefits from work. In fact, the consumption benefits are likely to be positive for high-income families and negative for low-income families.

Adding up all of these effects with some kind of rough implicit weights suggests that social and economic welfare is less equally distributed than money income, and possibly a great deal less so. Moreover, most of the trends over time in the importance of presently unmeasured aspects of welfare suggests that inequality may have been growing rather than decreasing. But these are very rough and largely intuitive guesses about magnitudes and probable incidence, and do no more than suggest a possibility that the rising incidence of social and economic discontent might be in part due to an increasingly less equal distribution of the output produced by the social and economic system.

RESEARCH STRATEGIES

Nonmarket Activities

The one indispensable ingredient for estimates of the level and distribution of nonmarket outputs seems self-evident: we need a carefully designed and large-scale study of time allocation among American families. A number of such studies have been conducted over the past years, although none seem to have the level of analytical detail required to fill in the kinds of estimates discussed above. We probably know enough at this point to be able to design such a study, although whether it can actually be carried out successfully and whether it can be financed are both open questions. As to the former, the level of required detail seems to suggest that we need some means of observing what people do virtually every minute of the day. The usual procedure is for interviewers to ask respondents what they were doing today or yesterday or last week, but that seems insufficient for the kind of activity-specific detail required by the analytical uses to be served. Having interviewers follow people around with a little notebook is certainly feasible, but it raises a serious question as to whether activities have been changed by the fact of their being observed—a problem that is unique to investigation in the social sciences. Ideally, something like closed-circuit television (without the unit being aware of the observation) seems like the kind of measurement technique needed. I do not make that suggestion seriously, but simply use it to illustrate the kind of observational environment that might be needed.

Environmental Asset Yields

There are two general ways in which one might estimate the welfare loss from environmental deterioration. One is to observe the specific physical consequences of environmental change, then translate these physical consequences into measures of social and economic cost. The appropriate cost measure is presumably the difference between the benefits yielded by activities that were carried out before the environmental change and the benefits yielded by the altered pattern of activities that result from the change. Measurement of the physical consequences (tons of soot per week) is relatively straightforward, at least compared with the measurement of the social and economic costs. Substitution of alternative activities and tradeoffs of one resource use against another is the rule rather than the exception in economic behavior, and we need to measure the net influence of all the adjustments that will be made to a change in environmental conditions.

For example, one solution to a heavy concentration of air pollutants in urban areas would be the imposition of sufficiently heavy taxes on the use of offending facilities to encourage investment in pollution control equipment. The optimal tax would be one that balances, at the margin, benefits (the reduced welfare loss associated with lower contamination levels) against costs (the investment in equipment). But neither the welfare loss from contamination nor the cost of lowering contamination levels is likely to be a simple linear function of contamination level.

Moreover, investment in pollution control equipment is only one of a number of ways in which welfare losses can be reduced. An alternative might be investment in transportation designed to facilitate a greater spatial separation of working and living areas and to encourage greater specialization of working versus living areas. This program would also reduce the welfare loss: the costs associated with any given volume of urban pollutants would be lowered because fewer consumption activities would be located in urban areas. In short, the creation of geographic areas with specialized uses might permit the simultaneous existence of high contamination levels in specified areas and relatively little welfare loss.

An interesting alternative strategy is to assume that the market has already done the measurements for us, via the generation of compensating income differentials. If the incidence of environmental deterioration is uneven across geographic units, if labor markets allocate resources in such a way as to equalize real wages at the margin, and if the optimum technology is a function of density, areas with a greater amount of environmental disamenities (to use the phrase employed by Tobin and Nordhaus in a recent paper which used this technique)[7] would have to offer a higher equilibrium wage to given quality labor. Given these assumptions, the real wage differential becomes a measure of the welfare loss associated with environmental deterioration. More precisely, it becomes a measure of the loss from structural shifts in location towards areas with greater amounts of environmental disamenities, but it cannot provide an estimate of the absolute (base) level of disamenity.

While there are serious problems with this approach, primarily in specifying the estimating equation, it seemed promising enough to warrant further work. One of its chief advantages is that it provides a global estimate of the net influence of all kinds of environmental deterioration, and thus provides a "ball park" estimate of the total social loss involved.

Income Distribution

Estimates of the distributional incidence of economic welfare must be based on microdata. Since an enormous amount of information will ultimately be required, and since the cost of new data is so high, one obvious solution is to tap existing data sets and attempt to synthesize information obtained from different households. For purposes of measuring social and economic welfare, the use of synthetically created households designed to represent different household types in the population does not present any problem, and additional microdata can easily be added to existing information provided the relevant household characteristics are known. Thus one research focus is to create a set of synthetic households with specified characteristics, for whom virtually all of the desired real income variables are known or can be obtained. There is no particular reason why this cannot be done, although it may turn out that a good many of the variables needed to measure welfare are not contained in any existing microdata set and therefore have to be obtained from new samples. There is also, of course, a major problem of concocting variables with which one can measure the various aspects of welfare deemed to be important.

---

[7] William Nordhaus and James Tobin, "Is Growth Obsolete?" paper prepared for Colloquium on Economic Growth, NBER Fiftieth Anniversary, December 10, 1970.

NEW PLANS AND RESULTS OF RESEARCH IN ECONOMIC FORECASTING—Victor Zarnowitz

ON MEASURING THE CONTRIBUTION AND PERFORMANCE OF FORECASTING METHODS

Three Approaches to the Evaluation of Forecasting Procedures

Recent research in economic forecasting has achieved significant success in developing criteria of predictive accuracy and applying them to new collections of diverse and interesting data.[1] It also demonstrated that useful information about the implicit structure of forecasts for which no explicit models are available can be distilled from comparisons of such forecasts with benchmark predictions derived by a known, replicable method [14] [13]. Further studies in this area should focus on the question of how predictive accuracy is related to the method and structure of forecasts.

Most economic forecasts are made to facilitate and improve business or economic policy decisions. In principle, then, the success of such forecasts should be measured by the degree to which they reduce errors in these decisions, allowing for any tradeoffs between the effects of the different errors and between the gains from better prediction and the time and dollar costs of forecast improvement. Full application of this criterion, however, poses high informational requirements, namely:

1. That the forecasts be all verifiable and their errors actually ascertained.
2. That the preferences of the decision maker and the constraints under which he operates be so specified as to permit evaluation of the consequences of forecasting errors in the context of the relevant "loss function."
3. That the effective costs of producing the forecasts be known or adequately estimated.

Typically, information on (1) and (2) is not available to an outside analyst. The costs of developing good quantitative data of this sort should vary with the complexity of the underlying problems, but they are probably often so large, even to the decision makers and forecasters who are directly involved, that the information may well be altogether incomplete. Studies of economic forecasting, therefore, must seek to advance as well as possible with little direct knowledge of costs and returns on types of forecasting activity.[2] It seems natural to view the accuracy of forecasts as being the single most important measurable aspect of their quality, even though it is recognized that knowledge of the *size* of forecasting errors may not be adequate to determine the *consequences* of forecasting errors for decisions based on the forecasts.

Forecasts of economic conditions that are actually proffered for business and public policy uses are generally the products of judgmental combinations of various techniques. This is so partly because forecasters face different problems which are believed to require different approaches; partly because they distrust relying on any single method; and partly because they wish to take account of information

NOTE: I am indebted to Robert C. Coates, James C. Ellert, and Mrs. Josephine Su for efficient help with the statistical work.

[1] Comprehensive work in this area was done at the National Bureau of Economic Research [19] [16] [6] [12] [3] [5] [23] [24]. Major contributions made elsewhere include those by Theil at the Econometric Institute of the Netherlands School of Economics, as well as by Christ, Klein, Okun, Suits, Stekler, and others. Space is not adequate here to give even a very incomplete listing of these works. For a fairly large but still far from exhaustive bibliography, see [20, pp. 437-439].

