Econometrics and Economic Theory in the 20th Century

The Ragnar Frisch Centennial Symposium

Edited by
Steinar Strøm
Contents

Introduction  
List of Contributors  

PART ONE: RAGNAR FRISCH AND HIS CONTRIBUTIONS TO ECONOMICS  
1 Ragnar Frisch at the University of Oslo  
   JENS C. ANDVIG AND TORE THONSTAD  
   3  
2 Ragnar Frisch and the Foundation of the Econometric Society and Econometrica  
   OLAV BJERKHOLT  
   26  
3 The Contributions of Ragnar Frisch to Economics and Econometrics  
   JOHN S. CHIPMAN  
   58  

PART TWO: UTILITY MEASUREMENT  
4 Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss  
   JERRY A. HAUSMAN AND WHITNEY K. NEWEY  
   111  
5 Consumer Demand and Intertemporal Allocations: Engel, Slutsky, and Frisch  
   RICHARD BLUNDELL  
   147  

PART THREE: PRODUCTION THEORY  
6 Production Functions: The Search for Identification  
   ZVI GRILICHES AND JACQUES MAIRESSE  
   169  
7 Investment and Growth  
   DALE W. JORGENSEN  
   204  

PART FOUR: MICROECONOMIC POLICY  
8 Evaluating the Welfare State  
   JAMES J. HECKMAN AND JEFFREY SMITH  
   241
Evaluating the Welfare State

James J. Heckman and Jeffrey Smith

Ragnar Frisch was a leading advocate of national economic planning in the service of the welfare state. His Nobel lecture (1970) stressed the value of interactions between economists and politicians in arriving at politically acceptable and economically viable national plans. A major theme of his lecture was that economists should act in the public interest and in so doing should recognize the diversity of policy objectives advocated by different groups in democratic societies. He made the important distinction between maximizing the mythical welfare function assumed in classical welfare economics and in the classic policy analysis of Tinbergen (1956) and reconciling and satisfying the diverse perceptions and values held by citizens of modern states. He stressed the role of economists in informing policy-makers and the general public about the relevant economic trade-offs and the costs of alternative policies.

Frisch's faith in the power of economics to supply the information required to make informed public choices seems wildly optimistic today.

Portions of this study were first presented in the Barcelona lecture to the World Econometric Society (Heckman, 1990b), and some of this material draws from Heckman (1992, 1993), Heckman, Smith, and Clements (1997), and Heckman and Smith (1993, 1995). The ideas not presented in those papers were first presented in informal discussions at a CEMFI conference in Madrid, Spain, in September 1993 and at a workshop at the Institute for Research on Poverty at the University of Wisconsin, Madison, in February 1994 and have been shared with colleagues at the University of Chicago. The research in this chapter was supported by grant NSF-SBR-93-21-048, by the Russell Sage Foundation, and by a grant from the American Bar Foundation. We thank the editor of the Econometric Monograph Series, Alberto Hall, and an anonymous referee for helpful comments, as well as Gary Becker, Olav Bjerkholt, Dragan Filipovich, Lars Hansen, Hidehiko Ichimura, Lance Lochner, Tom Macurdy, Derek Neal, Robert Pollak, Jose Scheinkman, Chris Taber, and Ed Vylacil. We also thank the seminar participants at the Oslo meeting in March 1995, as well as colleagues at Columbia University, the University of Chicago, University College London, the Canadian Econometric Studies Group in Montreal (September 1995), and Washington University, St. Louis. Ed Vylacil is singled out for a particularly close reading.
yet his message remains relevant. Economists are still asked to inform
the general public and policy-makers about the likely consequences of
alternative social programs. Social-welfare functions do not govern deci-
sion making in any democratic society, and it is clear that a variety of
criteria are relevant for evaluating alternative policies in democratic
societies composed of individuals and groups with diverse values and
perspectives.

Coercive redistribution and intervention are defining activities of the
welfare state. Principled redistribution and intervention are based on
interpersonal comparisons made by governments and groups of indi-
viduals in society. Different public policies typically have different
consequences for different citizens. Enumerating and evaluating these
consequences are important parts of social decision making, and differ-
ent criteria have been suggested.

This essay presents these criteria and considers the data required to
operationalize them. Some are so difficult to implement empirically
and so cannot serve as practical guides to social decision making. Other
criteria can be implemented, especially if economists have access to
microeconomic data, but such data must be supplemented by knowledge
of — or assumptions about — the choice processes of the agents being
studied or the dependence across potential outcomes associated with
different policies and by assumptions that relate partial-equilibrium evalu-
ations to general-equilibrium evaluations. We shall examine alternative
sets of identifying assumptions that bound, or exactly produce, the alter-
native criteria used in evaluating the welfare state.

Frisch was well aware that economists often need to supplement the
available data with assumptions in order to evaluate policies. His Nobel
lecture emphasized the "lure of unsolvable problems," and he advocated
evaluation of policies by making bold assumptions if necessary. His
famous article on circulation planning emphasized this point (Frisch,
1934). That essay examined how bold the assumptions have to be to
answer major economic evaluation questions given the type of data
available in many countries.

This essay considers the general policy-evaluation problem, but
focuses most of its attention on a specific version of it that is widely
studied in econometrics under the rubric of "the analysis of treatment
effects" (e.g., Heckman and Robb, 1985). In this version, a policy has a
voluntary component, sometimes called a program, and persons choose
to participate in it. A job-training program and a tuition subsidy for
college attendance are examples. There may be an involuntary com-
ponent to the policy as well, such as paying the taxes to finance the vol-
untary component or facing the wages produced by an increase in the
supply of trained workers resulting from the program. Having access to
data on the outcomes of participants and nonparticipants simplifies some
aspects of the policy-evaluation problem, compared with the case where

only participants or nonparticipants are observed, but also raises the
problem of self-selection bias.

From information on the program participation decisions of eligible
persons, it is sometimes possible to infer their preferences for the
outcomes produced by the program. This information is of value in
its own right. A strictly libertarian evaluation of a program stops with
determining individual subjective valuations for the program being
evaluated. Evaluation of the welfare state requires more information.
"Objective" evaluations of outcomes supplement revealed-preference
information to form the basis for policy discussions and interpersonal
comparisons. Even if such "objective" information is available about
outcomes under each policy, knowledge of individual preferences for
specific programs helps in constructing, or bounding, the distribution of
outcomes across alternative policy regimes that is required to implement
some of the criteria examined in this essay. In addition, if general-
equilibrium effects can be safely ignored, knowledge of the outcomes
for self-selected nonparticipants sometimes identifies the distribution of
outcomes for society at large in the absence of the program produced by
a policy.

In the course of examining these issues, we consider which policy
questions conventional econometric "treatment-effect" estimators
answer by embedding them in a simple general-equilibrium framework.
Most of the standard econometric estimators identify parameters of only
limited economic interest, but in certain special cases they provide partial
answers to economically interesting questions.

We use data from a social experiment designed to evaluate the gross
impact of training on the earnings and employment of participants in
order to examine the conflict or consistency among the various criteria
that have been proposed to evaluate policies. If all the criteria are in
agreement, their multiplicity is not a matter of practical importance. We
examine participant evaluations as revealed by their attrition from the
program instituted by a policy, by their responses to questionnaires,
and by econometric analyses of outcome and participation equations
under different identifying assumptions. We find that participant self-
evaluations disagree with revealed preferences as manifested by choices,
and with impacts objectively estimated using experimental data. The
criteria do not all agree, and there is scope for seriously conflicting
assessments of a program. We also present evidence that favorable
cost–benefit assessments of government training programs are consid-
erably weakened once the full cost of raising government revenue is taken
into account.

This chapter is organized in the following way: Section 1 presents
alternative criteria for evaluating the welfare state and the data required
to operationalize them. Both the general case and the specific case of
evaluating a program into which agents self-select are examined. Section
2 considers how alternative assumptions about decision processes and access to different sources of microeconomic data aid in the construction of the evaluation criteria. Section 3 examines the economic questions addressed by two widely used econometric evaluation estimators and relates them to a portion of what is required in a comprehensive cost–benefit analysis. Evidence is presented on how the inference from the most commonly used econometric evaluation estimator is modified when direct costs of the program are fully assessed, including the welfare costs of taxes raised to support the program. Section 4 presents empirical evidence on the consistency of alternative criteria derived from evaluations based on “objective” outcomes, evaluations inferred from self-selection and attrition decisions, and self-reported evaluations elicited with questionnaires regarding a prototypical job-training program. The chapter concludes with a summary in Section 5.

1 Alternative Criteria for Evaluating the Welfare State

1.1 The Origin of the Demand for Evaluations

Coercive redistribution and intervention are essential activities of the welfare state. Adopting any particular policy with a redistributational component involves weighing the subjective assessments made by different groups regarding the outcomes created by the policy using the political process. Coercion arises because the perceived benefit from a policy does not always equal or exceed its cost for all members of a society. If there were no coercion, redistribution and intervention would be voluntary activities, and apart from the free-rider problem there would be no scope for government activity in orchestrating redistribution and conducting interventions. There would be no need to publicly justify voluntary trades among individuals.

If government is producing a service for which there are good market substitutes, there is no need to resort to an elaborate evaluation procedure for the service. Market prices provide the right measure of marginal gain and cost, unless the usual problems of increasing returns, externalities, or public perception that private preferences are defective lead to mistrust of the signals produced from the market mechanism. The argument that justifies the welfare state denies the use of prices and private evaluations as the sole criteria for evaluation of governmental activities, but recognizes that they may be relevant inputs into the general policy-evaluation process.

The demand for publicly documented evaluations arises from a demand for information by rival parties in a democratic state. Even libertarians who do not accept coercion and who oppose government intervention evaluate policies in order to participate in the political dialogue of the welfare state.

Evaluating the Welfare State

The claims that markets fail or that consumer judgments are faulty often are made without a factual basis. If these claims are false, the case for a welfare state is weakened. In this study, we accept the reality of the welfare state, without necessarily endorsing the arguments for it. We do not consider the quality of the evidence supporting the premises of the welfare state. Instead, we consider the evaluation of specific policy proposals within its framework.

1.2 Alternative Criteria for Evaluating Programs

Let the outcome in the presence of policy $j$ for person $i$ be $Y_{ij}$, and let the personal preferences of person $i$ for outcome $Y$ be denoted $U_i(Y)$. A policy effects a redistribution from taxpayers to beneficiaries, and $Y_{ij}$ represents the flow of resources to $i$ under policy $j$. Some persons can be both beneficiaries and taxpayers. All policies we consider are assumed to be feasible. In the simplest case, $Y_{ij}$ is net income after tax and transfers, but it can also be a vector of incomes and benefits, including provision of in-kind services. Many criteria have been proposed to evaluate policies. Let “0” denote the no-policy state, and initially abstract from uncertainty. The standard model of welfare economics postulates a social-welfare function $W$ that is defined over the utilities of the $N$ members of society:

$$W(j) = W[U_i(Y_{ij}), \ldots, U_N(Y_{jN})]$$

Policy choice based on a social-welfare function picks that policy $j$ with the highest value for $W(j)$. A leading special case is the Benthamite social-welfare function:

$$B(j) = \sum_{i=1}^{N} U_i(Y_{ij})$$

Criteria (1) and (2) implicitly assume that social preferences are defined in terms of the private preferences of citizens, as expressed in terms of their own consumption; this principle is called welfarism (Sen, 1979). They could be extended to allow for interdependence across persons so that the utility of person $i$ under policy $j$ would be $U_i(Y_{ij}, \ldots, Y_{jN})$ for all $i$.

Conventional cost–benefit analysis assumes that $Y_{ij}$ is scalar income and orders policies by their contributions to aggregate income:

$$CB(j) = \sum_{i=1}^{N} Y_{ij}$$

Analysts who adopt criterion (3) implicitly assume that outputs are costlessly redistributed among persons via a social-welfare function, or else
they accept gross national product (GNP) as their measure of value for a policy.

While these criteria are traditional, they are not universally accepted, and they do not answer all of the interesting questions of political economy or "social justice" that arise in the political arena of the welfare state. In a democratic society, politicians and advocacy groups are interested in knowing the proportion of people who will benefit from policy $j$ as compared with policy $k$:

$$\text{PB}(j|j, k) = \frac{1}{N} \sum_{i=1}^{N} 1[U_i(Y_{j}) \geq U_i(Y_k)]$$

(4)

where "1" is the indicator function: $1(A) = 1$ if $A$ is true; $1(A) = 0$ otherwise. In the median-voter model, a necessary condition for $j$ to be preferred to $k$ is that $\text{PB}(j|j, k) \geq \frac{1}{2}$. Many writers on "social justice" are concerned about the plight of the poor, as measured in some base state $k$. For them, the gain from policy $j$ is measured in terms of the income or utility gains of the poor. In this case, interest centers on the gains to specific types of persons, such as the gains to persons with outcomes in the base state $k$ less than $y$: $\Delta_{ji} = Y_{ji} - Y_{ki}|Y_{ki} \leq y$, or their distribution

$$F(\Delta_{ji}|Y_{ki} = y_{ki}, Y_{ki} \leq y)$$

(5)

or the utility equivalents of these variables. Within a targeted subpopulation there is sometimes interest in knowing the proportion of people who gain relative to specified values of the base state $k$:

$$\Pr(\Delta_{ji} > 0|Y_{ki} \leq y)$$

(6)

In addition, measures (2) and (3) are often defined only for a target population, not for the full taxpayer population.

The existence of merit goods like education or health implies that specific components of the vector $Y_{j}$ are of interest to certain groups. Many policies are paternalistic in nature and implicitly assume that people make the wrong choices. "Social" values are placed on specific outcomes, often stated in terms of thresholds. Thus one group may care about another group in terms of whether or not they satisfy an absolute threshold requirement,

$$Y_{ji} \geq y \text{ for } i \in S$$

where $S$ is a target set toward which the policy is directed, or in terms of a relative requirement compared with a base state $k$,

$$Y_{ji} \geq Y_{ki} \text{ for } i \in S$$

Uncertainty introduces additional considerations. Participants in society typically do not know the consequences of each policy for each person. A fundamental limitation in applying these criteria is that, ex ante, these consequences are not known and, ex post, one may not observe all potential outcomes for all persons. If some states are not experienced, the best that agents can do is to guess about them. Even if, ex post, agents know their outcome in a benchmark state, they may not know it ex ante, and they may always be uncertain about what they would have experienced in an alternative state.

In the literature on welfare economics and social choice, one form of decision making under uncertainty has been extensively investigated. The "veil of ignorance" of Vickrey (1945) and Harsanyi (1955, 1973) postulates that decision-makers are completely uncertain about their positions in the distribution of outcomes under each policy, or should act as if they are completely uncertain, and they should use expected-utility criteria (Vickrey-Harsanyi) or a maximin strategy (Rawls, 1971) to evaluate their welfare under alternative policies. This form of ignorance is sometimes justified as an "ethically correct" position that captures how an "objectively detached" observer should evaluate alternative policies even if actual participants in the political process use other criteria. An approach based on the veil of ignorance is widely used in practical work in evaluating different income distributions (Sen, 1973) and requires information only about the marginal distributions of outcomes produced under different policies.

A less high-minded, but empirically more accurate, description of social decision making recognizes that persons act in their own self-interest and have some knowledge about how they will fare under different policies, but allows for the possibility that persons can only imperfectly anticipate their outcomes under different policy regimes. The outcomes in different regimes may be dependent, so that persons who benefit under one policy may also benefit under another. However, agents may not possess perfect foresight. Letting $I_i$ denote the information set available to agent $i$, agent $i$ will evaluate policy $j$ against $k$ using that information. Let $F(y_i, y_i|I_i)$ be the distribution of outcomes $(Y_{ji}, Y_{ki})$ as perceived by agent $i$. Under an expected-utility criterion, person $i$ prefers policy $j$ over $k$ if

$$E[U_j(Y_i|I_i)] > E[U_k(Y_i|I_i)]$$

Letting $\theta$, parameterize heterogeneity in preferences, so that $U(Y_i) = U(Y_i; \theta)$, and using integrals to simplify the expressions, the proportion of people who prefer $j$ is

$$\text{PB}(j|j, k) = \int [E[U_j(Y_i|\theta)|I_i] > E[U_k(Y_i|\theta)|I_i]]dF(\theta, I)$$

(7)
where $F(\theta, I)$ is the joint distribution of $\theta$ and $I$ in the population whose preferences over outcomes are being studied. In the special case where $I_i = (Y_{i_1}, Y_{i_2})$, so that there is no uncertainty about $Y_1$ and $Y_2$.

$$PB(j|k) = \int \{U(y_j; \theta) > U(y_k; \theta)\} dF(\theta, y_j, y_k)$$

Expression (8) is an integral version of (4) when outcomes are perfectly predictable and when preference heterogeneity can be indexed by vector $\theta$.

Adding uncertainty to the analysis makes it informative to distinguish between ex-ante and ex-post evaluations. Ex post, part of the uncertainty about policy outcomes is resolved, although individuals do not, in general, have full information about what their potential outcomes would have been in policy regimes they have not experienced, and they may have only incomplete information about the policy they have experienced (e.g., the policy may have long-run consequences extending past the point of evaluation). It is useful to index the information set $I_i$ by $t$, $I_n$, to recognize that information about the outcomes of policies may accrue over time. Ex-ante and ex-post assessments of a voluntary program need not agree. Ex-post assessments of a program through surveys administered to persons who have completed it (Katz et al., 1975) may disagree with ex-ante assessments of the program. Both may reflect honest valuations of the program, but they are reported at times when agents have different sets of information about the program. Before participating in a program, persons may be uncertain of the consequences of participation. Persons who have completed program $j$ may know $Y_j$, but can only guess at the alternative outcome $Y_k$, which they have not experienced. In this case, ex-post "satisfaction" for agent $i$ is synonymous with the following inequality:

$$U_i(Y_j) > E[U_i(Y_k)|I_a]$$

where $t$ is the post-program period in which the evaluation is made. In addition, survey questionnaires about "client" satisfaction with a program may capture subjective elements of program experience not captured by "objective" measures of outcomes, which usually exclude psychic costs and benefits.

In order to operationalize these notions empirically, it is useful to distinguish the effects of a policy as it impacts the tax-collection system from its effects operating through direct program participation. To this end, it is useful to isolate policy outcomes from alternative revenue-

---

1 We do not claim that persons will necessarily vote "honestly," although in a binary-choice setting they do, and there is no scope for strategic manipulation of votes (Moulin, 1983). "PB" is simply a measure of relative satisfaction and need not describe a voting outcome where other factors come into play.

2 A quotation from Knight 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).

3 The contrast between micro and macro analyses is overdrawn. The studies by Baumol and Quandt (1966), Lancaster (1971), and Domenich and McFadden (1975) were micro examples of attempts to solve what we have called a macro problem. Those authors considered the problem of forecasting the demand for a new good that had never previously been purchased.
output markets can be ignored and (b) that the problem of selection bias that arises from using self-selected samples of participants and nonparticipants to estimate population distributions can be ignored or surmounted.

More precisely, let \( j \) be the policy regime we seek to evaluate. Eligible person \( i \) in regime \( j \) has two potential outcomes: \((Y^o_i, Y^p_i)\), where the superscripts denote nondirect participation ("0") and direct participation ("1"). Noneligible persons have only one option: \( Y^o_i \). These outcomes are defined at the equilibrium level of participation under program \( j \). All feedback effects are incorporated in the definitions of the potential outcomes.

Let subscript "0" denote a policy regime without the program. Let \( D_j = 1 \) if person \( i \) participates in program \( j \). A crucial identifying assumption that is implicitly invoked in the microeconomic evaluation literature is that

\[
Y^0_{i} = Y^0_{0i}, \quad (A-1)
\]

and hence that \( F(y^0_i | D_j = 0, X) = F(y^0_i | D_j = 0, X) \) for \( y^0_i = y_0 \) given conditioning variables \( X \). The outcome for nonparticipants in policy regime \( j \) is the same in the no-policy state "0" and in the state in which policy \( j \) is operative. This assumption is consistent with a program that has "negligible" general-equilibrium effects and in which the same structure of tax-revenue collection is used in regimes \( j \) and "0."

An additional assumption sometimes invoked is that

\[
Y^1_i = Y^1_{i}, \quad (A-2)
\]

where \( Y_i \) is the outcome if the program is universally applied. This entails a different kind of general-equilibrium assumption – this time concerning expansion of program \( j \) to universal coverage. Making assumptions (A-1) and (A-2) together strains the imagination, for if a program is sufficiently small that (A-1) is plausible, its universal expansion may make it so large that (A-2) will not be plausible. Nonetheless, taken together, these assumptions, strengthened with additional assumptions about agent self-selection rules, enable analysts to generalize from self-selected samples within a given policy regime to choices across policy regimes. Assumption (A-2) is rarely used and plays only a minor role in this study. Assumption (A-1) plays a much more substantial role in this study and in the microeconomic evaluation literature.

