MICRODATA, HETEROGENEITY AND THE EVALUATION OF PUBLIC POLICY

Prize Lecture, December 8, 2000

by

JAMES J. HECKMAN[1]

Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, USA.

ABSTRACT

This paper summarizes the contributions of microeconometrics to economic knowledge. Four main themes are developed. (1) Microeconometricians developed new tools to respond to econometric problems raised by the analysis of the new sources of microdata produced after the Second World War. (2) Microeconometrics improved on aggregate time series methods by building models that linked economic models for individuals to data on individual behaviour. (3) An important empirical regularity detected by the field is the diversity and heterogeneity of behaviour. This heterogeneity has profound consequences for economic theory and for econometric practice. (4) Microeconometrics has contributed substantially to the scientific evaluation of public policy.

On behalf of all economists who analyze microeconomic data and who use microeconometrics to unite theory and evidence and to evaluate policy interventions of all kinds, I accept the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel.

The field of microeconometrics emerged in the past forty years to aid economists in providing more accurate descriptions of the economy, in designing and evaluating public policies and in testing economic theories and estimating the parameters of well posed economic models. It is a scientific field within economics that links the theory of individual behaviour to individual data where individuals may be firms, persons or households. Research in microeconometrics is data driven. The availability of new forms of data has raised challenges and opportunities that have stimulated all of the important developments in the field and have changed the way economists think about economic reality. Research in the field is also policy driven. Questions of eco-

[1] I am grateful to Jaap Abbring, Pedro Carneiro, Lars Hansen, Steve Levitt, Costas Meghir, Robert Moffitt, Jeffrey Smith and Edward Vytlacil for helpful comments. This research was supported by NSF 97-09-873 and NICHD-40-4043-000-85-261 and grants from the American Bar Foundation. I thank the National Science Foundation, the National Institutes of Health and the American Bar Foundation for their support over the years.

nomic policy that can be addressed with data motivate much of the research in this field. Research questions in this field are also motivated by the desire to test and implement new economic models.

In this lecture, I discuss four main themes in microeconometrics – a field that has been recognized by the Nobel committee for the first time this year. The first theme is that the post World War II development of rich new data on individuals and firms gave economists a deeper understanding of the economy. At the same time, it confronted econometricians with a host of unsolved problems that could not be adequately addressed with methods developed in traditional Cowles Commission simultaneous equations econometrics. Developments in microeconometrics have been stimulated by empirical problems that arise in analyzing economic data.

The second theme is closely related to the first. Microeconometrics grew out of Cowles econometrics in response to its perceived empirical failures. Cowles econometrics was aggregative in character and was first applied on a wide scale to economic time series. Many of the Cowles econometric models were not clearly motivated as solutions to precisely formulated individual decision problems. Even when they were, the literature on the aggregation problem in econometrics formally established that the link between the decision maker and the aggregate data used to estimate the models was not clear. Microeconometrics developed precisely formulated models of individual behavior and estimated models on individual data. The link between theory and data became much closer.

The third theme of this lecture is that a number of important empirical discoveries have emerged from microeconometric investigations. The most important discovery was the evidence on the pervasiveness of heterogeneity and diversity in economic life. When a full analysis of heterogeneity in responses was made, a variety of candidate averages emerged to describe the "average" person, and the long standing edifice of the representative consumer was shown to lack empirical support. This changed the way economists think about econometric models and policy evaluation. A new model of microeconomic phenomena emerged. In the context of regression analysis not only were intercepts variable but so were the slope coefficients and both slopes and intercepts could be correlated with regressors. Accounting for heterogeneity and diversity and its implications for economics and econometrics is a central message of this lecture and a main theme of my life's work.

The fourth theme of my lecture is that microeconometrics has contributed substantially to scientific policy evaluation based on econometric models which has always been a central problem in econometrics. As difficulties in identifying structural parameters became evident – whether in macro or micro data – microeconomists, following important suggestions by Marschak (1953) and Hurwicz (1962), began to ask whether it was necessary to recover complete structural models to answer specific policy questions in a principled way. This gave rise to a new emphasis on problem – specific parameters, or "treatment effects," which in general are distinct from structural parameters. These parameters answer more limited economic questions but are more easily identified or bounded. Understanding the advantages and limitations of these "treatment effects" and relating them to the structural parameters of the older literature is a recent advance.

1. MICROECONOMETRICS: ORIGINS AND A DEFINITION

Econometrics is a branch of economics that unites economic theory with statistical methods to interpret economic data and to design and evaluate social policies. Economic theory plays an integral role in the application of econometric methods because the data do not speak for themselves on many questions of interpretation. Econometrics uses economic theory to guide the construction of counterfactuals and to provide discipline on empirical research in economics.

The production of a large data base that can be used to describe the economy, to test theories about it and to evaluate public policy is a major development of Twentieth Century economics. Prior to the Twentieth Century, economics was largely a deductive discipline that drew on anecdotal observations and on introspection to test theory and evaluate public policies.

Alfred Marshall's theoretically fruitful notion of the "Representative Firm" and the "Representative Consumer" was firmly rooted in economic theory by the time economists began the systematic collection and analysis of aggregate economic data. The early econometricians focused on aggregate data to measure business cycles and to build models that could be the basis for an empirically based approach to macro policy evaluation. Using linear equation systems, these scholars developed a framework for analyzing causal models and producing policy counterfactuals. For the first time, causation was distinguished from correlation in a formally precise way that could be empirically implemented.

Despite these substantial intellectual contributions, empirical results from these methods proved to be disappointing. Almost from the outset, aggregate time series data were perceived to be weak and empirical macro models were perceived as ineffective in testing theories and producing policy advice (see Mary Morgan, 1990). With a few notable exceptions, macroeconometricians turned to using statistical time series methods where the link between the statistical model and economic theory was usually weak.[2]

Early on, Orcutt (1962) advocated a program of combining micro and macrodata to produce a more credible description of economic phenomena and to test alternative economic theories. At the time he set forth his views, the microdata base was small, computers had limited power and a whole host of econometric problems that arose in using microdata to estimate behavioral relationships were not understood, much less solved. Nonetheless, Orcutt's vision was a bold one and he helped set into motion the forces that produced modern microeconometrics.

[2] See however the important work of Hansen and Sargent (1980, 1991), Hansen and Singleton (1982) and Fair (1976, 1994) which constitutes an exception to this rule. Heckman (2000) discusses this development. These authors formulate well posed decision problems for individuals in deriving their estimating equations, but apply them to aggregate data.

Microeconometrics extended the Cowles theory by building richer economic models where heterogeneity of agents plays a fundamental role and where the equations being estimated are more closely linked to individual data and individual choice models. At its heart, economic theory is about individuals and their interactions in markets or other social settings. The data needed to test the micro theory are microdata. The econometric literature on the aggregation problem (Theil, 1954; see, *e.g.* Green, 1964, or Fisher, 1969 for surveys) demonstrated the fragility of aggregate data for inferring either the size or the sign of micro relationships. In the end, this literature produced negative results and demonstrated the importance of using microdata as the building block of an empirically based economic science. It provided a major motivation for the collection and analysis of microeconomic data.

Another motivation was the growth of the modern welfare state and the ensuing demand for information about the characterization, causation, and solutions to social problems and the public demand for the objective evaluation of social programs directed toward specific groups. Application of the principles of the Cowles paradigm and its extensions by Theil (1961), gave rise to a demand for structural estimation based on micro data. In the optimistic era of the 1960s and 1970s, estimation of policy-invariant structural parameters on micro data became a central goal of policy oriented econometric analysis to consider the effects of old policies in new environments and to consider the possible effects of new policies never tried.

Another use for structural models independent of interest in policy analysis was to test economic theory. Labor economics in particular had been enriched by the application of neoclassical theory to the labor market. This demand was further fueled by the emergence of a micro theory-based macroeconomics. The numerical magnitudes of individual level preference and production parameters played a crucial role in macro theory and macro policy debates.

Another demand for structural estimation arose from the need to synthesize and interpret the flood of microdata that began to pour into economics in the mid 1950s. The advent of micro surveys coupled with the introduction of the computer and the development and dissemination of multiple regression methods by Goldberger (1964) and Theil (1961, 1971) made it possible to produce hundreds, if not thousands, of regressions quickly. The resulting flood of numbers was difficult to interpret or to use to test theories or create an informed policy consensus. A demand for low dimensional economically interpretable models to summarize the growing mountains of micro data was created, and there was increasing recognition that standard regression methods did not capture all of the features of the data nor did they provide a framework for interpreting the data within well-posed economic models.

Before turning to specific developments in the field, it is useful to consider two distinct policy evaluation questions which differ greatly in the data and assumptions required to answer them. The evolution of microeconometrics in the past thirty years can be described as moving from answering the harder structural questions to the relatively easier treatment effect questions.

2. ECONOMIC POLICY, ECONOMIC MODELS AND ECONOMETRIC POLICY EVALUATION

Two conceptually distinct policy evaluation questions are often confused. Their careful separation is a major development in microeconomics and is a major theme of this lecture.

Those questions are

(1) "What is the effect of a program in place on participants and nonparticipants compared to no program at all or some alternative program?"

This is what is now called the "treatment effect" problem. The second and the more ambitious question raised is

(2) "What is the likely effect of a new program or an old program applied to a new environment?"

The second question raises the same type of problems as arise from estimating the demand for a new good.[3] Its answer usually requires structural estimation.

It is easier to answer the first question than the second although the early literature attempted to answer both by estimating structural models. A major development in policy evaluation research to which I have contributed has been clarification of the conditions that must be satisfied to answer both types of questions, and other related questions.

The goal of structural econometric estimation is to provide the ingredients to solve a variety of decision problems. Those decision problems entail such distinct tasks as (a) evaluating the effectiveness of an existing policy, (b) projecting the likely effectiveness of a policy in different environments from the one where it was experienced, or (c) forecasting the effects of a new policy, never previously experienced.[4] In this lecture I only consider decision problems that arise in policy analysis.

Additional benefits of structural models are that they can be used to test economic theory and make quantitative statements about the relative importance of causes within a theory. In addition, structural models based on invariant parameters can be compared across empirical studies. Empirical knowledge can be cumulated within structural frameworks. However, for certain important classes of decision problems, knowledge of all or even any

[3] This question is discussed in basic papers by Gorman (1980, first written in 1956), Quandt (1970), Lancaster (1966, 1971), McFadden (1974) and Domencich and McFadden (1975) among others.

[4] Marschak (1953) stressed these features of structural estimation. Similar issues arise in estimating the demand for new goods. Structural methods can be used to estimate the parameters of demand equations in a given economic environment, in forecasting the demand for goods in a different environment, and in forecasting the demand for a new good never previously consumed. Knowledge of the parameters of demand functions is crucial in testing alternative theories of consumer demand and measuring the strength of complementaries and substitution among goods.

structural parameters of a model is unnecessary. This is fortunate because recovering structural parameters is usually not an easy task.

In the recent literature on policy evaluation, the implicit goal has been to recover the ingredients of models required to solve more specific decision problems. This may entail knowing only combinations of structural parameters, or parameters that are not structural in any conventional sense of that term. Thus the modern treatment effect literature in economics takes as its main goal the estimation of treatment effects – not the full range of parameters pursued in structural econometrics – although the precise questions being answered in particular studies are often not clearly stated. These treatment effects are identified under weaker conditions than are required for recovering all of the structural parameters of the model. The Cowles distinctions between endogenous and exogenous variables and the later distinctions of weak, strong and superexogeneity developed in the literature on estimating structural parameters and policy forecasting (Engle, Hendry and Richard, 1983) are largely irrelevant in identifying certain widely used treatment parameters. By focusing on one particular decision problem, the treatment effect literature achieves its objectives under weaker and hence more credible conditions than are invoked in the structural econometrics literature. At the same time, the parameters so generated are less readily transported to different environments to estimate the effects of the same policy in a different setting or the effects of a new policy and they are difficult to compare across studies. The treatment effect literature has to be extended to make such projections and comparisons and, unsurprisingly, the required extensions are nonparametric versions of the assumptions used by structural econometricians.[5]

To make this discussion specific, but at the same time keep it simple, consider the prototypical problem of determining the impact of taxes and welfare payments on labor supply. This problem motivated the early literature in evaluating the welfare state, (Cain and Watts, 1973) motivated my own research, and remains an important policy problem down to this day.

Following the conventional theory of consumer demand, write an interior solution labor supply equation of hours of work H in terms of wages, W, and other variables including assets, demographic structure and the like. Denote these other variables by X. Let U denote an unobservable from the point of view of the observing economist. As we shall see, unobservables play a big role in microeconometrics. There is much evidence that unobservables are empirically important. Modern microeconometrics is devoted to accounting for them.

In the most general form for H,

$$H = \phi(W, X, U). \tag{1}$$

Assume for simplicity that ϕ is differentiable in all of its arguments. Equation

[5] This point is developed more fully in Heckman, LaLonde and Smith (1999) and Heckman and Vytlacil (2001a, 2001d, 2002).

(1) is a Marshallian causal function.[6] Its derivatives produce the *ceteris paribus* effect of a change in the argument being varied on H. Suppose that we wish to evaluate the effect of a change in a proportional wage tax on labor supply. Proportional wage taxes at rate t make the after tax wage $W(1-t)$. Assume that agents correctly perceive the tax and ignore any general equilibrium effects of the tax. In the language of treatment effects, the treatment effect or "causal effect" of a tax change on labor supply defined at the individual level is $\phi(W(1-t),X,U) - \phi(W(1-t'),X,U)$ for the same person subject to two different taxes, t and t'.

An additively separable version of the Marshallian causal function (1) is

$$H = \phi(W, X) + U, \; E(U) = 0. \tag{2}$$

This version enables the analyst to define the *ceteris paribus* effects of W and X on H without having to know the level of the unknown (to the econometrician) unobservable U. A parametric version of (1) is

$$H = \phi(W, X, U, \theta) \tag{1(a)}$$

where θ is a low dimensional parameter that generates the ϕ of equation (1). A parametric version of (2) is

$$H = \phi(W, X, \theta) + U. \tag{2(a)}$$

The parameters θ reduce the dimensionality of the identification problem from that of identifying an infinite-dimensional function to that of identifying a finite set of parameters. They play a crucial role in forecasting the effects of an old policy in different populations, in cumulating evidence across studies and in forecasting the effects of a new policy. A-linear- in-parameters representation of H writes:

$$H = \alpha' X + \beta \ln W + U \tag{3}$$

where we adopt a semi-log specification to represent models widely used in the literature on labor supply. (See Killingsworth, 1983.)

Following Marschak (1953), it is useful to distinguish three different policy evaluation problems. A tax is externally imposed on a population or a subpopulation of the economy. (Thus the tax is determined independently of U, but it may depend on X and W, variables which we observe and on which we can condition.) (1)The case where tax t has been implemented in the past and we wish to forecast the effects of the tax in a population with the same distribution of (W, X, U) variables as prevailed when historical measurements of tax variation were made. (2) A second case where tax t has been implemented in the past but we wish to project the effects of the same tax to a different population of (W, X, U) variables. (3) A case where the tax has never been implemented and we wish to forecast the effect of a tax either on an initial population used to estimate (1) or on a different population.

Suppose that the goal of the analysis is to determine the effect of taxes on

[6] See Heckman (2000a) or Heckman and Vytlacil (2001a, 2002) for a rigorous definition of Marshallian causal functions.

average labor supply on a relevant population with distribution $G(W, X, U)$. In case 1, we have data from the same population for which we wish to construct a forecast. Suppose we observe different tax regimes. Persons face externally imposed tax rate t_j in regime j, $j = 1, ..., J$. In the sample from each regime we can identify

$$(4) \qquad E(H \mid W, X, t_j) = \int \phi(W(1-t_j), X, U)dG(U \mid X, W).$$

For the entire population this function is

$$(5) \qquad E(H \mid t_j) = \int \phi(W(1-t_j), X, U)dG(U, X, W).$$

This function is assumed to apply to the target population of interest. Knowledge of (4) or (5) from the historical data can be projected into all future periods provided the joint distributions of data are temporally invariant. If one regime has been experienced in the past, lessons from it apply to the future, provided that the same $\phi\,(\cdot)$ and $G(\cdot)$ prevail. No explicit counterfactual state need be constructed. No knowledge of any Marshallian causal function or structural parameter is required to do policy analysis for case one. It is not necessary to break apart (4) or (5) to isolate ϕ from G.[7]

Case two resembles case one except for one crucial difference. Because we are now projecting the same policy onto a different population, it is necessary to break (4) or (5) into its components and determine $\phi(W(1-t_j),X,U)$ separately from $G(U, X, W)$.The problem of policy evaluation becomes much harder. A quotation from Frank Knight (1921) is apt:

The existence of a problem in knowledge depends on the future being different from the past, while the possibility of a solution of the problem depends on the future being like the past. (Knight, 1921, p. 313).

The assumptions required to project the effects of the old policy in a new regime require that we borrow from the past to determine the components of (4) or (5) on new populations.

Those assumptions are:

(a) Knowledge of $\phi(\cdot)$ is needed for the new population. This may entail determination of ϕ on a different support from that used to determine ϕ in an initial sample if the target population has a different support than the original source population. At this stage, structural estimation comes into its own. It sometimes enables us to extrapolate ϕ from a source population to a target population. A completely nonparametric solution to this problem is impossible even if we adopt structural additive separability assumption (2a) unless the supports of target and source populations coincide.

[7] It is not even required that t be externally specified. If a policy setting function $t = \eta(X, W, U)$ generates t, and is 1–1 in U and t given (X, W), then each t is associated with a unique U given (X, W). Provided that the goal of the analysis is to forecast the effects of future t generated by η, we can use historical data to do so. If is not 1–1 in (U,t) given (X, W), then it is not possible, in general, to use historical data to predict the effect of t variation generated by ϕ on mean H. If, however, the goal is to forecast policies generated by a new rule (including external variations of t unrelated to U), then case one no longer is relevant, and it is necessary to do structural estimation (Lucas, 1976).

Some structure must be placed on ϕ even if (2a) characterizes the labor supply model. Parametric structure (3) is traditional in the labor supply literature and versions of a linear in parameters model dominate applied econometric research.[8]

(b) Knowledge of $G(\cdot)$ for the target population is also required. In this context, exogeneity enters as a crucial facilitating assumption. If we define exogeneity by

$$(X, W) \perp\!\!\!\perp U \tag{A-1}$$

then

$$G(U \mid X, W) = G(U).^{9}$$

In this case, if we assume that the distribution of unobservables is the same in the sample as in the forecast or target regime, $G(U) = G'(U)$, where $G'(U)$ is the distribution of unobservables in the target population, we can project to a new population using the relationship

$$E(H \mid W, X, t_j) = \int \phi(W(1-t_j), X, U) dG(U) \tag{6}$$

provided we can determine $\phi(\cdot)$ over the new support of X, W, U. If, however, $G' \neq G$, G' must somehow be determined. This entails invoking some structural assumptions to determine the relationship between G and G'.

In the third case, where no tax has previously been introduced, knowledge of the target population is required. Taxes operate through the term $W(1 - t)$. If there is no wage variation in samples extracted from the past, there is no way to identify the effect of taxes on labor supply since by assumption $t = 0$, and it is not possible to determine the effect of the first argument on labor supply. The problem is only worse if we assume that taxes operate on labor supply independently of wages. Then, even if there is wage variation, it is impossible to identify tax effects or to project them to a new population.[10]

The preceding discussion applies with equal force to analyses of aggregate

[8] The assumption that $\phi(W, X)$ is real analytic so that it can be extended to other domains is another structural assumption. This assumption is exploited in Heckman and Singer (1984) to solve a censoring problem in duration analysis.

[9] There are many definitions of this term. Assumption (A-1) is often supplemented by the additional assumption that the distribution of X does not depend on the parameters of the model (*e.g.* θ in (1-a) or (2-a)). See Engle, Hendry and Richard (1983).

[10] If wages vary in the prepolicy period, it may not be necessary to decompose (4) into ϕ and G, or to do structural estimation, in order to estimate the effect of taxes on labor supply in a regime that introduces taxes for the first time. If the support of $W(1-t) \stackrel{\text{def}}{=} W^*$ in the target regime is contained in the support of W in the historical regime, the supports of the X are the same in both regimes and the conditional distributions of U given X, W and U given X, W^* are the same, then knowledge of (4) over the support of W in the historical or source regime is enough to determine the effect of taxes in the target regime. More precisely, letting "historical" denote the past data, and "target" denote the target population for projection, we may write these assumptions as:

(a) Support $(X, W^*)_{\text{target}} \subseteq$ Support $(X, W)_{\text{historical}}$ (b) $G(U \mid X, W^*)_{\text{target}} = G(U \mid X, W)_{\text{historical}}$ where $W^* = W(1-t)$ for random variables W defined in the new regime and $(W^*)_{\text{target}} = W_{\text{historical}}$.

data and to analyses of microdata. Using individual variation in micro surveys provides a new avenue of identification of ϕ and G not available in macrodata. It thus facilitates identification of structural parameters.

The treatment effect literature extends Marschak's first case by allowing the treatment (t) to be endogenous. Consider two populations. These can be subpopulations of a general population and will be referred to as the treatment group and the comparison group. In one population the tax is t_j and in the other the tax is t_k, which may be no tax at all. If the two populations are identical in terms of ϕ and G, and differ only in an externally imposed tax rate, then it is possible to determine the effect on mean hours of work of t_j relative to tax t_k for either population for any given X,W by simply contrasting mean hours in the two populations, $E(H \mid W,X, t_j) - E(H \mid W,X,t_k)$, over domains of common support for W, X. No knowledge of ϕ or G is required, so no structural estimation is required. Moreover, as previously noted, there are (stringent) conditions under which this exercise is valid even if t is endogenously determined by a stable policy rule provided that the rule is 1–1 in (t and U) for a given X,W.

In the context of the labor supply example, the literature on treatment effects seeks to identify the contrasts in mean hours worked on a given population of (X, W, U) that would arise from different externally imposed policy (t) regimes without decomposing mean hours into ϕ or G components using data from populations where t is not externally specified. Policy experiments (actual or natural) that change t and that do not change ϕ or G identify such effects. Instruments that shift t, keeping ϕ and G invariant, are also used. A variety of methods are used to control for observed and unobserved differences in outcomes across policy regimes that are unrelated to the policy being evaluated. The identifying conditions required to estimate treatment effects are generally weaker than those required to identify ϕ and G in the sense that fewer assumptions are required to identify the treatment effects. At the same time, the estimates produced are very problem specific and apply only to the populations being studied. The treatment effects lack the transportability of ϕ to new environments and the interpretability of ϕ in terms of *ceteris paribus* changes ("causal effects") for all of the conditioning variables except t.

This dualism between treatment effects and structural equations runs throughout the literature and my own work. I return to this theme but first I consider how the availability of microdata provided the impetus for the development of microeconometrics.

(Note 10 continued)

In this case, no structural estimation is required to forecast the effect of taxes on labor supply in the target population. A fully nonparametric policy evaluation is possible estimating (4) or (5) nonparametrically (and not decomposing $E(H \mid X, W)$ into the $\phi(\cdot)$ and $G(\cdot)$ components). Under assumption (a), we may find a counterpart value of $W(1-t) = W^*$ in the target population for each X to insert in the nonparametric version of (4) (or (5)). If these conditions are not met, it is necessary to build up the G and the ϕ functions over the new supports using the appropriate distributions. We enter the realm where structural estimation is required, either to extend the support of the $\phi(\cdot)$ functions or to determine $G(U \mid X,W)$ or both. It is still necessary to determine the relationship between W and X in the target population.

3. NEW FEATURES OF MICRO DATA

The micro data first produced on a large scale in the 1950s revealed patterns and features that were not easily rationalized by standard models of consumer demand and labor supply or that were well modeled by conventional regression analysis. Important dimensions of heterogeneity and diversity that are masked in macro data were uncovered. These findings challenged the standard econometric tool kit of the day.