[2] The economic concept of a rate-of-return criterion of forecast appraisal remains important, however. It implies first the insistence on point (1) in the text above, which provides the principle for the selection of forecasts to be analyzed. Second, it suggests that the choice of the measure of average forecast accuracy be linked to the assumed form of the loss function (e.g., if a quadratic loss criterion is adopted, the appropriate measure is the root mean square error; see [18]). For a discussion and application of several alternative error-cost functions, see [15, Chap. IX].

that does not fit into any preconceived model (e.g., the latest news about such exogenous or "random" events as strikes, wars, elections, etc., or "autonomous" expectations concerning developments of this kind). Thus, predictions of the economy's future course are seldom derived by methods that are fully specified, explicit, and reproducible. How, then, can one go about trying to answer the high-priority question about the relative efficiency of forecasting procedures?

The problem can be attacked in different ways, mainly directly or mainly indirectly, depending on the type of forecasts involved and the amount of information about them. I can perceive three workable approaches which are complementary rather than substitutive: (1) analysis of ex ante and ex post forecasts with econometric models; (2) comparison of authentic forecasts, which do not have a fully specified structure, with constructed benchmark forecasts, which do; and (3) study of surveys which include information about the procedures that the respondents use in deriving their reported predictions. The three sections that follow discuss these approaches briefly in the above order.

Sources of Errors in Econometric Model Forecasts

An econometric model is not, as a rule, sufficient for making predictions: The values of the exogenous variables derive equally from information from outside the model and from the forecaster's judgment. Typically, the interaction between the forecaster and the model is much more extensive and complex than that, however, since various judgmental adjustments of the constant terms (and possibly other parameter estimates) of the model are frequently carried out in the process of deriving the predictions. Rough estimates of the joint effects on the predictions of both such adjustments and the errors in the exogenous inputs can be obtained by comparing the accuracy of the reported ex ante forecasts and of the ex post forecasts which the model would have produced in the absence of any modifications of its stated structure and given the correct historical values of the exogenous variables. If full information were available on the adjustments that were actually made, considerably more could be done to decompose the over-all error of the ex ante forecasts into the parts attributable to (1) the model itself, i.e., its basic specifications and method of estimation; (2) the forecaster's own predictions of the exogenous variables; (3) his decisions to alter the constants in the model or to assign values other than zero, the expected value, to some disturbance terms; (4) any other sources of error including deficiencies in the data used; and (5) the interaction among these different types of error. To be sure, even with all the information about the model, data, and procedures that could conceivably be assembled, the present-day evaluation techniques would still not provide a complete anatomy of the errors of forecasts with econometric models.[3] But the potential gains from such assessments are probably large, and much remains to be done along these lines both to compile and to explore the record of econometric forecasting.

---

[3] Of the several reasons for this, the difficulty of testing the specification of the model appears to be particularly serious. Economic theory helps mainly qualitatively, by indicating what variables to choose and what signs of coefficients to expect, but it typically gives little information about the size of the parameters, the precise form of the relationships (nonlinearities), the lag structure, etc. This leaves much room for sample experimentation, which, however, may improve the fit of the model to the historical data rather than its predictive ability. Proper tests of the forecasting performance are therefore essential, but here again serious problems arise. If deviations from the structural continuity of the economy are substantial, a model may describe the historical experience well and yet err considerably in prediction. It is then difficult to distinguish between tests of predictive performance and tests of structural stability. There is the further complication that the exact distribution properties of estimators for small samples are not established—and the samples of observations from economic time series used in the econometric models are usually rather small. Also, the argument is made that the tests should be designed "in a true ex ante framework" (allowing fully for the interaction between the model and the model-builder acting as forecaster), but it is not clear how, precisely, this should be done. See [4] and [10].

Benchmark Models of Basic Forecasting Methods

The forecasting devices that are currently most relevant can be classified broadly as (1) extrapolative techniques; (2) surveys of intentions or anticipations by economic decision-making units; (3) business cycle indicators; and (4) econometric models. Each of the four methods provides materials that are widely used as ingredients in forecasting the course of the economy, although the ways of combining such diverse data vary greatly, as do the weights which are in effect assigned to the underlying procedures. Each of the methods can also be employed alone: consider projections of historical values of the series to be predicted, from the simplest "naive" models to the technically sophisticated representations of autoregressive and moving average processes; opinion polls or aggregations of the expectations of individual businessmen or consumers; strictly mechanical uses of indicators, say, in regressions of lagging or coincident indicator series on the leading indicators; and similar applications of reduced forms of predictive econometric models. Such pure or extreme forms are generally not presented as forecasts proper, but they often serve as benchmarks of predicting performance. For example, authentic forecasts of, say, GNP are compared with outputs of various extrapolative formulas applied to the past values of GNP; business expenditures on plant and equipment, as predicted from a regression equation with past values of selected explanatory variables, are compared with aggregated business anticipations of such outlays; judgmental forecasts of industrial production by a business economist are compared with predictions based on a lagged regression of production on new orders; and ex ante forecasts of any of these variables obtained by a team of econometricians armed with a large-scale model of the economy are compared with the corresponding ex post forecasts.[4] However, at present such comparisons tend to be used in limited and haphazard ways, chiefly as screening devices. The research strategy here proposed is to confront bona fide forecasts from various sources with a full array of these benchmark predictions so that as much information as possible can be extracted from the analysis about the relative accuracy, not only of the forecasts proper, but also, more importantly, of the forecasting methods. Furthermore, in doing so we should be gaining some knowledge of the *structure* of the forecasts proper which, as already noted, is for the most part (in particular, for all the "judgmental" business forecasts) unknown and not directly observable.

Surveys of Forecasts

Comprehensive, periodic surveys of economic forecasts offer another program for the comparative analysis of forecasting procedures: The participants can be asked to specify their methods, assumptions, etc.; their predictions can be grouped according to the resulting replies; and the forecasts in the different groups can then be studied and perhaps ranked by some criteria of relative performance. Notable among the evident advantages of this direct approach are that the forecasters themselves tell us about the methods they use and that this original information relates to the collected forecast data. It should certainly be possible to obtain valuable qualitative information in this way, although the process is time-consuming and may be relatively costly; but the real limitation here is that surveys are not apt to furnish *quantitative* data on the relative importance of the different forecasting techniques.[5] Moreover, one would not expect this approach

[4] It will be noted that the last item listed in this sentence refers to approach (1), discussed in the previous section. Approach (2), which is the subject of the present section, can in this sense be viewed as including elements of (1) as a special case.

[5] Consultations with business economists and study of forecast data and the literature (see, e.g., [2]) all leave little doubt that many forecasters would find it quite difficult and costly—perhaps altogether impractical—to quantify such information. Forecasters will, however, *rank* the principal forecasting methods according to the relative importance of these methods in their work. (Data of this sort are discussed in the next-to-last section of this paper.)

to give conclusive results for any sample in which either the forecasters failed to show significant differences in the choice of methods or the forecasts failed to vary significantly in their average accuracy.

Some forecasting procedures are designed mainly to deal with particular types of economic events or movements: thus, selected indicator series are used to predict business cycle turning points and, for certain variables, extrapolations of past trends are used to project future trends. It is therefore possible that such methods serve the forecasters much better at certain times than at others; e.g., an indicator "benchmark model" might prove most useful at revivals and recessions, while an extrapolative model might come out ahead in periods of continued growth. Periodic surveys of forecasts that include information about the methods used by each forecaster offer a way of testing this possibility. If the surveys were taken frequently enough over a sufficiently long and varied stretch of time, separate comparisons of forecasts based on different procedures could be made for periods with different characteristics such as short and protracted expansions, turning points, and recessions. This is clearly a potentially important application of a "cross-sectional" analysis of forecast data, and should complement the time series analysis, which is concerned with the over-all accuracy, bias, efficiency, and consistency of forecasts.

Further Steps

The preceding sections give merely a broad outline of some research undertakings which, I submit, are needed and promising but also rather ambitious. To elaborate a little, we are interested in the relative strengths and weaknesses of *specific* models and procedures. Evidently the models that are being used for macroeconomic forecasting, simulation, and analysis differ greatly in size, structure, and statistical implementation, although some of them overlap others and some are more interesting than others. Studies of the comparative performance of major econometric models can promote understanding of the role of such differences and improve ability to discriminate among alternative formulations of economic relationships. Similar observations about the need for specificity and diversity apply, *mutatis mutandis*, to the work on other forecasting procedures (extrapolations, anticipation surveys, indicators).