From data on individual program participation decisions it is possible to infer the implicit valuations of the program made by persons eligible for it. These evaluations constitute all of the data needed for a libertarian program evaluation, but something more than these will be required to evaluate programs in the interventionist welfare state. For certain decision rules, it is possible to use the data from self-selected samples to bound or estimate the joint distributions required to implement criterion (4) or (7), as we demonstrate later.

The existence of a voluntary-participation component for a program under policy \( j \) creates an option value that for eligible person \( i \) is

\[
\max \{Y^0_i, Y^1_i \} - Y^0_i
\]

By (A-1) this is the same as \( \max \{Y_{0i}, Y_{1i} \} - Y_{0i} \), which if strengthened by (A-2) is \( \max \{Y_{0i}, Y_{1i} \} - Y_{0i} \). The distribution of the value of this option for those who take it is

\[
F(Y^1_i - Y^0_i | Y^1_i > Y^0_i)
\]

For persons interested in the equity of program provisions, it is of interest to examine the dependence between the options offered and the nonparticipation outcomes, which are assumed to approximate the no-policy outcomes.

People who fear "cream skimming" by program administrators whose performance is evaluated on the basis of the outcomes for the participants they select claim that \( Y^1_i \) and \( Y^0_i \) are strongly positively dependent and that the gross value added, \( \Delta_i = Y^1_i - Y^0_i \), is unrelated or negatively related to \( Y^0_i \). To address these concerns, it is necessary to know the joint dependence between \( Y^0_i \) and \( Y^1_i \) and to compute the dependence between \( \Delta_i \) and \( Y^0_i \).

2 The Data Needed to Evaluate the Welfare State

To implement criteria (1) and (2), it is necessary to know the distribution of outcomes across the entire population and to know the utility functions of individuals. In the case where \( Y \) refers to scalar income, criterion (3) requires only GNP (the sum of the program-j outcome distribution). If interest centers solely on the distributions of outcomes for direct program participants, the measures can be defined solely for populations with \( D_j = 1 \). Criteria (4), (5), (6), and (8) require knowledge of outcomes and preferences across programs. Criterion (7) requires knowledge of the joint distribution of information and preferences across persons. Tables 8.1A and 8.1B summarize the criteria and the data needed to implement them.

This study has little to say about estimating preference functions or preference heterogeneity. We refer readers to Heckman (1974a) and the comprehensive survey by Browning, Hansen, and Heckman in press, who have documented the empirical importance of preference heterogeneity. Our focus is on estimating the distributions of outcomes across policy states as a first step toward empirically implementing the full criteria. This more modest objective can be fit into the framework of Section 1 by assuming that utilities are linear in their arguments and identical across persons.
### Table 8.1A. Population Data Requirements to Implement Criterion: General Population (Compulsory Programs); Program j Compared with Program k

<table>
<thead>
<tr>
<th>Cost–benefit criterion</th>
<th>Benthamite criterion</th>
<th>General social-welfare function with interdependent preferences</th>
<th>Selfish voting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
<td>$E(Y_j) - E(Y_k) \geq 0$</td>
<td>$E[U(Y_i, \theta)] - E[U(Y_k, \theta)] \geq 0$</td>
<td>$W(j) &gt; W(k)$, $W(\ell) = W[U(Y_i, \theta), \ldots, Y_{\alpha}, i, \ldots, Y_{\alpha}, \theta]$, $\ell = j, k$</td>
</tr>
<tr>
<td>Require</td>
<td>Population means $E(Y_j), E(Y_k)$</td>
<td>$U(Y_j, \theta)$ and distribution of $(Y_j, \theta), F(Y_j, \theta); \ell = j, k$</td>
<td>Need each $U(Y_i, \theta)$ for all $i$; need outcomes for each person$^b$</td>
</tr>
<tr>
<td>Estimable on aggregate time-series data?</td>
<td>Yes, if data exist on aggregate economy in both regimes and can eliminate trend</td>
<td>No, unless $\theta$ is the same for everyone (homogeneity); $U(Y_j, \theta)$ known and the moment $\int U(y, \theta) dF(y, \theta)$ known or estimable; $\ell = j, k$</td>
<td>Need $U(Y_j, \theta), F(Y_j, \theta)^c$</td>
</tr>
</tbody>
</table>

$^a$ This includes the special case where individual utility depends only on individual consumption.

$^b$ In special cases, summary statistics of the distribution of $Y$ may suffice.

$^c$ For altruistic voting, $U$ depends on outcomes for other persons, $Y_{\alpha}, Y_{\beta}, \ldots, Y_{\alpha}$ or various sub-aggregators.

### Table 8.1B. Population Data Requirements to Implement Criterion (Voluntary Programs, Conditional on $D_i = 1$); Program j Compared with Program k

<table>
<thead>
<tr>
<th>Cost–benefit criterion</th>
<th>Benthamite criterion</th>
<th>General social-welfare function with interdependent preferences$^a$</th>
<th>Selfish voting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
<td>$E(Y_j</td>
<td>D_i = 1) - E(Y_k</td>
<td>D_i = 1)$; what j participants gain over state k</td>
</tr>
<tr>
<td>Require</td>
<td>Population conditional means $E(Y_j</td>
<td>D_i = 1), E(Y_k</td>
<td>D_i = 1)$</td>
</tr>
<tr>
<td>Estimable on aggregate time-series data?</td>
<td>Yes, if aggregate data for participants exist in both regimes, can eliminate trend</td>
<td>No, unless $\theta$ is the same for everyone (homogeneity); $U(Y_j, \theta)$ known, and the moment $\int U(y, \theta) dF(y, \theta</td>
<td>D_i = 1)$ is known or estimable; $\ell = j, k$</td>
</tr>
</tbody>
</table>

$^a$ This criterion is not well defined when restricted to subsets of the population. If only the utility of voluntary participants is considered, some position about the utility of nonparticipants must be taken, and the feedback between participants and nonparticipants must be explicitly modeled. When individual utility depends only on individual consumption, the criterion is well defined.
This section considers the problem of constructing the distribution of $(Y_j^o, Y_j^n)$, the distribution of potential outcomes within policy regime $j$ in which direct participation is voluntary. Extension of the estimates of this distribution to other policy regimes follows by invoking the assumptions discussed in Section 1. We discuss how the widely invoked implicit assumption that responses to program treatment are homogeneous across persons greatly simplifies the construction of the joint distribution of potential outcomes and how explicit assumptions about the structure of voluntary program participation rules aid in identifying or reducing the uncertainty about the distributions of outcomes. We consider the information available from cross-section data, from social experiments, from panel data, and from repeated cross-section data.

2.1 The Microeconomic Evaluation Problem

To simplify the notation, we drop the policy-regime subscript $j$. All of the distributions we consider in this section are measured within that regime. The extrapolation of within-regime measures to across-regime measures is made using assumptions (A-1) and (A-2) discussed in Section 1. In a regime with voluntary participation, we have access to

$$F(y_j^o | D = 1, X) \quad \text{and} \quad F(y_j^n | D = 0, X)$$

the distributions of outcomes for participants and nonparticipants at time $t$, respectively. These embody both the direct and indirect effects of the program.

The fundamental evaluation problem arises from the fact that we do not observe $(Y_j^o, Y_j^n)$ for anyone – just one coordinate or the other of this pair. Given knowledge of individual preferences, and their joint distribution with the outcomes, all of the policy criteria discussed in Section 1 and summarized in Tables 8.1A and 8.1B can be implemented. Here we focus on recovering $F(y_j^o, y_j^n | D, X)$, from which all of the distributions discussed in Section 1 can be recovered. For evaluating the criteria only for program participants, it is enough to know $F(y_j^o, y_j^n | D = 1, X)$ – the potential outcomes for participants – or various marginal distributions formed from this distribution.

As previously noted, the different evaluation criteria require different data for their empirical implementation. Cost–benefit analysis can, in principle, be performed using a before–after analysis of aggregate time-series data. However, if a program has a small impact on the economy, and other policies are instituted coincident with the program being evaluated, or if the time series is nonstationary, aggregate data do not offer a reliable source of information.

The missing counterfactual for cost–benefit criterion (3) is the mean $E(Y_j^o)$, or $E(Y_j^o | D = 1, X)$ if the evaluation is conducted solely for participants. $E(Y_j^o | D = 1, X)$ is produced from data on program participants. Benthamite criterion (2) is more demanding and requires $F(y_j^o | X)$, or $F(y_j^o | D = 1, X)$ if the criterion is defined only for participants. The voting criterion (8) requires $F(y_j^o, Y_j^n | X)$, or $F(y_j^o, Y_j^n | D = 1, X)$ if the criterion is defined only for participants.

In this section we consider how to use cross-section data, data from ideal social experiments, panel data, and repeated cross-section data to construct the different evaluation criteria. Panel data can be used as repeated cross sections, and repeated cross sections can be used as cross sections. Thus it is natural to start with the cross-section case and then determine how access to other sources of data aids in securing identification of the evaluation criteria presented in Section 1.

2.1.1 Cross-Section Data

From cross-section data on $F(y_j^o | D = 1, X)$, $F(y_j^o | D = 0, X)$, and $Pr(D = 1 | X)$ we cannot directly construct the joint distribution $F(y_j^o, y_j^n, D, X)$. Using $F(y_j^o, D = 0, X)$ to proxy $F(y_j^o | D = 1, X)$ runs the risk of selection bias. Various different identifying assumptions have been used to recover the counterfactual distribution $F(y_j^o, D = 1, X)$ or the joint distribution $F(y_j^o, y_j^n, D, X)$. To simplify the notation in this subsection, we drop the $t$ subscript and assume that $(Y_j^o, Y_j^n)$ are measured after the program intervention.

Conditional Independence. One assumption that underlies the method of matching postulates conditioning variables $X$ such that

$$F(y_j^o | D = 1, X) = F(y_j^o | D = 0, X) = F(y_j^o | X) \quad (I-1a)$$

If this assumption is valid, we can safely use nonparticipants to measure what participants would have earned had they not participated, provided we condition on $X$. Using "$\perp$" to denote independence, this identifying assumption is equivalent to $Y_j^o \perp D | X$. To ensure that (I-1a) has an empirical counterpart, we also assume that

$$0 < Pr(D = 1 | X) < 1 \quad (I-1b)$$

over the support of $X$. This condition ensures that both sides of (I-1a) are well defined, (i.e., that for each $X$, there are both participants and nonparticipants). For computing counterfactual means, a simpler requirement is

$$E(Y_j^o | D = 1, X) = E(Y_j^o | D = 0, X) \quad (I-2)$$

Failure to satisfy this condition is an important source of failure in the use of matching to evaluate job-training programs (Heckman et al., 1996; in press).
This method underlies the intuitive principle of "controlling on observables" to eliminate selection bias (Heckman and Robb, 1985).

The identification assumption (I-1a) implies that $\Pr(D = 1|X, Y^0) = \Pr(D = 1|X)$ (i.e., that $Y^0$ does not determine participation in the program), although it does not exclude the possibility that participation in the program is based on $Y^1$. If we strengthen (I-1a) to read

$$(Y^0, Y^1) \perp D|X$$

we can recover $F(y^1|x)$ for the support of $X$ satisfying (I-1b). Thus for the entire population or for the sample conditional on $D = 1$, we can construct the cost–benefit criterion and the Benthamite criterion, but not the voting criterion, because there is no information on the joint distribution of $(Y^0, Y^1)$.

To recover the joint distribution, we need some way to associate values of $Y^0$ with $Y^1$. The dummy-endogenous-variables model (Heckman, 1978) assumes that

$$Y^1 = \alpha + Y^0$$

where $\alpha$ is a constant or a function of $X$. Defining $\alpha$ as the treatment effect, this assumption imposes homogeneity of responses to treatment. Everyone with the same $X$ value benefits or loses by the same amount. A generalization of this method developed by Heckman and Smith (1993) and Heckman, Smith, and Clements (1997c) assumes that the quantities of $Y^1$ and $Y^0$ are the same for each person with the same $X$. Equating quantiles across the two marginal distributions, we form pairs:

$$\left\{ (y^0(q), y^1(q)) | \inf_{y^1} F(y^1|x) > q \text{ and } \inf_{y^0} F(y^0|x) > q, 0 \leq q \leq 1 \right\}$$

(13)

Conditional on $X$, the quantile ranks are preserved, but the effect of treatment is not necessarily the same at all quantiles. More generally, we could assume that the quantiles are mapped in a general way $q_1 = \varphi(q_0)$, where $q_1$ is a quantile of $Y^1$ and $q_0$ is a quantile of $Y^0$. The gain to moving from "0" to "1" is

$$\Delta(q_0) = Y^1[\varphi(q_0)] - Y^0(q_0)$$

(14)

where $Y_i[\varphi(q_0)]$ is the $q_i$-th quantile of $Y_i$ expressed as a function of $q_0$, and $Y_0(q_0)$ is the $q_0$-th quantile of $Y_0$.

If $\varphi$ is a random function, then the mass at $q_0$ is distributed to different values of $q_1$, and $\varphi(q_0)$ has an interpretation as a probability density. If $\varphi$ is a uniform density mapping of $q_0$ to $q_1$ over the interval $[0, 100]$ for all $q_0$, $Y^1$ and $Y^0$ are stochastically independent. Provided the mapping

\[ \varphi \text{ is known, the assumption of conditional independence is sufficient to identify the joint distribution } F(y^1, y^0, D|X). \]

An alternative assumption about the dependence across outcomes is that $Y^1 = Y^0 + \Delta$, where $\Delta$ is stochastically independent of $Y^0$ given $X$. That is,

$$Y^0 \perp \Delta|X$$

(1-5)

This assumption states that the gain from participating in the program is independent of the base $Y^0$. If (I-3) and (I-5) are invoked jointly, we can identify $F(y^1, y^0|x)$ from the cross-section outcome distributions for participants and nonparticipants and estimate the joint distribution by deconvolution methods.

To see how to use this information, note that

$$Y = Y^0 + D\Delta$$

From $F(y|D = 0, X)$, we identify $F(y^0|x)$ as a consequence of (I-3). From $F(y|D = 1, X)$ we identify $F(y^1|x) = F(y^0 + \Delta|x)$. If $Y^0$ and $Y^1$ have densities, then, as a consequence of (I-5), the densities satisfy

$$f_i(y^i|x) = f_\Delta(\Delta|x) * f_0(y^0|x)$$

where "*" denotes convolution. The characteristic functions of the three random variables satisfy

$$E(e^{it_y} | X) = E(e^{it_\Delta} | X)E(e^{it_0} | X)$$

Because we can identify $F(y^1|x)$, we know its characteristic function. By a similar argument we can recover $E(e^{it_x} | X)$. Then

$$E(e^{it_\Delta} | X) = \frac{E(e^{it_y} | X)}{E(e^{it_0} | X)}$$

(15)

and by the inversion theorem (e.g., Kendall and Stuart, 1977) we can recover the density $f_\Delta(\Delta|x)$. We know the joint density

$$f(\Delta, y_0|x) = f_\Delta(\Delta|x)f(y^0|x)$$

From the definition of $\Delta$ we obtain

$$f(y^1 - y^0|x)f(y^0|x) = f(y^1, y^0|x)$$

Thus we can recover the full joint distribution of outcomes and the distribution of gains.

---

5 When $\varphi$ is random, and the random variables are discrete, the matrix mapping probability of $Y^1$ into $Y^0$ must be a Markov matrix to preserve probability. For continuous distributions we need a Markov operator.

6 Barros (1987) used this assumption in the context of an analysis of selection bias.
Under assumption (I-3), assumption (I-5) is testable. The ratio of two characteristic functions in (15) is not necessarily a characteristic function. If it is not, the estimated density \( f_0 \) recovered from the ratio of the characteristic functions need not be positive, and the estimated variance of \( \Delta \) can be negative.\(^7\)

In a regression setting in which means and variances are assumed to capture all of the relevant information, this approach is equivalent to the traditional normal-random-coefficient model. Letting

\[
Y^1 = \mu_0(X) + U_1, \quad E(U_1|X) = 0 \\
Y^0 = \mu_0(X) + U_0, \quad E(U_0|X) = 0
\]

this version of the model can be written as

\[
Y = \mu_0(X) + D[\mu_0(X) - \mu_0(X)] + D(U_1 - U_0) + U_0 \\
= \mu_0(X) + D\bar{\alpha}(X) + U_0
\]

where \( \bar{\alpha}(X) = \mu_0(X) - \mu_0(X) \), and \( \varepsilon = U_1 - U_0 \). By virtue of (I-3), \( U_0, U_1 \) \( \perp \! \! \! \perp \) \( D \) \( X \).

We can use nonparametric regression methods to recover \( \mu_0(X) \) and \( \mu_1(X) - \mu_0(X) \), or we can use ordinary parametric regression methods, assuming that \( \mu_1(X) = X\beta_1 \) and \( \mu_0(X) = X\beta_0 \). Equation (16) is a components-of-variance model, and a test of (I-5) is that

\[
\text{var}(Y|D = 1, X) = \text{var}(Y^0 + \Delta|D = 1, X) \\
= \text{var}(Y^0|X) + \text{var}(\Delta|X) \\
\leq \text{var}(Y|D = 0, X) = \text{var}(Y^0|X)
\]

Under standard conditions, each component of variance is identified and is estimable from the residuals obtained from the nonparametric regression of \( Y \) on \( D \) and \( X \).

An alternative approach relies only on the information contained in the marginal distributions obtained using the conditional independence assumption to bound the joint distribution conditional on \( D \). The Fréchet (1951) bounds inform us that

\[
\max\{F(y^0|X), F(y^1|X) - 1, 0\} \leq F(y^0, y^1|X) \\
\leq \min\{F(y^0|X), F(y^1|X)\} \quad (17)
\]

These bounds are purely statistical and assume no information about agent behavior. Combining the bounds with (I-1b) and (I-3) allows us to

\footnote{For the ratio of characteristic functions, \( r(\ell) \), to be a characteristic function, it must satisfy the requirements that \( r(0) = 1 \), that \( r(\ell) \) is continuous in \( \ell \), and \( r(\ell) \) is nonnegative definite. This identifying assumption can be tested using the procedures developed by Heckman, Robb, and Walker (1990).}

\footnote{An exception is that the bounds for low-probability events are informative.}

\footnote{Heckman (1974a,b) has demonstrated how access to censored samples on hours of work, wages for workers, and employment choices identifies the joint distribution of the value of nonmarket time and potential market wages under a normality assumption. Heckman and Honore (1990) considered nonparametric versions of this model without labor supply.}

\footnote{We could augment decision rule (18) to be \( D = I(Y^0 - Y^1 - k(Z) \geq 0) \). Provided that we measure \( Z \) and condition on it, and provided that \( (U_1 - U_0) \perp \! \! \! \perp (X, Z) \), the model remains nonparametrically identified. The crucial property of the identification result is that no new unobservable enters the model through the participation equation. However, if we add \( Z \), subjective valuations of gain \( (Y^0 - Y^1 - k(Z)) \) no longer equal "objective" measures \( (Y^0 - Y^1) \).}
### Table 8.2. What Cross-Section Data Nonparametrically Identify from \( F(y_1|D = 1, X) \) and \( F(y_0|D = 0, X) \), \( \Pr(D = 1|X) \) under Different Assumptions

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Recovers</th>
<th>Behavioral assumption</th>
<th>Criteria recovered</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) ( E(Y^D = 1, X) = E(Y^X) = E(Y^D = 0, X) ) (conditional mean independence)</td>
<td>( E(Y^D = 1, X) = E(Y^D = 0, X) )</td>
<td>( Y^D ) does not determine participation conditional on ( X )</td>
<td>Cost-benefit (for ( D = 1 ) and total population)</td>
</tr>
<tr>
<td>( (Y^D, Y^X) = D ) (conditional independence)</td>
<td>( (Y^D, Y^X) = D )</td>
<td>( Y^D ) does not determine participation conditional on ( X )</td>
<td>Cost-benefit (for ( D = 1 ) and total population); Benthamite (for ( D = 1 ))</td>
</tr>
<tr>
<td>(4) Assumption (3) plus ( \Delta ) ( = Y^D )</td>
<td>( (Y^D, Y^X) = D )</td>
<td>( Y^D ) does not determine participation conditional on ( X )</td>
<td>Cost-benefit and Benthamite (for ( D = 1 ) and total population)</td>
</tr>
<tr>
<td>( \Delta ) ( = Y^D )</td>
<td>( (Y^D, Y^X) = D )</td>
<td>( Y^D ) does not determine participation conditional on ( X )</td>
<td>All criteria</td>
</tr>
</tbody>
</table>

*If the IV assumption is interpreted to be \( (U_{x}, U_{i}) \) \( \perp \mid \) \( D \mid \) \( X, Z \), or \( (Y^D, Y^X) \) \( \perp \mid \) \( D \mid \) \( X, Z \), IV is just matching conditional on \( X \) and \( Z \), and line (3) applies for this conditioning set.