Inspection of cross section data reveals that otherwise observationally identical people make different choices, earn different wages and hold different levels and compositions of asset portfolios. These data reveal the inadequacy of the traditional representative agent paradigm.[11] Table 1 presents a typical sample of data on labor supply. A considerable fraction of people do not work, and we do not observe wages for nonworkers. The R^2 (measure of explained fit) of any micro relationship is typically low, so the unobservables account for a lot of the variability in hours of work. Different assessments of the unobservables have different effects on the interpretation of the evidence. For example, is joblessness due to unobserved tastes for leisure on the part of workers or a failure of the market to generate wage offers which are only observed if they are accepted?[12] Are all women transients in the labor market or do some women (or most) have a long term attachment to it?[13]

Table 1
Participation, Hours Worked and Wage Data
NLSY Data, 1979-1994

Demographic Group	% Working at Age 29
White Males	83.5%
Black Males	75.0%
Hispanic Males	80.0%
White Females	76.4%
Black Females	69.6%
Hispanic Females	66.6%

R^2 from Regressions

Demographic Group	Total Hours Worked on Education and Experience	Log Wage on Education and Experience
White Males	0.12	0.10
Black Males	0.15	0.14
Hispanic Males	0.11	0.10
White Females	0.15	0.17
Black Females	0.18	0.21
Hispanic Females	0.18	0.10

Source: National Longitudinal Survey of Youth, 1979-1994, as used in Carneiro, Heckman, and Vytlacil (2001).

[11] Lancaster (1966, 1971), Quandt (1970), McFadden (1974) and Domencich and McFadden (1975) were among the first to question the empirical validity of the representative agent empirical paradigm. See Kirman (1992) for a recent assessment of the representative agent paradigm.

[12] Flinn and Heckman (1982) analyze this question and show the difficulty of resolving it using data on market choices.

[13] Heckman and Willis (1977) and Heckman (1981c) analyze this question.

There are additional problems with using these data that are much less apparent in analyses of time series data. Wages are missing for nonworkers. How can one estimate the effect of wages on labor supply if wages are only available for workers? How can one interrelate the various dimensions of labor supply (hours of work, work or not work, number of periods worked) in order to do counterfactual policy analysis?

In confronting the new data, a variety of econometric problems arose: (a) accounting for discreteness of outcome variables; (b) rationalizing choices made at both the extensive and intensive margins (models for discrete choice and for joint discrete and continuous choices) within a common structural model and (c) accounting systematically for missing data where prices or wages are missing *because* of choices made by individuals.

Focusing solely on the statistical aspects of microeconometrics obscures its basic contributions. After all, many statisticians worried about some of these problems. Models for discrete data were analyzed by Goodman, 1968, Haberman, 1974, and Bishop, Fienberg and Holland, 1974, although it was economists who pioneered the study of models with jointly determined discrete and continuous outcomes (Heckman, 1974a,b) and models with systematically missing data (Gronau 1974, Heckman 1974a,b, 1976a,b, 1979).[14] An important contribution of microeconometrics was to clarify the limitations of, and to extend, these statistical frameworks for estimating economic models, making causal distinctions and solving various versions of the policy evaluation problem described in section 2.

Unlike the models developed by statisticians, the class of microeconometric models developed to exploit and interpret the new sources of microdata emphasized the role of economics and causal frameworks in interpreting evidence, in establishing causal relationships and in constructing counterfactuals, whether they were counterfactual missing wages in the analysis of female labor supply or counterfactual policy states that arise in evaluating social policies. Research in microeconometrics demonstrated that it was necessary to be careful in accounting for the sources of manifest differences among apparently similar individuals. Different assumptions about the sources of unobserved heterogeneity have a profound effect on the estimation and economic interpretation of empirical evidence, in evaluating programs in place and in using the data to forecast new policies and assess the effect of transporting existing policies to new environments.

Heterogeneity due to unmeasured variables became an important topic in this literature because its manifestations were so evident in the data and the consequences of ignoring it turned out to be so profound. The problem became even more apparent as panel micro data became available and it was possible to observe persistent differences over time for the same persons.

[14] See Holt (1985) for a discussion of the originality of the work of econometricians in analyzing models for missing data when the missing data are not random (*i.e.* are systematically related to the observables and unobservables of the model).

4. POTENTIAL OUTCOMES, COUNTERFACTUALS AND SELECTION BIAS

My initial efforts in the field of microeconometrics were focused on building models to capture the central features of data like that displayed in Table 1 within well posed choice theoretic models that also could be used to address structural policy evaluation problems (Marschak question 2 problems as defined in subsection 2). I was inspired by the work of Mincer (1962) on female labor supply and was challenged by the opportunity of building a precise econometric framework for analyzing the various dimensions of female labor supply and their relationship with wages. In accomplishing this task, I drew on two sets of econometric tools that were available and my attempts to fuse these tools into a common research instrument produced both frustration and discovery.

The two sets of tools available to me were (1) classical Cowles Commission simultaneous equations theory and (2) models of discrete choice originating in mathematical psychology that were introduced into economics by Quandt (1956, 1970), McFadden (1974, 1981), and Domencich and McFadden (1975). My goal was to unite these two literatures in order to produce an economically motivated, low dimensional, simultaneous equations model with both discrete and continuous endogenous variables that accounted for systematically missing wages for nonworkers and different dimensions of labor supply within a common framework, that could explain female labor supply, and that could be the basis for a rigorous analysis of policies never previously implemented.

The standard model of labor supply embodied in equations (1), (2) or (3) is not adequate to account for the data in Table 1. Neither is Cowles econometrics. Under standard conditions, Cowles methods can account for the correlation between W and U in equation (3) assuming that wages are measured for everyone. Such correlation can arise from measurement error in wages or because of common unobservables in the wage and labor supply equations (*e.g.* more motivated people work more and have higher wages and motivation is not observed). Cowles methods do not tell us what to do when wages are missing, how to account for nonworkers, or how to relate the decision to work with the decision on hours of work.

In a series of papers written in the period 1972–1975 (Heckman 1973, 1974a,b, 1976a,b, 1978a),[15] I developed index models of potential outcomes to unite Cowles econometrics and discrete choice theory as well as to unify the disjointed and scattered literature on sample selection, truncation and limited dependent variables that characterized the literature of the day.[16] I also developed a variety of two stage estimators for this class of models to circumvent computational difficulties associated with estimating these models by the method of maximum likelihood.

[15] The earliest papers were published in 1974 (Heckman 1974a,b) but widely circulated before then. Heckman (1976b, 1978a) was actually written in 1973 and widely circulated at that time to many senior econometricians.

[16] See Heckman and MaCurdy (1985) for a systematic development of index function models.

Following the literature in mathematical psychology and discrete choice as synthesized and extended by McFadden (1974, 1981) define

$$(7) \qquad Y_i = g_i(X_i, U_i) \quad i = 1, \ldots, I$$

as latent random variables reflecting potential outcomes. In the context of discrete choice, the Y_i are latent utilities associated with choice i and they depend on both observed(X_i) and unobserved (U_i) characteristics. These are also called index function models. Within each choice i, the level of the utility may vary. More generally, following the Cowles program, and in particular Haavelmo (1943), (7) may represent any potential outcome, including wages, hours of work and the like. Equations (1) or (7) are Marshallian causal relationships that tell us how hypothetical outcome Y_i varies as the arguments on the right hand side are manipulated holding everything else but the manipulated variable fixed.

Depending on the context, the Y_i may be directly observed or only their manifestations may be observed. In models of discrete choice, the Y_i are never observed but we observe $\underset{i}{\operatorname{argmax}}\{Y_i\}$. In the more general class of models I considered, some of the Y_i can be observed under certain conditions.

To consider these models in the most elementary setting, consider a version with three potential outcome functions. The literature analyzes models with many potential outcomes. Write the potential outcomes in additively separable form as

$$(8) \qquad \begin{aligned} Y_0 &= g_0(X) + U_0 \\ Y_1 &= g_1(X) + U_1 \\ Y_2 &= g_2(X) + U_2. \end{aligned}$$

These are latent variables that may be only imperfectly observed. In the context of the neoclassical theory of labor supply, the theory of search, and the theory of consumer demand, the reservation wage or reservation price at zero hours of work (zero demand for the good) plays a central role. It informs us what price it takes to induce someone to work the first hour or buy the first unit of a good. Denote this potential reservation wage function by Y_0. Let Y_1 be the market wage function – what the market offers. Ignoring any fixed costs of work, a person works ($D = 1$)if

$$(9) \qquad Y_1 \geq Y_0 \Leftrightarrow D = 1.$$

Otherwise the person does not work. Potential hours of work Y_2 are generated from the same preferences that produce the reservation wage function so Y_2 and Y_0 are generated by a common set of parameters. In my 1974a paper, I produced a class of simple tractable functional forms where

$$10(a) \qquad Y_0 = \ln R = \text{log reservation wage}$$

$$10(b) \qquad Y_1 = \ln W = \text{log market wage}$$

$$10(c) \qquad Y_2 = \frac{(\ell nW - \ell nR)}{\gamma} \qquad \gamma > 0$$

and observed hours of work are written as

$$H = Y_2 1(\ell nW \geq \ell nR) = \frac{(\ell nW - \ell nR)}{\gamma} 1(\ell nW \geq \ell nR),$$

where $1(A)$ is an indicator that equals 1 if A is true. Proportional taxes or transfers t introduce another source of variation into these equations so that in place of W one uses the after tax wage $W(1 - t)$. The unobservables U_1 and U_0 account for why otherwise observationally identical people (with the same X) make different choices.[17]

Closely related to this model is the pioneering model of Roy (1951) on self-selection in the labor market that was rediscovered in the 1970s. His model is a version of the model for index functions just presented.[18] From equation (8), Y_0 and Y_1 are potential outcomes and Y_2 is a latent utility.

$$(11) \qquad \begin{aligned} &Y_2 \geq 0 \Leftrightarrow D = 1, \text{ and } Y_1 \text{ observed} \\ &Y_2 < 0 \Leftrightarrow D = 0, \text{ and } Y_0 \text{ observed.} \end{aligned}$$

Thus observed Y is

$$(12) \qquad Y = DY_1 + (1 - D) Y_0.$$

In the original Roy model, $Y_2 = Y_1 - Y_0$. In the Generalized Roy Model, Y_2 is more freely specified but may depend on Y_1 and Y_0.

These models of potential outcomes contain several distinct ideas. (1) As in the Cowles Commission analyses, there is a hypothetical superpopulation of potential outcomes defined by possible values assumed by the Y_j for *ceteris paribus* changes in the X and the U. These are models of Marshallian causal functions usually represented by low dimensional structural models to facilitate forecasting and policy analysis. (2) Unlike the Cowles models, but like the models for discrete choice, some of the latent variables are not observed (*e.g. ln R* is not observed but is sometimes elicited by a questionnaire.) (3)

[17] In Heckman (1974b) I present a more explicit structural model of labor supply, child care and wages that develops, among others things, the first rigorous econometric framework for analyzing the effect of progressive taxes on labor supply and the effect of informal markets on labor supply. In that paper, I characterized preferences by the marginal rate of substitution function and generate Y_0 and Y_2 from the consumer indifference curves and produce Y_2, from a solution of consumer first order conditions and the budget constraint. In that model the unobservables affecting preferences translate into variation across consumers (or workers) in the slopes of indifference curves. Characterizing consumer preferences by the slopes of indifference curves facilitates the analysis of labor supply with linked income tax schedules and provides a more flexible class of preferences than is produced by simple linear or semilog specifications of labor supply equations. The first analysis of progressive taxes for the convex case appears in an appendix to Heckman (1974b). Because of space limitations, the editor, T.W. Schultz, requested a condensed presentation. The full formal analysis was published later in Heckman and MaCurdy (1981, 1985) and Heckman, Killingsworth and MaCurdy (1981a,b). Hausman (1980, 1985) extends this analysis to the nonconvex case.

[18] A.D. Roy develops an economic model of income inequality and sorting but does not consider any of the econometric issues arising from his model. Willis and Rosen (1979) and Lee (1978) are two applications of the Roy model.

Unlike either the Cowles model or the discrete choice model, some of the latent variables are observed, but only as a consequence of choices, *i.e.* they are observed selectively.

Thus we observe ($ln\ W - ln\ R$) up to scale only if $ln\ W \geq ln\ R$ ($D = 1$). We observe wages only if $ln\ W \geq ln\ R$ ($D = 1$). This selective sampling of potential outcomes gives rise to the problem of *selection bias.* We only observe selected subsamples of the latent population variables. In the context of the Roy model, we observe Y_0 or Y_1 but not both.

If there were no unobservables in the model, this selective sampling would not be a cause for any concern. Conditioning on X, we would obtain unbiased or consistent estimators of the missing outcomes for those who do not work using the outcomes of those who do work. Yet the data in Table 1, which are typical, reveal that the observables explain only a small fraction of the variance in virtually all microeconomic variables. It is necessary to account for heterogeneity in preferences and selective sampling on unobservables. As a consequence of selection rule (9), in general the wages and hours we observe are a selected sample of the potential outcomes from the larger population.[19] Accounting for this is a major issue if we seek to estimate structural relationships (the parameters of the causal functions) or describe the world of potential outcomes (the equations such as 10(a)–10(c)). This gives rise to the problem in selection bias. In solving this problem a new analysis of discrete choice and mixed continuous-discrete choice revised conventional Cowles econometrics, and demonstrated the inadequacy of conventional statistical models for discrete data in making causal distinctions. The theory of discrete choice and mixed discrete-continuous choice challenged the received Cowles paradigm by linking econometrics more closely to choice and decision processes. I consider these revisions in Appendix A-1. In brief, log linear models used by statisticians to model discrete data were unable to make the *ceteris paribus* distinctions between true and spurious causality that are required in econometric policy analysis and new conditions for coherence in simultaneous equations models were developed to make models probabilistically and economically well defined. (Heckman, 1976b, 1978a). Amemiya (1985) presents a masterful summary of the main developments in this literature.

5. SELECTION BIAS AND MISSING DATA

Selection bias arises in estimating structural models with partially observed potential outcomes. But the problem of selection bias is more general and can arise when a rule other than simple random sampling is used to sample the underlying population that is the object of interest. The distorted representation of a true population in a sample as a consequence of a sampling

[19] If (9) applies than there must be selection bias in observing wages or reservation wages except for degenerate cases (Heckman, 1993). The selection in potential hours is an immediate consequence of (9) since $D = 1 \Leftrightarrow \ln W - \ln R \geq 0$.

rule is the essence of the selection problem. The identification problem is to recover features of a hypothetical population from an observed sample. (See Figure 1). The hypothetical population can refer to the potential wages of all persons whether or not they work (and wages are observed for them) or to the potential outcomes of any choice problem where only actual choices are observed. Distorting selection rules may arise from decisions of sample survey statisticians, or the economic self selection decisions of the sort previously discussed where, as a consequence of self selection, we only observe subsets of a population of potential outcomes (*e.g.* Y_0 or Y_1 in the Roy model).

A random sample of a population produces a description of the population distribution of characteristics that provides a full enumeration of the models of potential outcomes presented in the previous sections. A sample selected by any rule not equivalent to random sampling produces a description of the population distribution of characteristics that does not accurately describe the true population distribution of characteristics no matter how big the sample size.

Two characterizations of the selection problem are fruitful. The first, which originates in statistics, involves characterizing the sampling rule depicted in Figure 1 as applying a weighting to hypothetical population distributions to produce observed distributions. The second, which originates in econometrics, explicitly treats the selection problem as a missing data problem and, in its essence, uses observables to impute the relevant unobservables.

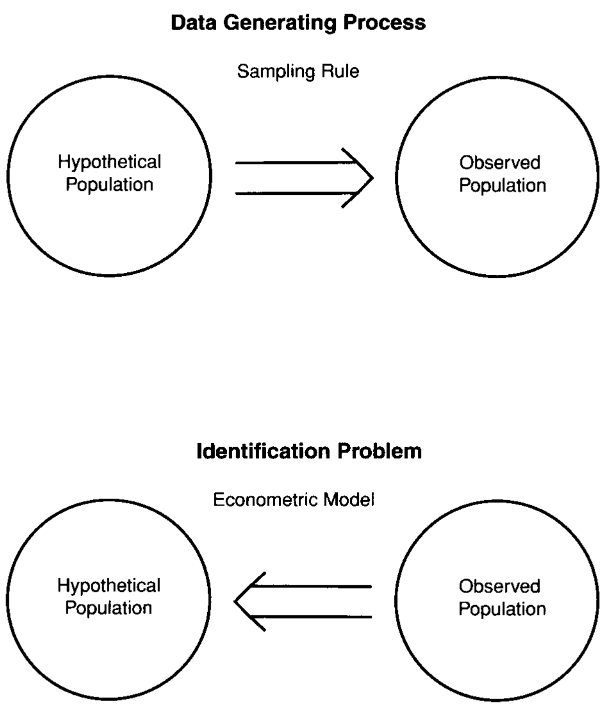

Figure 1. Relationship between Hypothetical (Counterfactual) Population and Observed Data.

(i) Weighted Distributions

Any selection bias model can be described in terms of weighted distributions. Let Y be a vector of outcomes of interest and let X be a vector of "control" or "explanatory" variables. The population distribution of (Y,X) is $F(y,x)$. To simplify the exposition, assume that the density is well defined and write it as $f(y, x)$.

Any sampling rule is equivalent to a non-negative weighting function $\omega(y, x)$ that alters the population density. People are selected into the sampled population by a rule that differs, in general, from random sampling. Let (Y^*,X^*) denote the random variables produced from sampling. The density of the sampled data $g(y^*,x^*)$ may be written as

$$g(y^*,x^*) = \frac{\omega(y^*,x^*)f(y^*,x^*)}{\int \omega(y^*,x^*)f(y^*,x^*)dy^*dx^*} \tag{13}$$

where the denominator of the expression is introduced to make the density $g(y^*,x^*)$ integrate to one as is required for proper densities. Simple random sampling corresponds to the case where $\omega(y, x) = 1$. Sampling schemes for which $\omega(y, x) = 0$ for some values of (Y, X) create special problems because not all values of (Y, X) are sampled.[20]

In many problems in economics, attention focuses on $f(y|x)$, the conditional density of Y given $X = x$. If samples are selected solely on the x variables ("selection on the exogenous variables"), $\omega(y, x) = \omega(x)$ and there is no problem about using selected samples to make valid inference about the population conditional density. Sampling on both y and x is termed *general stratified sampling*, and a variety of different sampling schemes can be characterized by the structure they place on the weights (Heckman, 1987).

From a sample of data, it is not possible to recover the true density $f(y, x)$ without knowledge of the weighting rule. On the other hand, if the weight $\omega(y^*,x^*)$ is known and the support of (y, x) is known and $\omega(y, x)$ is nonzero, then $f(y, x)$ can always be recovered because

$$\frac{g(y^*,x^*)}{\omega(y^*,x^*)} = \frac{f(y^*,x^*)}{\int \omega(y^*,x^*)f(y^*,x^*)dy^*dx^*} \tag{14}$$

and by hypothesis both the numerator and denominator of the left hand side are known, and we know $\int f(y^*,x^*)dy^*dx^* = 1$, so it is possible to determine $\int \omega(y^*,x^*)f(y^*,x^*)dy^*dx^*$. It is fundamentally easier to correct for sampling plans with known non-negative weights or weights that can be estimated separately from the full model than it is to correct for selection where the

[20] For samples in which $\omega(y,x) = 0$ for a non-negligible proportion of the population, it is useful to consider two cases. A *truncated sample* is one for which the probability of observing the sample from the larger random sample is not known. For such a sample, (13) is the density of all the sampled Y and X values. A *censored sample* is one for which the probability is known or can be consistently estimated.

weights are not known, and must be estimated jointly with the model.[21] Choice based sampling, length biased sampling and size biased sampling are examples of the former; sampling arising from selection in the model of equations 10(a)–10(c) or in the generalized Roy model are examples of the latter.

The requirements that (a) the support of (y, x) is known and (b) $\omega(y, x)$ is nonzero are not innocuous. In many important problems in economics, requirement (b) is not satisfied: the sampling rule excludes observations for certain values of (y, x) and hence it is impossible without invoking further assumptions to determine the population distribution of (Y, X) at those values. If neither the support nor the weight is known, it is impossible, without invoking strong assumptions, to determine whether the fact that data are missing at certain (y, x) values is due to the sampling plan or that the population density has no support at those values. Using this framework, Heckman (1987) analyzes a variety of sampling plans of interest in economics, showing what assumptions they make about the weights and the model to solve the inferential problem of going from the observed population to the hypothetical population.

Figures 2(a) and 2(b) illustrate the problem arising from $\omega(x, y) = 0$ in a simple way. In Figure 2(a), I depict a truncated distribution for Y with data missing for values of Y below c. Any shape of the true hyperpopulation density is possible below c. Figure 2(b) shows a regression version of the same problem for a labor supply function H written in terms of wage W. We can fit the regression within the sample, but how do we project it to new samples or to the hypothetical population?

(ii) A Regression Representation of the Selection Problem When There is Selection on Unobservables

A regression version of the selection problem when the weights $\omega(y, x)$ cannot be estimated independently of the model originates in the work of Gronau (1974), Heckman (1976a,b, 1978a, 1979) and Lewis (1974). It starts from the Roy model, using (8), assuming $(U_0, U_1, U_2) \perp\!\!\!\perp X, Z$. It is closely related to Lester Telser's characterization of simultaneous equations bias in a conventional Cowles system.[22] I use Z to denote variables that affect choices

[21] Selection with known weights has been studied under the rubric of the Horvitz-Thompson estimator since the mid 50s. Rao (1965, 1985) summarizes this research in statistics. Important contributions to the choice based sampling literature in economics were made by Manski and Lerman (1977), Manski and McFadden (1981) and Cosslett (1981). Length biased sampling is analytically equivalent to choice based sampling and has been studied since the late 19th Century by Danish actuaries. See Sheps and Menken (1973) and Trivedi and Baker (1983). Heckman and Singer (1985b) extend the classical analysis of length biased sampling in duration analysis to consider models with unobservables dependent across spells and time varying variables. In their more general case, simple weighting methods with weights determined independently from the model are not available.

[22] See equation system (A-1) in Appendix A. See Telser (1964) and the discussions in Heckman (1976b, 1978a, 2000).

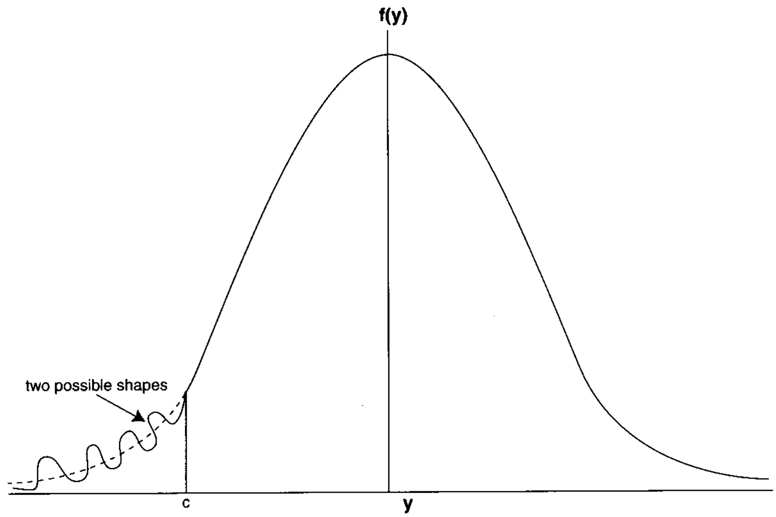

Figure 2a. Data for $y < c$ missing. Two possible slopes for density below c.

