1. Introduction

2. Public Job Training and Active Labor Market Policies

3. The Evaluation Problem and the Parameters of Interest in Evaluating Social Programs
3.1 The Evaluation Problem

Conditioning variables X.

Potential outcomes \( Y_0 \) and \( Y_1 \)

(3.1a) \( Y_0 = \mu_0(X) + U_0 \)

(3.1b) \( Y_0 = \mu_0(X) + U_1 \)

\[
E(Y_0 | X) = \mu_0(X)
\]

and \( E(Y_1 | X) = \mu_1(X) \)

\[
E(U_1 | X) = E(U_0 | X) = 0
\]

\( Y_0 = g(X) + U_0 \)
\[ Y_1 = g(X) + U_1 \]

Do Not Necessarily Require

\[ E(U_0 \mid X) \neq 0 \text{ and } E(U_1 \mid X) \neq 0 \]

But Then

\[ \mu_0(X) = g_0(X) + E(U_0 \mid X) \]

\[ \mu_1(x) = g_1(X) + E(U_1 \mid X). \]

Does not imply that \( E(U_1 - U_0 \mid X, D = 1) = 0. \)
Y may be a vector of outcomes

$$E(U_{0t}|X) = 0, E(U_{1t}|X) = 0$$

$$X = (X_1,\ldots,X_T)$$

$$E(Y_{1t} - Y_{0t}|X,D = 1) \quad t = 1,\ldots,T,$$
Assume A Noncausality Condition

\[ Y^P = ((Y_{01}, Y_{11}), ..., (Y_{0T}, Y_{1T})) \]

\[ f(X|D, Y^P) = f(X|Y^P) \]

D does not determine X. To obtain

\[ E(Y_{1t} - Y_{0t}|X_c, D = 1) \] integrate out \( \tilde{X}_c \).

\( \tilde{X}_c \) is the portion of X not in \( X_c \):

\[ X = (X_c, \tilde{X}_c). \]

\[ E(Y_{1t} - Y_{0t}|X_c, D) \]

\[ f(\tilde{X}_c|D = 1) \]
Roles of “0” and “1” can be reversed. “0” as a benchmark “no treatment ” state. “0” to “1” is given by

\[ \Delta = Y_1 - Y_0. \]

The gain \( \Delta \). The fundamental evaluation problem. Do not know both coordinates of \((Y_1, Y_0)\) and hence \( \Delta \). Missing data problem.
3.3 The Counterfactuals Most Commonly Estimated In The Literature

Direct Effects.

Indirect Effects.

(A) the proportion of people taking the program who benefit from it:

\[ \Pr(Y_1 > Y_0 \mid D = 1) = \Pr(\Delta > 0 \mid D = 1); \]
(B) the proportion of the total population benefiting from the program:

\[
\Pr(Y_1 > Y_0 \mid D = 1) \cdot \Pr(D = 1) = \Pr(\Delta > 0 \mid D = 1) \cdot \Pr(D = 1);
\]
(C) selected quantiles of the impact distribution

\[ \inf_{\Delta} \{ \Delta : F(\Delta \mid D = 1) > q \} , \]

where \( q \) is a quantile of the distribution and where “inf” is the smallest attainable value of \( \Delta \) that satisfies the condition stated in the braces;
(D) the distribution of gains at selected base state values:

\[ F(\Delta \mid D = 1, Y_0 = y_0); \]

(E) the increase in the level of outcomes above a certain threshold \( \bar{y} \) due to a policy:

\[ \Pr(Y_1 > \bar{y} \mid D = 1) - \Pr(Y_0 > \bar{y} \mid D = 1). \]
(3.3) \[ Y = DY_1 + (1 - D)Y_0. \]

When \( D = 1 \) we observe \( Y_1 \); when \( D = 0 \) we observe \( Y_0 \)

Random Coefficient: Switching Regression

Formulation

\[ E(Y_j|X) = \mu_j(X) = X\beta_j, j = 0, 1. \]