The suggested agenda for future work on economic forecasts and forecasting procedures builds on past and current research at the National Bureau and elsewhere. The new project is still in a very early stage, which as usual involves considerable uncertainty and experimentation. Nevertheless, some of the results are worth discussing now, and I shall draw on them in what follows to illustrate how our research plans are being implemented.

ON THE ACCURACY AND PROPERTIES OF FORECASTS WITH ECONOMETRIC MODELS

Ex Ante and Ex Post Forecast Errors

A novel and informative analysis of the record of predictions made with the aid of two econometric models, Wharton-EFU and OBE, is offered in [5]. The data used in Table I-2 were obtained through the courtesy of Haitovsky and Treyz from [9], an extended and updated version of the earlier work in [5].

Table I-2 shows selected measures relating to the accuracy of the Wharton and OBE model forecasts for GNP, namely, ratios of mean absolute errors of corresponding ex post *(XP)* and ex ante *(XA)* predictions and *XP* alone. The forecasts cover spans of time that vary from one to four quarters; the measures are listed separately for each span. The first two columns of the table are based on the errors of the authentic ex ante forecasts which incorporate the model builders' original judgmental adjustments *(OR)* of the constant terms; exactly the same *OR* adjustments were also applied in the computation of the mean absolute errors

TABLE I-2

Wharton and OBE Model Ex Post Forecasts of GNP over Spans of from One to Four Quarters, Comparisons with Ex Ante Forecasts With and Without Adjustments of Constant Terms

Span of Forecast (quarters)	Measures Relating to GNP Forecasts					
	With Original Judgmental Adjustments (*OR*)		With Mechanical Adjustments (*AR* or *GG*)[a]		Without Adjustments (*NO*)	
	XP[b] (1)	*XA/XP*[c] (2)	*XP*[b] (3)	*XA/XP*[c] (4)	*XP*[b] (5)	*XA/XP*[c] (6)
	Wharton Model, 1966 III-1969 III					
One	5.97	0.63	11.02	0.87	27.32	1.05
Two	10.05	0.78	17.12	0.90	35.26	1.08
Three	11.65	0.89	20.79	0.99	26.34	1.15
Four	14.20	0.89	22.36	1.07	24.98	1.10
Next year	8.68	0.76	13.78	1.04	23.82	1.17
	OBE Model, 1967 II-1969 III					
One	4.35	0.49	4.58	1.03	6.44	0.97
Two	8.76	0.89	9.26	1.12	13.78	0.99
Three	13.65	0.93	15.61	1.04	18.75	0.94
Four	19.37	1.01	22.43	1.00	23.54	1.04
Next year	11.18	1.03	13.29	1.13	16.65	1.05

[a] *AR* adjustments use the simple average of the last two observed residuals, $(e_t + e_{t-1})/2$. *GG* adjustments use ρ^T $(e_t + \rho e_{t-1})/2$, where ρ is the estimated first-order autocorrelation coefficient of the residual errors and T is the span of the forecast. See text and footnote 6.

[b] XP = mean absolute error (*MAE*) of the ex post forecasts of gross national product (GNP), in billions of current dollars.

[c] XA/XP = ratio of *MAE* of the ex ante forecasts of GNP to the corresponding *MAE* of the ex post forecast (*XP*).

of the ex post forecasts. Columns 3 and 4 relate to forecasts with two types of mechanical adjustment, *AR* or *GG*, described in footnote a of the table.[6] Whichever adjustment gives, on balance, lower errors in the given case is used, but always the same adjustment is applied to the corresponding ex ante and ex post predictions. Finally, in columns 5 and 6, the *XP* and *XA* measures are compared for the forecasts made without any adjustments of the constant terms (*NO*).

Reading down the *XP* columns, one finds the usual tendency of the average errors to increase with the length of the predictive span; the errors rise less than in proportion to the length, however.[7] The same applies to the corresponding summary measures of error for ex ante forecasts (*XA*). Hence the ratios *XA/XP* neither increase nor decrease systematically as the forecast span lengthens.

The original judgmental adjustments tend to result in lower average errors than the mechanical adjustments, and the ex post forecasts with no adjustments of the constant terms have by far the largest errors. That is, typically, $XP - OR < (XP - AR,\ XP - GG) < XP - NO$, as can be seen by comparing columns 1, 3, and 5. Again the same applies to the errors of the ex ante forecasts, as summarized in the *XA* measures. Indeed, the adjustments, particularly the judgmental ones,

[6] The *AR* procedure is characteristic of the type of mechanical adjustment often used by the Wharton forecasting team. This adjustment does not vary with the span of prediction. The *GG* procedure was developed by George Green in adaptation of the form of adjustment proposed in [7] and applied to the OBE model; see [8]. For the formulas used, see Table I-2, footnote a.

[7] In other words, errors cumulate over the successive quarterly periods within a span, but most often (although not very regularly) by decreasing increments. This behavior has been found in a variety of forecasts (see [19, Chap. 5]). Forecasts for one year ahead (column 5) can be viewed as referring to the average levels in the next four quarters and thus as having effective spans of approximately 2½ quarters.

reduce the XA errors even more than the XP ones. (Note that the XA/XP ratios in column 2 are, with rather few exceptions, smaller than their counterparts in columns 4 and 6.) This underscores the success of the *OR* adjustments which, of course, originate in the efforts of the econometric forecasters to improve their ex ante predictions.

The errors of the ex ante forecasts are on the average smaller than those of the ex post forecasts with the same judgmental adjustments: Most of the XA/XP ratios in column 2 are less than 1. With mechanical adjustments or with no adjustments at all, $XA > XP$ in most cases, but the deviations of the ratios from unity are here generally smaller (columns 4 and 6).

For most other variables, too, the XA/XP ratios are predominantly less than 1 (often by large margins) when the *OR* adjustments are applied. Measures for a collection of eight important variables[8] show that $XA < XP$ in nearly half the cases when the mechanical adjustments are used and in a smaller but still appreciable proportion of cases even when no adjustments at all are made. These general findings apply to both the Wharton and the OBE models, but the tendencies for the ex ante forecasts to be more accurate than the corresponding ex post forecasts, and for the *OR* adjustments to reduce the errors, are on the whole weaker for the OBE than for the Wharton model.[9]

An Interpretation of the Findings

The evidence on the Wharton and OBE forecasts, as presented in [5] and [9], suggests the following observations.

1. Mechanical adjustments of a very simple type, involving only the last two observed single-equation disturbances and sometimes the first-order autocorrelation of these terms, tend to reduce the errors of the ex post forecasts. This effect is especially strong for the Wharton predictions. It reflects the presence of substantial serial dependencies in the residuals, whether these are due to the omission of important variables or of minor random factors that nevertheless have some net lingering influence, to misspecification of the *form* of a relationship, or to measurement errors in the variables to be "explained." There is evidence that the gains from the mechanical adjustments are larger and more widespread for predictions over the short spans—one or two quarters—than over the longer spans. It appears that the autocorrelated errors here involved are mainly short-lived, particularly in the more volatile series, but that they are larger and more persistent in some of the relatively smooth, trend-dominated variables.

2. On balance of the evidence, judgmental alterations of constant terms reduce the XP errors still more than mechanical procedures, particularly for the shorter-span predictions. Inspection of residuals for any pattern that may have potential predictive value has long been recognized as a useful practice in forecasting with econometric models, so presumably the *OR* adjustments include much of the information that is contained in the *AR* or *GG* adjustments. In addition, however, the judgmental adjustments can incorporate information about exogenous developments already in progress or impending, such as a large strike, a major bankruptcy or sectoral financial difficulties, a war, a foreign crisis. They can be used to deal with autonomous elements in fiscal and monetary policies, e.g., changes in tax rates or reserve requirements. They can allow for small revisions of inputs of data and take advantage of the most recent information on the current state of

[8] The variables include GNP, personal consumption expenditures, gross private domestic investment, change in nonfarm business inventories, GNP implicit price deflator, unemployment rate, corporate profits before taxes (Wharton only), and net exports (OBE only). All of these are in billions of current dollars, except the price deflator (1958 = 100.0) and the unemployment rate (percentage of the labor force).