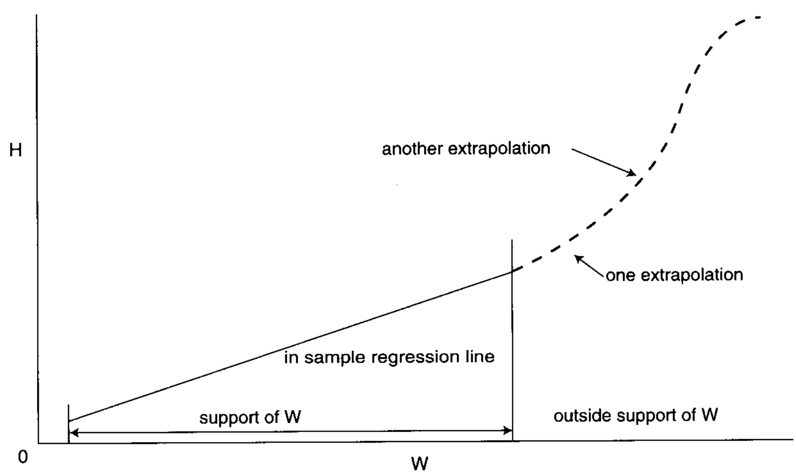

Figure 2b.

while the X affect outcomes. There may be variables in common in X and Z. We observe Y (see equation (12)). Then

15(a) $E(Y|X, Z, D = 1) = E(Y_1|X, Z, D = 1) = \mu_1(X) + E(U_1|X, Z, D = 1)$

and

15(b) $E(Y|X, Z, D = 0) = E(Y_0|X, Z, D = 0) = \mu_0(X) + E(U_0|X, Z, D = 1).$

The conditional means of U_0 and U_1 are the "control functions" or bias functions as introduced and defined in Heckman (1980a) and Heckman and Robb (1985, 1986). The mean observed outcomes (the left hand side variables) are generated by the mean of the potential outcomes plus a bias term.

Define $P(z) = \Pr(D = 1|Z = z)$. As a consequence of decision rule (11), in Heckman (1980a) I demonstrate that under general conditions we may always write these expressions as

16(a) $$E(Y|X, Z, D = 1) = \mu_1(X) + K_1(P(Z))$$

16(b) $$E(Y|X,Z, D = 0) = \mu_0(X) + K_0(P(Z))$$

where $K_1(P(Z))$ and $K_0(P(Z))$ are control functions and depend on Z only through P. The functional forms of the K depend on specific distributional assumptions. See Heckman and MaCurdy (1985) for a catalogue of examples.

The value of P is related to the magnitude of the selection bias. As samples become more representative, $P \to 1$, $K_1(P) \to 0$. See Figure 3 which plots control function $K_1(P)$ versus P. As $P \to 1$, the sample becomes increasingly representative since the probability of any type of person being included in the sample is the same (and $P = 1$). The bias function declines with P. We can compute the population mean of Y_l in samples with little selection. (High P). In general, regressions on selected samples are biased for $\mu_1(X)$. We conflate the selection bias term with the function of interest. If there are variables in Z not in X, regressions on selected samples would indicate that they "belong" in the regression. Representation 16(a) and 16(b) is the basis for an entire econometric literature on selection bias in regression functions.[23] The key idea in all this literature is to control for the effect of P on fitted relationships.[24]

The control functions relate the missing data (the U_0 and U_1) to observables. Under a variety of assumptions, it is possible to form these functions up to unknown parameters and identify the $\mu_0(X)$, $\mu_1(X)$ and the unknown parameters from regression analysis, and control for selection bias. (See Heckman (1976a), Heckman and Robb (1985, 1986) and Heckman and Vytlacil, 2002.)

In the early literature, specific functional forms for (15) and (16) were derived assuming that the U were joint normally distributed:

(A-2) $$(U_0, U_1, U_2) \sim N(0, \Sigma).$$

This assumption, coupled with the assumption

(A-3) $$(U_0, U_1, U_2) \perp\!\!\!\perp (X, Z),$$

produces precise functional forms for K_1 and K_0. For censored samples, a two step estimation procedure was developed. (1) Estimate $P(Z)$ from data on the decision to work and (2) using an estimated $P(Z)$ form $K_1(P(Z))$ and $K_0(P(Z))$ up to unknown parameters. Then 16(a) and 16(b) can be estimat-

[23] Heckman, Ichimura, Smith and Todd (1998) present methods for testing the suitability of this representation in a semiparametric setting.

[24] Heckman (1980a) suggests a series expansion of the K_1 and K_0 functions in terms of polynomials of P and suggests that a test for the absence of selection can be based on a test of whether the joint set of polynomials is statistically significant in an outcome equation. Andrews (1991) and Newey (1994) provide more general analyses.

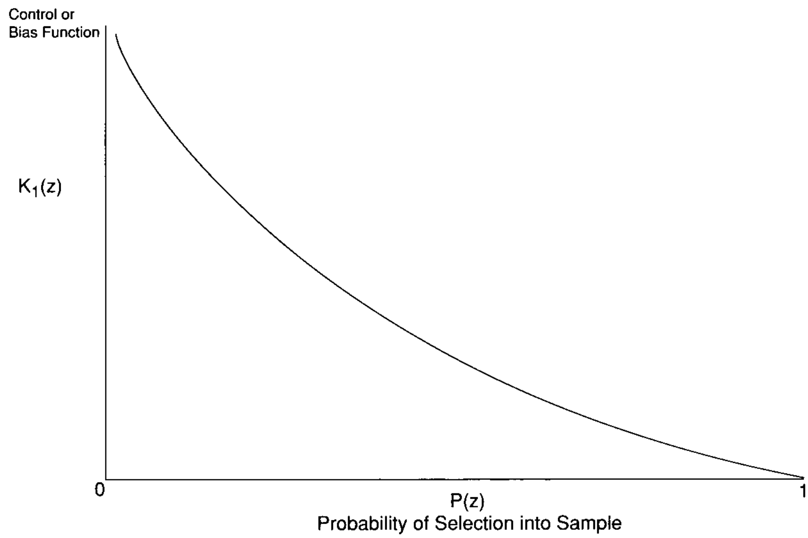

Figure 3. Control Function or Selection Bias as a Function of $P(z)$.

ed using regression. This produces a convenient expression linear in the parameters when $\mu_1(X) = X\beta_1$ and $\mu_0(X) = X\beta_0$.[25] A direct one step regression procedure was developed for truncated samples. (See Heckman and Robb, 1985, 1986.) Equations 16(a) and 16(b) became the basis for an entire literature which generalized and extended the early models, and remains active to this day.

(iii) Empirical Results From These Models and Their Consequences For Economics

The regression framework is useful for investigating microeconomic phenomena from selected samples in the general case of selection covered by the Roy model. In general, no simple weighting with weights that can be estimated separately from the complete model is available to solve the selection problem in the Roy model. Versions of this model have been applied to a variety of problems in economics besides investigations of labor supply and wages.

Recognizing the potential importance of selection shapes the way we interpret economic and social data and gauge the effectiveness of social policy. Consider, for example, the important question of whether there has been improvement in the economic status of African Americans. As depicted in Figure 4a, the median black-white male wage ratio increased in the U.S over the period 1940–1980 and then stabilized. (See the dark curve in Figure 4a.) This statistic is widely cited as justification for a whole set of social policies put

[25] Corrections for using estimated $P(Z)$ in first stage estimation are given in Heckman (1979) and Newey and McFadden (1994). Assumptions (A-2) and (A-3) were also used to estimate the model by maximum likelihood as in the papers of Heckman (1974a,b). The early literature was not clear about the sources of identification, whether exclusion restrictions were needed and the role of normality.

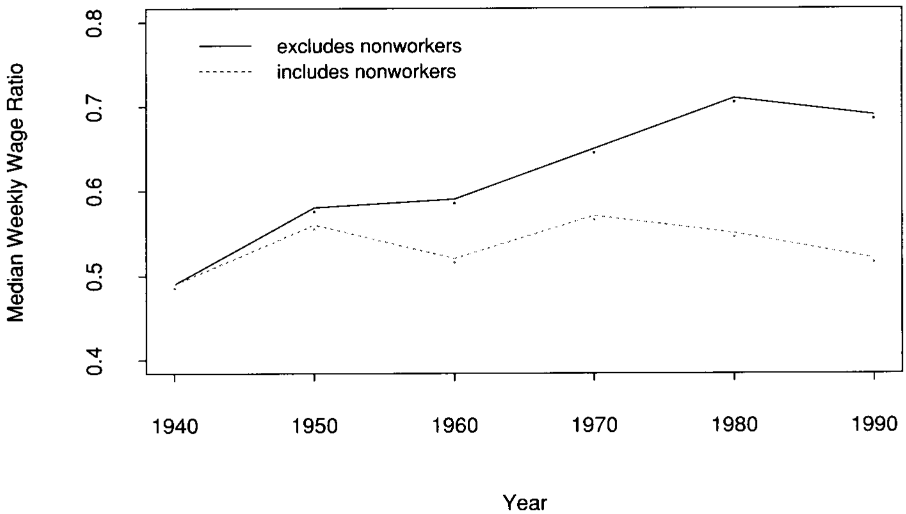

Figure 4a. Median Black-White Male Wage Ratio, 1940–1990. Source: Heckman and Todd (1999).

into place in this period. Over the same period, blacks were withdrawing from the labor force, ($P(Z)$ was going down) and hence from the statistics used to measure wages, at a much greater rate than were whites (see Figure 4b). Correcting for the selective withdrawal of low wage black workers from employment reduces and virtually eliminates black male economic progress compared to that of whites and challenges optimistic assessments of African-American economic progress.[26]

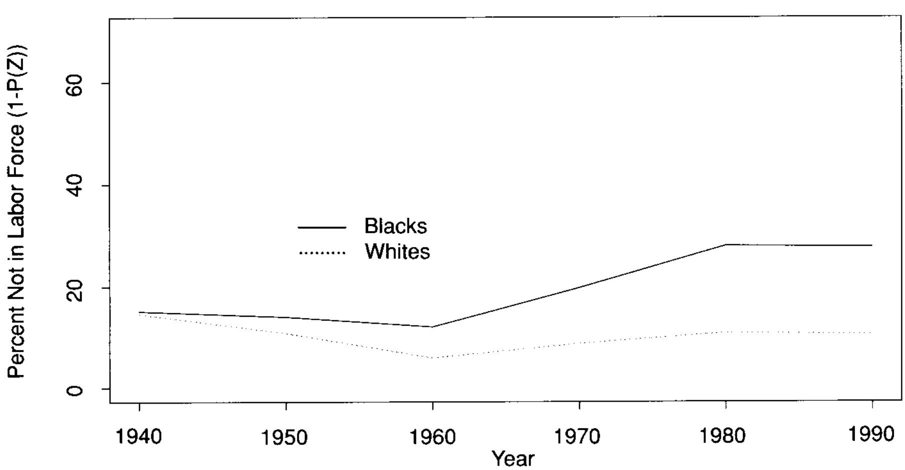

Figure 4b. Percent of Males Not in Labor Force, 1940–90. Source: Heckman and Todd (1999).

[26] The particular selection correction used to produce the numbers used in this figure is to use median wages of workers assuming that low wage workers are the ones who drop out and dropouts are less than 50 % of the entire population. Butler and Heckman (1977) first raised this issue. Subsequent research by Brown (1984), Juhn (1997), Chandra (2000) and Heckman, Lyons and Todd (2000) verify the importance of accounting for dropouts in analyzing black-white wage differentials. Research on this important question is very active.

Thinking about issues in this way has much wider generality. It affects the way we analyze inequality and the effects on employment and welfare of alternative ways of organizing the labor market. In European discussions, the low wage, high inequality U.S. labor market is often compared unfavorably to high wage, low inequality, European labor markets.

These comparisons founder on the same issues that arise in discussing black-white wage gaps. In Europe, the unemployed and the nonemployed are not counted in computing the wage measures used to gauge the performance of the labor market. This practice understates wage inequality and overstates wage levels for the entire population by counting only the workers. A recent paper by Blundell, Reed and Stoker (1999) indicates the importance of the selection problem in the British context. The British data reveal a growth in the real wages of workers over the period 1978–1994. (See the top curve in Figure 5(a)) At the same time, the proportion of persons working has declined (see Figure 5(b)) and accounting for dropouts reduces the level and rate of growth of real wages. The observed growth in real wages may be a consequence of improvements in skill endowments and skill prices (*e.g.* $\mu_1(X)$) or improvements in the nonmarket sector that change the conditional mean of the unobservables in the wage equation by eliminating workers with low potential wages from the labor market. Adjusting for selection (the lower two curves in Figure 5b) greatly reduces the wage growth.

Accounting for selection also affects measures of wage variability over the cycle (Bils, 1985). Low wage persons drop out of the work force (and hence the statistics used to measure worker wages in recessions) and they return to it in booms. Changing composition partly offsets measured wage variability.

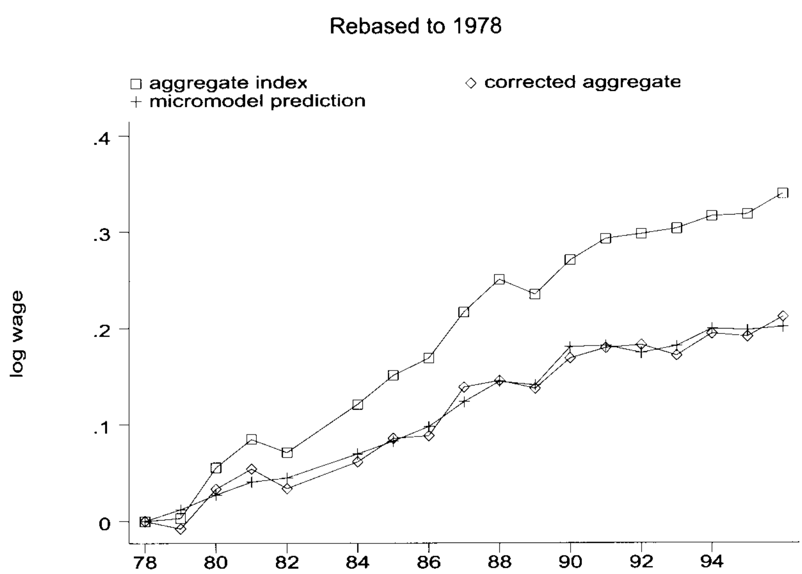

Figure 5a. Wage Predictions from Micromodel, Aggregate wage and Corrections. Source: Blundell, Reed and Stoker (1999).

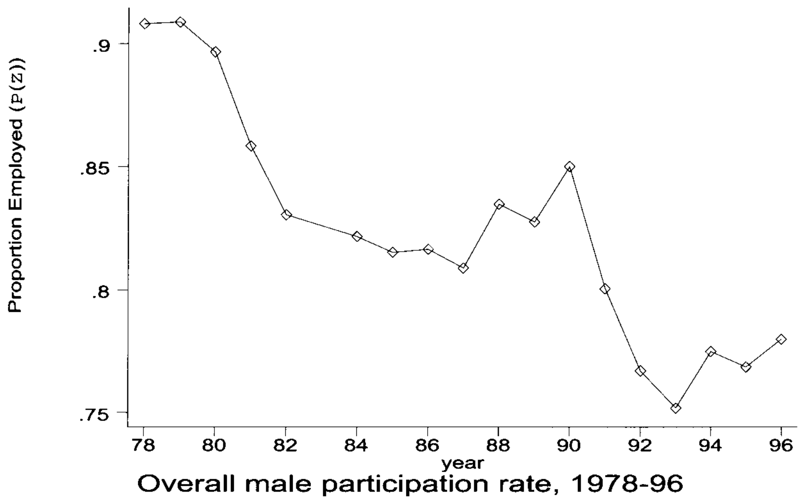

Figure 5b. British males – wages and labour market participation. Source: Blundell, Reed and Stoker (1999).

Thus measured wages appear to exhibit "too little variability" over the business cycle. When a Roy model of self selection is estimated with multiple market sectors, the argument becomes more subtle. Over the cycle, not only is there entry and exit from the workforce but there are movements of workers across sectors within the workforce. Measured wages are thus not simply the price of labor services. In addition to the standard selection effect, measured wages include the effect of weighting placed on different mixes of skills used by workers over the cycle (Heckman and Sedlacek, 1985).

Accounting for both the extensive and intensive margin affects our view of the operation of the labor market. Consider equation (2). We only observe hours of work for workers. To focus on the selection problem assume, contrary to the fact, that wages are observed for everyone, and ignore any endogeneity in wages (so $W \perp\!\!\!\perp U$). Then letting $D = 1$ denote work, the observed labor supply conditional on W and X is

$$E(H \mid W, X, D = 1) = H(W, X) + E(U \mid W, X, D = 1) \tag{17}$$

where the Marshallian labor supply parameter or causal parameter for wages is $\frac{\partial H}{\partial W}$, the *ceteris paribus* change in labor supply due to a change in wages. Compensating for income effects, we can construct a utility constant labor supply function from this to use to conduct a welfare analysis *e.g.* to compute measures of consumer surplus. But a labor supply function fit on a selected sample of workers identifies two wage effects: the Marshallian effect and a compositional or selection effect due to entry and exit from the work force (Heckman, 1978c). Thus

$$\frac{\partial E(H \mid W, X, D = 1)}{\partial W} = \frac{\partial H(W, X)}{\partial W} + \frac{\partial E(U \mid W, X, D = 1)}{\partial W}. \tag{18}$$

The second term is a selection effect, or compositional effect, arising from the change in the composition of the unobservables due to the entry or exit of people into the workforce induced by the wage change This is not a *ceteris paribus* change corresponding to the parameters of classical consumer theory. Equation (18) does not tell us how much a given worker would change her labor supply when wages change. However, it does inform us of what an in-sample wage change would predict for average labor supply. It answers a Marschak Question One-type evaluation question. Under proper conditions on the support of W, it can be used to estimate the within-sample effects of taxes on labor supply[27]

Aggregate labour supply elasticities are inclusive of the effects of entry and exit into the workforce as well the effects of movement along a Marshallian labour supply curve. This simple observation has had substantial effects on the specification, estimation and interpretation of labor supply in macroeconomic models.

In the early 1980s, a literature in macroeconomics arose claiming that aggregate labour supply elasticities were too small and wage movements were too large for a neoclassical model of the labour market to explain the U.S. time series data. I have already discussed why measured wages variation understates the variation in the price of labour. In Heckman (1984), I go on to note that the macro literature focused exclusively on the interior solution labour supply component – the first term on the right of (18) – and ignored the selection effect arising from entry and exit of workers from the labour force. Since half of the aggregate labour supply movements are at the extensive margin (Coleman, 1984), where the labour supply elasticity is higher, the standard 1980's calculations understated the true aggregate labour supply elasticity and hence understated the ability of a neoclassical labour supply model to account for fluctuations in aggregate labour supply. Accounting for choices at the extensive margin changed the way macroeconomists perceived and modeled the labour market. (See, *e.g.* Hansen, 1985, and Rogerson, 1988).

Empirical developments in the labour supply literature reinforced this conclusion. Early on, the evidence called into question the empirical validity of model (10). Fixed costs of work make it unlikely that the index for hours of work is as tightly linked to the participation decision as that model suggests. When workers jump into the labour market, they tend to work a large numbers of hours, not a small number of hours as (10) suggest if the U are normally distributed. Heckman (1976a, 1980b) and Cogan (1981) proposed a more general model with fixed costs in which participation and hours of work equations are less tightly linked. This produces an even greater elasticity for the second term in equation (18). The evidence also called into question the validity of the normality assumption, especially for hours of work data. Hours of work distributions from many countries reveal spiking at standard hours of work.[28] This led to developments to relax the normality assumption used in the early models.

[27] The required condition is (a) in footnote[10] of Section 2.
[28] See the articles in the *Journal of Human Resources*, 1990, special issue of taxes and labor supply.

(iv) Identification

Much of the econometric literature on the selection problem combines discussions of identification (going from populations generated by selection rules back to the source population) with discussions of estimation in solving the inferential problem of going from observed samples to hypothetical populations.[29] It is analytically useful to distinguish the conditions required to identify the selection model from ideal data from the numerous practical and important problems of estimating the model. Understanding the sources of identification of a model are essential to understanding how much of what we are getting out of an empirical model is a consequence of what we put into it.

A conference at the Educational Testing Service (ETS) in 1985 brought together economists and statisticians, and provided some useful contrasts in points of view on causal modelling and selection models (see Wainer, 1986, reissued, 2000).[30] At that conference Holland (1986) used the law of iterated expectations to write the conditional distribution of an outcome, say Y_1 on X in the following form:

$$(19) \qquad F(Y_1|X) = F(Y_1|X, D=1)\Pr(D=1|X) + F(Y_1|X, D=0)\Pr(D=0|X).$$

From the analysis of (11) and (12), we observe Y_1 only if $D = 1$. In a censored sample, we can identify $F(Y_1|X, D=1)$, $\Pr(D=1|X)$ and hence $\Pr(D=0|X)$. We do not observe Y_1 when $D = 0$. Hence, we do not identify $F(Y_1|X)$. In independent work, James Smith and Finis Welch (1986) made a similar decomposition of conditional means (replacing F with E).

Holland questioned how one could identify $F(Y_1|X)$ and a briefly compared selection models with other approaches. Smith and Welch (1986) and some of the authors at the ETS conference discussed how to bound $F(Y_1|X)$ (or $E(Y_1|X)$) by placing bounds on the missing components $F(Y_1|X, D = 0)$ $E(Y_1|X, D=0)$ respectively).[31] A clear precedent for this idea was the work of Peterson (1976) who developed nonparametric bounds for the competing risk model of duration analysis which is mathematically identical to the Roy model of equations (11) and (12).[32] I discuss some recent developments in this literature in Appendix A-2.

The normality assumption that was widely used in the early literature was called into question. Goldberger (1983) and Arabmazar and Schmidt (1981) presented Monte Carlo analysis of models showing substantial bias for models with continuous outcomes when normality was assumed but the true model was non-normal. The empirical evidence is more mixed. Normality is not a bad assumption for analyzing models of self selection for log wage outcomes

[29] See Heckman (2000) for one precise definition of identification.

[30] The exchange between Tukey and myself recorded in that volume highlights the contrast between statisticians and econometricians in the value placed on making identifying discussions explicit and in making causal distinctions.

[31] Smith and Welch use their analysis to bound the effects of dropping on the black-white wage gap discussed in subsection (iii).

[32] The competing risks model replaces $Max(Y_0,Y_1)$ with $Min(Y_0,Y_1)$.

once allowance is made for truncation and self selection.[33] See Figures 6(a) and 6(b) from Heckman and Sedlacek (1985), and the related analysis of Blundell, Reed and Stoker (1999). Olson (1980) and Lee (1982) present early non-normal but parameteric extensions of the early normal Roy framework. Heckman and MaCurdy (1985) present a synthesis of this early literature. Heckman (1980a) presents an early nonparametric estimator of the control function using a series expansion in P.