Obtain

\[ Y = D(\mu_1(X) + U_1) + (1 - D)(\mu_0(X) + U_0). \]

\[ Y = \mu_0(X) + D(\mu_1(X) - \mu_0(X) + U_1 - U_0) + U_0. \]
(3.4) \[ Y = X\beta_0 + D(X(\beta_1 - \beta_0) + (U_1 - U_0)). \]

\[ E(Y_1 - Y_0 | X) = E(\Delta | X). \]
(3.5) \[ Y = \mu_0(X) + D(\mu_1(X) - \mu_0(X)) \]
\[ + \{ U_0 + D(U_1 - U_0) \} \]
\[ = \mu_0(X) + D(E(\Delta|X)) \]
\[ + \{ U_0 + D(U_1 - U_0) \} \]
\[ = X\beta_0 + DX(\beta_1 - \beta_0) \]
\[ + \{ U_0 + D(U_1 - U_0) \} \]

If the model is specialized \( \beta_1 = (\beta_{10}, \ldots, \beta_{1K}) \)

and \( \beta_0 = (\beta_{00}, \ldots, \beta_{0K}) \),
$$\beta_{1j} = \beta_{0j} = \beta_j, \quad j = 1, \ldots, K$$

and $\beta_{00} = \beta_0$ and $\beta_{10} - \beta_{00} = \alpha$:

(3.6) \hspace{1cm} E(Y_1 - Y_0|X) = \beta_{10} - \beta_{00} = \alpha.

(3.7) \hspace{1cm} Y = X\beta + D\alpha + \{U_0 + D(U_1 - U_0)\}.
Mean is not zero in general.

\[ E[U_0 + D(U_1 - U_0)] = \]

\[ E(U_1 - U_0 | D = 1) \Pr(D = 1). \]

\[ E(U_1 - U_0 | D = 1) \neq 0. \]

Does not estimate \( E(\Delta | X) \). \( \text{(ATE|X)} \)

Instead

\[ E(\Delta | X, D = 1). \text{ (TT|X)} \]
\[(3.8) \ E(\Delta|X,D = 1) = E(Y_1 - Y_0|X,D = 1) \]
\[= X(\beta_1 - \beta_0) + E(U_1 - U_0|X, D = 1). \]
(3.9) \[ Y = \mu_0(X) + D(E(\Delta|X, D = 1)) \]

\[ +\{U_0+D[(U_1 - U_0) - E(U_1 - U_0|X,D=1)]\} \]

\[ = X\beta_0+D(X(\beta_1 - \beta_0)+E(U_1-U_0|X,D=1)) \]

\[ +\{U_0+D[(U_1- U_0)-E(U_1- U_0|X,D=1)]\}. \]

\[ E(\Delta \mid X, D = 1) \]

“structural” parameters \( X(\beta_1 - \beta_0) \) unobservables \((E(U_1 - U_0|X, D = 1))\).
Two evaluation parameters are the same if

\[ U_1 - U_0 = 0 \]

\[ Y_1 - Y_0 = \mu_1(X) - \mu_0(X) = X(\beta_1 - \beta_0). \]

(i.e. \( Y_1 - Y_0 = \alpha \)),

(3.10) \( Y = X\beta + D\alpha + U \), where \( E(U) = 0 \).

\[ X(\beta_1 - \beta_0), \ D \text{ and } U. \]

(3.10) is very special but conventional.

\( (Y_1 - Y_0 = \alpha) \). \( \alpha \) constant

Second Case When \( Y_1 - Y_0 = \alpha \), \( \alpha \) not con-
stant

\[ \mu_1(X) \text{ and } \mu_0(X) \text{ or } \beta_1 \neq \beta_0 \]

(3.11) \( E(U_1 - U_0|X, D = 1) = 0. \)

(3.11) arises when conditional on \( X, D \) does not explain or predict \( U_1 - U_0 \).

In that case

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

and
\begin{align*}
\mathbb{E}(U_0 + D[(U_1 - U_0)] - \\
\mathbb{E}(U_1 - U_0|X,D = 1)|X,D) &= \mathbb{E}(U_0|X,D) \\
\text{When } (3.11) \text{ holds} \\
\text{Var}(U_0 + D(U_1 - U_0)|X,D) &= \text{Var}(U_0|X,D) \\
+2\text{COV}(U_0, U_1 - U_0 | X, D)D \\
+\text{VAR}(U_1 - U_0 | X, D)D. \\
\end{align*}
Distinction between a model with $U_1 = U_0$, and one with $U_1 \neq U_0$, is fundamental. In the general case when $U_1 \neq U_0$ and (3.11) no longer holds.