[9] One possible reason for this is that the ex ante forecasts made by the OBE are not released to the public. Judgmental adjustments may play a smaller role in these forecasts and are in any event difficult to reconstruct.

the economy. Such adjustments may be made before the model is solved to obtain the forecast or after the forecasters have examined the model predictions and found any of them "unreasonable," given their prior concepts and external information. Clearly, mechanical adjustments can still be viewed as a part of the model, but the judgmental ones cannot. The success of the latter emphasizes the role of the forecaster in his interaction with the econometric model.

3. In this analysis, the *XA* and *XP* measures are strictly matched. The only source of difference between them is that the ex ante predictions are based on the forecasters' estimates of the values of the exogenous inputs; the ex post predictions, on historical data for these variables. Since the latter information is presumed correct, whereas the judgmental "guesstimates" of the exogenous values inevitably contain errors, this factor per se must work in favor of the ex post predictions. Yet the *XA* errors are frequently smaller than the corresponding *XP* errors. Some of these differences are minor and probably not significant, but some are substantial. Even where the ratios are close to unity, the question may be raised why they are not higher. The low *XA*/*XP* ratios are observed in the first place in connection with the judgmental (*OR*) adjustments, especially for the Wharton model. It is possible that these adjustments are frequently of a kind to offset the defective external predictions for the exogenous variables. Actually, if no other factors were at work here, strict offsets would merely prevent *XA* from exceeding *XP*, that is, they would equalize the two errors; the *OR* procedure would have to do more than that to draw *XA* below the levels of *XP*.

4. Even where only mechanical adjustments are used, $XA < XP$ in nearly half of the comparisons made. Conceivably, such procedures would sometimes work to cancel the errors in the predicted exogenous inputs into the model, but the validity of this conjecture is untested and uncertain. In any event, this factor could not fully account for the findings under discussion. Moreover, *XA*/*XP* ratios that are smaller than 1 are observed even for predictions made without any adjustments of the constant terms, especially for the OBE model (see, e.g., column 6 in Table I-2). This suggests that the models contain some misspecifications which are being more than offset by errors in exogenous variables. This could happen, for instance, when a sufficiently large underestimate of government spending is fed into a model with too high fiscal multipliers.

There is real need for tests designed to deal effectively with this analytical situation. What errors are actually being made in the exogenous forecasts? What precisely can be established about the role and effects of such errors in forecasting with an econometric model of a given structure? A detailed investigation of this subject for each of the major models could contribute much to our knowledge of the models and forecasting processes involved and perhaps even to their improvement.

AGGREGATE ECONOMIC FORECASTS AND BENCHMARK PREDICTIONS

Comparisons with Naive and Autoregressive Extrapolations

Forecasting accuracy is a relative concept. There is no absolute standard by which to judge the quality of economic forecasts. In any event, however, economic forecasts should not be inferior to mechanical extrapolations in serving the purposes of their users. In practice, this should mean that forecasters try to utilize efficiently the predictive power of the past values of the series they predict. It also means that economists qua forecasters must suffer their predictions to undergo the tests of "naive," and even of not so naive, extrapolative models. However, not all interpretations of such tests are fair and acceptable.

Haitovsky and Treyz [9] report on systematic comparisons of Wharton and OBE forecasts with predictions from three extrapolative models: *N*1, which projects forward the last known (base-period) level of the given variable; *N*2, which projects the last known change; and *N*3, which projects the average relation of the present level to the past levels of the series by means of fourth-order autoregressive equations estimated with quarterly data for the sample period of the model (Wharton: 1948-64; OBE: 1953–II-1967).[10] Table I-3 illustrates the results for the ex ante *(OR)* forecasts of GNP.

TABLE I-3

Wharton and OBE Model Ex Ante Forecasts of GNP over Spans of from One to Four Quarters, Comparisons with Two Types of Extrapolation
(billions of dollars)

Span of Forecast (quarters)	Mean Absolute Errors (*MAE*) 1966 III–1969 III: Wharton Model *(XA)*[a] (1)	*N2*[b] (2)	*N3*[c] (3)	1967 II–1969 III: OBE Model *(XA)*[a] (4)	*N2*[b] (5)	*N3*[c] (6)
One	3.75	3.57	5.32	2.12	3.17	2.38
Two	7.82	7.51	11.56	7.82	7.34	3.83
Three	10.37	12.62	17.80	12.73	12.92	4.48
Four	12.62	17.72	25.54	19.48	15.88	4.92
Next year	6.58	10.81	14.23	11.47	9.34	3.53

[a] *MAE* of ex ante forecasts with original adjustments.

[b] Extrapolations of the last-known change, according to the form $A_{t+1} = A_t + \Delta A_t + u_{t+1}$, where A is an actual value and u is a random error assumed to be zero.

[c] Fourth-order autoregressive extrapolation: $A_{t+1} = a_0 + a_1A_t + a_2A_{t-1} + a_3A_{t-2} + a_4A_{t-3} + u_{t+1}$, based on the assumption that $u = 0$.

The table lists the mean absolute errors for the forecasts and the corresponding statistics for the best and second-best of the three *N* models. These turn out to be *N2* and *N3*, but with different ranks for the two sets of comparisons which involve different sample and forecast periods (cf. columns 2–3 and 5–6). The Wharton forecasts have on the average smaller errors than the extrapolative benchmark models, except for the two shortest spans, where *N2* is somewhat more accurate. The OBE forecasts are less accurate than *N3* except for the one-quarter span, and less accurate than *N2* in a few instances.

According to corresponding measures for other variables, the errors of *N3* are on the average smaller than those of *N2* for consumption (C), but larger than *N2* in government expenditures. *N1* works best for gross private domestic investment (GPDI), but elsewhere tends to be the least accurate of the three benchmark models. The econometric forecasts are for the largest part more accurate than the extrapolative predictions, but *N3* has smaller average errors for most spans than Wharton for real GNP and than OBE for C and GPDI.

Table I-4 presents a more extensive array of summary statistics of error for still other types of extrapolations and some selected business forecasts. It includes, in addition to the naive models *N1* and *N2*, projections of the average value of past changes in the given series as available to the forecaster from the historical

[10] See Table I-3, footnotes b and c. An attempt was made in these tests to allow for the effects of data revisions. The ex ante forecasts are based on lagged values representing preliminary data. In [5], the values of revised actual changes are added to these preliminary figures on level to form the series of realization, A_t. This procedure has two drawbacks: First, certain inconsistencies can arise within the set of the actual values thus computed; e.g., as pointed out by George Green, ratios of the estimates for GNP and the price deflator would deviate from the estimates for real GNP. Second, the implicit assumption behind these calculations is that data revisions affect the levels rather than the changes in the series, but, as shown in [3, Chap. VI], the revisions of the quarterly estimates of GNP and its components do affect the major patterns of change as well as the levels. However, it is far from obvious how data revisions should be handled in the context of forecast evaluation; other approaches to the problem present different difficulties.

TABLE I-4

Summary Measures of Error for Extrapolations of Naive and Autoregressive Models and Selected Forecasts of Annual Levels of Four Aggregative Variables, 1953-63

Summary Statistics of Error[a]	Naive and Autoregressive Extrapolations				Selected Forecasts		
	N1 (1)	N2 (2)	N2*[b] (3)	N3[c] (4)	Set B (5)	Set F (6)	Set G (7)
	Gross National Product (billions of dollars)						
ME	–19.9	–8.8	–1.9	–6.4	–2.8	–4.8	1.8
SD	15.2	18.1	15.9	14.7	10.8	7.8	8.1
MAE	22.7	16.0	11.6	13.7	9.7	8.1	6.9
RMSE	24.6	19.3	15.3	15.4	11.2	9.1	8.3
	Industrial Production (1947–49 = 100)						
ME	–4.2	0.7	1.1	–4.5	n.a.	0.8	2.5
SD	10.2	12.3	10.3	8.9	n.a.	4.8	4.4
MAE	9.2	10.5	7.2	7.5	n.a.	3.9	4.2
RMSE	10.6	11.8	9.9	9.6	n.a.	4.9	5.0
	Personal Consumption Expenditures (billions of dollars)						
ME	–12.9	–3.8	–2.9	–1.4	–1.3	–3.0	n.a.
SD	4.9	8.3	5.0	6.1	5.6	4.8	n.a.
MAE	12.9	7.5	4.7	4.9	4.6	4.8	n.a.
RMSE	13.7	8.8	5.6	5.9	5.7	5.7	n.a.
	Gross Private Domestic Investment (billions of dollars)						
ME	–2.9	–5.4	0.1	–3.9	–2.7	–1.3	n.a.
SD	8.7	15.3	9.2	7.9	6.9	4.4	n.a.
MAE	7.1	11.0	6.4	7.4	6.3	3.9	n.a.
RMSE	9.2	16.2	9.2	8.8	7.4	4.6	n.a.

n.a. = not available.