---

This is a strong form of stochastic dominance. All of the mass of the \( Y \) distribution conditional on \( Y \) is to the right of the mass of the \( Y \) distribution conditional on \( Y \). More generally, because the \( Y \) distribution may not know \( (Y^D, Y^X) \), but may base their participation decisions on unbiased guesses \( (Y^D, Y^X) \), we can model the participation in the following way:

\[
\Pr(Y > y_0, Y > y_1 | D = 1, X) = \Pr(Y > y_0, Y > y_1 | D = 1, X) = 1
\]

This implies that \( Y \) is right-tail increasing in \( y \). That is, \( \Pr(Y > y_0, Y > y_1 | D = 1, X) \) is increasing in \( y \) for all \( y \). This in turn implies that \( \Pr(Y > y_0, Y > y_1 | D = 1, X) \) is increasing in \( y \).
\( D = 1, X \) is nondecreasing in \( y^0 \) for all \( y^1 \). Intuitively, the higher the value of \( y^0 \), the more the mass in the condition \( Y^0 \) distribution is shifted to the right, so that "high values of \( Y^0 \) go with high values of \( Y^1 \)." \( Y^1 \) being right-tail-increasing, given \( y^0 \), implies that \( Y^1 \) and \( Y^0 \) (given \( D = 1 \)) are positive-quadrant-dependent, so that \( \Pr(Y^1 \leq y^1 | Y^0 \leq y^0, D = 1, X) \geq \Pr(Y^1 \leq y^1 | D = 1, X) \) and \( \Pr(Y^1 \leq y^1 | Y^0 \leq y^0, D = 1, X) \geq \Pr(Y^1 \leq y^1 | D = 1, X) \). Common measures of dependence like the product-moment correlation, Kendall's \( \tau \), and Spearman's \( \rho \) are all positive when there is positive-quadrant dependence. Even under imperfect information, rationality in the form considered here can restrict the nature of the dependence between \( Y^1 \) and \( Y^0 \) given \( D = 1 \). Evidence against such dependence is evidence against the income-maximizing Roy model. Even if \( Y^0 \) and \( Y^1 \) are negatively correlated in the population, they are positively correlated given \( D = 1 \) if agents are income maximizers. This insight motivates our imposition of positive dependence between \( Y^0 \) and \( Y^1 \) in participant populations (\( D = 1 \)) to recover the joint distribution \( F(y^0, y^1 | D = 1, X) \) in the empirical analysis reported in Section 4.

Identification through the Instrumental-Variable Moment Condition and Extensions of the Condition. Taking (16) as a point of departure, it is possible under conditions we now specify to apply the method of instrumental variables to estimate \( E(Y^1 - Y^0 | D = 1, X) \) and \( E(Y^1 - Y^0 | X) = E(\Delta X) \). This allows implementation of the cost–benefit criterion provided that instrumental variables \( Z \) exist that satisfy the following conditions:

\[
E[U_0 + D(U_1 - U_0) | X, Z] = 0 \quad \text{for identifying } E(Y^1 - Y^0 | X) \quad (I-6a)
\]

or

\[
E[U_0 + D(U_1 - U_0) - E(U_1 - U_0 | D = 1, X) | X, Z] = 0 \quad \text{for identifying } E(Y^1 - Y^0 | X, D = 1) \quad (I-6b)
\]

A second condition is that \( D \) depend on \( Z \):

\[
\Pr(D = 1 | X, Z = z) \neq \Pr(D = 1 | X, Z = z') \quad \text{for some } z \neq z' \text{ for all } X
\]

Under condition (I-6a) we can write

\[
E(Y^1 | X, Z) = \mu_0(X) + E(\Delta X) \Pr(D = 1 | X, Z)
\]

Or, under condition (I-6b), we obtain

\[
E(Y^1 | X, Z) = \mu_0(X) + E(\Delta X) \Pr(D = 1 | X, Z)
\]

These implications are strict except in the case where \( Y^0 \) and \( Y^1 \) are binary random variables. In that case, Tong (1980) has shown that these notions of dependence are all equivalent.

Evaluating the Welfare State

\[
E(Y^1 | X, Z) = \mu_0(X) + E(\Delta | D = 1, X) \Pr(D = 1 | X, Z)
\]

Thus the population moment equation that identifies \( E(\Delta | X) \) under (I-6a) and (I-7) is

\[
E(\Delta | X) = \frac{E(Y^1 | X, Z = z) - E(Y^1 | X, Z = z')}{\Pr(D = 1 | X, Z = z) - \Pr(D = 1 | X, Z = z')} \quad (19)
\]

and the population moment equation that identifies \( E(\Delta | X, D = 1) \) under (I-6b) and (I-7) is the same:

\[
E(\Delta | X, D = 1) = \frac{E(Y^1 | X, Z = z) - E(Y^1 | X, Z = z')}{\Pr(D = 1 | X, Z = z) - \Pr(D = 1 | X, Z = z')} \quad (20)
\]

To satisfy condition (I-6b), it is required that the standard instrumental-variables condition \( E(U_0 | X, Z) = 0 \) be satisfied and, in addition, that

\[
E(U_1 - U_0 | D = 1, X, Z) = E(U_1 - U_0 | D = 1, X)
\]

Notice that condition (I-6b) is still satisfied if \( U_1 = U_0 \), so the response to treatment in (16) is the same for everyone, as assumed by Heckman (1978). This condition is also satisfied if

\[
(U_1 - U_0) \cup (X, Z, D)
\]

As a consequence of (I-8),

\[
\Pr(D = 1 | X, Z, Y^1 - Y^0) = \Pr(D = 1 | X, Z, U^1 - U^0) = \Pr(D = 1 | X, Z)
\]

The identifying assumption (I-8) will be satisfied if agents know \( X \) and \( Z \) but cannot predict \( (U_1 - U_0) \) at the time they make their decisions to participate in the program. Condition (I-8) will also be satisfied if (I-8) is weakened to a statement about mean independence:

\[
E(U_1 - U_0 | D = 1, X, Z) = E(U_1 - U_0)
\]

(1-8)'

which will be satisfied if the unobserved components of the gain do not determine program participation.\footnote{Condition (I-8) does not rule out that \( Y^0 \) determines \( D \), but if \( Y^0 \) does determine \( D \), it is required that, given \( X, Z \), and \( Y^0, \Delta \) does not determine \( D \). Under condition (I-8), \( E(Y^1 - Y^0 | D = 1, X) = E(Y^1 - Y^0 | X) \). The effect of "treatment" on the "treated" is the same as the effect of taking a person from the population at random and assigning that person to treatment. Moreover, (I-8) ensures that (I-6a) is satisfied as well.}

See Heckman (1977a) for further discussion of these conditions.
Notice that condition (19) is still satisfied if (I-6a) is weakened to
\[ E[U^0 + D(U_i - U_0)|X, Z] = M_r(X) \] (I-6a)' and condition (20) still holds if (I-6b) is weakened to
\[ E[U_0 + D(U_i - U_0) - E(U_i - U_0|D = 1, X)|X, Z] = M_r(X) \] The \( M_r(X) \) and \( M_r(X) \) terms difference out in the instrumental-variables moment conditions (19) and (20), respectively.

Invoking (I-6) and (I-7) under assumptions (A-1) and (A-2), we can answer the cost–benefit questions for the entire population if we assume (I-6a) and for populations for which \( D = 1 \) if we assume (I-6b). These assumptions are not strong enough to identify the Benthamite criterion or the voting criterion. Recovery of the full joint distribution of \((U_i, U_0, D)\) requires strengthening these assumptions. The conditional independence assumption that justifies matching (I-3) will suffice.

Thus, in place of (I-3), which is defined solely in terms of variables \(X\) in the outcome equations, we can assume that access to a variable \(Z\) produces conditional independence:
\[ (U_0, U_1) \perp D|X, Z, \quad \text{but} \quad (U_0, U_1) \not\perp D|X \] (I-9)
Equivalently, we can write
\[ (Y_0, Y_1) \perp D|X, Z, \quad \text{but} \quad (Y_0, Y_1) \not\perp D|X \] (I-9)'

Under these assumptions we can recover the marginal and joint distributions as discussed in the subsection on conditional independence. Interpreted in this way, the instrumental-variables method generalizes the matching method and extends the identification analysis based on conditional independence in terms of variables in the outcome equation to utilize a larger conditioning set beyond those variables.\(^{13}\)

2.1.2 Social Experiments

We consider randomization administered at two different points: (a) at entry or the stage where persons have applied to and been accepted into a program and (b) at eligibility. As noted by Heckman (1992) and Heckman and Smith (1993, 1995), and Heckman, LaLonde, and Smith (in press-b), social experiments with randomization administered at the stage where persons have applied to and been accepted into a program recover two marginal distributions conditional on \(D = 1:\)
\[ F(y_0^o|D = 1, X) \quad \text{and} \quad F(y_i^o|D = 1, X) \] (21)

\(^{13}\) Heckman, Ichimura, and Todd (1997b, 1986a) and Heckman et al. (1996; in press-a) have extended matching to consider variables in the program participation equation that are not in the outcome equation.

From such an experiment we obtain a truncated sample, and experiments administered at this stage do not identify \( \Pr(D = 1|X) \) (Heckman, 1992; Moffitt, 1992). The identifying assumptions that justify this method are

Randomization does not change the program being studied (no randomization bias), and no close substitutes for the treatment are available to persons randomized out (no substitution bias). (I-10)

Heckman (1992) and Heckman and Smith (1993) have discussed the need for the absence of substitutes for the program being evaluated and the failure of the no-randomization bias assumption. Heckman et al. (1997a) have provided evidence on the importance of substitution bias in an evaluation of a major job-training program. See also the evidence on these questions assembled in Heckman et al. (in press-b).

From the conditional distributions it is possible to recover the information required to construct the participant versions of the cost–benefit criteria
\[ F(Y^1 - Y^0|D = 1, X) \]
and the Benthamite criterion. Without further assumptions, social experiments do not recover the joint distribution
\[ F(y_0^o, y_i^o|D = 1, X) \] (22)

Any one of several additional assumptions can be used to supplement the information available from social experiments. The joint distribution (22) can be bounded from the experimentally determined marginals using the Fréchet bounds (Heckman and Smith, 1993; Heckman et al., 1997). Assumptions can be made about the association of quantile ranks (dependence) between outcomes across distributions to recover \( F(y_0^o, y_i^o|D = 1, X) \). An alternative assumption is (I-5).

With these assumptions we can construct or bound all of the evaluation criteria presented in Section 1 for the conditional (on \( D = 1 \)) distribution. Under conditional independence assumption (I-3), it is possible to recover the complete marginal distributions \( F(y_0^o|D = 1, X) = F(y_0^o|X) \) and \( F(y_i^o|D = 1, X) = F(y_i^o|X) \) and bound \( F(y_0^o, y_i^o|X) \) using the Fréchet bounds, as well as to identify \( F(y_0^o, y_i^o|X) \) by (a) making an assumption connecting the quantiles of the two marginal distributions or (b) assuming, as in (I-5), that gains \( \Lambda \) are unrelated to the base state \( Y^0 \).

If decision rule (18) is postulated, we can use the Roy model (under the conditions specified in Theorem A-1) to identify \( F(y_0^o, y_i^o|X) \) from the conditional distributions \( F(y_0^o|D = 1, X) \) and \( F(y_i^o|D = 1, X) \). Under assumptions (A-1) and (A-2) we can answer the evaluation questions comparing policy \( j \) with policy "0" that were posed in Section 1 for the entire population and the conditional (for participants) population.
Under more general participation rules we can apply Theorem A-2 to data from a social experiment with randomization administered at the point of entry into the program to identify $F(y^1_t, D_{tX})$ and $F(y^0_t, D_{tX})$ for both $D = 1$ and $D = 0$. Thus we can construct the cost–benefit and Benthamite criteria for the general population and for the participant populations, but not the general voting criterion or any other criterion requiring the joint distribution of outcomes.

One advantage of social experiments over conventional microeconomic data augmented with the conditional independence condition (I-3) is that experiments expand the range of the support over which the parameters can be identified. Thus, suppose that $\text{Support}(X|D = 1) \neq \text{Support}(X|D = 0)$. For the domains of $X$ in which there is no common support, Theorems A-1 and A-2 do not apply, and we cannot use conditional independence assumption (I-3). Randomization guarantees that in the population generating the experimental samples, $\text{Support}(X|D = 1)$ will be the same for participants and randomized-out persons. Thus, randomization ensures that the support conditions of Theorems A-1 and A-2 are satisfied for the population of participants. However, it may still happen that the support of $X$ for the population for which $D = 1$ is not the same as the support of $X$ for the whole population. Then, even with experimental data, the parameters of interest are identified only over the available support. For both experimental and nonexperimental data it may be necessary to sample more widely on $X$ coordinates to recover parameters defined for the entire population. Experiments have the advantage that they allow identification of impacts even for persons with values of $X$ such that $\text{Pr}(D = 1|X) = 1$, which is not possible using nonexperimental methods because there is no comparison group.

If randomization is performed on eligibility for the program, we recover $F(y^1_{tX})$, $F(y^1_{tD = 1, X})$, and $F(y^0_{tD = 1, X})$ (Heckman, 1992; Heckman and Smith, 1993; Heckman et al., in press-b). In addition, we recover $\text{Pr}(D = 1|X)$, at least for those values of $X$ possessed by eligible persons. Many would regard $F(y^1_{tX})$ as a better approximation to the no-policy outcome distribution than the approximation embodied in assumption (A-1). Although both approximations ignore general-equilibrium effects, $F(y^0_{tX})$ avoids self-selection bias. Randomization at eligibility does not recover the full joint distribution of outcomes unless additional assumptions of the type previously discussed are invoked. Table 8.3 summarizes the information obtained from the two types of experiments.

### 2.1.3 Panel Data

Panel data provide a new source of identifying information. Participation or nonparticipation outcomes in one period can proxy participation or nonparticipation outcomes in another period. Restoring the $t$ sub-

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Behavioral assumption</th>
<th>Recovers</th>
<th>Criteria recovered without further assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomization at entry</td>
<td>No randomization or substitution bias</td>
<td>Same as column 1</td>
<td>$F(y^1_{tD = 1, X})$, $F(y^0_{tD = 1, X})$</td>
</tr>
<tr>
<td>Randomization at eligibility</td>
<td>No randomization or substitution bias</td>
<td>Same as column 1</td>
<td>$F(y^1_{tX})$, $F(y^1_{tD = 1, X})$, $F(y^1_{tD = 1, X})$, $\text{Pr}(D = 1</td>
</tr>
</tbody>
</table>

script, panel data allow us to make the following approximations for person $i$:

\[ Y^1_{ni} = Y^1_n \quad (t \neq t') \]  

\[ Y^0_{ni} = Y^0_n \quad (t \neq t') \]  

Provided that the approximations are valid ("≈" is "≈"), we can substitute for the missing counterfactual outcome for each person and identify the joint distribution of $(Y^0_n, Y^1_n)$ for different conditioning sets. We can answer all of the questions posed in Section 1 for period-$t$ versions of the criteria presented there. It is the ability to directly estimate the dependence across potential outcomes without invoking additional assumptions that is the distinguishing feature of panel data.

When adding a temporal dimension to the analysis, it is useful to distinguish reversible programs from irreversible programs. Human-capital or personal-investment programs have certain irreversibility features, but it is typically assumed that they have no effect on preprogram outcomes. For such programs, we require $t' < t$ in (I-11), where $t$ is the period of participation. Reversible programs switch on and off and have no lasting effects. Examples may include job subsidies or unemployment-insurance benefits. With reversible programs we can go forward or back-

---

14 If agents anticipate participation in the program, they may take actions in the preprogram period that distinguish them from nonparticipants. The assumed absence of anticipatory behavior is central to received models of program evaluation.
ward in time in the search for a valid counterfactual state, so that we may have \( t' < t \) or \( t' > t \) in (I-11). We first consider reversible programs.

**Reversible Programs.** Nonstationarity in the external environment, the effects of aging and life-cycle investment, and idiosyncratic period-specific shocks render assumptions (I-11a) and (I-11b) suspect. To circumvent these problems, the identifying assumptions usually are reformulated at the population level, and conditioning variables \( X \) are assumed that "adjust" \( Y^0 \) and \( Y^1 \) to equality in distribution or conditional mean and allow for idiosyncratic fluctuations. For simplicity, we conduct only a two-period analysis, but estimation of the necessary adjustments may require more data.\(^{15}\) The modified identification conditions become

\[
F(y^0, y^1 | D_t = 0, D_r = 0, X) = F(y^0, y^1 | D_t = 0, X) \quad \text{for } y^0 = y^1 \quad (I-12a)
\]

We can use (I-11b) to construct all of the counterfactuals in period \( t \) conditional on \( D = 0 \) without any further adjustments if it is assumed that

\[
F(y^0, y^1 | D_t = 0, D_r = 1, X) = F(y^0, y^1 | D_t = 0, X) \quad \text{for } y^1 = y^1' \quad (I-12b)
\]

\[
F(y^0, y^1 | D_t = 0, D_r = 1, X) = F(y^0, y^1 | D_t = 1, X) \quad \text{for } y^0 = y^0', \text{ or } F(y^0, y^1 | D_t = 1, D_r = 1, X) = F(y^0 | D_t = 1, X) \quad (I-12b)
\]

Much less often is it also assumed that \( F(y^0, y^1 | D_t = 0, D_r = 0, X) = F(y^0 | D_t = 0, X) \) for \( y^0 = y^0' \), or \( F(y^0, y^1 | D_t = 1, D_r = 1, X) = F(y^1 | D_t = 1, X) \) for \( y^1 = y^1' \), although these assumptions seem equally plausible and are testable. They would require that the \( Y^0 \) and \( Y^1 \) be perfectly dependent, as are outcomes \( Y^0 \) and \( Y^0 \).

For means, the weaker versions of (I-12a) and (I-12b) are, respectively,\(^{15}\)

\[
E(Y^0 | D_t = 0, D_r = 1, X) = E(Y^0 | D_t = 1, X) \quad \text{[for (I-11a)']} \quad (I-12a')
\]

and

\[
E(Y^1 | D_t = 0, D_r = 1, X) = E(Y^1 | D_t = 0, X) \quad \text{[for (I-11b)']} \quad (I-12b')
\]

These are strong implicit behavioral assumptions. Assumptions (I-12a) and (I-12a)\(^{15}\) require that persons who participate in \( t \) but not in \( t' \) have the same no-treatment mean outcome in \( t' \) as persons who take treatment in period \( t \) would have in \( t \). It rules out that the switch from \( D_r = 0 \) to \( D_r = 1 \) is caused by differences in \( Y^0 \) between \( t' \) and \( t \). More precisely, it excludes \( Y^0 \) as a determinant of \( D_r \). Assumptions (I-12b)\(^{15}\) and (I-12b)\(^{15}\) are comparable assumptions about the lack of influence of \( Y^0 \) in determining participation in \( t' \).

One way to justify these identifying assumptions is to postulate a strengthened form of the conditional independence assumption used to justify matching:

\[
(Y^0_{t'}, Y^1_{t'}) \perp (D_t, D_r) | X, \quad t \neq t' \quad (I-13)
\]

This condition rules out any dependence between \( D_t \) and \( D_r \) and the components of \( (Y^0_{t'}, Y^0_{t'}) \) that cannot be predicted by \( X \). This assumption rules out selection on any unobserved components of potential outcomes. It is inconsistent with the Roy model. A weaker version of (I-13) is that conditional on \( D_t \) and \( X \), \( (Y^0_{t'}, Y^0_{t'}) \) are independent of \( D_r \):

\[
(Y^0_{t'}, Y^0_{t'}) \perp D_r | X, D_t \quad (I-14)
\]

This condition rules out any dependence between the components of \( (Y^0_{t'}, Y^0_{t'}) \) that cannot be predicted by \( D_t \) and \( X \) and the random variable \( D_r \). (I-12a)\(^{15}\) and (I-12b)\(^{15}\) can be justified by these assumptions.

---

\(^{15}\) See Heckman and Robb (1986, pp. 210–15), where these adjustments are discussed in detail. See also Heckman et al. (in press-b).