The distinctions among these three models:

(a) the coefficient on D is fixed (given X) for everyone;

(b) the coefficient on D is variable (given X),

(c) the coefficient on D is variable (given X)
and does help determine program participation.
3.4 Is Treatment on the Treated an Interesting Economic Parameter?

What economic question does parameter (3.2) answer?

$Y_1$ and $Y_0$. (Two outcomes)

Let $\varphi$ be a vector of policy variables

$c(\varphi)$ is the social cost of $\varphi$

$c(0) = 0$ and $c$ is convex and increasing in
Let $N_1(\varphi) N_0(\varphi)$ be the number of persons in state “0”.

**Total Net Output In Society:**

$$N_1(\varphi)E(Y_1 \mid D = 1, \varphi) +$$

$$N_0(\varphi)E(Y_0 \mid D = 0, \varphi) - c(\varphi),$$

$$N_1(\varphi) + N_0(\varphi) = \bar{N} \quad \text{(Full Employment)}$$
\[
\Delta(\varphi) = \\
\frac{\partial N_1(\varphi)}{\partial \varphi} [E(Y_1 | D=1, \varphi) - E(Y_0 | D=0, \varphi)] + \\
N_1(\varphi) \left[ \frac{\partial E(Y_1 | D = 1, \varphi)}{\partial \varphi} \right] + \\
N_0(\varphi) \left[ \frac{\partial E(Y_1 | D = 0, \varphi)}{\partial \varphi} \right] - \frac{\partial c(\varphi)}{\partial \varphi}.
\]
Parameters of interest:

(i) \( \frac{\partial N_1(\varphi)}{\partial \varphi} \);

the number of people entering or leaving state 1.

(ii) \( E(Y_1 \mid D = 1, \varphi) - E(Y_0 \mid D = 0, \varphi) \);

the mean output difference between sectors.

(iii) \( \frac{\partial c(\varphi)}{\partial \varphi} \);

the social marginal cost of the policy.
Nowhere on this list these are “the effect of treatment on the treated”:

(a) $E(Y_1 - Y_0 \mid D = 1, \varphi)$

or

(b) $E(Y_1 \mid \varphi = \bar{\varphi}) - E(Y_0 \mid \varphi = 0)$

where $\varphi = \bar{\varphi}$ sets $N_1(\bar{\varphi}) = \bar{N}$. This is the effect of universal coverage for the program.
Place of Context of

Binary choice random utility framework.

\[ U = X + k(\varphi) \]

\( k \) is monotonic in \( \varphi \) the joint distributions of

\((Y_1, X)\) and \((Y_0, X)\) are \( F(y_1, x) \) and \( F(y_0, x) \),

\( X = Y_1 - Y_0 \)

\[ D = 1(U \geq 0) = 1(X \geq -k(\varphi)) \]
\((1(Z > 0) = 1 \text{ if } Z > 0; = 0 \text{ otherwise})\)

\[ N_1(\varphi) = \bar{N} \Pr(U \geq 0) = \bar{N} \int_{-k(\varphi)}^{\infty} f(x)dx, \]

and

\[ N_0(\varphi) = \bar{N} \Pr(U < 0) = \bar{N} \int_{-\infty}^{-k(\varphi)} f(x)dx. \]

Total output is

\[ \bar{N} \int_{-\infty}^{\infty} y_1 \int_{-k(\varphi)}^{\infty} f(y_1, x \mid \varphi)dx dy_1 \]

\[ + \bar{N} \int_{-\infty}^{\infty} y_0 \int_{-\infty}^{-k(\varphi)} f(y_0, x \mid \varphi)dx dy_0 - c(\varphi). \]
(3.13) \( \Delta(\varphi) \)

\[
\begin{align*}
\Delta(\varphi) &= \bar{N}k'(\varphi)f_x(-k(\varphi)) \\
\frac{\partial}{\partial \varphi} &+ \frac{\partial}{\partial \varphi}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial}{\partial \varphi} &+ \frac{\partial}{\partial \varphi}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial}{\partial \varphi} &+ \frac{\partial}{\partial \varphi}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial}{\partial \varphi} &+ \frac{\partial}{\partial \varphi}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial}{\partial \varphi} &+ \frac{\partial}{\partial \varphi}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial}{\partial \varphi} &+ \frac{\partial}{\partial \varphi}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial}{\partial \varphi} &+ \frac{\partial}{\partial \varphi}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial}{\partial \varphi} &+ \frac{\partial}{\partial \varphi}.
\end{align*}
\]