[a] *ME* = mean error; *SD* = standard deviation or errors; *MAE* = mean absolute error; *RMSE* = root mean square error.

[b] The assumption of $N2^*$ is that $A_{t+1} = A_1 + \overline{\Delta A} + u_{t+1}$, where $\overline{\Delta A}$ is the average value of past changes in A, computed from some starting date (here 1947) up to and including the most recent value known at the time of the forecast. Again, in the prediction $u_{t+1} = 0$.

[c] Based on regressions of A_t on A_{t-1}, where the A's denote the actual values of the given variable; $i = 1, 2, \ldots, 5$ for GNP and industrial production; $i = 1, 2$ for consumption; and $i = 1$ for GPDI. See text for more detail.

record ($N2^*$; see Table I-4, footnote b). The autoregressive predictions ($N3$) are based on quarterly series in a moving sample period, that is, on equations that are re-estimated so as to add information from the recent data points that could have been used in the corresponding annual forecast.[11]

The simple trend projections ($N2^*$) produce smaller mean absolute errors than any of the other N models for each of the four variables covered, but in terms of the root mean square errors the autoregressive models ($N3$) are more accurate than $N2^*$ for industrial production and gross private domestic investment (columns 3 and 4). Except for consumption, $N3$ shows more indication of bias (larger

[11] Extrapolated values are substituted for the as yet unknown actual values in the autoregression designed to predict more than one quarter ahead. For example, if a third-order form is used, the prediction for two quarters ahead would be: $A_{t+2} = a + b_1\hat{A}_{t+1} + b_2A_t + b_3A_{t-1}$, where $\hat{A}_{t+1} = a + b_1A_t + b_2A_{t-1} + b_3A_{t-2}$ (this assumes that A_t is the last-known value of the A series). It is a stepwise process of "chaining" the extrapolated values $\hat{A}_{t+i}$ and the actual values A_{t-j}, in which more of the former and fewer of the latter terms are being used as the predictive span is increased. To derive annual benchmark predictions, the coefficients $a, b_1, b_2, \ldots,$ are typically re-estimated each year in the first step of the process, relating to the fourth quarter of the base year, and are retained in the following four steps which refer, respectively, to each of the successive quarters of the target year. The scheme implies an increase in the variance of the residual errors u_{t+i} as the span i is lengthened [19, p. 102].

absolute values of *ME*) but less of inefficiency (smaller *SD*) in comparison with $N2^*$. $N1$ and $N2$ yield for the most part considerably larger errors than either $N2^*$ or $N3$ (note, however, the relatively good showing of $N1$ for GPDI).

The three sets of authentic ex ante forecasts labeled B, F, and G come from the collection examined in my 1967 study [19] and represent well-known financial and publishing sources. These forecasts have on the average smaller errors than any of the included N models, except for consumption, where they are about as accurate as $N2^*$ and $N3$.

It might seem that the comparisons with extrapolations are more favorable to these business forecasts than to the econometric model forecasts, but the differences between the measures in Tables I-3 and I-4 are such as to preclude any reliable conclusions on the matter. The autoregressive models ($N3$) are not computed alike in the two sets of comparisons. Also, and this is probably more important, the forecasting periods differ in ways that are apt to affect the results.

The tabulation below of summary statistics of errors in annual GNP forecasts illustrates how sensitive the measures of relative accuracy of forecasts and extrapolations of various types can be with regard to differences in the periods (M_P is in billions of dollars):

	Set B		Set F		Set G	
	1953-63	1964-69	1953-63	1964-69	1953-63	1964-69
M_P	10.7	15.2	8.8	14.3	7.9	17.0
M_P/M_{N1}	0.435	0.276	0.359	0.258	0.322	0.306
M_P/M_{N2}	0.553	1.024	0.457	0.959	0.410	1.140
M_P/M_{N2^*}	0.699	0.421	0.578	0.394	0.518	0.467

The years 1953-63 witnessed three business recessions; in the years 1964-69 the growth in GNP was relatively steady and uninterrupted by any major declines. The errors of the GNP forecasts, not surprisingly, tend to be considerably larger in the latter than in the former period, reflecting the strong upward trend in the GNP series; this is clearly exemplified by the root mean square errors (M_P) of the three sets of business forecasts.[12] But such an increase in absolute errors need not denote a worsening of the forecasts. In terms of the ratios of M_P to the corresponding measures for the naive model $N1$, the performance of the forecasters improved (the M_P/M_{N1} figures are smaller for 1964-69 than for 1953-63). Similarly, the root mean standard error (RMSE) ratios declined when the $N2^*$ model was used as a yardstick. Thus it can be said that the GNP forecast errors increased on the average less between the two periods than did the mean annual change in GNP measured either individually in the years covered or cumulatively over the postwar years.[13]

[12] Being based on small numbers of observations, the root mean square error statistics differ appreciably depending on whether they are computed according to the expression $[(1/n)\Sigma(E_t)^2]^{1/2}$ or as

$$\left\{(\bar{E}_t)^2 + [1/(n-1)]\Sigma(E_t - \bar{E}_t)^2\right\}^{1/2},$$

where $E_t = P_t - A_t$, $\bar{E}_t = (1/n)\Sigma E_t$, and the Σ's denote summations over $t = 1, \ldots, n$. The former, unadjusted estimates are listed as M_P in the preceding tabulation; the latter ones, with an adjustment for sample size, are used in the RMSE lines in Table I-4.

[13] The relative deterioration of the benchmark extrapolations $N2^*$, which is indicated by these results, suggests that it might be better to use a moving period for the computation of the average historical change rather than a period with a fixed origin in the increasingly remote past. To quote Samuelson [17], "just as there is a time for remembering, there is a time for forgetting." Of course, the $N2^*$ model represents an extremely simple type of trend extrapolation. There are various ways by which the recent past can be emphasized and the distant past de-emphasized in extrapolative models. (In one of the most popular formulas, prediction is by means of a linear adaptation model with coefficients declining exponentially in the direction of the past.) For series that show fairly steady *rates* of growth much of the time, it may also be better to use relative rather than absolute changes as the basis for trend projections.

On the other hand, comparisons of the same forecasts with simple extrapolations of last-known changes (M_P/M_{N2}) clearly show the more recent judgmental predictions in much worse light than the earlier ones. (It will be recalled that such naive projections also proved relatively strong when confronted with the econometric forecasts in Table I-3.) In times of sustained growth, the $N2$ extrapolations are at their best (the model performs much worse, of course, in times of weaker trends and stronger cyclical movements, as during the post-Korean decade).

Factors Influencing the Relative Accuracy of Forecasts and Extrapolations

Many important economic aggregates, particularly such nominal quantities as GNP, consumption expenditures, and the price level, contain strong trend and autoregressive elements which should be exploited in the forecasts. However, forecasters must also predict the course of important variables that are much more volatile and sensitive to business fluctuations, e.g., the unemployment rate, inventory investment, and interest rates, which may often be strongly influenced by regressive rather than extrapolative expectations. Furthermore, forecasters' services appear to be particularly sought after—and tested—in the neighborhood of actual and suspected turning points in business and financial activity. Simple naive models either do not predict any turns at all or merely reproduce the actual turns with lags. Higher-order autoregressive models can, in principle, produce forecasts of turning points, but there is some evidence that they are poor predictors of the timing of such events [19, Chap. 6]. Actual changes often show much higher correlations with the changes predicted by forecasters than with the changes produced by autoregressive extrapolations; e.g., for GNP, 1953-63, the r^2 coefficient measuring the predictive efficiency of set F in this sense is .814, and r^2 for set G is .778, whereas, for the best of the autoregressive models examined, $r^2 = .179$ (see [14, Table I-5 and text]). To be sure, extrapolative models better than $N3$ can frequently be constructed, although this is likely to require considerably more information and computational processing. However, such improvements are much more likely to take the form of reduced systematic errors than of reduced predictive inefficiency of the extrapolations.[14]

If forecasts are apt to prove better at turning points that did actually occur, they are also apt to prove worse than extrapolations at times when turning points were expected, but failed to materialize. The two types of directional error may have different costs or consequences, and, if so, should be given correspondingly different weights. With equal weights, at least, the evidence is predominantly favorable to the forecasts. Because false warnings are much less numerous than missed turns, directional errors in general are considerably more frequent in extrapolations than in forecasts [19, pp. 114-20].