\(^{16}\) In the method of "difference in differences" it is assumed that a common trend operates on all persons, irrespective of their participation status. The trend is eliminated from the means by comparing participant change to nonparticipant change. More generally, nonparticipants in \( t \) and \( t' \) can be used to identify the common trend.
We could augment (I-13) or (I-14) to include matching variables Z not included in X. Thus it may happen that (I-13) does not hold, but

\[(Y^0_{\tau'}, D^0_{\tau'}) \perp (D_{\tau}, D_t)|X, Z\]  \hspace{1cm} (I-13)'

Similarly, (I-14) may be invalid, but it may happen that

\[(Y^0_{\tau'}, D^0_{\tau'}) \perp D_{\tau}|X, D_t, Z\]  \hspace{1cm} (I-14)'

is valid. We could also invoke other assumptions patterned after our cross-section analysis to recover the missing counterfactual state. We could model participation in periods \(t\) and \(t'\) using dynamic selection models. Each cross-section estimator has a panel-data counterpart, which, for the sake of brevity, we do not develop in this study.

If the date of enrollment into the program is exogenous, it is incorrect to simply condition on it, and conditions (I-13) and (I-14) have to be strengthened in order to avoid building an explicit model of the date of enrollment.\(^{17}\) Let \(\tau\) be the date of enrollment into the program. Then to use (I-12a) and (I-12b) without modification, we need to augment the conditional independence assumptions to read

\[(Y^0_{\tau'}, Y^0_{\tau}) \perp (D_{\tau}, D_t, \tau)|X\]  \hspace{1cm} (I-13)''

or, in the weaker form,

\[(Y^0_{\tau'}, Y^0_{\tau}) \perp (D_{\tau}, \tau)|X, D_t\]  \hspace{1cm} (I-14)''

These conditions rule out dependence between potential outcomes and the set of participation variables conditional on \(X\) [[(I-13)''], or dependence between potential outcomes and non-\(t\) participation variables conditioned on \(X\) and \(D_t\) [[(I-14)'']]. Under either set of assumptions, we can ignore the date of enrollment as a factor in producing the counterfactual distributions.

Other types of identifying assumptions could be invoked. Cameron and Heckman (1991a,b) developed a multivariate version of the Roy model that explicitly models \(\tau\) and showed that its parameters can be identified. Those models are closely related to standard panel-data attrition models (e.g., Ridder, 1990).

The Irreversible Case. In the irreversible case, there are no counterparts for (I-11b)'', (I-11b)''' (I-12b)', or (I-12b)''' because there are no observations on treated persons in the preprogram period \(t'\). First consider the

case in which program enrollment date \(\tau\) is fixed and is common for all persons. The probability space is restricted, so \(Pr(D_{\tau} = 1|X) = 0\), and no value of \(Y^0_{\tau}\) is defined. \(F(y^0_{\tau}, y_{\tau}|D_{\tau} = 1, X)\) can be identified from \(F(y^0_{\tau}, y_{\tau}|D_{\tau} = 1, X)\) if the preprogram outcomes of participants have the same relationship to program outcomes in \(t\) as their non-program outcomes in period \(t\). [This is just assumption (I-12a).]

We cannot use (I-12b) to construct \(F(y^0_{\tau}, y_{\tau}|D_{\tau} = 0, X)\) because no value of \(Y^0_{\tau}\) is defined. In the irreversible case, we have a truncated sample.

If we invoke a conditional independence assumption and assume a counterpart to (I-12) defined for the reversible case,

\[(Y^0_{\tau}, Y^0_{\tau}) \perp D_{\tau}|X]\]  \hspace{1cm} (I-15)

we can identify the full joint distribution.\(^{19}\) Otherwise we can identify only the evaluation criteria for the population conditional on \(D_{\tau} = 1\). Because we know \((Y^0_{\tau}, Y^0_{\tau})\) conditional on \(D_{\tau} = 1\), we can use a vector generalization of Theorem A-2, presented in Appendix A as Theorem A-3, to identify \(F(y^0_{\tau}, y_{\tau}|X)\) and \(F(y^0_{\tau}, y_{\tau}|D_{\tau}, X)\). What is required is a set of \(X\) values such that \(Pr(D_{\tau} = 1|X) = 0\). Under the assumptions made in Theorem A-3, it is possible to recover the full distribution of outcomes even in the irreversible case.

If \(\tau\) is not the same for everyone, and is random, but \(t' < \tau < t\), then to use (I-15) we need to assume

\[(Y^0_{\tau}, Y^0_{\tau}) \perp D_{\tau}, \tau|X]\]  \hspace{1cm} (I-16)

or

\[(Y^0_{\tau}, Y^0_{\tau}) \perp D_{\tau}|\tau, X]\]  \hspace{1cm} (I-16)'

These assumptions enable us to ignore the date of enrollment as a determinant of outcomes in constructing the counterfactual distributions.

Heckman and Robb (1985) discussed more general uses of panel data to proxy unobservables to eliminate selection bias. The leading cases are fixed-effects or autoregressive models that transform equations by differencing or generalized differencing to eliminate unobserved components that would produce selection bias. All of the conventional “proxy-variable” econometric methods that eliminate selection bias through some transformation of the original equations can be shown to be equivalent to constructing counterfactual outcomes (i.e., producing predicted values of the outcomes needed to form the missing component of the

\(^{17}\) In a fully dynamic model in which enrollment dates are endogenous, the date of enrollment would be a further source of information about revealed preferences, which we do not pursue in this study. Qualitatively, it conveys information on subjective evaluations in the same way attrition and self-selection decisions convey information about choices.

\(^{18}\) The required modification for conditional means is obvious and thus is omitted.

\(^{19}\) This assumption could be augmented to allow for \(Z\) to be added to the conditioning set so that we would have \((Y^0_{\tau}, Y^0_{\tau}) \perp D_{\tau}|X, Z\), but (I-15) would be invalid.
counterfactual). More generally, if the original equations are subject to transformations, the previously stated identification conditions apply to the transformed equations (Heckman, 1997b). Summaries of the main identification results for joint distributions and means and marginal distributions that exploit panel data are given in Tables 8.4 and 8.5, respectively.

2.1.4 Repeated Cross-Section Data

Heckman and Robb (1985) demonstrated that all panel-data identification assumptions about means, variances, and covariances have counterparts in repeated cross-section data. Conditional-mean versions of all of the identification assumptions presented in Section 2.1.3 have counterparts in repeated cross sections of unrelated persons sampled from the same populations. We first consider the reversible case.

Identification conditions (I-11a)" and (I-11b)" can be defined for a common population and do not require that the same persons be followed over time. The same is true for (I-12a)" and (I-12b)" and the other identifying assumptions for conditional means. However, it now becomes necessary to classify persons in t' as program participants or nonparticipants in t. This is not so easy to do in the repeated cross-section case, because different persons are sampled in t and t'. What is lost when the analyst is restricted to using repeated cross-section data is the ability to construct joint distributions (Yₜ, Yₜ') without invoking the cross sectional assumptions made in Section 2.1.1.

Without invoking additional assumptions about dependence between the two potential outcomes, the identifying assumptions for conditional means enable us to recover only the cost–benefit and Benthamite criteria, not the voting criterion, which is based on the full joint distribution of potential outcomes. An essential benefit of panel data—that they afford nonparametric identification of the joint distribution of potential outcomes under the identifying assumptions made in Section 2.1.3—is lost when the analyst has access only to repeated cross-section data. A summary of the main cases for panel data and repeated cross sections is presented in Table 8.5.

Table 8.4: Panel Data Main Cases for Distributions

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Recovery</th>
<th>Criteria Recovered</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Fᵧₑ₀</td>
<td>Fₓₑ₀ = Fₓᵧₑ₀(X); ŷ = yₑ₀</td>
<td>Stationarity or sufficient information to adjust for stationarity</td>
</tr>
<tr>
<td>(2) Fᵧₑ₀</td>
<td>Fₓₑ₀ = Fₓᵧₑ₀(X); ŷ = yₑ₀</td>
<td>Stationarity plus reversibility</td>
</tr>
<tr>
<td>(3) Fᵧₑ₀</td>
<td>Fₓₑ₀ = Fₓᵧₑ₀(X); ŷ = yₑ₀</td>
<td>Among participants in t, participants in t' consistent with both irreversibility and reversibility of program for generalized x</td>
</tr>
<tr>
<td>(4) Fᵧₑ₀</td>
<td>Fₓₑ₀ = Fₓᵧₑ₀(X); ŷ = yₑ₀</td>
<td>Among nonparticipants in t, participants in t' consistent with both irreversibility and reversibility of program for generalized x</td>
</tr>
</tbody>
</table>

For example, in the method of fixed effects without regressors, Yᵢ = αᵢ + εᵢ, and Yᵢ' = εᵢ, where E(εᵢ) = 0 and εᵢ ~ N(0, σₑ). If Pr(D = 1 | εᵢ) = P, which is not a function of εᵢ, we can identify E(αᵢ | D = 1) = E(Yᵢ - Yᵢ') / D = P. Observe that P can depend on εᵢ; see Heckman and Robb (1985). Heckman (1997b) demonstrated that Yₑ₀ is properly interpreted as a proxy for Yᵢ. If there are regressors, we can modify this example to allow for use of X-adjusted (Yᵢ - Yᵢ') ([(Yᵢ - Xᵢβ) - (Yᵢ' - Xᵢβ)]), where for convenience we assume a common β.

21 The modification of the analysis in this subsection to account for irreversibility is straightforward and thus is omitted.
3 The Relationship between Traditional Cost–Benefit Analysis and the Parameters Widely Used in the Econometric Evaluation Literature

In this section we relate the parameters estimated in the microeconomic evaluation literature to the parameters needed to perform cost–benefit analysis. We present empirical evidence on the importance of accounting for the direct costs of a program and the marginal welfare costs of taxation in assessing the net benefits of a policy. We follow the literature in cost–benefit analysis and assume that the policy being evaluated has a voluntary component and that valid evaluations of a policy can be derived from looking at the impact of the policy on self-selected participants and nonparticipants.

We postulate the following framework: For a given program associated with policy j there are two discrete outcomes corresponding to direct receipt of treatment \((D_j = 1,\) for program participation) and no treatment \((D_j = 0),\) and there is a set of program-intensity variables \(q_j\) defined under policy j that affect outcomes in the two states and the allocation of persons to treatment or nontreatment. The program-intensity variables \(q_j\) may be discrete or continuous. Policy "0" is a no-intervention benchmark with program intensity \(q_0.\)

Assuming that costless lump-sum transfers are possible, that a single social-welfare function governs the distribution of resources, and that prices reflect true opportunity costs, traditional cost–benefit analysis (e.g., Harberger, 1971; Boudreaux and Bruce, 1984) seeks to determine the impact of programs on the total output of society. Efficiency becomes the paramount criterion in this framework, with the distributional concerns assumed to be taken care of through lump-sum transfers and taxes. In this framework, impacts on total output, as in the evaluation criterion (3), are the only objects of interest in evaluating policies.

For policy j, let \(y_{ij}^r\) and \(y_{ij}^n\) be individual outputs for person i in the direct-participation state \((D_j = 1)\) and direct-nondirect state \((D_j = 0),\) respectively. The vector of program-intensity variables \(q_j\) operates on all persons within the context of program j, although its effects need not be uniform. It determines, in part, participation in the program. We can write \(D_j(q_j)\) as the indicator for participating in program j when the program intensity is \(q_j.\) To simplify notation, we keep implicit any conditioning on personal characteristics that may affect both participation and outcomes. We define \(c(q_j)\) as the social cost of \(q_j\) denominated in units of output. In general, policies could be designed for specific persons, but we do not consider that possibility here. We assume that \(c(0) = 0\) and that \(c_j\) is convex and increasing in \(q_j.\) The value \(q_0\) defines another benchmark policy, "0", in which there is no program and therefore no participants. This policy has the associated cost function \(c_0(0).\)
When \( \varphi_j = 0 \), there might be effects of policy \( j \) on output that would distinguish that policy from the no-policy regime "0." A law that is universally assented to and accepted might raise output at no cost (e.g., adopting a convention about driving on the right-hand side of the road). Output could be different in a policy without the law (policy "0"), but the direct costs of enforcement would be the same under both policies.

Letting \( N_i(\varphi) \) be the number of direct program participants, and \( N_0(\varphi) \) the rest of the population, the total output of society under policy \( j \) at program-intensity level \( \varphi_j \) is

\[
N_i(\varphi_j)E[Y_j|D(\varphi_j) = 1, \varphi_j] + N_0(\varphi_j)E[Y_j^0|D(\varphi_j) = 0, \varphi_j] - c(\varphi_j)
\]

where \( N_i(\varphi) + N_0(\varphi) = \bar{N} \) is the total number of persons in society. Vector \( \varphi \) appears twice in the conditioning arguments as a determinant of \( D \) and as a determinant of the output levels in the different states. Vector \( \varphi_j \) is general enough to include financial-incentive variables as well as mandates that assign persons to a particular treatment state. Recall that we keep conditioning on personal characteristics implicit.

Assume, for simplicity, differentiability of the treatment choice and mean-outcome functions, and further assume that \( \varphi_j \) is a scalar, a simplifying assumption that is easily relaxed. The change in output in response to a marginal increase in the policy-intensity parameter \( \varphi_j \) from any given position is

\[
M(\varphi_j) = \frac{\partial N_i(\varphi_j)}{\partial \varphi_j} \{E[Y_j|D(\varphi_j) = 1, \varphi_j] - E[Y_j|D(\varphi_j) = 0, \varphi_j]\} + N_i(\varphi_j) \frac{\partial E[Y_j|D(\varphi_j) = 1, \varphi_j]}{\partial \varphi_j} + N_0(\varphi_j) \frac{\partial E[Y_j^0|D(\varphi_j) = 0, \varphi_j]}{\partial \varphi_j} - c'(\varphi_j)
\]

The first term arises from the change in the number of participants induced by the policy change. The second and third terms arise from changes in output among the policy participants and nonparticipants induced by the policy change. The fourth term is the marginal direct-output cost of the change in the intensity of policy \( \varphi_j \).

In principle, this measure could be estimated from time-series data on the change in aggregate GNP occurring after the policy-intensity parameter is varied. Under the assumption of a well-defined social-welfare function with interior solutions and the additional assumption that prices are constant at initial values, an increase in GNP at base-period prices will raise social welfare.\(^{22}\)

\(^{22}\) See, e.g., Laffont (1989, p. 155) or the comprehensive discussion by Chipman and Moore (1976).

---

Evaluating the Welfare State

If marginal program-intensity changes under policy regime \( j \) have no effect on intrasector mean output, the bracketed expressions in the second and third terms are zero. In this case, the parameters of interest are as follows:

(i) \( \frac{\partial N_i(\varphi_j)}{\partial \varphi_j} \), the number of people induced into program \( j \) by the change in \( \varphi_j \)

(ii) \( E[Y_j|D(\varphi_j) = 1, \varphi_j] - E[Y_j^0|D(\varphi_j) = 0, \varphi_j] \), the mean output difference between participants and nonparticipants

(iii) \( c'(\varphi_j) \), the direct social marginal cost of policy \( j \) at program-intensity level \( \varphi_j \)

It is revealing that nowhere in this list do we see the parameters that receive the most attention in the econometric policy-evaluation literature (e.g., Heckman and Robb, 1985). These are as follows:

(a) \( E[Y_j] - E[Y_j^0|\varphi_j = 1, \varphi] \), "the effect of treatment on the treated" for persons in regime \( j \) at policy intensity \( \varphi_j \)

(b) \( E[Y_j^0|\varphi_j = \bar{\varphi}] \), where \( \varphi_j = \bar{\varphi} \) sets \( N_i(\bar{\varphi}) = \bar{N} \); this is the effect of universal direct participation in program \( j \) compared with universal nonparticipation in \( j \) at a level of program intensity \( \bar{\varphi} \)

(c) \( E[Y_j] - E[Y_j^0|\varphi] \), the effect of randomly selecting persons for direct treatment and forcing their compliance with this treatment compared with their position in the no-participation state under policy \( j \) at program-intensity level \( \varphi_j \)

Parameter (ii) can be obtained from simple mean differences between the treated and the nontreated. No adjustment for selection bias is required. Parameter (i) can be obtained from knowledge of the net movement of persons into or out of direct participation in the program in response to the policy change, something usually not measured in microeconomic policy evaluations; for discussions of this problem, see Moffitt (1992) or Heckman (1992). Parameter (iii) can be obtained from cost data. It should include the full social costs of the program, including the welfare cost of raising public funds, although these are often ignored.

It is informative to place additional structure on this model. This leads to a representation of a criterion that is widely used in the literature on microeconomic program evaluation and also establishes a link with the discrete-choice literature in econometrics. Assume a binary-choice random-utility framework like that used in the Roy model. Suppose that under policy regime \( j \) with program-intensity level \( \varphi_j \), agents make choices to directly participate or not based on net utility and that policies affect participant utility through an additively separable term \( k(\varphi_j) \) that is assumed scalar and differentiable. Net utility from participating in the program is \( U_j = X + k(\varphi_j) \), where \( k \) is monotonic in \( \varphi_j \) and where
the joint distributions of \((Y^j, X)\) and \((Y^0, X)\) are \(F(y^j, X)\) and \(F(y^0, X)\), respectively. In the special case of the Roy model, \(X = Y^j - Y^0\), and \(k = 0\). In general, \(D(\varphi_i) = 1(U_i \geq 0) = 1[X \geq -k(\varphi_i)]\), so that

\[
N_i(\varphi_i) = \bar{N} \Pr(U_i \geq 0) = \bar{N} \int_{-k(\varphi_i)} f(x) \, dx
\]

\[
N_0(\varphi_i) = \bar{N} \Pr(U_i < 0) = \bar{N} \int_{-k(\varphi_i)} f(x) \, dx
\]

The total output is

\[
\bar{N} \int \int f(y^j, x|\varphi_i) \, dx \, dy^j
\]

\[
+ \bar{N} \int \int f(y^0, x|\varphi_i) \, dx \, dy^0 - c_i(\varphi_i)
\]

Under standard conditions\(^\text{24}\) we can differentiate under the integral sign to obtain the following expression for the marginal change in output with respect to a change in intensity parameter \(\varphi_i\) within policy regime \(j\):

\[
M(\varphi_i) = \bar{N}k(\varphi_i) f_i[-k(\varphi_i)] \left[ E[Y^j|D(\varphi_i) = 1, X = -k(\varphi_i), \varphi_i] - E[Y^0|D(\varphi_i) = 0, X = -k(\varphi_i), \varphi_i] \right]
\]

\[
+ \bar{N} \left[ \int \int \frac{\partial f(y^j, x|\varphi_i)}{\partial \varphi_i} \, dx \, dy^j \right] + \int \int \frac{\partial f(y^0, x|\varphi_i)}{\partial \varphi_i} \, dx \, dy^0 - c_i(\varphi_i)
\]

where \(f_i\), the marginal density of \(X\), is evaluated at \(X = -k(\varphi_i)\).

This model has a well-defined marginal-entry condition: \(X \geq -k(\varphi_i)\). The first set of terms corresponds to the gain arising from the movement of persons at the margin (the term in curly brackets) weighted by the proportion of the population at the margin, \(f_i[-k(\varphi_i)]\), times the number of people in the population. This term is the net gain from switching from nonparticipant to participant status. The expression in curly brackets in the first term is a limit form of the “local average treatment effect” of Imbens and Angrist (1994). The second set of terms is the within-treatment-status change in output resulting from the change in the program-intensity parameter. This term is ignored in many evaluation studies. It describes how people who do not switch their participation status are affected by the policy change. The third term is the direct marginal social cost of the policy change, which is rarely estimated. At a social planner’s optimum, \(M(\varphi_i) = 0\), provided standard second-order conditions are satisfied. Marginal benefit should equal marginal cost.

\(^{23}\) These are assumed to be absolutely continuous with respect to Lebesgue measure.

\(^{24}\) See, e.g., Royden (1968) for the required domination conditions.

---

**Evaluating the Welfare State**

Either a cost-based measure of marginal benefit or a benefit-based measure of cost can be used to evaluate the marginal gains or costs of the change in policy intensity.

Observe that the local average treatment effect is simply the effect of treatment on the treated at the margin \(X = -k(\varphi_i)\):

\[
E[Y^j|D(\varphi_i) = 1, X = -k(\varphi_i), \varphi_i] - E[Y^0|D(\varphi_i) = 0, X = -k(\varphi_i), \varphi_i]
\]

\[
= E[Y^j - Y^0|D(\varphi_i) = 1, X = -k(\varphi_i), \varphi_i]
\]

The proof of this result is immediate once it is recognized that the set \(X = -k(\varphi_i)\) is the indifference set for this problem. Thus the Imbens and Angrist parameter is a marginal version of the conventional evaluation parameter (“treatment effect on the treated”) for gross outcomes. This parameter is but one of the three ingredients required to produce an evaluation of social welfare under the cost–benefit criterion.