Heckman and Honoré (1990) consider identification of the Roy model under a variety of conditions. They establish that under normality, the model is identified even if there are no regressors so there are no exclusion restrictions. They further establish that the model is identified (up to subscripts) even if one observes only Y, but does not know if it is Y_1 or Y_0. The original normality assumption used in selection models was based on powerful functional form assumptions.[34]

They develop a nonparametric Roy model and establish conditions under which variation in regressors over time or across people can identify the model nonparametrically. One can replace distributional assumptions with different types of variation in the data to identify the Roy version of the selection model. Heckman and Smith (1998) extend this line of analysis to the Generalized Roy model. It turns out that decision rule (11) plays a crucial role in securing identification of the selection model. In a more general case,

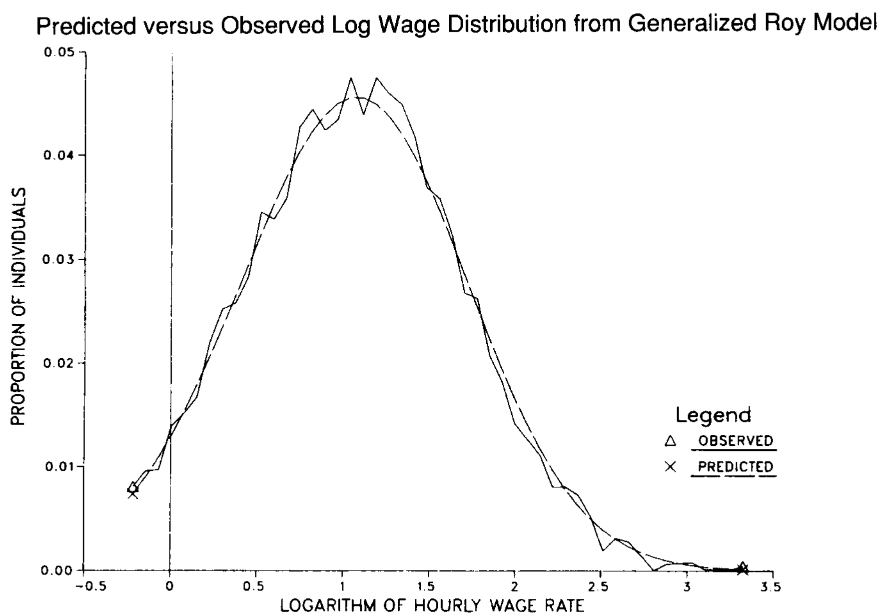

Figure 6a. Nonmanufacturing Sector. Source: Heckman and Sedlacek (1985).

[33] Normality of latent variables turns out to be an acceptable assumption for discrete choice models except under extreme conditions (Todd, 1996).
[34] Powerful, but testable. The model is overidentified. See for example the tests by Bera, Jarque and Lee (1984) for the tests of distributional assumptions within a class of limited dependent variable models.

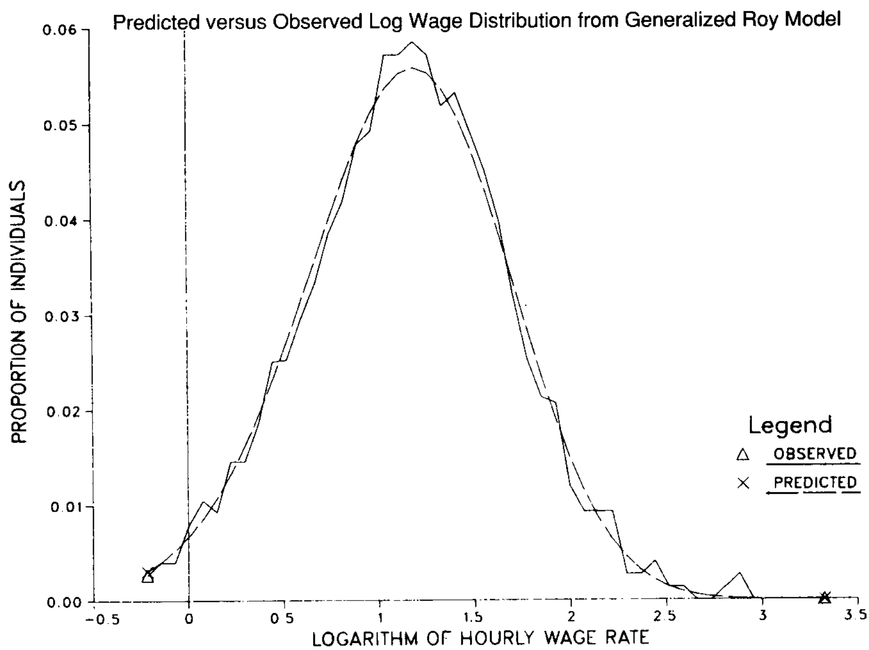

Figure 6b. Manufacturing Sector. Source: Heckman and Sedlacek (1985).

where Y_2 may depend on $Y_1 - Y_0$ but on other unobservables as well, even with substantial variation in regressors across persons or over time, only partial identification of the full selection model is possible. When the models are not identified, it is still possible to bound crucial parameters and an entire literature has grown up elaborating this idea. See Appendix (A-2) for a discussion of this literature. Heckman, Ichimura, Smith and Todd (1998), among others, discuss semiparametric estimation of selection models. See also Robinson (1988) and Ahn and Powell (1993).

6. MICRODYNAMICS AND PANEL DATA: HETEROGENEITY VS. STATE DEPENDENCE AND LIFE CYCLE LABOR SUPPLY

The initial microdata were cross sections. Thus early work on discrete choice, limited dependent variables, and models with mixed continuous-discrete endogenous variables was cross sectional in nature and focused exclusively on explaining variation over people at a point in time. This gave rise to multiple interpretations of the sources of the unobservables in (7) and (8). The random utility models introduced in the literature in discrete choice interpreted these as temporally independent preference shocks (McFadden, 1974) especially when discrete choice was considered. Other interpretations were (a) systematic variations in unobserved preferences that were stable over time and (b) omitted characteristics of choices and agents which may or may not be stable over time.[35]

[35] Heckman and Snyder (1997) consider the history of these ideas.

With the advent of panel data in labor economics, an accomplishment due in large part to James Morgan and his group at the Institute For Survey Research at the University of Michigan, it was possible to explore these sources of variation more systematically.[36] The issue was especially important in the study of the female labour supply.

Mincer (1962) used an implicit version of the random utility model to argue that cross section labor force participation data could be used to estimate Hicks-Slutsky income and substitution effects. His idea was that H in equation (1) measured the fraction of lifetime that people worked and that if leisure time is perfectly substitutable over time the timing of labor supply is irrelevant and could be determined by the flip of a coin. Then a regression of labour force participation rates, on W would identify a Hicks-Slutsky wage effect.[37] Ben-Porath (1973) assumed instead that shocks were permanent stable traits of individuals, interpreted labour force participation as a corner solution, and showed that a regression of labour force participation rates on wages would identify parameters from a distribution of tastes for work, rather than the Hicks-Slutsky substitution effect (*i.e.* it would define the parameters of $\Pr(D = 1 | X)$ from equation (9)).

This issue is also important in understanding employment and unemployment data. A frequently noted empirical regularity in the analysis of unemployment data is that those who are unemployed in the past or have worked in the past are more likely to be unemployed (or work) in the future. Is this due to a causal effect of being unemployed or (working) or is it a manifestation of a stable trait *e.g.* some people are lazier than others and observables are persistent? One theory of macroeconomics was built around the premise that promoting work through macro policies would foster higher levels of employment (Phelps, 1972). The distinction between true and spurious effects is the distinction between true and spurious state dependence.

In a series of papers, I developed econometric models to use panel data to investigate these issues. One set of studies builds on the model of equations 10(a)–10(c) but places them in a life cycle setting. My work on life cycle labor supply (Heckman, 1974c, 1976c), demonstrated that the marginal utility of wealth constant ("Frisch") demand functions were the relevant concept for analyzing the evolution of labor supply over the life cycle in environments of perfect certainty or with complete contingent claims markets. Building on this work, MaCurdy and I (1980), drawing on Heckman (1974c) and thesis research by MaCurdy (1978, 1981), formulated and estimated a life cycle version of the model of equations (10) that interpreted one of the key unobservables in the model as the marginal utility of wealth, λ. In the economic settings we assumed, λ is a stable unobservable or fixed effect derived from economic theory. The models we developed extended, for the first time, models for limited dependent variables, systematically missing data and joint

[36] See Stafford (2001).

[37] Heckman (1978c) provides a formal analysis and the relationship to the random utility model studied by McFadden.

continuous-discrete endogeneous variables to a panel setting.[38] Our evidence and my related joint work with Willis (Heckman and Willis, 1977) suggests that a synthesis of the views of Ben-Porath and Mincer was appropriate, and a pure random utility specification was inappropriate. This framework has been extended to account for human capital and uncertainty in important papers by Altug and Miller (1990, 1998).

In related work, I generalized the static cross sectional models of discrete choice to a dynamic setting, and used this generalization to address the problem of heterogeneity vs. state dependence. This fundamental problem can be understood most simply by considering the following urn schemes (Heckman, 1981c).

In the first scheme there are I individuals who possess urns with the same content of red and black balls. On T independent trials individual i draws a ball and then puts it back in his or her urn. If a red ball is drawn at trial t, person i experiences the event (*e.g.* is employed, is unemployed, *etc.*). If a black ball is drawn, person i does not experience the event. This model corresponds to a simple Bernoulli model and captures the essential idea underlying the choice process in McFadden's (1974) work on discrete choice. From data generated by this urn scheme, one would not observe the empirical regularity that a person who experiences the event in the past is more likely to experience the event in the future. Irrespective of their event histories, all people have the same probability of experiencing the event.

A second urn scheme generates data that would give rise to a measured effect of past events on current events solely due to heterogeneity. In this model, individuals possess distinct urns which differ in their composition of red and black balls. As in the first model, sampling is done with replacement. However, unlike the first model, information concerning an individual's past experience of the event provides information useful in locating the position of the individual in the population distribution of urn compositions.

The person's past record can be used to estimate the person-specific urn composition. The conditional probability that individual i experiences the event at time t is a function of his past experience of the event. The contents of each urn are unaffected by actual outcomes and in fact are constant. There is no true state dependence.

The third urn scheme generates data characterized by true state dependence. In this model individuals start out with identical urns. On each trial, the contents of the urn change *as a consequence of the outcome of the trial.* For example, if a person draws a red ball, and experiences the event, additional new red balls are added to his urn. Subsequent outcomes are affected by previous outcomes because the choice set for subsequent trials is altered as a consequence of experiencing the event.

A variant of the third urn scheme can be constructed that corresponds to a renewal model. In this scheme, new red balls are added to an individual's urn

[38] Browning, Deaton and Irish (1985) adapt this idea to repeated cross section data using standard methods for analyzing synthetic cohorts.

on successive drawings of red balls until a black ball is drawn, and then all of the red balls added as a result of the most recent continuous run of drawings of red balls are removed from the urn. The composition of the urn is then the same as it was before the first red ball in the run was drawn. A model corresponding to fixed costs of labor force entry is a variant of the renewal scheme in which new red balls are added to an individual's urn only on the first draw of the red ball in any run of red draws.

The crucial feature that distinguishes the third scheme from the second is that the contents of the urn (the choice set) are altered as a consequence of previous experience. The key point is not that the choice set changes across trials but that it changes in a way that depends on previous outcomes of the choice process. To clarify this point, it is useful to consider a fourth urn scheme that corresponds to models with more general types of heterogeneity to be introduced more formally below.

In this model, individuals start out with identical urns, exactly as in the first urn scheme. After each trial, but independent of the outcome of the trial, the contents of each person's urn are changed by discarding a randomly selected portion of balls and replacing the discarded balls with a randomly selected group of balls from a larger urn (say, with a very large number of balls of both colors). Assuming that the individual urns are not completely replenished on each trial, information about the outcomes of previous trials is useful in forecasting the outcomes of future trials, although the information from a previous trial declines with its remoteness in time. As in the second and third urn models, previous outcomes give information about the contents of each urn. Unlike the second model, the fourth model is a scheme in which the information depreciates since the contents of the urn are changed in a random fashion. Unlike in the third model, the contents of the urn do not change as a consequence of any outcome of the choice process.

In the literature on female labor force participation, models of extreme homogeneity (corresponding to urn model one) and extreme heterogeneity (corresponding to urn model two with urns either all red or all black) are both consistent with the cross sectional evidence. This is the contrast between Mincer and Ben-Porath. Heckman and Willis (1977) estimate a model of heterogeneity in female labor force participation probabilities that is a probit analogue of urn model two.

Urn model three is of special interest. It is consistent with human capital theory, and other models that stress the impact of prior work experience on current work choices. Human capital investment acquired through on the job training may generate structural state dependence. Fixed costs incurred by labor force entrants may also generate structural state dependence as a renewal process. So may spell-specific human capital. This urn model is also consistent with psychological choice models in which, as a consequence of receiving a stimulus of work, women's preferences are altered so that labor force activity is reinforced. (Atkinson, Bower and Crothers 1965) or economic models of habit formation.

Panel data can be used to discriminate among these models. For example, an implication of the second urn model is that the probability that a woman participates does not change with her labor force experience. An implication of the third model in the general case is that participation probabilities change with work experience. One method for discriminating between these two models utilizes individual labor force histories of sucient length to estimate the probability of participation in different subintervals of the life cycle. If the estimated probabilities for a given woman do not differ at different stages of the life cycle, there is no evidence of structural state dependence.

Heckman (1981a,b) develops a class of discrete data stochastic processes that generalize the discrete choice model of McFadden (1974) to a dynamic setting. That set up is sufficiently general to test among all four urn schemes and present a framework for dynamic discrete choice.[39] Heckman and Singer (1985a) present an explicit generalization of the McFadden model in which Weibull shocks arrive at Poisson arrival times. Dagsvik (1994) presents a generalization to a continuous time setting.

Heckman (1978d) shows that it is possible to use runs tests to distinguish between heterogeneity and state dependence. The intuition behind the test is that just the total number of past occurrences of an event, and not the order of occurrence of past events, is relevant to predicting the current probability of an outcome if a pure heterogeneity explanation is appropriate. If state dependence matters, then the timing of past events matters. This insight is the basis for recent work by Honoré and Kyriazidou (2000) in analyzing discrete data models with lagged dependent variables. It is also used by Chiappori and Heckman (2000) in distinguishing adverse selection from moral hazard models. Heckman and Borjas (1980) present a related analysis for continuous time duration analysis and apply this to analyzing the important policy question of whether or not past unemployment causes future unemployment or whether past unemployment is just a signal of a propensity to be unemployed. For young men in the U.S., they find that the latter story is more appropriate. The experience of unemployment has no lasting effects on future employment. Heckman (1981c) analyzes female employment, and finds that for older married women (age 45–59), past employment raises future employment even controlling for unobservables. Layard, Nickell and Jackman (1991) summarize the recent evidence on true state dependence in unemployment histories.

Table 2 from Heckman (1981c) reveals that heterogeneity and temporal persistence are important features of the data on the labor force participation of women in the data. The table records the work history of women over three year periods with "1" denoting working in the year and "0" not working. If the classical random utility model characterized the data, all the rows with the same number of "1*s*" and "0*s*", irrespective of their order, should be roughly equal, which they clearly are not, especially for the younger women. Those who work tend to persist in work while those who do not work persist

[39] Heckman and Willis (1974, 1975) present a protype for this class of models in their analysis of fertility dynamics. Lillard and Willis (1978) apply this model to the analysis of earnings dynamics.

Table 2
Runs Patterns in the Data
(1 corresponds to work in the year, 0 corresponds to no work)

Runs Pattern 1968	1969	1970	No. of Observations	Runs Pattern 1971	1972	1973	No. of Observations
A. Women Aged 45-59 in 1968							
0	0	0	87	0	0	0	96
0	0	1	5	0	0	1	5
0	1	0	5	0	1	0	4
1	0	0	4	1	0	0	8
1	1	0	8	1	1	0	5
0	1	1	10	0	1	1	2
1	0	1	1	1	0	1	2
1	1	1	78	1	1	1	76
B. Women Aged 30-44 in 1968							
0	0	0	126	0	0	0	133
0	0	1	16	0	0	1	13
0	1	0	4	0	1	0	5
1	0	0	12	1	0	0	16
1	1	0	24	1	1	0	8
0	1	1	20	0	1	1	19
1	0	1	5	1	0	1	8
1	1	1	125	1	1	1	130

Source: Heckman, James J., "Heterogeneity and State Dependence," in *Studies in Labor Markets*, S. Rosen, ed. Chicago: University of Chicago Press. 1981a.

in not working. This is true even after conditioning on observables. Temporally persistent unobservables are a central feature of microeconomic data which affect how we estimate models and interpret data.

Early work on this topic accounted for persistent heterogeneity using parametric distributions for the unobservables.[40] An important issue was whether it was possible to distinguish heterogeneity from state dependence without using parametric structure. Using analyses of runs tests, Heckman (1978c) showed it was possible to test between the two sources of dependence. The open question was whether it was possible to measure the relative contributions the way Heckman (1981a,b) had done using parametric structures.

Elbers and Ridder (1982) showed that this was theoretically possible in the context of a proportional hazard model in duration analysis but did not provide an estimation algorithm. Heckman and Singer (1984) extend their analysis to develop a consistent estimation procedure. Cameron and Heckman (1998, 2001) and Hansen, Heckman and Vytlacil (2000) extend and generalize this analysis models to discrete time outcomes with jointly determined discrete and continuous outcome variables. A major empirical finding

[40] For an overview of mixing models in demography, see Sheps and Menken (1973). Heckman and Willis (1974, 1975, 1977) estimate models of fertility and labour supply with parametric mixing models. Lancaster (1979, 1990) and Lancaster and Nickell (1980) apply these models to analyze unemployment spells.

in the work of Heckman and Singer that has been replicated in numerous subsequent studies is that distributions of unobservables can be approximated by low dimensional finite mixtures or "types." This has proved fruitful in the analysis of discrete dynamic choice. (See, *e.g.* Eckstein and Wolpin, 1999.)

The field of discrete dynamic choice has progressed enormously and is an active area of research. Flinn and Heckman (1982) present the first rigorous structural model of discrete dynamic choice in the context of a search model of unemployment. They investigate the nonparametric identifiability of the search model and present a class of identification theorems that have also proved useful in the context of identifying auction models. (See, *e.g.* Laffont, Ossard and Vuong, 1995; Donald and Paarsch, 1996; Hong, 1998). Later work in dynamic discrete choice by Pakes (1986), Rust (1987), Eckstein and Wolpin (1989, 1999) and Keane and Wolpin (1997, 1999) and others has focused more exclusively on computational methods for parametric models.[41] See Rust (1996) for a recent overview of this work.

7. TREATMENT EFFECTS

The identification and estimation of structural econometric models are challenging tasks, especially for models with dynamics and uncertainty. The value of this research is beyond question. But the difficulty of this task is also clear. The task has been made harder by the empirical findings showing the importance of heterogeneity in economic data.

Cowles econometrics focused on a linear-in-parameters model for person i:

$$Y_i = X_i\beta + U_i \tag{20a}$$

where $E(U_i) = 0$ and analyzed the problems that arose when $E(U_i|X_i) \neq 0$. Heterogeneity among individuals was modeled as heterogeneity in intercepts. The empirical evidence from the entire body of research in microeconometrics in the past 30 years supports a more encompassing view of heterogeneity in both slopes and intercepts:

$$Y_i = X_i\beta_i + U_i,\ E(U_i) = 0, \tag{20b}$$

where the slopes vary among individuals. Letting $\bar{\beta} = E(\beta_i)$ and $v_i = \beta_i - \bar{\beta}$, a lot of recent evidence in a number of areas of microeconometrics points to correlation between X_i and β_i so in addition to $E(U_i|X_i) \neq 0$, $E(v_\iota|X_i) \neq 0$. (Carneiro, Heckman and Vytlacil, 2001c,d). This is the "correlated random coefficient model". It arises naturally in the analysis of the economic returns schooling. If schooling is an X and Y is log earnings, the component of β associated with schooling is a rate of return which may vary among individuals and which is plausibly correlated with the level of schooling. Standard instrumental variables methods for $\bar{\beta}$ break down when agents act on v_i in selecting X_i (Heckman, 1997 and Heckman and Vytlacil, 1998). Models like (20b) are

[41] Related research on auctions by Athey and Haile (2000) and others considers nonparametric identification more seriously.

only the tip of the empirical iceberg.[42] Most structural models have a general nonlinear form for the estimating equations which can often only be defined recursively (see, *e.g.* Rust, 1996).[43] This adds further difficulty to the estimation of structural equations.

Given the complexity of estimating structural models with heterogeneity in slopes and intercepts, it is not surprising that econometricians and empirical economists have sought simpler methods for answering certain narrowly focused questions, rather than the full array of questions that can be addressed by structural equations methods. The literature on treatment effects investigates a class of interventions with partial coverage so there is a "treatment" and "control" group. It is not helpful in evaluating interventions that apply universally within an economy unless there are data on separate economies experiencing different interventions and the economies are segregated from each other. It finesses general equilibrium problems by assuming that the outcomes of nonparticipants (control group members) are the same as what they would experience in the absence of the intervention. (Heckman and Smith, 1998).[44]

The treatment effect literature approaches the problem of policy evaluation in the same way that biostatisticians approach the problem of evaluating a drug. Outcomes of persons exposed to a policy are compared to outcomes of those that are not. The analogy is more than a little strained in the context of evaluating many social policies because in a modern economy outcomes of persons are linked through markets and other forms of social interaction. This gives rise to a distinction between those "directly" affected and those only "indirectly" affected. Thus those who attend college because of a tuition subsidy program are directly affected. The rest of society is indirectly affected by the cost of taxation to finance the subsidy and by the effects of an expansion of the stock of education on the prices of educated and uneducated labor.[45]

The essential differences between the treatment effect literature and the structural equations literature are conveyed in the following simple analysis. Consider the models for potential outcomes given in (7). Those functions are nonlinear and the U_i may be vectors of outcome-specific unobservables. The U_i may be stochastically dependent on the X_i. The random coefficient model of (20a) is a special case of (7) interpreting the U_i as vectors. Consider a mo-

[42] Carneiro, Heckman and Vytlacil (2001c,d) and Dustmann and Meghir (2001) present estimates based on a correlated random coefficient models.

[43] Random coefficient models with X_i independent of β_i were introduced into econometrics in Rubin (1950). Random coefficient probit models with X_i independent of β_i in linear indices are implicit in Thurstone (1927) and were introduced into econometrics by Domencich and McFadden (1975). Ichimura and Thompson (1998) provide a recent insightful contribution to this literature.

[44] The book by Campbell and Stanley (1966) is a classic reference in educational statistics. Early contributions by economists include papers by Goldberger (1972), Barnow, Cain and Goldberger (1980), and Cain (1975).

[45] See Heckman and Smith (1998). Heckman, Lochner and Taber (1998a,b,c, 1999) demonstrate how substantial these "indirect" effects can be. Lewis (1963) addresses these general equilibrium effects in the context of evaluating the impact of unionism on the economy, and introduces the notion of direct and indirect effects.

del with two potential outcomes (Y_0, Y_1). Write the outcomes as Marshallian causal functions

(21a) $$Y_0 = g_0(X, U_0)$$

(21b) $$Y_1 = g_1(X, U_1).$$

For specificity think of (Y_0, Y_1) as potential earnings of a person as a high school or as a college graduate.

The structural approach seeks to determine g_0 and g_1, usually by invoking additive separability, *e.g.* in the case of scalar (U_0, U_1)

(22a) $$Y_0 = g_0(X) + U_0$$

(22b) $$Y_1 = g_1(X) + U_1.$$

Determination of these functions and the economic mechanism selecting which component of (Y_0, Y_1) is observed enables the analyst to answer the full array of policy counterfactuals considered in Section 2, subject to the conditions on support discussed there.

The treatment effect literature focuses on a narrower range of questions and answers them under weaker conditions than are required to identify and estimate structural equations. In the context of equations (21a)–(21b), it is natural to think of the intervention as movement of a person from the "0" state to the "1" state. $D = 1$ if a person is in 1 (*e.g.* college educated) and $D = 0$ otherwise (*e.g.* high school).