\[
\begin{align*}
\frac{\partial}{\partial \varphi} &+ \frac{\partial}{\partial \varphi}.
\end{align*}
\]
A well-defined margin:

\[ X = -k(\varphi), \ \Delta(\varphi) = 0, \]
(3.14) \( E(Y_1 \mid D = 1, X = -k(\varphi), \varphi) \)

\[-E(Y_0 \mid D = 0, X = -k(\varphi), \varphi)\]

\[= E(Y_1 - Y_0 \mid D = 1, X = -k(\varphi), \varphi).\]

(Marginal Effect)
The conventional evaluation parameter,

\[ E(Y_1 - Y_0 \mid D = 1, x, \varphi) \]

Can be justified as an all or nothing evaluation of a policy at level \( \varphi = \tilde{\varphi} \):

\[ A(\tilde{\varphi}) = \{ N_1(\tilde{\varphi})E(Y_1 \mid D = 1, \varphi = \tilde{\varphi}) \]

\[ + N_0(\tilde{\varphi})E(Y_0 \mid D = 0, \varphi = \tilde{\varphi}) - c(\tilde{\varphi}) \}

\[ - \{ N_1(0)E(Y_1 \mid D = 1, \varphi = 0) \]

\[ + N_0(0)E(Y_0 \mid D = 0, \varphi = 0) \} \]
\[ N_0(0) = \bar{N}. \] If \( A(\tilde{\varphi}) > 0 \), undertake the project.

**Crucial Assumption:** That

\[(3.15)E(Y_0 \mid D=0, \varphi = \tilde{\varphi}) = E(Y_0 \mid D=0, \varphi=0).\]

(No treatment state = No Program State)
Absence of general equilibrium effects. When (3.15) holds we have

\begin{equation}
(3.16) A(\tilde{\varphi}) = N_1(\tilde{\varphi}) E(Y_1 - Y_0 \mid D = 1, \varphi = \tilde{\varphi}) - c(\tilde{\varphi}).
\end{equation}

For evaluating the effect of marginal variation or “fine-tuning” of existing policies, measure \( \Delta(\varphi) \) is more appropriate.
4. Prototypical Solutions to the Evaluation Problem

Three widely-used comparisons

\[ E(Y_1 - Y_0 \mid X, D = 1). \]

4.1 The Before-After Estimator

Let \( Y_{1t} \) be the post-program earnings of a person longitudinal data are available, \( Y_{0t'} \) is the pre-program outcome \( t > k > t' \).

(4.A.1) \[ E(Y_{0t} - Y_{0t'} \mid D = 1) = 0. \]
“Before-After” estimator

\[(4.1) \quad (\overline{Y}_{1t} - \overline{Y}_{0t'})_1,\]

“1” denotes conditioning on \(D = 1\), and the “−” denotes

\[Y_{1t} - Y_{0t} = (Y_{1t} - Y_{0t'}) + (Y_{0t'} - Y_{0t}).\]

The second term \((Y_{0t'} - Y_{0t})\) is the approximation error.
4.2. The Difference-in-Differences Estimator

\( (4.2) E(Y_{0t} - Y_{0t'} | D = 1) \)

\[ = E(Y_{0t} - Y_{0t'} | D = 0) \]

difference-in-differences estimator given by

\( (4.2) (\bar{Y}_{1t} - \bar{Y}_{0t'})_1 - (\bar{Y}_{0t} - \bar{Y}_{0t'})_0 \) \( t > k > t' \)

\( E(\Delta_t | D = 1) = E(Y_{1t} - Y_{0t} | D = 1) \) where

\( \Delta_t = Y_{1t} - Y_{0t} \)

\[ E[(\bar{Y}_{1t} - \bar{Y}_{0t'})_1 - (\bar{Y}_{0t} - \bar{Y}_{0t'})_0] = E(\Delta_t | D = 1). \]
\[(Y_{1t} - Y_{0t'})_1 - (Y_{0t} - Y_{0t'})_0.\]
4.3 The Cross-Section Estimator

(4.A.3) \( E(Y_{0t} \mid D = 1) = E(Y_{0t} \mid D = 0) \),

(4.3) \( (\bar{Y}_{1t})_1 - (\bar{Y}_{0t'})_0 \).