Average errors of judgmental business forecasts often increase more rapidly with lengthening span than do the average errors of trend projections $N2^*$ and autoregressive predictions $N3$. For example, the ratio M_P/M_{N2}^* is .67 for six-month and .82 for twelve-month predictions of industrial output, according to measures for mean forecasts of a large group of business economists, 1947-63. Apparently, many forecasters fail to use effectively the past record of their series in predicting developments beyond the first two or three quarters ahead.

There are also some contrary considerations, however, which apply to forecasts based on more sophisticated models. Some economic relationships involve long distributed lags, and these may be more helpful in forecasting over intermedi-

[14] For a discussion of sophisticated techniques recently developed for optimal characterization of discrete linear processes, see [1]. Some support for the statement in the text is found in the preliminary results obtained by Mervin Daub in "An Appraisal of Canadian Short-Term Aggregate Economic Forecasts," a proposed dissertation draft, Graduate School of Business, University of Chicago (1970). This work, among others, applies the Box-Jenkins techniques in comparisons with Canadian GNP forecasts. Further study of the relative performance of these models, with applications to U.S. forecasts, is planned.

ate time spans than extrapolations based on trend projections or autoregressions. The lagged values of the dependent variables in an autoregressive equation may serve as strong proxies for the true "causal" variables of a structural model, which have similar timing. But if the autoregressive model contains only a few terms for the shortest lags and the structural model also contains longer lags in exogenous and other endogenous variables, the two matrices of the respective "explanatory" variables will be less well correlated. The longer lags of the structural model may contribute mainly to the accuracy of the longer forecasts and, if so, especially for these forecasts, the autoregressive scheme should be less useful. Accordingly, in a recent debate on ways and results of testing the predictive performance of econometric models vs. autoregressive extrapolations, several model builders and critics have argued that such tests are more powerful when applied to multiperiod forecasts with longer spans than when applied to short single-period forecasts only.[15] Furthermore, the expectation was expressed that the comparisons involving longer predictions would prove to be substantially more favorable to the econometric models.

Comparisons of the type presented in Table I-3 provide tests of the validity of this contention. Their evidence, however, is mixed. Ratios of the corresponding mean absolute errors of ex ante forecasts and autoregressive extrapolations decline as the length of the span increases for the Wharton predictions of several variables (GNP, GPDI, and the exogenous forecasts of government spending). For some other variables and, particularly, for the OBE predictions, the ratios behave irregularly or even increase with the length of the span.[16]

QUARTERLY SURVEYS OF MULTIPERIOD FORECASTS BY DIFFERENT METHODS

General Characteristics and Results of the ASA-NBER Surveys

Beginning in the last quarter of 1968, the American Statistical Association has conducted quarterly surveys of forecasts by those members of its Business and Economics Statistics Section who are professionally engaged in a continuing analysis of the business outlook. The National Bureau cooperated with the ASA in designing the questionnaire for the surveys and it regularly processes and evaluates the results.[17]

The surveys cover total GNP and some of its more sensitive or autonomous expenditure components, and, also, industrial production, unemployment, corporate profits, housing starts, and the implicit price deflator. Forecasts for the current and the following three quarters are produced on each occasion, but at mid-year and year-end predictive spans of five and six quarters are used. Information is also collected on the composition of the sample, the methods and assumptions of the forecasters, and the probabilities they attach to different expected outcomes of anticipated changes in GNP and the price level.

Over fifty forecasters took part in each of the recent ASA-NBER surveys. They represent a broadly based and diversified group, in terms of their primary affiliation and location. The rates of turnover and attrition were initially rather high, but they quickly declined, and most of the present members have excellent or good records of participation in the past surveys. The group averages presumably reflect

[15] See the comments on Cooper [4] by Goldfeld, McCarthy, and the OBE Econometric Branch staff [9a]. Also see [10].

[16] These results can be linked to differences between the two sets of econometric forecasts—the OBE predictions of endogenous variables appear relatively more successful for the short spans, the Wharton predictions for the longer spans; and to differences between the two applications of the *N*3 model—the errors of the latter are on the whole smaller in the set compared with the OBE forecasts than in the set compared with the Wharton forecasts. This last finding is presumably due mainly to the fact that the Wharton sample period (1948-64) differs much more than the OBE sample period (1953-67) from the forecast periods of 1966 (1967)-69.

[17] Charlotte Boschan and I share the responsibility for this analysis. For a description of the new survey and an assessment of some early results, see [21] [22]; also, see the press releases in successive issues of the *American Statistician* from April 1969 on.

well the trends in what is sometimes called the "standard forecast," but the individual forecasts and their distributions deserve no less attention. For example, changes in the dispersion of member forecasts around the group averages may have interesting implications as indications of the varying degree of uncertainty and consensus among the forecasters.

Table I-5 presents the means and standard deviations of errors for all forecasts of three selected variables from the first seven ASA-NBER surveys. When read line by line across the table, the figures show the accuracy of predictions made at a given date for successively more distant target periods. There is the usual tendency for the average errors and the dispersion of errors to increase with the span of forecast, but it is easy to find exceptions to this rule in the average errors relating to particular surveys. It appears that these cases reflect mainly the forecasters' failure to anticipate major deviations from recent trends such as the slowdown in 1969III-1970I: note, e.g., the results for the May 1969 survey in line 3.

When read column by column, the table tells us how the averages and dispersion of errors of forecasts for a given variable change as the target period draws nearer in each successive survey. Here, too, the predictive span varies, but the analytical situation differs from that involved in the line-by-line comparisons. There the total fund of information available to any forecaster can be taken to be fixed as of the date at which the survey questionnaire was filled out, for all spans. Here new information becomes available to the forecaster between the surveys, including, importantly, any lessons from the errors of the preceding forecast. This information appears to yield considerable advantages. Thus, reading down any of the columns 2-7 in Table I-5, one finds a systematic improvement in the forecasts for any of the variables covered, that is, progressively smaller average errors (taken without regard to sign) and smaller standard deviations of errors. These gains in accuracy are on the whole more pronounced and regular than the intrasurvey gains from reductions in the time span of the forecast (which can be observed by reading the entries in each line from right to left).

A systematic learning process is strongly indicated by the results for GNP. As long as underestimates (indicated by the negative signs in Table I-5) prevailed, each survey improved upon the previous one by revising the forecasts upward: this applies to the predictions of December 1968 and February, May, and August 1969. The December 1969 forecasts were also higher than the corresponding averages from the August 1969 survey, but this time the upward adjustment proved counterproductive: The forecasters overshot the actual values by failing to recognize adequately the deceleration of GNP. In the next two surveys, however, this error was largely corrected. Apparently noting the emergence of overestimates in the period, the forecasters revised their predictions *downward* in these surveys, thereby tending to reduce the errors. These results are suggestive of elements of partially successful adaptive forecasting.

Finally, moving along any of the diagonals in Table I-5 allows us to compare the errors in successive forecasts for approximately the same span. For GNP and some other variables such as the price deflator, this exercise shows that the predictions from the three earliest surveys have been substantially less accurate than the more recent ones. However, the arrays of the error measures for some volatile series, such as the changes in inventories and the unemployment rate, show much less evidence of this and the other regularities here discussed.