Two treatment effects receive the most attention in the current literature: the average treatment effect for $\Delta = Y_1 - Y_0$, *ATE*, which is the effect of picking someone at random to get treatment:

$$E(\Delta | X = x) = E(Y_1 - Y_0 | X = x) = ATE(x)$$

or treatment on the treated, TT, the effect of treatment on those who actually are treated:

$$E(\Delta | X = x, D = 1) = E(Y_1 - Y_{0z} | X = x, D = 1) = TT(x).$$

In terms of the structural model of equations (21a)–(21b),

$$ATE(x) = E(Y_1 - Y_0 | X = x) = E[g_1(X, U_1) - g_0(X, U_0) | X = x]$$

$$TT(x) = E(Y_1 - Y_0 | X = x, D = 1) = E[g_1(X, U_1) - g_0(X, U_0) | X = x, D = 1].$$

In terms of a more familiar additively separable representation

$$ATE(x) = g_1(x) - g_0(x) + E[U_1 - U_0 | X = x]$$

$$TT(\mathrm{x}) = g_1(x) - g_0(x) + E[U_1 - U_0 | X = x, D = 1].$$

The definition of these parameters does not require that the X are exogenous $(E(U_1 - U_0 | X = x) \neq 0$ or $E(U_1 - U_0 | X = x, D = 1) \neq 0)$ or that the structural functions g_1 and g_0 are identified. Of course, if one knows the structural functions and the dependence between (U_0, U_1) and X, one can identify the treatment parameters, so the structural approach is more general.

It is instructive to write the treatment parameters within a correlated random coefficient framework 20(b). Define observed Y as

$$Y=(1-D)Y_0+DY_1. \tag{23}$$

Y may be discrete, continuous or mixed discrete-continuous. Define the residuals from the conditional expectations of (Y_0, Y_1) as

$$\varepsilon_0 = Y_0 - E(Y_0 \mid X = x) = Y_0 - E(g_0(X, U_0) \mid X = x)$$

$$\varepsilon_1 = Y_1 - E(Y_1 \mid X = x) = Y_1 - E(g_1(X, U_0) \mid X = x).$$

In the additively separable case

$$\varepsilon_0 = Y_0 - g_0(x) - E(U_0 \mid X = x)$$

$$\varepsilon_1 = Y_1 - g_1(x) - E(U_1 \mid X = x).$$

Even in this case, only if the errors are mean independent of X and mean zero will (U_0, U_1) coincide with $(\varepsilon_0, \varepsilon_1)$.In the treatment effect literature there is no mean independence requirement for (U_0, U_1). For notational simplicity, define $E(Y_1|X = x) = \mu_1(x)$ and $E(Y_0|X = x) = \mu_0(x)$. Observe that $\mu_1(x)$ and $\mu_0(x)$ are not Marshallian causal functions except when $E(U_0|X) = 0$ and $E(U_1|X) = 0$.

In this notation

$$ATE(x) = \mu_1(x) - \mu_0(x)$$

$$TT(x) = \mu_1(x) - \mu_0(x) + E[\varepsilon_1 - \varepsilon_0 \mid X = x, D = 1].$$

We may write (23) as

$$Y = \mu_0(X) + D(\mu_1(X) - \mu_0(X) + \varepsilon_1 - \varepsilon_0) + \varepsilon_0. \tag{24}$$

This is a nonparametric random coefficient model of the form 20(b) where the random coefficient is on D and D may be correlated with ε_0 and $\varepsilon_1 - \varepsilon_0$. Observe that the coefficient on D is $\Delta = Y_1 - Y_0 = \mu_1(x) - \mu_0(x) + \varepsilon_1 - \varepsilon_0$, the individual level treatment effect. It has mean $ATE(x) = E(Y_1 - Y_0|X = x) = \bar{\Delta}(x)$. For notational convenience, I keep the dependence on X implicit in the ensuing analysis.

We may write (24) in two different ways. First, in terms of ATE we obtain

$$Y = \mu_0 + (\mu_1 - \mu_0)D + \{\varepsilon_0 + D(\varepsilon_1 - \varepsilon_0)\} = \mu_0 + (ATE)\, D + \{\varepsilon_0 + D(\varepsilon_1 - \varepsilon_0)\}.$$

In terms of TT we obtain

$$\begin{aligned} Y &= \mu_0 + [\bar{\Delta} + E(\varepsilon_1 - \varepsilon_0 \mid D = 1)]D + \{\varepsilon_0 + [(\varepsilon_1 - \varepsilon_0) - E(\varepsilon_1 - \varepsilon_0 \mid D = 1)]D\} \\ &= \mu_0 + (TT)D + \{\varepsilon_0 + [(\varepsilon_1 - \varepsilon_0) - E(\varepsilon_1 - \varepsilon_0 \mid D = 1)]D\}. \end{aligned}$$

Now the econometric problem of identifying these treatment effects is localized to the problem that D may be correlated with the error terms in braces.

Observe that a variety of means can be defined over different conditioning sets. Thus there are different mean responses and different means answer different questions. There is no single average or "representative" agent that is useful for answering all policy evaluation questions except in special cases noted below. In the presence of selection on idiosyncratic unobservables, no single "effect" describes a program or intervention. A variety of treatment effects can be defined that depend on the conditioning sets used to define "the" effect.

Picking persons at random and entering them into a program and comparing their mean outcomes with those of other randomly selected persons denied access produces the Average Treatment Effect (*ATE*). Picking persons at random who go into the program and comparing their average outcomes with those of the same type of people denied access to the programs defines the parameter Treatment on the Treated (*TT*). Assuming full compliance, this is the implicit parameter of interest in recent social experiments that deny access to otherwise acceptable applicants.

It is useful to distinguish three cases of equation (24) and relate them to more conventional econometric models. The first case arises when responses to treatment are the same for everyone (given $X = x$):

(C-1) $\varepsilon_1 = \varepsilon_0$ and thus Δ is a constant.

In this case $E(\Delta|D = 1) = E(\Delta) = \Delta$. One representative agent summarizes the average outcomes of the program. There is a single mean "effect." The problem of selection bias comes down to the Cowles problem that D may be correlated with ε_0.[46]

The second case arises when responses to treatment vary among people, but decisions to take treatment are not based on these variable responses:

(C-2) $\varepsilon_1 \neq \varepsilon_0$, and thus Δ varies among people, but $\varepsilon_1 - \varepsilon_0$ is mean independent of D so that $E(\Delta|D = 1) = E(\Delta)$.

In this case returns to participation in the activity being evaluated vary *ex post* but are the same on average for all persons with the same values of X. Like the case with a common coefficient, this case is favorable to a representative agent description of the intervention being evaluated. Again, the econometric problem comes down to the problem that D may be correlated with ε_0.

The third case arises when responses to treatment vary among people and decisions to take treatment are based on this variation:

(C-3) $\varepsilon_1 \neq \varepsilon_0$ and thus Δ varies among people, and $\varepsilon_1 - \varepsilon_0$ is not mean independent of D and thus $E(\Delta|D = 1) \neq E(\Delta)$.

People sort into treatment status based at least in part on unobserved gains. Now the econometric problem entails accounting not only for the correlation between ε_0 and D but also accounting for the correlation between Δ and D and (Δ, D) and ε_0.

[46] Observe that $U_1 = U_0$ does not imply that $\varepsilon_1 = \varepsilon_0$.

Heterogeneity in response to treatment on which agents act violates the common parameter assumption in Cowles econometrics. There is no single "effect" of treatment but, rather, a variety of effects depending on the conditioning variables.[47] This represents a radical departure from the policy invariant structural parameters based on g_0 and g_1 that are the hallmark of Cowles econometrics. The unity and simplicity of the structural literature in producing parameters that can be transported and compared across economic environments appears to be lost in the literature on treatment effects.

Recovering these evaluation parameters from data requires making fewer assumptions than are required to recover structural parameters. But nontrivial problems still remain that take us back to the selection problem for estimating structural parameters albeit in a different form. Retaining the focus on means, from observational data we can consistently estimate $E(Y|X=x, D=1) = E(Y_1|X=x, D=1)$ (*e.g.* earnings of college graduates) and $E(Y|X=x, D=0) = E(Y_0|X=x, D=0)$ (*e.g.* earnings of high school graduates).

Using the earnings of high school graduates to proxy what college graduates would earn if they were high school graduates is problematic. Comparing the mean earnings of the two groups:

$$E(Y_1 \mid X = x, D = 1) - E(Y_0 \mid X = x, D = 0) =$$

$$\underbrace{E(Y_1 - Y_0 \mid X = x, D = 1)}_{TT(X)} + \underbrace{E(Y_0 \mid X = x, D = 1) - E(Y_0 \mid X = x, D = 0).}_{\textbf{Selection Bias}}$$

The second term is the difference between what college graduates would earn if they were high school graduates and what high school graduates would earn. Another definition of selection bias would arise if we chose *ATE* as our parameter.

If there were no unobservables, or if fortuitously conditioning on X eliminated mean differences in unobservables, as is assumed by statisticians who advocate the method of matching, then the selection bias term vanishes. Yet the poor fit of most microdata equations suggests that the assumption of no observables is unacceptable. Reliance on matching is an act of faith.[48] Many different approaches have been proposed to eliminate these selection biases. Joint work with Robb (1985, 1986) developed a variety of different estimators for different economic models, sampling plans and treatment parameters. We consider a variety of identifying assumptions and use economics and statistics to justify choices of estimators.

If different environments are characterized by different (X, U_0, U_1) dependencies, the treatment effects estimated in one environment do not transport to other environments.[49] Variations in X are not *ceteris paribus* changes and do

[47] This point was recognized early in Heckman and Robb (1985, 1986) and Björklund and Moffitt (1987).

[48] The classical model of ability bias in earnings equations relating schooling to earnings assumes that the selection bias in the return to schooling is positive even conditioning on X so is inconsistent with matching. The Roy model is inconsistent with matching. (See Heckman and Vytlacil, 2002).

[49] This is a treatment effect version of the point originally made by Haavelmo (1943).

not answer structural questions. The treatment effect literature evades these problems by focusing on estimating one effect, or a limited set of effects, that apply to one environment and cannot be applied to other environments.[50,51] This approach avoids most of the problems of structural estimation but at the cost of producing estimates that do not answer structural questions. Not only are estimates of these parameters in general incomparable across studies based on different samples (so it is difficult to cumulate knowledge across studies) but it is diffcult to relate the different treatment effects estimated on the same sample.

(i) Using the Marginal Treatment Effect to Unify the Literature on Treatment Effects
Recent joint work with Vytlacil (1999, 2000a,b, 2001a,b,c,d, 2002) unifies the literature on treatment effects using an economically interpretable treatment parameter, the Marginal Treatment Effect (*MTE*). We link the treatment effect literature to the more conventional structural equations literature to harvest some of the benefits of the structural approach.

The *MTE* is the mean effect of the program for those at the margin of participation in it for given values of observables and conditioning on the unobservables in the program participation equation.[52]

We present conditions under which it is possible to represent all of the conventional treatment parameters as weighted averages of the *MTE*, where different parameters correspond to different weights. See Heckman and Vytlacil (1999, 2000a, 2001c,d, 2002). Under the same conditions, we organize the econometric evaluation literature by classifying estimators on the basis of whether or not they assume that the *MTE* depends on the unobservables in the equation determining participation in the program. Ordinary linear instrumental variables (*IV*) is characterized as a weighted average of *MTE* where the weights in general differ from those used to define the standard treatment parameters. (Heckman and Vytlacil, 2000a,b, 2001c,d, 2002).

More precisely, our work starts with 21(a) and 21(b) and links up to the discrete choice literature by postulating a latent variable $D^* = \mu_D(Z) - U_D$ such that $D = 1$ if $D^* \geq 0$; $D = 0$ otherwise. Thus the choice mechanism in the Roy model and its generalizations is retained but in a semiparametric setting. We assume: (a) $\mu_D(Z)$ is a nondegenerate random variable conditional on X; (b) U_D is absolutely continuous with respect to Lebesgue measure; (c) (U_1, U_D) and (U_0, U_D) are independent of Z conditional on X; (d) Y_1 and Y_0 have finite first moments; and (e) $1 > \Pr(D = 1 | X = x) > 0$ for every $x \in Supp(X)$. Assumptions (a) and (c) are "instrumental variable" assumptions that there is

[50] Heckman (1992) makes this point in the context of estimating parameters determined from random assignment to treatment. See Heckman and Vytlacil (2001d, 2002) for a more extensive discussion.
[51] Marschak (1953) and Hurwicz (1962) consider economic decision problems that do not consider full knowledge of structural parameters.
[52] This parameter was introduced into the evaluation literature by Björklund and Moffitt (1987). It is the limit form of the LATE parameter, where the LATE parameter was introduced in Imbens and Angrist (1994) and the limit form of the LATE was introduced in Heckman (1997) and Angrist, Graddy and Imbens (2000). See also Heckman and Smith (1998).

at least one instrument that determines participation in the program but not outcomes. Assumption (b) is a technical assumption made primarily for convenience. Assumption (d) guarantees that the parameters of interest will be well defined. Assumption (e) is the assumption in the population of both a treatment and a control group for each X. These conditions impose testable restrictions on the data (Heckman and Vytlacil, 2000a). X does not have to be exogenous as long as one is evaluating programs in place rather than projecting to new populations.[53]

Without loss of generality, we include the elements of X in Z. We define $P(z)$ as the probability of receiving treatment conditional on $Z = z$: $P(z) \equiv \Pr(D = 1|Z = z) = F_{U_D}|X(\mu_D(z)|x)$, where $F_{U_D}|X(\cdot|x)$ denotes the distribution of U_D conditional on $X = x$. Without loss of generality, we impose the normalization that $U_D \sim \text{Unif}[0,1]$ so $\mu_D(z) = \text{P}(z)$. Vytlacil (2001) proves under assumptions (a)–(e) the selection model is equivalent to the LATE model of Imbens and Angrist (1994).

The average effect of treatment on those at the margin of participation in the program at level $U_D = u_D$ is the Marginal Treatment Effect (*MTE*): $\Delta^{MTE}(x, u_D) \equiv E(\Delta|X = x, U_D = u_D)$. It is the basis for unifying both estimators and treatment parameters. We develop the method of local instrumental variables (*LIV*) to estimate this parameter. It can be estimated from the derivative of $E(Y|X = x, P = p)$ with respect to p as I note below.

We establish that under conditions (a)–(e) all of the population treatment parameters used in the evaluation literature are weighted versions of the *MTE*. Thus for treatment parameter j,

$$\text{(25) Parameter}_j(x) = \int_0^1 \Delta^{MTE}(x, u_D)\omega_j(x, u_D)du_D.$$

More specifically we have

$$ATE(x) = \int_0^1 MTE(x, u_D)du_D$$

and

$$TT(x) = \int_0^1 MTE(x, u_D)g_x(u)du_D.$$[54]

The *LATE* parameter of Imbens and Angrist (1994) is

$$LATE(x, P(z), P(z')) = \left[\int_{P(z')}^{P(z)} MTE(X = x, U_D = u_D)du\right] \frac{1}{P(z) - P(z')}.$$

[53] Observe that there are no exogeneity requirements concerning X. A counterfactual "no feedback" condition corresponding to the classical noncausality assumptions of structural econometrics is required for interpretability so that conditioning on X does not mask the effects of D. Letting X_d denote a value of X if D set to d, a sufficient condition that rules out feedback from D to X is: $X_1 = X_0$ *a.e.*. Heckman and Vytlacil (2001d, 2002) discuss the role of exogeneity assumptions in projecting estimated treatment effects to new environments.

[54] $g_x(u)$ is defined in Heckman and Vytlacil (1999, 2000a).

The probability limit of *IV* estimators of treatment effects (as well as other estimators), may also be written as weighted averages of *MTE*.[55] Thus conditional on $X = x$, the probability limit of estimator k is

$$\text{(26) plim estimator } k(x) = \int_0^1 \Delta^{MTE}(x, u_D)\omega_k(x, u_D)du_D.$$

One can show that the weights corresponding to the conventional treatment parameters integrate to unity, as do the weights corresponding to many of the estimators including *IV* using $P(Z)$ as an instrument. In general the probability limits of the various estimators weight MTE differently than do the parameters.[56] Notice that the parameters and estimators of the form given above coincide if responses to treatment do not vary among individuals (given $X = x$) or if they do, that agents do not participate in the program on the basis of such variation. In the more general case, which describes most of the studies in the literature, the estimators and parameters differ.

(ii) Estimating MTE and Understanding Instrumental Variables

Consider applying the method of instrumental variables to estimate the standard treatment effect parameters. Economists have been using instrumental variables (*IV*) for over 70 years.[57] The intuition supporting the instrumental variables method is widely understood. It mimics experimental variation by using instrumental variable variation. In correlated coefficient regression model (C-1), *IV* is a solution to the problem that D is correlated with the error term ε_0. Standard instrumental variables methods break down in the more general case considered in case (C-3) (Heckman, 1997).

Consider using $P(z)$ as an instrument for D. Suppose that we seek to estimate the parameter treatment on the treated (TT) in case (C-3). Suppress the explicit dependence on X to simplify the notation. Under the assumptions made in Heckman and Vytlacil, the TT parameter may be written as

$$\begin{aligned} TT = E(\Delta | Z = z, D = 1) &= E(\Delta | P(z) \geq U_D) \\ &= \mu_1 - \mu_0 + E(\varepsilon_1 - \varepsilon_0 | P(z) \geq U_D) \\ &= \bar{\Delta} + E(\varepsilon_1 - \varepsilon_0 | P(z) \geq U_D). \end{aligned} \tag{27}$$

In case (C-1) and (C-2), the final term in the last two expressions is zero. Using (24), we may write

$$E(Y | Z = z) = \mu_0 + \bar{\Delta} + E(\varepsilon_1 - \varepsilon_0 | P(z) \geq U_D)P(z). \tag{28}$$

[55] Expression (26) below is related to similar expressions in Yitzhaki (1996) and Angrist, Graddy and Imbens (2000).

[56] Under conditions (a)–(e) the maximum possible difference between any two policy parameters representable in the form of (25) or estimators and parameters can be written as a product of the difference between the largest and smallest possible value of *MTE* and a measure of the distance between the two weights.

[57] See the history in Morgan (1990). The earliest use was by Wright (1928).

If $E(\varepsilon_1 - \varepsilon_0 | P(z) \geq U_D) = 0$, as in case (C-1) or (C-2), then we may apply the logic of instrumental variables to write for two values of z, z' such that $P(z) \neq P(z')$

$$E(Y|Z=z) = \mu_0 + [\bar{\Delta} + E(\varepsilon_1 - \varepsilon_0 | P(z) \geq U_D)]P(z)$$
$$E(Y|Z=z') = \mu_0 + [\bar{\Delta} + E(\varepsilon_1 - \varepsilon_0 | P(z') \geq U_D)]P(z')$$

and subtract the bottom equation from the top and divide by $P(z) - P(z')$ to obtain

$$\frac{E(Y \mid Z=z) - E(Y \mid Z=z')}{P(z) - P(z')} = \bar{\Delta} + \frac{E(\varepsilon_1 - \varepsilon_0 \mid P(z) \geq U_D)P(z) - E(\varepsilon_1 - \varepsilon_0 \mid P(z') \geq U_D)P(z')}{P(z) - P(z')}. \tag{29}$$

If $E(\varepsilon_1 - \varepsilon_0 | P(z) \geq U_D) = 0$, *IV* identifies "the" effect of treatment $\bar{\Delta}$, and all mean treatment parameters are the same. In the more general case, the method only identifies the combination of parameters in (29).

In case (C-1) and (C-2), $P(z)$ enters (28) linearly, as the variable multiplying the term in brackets. In case (C-3), $P(z)$ enters the expression in two places. Thus a test of whether (C-1) or (C-2) is valid, is a test of the linearity of (28) in terms of $P(z)$. This is also a test of the validity of the standard *IV* procedure for identifying $\bar{\Delta}$. See Figure 7 which shows two cases: (a) one where $E(Y|P=p)$ is linear in p so (C-1) and (C-2) apply and the general case (C-3) (corresponding to the dotted line) where $E(Y|P=p)$ is a nonlinear function of P.

Even if the test is failed so linear *IV* does not identify $\bar{\Delta}$, it is possible to extract important information from (28). To show this, use separability to write (28) in equivalent form:

$$E(Y \mid P(Z) = P(z)) = \mu_0 + \bar{\Delta}P(z) + \int_{-\infty}^{\infty}\int_{0}^{P(z)} (\varepsilon_1 - \varepsilon_0) f(\varepsilon_1 - \varepsilon_0 \mid U_D = u_D) du_D d(\varepsilon_1 - \varepsilon_0).$$

Differentiate with respect to $P(z)$ to obtain

$$\frac{\partial E(Y \mid P(Z) = P(z))}{\partial P(z)} = \bar{\Delta} + \int_{-\infty}^{\infty} (\varepsilon_1 - \varepsilon_0) f(\varepsilon_1 - \varepsilon_0 \mid U_D = P(z)) d(\varepsilon_1 - \varepsilon_0) = MTE.$$

From knowledge of the *MTE*, we can recover all of the treatment parameters by integrating up *MTE* so identified using the relationships given in subsection (i). From the derivative of the $E(Y|P=p)$ function with respect to p we can recover all of the parameters of the model. This is the method of local instrumental variables (*LIV*) introduced into this literature by Heckman and Vytlacil (1999, 2000a, 2001d).

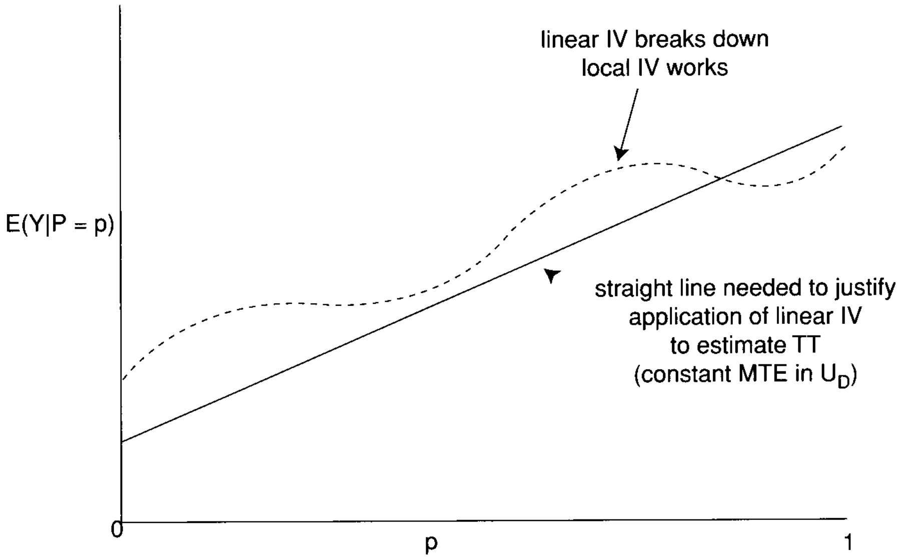

Figure 7. When Conventional *IV* Fails and When it Works.

When the support of $P(Z)$ is the full unit interval, we can estimate the treatment parameters by estimating *MTE* and using equation (25) to recover the parameter of interest. We also develop more general methods for estimating the parameters which do not require the support of $P(Z)$ to be the full unit interval (e.g., we allow Z to be discrete) and do not require estimating a derivative of a conditional expectation. We can replace *MTEs* by *LATEs* and integrals by sums. However, these methods still require support conditions on $P(Z)$, with the support condition depending on the particular parameter of interest. When these support conditions do not hold, we develop sharp bounds on the treatment parameters that exploit all of the information in the model and in the available data (Heckman and Vytlacil, 2000a, 2001b).

Using the *MTE* function we can organize all of the econometric estimators in the evaluation literature on the basis of whether or not they allow for selection into the program being evaluated on the basis of unobservable gains. See Figure 8. Conventional matching and conventional IV estimators assume no selection on the gains.[58] Thus *MTE* is flat for that class of estimators. Selection models, *LATE* and *LIV* allow for selection unobservables gains and hence are consistent with the Roy model. In this case *MTE* is a nontrivial function of u_d. The evidence from the microeconometric literature on the *MTE* that is reported in Table 3 suggests that across a variety of studies of economic phenomena the *MTE* is not constant and hence that matching and con-

[58] By matching, I mean estimators that exploit the full Rosenbaum-Rubin (1983) conditions $(Y_0,Y_1) \perp\!\!\!\perp D|X$. As noted in Heckman, Ichimura and Todd (1997) and Heckman, Ichimura, Smith and Todd (1998) to estimate *TT* one can get by with the weaker conditions $Y_0 \perp\!\!\!\perp D|X$ or $E(Y_0|X,D=1)=E(Y_0|X)$ which allows for selection gains $(Y_1 - Y_0)$.

