(Information for the public)

The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel, 2000

James Heckman and **Daniel McFadden** have each developed theory and methods that are widely used in the statistical analysis of individual and household behavior, within economics as well as other social sciences.

Microeconometrics and Microdata

Microeconometrics is an interface between economics and statistics. It encompasses economic theory and statistical methods used to analyze microdata, i.e., economic information about individuals, households and firms. Microdata appear as cross-section data which refer to conditions at the same point in time, or as longitudinal data (panel data) which refer to the same observational units over a succession of years. During the last three decades, the field of microeconometrics has expanded rapidly due to the creation of large databases containing microdata.

Greater availability of microdata and increasingly powerful computers have opened up entirely new possibilities of empirically testing microeconomic theory. Researchers have been able to examine many new issues at the individual level. For example: what factors determine whether an individual decides to work and, if so, how many hours? How do economic incentives affect individual choices regarding education, occupation or place of residence? What are the effects of different labor-market and educational programs on an individual's income and employment?

The use of microdata has also given rise to new statistical problems, owing primarily to the limitations inherent in such (non-experimental) data. Since the researcher can only observe certain variables for particular individuals or households, a sample might not be random and thereby not representative. Even when samples are representative, some characteristics that affect individuals' behavior remain unobservable, which makes it difficult, or impossible, to explain some of the variation among individuals.

This year's laureates have each shown how one can resolve some fundamental statistical problems associated with the analysis of microdata. James Heckman's and Daniel McFadden's methodological contributions share a solid foundation in economic theory. They emerged in close interaction with applied empirical studies, where new databases served as a definitive prerequisite. The microeconometric methods developed by Heckman and McFadden are now part of the standard tool kit, not only of economists, but also of other social scientists.
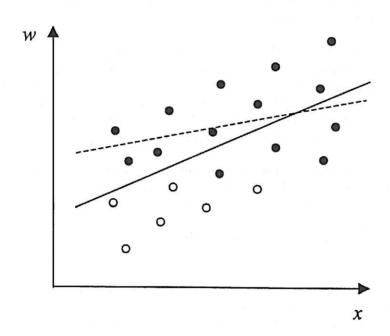
James J. Heckman

James Heckman has made many significant contributions to microeconometric theory and methodology, with different kinds of selection problems as a common denominator. He developed his methodological contributions in conjunction with applied empirical research, particularly in labor economics. Heckman's analysis of selection problems in microeconometric research has had profound implications for applied research in economics as well as in other social sciences.

Selection Bias and Self-selection
Selection problems are legion in microeconometric studies. They can arise when a sample available to researchers does not randomly represent the underlying population. Selective samples may be the result of rules governing collection of data or the outcome of economic agents' own behavior. The latter situation is known as self-selection. For example, wages and working hours can only be observed in the case of individuals who have chosen to work; the earnings of university graduates can only be observed for those who have completed their university education, etc. The absence of information regarding the wage an individual would earn, had he or she chosen otherwise, creates problems in many empirical studies.

The problem of selection bias may be illustrated by the following figure, where w denotes an individual's wage and x is a factor that affects this wage, such as the individual's level of education. Each point in the figure represents individuals with the same education and wage levels in a large and representative sample of the population. The solid line shows the statistical (and true) relationship that we would estimate if we could indeed observe wages and education for all these individuals. Now assume – in accordance with economic theory – that only those individuals whose market wages exceed some threshold value (the reservation wage) choose to work. If this is the case, individuals with relatively high wages and relatively long education will be overrepresented in the sample we actually observe: the dark points in the figure. This selective sample creates a problem of selection bias in the sense that we will estimate the relation between wage and education given by the dashed line in the figure. We thus find a relationship weaker than the true one, thereby underestimating the effect of education on wages.

Heckman's Contributions
Heckman's methodological breakthroughs regarding self-selection took place in the mid-1970s. They are closely related to his studies of individuals' decisions about their labor-force participation and hours worked. As we observe variations in hours of work solely among those who have chosen to work, we could – again – encounter samples tainted by self-selection. In an article on the labor supply of married women, published in 1974, Heckman devised an econometric method to handle such self-selection problems. This study is an excellent illustration of how microeconomic theory can be combined with microeconometric methods to clarify an important research topic.