The Methods Reported and the Accuracy of Forecasts

In each survey, the participants are asked to rank the methods identified in Table I-6 according to their relative importance in the forecaster's own work. In December 1968, over 80 per cent reported using an informal GNP model, and about 70 per cent reported using leading indicators and anticipation surveys. In

TABLE I-5

Mean Errors and Standard Deviations of Errors in Forecasts of Three Variables, ASA-NBER Quarterly Business Outlook Surveys, 1968-70

Date of Survey	Forecast Quarter[a] 1968 IV (1)	1969 I (2)	1969 II (3)	1969 III (4)	1969 IV (5)	1970 I (6)	1970 II (7)
			Gross National Product (billions of dollars)				
Dec. 1968	–3.2(3.6)	–6.5(6.4)	–16.2(8.9)	–19.2(12.2)			
Feb. 1969		–2.3(3.5)	–13.0(5.6)	–16.2 (9.7)	–12.0(11.6)		
May 1969			–8.1(2.8)	–13.0 (5.5)	–11.3 (8.6)	–5.8(12.2)	
Aug. 1969				–3.9 (5.5)	–3.1 (5.6)	+1.1 (8.9)	+3.1(12.3)
Dec. 1969					+1.3 (3.2)	+3.8 (6.0)	+3.3 (9.2)
Feb. 1970						+1.8 (3.3)	+1.7 (5.3)
May 1970							+1.0 (3.9)
			Index of Industrial Production (1957–59 = 100)				
Dec. 1968	–1.3(1.0)	–4.0(2.0)	–6.5(2.6)	–5.8 (3.0)			
Feb. 1969		–2.0(1.8)	–4.7(2.5)	–4.6 (2.6)	–1.1 (3.0)		
May 1969			–1.9(0.9)	–2.3 (1.7)	–0.1 (2.3)	+2.6 (2.9)	
Aug. 1969				0 (1.0)	1.6 (2.1)	+3.5 (2.9)	+5.0 (3.5)
Dec. 1969					0.9 (0.7)	+1.7 (1.8)	+2.3 (3.1)
Feb. 1970						+0.1 (1.4)	+0.8 (2.3)
May 1970							+0.7 (1.3)
			GNP Implicit Price Deflator (1958 = 100)				
Dec. 1968	–0.7(0.5)	–0.7(0.8)	–1.8(0.9)	–3.0 (1.1)			
Feb. 1969		–0.3(0.6)	–1.5(0.7)	–2.5 (0.7)	–3.7 (0.9)		
May 1969			–1.1(0.4)	–2.0 (0.6)	–3.2 (0.7)	–3.4 (0.9)	
Aug. 1969				–0.5 (0.6)	–1.4 (0.7)	–1.4 (1.0)	–2.5 (1.2)
Dec. 1969					–0.5 (0.5)	–0.2 (0.6)	–1.0 (0.9)
Feb. 1970						+0.1 (0.5)	–0.8 (0.6)
May 1970							–0.4 (0.8)

[a] The target of the forecast. For example, the entries in the first line of column 1 refer to the forecasts (in this case, estimates for the current or base quarter) made in the December 1968 survey for the last quarter of 1968. The entries in the first line of column 2 refer to the forecasts from the same survey for the first quarter of 1969. The first figure is the mean, the second figure (in parentheses) is the standard deviation, of errors of all individual forecasts in the given category (as defined by the date of survey, the date of the target quarter, and the predicted variable).

addition, about one-fifth of the survey members had their own econometric models, and more than half made use of models designed by others. These proportions, which disregard the ranks assigned to these methods or procedures, are also roughly indicative of the corresponding distributions in the later surveys.

What is here labeled an informal GNP model embraces, one suspects, a variety of procedures whereby the forecasters predict the major expenditure components of GNP, add these figures to obtain an over-all forecast, finally check and, if needed, repeatedly adjust the results for plausibility and consistency. It is an eclectic, "intuitive," or "opportunistic" approach,[18] in which judgment typically plays a large role, although various models of macroeconomic relationships are probably often employed, at least implicitly. From about half to two-thirds of the participants in each survey ranked this approach as first. On the average, approximately 13 per cent of the forecasters ranked the leading indicators as first; 8 per cent, outside models; and 6 per cent, own econometric models.

Considerable effort has been made to establish whether the different methods or groupings of methods were associated with significantly different degrees of forecasting accuracy. This analysis relied mainly on regressions of forecast errors on dummy variables representing the different methods, with separate estimates being made for each survey and target quarter (predictive span). Table I-6 presents a sample of results for two of the surveys, using forecasts of GNP. In each of these regression equations, the individual forecast errors serve as the dependent variable and the forecasters are classified by the first-ranked methods. The constant terms a (column 1) are estimates of the expected values of the errors for those forecasters who ranked the informal GNP model first, always the largest group.[19] The b_i coefficients of the dummy variables are estimates of the differences between these average errors and those of the groups who ranked the other (ith) method first. The t statistics (ratios of the b_i coefficients to their respective standard errors) indicate the statistical significance of these differences.

As illustrated in Table I-6, relatively few of the b_i coefficients appear to be significant (less than one in four at the 10 per cent level, one in six at the 5 per cent level, according to a count that includes also the other surveys not included in the table). Over 60 per cent of all significant coefficients are concentrated in the very thinly populated groups: anticipations surveys, other methods, and methods not elsewhere classified (columns 5-7). These are mostly negative, often large, deviations, which suggests that the forecasters in these groups had a relatively poor record because of large underestimation errors.

According to these results, no systematic differences appear between the accuracy of the forecasters who used principally the informal GNP model, those who favored the econometric models, and those who favored the leading indicators approach. Tests designed to include the evidence of all ranks (rather than only the first ranks) lead to the same general conclusion. To be sure, the number of different rankings is so large that the dummy-variables approach is not applicable to all the combinations that can be distinguished, given the size of the available samples. But it was possible to try out a few alternatives based on simplified classification schemes, and these revealed no significant differences between the three principal methods (informal GNP model; econometric models, own and outside; and leading indicators combined with anticipations surveys) as far as the accuracy of forecasting GNP is concerned.

Should we then read these results as saying that it does not matter, at least for the accuracy of the GNP forecasts, which methods are being used and how they

[18] These descriptive terms have been used by some practitioners or sympathetic observers of this type of forecasting. See [2, pp. 137-59] [11, pp. 389ff.].

[19] As elsewhere, the errors are computed by subtracting the actual from the predicted values. Hence, the negative signs of the coefficients indicate average underestimates; positive signs, average overestimates. The constant terms often fairly well resemble the corresponding over-all mean errors, and they are generally highly significant, as would be expected.

TABLE I-6

Forecasting Accuracy and Forecasting Method, Evidence from ASA-NBER Business Outlook Surveys, Regressions with Dummy Variables, 1968-70

	Estimated Coefficients of the Regression Equation:[a] $E = a + \Sigma\, b_i D_i$						
Forecast Quarter[b]	Informal GNP Model a (1)	Econo-metric Model—Own b_1 (2)	Econo-metric Model—Outside b_2 (3)	Leading Indica-tors b_3 (4)	Anticipa-tions Surveys b_4 (5)	Other Methods b_5 (6)	N.E.C. b_6 (7)
			December 1968 Survey				
IV '68 (0)	–3.10	0.50	–0.23	–1.65	0.60	2.10	0.77
	(5.99)	(0.29)	(0.20)	(1.40)	(0.23)	(0.97)	(0.35)
I '69 (1)	–6.02	0.82	–3.56[c]	–1.06	1.02	1.35	2.35
	(6.55)	(0.27)	(1.71)	(0.51)	(0.22)	(0.35)	(0.61)
II '69 (2)	–15.68	–0.52	–5.15[c]	0.76	3.68	–3.99	4.01
	(12.37)	(0.12)	(1.79)	(0.26)	(0.57)	(0.75)	(0.75)
III '69 (3)	–18.64	–1.36	–2.61	–1.53	6.14	–12.36[c]	9.64
	(10.84)	(0.24)	(0.67)	(0.39)	(0.70)	(1.71)	(1.33)
IV '69 (4)	–11.58	–5.82	–2.84	–3.67	5.08	–24.75[d]	9.91
	(5.09)	(0.77)	(0.55)	(0.71)	(0.44)	(2.59)	(1.04)
			February 1970 Survey				
I '70 (0)	2.68	–3.18	–1.18	–3.39[d]	0.32	–1.43	–1.01
	(4.89)	(1.88)	(0.51)	(2.56)	(0.10)	(0.85)	(0.72)
II '70 (1)	3.06	–4.31	–2.06	–5.34[d]	–0.06	–3.81	–0.73
	(3.40)	(1.55)	(0.54)	(2.46)	(0.01)	(1.37)	(0.31)

[a] E is the individual forecast error (predicted minus actual value) for the given survey and target quarter. The dummy variables D_i assume the value 1 for the method i, zero otherwise. The methods for $i = 1, \ldots, 6$ are identified below in the headings for columns 2-7 (e.g., econometric models, own and outside, have indexes 1 and 2, respectively, etc.). N.E.C. = not elsewhere classified. The method labeled "informal GNP model" is represented by the constant (column 1). The classification is by first ranks only. The figures in parentheses show the t ratios for the corresponding coefficients in the line above (the signs of these ratios, which are the same as those of the coefficients, are omitted).