The conventional evaluation parameter (“treatment effect on the treated”) does not incorporate costs, does not correspond to a marginal change, and includes the effect of intramarginal changes. This parameter is, in general, inappropriate for evaluating the effect of a policy change on GNP. However, under certain conditions, which we shall now make precise, it is sometimes informative about the gross gain accruing to the economy from the existence of program \(j\) at level \(\varphi_i\) compared with the alternative of shutting it down and switching to policy “0.” The social cost associated with policy “0” is \(c_0(\varphi_i)\), which we assume to be zero: \(c_0(\varphi_i) = 0\).

The appropriate criterion for an all-or-nothing evaluation of a policy at level \(\varphi_i\) is

\[
A(\varphi_i) = \{N_i(\varphi_i) E[Y|D(\varphi_i) = 1, \varphi_i] + N_0(\varphi_i) E[Y|D(\varphi_i) = 0, \varphi_i] - c(\varphi_i)\}
\]

\[
- \bar{N} E[Y_0|\varphi_i]
\]

In the no-policy regime, there is only one output, \(Y_0\), and everyone is in the no-program state. If \(A(\varphi_i) > 0\), total output is increased by establishing program \(j\) at level \(\varphi_i\). In the special case in which the outcome in the nonparticipation state under regime \(j\) (\(Y_0^j\)) is the same as the outcome in the no-program state (\(Y_0\)) for both participants and nonparticipants under regime \(j\), we have

\[
E[Y^j|D(\varphi_i) = 0, \varphi_i] = E[Y_0|D(\varphi_i) = 0, \varphi_i]
\]

(23a)
and

\[ E[Y_1^* | D_j(\varphi_j) = 1, \varphi_j] = E[Y_0 | D_j(\varphi_j) = 1, \varphi_0] \]  \hspace{1cm} (23b)

The right-hand sides of both expressions describe hypothetical conditional expectations. The right-hand side of (23a) is what the outcome in the no-program state would be for persons who would not directly participate in the program under policy \( j \) with parameters \( \varphi_j \) [i.e., those for whom \( D_j(\varphi_j) = 0 \)]. The right-hand side of (23b) is the corresponding expression for persons who would participate in the program under policy \( j \) with intensity parameters \( \varphi_j \) [i.e., those for whom \( D_j(\varphi_j) = 1 \)]. These conditioning statements select out, respectively, nonparticipants and participants in policy regime \( j \) and compute the expected values of output in the policy-“0” regime.

Assuming that the probability of participation in regime \( j \) under program-intensity level \( \varphi_j \) does not depend on the value of \( \varphi_0 \) in the no-program state,

\[ \Pr(D_j = 1 | \varphi_j, \varphi_0) = \Pr(D_j = 1 | \varphi_j) \]  \hspace{1cm} (A-3)

then under assumption (A-1) we can use the law of iterated expectations to write

\[ E(Y_0^* | \varphi_0) = E[Y_0 | D_j(\varphi_j) = 1, \varphi_0] \Pr[D_j(\varphi_j) = 1 | \varphi_j] \]

\[ + E[Y_0 | D_j(\varphi_j) = 0, \varphi_0] \Pr[D_j(\varphi_j) = 0 | \varphi_j] \]

From (23a) and (23b) and (A-3) we obtain

\[ E(Y_0^* | \varphi_0) = E[Y_1^* | D_j(\varphi_j) = 1, \varphi_j] \Pr[D_j(\varphi_j) = 1 | \varphi_j] \]

\[ + E[Y_1^* | D_j(\varphi_j) = 0, \varphi_j] \Pr[D_j(\varphi_j) = 0 | \varphi_j] \]

Substituting for \( E(Y_0^* | \varphi_0) \) in the expression for \( A(\varphi_j) \), we obtain

\[ A(\varphi_j) = N(\varphi_j) E[Y_1^* - Y_1^* | D_j(\varphi_j) = 1, \varphi_j] - c_j(\varphi_j) \]  \hspace{1cm} (24)

which vindicates the use of the parameter “treatment effect on the treated” as an evaluation parameter in the case in which there are no general-equilibrium effects in the sense of assumption (A-1). This important case is applicable to small-scale social programs with partial participation. For evaluating the effects of fine-tuning the intensity levels of existing policies, measure \( M(\varphi_j) \) is more appropriate.

3.1 Empirical Evidence on the Importance of Adjusting for Direct Costs and the Welfare Costs of Taxation in Cost–Benefit Analysis

This subsection examines the effects of accounting for both the direct costs and the welfare costs of raising government tax revenue in computing benefit–cost estimates for a prototypical government training program. Accounting for direct costs and the welfare costs of government revenue substantially reduces the estimated returns to government training programs, correcting the use of the difference between costs and benefits for the Job Training Partnership Act (JTPA) program appear in Table 8.6. Benefits are measured using the difference in mean earnings between the experimental-treatment and control groups in the JTPA data, which are described more fully in Section 4 and in Appendix B. Direct costs represent the estimated difference in training costs between the treatment and control groups\(^2\) and are assumed to be incurred within the first 18 months of the program.


<table>
<thead>
<tr>
<th>Benefit duration</th>
<th>Direct costs included?</th>
<th>6-month interest rate</th>
<th>Welfare cost of taxes</th>
<th>Adult males</th>
<th>Adult females</th>
<th>Male youth</th>
<th>Female youth</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 months</td>
<td>No</td>
<td>0.000</td>
<td>0.000</td>
<td>1,354</td>
<td>1,703</td>
<td>−967</td>
<td>136</td>
</tr>
<tr>
<td>30 months</td>
<td>Yes</td>
<td>0.000</td>
<td>0.000</td>
<td>523</td>
<td>532</td>
<td>−2,922</td>
<td>−1,180</td>
</tr>
<tr>
<td>30 months</td>
<td>Yes</td>
<td>0.050</td>
<td>0.500</td>
<td>108</td>
<td>−54</td>
<td>−3,900</td>
<td>−1,828</td>
</tr>
<tr>
<td>30 months</td>
<td>Yes</td>
<td>0.025</td>
<td>0.000</td>
<td>433</td>
<td>432</td>
<td>−2,859</td>
<td>−1,195</td>
</tr>
<tr>
<td>30 months</td>
<td>Yes</td>
<td>0.025</td>
<td>0.500</td>
<td>17</td>
<td>−154</td>
<td>−5,836</td>
<td>−1,853</td>
</tr>
<tr>
<td>7 years</td>
<td>No</td>
<td>0.000</td>
<td>0.000</td>
<td>5,206</td>
<td>5,515</td>
<td>−5,843</td>
<td>865</td>
</tr>
<tr>
<td>7 years</td>
<td>Yes</td>
<td>0.000</td>
<td>0.000</td>
<td>4,375</td>
<td>4,344</td>
<td>−5,798</td>
<td>−451</td>
</tr>
<tr>
<td>7 years</td>
<td>Yes</td>
<td>0.050</td>
<td>0.500</td>
<td>3,960</td>
<td>3,758</td>
<td>−6,775</td>
<td>−1,109</td>
</tr>
<tr>
<td>7 years</td>
<td>Yes</td>
<td>0.025</td>
<td>0.000</td>
<td>3,523</td>
<td>3,490</td>
<td>−5,166</td>
<td>−610</td>
</tr>
<tr>
<td>7 years</td>
<td>Yes</td>
<td>0.025</td>
<td>0.500</td>
<td>3,108</td>
<td>2,905</td>
<td>−6,143</td>
<td>−1,268</td>
</tr>
</tbody>
</table>

Notes: (1) “Benefit duration” indicates how long the estimated benefits from JTPA are assumed to persist. Actual estimates are used for the first 30 months. For the 7-year duration case, the average of the benefits in months 18–24 and 25–30 is used for the benefits in each future period.

(2) “Welfare cost of taxes” indicates the additional cost in terms of lost output due to each additional dollar of taxes raised. The value 0.50 lies in the range suggested by Browning (1987).

(3) Estimates are constructed by breaking up the time after random assignment into 6-month periods. All costs are assumed to be paid in the first 6-month period, while benefits are received in each 6-month period and discounted by the amount indicated for each row of the table.

\(^2\) Both the impact and cost estimates are drawn from the analysis of Orr et al. (1995). The first row of Table 8.6 corresponds to the case they consider. The impact estimates of Orr
extend at least 5 years after random assignment for a subset of the experimental sample.

4 Evidence on Impact Heterogeneity and the Value of Self-assessments and Revealed-Preference Information

This section of the study addresses three questions. Question (1): What is the empirical evidence on heterogeneity in program impacts among persons? The conventional approach implicitly assumes impact homogeneity conditional on observables. This assumption greatly simplifies the task of evaluating the welfare state. Using data from an experimental evaluation of a prototypical job-training program, we use many of the assumptions presented in Section 2 to bound or identify the joint distribution of outcomes conditional on \( D = 1 \). We find considerable evidence of heterogeneity of program impacts, so that conventional econometric methods do not take us very far in constructing the evaluation criteria discussed in Section 1. Use of experimental data enables us to avoid the self-selection problems that plague ordinary observational data and simplifies our analysis.

Given our evidence on impact heterogeneity, we ask question (2): How sensitive are the estimates of the proportion of people who gain from the program – what we have called the “voting criterion” – to alternative assumptions about the dependence between \( Y^o \) and \( Y^i \)? We find that the estimates are very sensitive to alternative assumptions. At the same time, for adult women, the estimate of those who benefit from the program exceeds 50 percent in every case we consider except one, and it is close to 100 percent in some cases.

Some of the estimates used to answer question (2) assume that \( Y^o \) and \( Y^i \) are positively dependent given \( D = 1 \). We established in Section 2 that under purposive selection based on outcomes in the treated and untreated states, such dependence among participants arises even if \( Y^i \) and \( Y^o \) are independent or are negatively correlated in the population as a whole. An alternative to imposing a particular decision rule is to infer it from self-assessments of the program. These assessments are all that is required for a libertarian evaluation of the welfare state. We examine the implicit value placed on the program by addressing the following: Question (3a): Are persons who applied to the program and were accepted into it, but then randomized out of it, placed in an inferior position relative to those accepted applicants who were not randomized out? We measure ex-ante rational regret using second-order stochastic dominance, which is an appropriate measure under the assumption that individuals are completely uncertain of both \( Y^o \) and \( Y^i \) before going into the program. We also consider ex-post evaluations of participants by asking question (3b): How “satisfied” are participants with their experience in the program? Self-assessments of programs are widely used in evalu-
tion research (e.g., Katz et al., 1975), but the meaning to be placed on them is not clear. Do they reflect an evaluation of the experience of the program (its process) or an evaluation of the benefits of the program? Our evidence suggests that respondents report a net benefit inclusive of their costs of participating in the program. Groups for whom the program has a negative average impact, as estimated by the "objective" experimental data, express as much (or more) enthusiasm for the program as groups with positive average impacts. A third source of revealed-preference evaluations uses the revealed choices of attriters from the program. Econometric models of self-selection, since Heckman (1974a,b), have used revealed-choice behavior to infer the evaluations people place on programs either by selecting into them or dropping out of them. Finally, we come to question (3c): What implicit valuation of the program do attriters place on it?

4.1 Data

Our estimates are based on data from a recent experimental evaluation of the employment and training programs funded under title II-A of the U.S. Job Training Partnership Act (JTPA) (Orr et al., 1995). This program provides classroom training, on-the-job training, and job-search assistance to the economically disadvantaged. We focus primarily, but not exclusively, on adult women (age 22 and older) for many, but not all, of our analyses. We also present results for other demographic groups: adult men (age 22 and older) and male and female out-of-school youth (ages 16–21). Our largest samples are for adult women. Given that many of the adult women in the program are welfare recipients, their experiences with training are of special interest, given recent reforms in the U.S. welfare system. Appendix B describes the JTPA data in greater detail.

4.2 Evidence on Impact Heterogeneity

This subsection presents evidence on variability in the response to training. We find strong evidence against homogeneity. However, unless the dependence across outcomes in the treated and untreated states is very high, the estimated variability in program gains is implausibly large.

Suppose that the JTPA experiment satisfies (1-10). Suppose that there are \( N \) treated persons and \( N \) nontreated persons. Suppose that the outcomes are continuously distributed. Rank the individuals in each treatment category in the order of their outcome values from the highest to the lowest. Define \( Y_{ij} \) as the \( i \)th highest ranked person in the \( j \) distribution. Ignoring ties, we obtain two data distributions:

\[
Y^1 = \begin{pmatrix} Y^1_{(1)} \\ \vdots \\ Y^1_{(N)} \end{pmatrix} \quad Y^0 = \begin{pmatrix} Y^0_{(1)} \\ \vdots \\ Y^0_{(N)} \end{pmatrix}
\]

We know the marginal data distributions \( F(y^1|D = 1) \) and \( F(y^0|D = 1) \), but we do not know where person \( i \) in the treatment distribution would appear in the nontreatment distribution. These distributions can also be defined conditional on \( X \). Corresponding to the ranking of the treatment outcome distribution, there are \( N! \) possible patterns of outcomes in the associated nontreatment outcome distribution. By considering all possible permutations, we can form a collection of possible impact distributions, that is, alternative distributions of

\[
\Delta = Y^1 - \Pi \cdot Y^0 \quad (\ell = 1, \ldots, N!)
\]

where \( \Pi \) is a particular \( N \times N \) permutation matrix of \( Y^0 \) in the set of all \( N! \) permutations associating the ranks in the \( Y^1 \) distribution with the ranks in the \( Y^0 \) distribution, and \( \Delta \), \( Y^1 \), and \( Y^0 \) are \( N \times 1 \) vectors of impacts and treated and untreated outcomes. By considering all possible permutations, we obtain all possible sorts of treatment \( Y^1 \) and nontreatment \( Y^0 \) outcomes, using realized values from one distribution as counterfactuals for the other.

The dummy-endogenous-variable model assumes a constant treatment effect for all persons. This model admits only one permutation: \( \Pi = 1 \) for each \( X \). The best in one distribution is the best in the other distribution. In the common-effect case, \( Y^1 \) and \( Y^0 \) differ by a constant for each person. A generalization of that model preserves perfect dependence in the ranks between the two distributions, but does not require the impact to be the same at all quantiles of the base-state distribution.

In place of ranks, we work with the percentiles of the \( Y^1 \) and \( Y^0 \) distributions, which have much better statistical properties (Heckman and Smith, 1993; Heckman et al., 1997c). Equating percentiles across the two distributions, we form the pairs given in expression (13) and obtain the deterministic gain function given in (14). For the case of absolutely continuous distributions with positive density at \( y^0 \), the gain function (14) can be written as \( \Delta(y^0) = F_{Y^1}(y^0|D = 1) - y^0 \). We can test nonparametrically for the classic common-effect model by determining whether or not percentiles are uniformly shifted at all points of the distribution. We can form other pairings across percentiles by mapping percentiles from the \( Y^1 \) distribution into percentiles from the \( Y^0 \) distribution using the map \( T: q_1 \rightarrow q_0 \). The data are consistent with all admissible transformations, including \( q_0 = 100 - q_1 \), where the best in one distribution
is mapped into the worst in the other. They cannot reject any of these models nor more general models in which Π_1 is a Markov transition matrix and we consider all possible Markov matrices.

Figure 8.1 presents empirical evidence on the question of the constancy of the gain effect across quantiles. It shows the estimate of Δ(y₀) for adult women assuming that the best persons in the "1" distribution are the best in the "0" distribution. More formally, it assumes that the permutation matrix Π = I. No conditioning is made, so the full sample is utilized. Between the 25th and 85th percentiles the assumption of a constant impact is roughly correct. It is grossly at odds with the data at the highest and lowest percentiles. Heckman et al. (1997c) and Heckman and Smith (1993) have presented a more extensive empirical analysis of this model for different conditioning sets and have reached essentially the same conclusions.

#### 4.2.1 Fréchet Bounds

The Fréchet bounds of expression (17) can also be applied to conditional (on D = 1) distributions. Both the lower and the upper Fréchet bounds are proper probability distributions. At the upper bound, Y¹ is a non-decreasing function of Y₀. At the lower bound, Y₀ is a non-increasing function of Y¹. These bounds are not helpful in bounding the distribution of gains Δ = Y¹ − Y₀, although the bound certain features of it. From a theorem of Cambanis, Simons, and Stout (1976), if k(Y¹, Y₀) is superadditive (or subadditive), then extreme values of E[k(Y¹, Y₀)| D = 1] are obtained from the upper and lower bounding distributions obtained from the experimental data.

Because k(Y¹, Y₀) = Y¹Y₀ is superadditive, the maximum attainable product-moment correlation ρ_Y₁Y₀ is obtained from the upper-bound distribution, and the minimum attainable product-moment correlation is obtained at the lower-bound distribution. Because var(Δ) is a subadditive function, it is possible to bound the variance of Δ = var(Y¹) + var(Y₀) − 2ρ_Y₁Y₀[ρ_Y₁Y₀var(Y¹)var(Y₀)]^1/2 and thus determine whether or not the data are consistent with the common-effect model in which Y¹ − Y₀ = a, a constant, which implies var(Δ) = 0. Kendall's τ and Spearman's ρ also attain their extreme values at the bounding distributions (Tchen, 1980). However, the Fréchet inequalities do not provide bounds on the quantiles of the Δ = (Y¹ − Y₀) distribution. Only the extreme high and extreme low quantile values are obtained from the Fréchet bounds of the joint distribution. Table 8.7 presents the ranges of values of ρ_Y₁Y₀, Kendall’s τ, Spearman’s ρ, and [var(Δ)]^1/2 for the JTPA data for adult women. The ranges are rather wide, but it is interesting to observe that the Fréchet bounds rule out the common-effect model, as var(Δ) is bounded away from zero. They clearly do not rule out the deterministic case of perfect correlation in the ranks across outcome distributions as long as Δ is not a constant.

#### 4.2.2 Sensitivity to Alternative Assumptions about Dependence across the Distributions

Using the sample data, we can pair percentiles of the Y¹ and Y₀ distributions for any choice of rank correlation τ between −1.0 and 1.0.

---

27 A function k(x, y) is superadditive if x > x' and y > y' implies that k(x, y) + k(x', y') > k(x', y) + k(x, y'). Subadditively reverses the inequality. Strict forms of these ideas convert weak inequalities into strong ones.

28 Heckman et al. (1997c) conducted a Monte Carlo analysis of the standard errors of the standard deviation of Δ. They found that these standard errors did not provide a reliable guide to inferences regarding the null hypothesis that the true impact standard deviation is zero, using inference based on asymptotic normality of the test statistics. However, Monte Carlo estimation of the sampling distribution under the null that var(Δ) = 0 indicates that the null can be rejected in these data at the 0.0001 level.
Table 8.7. Characteristics of the Distribution of Impacts on Earnings in the 18 Months after Random Assignment at the Fréchet-Hoeffding Bounds (National JTPA Study, 18-Month Impact Sample, Adult Females)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact standard deviation</td>
<td>14,968.76</td>
<td>674.50</td>
</tr>
<tr>
<td></td>
<td>(211.08)</td>
<td>(137.53)</td>
</tr>
<tr>
<td>Outcome correlation</td>
<td>-0.760</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Spearman’s ρ</td>
<td>-0.9776</td>
<td>0.9867</td>
</tr>
<tr>
<td></td>
<td>(0.0016)</td>
<td>(0.0013)</td>
</tr>
</tbody>
</table>

Notes: (1) These estimates differ slightly from those reported for $r = 1.0$ and $r = -1.0$ in Table 8.8 because they were obtained using the empirical CDFs calculated at 100 earnings intervals rather than using the percentiles of the two CDFs. See Heckman, Smith, and Clements (1997) for details.

(2) Bootstrap standard errors in parentheses.

The case of $r = 1.0$ corresponds to the case of perfect positive dependence, where $\Pi = I$ and $q_1 = q_0$. The case where $r = -1.0$ corresponds to the case of perfect negative dependence, where $q_1 = 100 - q_0$. The first and last rows of Table 8.8 display estimates of quantiles of the impact distribution and other features of the impact distribution for these two cases.