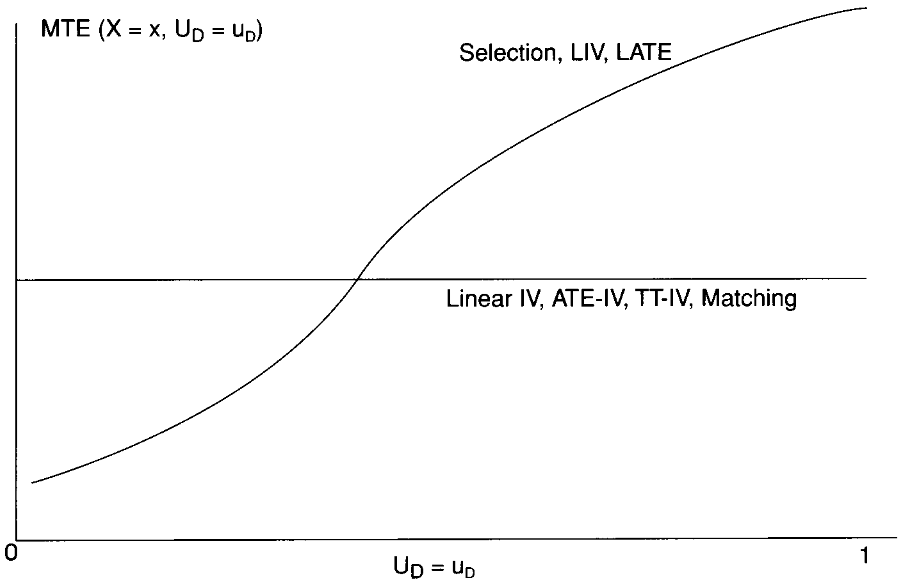

Figure 8. Conditions on *MTE* to Justify Alternative Estimation Strategies.

ventional IV do not identify any treatment parameter of interest.[59] When the support of $P(Z)$ is only partial, tight, simple and easily implemented bounds on all of the treatment parameters can be constructed. (See Heckman and Vytlacil 2000a, 2001b).

(iii) Policy Relevant Treatment Parameters

The conventional treatment parameters are justified on intuitive grounds. The link to cost benefit analysis and interpretable economic frameworks is obscure. Heckman and Smith (1998) develop the relationship between these parameters and the parameters of cost benefit analysis. Sometimes the traditional parameters answer interesting policy questions and sometimes they do not.

A more direct approach to defining economic treatment parameters pursued in Heckman and Vytlacil (2001a,c,d, 2002) is to postulate a policy question or decision problem of interest and to derive and estimate the parameter that answers it. Taking this approach does not in general produce the conventional treatment parameters.

We consider a class of policies that affect P, the probability of participation in a program, but do not directly affect *MTE*. An example from the economics of education would be policies that change tuition or distance to school but do not directly affect the gross returns to schooling. Define P as the baseline probability and define P^* as the probability produced under an alternative policy regime. For simplicity, compare policies using a Benthamite criterion and consider the effect of the policies on the mean utility of individuals

Table 3
Evidence of Selection on Unobservables

$$Y = DY_1 + (1 - D)Y_0$$
$$Y_1 = \mu_1(x) + U_1$$
$$Y_0 = \mu_0(x) + U_0$$
$$Z \perp\!\!\!\perp (U_0, U_1),\ Z \not\perp\!\!\!\perp D$$
$$D = 1(Y_2 = \mu_2(Z) + U_2 \geq 0),\ \text{where}$$

$\mu_2(Z) + U_2$ is the index determining selection into "1" or "0"

Hypothesis: No Selection on Unobservables

$$H_0 : E(U_1 - U_0 \mid D = 1, Z, X) = M(X)$$

(in normal model, $\sigma_{12} = \sigma_{02}$)

Study	Method	Finding
	Unionism	
Lee (1978)	Normal Selection Model $(H_0 : \sigma_{12} = \sigma_{02})$	$\sigma_{12} = \sigma_{02}$ Do not reject
Farber (1983)	Normal Selection Model $(H_0 : \sigma_{12} = \sigma_{02})$	$\sigma_{12} = \sigma_{02}$ Do not reject
Duncan and Leigh (1985)	Normal Selection Model $(H_0{:}\sigma_{12} = \sigma_{02})$	$\sigma_{12} = \sigma_{02}$ Do not reject
Robinson (1989)	Normal Selection Model $(\mu_1 - \mu_0)_{IV} = (\mu_1 - \mu_0)_{\text{normal}}$	$\sigma_{12} \neq \sigma_{02}$ Reject
	Schooling (College vs. High School)	
Willis and Rosen (1979)	Normal Selection Model $(H_0{:}\sigma_{12} = \sigma_{02})$	$\sigma_{12} \neq \sigma_{02}$ Reject
Heckman, Tobias and Vytlacil (2000)	Normal Selection Model $(H_0{:}\sigma_{12} = \sigma_{02})$	$\sigma_{12} \neq \sigma_{02}$ Reject
	Job Training	
Bjorklund and Moffitt (1987)	Normal Selection Model $(H_0{:}\sigma_{12} = \sigma_{02})$	$\sigma_{12} \neq \sigma_{02}$ Reject
Heckman, Ichimura, Smith and Todd (1998; Supplement)	$E(U_1 - U_0 \mid D = 1, Z, X) =$ $E(U_1 - U_0 \mid D = 1, X)$	Reject selection on unobservables
	Sectoral Choice	
Heckman and Sedlacek (1990)	Normal Selection Model $(H_0{:}\sigma_{12} = \sigma_{02})$	$\sigma_{12} \neq \sigma_{02}$ Reject
	Migration	
Pessino (1991)	Normal Selection Model $(H_0{:}\sigma_{12} = \sigma_{02})$	$\sigma_{12} \neq \sigma_{02}$ Reject
Tunali (2000)	$H_0 : E(U_1 - U_0 \mid D = 1) = 0$ (estimated using robust selection)	Cannot reject

with a given level of $X = x$. For utility V, assuming that $E(V(Y_1))$ and $E(V(Y_0))$ exist and are finite, we have that

$$(30)\quad E(V(Y) \mid \text{under policy*}, X = x) - E(V(Y) \mid \text{under baseline}, X = x)$$
$$= \int_0^1 \Delta_V^{MTE}(x, u_D)\omega^*(x, u_D)du_D,$$

where the policy weights are

$$\omega^*(x, u_D) = F_{P|X}(u_D|x) - F_{P^*|X}(u_D|x),$$

[59] In this table $Y_2 = \mu_2(Z) + U_2$. Most of the estimates in that table are derived from structural models based on normality assumptions described by (A-2) and (A-3) of Section 5 using the index models defined in Section 4, especially equations 8–11 and so are subject to the criticism that they are produced from functionally from dependent methods. See Heckman, Tobias and Vytlacil (2000) for the definition of the *MTE* parameter for the classical normal model. They show that if $\sigma_{12} \neq \sigma_{02}$, *MTE* is a nonconstant function of u_D.

where $F_{P\,|\,X}(\cdot\,|x)$ is the distribution of P conditional on $X = x$,[60] and $\Delta_V^{MTE}(x, u_D) = E(V(Y_1) - V(Y_0)\,|X = x,\ U_D = u_D) = E(\Delta_V|X = x,\ U_D = u_D)$ where $\Delta_V = V(Y_1) - V(Y_0)$. When V is the identity function, we compare mean outcomes, as in conventional cost benefit analysis. The policy parameter is a weighted average of the *MTE* as previously defined. Instead of hoping that conventional treatment parameters answer interesting economic questions, a better approach is to estimate Δ_V^{MTE} and weight it by the appropriate weight that is determined by how the policy changes the distribution of P.

An alternative approach to policy evaluation is to produce a policy weighted instrumental variable based on a specific choice for $\omega(x, u_D)$ that captures the effect of the policy change. If we choose the weights for the estimator $\omega_k(x, u_D)$ in (2) to coincide with the weights for the policy change, $\omega^*(x, u_D)$, in (3) we can produce an estimator that is tailored to the policy change of interest. Heckman and Vytlacil (2001c,d, 2002) establish that the policy relevant instrumental variable is $\left[\frac{f_{P*}(P)}{f_P(P)} - 1\right]$ where f_{P*} and f_P are the densities of P^* and P respectively. It is possible to determine the distribution of P and P^* independently of determining the other ingredients required for forming the policy relevant treatment parameter.

Figure 9, taken from the research of Carneiro, Heckman and Vytlacil (2001), plots the estimated *MTE* as a function of u_D for the returns to college education for a sample of white males in the United States in the late 1980s and early 1990s. It is increasing in u_D suggesting that monetary returns are the highest for those who are the least likely to go to school. Also plotted are the weighting functions for *MTE* that are implicit in defining *TT* and *ATE* and in using conventional linear IV with $P(z)$ as an instrument to estimate "the" effect of schooling on education. The weights for the different treatment parameters differ from each other and from those for *IV*. The fact that the *MTE* is rising implies that conventional methods of matching and linear *IV* do not identify *TT* or *ATE* in these data.

Figure 10, also taken from Carneiro, Heckman and Vytlacil (2001), displays the policy weight of *MTE* for three different policies defined at the base of the figure.[61] The treatment effect produced by *IV* weights *MTE* close to the weight required to evaluate policy III but it is far off the mark in evaluating policies I and II. The agreement between the *IV* weights for the *MTE* and the

[60] Keeping the conditioning on X implicit, we have

$$E(V(Y)\,|\text{baseline}) = \int_0^1 E(V(Y)\mid P(Z) = p)dF_P(p) =$$

$$\int_0^1 \left[\int_0^1 \mathbf{1}_{[0,p]}(u)E(V(Y_1)\mid U = u) + \mathbf{1}_{(p,1]}(u)E(V(Y_0)\mid U = u)du\right] dF_P$$

$$= \int_0^1 [(1 - F_P(u))E(V(Y_1)\mid U = u) + F_P(u)E(V(Y_0)\mid U = u)]\,du$$

where $\mathbf{1}_{\mathcal{A}}(u)$ is an indicator function for the event $u \in \mathcal{A}$. Thus comparing the baseline to the new regime

$$E_{P^*}(V(Y)) - E_P(V(Y)) = \int_0^1 E(\Delta_V\mid U = u)(F_P(u) - F_{P^*}(u))du.$$

[61] These policy weights are normalized by $\Delta\bar{P}$, the proportion of people in the population induced to change their schooling status by the intervention. This makes the policy weights comparable to the *IV* weights.

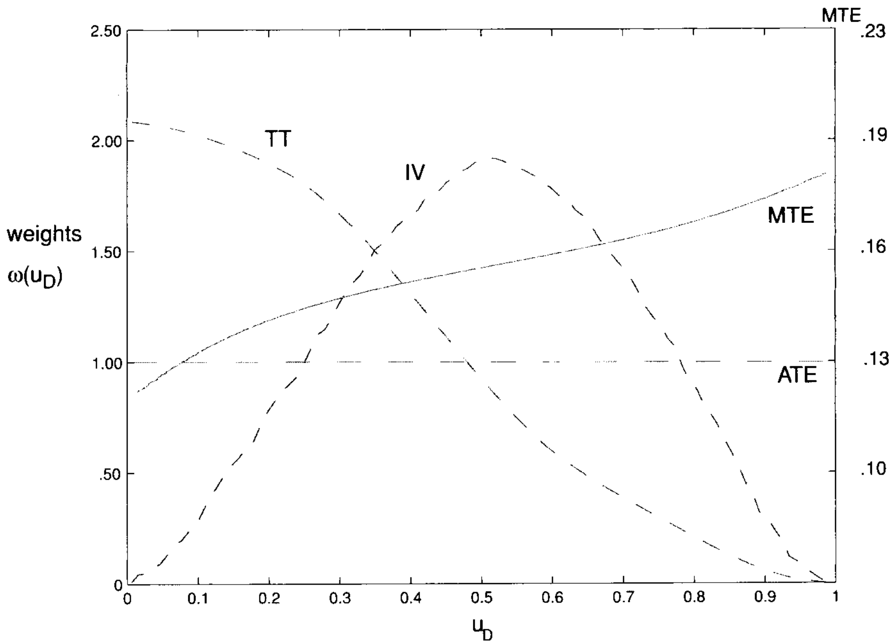

Figure 9. Marginal Treatment Effect vs *ATE, IV* and *TT* Weights. Source: Carneiro, Heckman, and Vytlacil (2001).

policy III weights is fortuitous. Social experiments that randomize people out of a program at the point where they apply and are accepted into it estimate *TT* under full compliance. (See Heckman, LaLonde and Smith, 1999; Heckman and Vytlacil, 2002). Given the shape of the weights for *TT*, such experiments do not accurately evaluate the effect of any of three policies on mean outcomes that are considered in Figure 10.

If the *MTE* is flat, there is one single "effect" of all policies and *IV* estimates that effect. In the general case, which covers most of the studies surveyed by Heckman and Vytlacil (2002) and in Table 3, different policies are associated with different weights and only by accident would linear *IV* identify the appropriate policy response.

When the support of P^* is not contained in the support of P, so that the policy intervention being studied extends P^* outside of historical data, it is necessary to make additional assumptions in order to perform a principled policy analysis. If parametric assumptions are made about $P(z)$, the probability is determined by historical data, and the intervention being studied changes the distribution of Z in a known way, it is straightforward to determine the distribution of $P^*(z)$, including the new support. However, *MTE* is only nonparametrically identified over the support of P. See Heckman and Vytlacil (2000a, 2002).

(iv) Extrapolating To New Populations

If we seek to apply the estimated *MTE* to new populations where the dependence between X and (U_0, U_1, U_D) is different from that in the sample used to

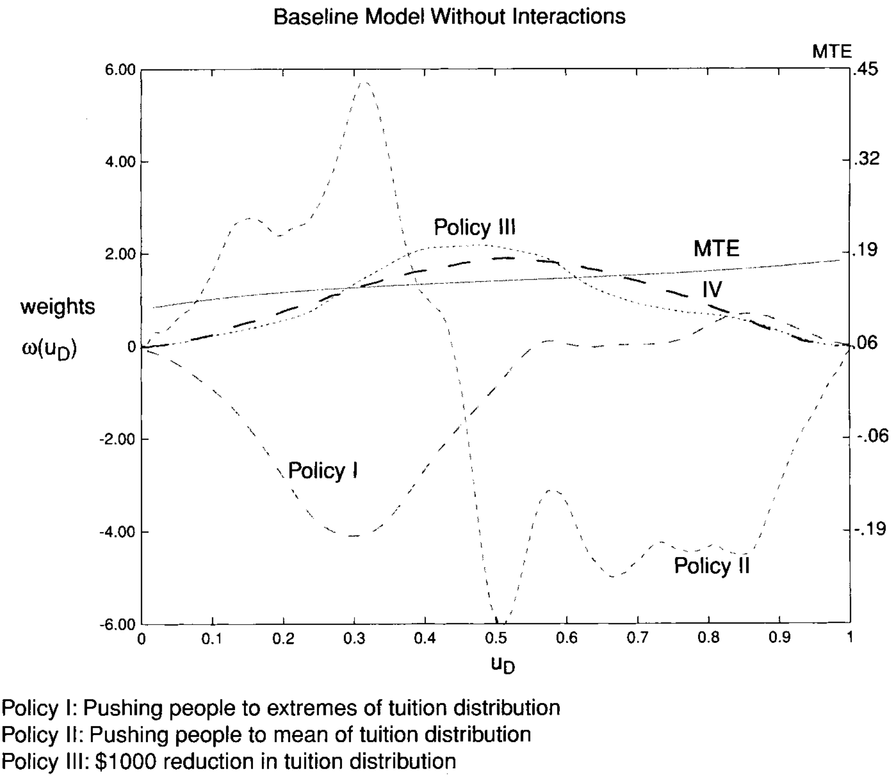

Figure 10. Marginal Treatment Effect vs Policy and *IV* weights. Source: Carneiro, Heckman, and Vytlacil (2001).

estimate it, or where the support of (X, U_D) is different from that used in the estimation sample, it is necessary to make the same types of independence and support assumptions previously discussed in Section 2. This point has been recognized by econometricians since the time of Haavelmo (1943). The treatment effect literature avoids structural assumptions by evading the questions addressed by structural econometrics.

(v) A Comparison of Three Approaches

Table 4 compares the strengths and limitations of the three approaches to policy evaluation that I have discussed in this lecture: the structural approach, the conventional treatment effect approach and the recently developed approach to treatment effects based on the *MTE*.

The approach based on the *MTE* shares with the structural approach interpretability of parameters. Like the structural approach it addresses a range of policy evaluation questions. The parameter is less comparable and less easily extrapolated across environments than are structural parameters, unless explicit assumptions are made about relationships between observables and unobservables both in estimation samples and target samples. It is, however, comparable across populations with different distributions of P and results from one population can be applied to another population provided the P-X dependence is controlled for. The cost of these advantages of the structural

Table 4. Comparisons of Three Approaches

	Structural Econometric Approach	Treatment Effect Approach	Approach Based on MTE
etability	Well defined economic parameters and welfare comparisons	Link to economics and welfare comparisons obscure	Interpretable in terms of willingness to pay; weighted averages of the MTE answer well posed economic questions
ɔf Questions ɜed	Answers many counterfactual questions	Focuses on one treatment effect or narrow range of effects	With support conditions, generates all treatment parameters
ılation to vironments	Provides ingredients for extrapolation	Evaluates one program in one environment	Can be partially extrapolated; extrapolates to new environments with different distributions of the probability of participation due solely to differences in distributions of Z.
ıbility ;udies	Policy invariant parameters comparable across studies	Not generally comparable	Partially comparable; comparable across environments with different distributions of the probability of participation due solely to differences in distributions of Z.
ımetric	Exogeneity, super exogeneity and selection bias	Selection bias	Selection bias
ɔlicies	Programs with either partial or universal coverage, depending on variation in data (prices/endowments)	Programs with partial coverage (treatment and control groups)	Programs with partial coverage (treatment and control groups)
ibrium	Need to link to time series data; parameters compatible with general equilibrium theory	Difficult because link to economics is not precisely specified	Can be linked to nonparametric general equilibrium models

approach is the greater range of econometric problems that must be solved. The conventional treatment approach produces parameters that cannot be linked to well posed economic models and hence do not provide building blocks for an empirically motivated general equilibrium analysis. The *MTE* estimates the preferences of the agents being studied and provides a basis for integration with well posed economic models.

(vi) Evidence From The Literature on Treatment Effects

Despite its limitations, the literature on treatment effects has produced a large number of important studies of economic policies in place that have changed the way we think about the effectiveness of economic policies. The literature is too vast to summarize in this lecture.

I briefly sketch some of the main empirical results from the literature on evaluating active labor market policies, drawing on my survey of this field written with LaLonde and Smith (Heckman, LaLonde and Smith, 1999). Looking across countries and over time, most active labor market policies are ineffective in promoting long term wage growth and employment. Application of very basic empirical methods shows that comparing post-program outcomes of trainees with preprogram outcome measures – the common practice in many evaluation schemes – dramatically overstates the gains from participation in training. This occurs because trainees typically experience a decline in their employment prior to entering training and training is a form of job search.[62] Much of the post program improvement in outcome measures would have occurred in the absence of any training program.

Application of basic principles, such as using a comparison group of non-trainees that is comparable in terms of geographic location, labor market histories and questionnaires administered to samples of trainees, goes a long way toward eliminating selection bias in evaluating social programs. However, such matching does not eliminate all of the bias, and accounting for selection on unobservables is important in getting accurate estimates of the commonly used "treatment on the treated" parameter.[63] (See Heckman, Ichimura, Smith and Todd, 1996, 1998, and Heckman, LaLonde and Smith, 1999).

[62] Ashenfelter (1978) demonstrated that trainees suffer a decline in earnings prior to enrollment in the program. Heckman (1978b) and Heckman and Robb (1985) present the first economic models of this phenomenon using a model based on present value earnings maximization as the criterion for program participation. Heckman, Ichimura, Smith and Todd (1998), Heckman and Smith (1998) and Heckman, LaLonde and Smith (1999) demonstrate that it is employment dynamics, not earnings dynamics, that affects enrollment into programs. Heckman, LaLonde and Smith (1999) present a search model that extends the analysis of Heckman and Robb (1985).

[63] An influential study of LaLonde (1986) cast doubt on the ability of nonexperimental econometric methods to evaluate social programs. Subsequent analysis by Heckman, Ichimura, Smith and Todd (1996, 1998) demonstrated that LaLonde's findings are generated by comparing incomparable people. His trainees live in different labor markets, are administered different questionaires and have different X values than his comparison group. Making trainees comparable to nontrainees eliminates most but not all of LaLonde's bias. See the important papers by Smith and Todd (2000, 2001) who reanalyze LaLonde's data and demonstrate that matching methods *do not* eliminate empirically important components of LaLonde's selection bias. For problems with social experiments see Heckman (1992), Heckman and Smith (1993, 1995), Manski, (1996), Heckman, LaLonde and Smith (1999), and Heckman and Vytlacil (2001a).

8. UNITING MACRO AND MICROECONOMETRICS: GENERAL EQUILIBRIUM POLICY EVALUATION

The evidence from microeconomic data has already had a substantial effect on the development of macroeconomic theory which is slowly abandoning the representative agent paradigm. I have already discussed how recognition of choices at the extensive margin have altered macroeconomic discussions of the labor market. Numerous other examples could be presented of evidence from micro data that has affected the development of macro theory. (See the survey in Browning, Hansen and Heckman, 1999.)

Microeconometric methods have been developed for evaluating a host of social programs put in place by the modern welfare state and for interpreting empirical economic relationships. But the scope of microeconometrics alone is necessarily limited. Many programs, like the tuition subsidy programs discussed in Section 7, are national in character and likely have general equilibrium effects. Expanding the stock of educated people is likely to reduce the return to educated labor. Reducing taxes on labor expands the supply of labor and reduces the real wage. Partial equilibrium methods can only go so far in evaluating the full impacts of large scale public programs. The treatment effect methodology is ineffective in analyzing programs with universal coverage unless entire economies can be used for the treatment and control groups.

A synthesis of macro and micro approaches is required to analyze policies instituted at the national level with general equilibrium impacts and to interpret equilibrium pricing relationships such as earnings functions or asset pricing equations. Cross sectional variation cannot identify the effects of prices and interest rates that are common across persons. Yet accounting for the feedback of capital markets (borrowing costs) on human capital or physical capital investment decisions is essential in investigating the full effects of national policies on skill formation. Moreover, cross sectional variation in wages, prices, or interest rates is unlikely to be exogenous and this raises separate econometric issues.

The required synthesis of macro general equilibrium and microeconometrics proposed in Orcutt (1962) has just begun but the first results from this research program are promising. The goal of this line of research is to develop an empirically grounded general equilibrium theory that will improve on calibration as a source of estimates for the parameters of general equilibrium models, and that will provide a rigorous empirical and theoretical foundation for evaluating large scale social programs like educational subsidies that alter prices and social security reforms that have universal coverage.