[b] Span of forecast, in quarters, is indicated in parentheses.

[c] b_i coefficient significant at the 10 per cent level.

[d] b_i coefficient significant at the 5 per cent level.

are combined? It is possible that this is so, but quite unlikely, if the statement were to apply strictly to all kinds of well-defined methods. The survey questionnaire refers to general approaches used in macroeconomic forecasting rather than to any specific models and procedures. It may be that such methodological distinctions among practicing economic forecasters are seldom sufficiently large and systematic to produce clear-cut differences in the observed accuracy of aggregative forecasts. In other words, it is possible that the findings illustrated in Table I-6 reflect mainly the vagueness of the methodological classification used. If sharper distinctions were available, stronger elements of relationship between forecasting accuracy and forecasting method might be found. However, the very fact that no single method commands the general adherence of regular economic forecasters indicates that none was expected to be consistently more accurate than others.

CONCLUSION

In an era of rapid growth of national economic statistics, econometrics, and computer technology, complex models of the economy can be built and used in forecasting, and they are. Yet, in practice, economic forecasting is still very much the domain of individual accomplishment, of experience and judicious application of past and current data in drawing probabilistic inferences about the future. At least for two of the more important U.S. macroeconometric models, the major importance of the interaction between the model and its builders and operators is clearly demonstrated by the advantages of the ex ante forecasts over the ex post ones. Business economists' forecasts, which are not based on formal, explicit models but rely largely on general economic theory and judgment in interpreting the data, often include predictions that are about as accurate as the over-all forecasts made with econometric models. Mechanistic predictions, whether based on extrapolative benchmark models or on econometric models applied ex post without any adjustments (but with the advantage of correct exogenous inputs!) have on the whole been less accurate. (However, in particular periods even some naive models may often "outguess" all kinds of bona fide forecasts, as did the same-change extrapolations—*N*2—during the long expansion of the 1960's.)

No single, well-defined, and reproducible model commands the loyalty of any large group of economic forecasters in business, government, and academic institutions. The main reason for this is probably that none has been proved to be consistently superior. Forecasters are properly interested first of all in better forecasting, and any approach that held out a good promise to result in more accurate predictions would fairly promptly attract many followers, regardless of its other characteristics.

This, of course, is not to deny that forecasters, like economists in general, do have their theoretical and ideological preferences; for example, some are "monetarists" and pay great attention to changes in money supply, others of Keynesian persuasion place more emphasis on fiscal variables. Accordingly, a broad distinction can be made between such groups of forecasters, and comparisons between the results of their respective approaches should be instructive, especially where "families" of explicit models can be studied on both sides of the issue.[20]

No scientific progress can be accomplished by economists qua forecasters, as a group, without systematic testing of the different verifiable and replicable forecasting methods. Hence a major question for future research is how to forecast best (given a number of well-defined procedures or "models"); the question of who forecasts best (given predictions from various sources based on undefined combinations of various approaches) is per se of much less interest. This implies the need to study the performance of diverse, specific models. The main practical difficulty here is that the available samples of observations are small, the series of interesting ex ante forecasts being as a rule quite short. This can be fully corrected only by the progression of time in which more data accumulate, but meanwhile partial gains might be made by intensive studies of materials from new and underexplored sources and of various simulations and ex post predictions.

REFERENCES

1. Box, G. E. P., and G. M. Jenkins, *Time Series Analysis, Forecasting and Control.* San Francisco, Holden-Day, 1970.
2. Butler, William F., and Robert A. Kavesh, eds., *How Business Economists Forecast.* Englewood Cliffs, N. J., Prentice-Hall, 1966.
3. Cole, Rosanne, *Errors in Provisional Estimates of Gross National Product.* New York, NBER, 1969.
4. Cooper, Ronald L., "The Predictive Performance of Quarterly Econometric Models of the United States." In [25].
5. Evans, Michael K., Yoel Haitovsky, and George I. Treyz, assisted by Vincent Su, "An Analysis of the Forecasting Properties of the U.S. Econometric Models." In [25].

[20] Although much has been written about the relative merits of the monetarist and "fiscalist" positions, an objective and comprehensive analysis that would have a chance of being conclusive is still a task for further data collection and research (see also the last two sentences in the text below).

6. Fels, Rendigs, and C. Elton Hinshaw, *Forecasting and Recognizing Business Cycle Turning Points*. New York, NBER, 1968.
7. Goldberger, Arthur S., "Best Linear Unbiased Prediction in the Generalized Linear Regression Model." *Journal of the American Statistical Association*, June 1962, pp. 369-75.
8. Green, George R., in association with Maurice Liebenberg, and Albert A. Hirsch, "Short- and Long-Term Simulations with the OBE Econometric Model." In [25].
9. Haitovsky, Yoel, and George Treyz, "Forecasting with Quarterly Macroeconometric Models, Constant Adjustments, Autoregressive Extrapolations, and Naive Model Predictions: The U.S. Experience." Paper prepared for the Second World Congress of the Econometric Society. Cambridge, England, September 1970.
9a. Hickman, Bert G., ed., *Econometric Models of Cyclical Behavior*. Studies in Income and Wealth, Vol. 36. New York, NBER, 1971.
10. Howrey, Philip E., Lawrence R. Klein, and Michael D. McCarthy, "Notes on Testing the Predictive Performance of Econometric Models." Discussion Paper No. 173, Wharton School, Department of Economics, University of Pennsylvania, Philadelphia, 1970.
11. Lewis, John P., and Robert C. Turner, *Business Conditions Analysis*, 2nd ed. New York, McGraw-Hill, 1967.
12. Mincer, Jacob, ed., *Economic Forecasts and Expectations: Analyses of Forecasting Behavior and Performance*. New York, NBER, 1969.
13. Mincer, Jacob, "Models of Adaptive Forecasting." In [12].
14. Mincer, Jacob, and Victor Zarnowitz, "The Evaluation of Economic Forecasts." In [12].
15. Meyer, John R., and Robert R. Glauber, *Investment Decisions, Economic Forecasting and Public Policy*. Boston, Graduate School of Business Administration, Harvard University, 1964.
16. Moore, Geoffrey H., and Julius Shiskin, *Indicators of Business Expansions and Contractions*. New York, NBER, 1967.
17. Samuelson, Paul A., "Economic Forecasting and Science." *Michigan Quarterly Review*, October 1965.
18. Theil, Henri, *Applied Economic Forecasting*. Chicago, Rand-McNally, 1966.
19. Zarnowitz, Victor, *An Appraisal of Short-Term Economic Forecasts*. New York, NBER, 1967.
20. ———, "Prediction and Forecasting, Economic." *International Encyclopedia of the Social Sciences*, 1968, Vol. 12, pp. 425-39.
21. ———, "The New ASA-NBER Survey of Forecasts by Economic Statisticians." *American Statistician*, February 1969, pp. 12-16.
22. ———, "The ASA-NBER Quarterly Survey of the Economic Outlook: An Early Appraisal." *1969 Proceedings of the Business and Economic Statistics Section, American Statistical Association*, Washington, D. C., 1969, pp. 241-49.
23. ———, "Forecasting Economic Conditions: The Record and the Prospect." *The Business Cycle Today*. New York, NBER, forthcoming.
24. Zarnowitz, Victor, Charlotte Boschan, and Geoffrey H. Moore, with the assistance of Josephine Su, "Business-Cycle Analysis of Econometric Model Simulations." In [9a].