Heckman et al. (1997c) show how to obtain random samples of permutations conditional on values of $r$ between 1.0 and -1.0. We display two sets of estimates from their work. The first set assumes positive but not perfect dependence between the percentiles of $Y^1$ and $Y^0$, with $r = 0.95$. Estimates based on a random sample of 50 percentile permutations with this value of $r$ appear in the second column of Table 8.8. These results show that even a modest departure from perfect positive dependence substantially widens the distribution of impacts. More striking still are the results in the third column of Table 8.8, which correspond to the case where $r = 0.0$. This value of $r$ is implied by independence between the percentiles of $Y^1$ and $Y^0$. Here (as in the case with $r = -1.0$) the distribution of estimated impacts is implausibly wide, with large positive values in each distribution often matched with zero or small positive values in the other. However, the conclusion that a majority

Table 8.8. Estimated Parameters of the Impact Distribution: Perfect Positive Dependence, Positive Dependence with $r = 0.95$, Independence, and Perfect Negative Dependence Cases (National JTPA Study, 18-Month Impact Sample, Adult Females)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Perfect positive dependence ($r = 1.0$)</th>
<th>Positive dependence with $r = 0.95$</th>
<th>Independence of $Y^1$ and $Y^0$ ($r = 0.0$)</th>
<th>Perfect negative dependence ($r = -1.0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5th percentile</td>
<td>0.00</td>
<td>0.00</td>
<td>-18,098.50</td>
<td>-22,350.00</td>
</tr>
<tr>
<td></td>
<td>(47.50)</td>
<td>(360.18)</td>
<td>(630.73)</td>
<td>(547.17)</td>
</tr>
<tr>
<td>25th percentile</td>
<td>572.00</td>
<td>125.50</td>
<td>-6,043.00</td>
<td>-11,755.00</td>
</tr>
<tr>
<td></td>
<td>(232.90)</td>
<td>(124.60)</td>
<td>(300.47)</td>
<td>(411.83)</td>
</tr>
<tr>
<td>50th percentile</td>
<td>864.00</td>
<td>616.00</td>
<td>0.00</td>
<td>580.00</td>
</tr>
<tr>
<td></td>
<td>(269.26)</td>
<td>(280.19)</td>
<td>(163.17)</td>
<td>(389.51)</td>
</tr>
<tr>
<td>75th percentile</td>
<td>966.00</td>
<td>867.00</td>
<td>7,388.50</td>
<td>12,791.00</td>
</tr>
<tr>
<td></td>
<td>(305.74)</td>
<td>(272.60)</td>
<td>(263.25)</td>
<td>(253.18)</td>
</tr>
<tr>
<td>95th percentile</td>
<td>2,003.00</td>
<td>1,415.50</td>
<td>19,413.25</td>
<td>23,351.00</td>
</tr>
<tr>
<td></td>
<td>(543.03)</td>
<td>(391.51)</td>
<td>(423.63)</td>
<td>(341.41)</td>
</tr>
<tr>
<td>Percentage positive</td>
<td>100.00</td>
<td>96.00</td>
<td>54.00</td>
<td>52.00</td>
</tr>
<tr>
<td></td>
<td>(1.60)</td>
<td>(3.88)</td>
<td>(1.11)</td>
<td>(0.81)</td>
</tr>
<tr>
<td>Impact S.D.</td>
<td>1,857.75</td>
<td>6,005.96</td>
<td>12,879.21</td>
<td>16,432.43</td>
</tr>
<tr>
<td></td>
<td>(480.17)</td>
<td>(776.14)</td>
<td>(259.24)</td>
<td>(265.88)</td>
</tr>
<tr>
<td>Outcome correlation</td>
<td>0.9903</td>
<td>0.7885</td>
<td>-0.0147</td>
<td>-0.6592</td>
</tr>
<tr>
<td></td>
<td>(0.0048)</td>
<td>(0.0402)</td>
<td>(0.0106)</td>
<td>(0.0184)</td>
</tr>
</tbody>
</table>

Notes: (1) The values in this table are calculated using percentiles of the two distributions. The perfect-positive-dependence case matches the top percentile in the $Y^1$ distribution with the top percentile in the $Y^0$ distribution, the second percentile of the $Y^1$ distribution with the second of the $Y^0$ distribution, and so on. The perfect-negative-dependence case matches the percentiles in reverse order, so that the lowest percentile of the $Y^1$ distribution is matched with the highest percentile of the $Y^0$ distribution, and so on. The two intermediate cases match the percentiles of the $Y^1$ distribution with percentiles of a permutation of the $Y^0$ distribution such that the rank correlation of the matched percentiles has the value indicated.

(2) The perfect-positive- and perfect-negative-dependence cases are based on the single permutation having this characteristic in the sample. The values reported for the intermediate cases represent means of random samples of 50 permutations with the indicated value of $r$.

(3) For each case, the difference between each percentile of the $Y^1$ distribution and the associated percentile of the $Y^0$ distribution is the impact for that percentile. Taken together, the percentile impacts form the distribution of impacts. It is the percentiles of these impact distributions that are reported in the upper portion of the table. The impact standard deviation, outcome correlation, and percentage positive are calculated using the percentile impacts. The impact standard deviation (S.D.) is the standard deviation of the percentile differences. The outcome correlation is the correlation of the matched percentile from the two distributions. The percentage positive is the percentile of the impacts greater than or equal to zero.

(4) Bootstrap standard errors in parentheses.
of adult female participants benefit from the program is robust to the choice of \( r \).\(^9\)

### 4.2.3 Assuming the Gain Is Independent of the Base

Another source of identifying information for the joint distribution of outcomes and the distribution of impacts postulates that the gain \( \Delta \) is independent of the base \( Y^0 \), so that \( Y^0 \perp \Delta D = 1 \). Letting \( R = 1 \) if a person who applies and is provisionally accepted into the program is randomized into the program, and \( R = 0 \) if a provisionally accepted applicant is randomized out, \( Y = Y^0 + R\Delta \), and \( R\Delta \perp Y^0 \). Throughout we condition on \( D = 1 \). This identifying condition will be satisfied if \( Y^0 \) is known, but the gain \( \Delta \) cannot be forecast at the time decisions are made about program participation. This case has been extensively discussed by Heckman and Robb (1985, p. 181), and it produces a model that is intermediate between the common-effect model and the variable-impact model when the impact is anticipated by agents.

Setting \( Y^0 = X\beta + U_0 \), we obtain a conventional random-coefficient model for a regression: \( Y = RY^0 + (1 - R)Y^0 = X\beta + R\Delta + U_0 \). Using a components-of-variance model, we can write \( E(\Delta) = \bar{\Delta} \) and \( \epsilon = \Delta - \bar{\Delta} \) to obtain

\[
Y = X\beta + R\bar{\Delta} + \epsilon R + U, \quad E(\epsilon R + UX, R) = 0
\]

Following the analysis presented in Section 2.1.1, we estimate the variance of \( \epsilon \).

The first row of Table 8.9 presents estimates of the random coefficient based on the identifying assumption \( \Delta \perp Y^0 D = 1 \). The evidence supports the hypothesis that \( \text{var}(\Delta) > 0 \), suggesting that a more elaborate approach to estimating the distribution of \( \Delta \) based on deconvolution is likely to be fruitful. If we maintain normality of \( Y^0 \) and \( Y^0 \) (given \( D = 1 \) and \( X \)), the distribution of \( \Delta \) is normal, with mean \( \bar{\Delta} \) and variance \( \text{var}(\Delta) \), and deconvolution is easy to perform. Under this assumption we can estimate the voting criterion and determine the estimated proportion of people who benefit from the program.

More generally, it is not necessary to assume that the distribution of \( \Delta \) is normal. We use the deconvolution procedure discussed by Heckman et al. (1997c) to estimate the distribution of impacts nonparametrically. Table 8.9 presents parameters calculated from this distribution. The evidence suggests that under this assumption, about 43 percent of adult women were harmed by participating in the program. The estimated density is presented in Figure 8.2 and is clearly nonnormal. Nonetheless, the estimated variance of the nonparametric gain distribution matches the variance for the gain distribution obtained from the random-coefficient model within the range of the sampling error of the two estimates. The estimates of the proportion who benefit are in close agreement across the two models when normality is imposed on the random-coefficient model. The fact that we obtain a positive density indicates that the assumption \( Y^0 \perp \Delta D = 1 \) is consistent with the data for women and provides support for the hypothesis that agents do not select into the program based on \( \Delta \).

### 4.3 Evidence from Participant Behavior

#### 4.3.1 Testing for Ex Ante Stochastic Rationality of Participants

If individuals choose whether or not to participate in the program based on the gross gains from the program, if they possess a common, but unknown, concave utility function, and if they know the marginal distributions of outcomes in the participation and nonparticipation states, then second-order stochastic dominance should order the distributions...
of outcomes for persons who seek to go into the program. For nonnegative $y^1$ and $y^0$ this form of rationality implies
\begin{equation}
\int_0^\alpha F_0(y^1|D = 1)dy^1 < \int_0^\alpha F_0(y^0|D = 1)dy^0 \quad \text{for all } \alpha \in R.
\end{equation}

Draws from the $Y^1$ distribution produce higher expected utility than draws from the $Y^0$ distribution among participants. The difference between the two integrals is a measure of regret among persons randomized out from the program and forced into the no-treatment state. This condition may fail for many reasons: Persons may possess more information about their potential outcomes than just the marginal distributions; persons may have different utility functions; and persons may participate in the program on a principle other than expected utility formulated in terms of gross outcomes.

We test condition (25) by comparing the integrals of the empirical CDFs of the control-group and treatment-group earnings distributions for various values of $\alpha$. Table 8.10 displays the results of tests of the null hypothesis of equality of the integrated distributions in (25) for adult males and females and for male and female youth using self-reported earnings in the 18 months after random assignment. The table shows test results for $\alpha \in \{2,500, 5,000, 10,000, 15,000, 20,000, 25,000\}$. Standard errors are obtained by bootstrapping. For adult males, the integrated CDF of earnings for the control group exceeds that for the treatment group at every point, with a $p$ value below 0.05 for $\alpha < 16,500$, and below 0.10 for $\alpha < 22,500$, which includes most of the supports of the two earnings distributions. The data for adult females provide strong evidence of rational behavior in the sense of (25), passing the test at the 5 percent level or better for every value of $\alpha$. This evidence suggests that personal objectives and program objectives are aligned for adult women. Results for youth are mixed. For male youth, for whom the mean experimental impact is significantly negative, the difference in integrated CDFs is negative for most values of $\alpha$, though not statistically significant. For female youth, the difference switches signs at around $\alpha = 11,000$, but is never close to statistical significance.
4.3.2 Evidence from Self-assessments of Program Participants

Self-assessments of program participants represent an alternative to comparison of observed outcomes as a measure of program impact. Unlike the ex-ante measures based on second-order stochastic dominance, these measures are statements about ex-post expectations. There is no reason that the two measures should agree if people revise their assessments based on what they learn about a program by participating in it. In this section we consider the strengths and limitations of self-reported assessments of satisfaction with the program as evaluation criteria and report on self-evaluations by participants in the JTPA experimental-treatment group. We also consider what can be learned from self-assessment data regarding the heterogeneity of individual treatment effects and the rationality of program participants.

Using participant assessments to evaluate a program has two main advantages relative to the approaches already discussed. First, participants have information not available to external program evaluators. They typically know more about certain components of the cost of program participation than do evaluators. Most evaluations, including the National JTPA Study, do not even attempt to value participant time, transportation, child care, or other costs in evaluating program effectiveness, unless they are paid by the program through subsidies. Participants are likely to include such information in arriving at their self-assessments of the program. Second, participant evaluations provide information about the values placed on outcomes by participants relative to their perceived costs. They have the potential of providing a more inclusive measure of the program’s effects than would be obtained from looking only at gross outcomes—ones that include “client satisfaction.” To some parties in the welfare state, “customer satisfaction” is an important aspect of a program.

However, participant self-assessments may not be informative on the outcomes of interest to other parties in the welfare state. In evaluations of medical interventions, for example, treatment effects may not be observed by participants or may be difficult to assess, compared with what observing scientists might report. Participant assessments of the counterfactual state may be faulty because their judgments are based on inputs or on outcome levels rather than on gains over alternative levels. Persons who choose to go into the program may rationalize their participation in it in responding to questions. In addition, self-assessments, like all utility-based measures, are difficult to compare across individuals.

The top panel of Table 8.11 reports JTPA participant responses to a question about whether or not the program made them better off.30

30 The exact wording was “Do you think that the training or other assistance you got from the program helped you get a job or perform better on the job?” The question was asked only of treatment-group members who reported receiving JTPA services.

<table>
<thead>
<tr>
<th></th>
<th>Adult males</th>
<th>Adult females</th>
<th>Male youth</th>
<th>Female youth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage who self-report participating</td>
<td>61.63</td>
<td>68.10</td>
<td>62.62</td>
<td>66.29</td>
</tr>
<tr>
<td>(0.81)</td>
<td>(0.68)</td>
<td>(1.29)</td>
<td>(1.09)</td>
<td></td>
</tr>
<tr>
<td>Percentage of self-reported participants with a positive self-assessment</td>
<td>62.46</td>
<td>65.21</td>
<td>67.16</td>
<td>71.73</td>
</tr>
<tr>
<td>(1.04)</td>
<td>(0.85)</td>
<td>(1.59)</td>
<td>(1.29)</td>
<td></td>
</tr>
<tr>
<td>Overall percentage with positive self-assessments</td>
<td>38.49</td>
<td>44.41</td>
<td>42.06</td>
<td>47.55</td>
</tr>
<tr>
<td>(0.81)</td>
<td>(0.73)</td>
<td>(1.32)</td>
<td>(1.16)</td>
<td></td>
</tr>
<tr>
<td>Percentage of self-reported participants with a positive self-assessment by primary treatment received</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None (dropouts)</td>
<td>48.89</td>
<td>51.44</td>
<td>58.90</td>
<td>61.56</td>
</tr>
<tr>
<td>(2.07)</td>
<td>(1.85)</td>
<td>(3.33)</td>
<td>(2.79)</td>
<td></td>
</tr>
<tr>
<td>Classroom training in occupational skills</td>
<td>74.10</td>
<td>73.47</td>
<td>72.73</td>
<td>75.28</td>
</tr>
<tr>
<td>(2.15)</td>
<td>(1.36)</td>
<td>(3.60)</td>
<td>(2.30)</td>
<td></td>
</tr>
<tr>
<td>On-the-job training at private firm</td>
<td>75.13</td>
<td>78.90</td>
<td>71.00</td>
<td>75.00</td>
</tr>
<tr>
<td>(2.18)</td>
<td>(2.14)</td>
<td>(4.56)</td>
<td>(4.04)</td>
<td></td>
</tr>
<tr>
<td>Job-search assistance</td>
<td>59.27</td>
<td>59.80</td>
<td>68.09</td>
<td>68.94</td>
</tr>
<tr>
<td>(2.27)</td>
<td>(2.18)</td>
<td>(3.94)</td>
<td>(4.04)</td>
<td></td>
</tr>
<tr>
<td>Basic education</td>
<td>62.96</td>
<td>56.55</td>
<td>70.97</td>
<td>78.44</td>
</tr>
<tr>
<td>(4.67)</td>
<td>(3.84)</td>
<td>(4.09)</td>
<td>(3.19)</td>
<td></td>
</tr>
<tr>
<td>Work experience</td>
<td>66.67</td>
<td>68.75</td>
<td>82.76</td>
<td>73.17</td>
</tr>
<tr>
<td>(9.83)</td>
<td>(5.84)</td>
<td>(7.14)</td>
<td>(7.01)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>58.47</td>
<td>66.40</td>
<td>62.50</td>
<td>77.98</td>
</tr>
<tr>
<td>(3.65)</td>
<td>(2.98)</td>
<td>(4.77)</td>
<td>(3.99)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: (1) Reported proportions are based on responses to the question “Do you think that the training or other assistance you got from the program helped you get a job or perform better on the job?” This question was asked only of self-reported participants within the treatment group. The overall fraction of positive self-assessments assumes that self-reported nonparticipants would have provided negative self-assessments.

(2) The primary treatment is the one in which the trainee participated for the most hours according to the administrative records of the JTPA sites. Most trainees received only one service; few received more than two. See Smith (in press) for a detailed discussion. Note that for some self-reported participants the JTPA administrative records indicate that no services were received.

(3) Estimated standard errors in parentheses.

Assuming that people answered honestly, and were reporting a gross impact, the self-assessment data clearly contradict the hypothesis of impact homogeneity. For all four demographic groups, 65 to 70 percent of self-reported participants gave positive self-assessments, not the 100 percent or zero percent that would be predicted if impacts were homo-
is all the more striking when it is realized that self-assessments were recorded only for people who reported receiving training, whereas the gross outcome data for participants included those who left the program, and the attritors had lower earnings than the nonattritors.

4.3.3 Evidence from Program Dropouts

As a result of the relatively early implementation of random assignment in the JTPA participation process, many treatment-group members never enrolled in the JTPA program and so did not receive JTPA services.31 In this section we investigate what the information on dropout behavior reveals about treatment heterogeneity and participant rationality. A key limitation in doing this is that the enrollment decision did not depend simply on agent choices, but was a joint decision of the potential participant and of JTPA staff members. The JTPA performance-standards system, which rewards individual training centers based on the labor-market outcomes for their enrollees, provides both a mechanism and incentive for manipulation of the enrollment decision in order to increase center performance.32 Because we have no data on the preferences of bureaucrats, we ignore this problem and assume that the decisions we observe were solely those of the potential participants.

If anticipated discounted net impacts are the same across all persons, then everyone either participates in the program or drops out of it. The substantial dropout rates reported in the first column of Table 8.12 for all four demographic groups provide evidence that anticipated discounted impacts were heterogeneous.

Next consider the implications of these data for participant rationality. Assume a common discount rate and constant returns per period from the program. Suppose that persons apply and are accepted into the program if $E(\Delta I) > 0$, where $I$ is the information available at application. Suppose further that $\Delta$ is revealed at the time of acceptance into the program, that persons drop out whenever $\Delta \leq 0$, and that $\Delta \perp Y^0$ [this is identifying assumption (I-5)]. If persons entering the program cannot forecast $\Delta$, then, letting $e = 1$ if a person enrolls in the program and $e = 0$ if the person drops out, $E(Y^0|e = 1, D = 1) = E(Y^0|e = 0, D = 1)$, and $E(Y^0|e = 1, D = 1) = E(Y^0|e = 0, D = 1) + E(\Delta|A > 0, D = 1)$.

Table 8.12 presents the mean earnings of JTPA enrollees and dropouts for the 18 months after random assignment for the four demographic groups, along with the experimental impact estimates and the implied differences in $Y^0$ between the two groups. For adult females and for female youth, the data are consistent with this model, because the dif-

---

31 Heckman and Smith (1993, 1995) discussed this phenomenon.
32 This was discussed by Heckman, Smith, and Taber (1998b).
Table 8.12. Comparisons of Post-Random-Assignment Earnings of Treatment Group Enrollees and Dropouts (National JTPA Study, 18-Month Impact Sample)

<table>
<thead>
<tr>
<th>Percentage dropping out</th>
<th>Mean earnings of enrollees</th>
<th>Mean earnings of dropouts</th>
<th>Mean earnings difference</th>
<th>Experimental impact estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult males</td>
<td>40.92</td>
<td>12,638</td>
<td>12,181</td>
<td>0.0010</td>
</tr>
<tr>
<td>Adult females</td>
<td>37.39</td>
<td>(261)</td>
<td>(355)</td>
<td>(0.0975)</td>
</tr>
<tr>
<td>Male youth</td>
<td>34.83</td>
<td>10,275</td>
<td>9,442</td>
<td>0.1231</td>
</tr>
<tr>
<td>Female youth</td>
<td>37.02</td>
<td>6,108</td>
<td>6,251</td>
<td>0.6844</td>
</tr>
</tbody>
</table>

Notes: (1) The p values are from t tests of equality of means assuming unequal variances in the two groups.
(2) Mean earnings difference indicates the difference in mean earnings between the enrollees and the dropouts in the experimental treatment group.
(3) Experimental impact estimates differ from those of Bloom et al. (1993) because they are not regression-adjusted and because imputed values for adult female nonrespondents were not used.
(4) Estimated standard errors in parentheses.

The difference between the mean earnings of enrollees and dropouts is not statistically distinguishable from the experimental impact estimate. However, the data for adult males and male youth are not consistent with this model.

If, however, we relax the assumption of independence between $\Delta$ and $Y^e$, we can rationalize the male data. Suppose that $\Delta = \Delta(Y^e)$. If $\Delta(Y^e)$ is an increasing function of $Y^e$, this implies that

$$E(Y^t|e = 1, D = 1) - E(\Delta|D > 0, D = 1)$$

which is consistent with the patterns in Table 8.12 for adult males and for male youth.