Heckman, Lochner and Taber (1998a,b,c, 1999) present a prototype for such a synthesis of micro and macro data. In that work we generalize the framework of Auerbach and Kotlikoff (1987) to formulate and estimate a perfect foresight, dynamic general equilibrium model of skill formation that generalizes the Roy model to a full dynamic setting with endogenous skill accumulation. Micro and macro data are combined to determine the parameters of the model that accounts for heterogeneity and self selection in the labor market, and explains rising wage inequality in the U.S over the past two decades. Policy analysis using this model indicates that failure to account for general equilibrium effects in the fashion common in the treatment effect literature and in much of the partial equilibrium microeconometric literature overstates, by an order of magnitude, the effects of tuition reductions on college enrollment. Accounting for general equilibrium effects is both substantively and theoretically important.[64] The challenge in this literature is to develop empirically credible structural relationships based on microdata that can be linked to macro aggregates.

9. SUMMARY AND CONCLUSIONS

In the past half century, economics has been enriched by vast new resources of microeconomic data. These data have opened the eyes of economists to the diversity and heterogeneity of economic life. They have enabled economists to understand more fully a vast array of social problems, and to evaluate social programs designed to solve those problems. Those who initiated the extensive collection of microeconomic data deserve our sincere gratitude.

These data challenged traditional econometric methodologies. Problems that appear to be unimportant when examining aggregate averages become central in analyzing micro data. These problems, and the policy concerns that motivated the systematic collection of microeconomic data, gave birth to modern microeconometrics. This field unites economics and statistics to produce interpretable summaries of microdata, to test theories of the individual using data on individuals and to construct economic counterfactuals.

The field of microeconometrics is flourishing. Substantial progress has been made in understanding the sources of identification of models, in relaxing arbitrary functional form and distributional assumptions, and in the design and analysis of new surveys. A more robust approach to policy evaluation is being developed that will make policy analysis more objective and empirically rigorous.

The field will continue to flourish if it renews itself by tackling new econometric problems that arise from new problems in economics. It will die if it seeks only to refine the original models that launched the field.

Important challenges to the field include the development of a microeconomic-data based general equilibrium theory for testing theory and evaluating the impacts of large scale policies. They also include the development of empirically credible econometric cost benefit schemes for the evaluation of micro policies that link the program evaluation literature more closely to economics. I am sure that microeconometricians will rise to these and other challenges and in future years will give you updates on research in this field from this podium.

[64] Lee (2000) presents an extension of this model for occupational choice rather than educational choice. Studies by Calmfors (1994) and Davidson and Woodbury (1993) demonstrate the importance of accounting for displacement in evaluating various active labor market policies. See the survey in Heckman, LaLonde and Smith (1999).

REFERENCES

1. Ahn, Hyungtaik and Powell, James. "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58 (1993): 3–29.
2. Altug, Sumru and Miller, Robert A. "Household Choices in Equilibrium." *Econometrica* 58 (May 1990): 543–570.
3. Altug, Sumru and Miller, Robert A."The Effect of Work Experience on Female Wages and Labour Supply." *Review of Economic Studies* 65 (January 1998): 45–85.
4. Amemiya, Takeshi. "Regression Analysis When the Dependent Variable is Truncated Normal." *Econometrica* 41 (November 1973): 997–1016.
5. Amemiya, Takeshi. "Multivariate Regression and Simultaneous Equation Models When The Dependent Variables Are Truncated Normal." *Econometrica* 46 (November 1974): 999–1012.
6. Amemiya, Takeshi. *Advanced Econometrics.* Cambridge, Massachusetts: Harvard University Press, 1985.
7. Andrews, Donald W. K. "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models." *Econometrica* 59 (March 1991): 307–45.
8. Angrist, Joshua; Graddy, Kathryn; and Imbens, Guido. "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish." *Review of Economic Studies* 67 (July 2000): 499–528.
9. Arabmazar, Abbas and Schmidt, Peter. "Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity, *Journal of Econometrics* 17 (November 1981): 253–58.
10. Ashenfelter, Orley. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics* 6 (February 1978): 47–57.
11. Athey, Susan and Haile, Philip. "Identification of Standard Auction Models," Unpublished manuscript, MIT, 2000.
12. Atkinson, Richard C.; Bower, Gordon H.; and Crothers, Edward J. *An Introduction to Mathematical Learning Theory.* New York: Wiley, 1965.
13. Auerbach, Allen and Kotlikoff, Lawrence. *Dynamic Fiscal Policy.* Cambridge: Cambridge University Press, 1987.
14. Barnow, Burt S.; Cain, Glen; and Goldberger, Arthur. "Issues in The Analysis of Selectivity Bias." In *Evaluation Studies,* Volume 5, edited by E. Stromsdorfer and G. Farkas. Beverly Hills, CA: Sage Publications, 1980.
15. Ben-Porath, Yoram. "Labor-Force Participation Rates and the Supply of Labor." *Journal of Political Economy* 81 (May–June 1973): 697–704.
16. Bera, Anil K.; Jarque, Carlos M.; and Lee, Lung-Fei "Testing the Normality Assumption in Limited Dependent Variable Models." *International Economic Review* 25 (October 1984): 563–578.
17. Bils, Mark, "Real Wages over the Business Cycle: Evidence from Panel Data." *Journal of Political Economy* 93 (August 1985): 666–689.
18. Bishop, Yvonne M.; Fienberg, Stephen E.; and Holland, Paul W. *Discrete Multivariate Analysis: Theory and Practice,* Cambridge, MA: MIT Press, 1974.
19. Björklund, Anders and Moffitt, Robert. "Estimation of Wage Gains and Welfare gains in Self-Selection Models." Review of Economics and Statistics 69 (February 1987): 42–49.
20. Blundell, Richard; Reed, Howard; and Stoker, Thomas. "Interpreting Movements in Aggregate Wages: The Role of Labor Market Participation." Center for Labor Economics, University of California, Berkeley, 1999.
21. Blundell, Richard and Smith, Richard J. "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables." *Journal of Econometrics* 64 (September–October 1994): 355–373.
22. Brown, Charles. "Black-White Earnings Ratios Since the Civil Rights Act of 1964: The Importance of Labor Market Dropouts." *Quarterly Journal of Economics* 99 (February 1984): 31–44.

23. Browning, Martin; Deaton, Angus; and Irish, Margaret. "A Profitable Approach to Labor Supply and Commodity Demands Over the Life Cycle." *Econometrica* 53 (May 1985): 503–543.
24. Browning, Martin; Hansen, Lars; and James Heckman. "Micro Data and General Equilibrium Models." In J. B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics:* Volume 1, Chapter 8, 543–633, 1999.
25. Butler, Richard and Heckman, James. "Government's Impact on the Labor Market Status of Black Americans: A Critical Review." in: *Equal Rights and Industrial Relations* (Industrial Relations Research Association, Madison, WI) 1977.
26. Cain, Glenn. "Regression and Selection Models to Improve Nonexperimental Comparisons." In *Evaluation and Experiment,* edited by Carl Bennett and Arthur Lumsdaine. New York: Academic Press, 1975.
27. Cain, Glenn and Watts, Harold. *Income Maintenance and Labor Supply,* Chicago; Merkham, 1973.
28. Calmfors, Lars. "Active Labor Market Policy and Unemployment – A Framework For The Analysis of Crucial Design Features." *OECD Economic Studies* No. 22 (Spring 1994): 7–47.
29. Cameron, Stephen and Heckman, James. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence For Five Cohorts of American Males." *Journal of Political Economy* 106 (April 1998): 262–333.
30. Cameron, Stephen and Heckman, James. "The Dynamics of Educational Attainment For Black, White and Hispanic Males." Forthcoming, *Journal of Political Economy* (June 2001): pp. 445–499.
31. Campbell, Donald T. and Stanley, Julian C. *Experimental and Quasi-Experimental Designs For Research.* Chicago, IL: Rand McNally, 1966.
32. Carneiro, Pedro; Heckman, James; and Vytlacil, Edward. "Estimating The Return to Education When It Varies Among Individuals." Unpublished working paper, University of Chicago, 2001, presented as *Economic Journal Lecture* at Royal Economic Society Meeting, Durham England, April, 2001.
33. Chandra, Amitabh. "Labor-Market Dropouts and the Racial Wage Gap: 1940–1990." *American Economic Review.* 90 (May 2000): 333–338.
34. Chiappori, Andre and Heckman, James. "Testing Moral Hazard vs. Adverse Selectionin Insurance Markets." Unpublished manuscript, University of Chicago, 2000.
35. Cogan, John F. "Fixed Costs and Labor Supply." *Econometrica* 49 (June 1981): 945–963.
36. Coleman, Thomas. "Essays on Aggregate Labor Market Business Cycle Fluctuations." Ph.D. Dissertation, University of Chicago, 1984.
37. Cosslett, Steven. "Maximum Likelihood Estimation From Choice Based Samples." *Econometrica* 49 (September 1981): 1289–1316.
38. Dagsvik, John K. "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes." *Econometrica* 62 (September 1994): 1179–1205.
39. Davidson, Carl and Woodbury, Stephen. "The Displacement Effect of Reemployment Bonus Programs." *Journal of Labor Economics* 11 (October 1993): 575–605.
40. Domencich, Thomas and McFadden, Daniel. *Urban Travel Demand.* Amsterdam: North Holland, 1975.
41. Donald, Stephen G. and Paarsch, Harry, J. "Identification, Estimation and Testing in Parametric Empirical Models of Auctions Within the Independent Private Values Paradigm Source." *Econometric Theory* 12 (August 1996): 517–67.
42. Duncan, Gregory M. and Leigh, Duane E. "The Endogeneity of Union Status: An Empirical Test." *Journal of Labor Economics* 3 (July 1985): 385–402.
43. Dustmann, Christian and Costas Meghir. "Wages, Experience and Seniority," London: Institute for Fiscal Studies, Working paper, 01/01, (January, 2001).
44. Eckstein, Zvi and Wolpin, Kenneth "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience." *Review of Economic Studies* 56 (July 1989): 375–390.

45. Eckstein, Zvi and Wolpin, Kenneth. "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities." *Econometrica* 67 (November 1999): 1295–1339.
46. Elbers, Chris and Ridder, Geert. "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *Review of Economic Studies* 49 (July 1982): 403–409.
47. Engle, Robert; Hendry, David F.; and Richard, Jean-Francois. "Exogeneity." *Econometrica* 51 (March 1983): 277–304.
48. Fair, Ray. *A Model of Macroeconomic Activity.* Cambridge: Ballinger Publishing Company, 1976.
49. Fair, Ray. *Testing Macroeconometric Models.* Cambridge: Harvard University Press, 1994.
50. Farber, Henry. "Worker Preferences For Union Representation." In *New Approaches to Labor Unions,* edited by Joseph Reid, *Research in Labor Economics,* Suppl. 2, Greenwich, Conn: JAI Press, 1983.
51. Fisher, Walter, D. *Clustering and Aggregation in Economics.* Baltimore: Johns Hopkins Press, 1969.
52. Flinn, Christopher and Heckman, James. "New Methods for Analyzing Structural Models of Labor Force Dynamics." *Journal of Econometrics* 18 (January 1982): 115–168.
53. Glynn, Robert; Laird, Nan; and Rubin, Donald. "Selection Modeling vs. Mixture Modeling with Nonignorable Response." In *Drawing Inferences from Self-Selected Samples,* edited by Howard Wainer. New York: Springer-Verlag, 1986.
54. Goldberger, Arthur. *Econometric Theory. New York:* John Wiley, 1964.
55. Goldberger, Arthur. "Selection Bias in Evaluating Treatment Effects." Discussion paper No. 123–172, Institute for Research on Poverty, University of Wisconsin, 1972.
56. Goldberger, Arthur. "Abnormal Selection Bias." In *Studies in Econometrics, Time Series and Multivariate Statistics.* edited by Samuel Karlin, Takeshi Amemiya, Leo Goodman, and T.W. Anderson. New York: Academic Press, 1983.
57. Goodman, Leo. "The Analysis of Cross Classified Data: Independence, Quasi-Independence and Interactions in Contingency Tables With or Without Missing Entries." *Journal of The American Statistical Association* 63 (1968): 1091–1131.
58. Goodman, Leo. "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications." *Journal of The American Statistical Association* 65 (March 1970): 226–256.
59. Gorman, W. M. "A Possible Procedure for Analyzing Quality Differentials in the Egg Market." *Review of Economic Studies* 47 (October 1980): 843–856.
60. Gourieroux, C., Laffont, J. J. and Monfort, A. "Coherency Conditions in Simultaneous Linear Equation Models With Endogenous Switching Regimes." *Econometrica* 48 (April 1980): 675–695.
61. Green, H.A. John. *Aggregation in Economics.* Princeton: Princeton University Press, 1964.
62. Gronau, Reuben. "Wage Comparisons – A Selectivity Bias." *Journal of Political Economy* 82 (1974): 1119–1144.
63. Haavelmo, Trygve. "The Statistical Implications of a System of Simultaneous Equations." *Econometrica* 11 (1943): 1–12.
64. Haberman, Shelby. *Analysis of Frequency Data.* Chicago: University of Chicago Press, 1974.
65. Hansen, Gary D. "Indivisible Labor and the Business Cycle." *Journal of Monetary Economics* 16 (November 1985): 309–327.
66. Hansen, Karsten; Heckman, James; and Vytlacil, Edward. "Dynamic Counterfactuals and Dynamic Treatment Effects." Unpublished manuscript, University of Chicago, 2000.
67. Hansen, Lars and Sargent Thomas. "Formulating and Estimating Dynamic Linear Rational Expectations Models." *Journal of Economic Dynamics and Control* 2 (February 1980): 7–46.
68. Hansen, Lars and Sargent, Thomas. "Two Difficulties in Interpreting Vector Autoregressions." In *Rational Expectations Econometrics,* edited by Lars Hansen and Thomas Sargent. Boulder, Colorado: Westview Press, 1991.

69. Hansen, Lars and Singleton, Kenneth J. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica.* 50 (1982): 1269–1286.
70. Hausman, Jerry, A. "The Effects of Wages, Taxes, and Fixed Costs on Women's Labor Force Participation." *Journal of Public Economics* 14 (October 1980): 161–194.
71. Hausman, Jerry, A. "Taxes and Labor Supply." In *Handbook of Public Economics,* vol. 1, edited by A.J. Auerbach and M. Feldstein. New York: North Holland, 1985.
72. Heckman, James. "Dummy Endogenous Variables." Unpublished manuscript, Columbia University, 1973.
73. Heckman, James. "Shadow Prices, Market Wages and Labor Supply." *Econometrica* 42 (July 1974a): 679–694.
74. Heckman, James. "Effects of Child-Care Programs on Women's Work Effort." *Journal of Political Economy,* vol. 82 (2) (March/April 1974): 5136–5143. Reprinted in *Economics of the Family: Marriage, Children, and Human Capital,* edited by Theodore W. Schultz. Chicago: University of Chicago Press, 1974b.
75. Heckman, James. "Life Cycle Consumption and Labor Supply: An Explanation of the Relationship Between Income and Consumption over the Life Cycle." *American Economic Review* 64 (March 1974c): 188–194.
76. Heckman, James. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and A Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5 (Fall 1976a): 475–492.
77. Heckman, James. "Simultaneous Equations Models with Continuous and Discrete Endogenous Variables and Structural Shifts." In *Studies in Nonlinear Estimation,* edited by S. Goldfeld and R. Quandt. Cambridge, Ballinger, 1976b.
78. Heckman, James. "A Life Cycle Model of Earnings, Learning and Consumption." *Journal of Political Economy* 84 (August 1976c): S11–S44.
79. Heckman, James. "Dummy Endogenous Variables in a Simultaneous Equations System." *Econometrica* 46 (July 1978a): 931–959.
80. Heckman, James. "Longitudinal studies in labor economics: a methodological review." Unpublished manuscript, presented at SSRC Conference, Mt. Kisco NY, September, 1978b.
81. Heckman, James. "A Partial Survey of Recent Research on the Labor Supply of Women." *American Economic Review* 68 (May 1978c):200–207.
82. Heckman, James. "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence Against The Hypothesis of Spurious State Dependence." *Annals de INSEE* Paris, 1978d.
83. Heckman, James. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (January 1979): 153–161.
84. Heckman, James. "Addendum to Sample Selection Bias As A Specification Error." In *Evaluation Studies Review Annual,* Volume 5 edited by Ernst Stromsdorfer and George Farkas. San Francisco, CA: Sage Publications, 1980a.
85. Heckman, James. "Sample Selection Bias as a Specification Error with an Application to the Estimation of Labor Supply Functions." In *Female Labor Supply: Theory and Estimation,* edited by James Smith. Princeton: Princeton University Press, 1980b.
86. Heckman, James. "Statistical Models for Discrete Panel Data." In *Structural Analysis of Discrete Data with Economic Applications* edited by Charles Manski and Daniel McFadden. Cambridge: MIT Press, 1981a.
87. Heckman, James. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence." In *Structural Analysis of Discrete Data with Econometric Applications,* edited by Charles Manski and Daniel McFadden. Cambridge: M.I.T. Press, .1981b.
88. Heckman, James. "Heterogeneity and State Dependence." In *Studies in Labor Markets,* edited by Sherwin Rosen. Chicago: University of Chicago Press, 1981c.

89. Heckman, James. "Comments on the Ashenfelter and Kydland Papers" ["Macroeconomic Analyses and Microeconomic Analyses of Labor Supply"] and ["Labor-Force Heterogeneity and the Business Cycle"]. Carnegie-Rochester Conference Series on Public Policy 21 (Autumn 1984): 209–224.
90. Heckman, James. "Selection Bias." In *The New Palgrave: A Dictionary of Economics*, edited by John Eatwell, Murray Milgate, and Peter Newman. New York: Stockton, 1987.
91. Heckman, James. "Randomization and Social Policy Evaluation." In *Evaluating Welfare and Training Programs*, edited by Charles Manski and Irwin Garfinkel. Cambridge: Harvard University Press, 1992.
92. Heckman, James. "What Has Been Learned About Labor Supply In the Past Twenty Years?" *AEA Papers and Proceedings*, 83(May 1993):116–121.
93. Heckman, James. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32 (Summer 1997): 441–462.
94. Heckman, James. "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective." *Quarterly Journal of Economics* 115 (February 2000):45–97.
95. Heckman, James and Borjas, George."Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence." *Economica* 47 (August 1980): 247–283
96. Heckman, James and Honoré, Bo. "The Empirical Content of the Roy Model." *Econometrica* 58 (September 1990): 1121–1149.
97. Heckman, James; Ichimura, Hidehiko; Smith, Jeffrey; and Todd, Petra. "Sources of Selection Bias in Evaluating Social Programs: An Interpretation of Conventional Measures and Evidence on the Effectiveness of Matching as a Program Evaluation Method." *Proceedings of the National Academy of Sciences* 93 (November 1996): 13416–13420.
98. Heckman, James; Ichimura, Hidehiko; Smith, Jeffrey and Todd, Petra. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66 (September 1998): 1017–1098.
99. Heckman, James; Ichimura, Hidehiko; and Todd, Petra. "Matching As An Econometric Evaluation Estimator: Evidence From Evaluating A Job Training Programme." Review of Economic Studies 64 (October 1997): 605–654.
100. Heckman, James; Killingsworth, Mark and MaCurdy, Thomas. "Recent Theoretical and Empirical Studies of Labor Supply: A Partial Survey." In T*he Economics of the Labour Market*, edited by Zmira Hornstein, Joseph Grice and Alfred Webb. London: Her Majesty's Stationery Office, 1981b.
101. Heckman, James; LaLonde, Robert; and Smith, Jeffrey. "The Economics and Econometrics of Active Labor Market Programs." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Chapter 31, Amsterdam: North Holland, 1999.
102. Heckman, James; Lochner, Lance; and Taber, Christopher. "Explaining Rising Wage Inequality: Explorations With A Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents." *Review of Economic Dynamics* 1 (January 1998a): 1–58.
103. Heckman, James; Lochner, Lance; and Taber, Christopher. "General Equilibrium Treatment Effects: A Study of Tuition Policy." *American Economic Review* 88 (May 1998b): 381–386.
104. Heckman, James; Lochner, Lance; and Taber, Christopher. "Tax Policy and Human Capital Formation." *American Economic Review*, 88(May 1998c): 293–297.
105. Heckman, James; Lochner, Lance; and Taber, Christopher. "General Equilibrium Cost Benefit Analysis of Education and Tax Policies." in Gustav Ranis and Lakshmi K. Raut, eds., *Trade, Growth and Development: Essays in Honor of T.N. Srinivasan*, Elsevier Science, B.V., Amsterdam, 1999.
106. Heckman, James, Lyons, Thomas; and Todd, Petra. "Understanding Black-White Wage Differentials, 1960–1990." *American Economic* Review 90 (May 2000): 344–49.

107. Heckman, James and MaCurdy, Thomas. "A Life Cycle Model of Female Labor Supply." *Review of Economics and Statistics* 47 (January 1980): 47–74.
108. Heckman, James and MaCurdy, Thomas. "New Methods For Estimating Labor Supply Functions." In *Research in Labor Economics,* Volume 4, edited by R. Ehrenberg. Greenwich, Conn.: JAI Press, 1981.
109. Heckman, James and MaCurdy, Thomas. "A Simultaneous Equations Linear Probability Model." *Canadian Journal of Economics* 18 (February 1985): 28–37.
110. Heckman, James and MaCurdy, Thomas. "Labor Econometrics." In *Handbook of Econometrics,* vol. 3, edited by Zvi Griliches and Michael Intriligator. Amsterdam: North-Holland, 1986.
111. Heckman, James and Robb, Richard. "Alternative Methods for Evaluating the Impact of Interventions." In *Longitudinal Analysis of Labor Market Data,* edited by James Heckman and Burton Singer. Cambridge University Press for Econometric Society Monograph Series, New York, NY, 1985.
112. Heckman, James and Robb, Richard. "Alternative Methods for Solving the Problem of Selection Bias in Evaluating The Impact of Treatments on Outcomes." In *Drawing Inferences From Self-Selected Samples,* edited by Howard Wainer.New York: Springer-Verlag, 1986.
113. Heckman, James and Sedlacek, Guilherme. "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market." *Journal of Political Economy* 98 (December 1985): 1077–1125.
114. Heckman, James and Sedlacek, Guilherme. "Self Selection and The Distribution of Hourly Wage Rates." *Journal of Labor Economics,* 8, Part 2 (January 1990): S329–S363.
115. Heckman, James and Singer, Burton. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52 (March 1984):271–320.
116. Heckman, James and Singer, Burton. "Econometric Analysis of Longitudinal Data." in Zvi Griliches and Michael Intriligator, *Handbook of Econometrics,* Volume 3, Amsterdam: North-Holland, 1985a.
117. Heckman, James and Singer, Burton. "Social Science Duration Analysis." In *Longitudinal Analysis of Labor Market Data,* edited by J. Heckman and B. Singer. Cambridge, UK: Cambridge University Press, 1985b.
118. Heckman, James and Smith, Jeffrey. "Assessing The Case For Randomized Evaluation of Social Programs." In *Measuring Labour Market Measures,* edited by Karsten Jensen and Per Konstoj Madsen. Copenhagen: Ministry of Labour, 1993.
119. Heckman, James and Smith, Jeffrey. "Assessing The Case for Social Experiments." *Journal of Economic Perspectives* 9 (Spring 1995): 85–100.
120. Heckman, James and Smith, Jeffrey. "Evaluating The Welfare State." *In Econometrics and Economics in the 20th Century:* The Ragnar Frisch Centenary, edited by S. Strom. New York: Cambridge University Press for Econometric Society Monograph Series, 1998.
121. Heckman, James and Snyder, James. "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating The Preferences of Legislators." *RAND Journal of Economics* 28 (Special Issue 1997): S142–S189.
122. Heckman, James; Tobias, Justin; and Vytlacil, Edward. "Simple Estimators for Alternate Treatment Parameters in a Latent Variable Framework with an Application to Estimating the Returns to Schooling." NBER Working Paper, No. W7950, 2000.
123. Heckman, James, and Todd, Petra. "Understanding Black-White Wage Differentials over the Last Fifty Years," Unpublished manuscript, University of Chicago, 1999.
124. Heckman, James and Vytlacil, Edward. "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling when the Return is Correlated with Schooling." *Journal of Human Resources* 33 (Fall 1998): 33:974–987.
125. Heckman, James and Vytlacil, Edward. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences* 96 (April 1999): 4730–4734.