In subsequent work, Heckman proposed yet another method for handling self-selection: the well-known Heckman correction (the two-stage method, Heckman's lambda or the Heckit method). This method has had a vast impact because it is so easy to apply. Suppose that a researcher – as in the example above – wants to estimate a wage relation using individual data, but only has access to wage observations for those who work. The Heckman correction takes place in two stages. First, the researcher formulates a model, based on economic theory, for the probability of working. Statistical estimation of the model yields results that can be used to predict this probability for each individual. In the second stage, the researcher corrects for self-selection by incorporating these predicted individual probabilities as an additional explanatory variable, along with education, age, etc. The wage relation can then be estimated in a statistically appropriate way.

Heckman's achievements have generated a large number of empirical applications in economics as well as in other social sciences. The original method has subsequently been generalized, by Heckman and by others.

Duration Models
Duration models have a long tradition in the engineering and medical sciences. They are frequently used by social scientists, such as demographers, to study mortality, fertility and migration. Economists apply them, for instance, to examine the effects of the duration of unemployment on the probability of getting a job. A common problem in such studies is that individuals with poor labor-market prospects might be overrepresented among those who remain unemployed. Such selection bias gives rise to problems similar to those encountered in self-selected samples: when the sample of unemployed at a given point in time is affected by unobserved individual characteristics, we may obtain misleading estimates of "duration dependence" in unemployment. In collaboration with Burton Singer, Heckman has developed econometric methods for resolving such problems. Today, this methodology is widely used throughout the social sciences.

Evaluation of Active Labor Market Programs
Along with the spread of active labor-market policy – such as labor-market training or employment subsidies – in many countries, there is a growing need to evaluate these programs. The classical approach is to determine how participation in a specific program affects individual earnings or employment, compared to a situation where the individual did *not* participate. Since the same individual cannot be observed in two situations simultaneously, information about non-participation has to be used, thereby – once again – giving rise to selection problems. Heckman is the world's leading researcher on microeconometric evaluation of labor-market programs. In collaboration with various colleagues, he has extensively analyzed the properties of alternative non-experimental evaluation methods and has explored their relation to experimental methods. Heckman has also offered numerous empirical results of his own. Even though results vary a great deal across programs and participants, the results are often quite pessimistic: many programs have only had small positive – and sometimes negative – effects for the participants and do not meet the criterion of social efficiency.

Daniel L. McFadden

Daniel McFadden's most significant contribution is his development of the economic theory and econometric methodology for analysis of discrete choice, i.e., choice among a finite set of decision alternatives. A recurring theme in McFadden's research is his ability to combine economic theory, statistical methods and empirical applications, where his ultimate goal has often been a desire to resolve social problems.

Discrete Choice Analysis
Microdata often reflect discrete choices. In a database, information about individuals' occupation, place of residence, or travel mode reflects the choices they have made among a limited number of alternatives. In economic theory, traditional demand analysis presupposes that individual choice be represented by a continuous variable, thereby rendering it inappropriate for studying discrete choice behavior. Prior to

McFadden's prizewinning achievements, empirical studies of such choices lacked a foundation in economic theory.

McFadden's Contributions

McFadden's theory of discrete choice emanates from microeconomic theory, according to which each individual chooses a specific alternative that maximizes his utility. However, as the researcher cannot observe *all* the factors affecting individual choices, he perceives a random variation across individuals with the same observed characteristics. On the basis of his new theory, McFadden developed microeconometric models that can be used, for example, to predict the share of a population that will choose different alternatives.

McFadden's seminal contribution is his development of so-called *conditional logit analysis* in 1974. In order to describe this model, suppose that each individual in a population faces a number (say, J) of alternatives. Let X denote the characteristics associated with each alternative and Z the characteristics of the individuals that the researcher can observe in his data. In a study of the choice of travel mode, for instance, where the alternatives may be car, bus or subway, X would then include information about time and costs, while Z might cover data on age, income and education. But differences among individuals and alternatives other than X and Z, although unobservable to the researcher, also determine an individual's utility-maximizing choice. Such characteristics are represented by random "error terms". McFadden assumed that these random errors have a specific statistical distribution (termed an extreme value distribution) in the population. Under these conditions (plus some technical assumptions), he demonstrated that the probability that individual i will choose alternative j can be written as:

$$P_{ij} = \frac{e^{X_{ij}\beta + Z_i\delta_j}}{\sum_{k=1}^{J} e^{X_{ik}\beta + Z_i\delta_k}}.$$

In this so-called multinomial logit model, e is the base of the natural logarithm, while β and δ are (vectors of) parameters. In his database, the researcher can observe the variables X and Z, as well as the alternative the individual in fact chooses. As a result, he is able to estimate the parameters β and δ using well-known statistical methods. Even though logit models had been around for some time, McFadden's derivation of the model was entirely new and was immediately recognized as a fundamental breakthrough.