Another model assumes that the true treatment effect is revealed after random assignment and the net response varies over time. In this case, a person who values only the impacts from the program will remain in it if

$$\sum_{t=1}^{T} \delta^{t-1} \Delta_t > 0$$

Table 8.13. Six-Month Interest Rates and Discount Factors That Equalize the Discounted Present Value of Mean Earnings of Enrollees and Dropouts in the CT-OS Treatment Stream under Two Assumptions about Benefit Duration (National JTPA Study, 18-Month Impact Sample)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Adult males</th>
<th>Adult females</th>
<th>Male youth</th>
<th>Female youth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benefits fall to zero after 18 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equalizing $r$</td>
<td>-0.0627</td>
<td>-0.7720</td>
<td>0.2082</td>
<td>-0.2677</td>
</tr>
<tr>
<td>Equalizing $\delta$</td>
<td>1.0669</td>
<td>4.3858</td>
<td>0.8277</td>
<td>1.3656</td>
</tr>
<tr>
<td>Benefits continue for 7 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equalizing $r$</td>
<td>1.1250</td>
<td>0.8720</td>
<td>1.4520</td>
<td>0.9580</td>
</tr>
<tr>
<td>Equalizing $\delta$</td>
<td>0.4706</td>
<td>0.5342</td>
<td>0.4078</td>
<td>0.5107</td>
</tr>
</tbody>
</table>

Notes: (1) The CT-OS treatment stream refers to persons recommended for classroom training in occupational skills prior to random assignment in the JTPA experiment. This group comprises roughly one-third of the full experimental sample.
(2) Negative values for the interest rate indicate that positive future preference is required to rationalize choosing the enrollee earnings stream over the dropout earnings stream.
(3) Earnings streams are broken up into 6-month pieces in calculating the discount rates.

where $\Delta_t$ is the impact in the $t$th period after random assignment, and $\delta$ is a discount rate. If the inequality is reversed, or becomes an equality, the person drops out.

The implications of this model depend on the temporal pattern of the $\Delta_t$ values. For example, in classroom training, where the trainee forgoes earnings initially in order to invest in human capital, we would expect $\Delta_t < 0$ for $t' < t$, and $\Delta_t > 0$ for $t' > t$, where $t' \leq t$ are periods of human-capital accumulation. In the case of a constant $\Delta_t$, there would be perfect sorting by discount rate into the dropout and enrollee categories. Persons with low $\delta$ values would drop out, and those with high $\delta$ values would complete the training.

We calculate the interest rate $r$ [where $\delta = 1/(1 + r)$] required to equate the discounted present value of mean earnings in the dropout and enrollment states for persons in the classroom-training treatment stream under two sets of assumptions about the time pattern of impacts more than 18 months after random assignment. The top panel of Table 8.13 shows the interest rate necessary to equalize the present value of dropout and enrollee earnings if the impact falls to zero after 18 months. That these estimated rates are sometimes negative reflects the fact that the returns to training for some groups are insufficient to balance out the earnings lost in the initial period unless there is negative discount rate. The lower panel of Table 8.13 shows the interest rate necessary to
equate the present value of dropout and enrollee earnings under the assumption that the impact in the final 6-month period persists through 7 years.

Potential trainees exhibit high rates of time preference. Discount rates of this magnitude have been reported by Thaler (1992). Such high rates are consistent with the view that the poor, who are the primary targets of the JTPA program, are poor because they discount the future heavily.

4.4 Summary of the Evidence on Impact Heterogeneity and Its Consequences

Table 8.14 presents a summary of the main findings of this section: (1) Under a variety of assumptions, we find evidence of heterogeneity in net impacts $\Delta$. (2) The analysis of self-assessments suggests that respondents were reporting impacts different from the "objective" impacts determined from experimental data. This is a further source of heterogeneity and a source of disparity across studies. (3) Departures from high levels of positive dependence between $Y^p$ and $Y^o$ produced absurd ranges of impacts on gross outcomes. (The implicit correlations between $Y^p$ and $Y^o$ produced under different identifying assumptions are given in the last column of the table.) (4) The ranges of the estimated proportions of people benefiting from the program, in the sense of gross outcomes (the "voting criterion"), varied widely under different assumptions about the dependence in outcomes. The data from the self-report and attrition studies show a lower proportion benefiting—a phenomenon consistent with the hypothesis that net returns (not gross returns) were being reported by participants.

5 Summary

In his Nobel lecture, Ragnar Frisch (1970) recognized the diversity of preferences regarding the outcomes of public policies that characterizes participants in welfare states. This diversity in values gives rise to a multiplicity of criteria for evaluating policies. This study has considered these criteria and presented a formal analysis of the information required to evaluate public policies under different criteria. We have presented the approximations required to go from microeconomic evaluations to conclusions about the general-equilibrium outcomes from alternative policies. We have described conditions under which conventional econometric analyses of "treatment effects" provide part of the information required to conduct general-equilibrium cost-benefit analyses. We note that personal evaluations of policies may not coincide with the evaluations useful in the political arena of the welfare state, and we have presented methods to reveal private or "subjective" evaluations to supplement and complement the "objective" evaluations.
<table>
<thead>
<tr>
<th>Description of analysis</th>
<th>Evidence of heterogeneity?</th>
<th>Standard deviation of impacts</th>
<th>Evidence on voting criterion</th>
<th>Dependence between $Y^c$ and $Y^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive percentile dependence with rank correlation $\tau = 0.95$</td>
<td>Assumes that the percentiles of $Y^c$ and $Y^d$ given $D = 1$ have a rank correlation of 0.95; estimates are based on a random sample of 50 such permutations</td>
<td>Yes; the average minimum is $-14,504$, and the average maximum is $48,544$</td>
<td>Average standard deviation of $1,857$ (with standard deviation of $480$)</td>
<td>Average of 93 percent positive (with standard deviation of 3.88)</td>
</tr>
<tr>
<td>Independence of percentiles of $Y^c$ and $Y^d$, which implies a percentile rank correlation $\tau$ of 0.0</td>
<td>Assumes that the percentiles of $Y^c$ and $Y^d$ given $D = 1$ have a rank correlation $\tau$ of 0.0, which is implied by independence between them; estimates are based on a random sample of 50 such permutations</td>
<td>Yes; the average minimum is $-44,175$, and the average maximum is $60,599$</td>
<td>Average standard deviation of $12,879$ (with standard deviation of $259$)</td>
<td>Average of 54 percent positive (with standard deviation of 1.11)</td>
</tr>
<tr>
<td>Random-coefficient model</td>
<td>Assumes that $\Delta \neq Y_0</td>
<td>D = 1$</td>
<td>Yes; see Figure 8.2</td>
<td>Standard deviation is $2,271$</td>
</tr>
</tbody>
</table>

$^a$ A function $k(x,y)$ is superadditive if $x > x'$ and $y > y'$ implies that $k(x,y) + k(x',y') > k(x,y') + k(x',y)$. Subadditivity reverses the inequality.

$^b$ Results are for adult women only. Similar results are obtained for adult men and for male and female youth.

$^c$ The standard deviation is calculated over the random sample of 50 permutations with the indicated value of $\tau$.

$^d$ Results are for adult women only. For the remaining demographic groups, $\text{var}(Y^c) < \text{var}(Y^d)$, which indicates that neither the random-coefficient model nor deconvolution is appropriate.

$^e$ N.A., not applicable.

The product-moment correlation $\rho = 0.9774$. The product-moment correlation $\rho = 0.9595$. The product-moment correlation $\rho = 0.0147$ (with standard deviation of 0.0106); Kendall's rank correlation $\tau$ fixed at 0.0; both are calculated using the percentiles of the two distributions.

The product-moment correlation $\rho = 0.0147$ (with standard deviation of 0.0106); Kendall's rank correlation $\tau$ fixed at 0.0; both are calculated using the percentiles of the two distributions.

The product-moment correlation $\rho = 0.9774$. The product-moment correlation $\rho = 0.9595$. The product-moment correlation $\rho = 0.0147$ (with standard deviation of 0.0106); Kendall's rank correlation $\tau$ fixed at 0.0; both are calculated using the percentiles of the two distributions.
Implementation of many of the criteria used to evaluate the welfare state requires information on the joint distribution of outcomes across policies. Traditional cost–benefit analysis avoids this problem by assuming that a background social-welfare function automatically solves all of the distributional problems of the welfare state. In this case, which is assumed in much of the microeconometric evaluation literature, simple per-capita measures of economic efficiency based on the change in aggregate output attributable to a policy suffice to evaluate the welfare state. However, even in this case we note that estimators widely used in the econometric evaluation literature do not provide the ingredients required for a comprehensive cost–benefit analysis. In an empirical analysis, we have demonstrated that when conventional estimators are modified to account for direct costs and the welfare costs of taxation, they produce inferences about program impacts very different from those produced using standard econometric methods. We have described conditions under which standard econometric estimators provide reliable answers to well-posed general-equilibrium evaluation questions.

Homogeneity in the response to a policy across persons with the same observed characteristics is the central implicit identifying assumption that underlies most of the widely used econometric policy-evaluation methods. The assumption of response homogeneity greatly simplifies the evaluation problem. Part of the conflict in the estimates produced from different evaluation criteria arises from heterogeneity in the impacts of a given program across persons. We have presented evidence from a major social experiment that heterogeneity in the responses to treatment is an empirically important phenomenon.

An evaluation strategy that properly accounts for individual heterogeneity requires more information than traditional econometric evaluation methods. We have demonstrated how information about participant self-selection choices and program-participation rules aids in identifying the distributions of outcomes across policies and also provides information on personal valuations of program outcomes. We have discussed how social experiments and different types of microeconomic data can be used to identify the criteria considered in this study and how they can be supplemented with additional behavioral and statistical assumptions to construct all of the criteria. Unless special individual decision rules characterize program participation, these sources of data do not resolve the fundamental evaluation problem that persons cannot occupy mutually exclusive outcome states at the same time.

We have applied some of the methods developed in this study to data from a major job-training program. For adult females, we conclude that the program benefited most participants according to the “objective” evaluation criteria based on gross outcomes, but did not benefit a majority of participants according to self-assessments or the revealed-preference behavior of attritors from the program. The disagreement among the alternative criteria highlights the need for providing information about all of them to satisfy the different parties in the welfare state.

Appendix A

Let outcomes $Y^i$ and $Y^0$ be written as functions of observed variables $X$ and unobserved variables $U_i$ and $U_0$, respectively:

$$Y^i = g_i(X_i, X_c) + U_i \quad (A-1a)$$
$$Y^0 = g_0(X_0, X_c) + U_0 \quad (A-1b)$$

where $X_i$ (a $k_i$-dimensional vector) and $X_0$ (a $k_0$-dimensional vector) are variables unique to $g_i$ and $g_0$, respectively, and $X_c$ (a $k_c$-dimensional vector) includes variables common to the two functions. The variables $U_0$ and $U_i$ are unobserved from the point of view of the econometrician.

To simplify the analysis, we write program-specific subscripts.

The decision rule for program participation is given by

$$I = g_i(X_i, X_c) + U_i \quad \text{and} \quad D = I(I \geq 0) \quad (A-1c)$$

where $X_i$ (a $k_i$-dimensional vector) consists of measured variables, some of which may appear in $X_i$ and $X_0$, and where $U_i$ is unobserved by the econometrician. $I$ is a latent index or net utility of the relevant decision-maker. The joint distribution of $(U_0, U_i, D)$ is denoted by $F(u_0, u_i, u_D)$. These variables are statistically independent of $(X_0, X_i, X_c)$.

The Roy model is a special case of this framework in which selection into the program depends only on the gain from the program. In this case,

$$g_i = g_i(X_i, X_c) - g_0(X_0, X_c) \quad \text{and} \quad U_i = U_i - U_0$$

For convenience, we write $F(u_0, u_i)$ as the joint distribution of $(U_0, U_i)$.

The following theorem can be proved for the Roy model. It extends and clarifies a theorem presented by Heckman and Honoré (1990).

**Theorem A.1.** Let $Y^i = g_i(X_i, X_c) + U_i$ and $Y^0 = g_0(X_0, X_c) + U_0$.

Assume the following:

(i) $(U_0, U_i) \perp (X_0, X_i, X_c)$

(ii) $D = 1(Y^i \geq Y^0)$

(iii) $(U_0, U_i)$ is absolutely continuous, with Support$(U_0, U_i) = R_0 \times R_1$

(iv) For each fixed $X_c$,

$g_0(X_0, X_c) : R_0 \rightarrow R_1$ for all $X_0$

$g_i(X_i, X_c) : R_i \rightarrow R_1$ for all $X_i$
and
\[
\text{Support}(g_0(X_0, X_1) | X_1, X_0) = R_1 \quad \text{for all } X_1, X_0
\]
\[
\text{Support}(g_1(X_1, X_0) | X_1, X_0) = R_1 \quad \text{for all } X_1, X_0
\]
\[
\text{Support}(X_0 | X_1, X_0) = \text{Support}(X_0) = R_1 \quad \text{for all } X_1, X_0
\]
\[
\text{Support}(X_1 | X_0, X_1) = \text{Support}(X_1) = R_1 \quad \text{for all } X_0, X_1
\]

(v) The marginal distributions of $U_0$ and $U_1$ have zero medians.

Then $g_0$, $g_1$, and $F(u_0, u_1)$ are nonparametrically identified from data on participation choices and outcomes.

**Proof.** By assumption, we know that for all $(X_0, X_1, X_c)$ in the support of $(X_0, X_1, X_c)$ and for all $y$,

\[
\Pr(Y^1 \leq Y^0 | X_0, X_1, X_c) = \Pr[g_1(X_1, X_c) + U_1 \leq g_0(X_0, X_c) + U_0] \tag{A}
\]

\[
\Pr(Y^1 = y, Y^1 > Y^0 | X_0, X_1, X_c) = \Pr[g_1(X_1, X_c) + U_1 = y, g_1(X_1, X_c) + U_1 > g_0(X_0, X_c) + U_0] \tag{B}
\]

\[
\Pr(Y^0 \leq y, Y^0 > Y^1) = \Pr[g_0(X_0, X_c) + U_0 \leq y, g_0(X_0, X_c) + U_0 > g_1(X_1, X_c)
+ U_1] \tag{C}
\]

Fix $X_1$. Let $\bar{X}_0$ and $\bar{X}_0$ be in the support of $X_0$ and $X_0$, respectively. Using the information in (A), we can define sets of values $(X_0, X_1)$ corresponding to contours of constant probability:

\[
S(X_0, X_1 | S, \bar{p}) = \{(X_0, X_1) : \Pr[g_1(X_1, X_c) + U_1 > g_0(X_0, X_c) + U_0] = \Pr[g_1(X_1, X_c) + U_1 > g_0(X_0, X_c) + U_0] = \bar{p}\} \tag{A-2}
\]

\[
= \{(X_0, X_1) : g_1(X_1, X_c) + \ell = g_0(X_0, X_c)\}
\]

for some unknown constant $\ell$.

For any point in $S$ we can use the information in (B) to write

\[
\Pr[g_1(X_1, X_c) + U_1 \leq y, U_1 > U_0 + \ell] \tag{A-3}
\]

for all $y$. Varying $X_1$ over its full support from assumption (iv), we can find a compensating value $X_0$ within the set defined by (A-2) so that $\Pr(D = 1 | X_0, X_1) = \bar{p}$ is constant. This keeps fixed the second argument in (A-2). The variation in $X_1$ produces a set of $(y, X_1)$ values for each value of $X_1$ that identifies the function $g_1(X_1, X_c)$ over the support of $X_1$ up to an unknown constant. By similar reasoning, we can identify $g_0(X_0, X_c)$ up to an unknown constant using (C) over the support of $X_0$.

Tracing out (A-3) for all values $g_1$ and $y$ identifies $F(u_1, u_0 - u_1)$, except for a location parameter. Using (C), we identify $F(u_0, u_1 - u_0)$.

The content of this theorem is that if there is sufficient variation in $X_1, X_0$, and $X_c$, and if we know that program participation is based solely on outcome maximization, no arbitrary parametric structure on the outcome equations or on the distribution of the unobservables generating outcomes needs to be imposed to recover the full distribution of outcomes using ordinary microeconomic data.

Note that we can obtain $F(y^D | D = 1, X)$ from $F(y^D | D = 0, X)$, where $X$ denotes the full set of conditioning variables. Using the law of iterated expectations,

\[
F(y^D | X) = F(y^D | D = 1 | X) \Pr(D = 1 | X)
+ F(y^D | D = 0 | X) \Pr(D = 0 | X)
\]

From Theorem A-1 we can recover $F(y^D | X)$. Because we know $\Pr(D = 1 | X) = 1 - \Pr(D = 0 | X)$, we can recover $F(y^D | D = 1)$.

The assumptions made in Theorem A-1 about the supports of $X_1, X_0$, $g_1, g_0, U_1$, and $U_0$ are made for convenience, in an effort to focus on main ideas. Nowhere is it literally required that any function or variable "go to infinity" as some authors have claimed (e.g., Imbens and Angrist, 1994). A version of Theorem A-1 can easily be proved under the following alternative conditions:

\[
\text{Support}(U_0) = \text{Support}(U_1) = \text{Support}(g_0) = \text{Support}(g_1)
\]

where all of the supports are finite. Under these conditions, and assuming that all of the other conditions hold, it is possible to retrace the argument of Theorem A-1 and produce essentially the same theorem, provided that $\text{Support}(X_0 | X_1, X_0) = \text{Support}(X_0) \text{ and Support}(X_1 | X_0, X_c) = \text{Support}(X_1)$.

In more general cases where the supports of $g_0, g_1, U_1$, and $U_0$ do not coincide, or where there are restrictions on the supports of $X_1$ and $X_0$, a modified version of the theorem can be proved. It may be possible to construct a set of $(X_1, X_0)$ values that satisfy a condition like (A-2), except now it is no longer necessarily true that we can vary $X_1$ over its full support within any isoprobability set [the set of values of $(X_1, X_0)$ that set $\Pr(D = 1 | X_0, X_1) = \bar{p}$ for any $\bar{p}$]. That is, for each $\bar{p}$, we are no longer guaranteed to be able to find a compensatory value of $X_0$ to
ensure that for each \( X_i \) we can keep the probability fixed. Suppose \( g_1 \leq g_i \leq \bar{g}_i \) and \( g_0 \leq g_0 \leq \bar{g}_0 \). For each \( X_i \) and \( \bar{p} \), the support of \( g_1 - g_0 \) is \( (\bar{g}_0 - g_0, \bar{g}_1 - g_1) \), provided that \( \text{Support}(X_0, X_i) = \text{Support}(X_0) \) and \( \text{Support}(X_0, X_i) = \text{Support}(X_i) \). Only for subsets of the support of \( X_0 \) and \( X_i \) can the argument below (A-2) in the proof be invoked. Because \( g_0(X_0, X_i) \) and \( g_1(X_0, X_i) \) are not necessarily identified over their full support, it follows that we are not necessarily guaranteed to be able to identify \( F_{u_t}(u_t) \) over its full support. Moreover, in general, we shall be able to identify the \( g_i \) and \( g_0 \) functions only up to unknown scale parameters. With these qualifications about the support, the conclusions of Theorem A-1, restated to include the restrictions on supports, remain intact.

The Roy model has an unusual structure, because the participation rule and the outcome equations are tightly linked. As a consequence, we can recover the full joint distribution of \( F(y^1, y^0 | X) \) and the decision rule knowing only conditional distributions (B) and (C) available from cross-section data. For more general decision rules, such as (A-1c), that break the tight link between outcomes and participation decisions, it is not possible to use (B) and (C) to address those questions that can be answered only from the full joint distribution of \( (y^1, y^0) \). Even access to the data obtained from social experiments - \( F(y^1 | D = 1, X) \) - does not suffice to solve the fundamental evaluation problem that both \( y^1 \) and \( y^0 \) are never observed for a single person. However, a theorem analogous to Theorem A-1 can be proved that demonstrates that with sufficient variation in the \( X \) variables, it is possible to recover \( F(y^1 | D = 1, X) \) from nonexperimental data. Before presenting a more general version of Theorem A-1, it will be useful to review some recent results on the estimation of nonparametric and semiparametric discrete-choice models that are required in the proof of the theorem.

We consider the nonparametric identification of decision rule (A-1c) under the assumption that

\[
(X_1, X_e) \parallel U_t \tag{A-4}
\]

The original proof is due to Coslett (1983), who assumed \( g_i = (X_i, X_e) \beta_i \). Matzkin (1990, 1992) considered the more general case that will be used here.\(^{32}\) In the Roy model, (A-1c) is tightly linked to (A-1a) and (A-1b), and we observe \( y^1 \) and \( y^0 \) in censored samples. In the general case, nonparametric identification of \( g_i \) requires a separate argument.