126. Heckman, James and Vytlacil, Edward. "Local Instrumental Variables." In *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Theory and Econometrics: Essays in Honor of Takeshi Amemiya,* edited by C. Hsiao, K. Morimune, and J. Powell. Cambridge: Cambridge University Press, 2000a.
127. Heckman, James and Vytlacil, Edward. "The Relationship Between Treatment Parameters Within a Latent Variable Framework." *Economics Letters* 66 (January 2000b): 33–39.
128. Heckman, James and Vytlacil, Edward. "Causal Parameters, Treatment Effects and Randomization." 2001a. unpublished ms., University of Chicago.
129. Heckman, James and Vytlacil, Edward. "Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effects." In *Econometric Evaluations of Active Labor Market Policies in Europe,* edited by M. Lechnerand F. Pfeiffer. Forthcoming, 2001b.
130. Heckman, James and Vytlacil, Edward. "Policy Relevant Treatment Parameters." forthcoming, *American Economic Review* (May 2001c).
131. Heckman, James and Vytlacil, Edward. "Structural Equations, Treatment Effects and Econometric Policy Evaluation." Fisher-Schultz Lecture. Revised 2001. Forthcoming, *Ecometrica* (2001d).
132. Heckman, James and Vytlacil, Edward. "Econometric Evaluation of Social Programs." In *Handbook of Econometrics,* Volume 6, edited by James Heckman and Edward Leamer. Amsterdam: North Holland, 2002.
133. Heckman, James and Willis, Robert. "A Stochastic Model of Reproduction: An Econometric Approach." *Papers and Proceedings of the American Statistical Association, Social Statistics Section,* 1974
134. Heckman, James and Willis, Robert. "Estimation of a Stochastic Model of Reproduction: An Econometric Approach." In *Household Production and Consumption,* vol. 139, edited by Nestor Terleckyj. New York: Columbia University Press, 1975.
135. Heckman, James and Willis, Robert. "A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women." *Journal of Political Economy* 85 (February 1977): 27–58.
136. Holland, Paul. "A Comment on Remarks by Rubin and Hartigan." In *Drawing Inferences from Self Selected Samples,* edited by Howard Wainer. New York: Springer-Verlag, 1986
137. Holt, D. "Review of *Planning and Analysis of Observational Studies.*" *Journal of the American Statistical Association* 80 (September 1985): 772–773.
138. Hong, Han. "Equilibrium and Econometric Models of Asymmetric Auctions." Ph.D. dissertation, Stanford University, 1998.
139. Honoré, Bo and Kyriazidou, Ekaterini. "Panel Data Discrete Choice Models with Lagged Dependent Variables." *Econometrica* 68 (July 2000): 839–874.
140. Horowitz, Joel and Manski, Charles. "Identification and Robustness with Contaminated and Corrupted Data." *Econometrica* 63 (March 1995): 281–302.
141. Hurwicz, Leonid. "On the Structural Form of Interdependent Systems." In *Logic, Methodology and Philosophy of Science,* edited by Ernst Nagel, Patrick Suppes and Alfred Tarski. Stanford, California: Stanford University Press, 1962.
142. Ichimura, Hidehiko and Thompson, T. Scott. "Maximum Likelihood Estimation of a Binary Choice Model with Random Coecients of Unknown Distribution." *Journal of Econometrics* 86 (October 1998): 269–295.
143. Imbens, Guido and Angrist Joshua. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (March 1994): 467–475.
144. Juhn, Chinhui. "Labor Market Dropouts, Selection Bias, and Trends in Black and White Wages." Unpublished manuscript, University of Houston, 1997.
145. Keane, Michael and Wolpin, Kenneth "The Career Decisions of Young Men." *Journal of Political Economy* 105 (June 1997): 473–522
146. Keane, Michael and Wolpin, Kenneth. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." Unpublished manuscript Unversity of Pennsylvania, 1999.
147. Killingsworth, Mark. *Labor Supply.* (Cambridge: Cambridge University Press, 1983).

148. Kirman, Alan P. "Whom or What Does the Representative Individual Represent?" *Journal of Economic Perspectives* 6 (Spring 1992):117–36.
149. Knight, Frank. *Risk, Uncertainty and Profit.* New York: Houghton Mifflin Co., 1921.
150. Laffont, Jean-Jacques; Ossard, Herve; and Quang Vuong. "Econometrics of First-Price Auctions." *Econometrica* 63 (July 1995): 953–980.
151. LaLonde, Robert. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76 (September 1986): 604–620.
152. Lancaster, Kelvin. "A New Approach To Consumer Theory." *Journal of Political Economy* 74 (April 1966): 132–157.
153. Lancaster, Kelvin. *Consumer Demand: A New Approach.* New York: Columbia University, 1971.
154. Lancaster, Tony "Econometric Methods for the Duration of Unemployment," *Econometrica* 47 (July 1979): 939–56.
155. Lancaster, Tony. *Econometric Analysis of Transition Data.* Cambridge, UK: Cambridge University Press for Econometric Society Monograph Series, 1990.
156. Lancaster, Tony and Nickell, Stephen "The Analysis of Re-Employment Probabilities for the Unemployed." *Journal of Royal Statistical Society, Series A* 134 (Part 2 1980): 141–52.
157. Layard, Richard; Nickell, Stephen; and Jackman, Richard. *Unemployment: Economic Performance and The Labor Market.* Oxford: Oxford University Press, 1991.
158. Lee, Donghoon. "An Estimable Dynamic General Equlibrium Model of Work, Schooling and Occupational Choice." Unpublished working paper, University of Pennsylvania, 2000.
159. Lee, Lung-Fei. "Unionism and Wage Rates; A Simultaneous Equations Model with Qualitative and Limited Dependent Variables." *International Economic Review* 19 (June 1978): 415–433.
160. Lee, Lung-Fei. "Some Approaches To The Correction of Selectivity Bias." *Review of Economic Studies* 49 (April, 1982): 355–372.
161. Lewis, H. Gregg. *Unionism and relative wages in the United States; an empirical inquiry.* Chicago: University of Chicago Press, 1963.
162. Lewis, H. Gregg. "Comments on Selectivity Biases in Wage Comparisons." *Journal of Political Economy* 82 (November–December 1974):1145–55.
163. Lillard, Lee and Willis, Robert. "Dynamic Analysis of Earnings Mobility." *Econometrica* 46 (September 1978): 985–1012.
164. Lucas, Robert. "Econometric Policy Evaluation: A Critique." In *The Phillips Curve and Labor Markets,* Carnegie Rochester Conference Series on Public Policy, 1976.
165. MaCurdy, Thomas "Two Essays on The Life Cycle." Ph.D. Thesis, University of Chicago, 1978.
166. MaCurdy, Thomas. "An Empirical Model of Labor Supply in a Life-Cycle Setting." *Journal of Political Economy* 89 (December 1981): 1059–85.
167. McFadden, Daniel. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics,* edited by P. Zarembka. New York: Academic Press, 1974.
168. McFadden, Daniel. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data with Econometric Applications,* edited by Charles F. Manski and Daniel McFadden. Cambridge, MA: MIT Press, 1981.
169. Mallar, Charles D. "The Estimation of Simultaneous Probability Models." *Econometrica* 45 (October 1977): 1717–22.
170. Manski, Charles F. "Anatomy of the Selection Problem." *Journal of Human Resources* 24 (Summer 1989): 343–60.
171. Manski, Charles F. "Nonparametric Bounds on Treatment Effects." *American Economic Review* (May 1990): 319–323.
172. Manski, Charles F. "The Selection Problem." In *Advances in Econometrics: Sixth World Congress,* edited by Christopher Sims. Cambridge: Cambridge University Press, 1994.
173. Manski, Charles F. *The Identification Problem in the Social Sciences.* Cambridge: Harvard University Press, 1995.

174. Manski, Charles F. "Learning about Treatment Effects from Experiments with Random Assignment of Treatments." *Journal of Human Resources* 31 (Fall 1996): 709–33.
175. Manski Charles F. and Lerman, Steven. "The Estimation of Choice Probabilities From Choice-Based Samples." *Econometrica* 45 (November 1977): 1977–1988.
176. Manski, Charles F. and McFadden, Daniel."Alternative Estimators and Sample Designs for Discrete Choice Analysis." In *Structural Analysis of Discrete Data With Econometric Applications,* edited by Charles F. Manski and Daniel McFadden. Cambridge, MA: MIT Press,1981.
177. Marschak, Jacob. "Economic Measurements for Policy and Prediction." In *Studies in Econometric Method,* edited by William Hood and Tjalling Koopmans. New York: John Wiley, 1953.
178. Mincer, Jacob. "Labor Force Participation of Married Women." In *Aspects of Labor Economics,* edited by H. Gregg Lewis. Princeton: Princeton University Press, 1962.
179. Morgan, Mary. *The History of Econometric Ideas.* Cambridge: Cambridge University Press, 1990.
180. Nelson, Forrest and Olson, Lawrence. "Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables." *International Economic Review* 19 (October 1978): 695–709.
181. Newey, Whitney K. "The Asymptotic Variance of Semiparametric Estimators." *Econometrica* 62 (November 1994): 1349–1382.
182. Newey, Whitney K. and McFadden, Daniel. "Large Sample Estimation and Hypothesis Testing." In *Handbook of Econometrics,* edited by Robert Engle and Daniel McFadden. New York: Elsevier-North Holland, 1994.
183. Olsen, Randall, "A Least Squares Correction for Selectivity Bias," *Econometrica* 48 (November, 1980): 1815–1820.
184. Orcutt, Guy H. "Microanalytic Models of the United States Economy: Need and Development." *American Economic Review* 52 (May 1962): 229–240.
185. Pakes, Ariel. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica* 54 (July 1986): 755–84.
186. Pessino, Carola. "Sequential Migration Theory and Evidence from Peru." *Journal of Development Economics* 36 (1991): 55–87.
187. Peterson, Arthur. "Bounds For A Joint Distribution Function With Fixed Sub-Distribution Functions: Application To Competing Risks." *Proceedings of the National Academy of Sciences.* 73(January 1976): 11–13.
188. Phelps, Edmund. *Inflation Policy and Unemployment Theory* New York: W.W. Norton and Company, 1972
189. Quandt, Richard. "A Probabilistic Model of Consumer Behavior." *Quarterly Journal of Economics* 70 (November 1956): 507–536.
190. Quandt, Richard. *The Demand for Travel.* Lexington, Mass: D.C. Health, 1970.
191. Ransom, Michael R. "A Comment on Consumer Demand Systems with Binding Nonnegatively Constraints. *Journal of Econometrics* 34 (March 1987): 355–59.
192. Rao, C. R. "On Discrete Distributions Arising Out of Methods of Ascertainment." In *Classical and Contagious Distributions,* edited by Ganapati Patil, Calcutta: Pergamon Press, 1965.
193. Rao, C. R. "Weighted Distributions." In *A Celebration of Statistics,* edited by Anthony Atkinson and Stephen Fienberg. Berlin: Springer-Verlag, 1985.
194. Robins, James, "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies." In *Health Service Research Methodology: A Focus on AIDS,* edited by L. Sechrest, H. Freeman, and A. Mulley. Washington, DC: U.S. Public Health Service, 1989.
195. Robinson, Chris. "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models." *Journal of Political Economy* 97 (June 1989): 639–67.
196. Robinson, Peter. "Root-*N*-Consistent Semiparametric Regression." *Econometrica* 56 (1998): 931–954.

197. Rogerson, Richard. "Indivisible Labor, Lotteries and Equilibrium." *Journal of Monetary Economics* 21 (January 1988): 3–16.
198. Rosenbaum, Paul. *Observational Studies*, Chapter 4. Springer-Verlag, Leipzig, Germany, 1995.
199. Rosenbaum, Paul and Rubin, Donald. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (April 1983): 41–55.
200. Roy, A. D. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (January 1951):135–146.
201. Rubin, Herman. "Note on Random Coefficients" in *Statistical Inference in Dynamic Economic Models,* edited by Tjalling C. Koopmans, Cowles Commission For Research in Economics, Monograph 10, New York, Wiley, 1950.
202. Rust, John. "Optimal Replacement of GMC Bus Engines: An empirical model of Harold Zurcher." *Econometrica* 55 (September 1987): 999–1033.
203. Rust, John. "Numerical Dynamic Programming in Economics." *Handbook of Computational Economics,* Vol. 1. New York: Elsevier Science, 1996.
204. Schmidt, Peter. "Constraints on the Parameters in Simultaneous Tobit and Probit Models." In *Structural Analysis of Discrete Data with Econometric Applications,* edited by Charles F. Manski and Daniel McFadden. Cambridge, MA: MIT Press, 1981
205. Sheps, Mindel and Menken, Jane. *Mathematical Models of Conception and Birth.* Chicago: University of Chicago Press, 1973.
206. Smith, James and Welch, Finis. "Closing The Gap: Forty Years of Economic Progress for Blacks." Santa Monica, Ca: Rand Corporation, 1986.
207. Smith, Jeffrey and Todd, Petra. "Does Matching Overcome LaLonde's Critique of Nonexperimental Evaluation Methods?" Unpublished manuscript, University of Western Ontario, 2000.
208. Smith, Jeffrey and Todd, Petra. "Reconciling Conflicting Evidence on the Performance of Propensity Score Matching Methods." Papers and Proceedings, *American Economic Review* (May 2001), 112–118.
209. Stafford, Frank. "How the PSID Came to Be: A Life Course and Genealogical Panel of Families." Working paper, Survey Research Center, University of Michigan, 2001.
210. Telser, Lester. "Iterative Estimation of A Set of Linear Regression Equations." *Journal of the American Statistical Association* 59 (September 1964): 845–862
211. Theil, Henri. *Linear Aggregation of Economic Relations.* Amsterdam: North Holland, 1954
212. Theil, Henri. *Economic Forecasts and Policy.* second edition. Amsterdam: North Holland, 1961.
213. Theil, Henri. *Principles of Econometrics.* New York: Wiley, 1971.
214. Thurstone, Lewis. "A Law of Comparative Judgement." *Psychophysical Review* 34 (1927): 273–286.
215. Todd, Petra. "Essays on Empirical Methods for Evaluating the Impact of Policy Interventions in Education and Training." Ph.D. Dissertation, University of Chicago, 1996.
216. Tunali, Insan. "Rationality of Migration." *International Economic Review* 41 (November 2000): 893–920.
217. Trivedi, Pravin and Baker, G. M. "Unemployment in Australia: Duration and Recurrent Spells." *Economic Record* 59 (June 1983): 132–148.
218. Vytlacil, Edward "Independence, Monotonicity, and Latent Index Models: An Equivalence Result." forthcoming *Econometrica* (2001).
219. Wainer, Howard. *Drawing Inferences from Self-Selected Samples.* Mahwah,N.J.: Lawrence Erlbaum Associates, 2000.
220. Willis, Robert and Rosen, Sherwin. "Education and Self Selection." *Journal of Political Economy* 87 (Part 2, October 1979): S7–S36.
221. Wright, Philip Green. *The Tariff on Animal and Vegetable Oils.* New York: MacMillan, 1928.
222. Yitzhaki, Shlomo. "On Using Linear Regression in Welfare Economics." *Journal of Business and Economic Statistics* 14 (October 1996): 478–86.

APPENDIX A-1

Coherency Conditions and Causality

Consider a prototypical simultaneous equations model for two endogenous variables (Y_1, Y_2) written as a function of exogenous variables X:

$$\text{(A-1)} \qquad \begin{aligned} \alpha_{11} Y_1 + \alpha_{12} Y_2 &= X\beta_1 + U_1 \\ \alpha_{21} Y_1 + \alpha_{22} Y_2 &= X\beta_2 + U_2. \\ E(U_1, U_2 | X) &= (0,0). \end{aligned}$$

The Cowles group developed an elaborate theory for identification and estimation in this class of models when Y_1 and Y_2 are continuous variables of the type that appear in conventional market equilibria or interior solution demand equations.

How well does this theory transport to settings where (Y_1, Y_2) are discrete or mixed discrete-continuous? These questions were addressed in papers by Amemiya (1973, 1974), Heckman (1976b, 1978a), Gourieroux, Laffont and Monfort (1980) and Schmidt (1981), in different settings. First note that if Y_1 and Y_2 are discrete, (U_1, U_2) cannot be continuous. In particular, conventional normality assumptions for (U_1, U_2) are inappropriate. Redefining (A-1) to be a model where Y_1 and Y_2 are latent variables with

$$D_1 = 1(Y_1 \geq 0) \text{ and } D_2 = 1(Y_2 \geq 0)$$

preserves the Cowles paradigm.[65] However, the behavioral content of the model is not clear and involves relationships among latent variables that are difficult to motivate with a precise theory. This model is more an analogy with conventional simultaneous equations methods than a model that has a firm economic motivation.

A model with a clearer behavioral foundation writes dummy variables shifting the equations determining the latent indices.[65,66] Thus (Y_1, Y_2) are latent variables and

$$\text{(A-2a)} \qquad \begin{aligned} Y_1 &= \gamma_{11} D_1 + \gamma_{12} D_2 + X\beta_1 + U_1 \\ Y_2 &= \gamma_{21} D_1 + \gamma_{22} D_2 + X\beta_2 + U_2 \end{aligned}$$

where

$$\text{(A-2b)} \qquad D_1 = 1(Y_1 \geq 0) \text{ and } D_2 = 1(Y_2 \geq 0).$$

This model has the strange feature that Y_i causes D_i but D_i also causes Y_i. Since the two are mechanically related from (A-2b), it seems more natural to set $\gamma_{11} = \gamma_{22} = 0$. But even in this case, one encounters logical problems. $D_1 = 0$ can coexist with $Y_1 \geq 0$.This produces models with negative probabilities or pro-

[65] This model was developed in Heckman (1973, 1976b, 1978a) and Mallar (1977).
[66] The tobit version of this model was developed in Heckman (1973, 1976b, 1978a) and applied by my student Olson working with Nelson (1978).

babilities that exceed one. Mechanical application of Cowles methods breaks down. To rule out these pathologies for a general (U_1, U_2) requires

(A-3) $$\gamma_{12}\,\gamma_{21} = 0 \text{ (Coherency Condition).}$$

This condition seems to rule out true simultaneity and forces models into a recursive form. In fact, it eliminates bad economic models. In every application, the coherency condition has a clear economic interpretation (see, *e.g.* Heckman (1976b, 1978a), Ransom (1987) and Blundell and Smith (1994).) Heckman and MaCurdy (1985) present a comprehensive discussion of the coherency condition.

The model (A-2a) and (A-2b) allows one to make the spurious *vs.* true causality distinctions central to Cowles econometrics but with more general types of endogenous variables. Assuming that $\gamma_{11} = \gamma_{12} = \gamma_{22} = 0$, but $\gamma_{21} \neq 0$, a true effect of D_1 on D_2 exists if $\gamma_{21} \neq 0$ but a spurious effect arises if $D_1 \not\perp\!\!\!\perp U_2$. A properly reformulated Cowles model can still make the crucial Cowles distinctions between causal and statistical associations.[67]

This type of analysis also illustrates the benefits of the econometric approach linking statistics to economics. Log linear models for analyzing discrete data developed by statisticians (see *e.g.*, Goodman (1970) and Bishop, Fienberg and Holland, 1974) cannot distinguish the γ_{21} effect from the $U_2 \not\perp\!\!\!\perp D_1$ effect and so cannot be used to make causal distinctions. (Heckman, 1978a). Panel data extensions of these models can be used to distinguish whether past occurrences of an event causally affect the probability of occurrence of future events or just stand in for unmeasured components *i.e.* they can address causal questions in a dynamic setting, a topic addressed in Section 6.

[67] Heckman (1976b), estimates (A-2a) and (A-2b) under the coherency condition. Y_1 is an index of sentiment in favor of passing an anti-discrimination law and Y_2 is a measure of the impact such as wages or employment.

APPENDIX A-2

Bounding and Sensitivity Analysis

Starting from equation (19) or its version for conditional means, the papers by Smith and Welch (1986), Holland (1986) and Glynn, Laird and Rubin (1986) characterize the selection problem more generally without the index structure, and use either Bayesian or classical methods for performing sensitivity analyses for the effects of different identifying assumptions on inferring the population mean.

Selection on observables solves the problem of selection by assuming that $Y_1 \perp\!\!\!\perp D|X$ so $F(Y_1|X, D = 1) = F(Y_1|X)$. This is the assumption that drives matching models. It is inconsistent with the Roy model of self selection (Heckman and Vytlacil, 2001d, 2002).

Various approaches to bounding this distribution, or moments of the distribution, have been proposed in the literature all building on insights by Holland (1986) and Peterson (1976). To illustrate these ideas in the simplest possible setting, let $g(Y_1|X,D = 1)$ be the density of outcomes (*e.g.* wages) for persons who work ($D = 1$ corresponds to work). Assume censored samples. Missing is $g(Y|X, D = 0)$ *e.g.* the density of the wages of non-workers.

In order to estimate $E(Y_1|X)$, Smith and Welch (1986) use the law of iterated expectations to obtain

$$E(Y_1|X) = E(Y_1|X, D = 1)\Pr(D = 1|X) + E(Y_1|X, D = 0)\Pr(D = 0|X).$$

To estimate the left hand side of this expression, it is necessary to obtain information on the missing component $E(Y_1|X, D = 0)$. Smith and Welch propose and implement bounds on $E(Y_1|X,D = 0)$ *e.g.*

$$Y_L \leq E(Y_1|X,D = 0,Z) \leq Y^U$$

where Y^U is an upper bound and Y_L is a lower bound.[68] Using this information, they construct the bounds

$$E(Y_1|X,D = 1)\Pr(D = 1|X) + Y_L \Pr(D = 0|X) \leq E(Y_1|X)$$
$$\leq E(Y_1|X,D = 1)\Pr(D = 1|X) + Y^U \Pr(D = 0|X).$$

By doing a sensitivity analysis, they produce a range of values for $E(Y|X)$ that are explicitly dependent on the range of values assumed for $E(Y|X, D = 0)$. Later work by Manski (1989, 1990, 1994, 1995), Horowitz and Manski (1995) and Robins (1989) develop this type of analysis more systematically for a variety of models.

Glynn, Laird, and Rubin (1986) present a sensitivity analysis for distributions using Bayesian methods under a variety of different assumptions about $F(Y_1|X, D = 0)$ to determine a range of values of $F(Y|X)$. Holland (1986) proposes a more classical sensitivity analysis that vary the ranges of parameters of models. Rosenbaum (1995) discusses a variety of sensitivity analyses.

[68] In their problem there are plausible ranges of wages which dropouts can earn.

The objective of these analyses of bounds and the Bayesian and classical sensitivity analyses is to clearly separate what is known from what is conjectured about the data, and to explore the sensitivity of reported estimates to the assumptions used to secure them. Manski (1990, 1994) and Heckman and Vytlacil (2000a,b, 2001b) demonstrate the extra restrictions that come from using index models to produce bounds on outcomes.

Much of the theoretical analysis presented in the recent literature is nonparametric although in practice, much practical experience in statistics and econometrics demonstrates that high-dimensional nonparametric estimation is not feasible for most sample sizes available in cross sectional econometrics. Some form of structure must be imposed to get any reliable nonparametric estimates. However, feasible parametric versions of these methods run the risk of imposing false parametric structure.[69]

[69] The methods used in the bounding literature depend critically on the choice of conditioning variables X. In principle, all possible choices of the conditioning variables should be explored especially in computing bounds for all possible models, although in practice this is never done. If it were, the range of estimates produced by the bounds would be substantially larger than the wide bounds already reported in this literature.