Such models are highly useful and are routinely applied in studies of urban travel demand. They can thus be used in traffic planning to examine the effects of policy measures as well as other social and/or environmental changes. For example, these models can explain how changes in price, improved accessibility or shifts in the demographic composition of the population affect the shares of travel using alternative means of transportation. The models are also relevant in numerous other areas, such as in studies of the choice of dwelling, place of residence, and education. McFadden has applied his own methods to analyze a number of social issues, such as the demand for residential energy, telephone services and housing for the elderly.

Methodological Elaboration

Conditional logit models have the peculiar property that the relative probabilities of choosing between two alternatives, say, travel by bus or car, are independent of the price and quality of other transportation options. This property – called *independence of irrelevant alternatives* (IIA) – is unrealistic in certain applications. McFadden not only devised statistical tests to ascertain whether IIA is satisfied, but also introduced more general models, such as the so-called nested logit model. Here, it is assumed that individuals' choices can be ordered in a specific sequence. For instance, when studying decisions regarding place of residence and type of housing, an individual is assumed to begin by choosing the location and then the type of dwelling.

Even with these generalizations, the models are sensitive to the specific assumptions about the distribution of unobserved characteristics in the population. Over the last decade, McFadden has elaborated on simulation models (*the method of simulated moments*) for statistical estimation of discrete choice models allowing much more general assumptions. Increasingly powerful computers have enhanced the practical applicability of these numerical methods. As a result, individuals' discrete choices can now be portrayed with greater realism and their decisions predicted more accurately.

Other Contributions

In addition to discrete choice analysis, McFadden has made influential contributions in several other fields. In the 1960s, he devised econometric methods to assess production technologies and examine the factors behind firms' demand for capital and labor. During the 1990s, McFadden contributed to environmental economics, in particular to the literature on contingent-valuation methods for estimating the value of natural resources. A key example is his study of welfare losses due to the environmental damage along the Alaskan coast caused by the oil spill from the tanker Exxon Valdez in 1989. This study provides yet another example of McFadden's masterly skill in integrating economic theory and econometric methodology in empirical studies of important social problems.

****

James J. Heckman was born in Chicago, IL in 1944. After completing his undergraduate education at Colorado College, having majored in Mathematics, he attended Princeton University for graduate studies in Economics and received his Ph.D. there in 1971. Since then, Heckman has held professorships at Columbia University and Yale University. Since 1995, he is Henry Schultz Distinguished Professor of Economics at the University of Chicago.

James Heckman
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
USA
http://lily.src.uchicago.edu

Daniel L. McFadden was born in Raleigh, NC in 1937. He attended the University of Minnesota, where he received both his undergraduate degree, with a major in Physics and, after postgraduate studies in Economics, his Ph.D. in 1962. McFadden has held professorships at the University of Pittsburgh, Yale University and MIT. Since 1990, he is E. Morris Cox Professor of Economics at the University of California, Berkeley.

Daniel McFadden
Department of Economics
University of California
Berkeley, CA 94720
USA
http://emlab.berkeley.edu/users/mcfadden/index.html

Suggested Reading

Advanced information on the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel, 2000, The Royal Swedish Academy of Sciences, http://www.nobel.se/announcement/2000

Amemiya, T. (1987), Discrete Choice Analysis, in P. Newman, M. Milgate and J. Eatwell (eds.), *The New Palgrave – A Dictionary of Economics*, Macmillan.

Heckman, J.J. (1987), Selection Bias and Self-Selection, in P. Newman, M. Milgate and J. Eatwell (eds.), *The New Palgrave – A Dictionary of Economics*, Macmillan.

Heckman, J.J. and J. Smith (1995), Assessing the Case for Social Experiments, *Journal of Economic Perspectives 9*, 85-110.

Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.

McFadden, D.L. (2000), Disaggregate Travel Demand's RUM Side: A 30-Year Retrospective, manuscript, Department of Economics, University of California, Berkeley (http://emlab.berkeley.edu/users/mcfadden/index.html).

Small, K.A. (1992), *Urban Transportation Economics*, Fundamentals of Pure and Applied Economics Series, Harwood Academic Publishers.