From inspection of

\[
\Pr(D = 1 | X_i, X_e) = \Pr(g_i(X_i, X_e) + U_t \geq 0) = 1 - F_{u_t}(-g_i)
\]

it is clear that without further restrictions on the set of candidate \( g_i \) functions, it will be impossible to identify a unique member of the set. For any alternative distribution function \( F^* \), we can define \( g_i^* \) so that

\[
F_{u_t}(-g_i) = F_{u_t}(-g_i^*)
\]

with the result that \( (g_i, F_t) \) cannot be distinguished from \( (g_i^*, F_t^*) \).

Let \( G \) be the set of admissible functions. Matzkin (1990) showed that under the independence assumption, if \( \bar{g}_i \) is a least-concave representation of \( g_i \in G \), then \( (\bar{g}_i, F_t) \) is identified, where \( F_t \) is the associated distribution function.\(^{34}\) Because concavity naturally arises in many economic settings of consumer and producer choice, her assumption is an attractive one. We record Matzkin's basic assumptions, in addition to (A-4):

- (M-1) \( g_i \) is concave
- (M-2) \( \text{Support}[g_i(X_0, X_e)] \supset \text{Support}(U_t) \)

**Theorem M-1 (Matzkin 1).** Under (M-1) and (M-2), \( \bar{g}_i \) and the associated \( F_i \) are identified subject to a scale normalization for \( g_i \).

**Proof.** See Matzkin (1990). \( \square \)

Matzkin (1992) also considered an alternative identifying assumption that can substitute for (M-1) and (M-2).

- (M-3) There exists a subset \( T \) of the support of \( X = (X_i, X_e) \) such that
  - (i) for all \( g_i, g_j \in G \), and all \( X \in T, g_i(X) = g_j(X) \), and
  - (ii) for all \( t \) in the support of \( U_t \), there exists \( X \in T \) such that \( g_i(X) = t \).

In the estimation of production functions there is a natural set of values \( X = 0 \) where no input produces no output. Similarly, for cost functions, \( c(0) = 0 \) is a natural assumption. Matzkin developed a consistent estimator for \( g_i \) and \( F_t \) under additional assumptions.

**Theorem M-2 (Matzkin 2).** Under (M-2) and (M-3), and with \( X \parallel U_t \), \( (g_i, F_t) \) is identified up to a scale normalization for \( g_i \).

**Proof.** See Matzkin (1992). \( \square \)

We use Theorem M-1 or M-2 to claim that we can nonparametrically identify \( g_i \) and \( F_t \) over the support of the data. Obviously, if

\[
0 < p \leq \Pr(D = 1 | X_i, X_e) \leq \bar{p} < 1
\]

\(^{32}\) A function \( \bar{g}_i \) is a least-concave representation of concave function \( g_i \) if for any strictly increasing function \( h \) such that \( h \circ g_i \) is concave, there exists a concave function \( i \) such that \( h \circ g_i = i \circ \bar{g}_i \). Because \( g_i \) is a monotonic transformation of \( \bar{g}_i, \bar{g}_i \) and \( \bar{g}_i \) must have the same iso-value sets.

\(^{33}\) Heckman and Taber (1994) surveyed alternative approaches to identifiability in discrete-choice and duration models.
and the support of \( g_i \) is bounded and strictly contained in the support of \( U_i \), we may be able to identify \( F(U_i) \) only over a subset of its true support. We are now ready to prove Theorem A-2, which extends and clarifies a result of Heckman (1990a) that generalizes the proof of non-parametric identifiability of the Roy model.

**Theorem A-2.** Let \((U_0, U_1, U_2)\) be median-zero, independently and identically distributed random variables with distribution \( F(u_0, u_1, u_2) \). Assume structure \((A-1a)-(A-1c)\) and knowledge of \( F(Y|D = 0, X_0, X_1, X_2) \), \( F(Y|D = 1, X_1, X_2, X_3) \), and \( \Pr(D = 1|X_1, X_2) \). Assume the following:

(a-1) \((U_0, U_1) \perp \! \! \! \perp (X_0, X_1, X_2)\) or
(a-2) \((U_1, U_2) \perp \! \! \! \perp (X_1, X_2, X_3);\)
(b-1) \((M-1)\) and \((M-2)\) or
(b-2) \((M-3)\)

(c) Support\((U_1 \times U_0) = R_1 \times R_1\)
Support\((U_1 \times U_0) = R_1 \times R_1\); and
(d) \(g_i(X_i, X_2) = R_i \rightarrow R_1\) for all \( X_i \)
\(g_o(X_0, X_2) = R_o \rightarrow R_1\) for all \( X_0 \)
\(g_i(X_1, X_2) = R_i \rightarrow R_1\) for all \( X_1 \)
Support\((g_0(X_0, X_2)|X_2) = R_0\) for all \( X_0 \)
Support\((g_1(X_1, X_2)|X_2) = R_1\) for all \( X_1 \)
Support\((X_2|X_1) = \operatorname{Support}(X_2)\)
Support\((X_1|X_2) = \operatorname{Support}(X_1)\)

Then

(I) Under (a-1) or (a-2), (b-1) or (b-2), (c), and (d), \( F_1 \) and \( g_1 \) are identified. If \((M-1)\) and \((M-2)\) are used, \( g_i \) is understood to be the least-concave version of the original \( g_i \).

(II) Under (a-1), (b-1) or (b-2), (c), and (d), \( g_0(X_0, X_2) \) and \( F(u_0, u_2) \) are identified over the supports of \((X_0, X_2)\) and \((U_0, U_2)\), respectively.

(III) Under (a-2), (b-1) or (b-2) and (c), \( g_i(X_i, X_2) \) and \( F(u_1, u_2) \) are identified over the supports of \((X_2, X_2)\) and \((U_1, U_2)\), respectively.

**Proof.** Claim (I) is established in the theorems by Matzkin previously summarized.\(^{35}\) If we establish either the second or third claim, it is clear that the other claim can be proved by a similar argument. We consider only claim (II). Fix \( X_2 \). Observe that for \( X_0, X_1, X_2 \) in the support of \((X_0, X_1, X_2),\)

\[
F(Y|D = 0, X_0, X_1, X_2) = \frac{\int_{-\infty}^{x_0} f(u_0, u_1) \, du_0 \, du_1}{\int_{-\infty}^{x_2} f(u_1) \, du_1}
\]

and further observe that we know the denominator on the right-hand side. Thus we know the left-hand side of

\[
F(Y|D = 0, X_0, X_1, X_2) \Pr(D = 0|X_1, X_2)
\]

\[
= \int_{-\infty}^{x_0} \int_{-\infty}^{x_2} f(u_0, u_1) \, du_0 \, du_1
\]

Under condition (d), for each \( X_2 \), we can vary \( X_2 \) freely and trace out \( g_0, g_0 + \lambda \) for each \( \lambda = \Pr(D = 0|X_2) \). That is, we can vary \( X_0 \) as required to fix \( F(Y|D = 0, X_0, X_1, X_2) \) at a given value when \( y_0 \) is varied, and in this way we determine \( g_0 \) up to scale \( \lambda \). Tracing out \( g_0 \) for all values of \( y_0 \) and \( X_0 \) for each value of \( \lambda \) identifies \( F(u_0, u_2) \) up to scale for \( u_0 \). Setting \( \lambda = 1 \) (i.e., letting \( g_1 \rightarrow -\infty \)), we obtain \( F(u_0, u_2) \) by virtue of assumption (c). The location of \( U_0 \) is obtained from the assumption of a zero median. This pins down the constant \( \lambda \) and \( g_0 \). With \( g_0 \) in hand, we can recover \( F(u_0, g_0(X_0, X_2)) \). Varying \( X_2 \) over its full support, we can identify \( F(u_0, u_2) \). As this is true for each value of \( X_2 \), we have established (II), and by similar reasoning we can establish (III). Thus we establish the theorem.

Observe that the theorem can be modified so that the variables in common between \( g_i \) and \( g_j \) are different from the variables in common between \( g_0 \) and \( g_2 \). Furthermore, the supports of \((U_0, U_1)\) and the conditional supports of \((X_0)\) and \((X_1)\) do not have to be \( R_i \). It is enough to have \( \operatorname{Support}(U_0) = \operatorname{Support}(g_0(X_0, X_2)|X_2) \) and \( \operatorname{Support}(U_1) = \operatorname{Support}(g_1(X_1, X_2)|X_2) \). Theorem A-2 has a simpler structure than Theorem A-1. A discussion similar to that conducted after Theorem A-1 regarding the support of the \( X \) applies.

First, if \( 0 < \operatorname{Support}(\nu) < 1 \), then it is not possible to trace out the full distribution of \( F(u_i) \) nor is it possible to identify \( F(u_i) \) using the limit \( \lambda = 1 \). This could happen, for example, if the support of \( X_i \) is restricted for all \( X_i \) such that

\[
\operatorname{Support}(g_i(X_i, X_2)) \subset \operatorname{Support}(\nu)
\]

Under this restriction, it is not possible to trace out the full distribution of \( U_i \). Alternatively, even if the support of \( X_i \) is \( R_i \) for all \( X_i \), it is possible that (A-5) is satisfied by virtue of restrictions on the function \( g_i \). Notice also that (A-5) might be satisfied for some values of \( X_i \), but not for others. Similar remarks apply to \( g_0 \). Again, restrictions on the range of \( X_0 \) may prohibit recovery of \( F(u_0, g_0(X_0, X_2)|X_2) \) even if (A-5) does not hold. Thus it may happen that

\[
\operatorname{Support}(g_0(X_0, X_2)) \subset \operatorname{Support}(U_0)|g_0 + U_0 < 0
\]

which might arise because of restrictions on the support of \( X_0 \) or because of restrictions on \( g_0 \). Even if (A-6) does not hold for some \( X_i \), it may hold for others.
Theorem A-2 is weaker than Theorem A-1. It implies that we can recover $F(y|D = 1, X_0, X_1)$ from the available cross-section data, provided that its conditions are satisfied. To see this, recall that we can obtain $F(y|X)$ by letting $p \to 1$, which we are free to do because $g_1$ can be varied independently of $g_0$, and the support of $g_1$ is the whole real line. Because by hypothesis we know $F(y|D = 0, X) = 0$, we can apply the identity

$$F(y|X) = F(y|D = 0, X) \Pr(D = 0|X) + F(y|D = 1, X) \Pr(D = 1|X)$$

to solve for $F(y|D = 1, X)$ provided that $\Pr(D = 1|X) \neq 0$.

Using nonexperimental data, we are in the same position as we would be in if we ran an experiment that satisfied assumption (I-10) in the text. In particular, we can identify the mean impact of treatment on the treated, $E(Y - Y^0|D = 1, X)$. In the general case covered by Theorem A-2, social experiments do not solve the fundamental evaluation problem that we cannot observe the same person in both states simultaneously, and so cannot observe both components of $(Y, Y^0)$.

Collecting all of the subscripted variables into a common vector $X$, under the conditions of Theorem A-1 it is possible to generalize from the data recovered from a social experiment, $F(y|x|D = 1, X)$, combined with data on nonparticipants, $F(y|x|D = 0, X)$, to recover the entire distribution $F(y|x|D = 1, X)$ provided that assumption (I-10) is satisfied, and provided that there are no general-equilibrium effects. Thus it is possible to answer all of the questions posed in Section 1 of the text if agents are income maximizers. In the case of the Roy model described by Theorem A-1, social experiments are not required to answer these questions because $F(y|x|D = 1, X)$ is redundant information.

An extension of Theorem A-2 is useful in identifying the full distribution of outcomes conditional on $X$ in the panel-data case when treatments are irreversible.

**Theorem A-3.** Let $(U_0, U_1, U_1)$ be median-zero, independently and identically distributed random variables with distribution $F(u_0, u_1, u_1)$. Assume outcome and decision structure (A-1a)–(A-1c) and knowledge of $F(y|x, y|x'|D = 1, X_0, X_1, X_c)$ and $\Pr(D = 1|X_0, X_1, X_c)$. Assume the following:

(a) $(U_0, U_1, U_1) \perp (X_0, X_1, X_c)$
(b-1) $(M-1)$ and $(M-2)$ or
(b-2) $(M-3)$
(c) $\Pr(U_0, U_1, U_1) = R_0 \times R_1 \times R_1$
(d) Partition $X_1$ into $(X_{10}, X_{11}, X_{10})$, where $X_{10}$ is a subset of $X_1$ (with $k_{10}$ variables) not in $X_0$, $X_{11}$ or $X_c$, $X_{10}$ are variables in common with $X_0$ ($k_{10}$ in number), and $X_{10}$ are variables in common in $X_1$ and $X_c$. Partition $X_0$ into $(X_{00}, X_{01}, X_{00})$, where $X_{00}$ is a subset of $X_0$ (with $k_{00}$ variables) not in $X_1$ or $X_c$, and $X_{00}$ is

Then

(I) Under (a) and (b-1) or (b-2), (c), and (d), $g_1$ is identified where, if (M-1) and (M-2) are used, $g_1$ is understood to be the least-concave version of the original $g_1$. (This follows from Theorem M-1.)

(II) Under (a), (b-1) or (b-2), (c), and (d), $g_1(X_1, X_c), g_0(X_0, X_c), X_{00}$, and $F(U_0, U_1, U_1)$ are identified.

**Proof.** Claim (I) is established in the same way that claim (II) of Theorem A-2 is established. All that is needed to establish this claim is that $U_1 \perp (X_1, X_c)$ (b-1) or (b-2) from Matzkin, and $g_1(X_1, X_c) : R_1 \to R_1$ for all $X_0$.

Claim (II) is established in essentially the same way that claim (II) of Theorem A-2 is established. From assumption (c),

$$F(y|x, y|x' = 1, X_0, X_1, X_c) \Pr(D = 1|X_1, X_c) = F_{U_0, U_1, U_1}[y|X_0, X_1, X_c] = F_{U_0, U_1, U_1}[- \tilde{g}_0(X_0, X_1, X_c), - \tilde{g}_1(X_1, X_c), - g_1(X_1, X_c)]$$

From condition (d), we can vary the components $X_{00}, X_{10}$, and $X_{11}$ freely. If we set $X_{10}$ to the value such that $g_1 \to - \tilde{g}_1$, and if we set $X_{11}$ to the value such that $g_1 \to - \tilde{g}_0$ so that $\Pr(D = 1|X) = 1$, we can trace out $g_0(X_{00}, X_{01}, X_{00})$ given the remaining arguments for each value of the
conditioning arguments. [This limit operation “zeros out” the last two arguments of (A-7).] We can repeat this argument for all conditioning subsets and recover \( g_0(x_{00}, x_{01}, x_{02}) \) up to a constant. Tracing out \( g_0 \) for all values of \( y_0 \) identifies \( F(u_0) \) up to scale. The scale is identified by the median-zero assumption. By a parallel argument, but reversing the roles of \( X_1 \) and \( X_0 \), and \( g_1 \) and \( g_0 \), we obtain \( g_1 \) and \( F(u_1) \). Staying in the set where \( \Pr(D = 1|x_1, x_2) = 1 \), we can construct \( F(u_0, u_1) = F(y_0 - g_0, y_1 - g_1) \) by independently varying \( g_0 \) and \( g_1 \), which we are free to do as a consequence of assumption (d). More generally, we can repeat this argument for all values of \( p = \Pr(U_i > -g_i) \). We can vary \( x_{ii} \) to offset any changes induced in \( x_{00} \) and \( x_{11} \). Thus we can identify

\[
F(u_0, u_1, -g_0(x_{ii}, x_{i0}, x_{i1}, x_{i2}) = F(y_0 - g_0, y_1 - g_1 | D = 1, x_1, x_2)
\]

and hence we can identify the full joint distribution \( F(u_0, u_1, u_2) \) by tracing out \( x_{00}, x_{11}, \) and \( x_{ii} \). We can do this for all \( X_{ii}, X_{i0}, X_{i1}, \) and \( X_{i2} \) and hence the theorem is proved. \( \Box \)

Observe that the theorem does not require infinite supports. Thus it is enough to have

- Support \( (U_i) = \text{Support}[g_0(x_{ii}, x_{i0}, x_{i1}, x_{i2})] \)
- Support \( (U_0) = \text{Support}[g_0(x_{00}, x_{01}, x_{02}, x_{03}, x_{04}, x_{05}, x_{06})] \)
- Support \( (U_1) = \text{Support}[g_0(x_{11}, x_{10}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16})] \)

### Appendix B

The data analyzed in this study were gathered as part of an experimental evaluation of the training programs financed under Title II-A of the Job Training Partnership Act (JTPA). The experiment was conducted at a sample of 16 JTPA training centers around the country. Data were gathered on JTPA applicants randomly assigned either to a treatment group allowed access to JTPA training services or to a control group denied access to JTPA services for 18 months. Random assignment covered some or all of the period from November 1987 to September 1989 at each center. In total, 20,601 persons were randomly assigned.

Follow-up interviews were conducted with each person in the experimental sample during the period from 12 to 24 months after random assignment. Those interviews gathered information on employment, earnings, participation in government transfer programs, schooling, and training during the period after random assignment. The response rate for this survey was around 84 percent. The sample used here includes only those adult women who (1) had a follow-up interview scheduled at least 18 months after random assignment, (2) responded to the survey, and (3) had usable earnings information for the 18 months after random assignment.

The sample was chosen to match that used in the 18-month experimental-impact study by Bloom et al. (1993). As in that report, the earnings measure is the sum of self-reported earnings during the 18 months after random assignment. This earnings sum is constructed from survey questions about the length, hours per week, and rate of pay on each job held during this period. Outlying values for the earnings sum are replaced by imputed values as in the impact report. However, imputed earnings values used in the report for adult female nonrespondents are not used.

### References


Evaluating the Welfare State

evaluation estimator: evidence from evaluating a job training programme. 
(1998a). Matching as an econometric evaluation estimator. Review of Eco-
nomic Studies 65:261–94.
Heckman, J., LaLonde, R., and Smith, J. (in press-b). The economics and econo-
metrics of active labor market programs. In: Handbook of Labor 
Economics, Vol III, ed. O. Ashenfelter and D. Card. Amsterdam: North 
Holland.
interventions. In: Longitudinal Analysis of Labor Market Data, ed. J. 
hypothesis and estimating the mixing distribution by the method of 
of social programs. In: Measuring Labour Market Measures: Evaluating the 
Effects of Active Labour Market Policies, ed. K. Jensen and P. K. Madsen, 
Perspectives 9:85–110.
(1996). Experimental and nonexperimental evaluation. In: International Hand-
book of Labour Market Policy and Evaluation, ed. G. Schmidt, J. O'Reilly, 
Heckman, J., Smith, J., and Clements, N. (1997c). Making the most out of program 
evaluations and social experiments: accounting for heterogeneity in 
general models for unobservables in duration analysis. Mimeoographed man-
uscript, Department of Economics, University of Chicago.
Pilot Study in the Evaluation of Government Services. Ann Arbor, MI: Insti-
tute for Social Research.
Massachusetts Institute of Technology Press.
LaLonde, R. (1986). Evaluating the econometric evaluations of training pro-
Columbia University Press.
tations and Econometric Practice, pp. xi–xi. Minneapolis: University of Min-
nesota Press.
CHAPTER 9

Frisch, Hotelling, and the Marginal-Cost Pricing Controversy

Jean-Jacques Laffont

Le meilleur de tous les tarifs serait celui qui ferait payer à ceux qui passent sur une voie de communication un péage proportionnel à l’utilité qu’ils retirent du passage....

Il est évident que l’effet d’un tel tarif serait: d’abord de laisser passer autant de monde que si le passage était gratuit; ainsi point d’utilité perdue pour la société; ensuite de donner une recette toujours suffisante pour qu’un travail utile puisse être fait.

Je n’ai pas besoin de dire que je ne crois pas à la possibilité d’application de ce tarif volontaire; il rentrerait un obstacle insurmontable dans l’improbabilité universelle des pressions, mais c’est là le type dont il faut chercher à s’approcher par un tarif obligatoire.

Jules Dupuit (1849, p. 223)

1 Introduction

In an elegant *Econometrica* paper, Hotelling (1938) provided the appropriate formulas to assess the social costs of marginal departures from marginal-cost pricing when the interrelations between commodities are taken into account. In so doing he generalized the work of Dupuit (1844) and Marshall (1890). But he went further. He advocated marginal-cost pricing for those industries with large fixed costs and more generally increasing returns:

This proposition has revolutionary implications, for example in electric power and railway economics, in showing that society would do well to cut rates drastically and replace the revenue thus lost by subsidies

This essay was written for the Ragnar Frisch Centennial Symposium in Oslo, March 1995. Steinar Strøm’s friendly pressure is responsible for its existence. I am grateful to M. Ivaldi for helping me with the simulations and J. Tirole for helpful comments. I thank the participants in the conference for useful discussions, in particular G. Bjerkholt, J. Hausman, J. Heckman, K. Moene, and A. Sandmo. Thanks also to the referee.