\[
E((\bar{Y}_{1t})_1 - (\bar{Y}_{0t})_0) = E(\Delta_t \mid D = 1).
\]
5. Social Experiments

5.1 How Social Experiments Solve the Evaluation Problem.

"*" denote outcomes in the presence of random assignment. Conditional on $X$ for each person we have $(Y_1^*, Y_0^*, D^*)$

$(Y_1, Y_0, D)$ when the program operates normally without randomization. Let $R = 1$ if a person for whom $D^* = 1$
Thus, $R = 1$ and $R = 0$

(5.A.1) $E(Y_1^* - Y_0^* | X, D^* = 1) = E(Y_1 - Y_0 | X, D = 1)$.

(5.A.2a) $E(Y_1^* | X, D^* = 1) = E(Y_1 | X, D = 1)$

and

(5.A.2b) $E(Y_0^* | X, D^* = 1) = E(Y_0 | X, D^* = 1)$.

If (5.A.2a) is true, among the population for whom $D = 1$ and $R = 1$ we can identify

42
$E(Y_1 | X, D = 1, R = 1) = E(Y_1 | X, D = 1).$

$E(Y_0 | X, D = 1, R = 0) = E(Y_0 | X, D = 1).$

Assumption (4.A.3).

$E(\Delta | X, D = 1) = E(Y_1 - Y_0 | X, D = 1).$

An experiment that satisfies (5.A.1) or (5.A.2a)

and (5.A2b) does not make $D$ orthogonal to $U$.

Equates the bias in the two groups $R = 1$ and $R = 0$. 
\[
E(Y|X, D = 1, R = 1) = \\
g_1(X) + E(U_1|X, D = 1)
\]

and \( E(Y|X, D = 1, R = 0) = \\
g_0(X) + E(U_0|X, D = 1). \)
\[ E(Y|X, D = 1, R = 1) = g_1(X) \]
\[ + E(U_1 - U_0|X, D = 1) + E(U_0|X, D = 1). \]

Subtracting the second mean from the first eliminates bias:

\[ E(Y|X,D=1,R=1) - E(Y|X,D = 1,R = 0) \]
\[ = g_1(X) - g_0(X) + E(U_1 - U_0|X, D = 1). \]

If assumption (5.A.1) or assumptions (5.A.2a) and (5.A.2b) fail to hold because the program participation probabilities are affected, so \( D^* \)
and \( D \) are different, then the composition of the participant population differs in the presence of random assignment. In two cases, experimental data still provide unbiased estimates of the effect of treatment on the treated. First, if the effect of training is the same for everyone:

\[
(5.\text{A.3}) \quad Y_{1i} - Y_{0i} = \Delta_i \equiv \Delta \text{ for all } i.
\]

\[U_1 = U_0 \text{ in (3.9).}\]
Second Case:

(5.A.4) \[ E(\Delta \mid X, D = 1) = E(\Delta \mid X), \]

(5.A.3) or (5.A.4) holds, \( E(\Delta \mid X, D = 1). \)

\[ E(\Delta \mid X, D = 1) = E(Y_1 - Y_0 \mid X, D = 1). \]

Assumption (4.A.3) of \( X \) such that

\[ \Pr(D = 1 \mid X) = 1. \]

Balancing the distribution of \( X \) such that

\[ \Pr(D = 1 \mid X) = 0. \]

\[ E(\Delta \mid X, D = 1) = E(Y_1 - Y_0 \mid X, D = 1). \]
We do not directly estimate the effect of randomly selecting a person to go into the program:

$$E(\Delta \mid X) = E(Y_1 - Y_0 \mid X)$$
To secure compliance given payment assume

\[(5.\text{A}.5a) \quad \Pr(D = 1 \mid X, c) = 1\]

and

\[(5.\text{A}5b) \quad E(\Delta \mid X, c) = E(\Delta \mid X)\]
Randomization of eligibility is sometimes

Let $e = 1$

$(Y_0, Y_1, D, X)$ and that

$\Pr(D = 1 \mid X) \neq 0$. Can Form

\[
\frac{E(Y \mid X, e = 1) - E(Y \mid X, e = 0)}{\Pr(D = 1 \mid X, e = 1)} = E(\Delta \mid X, D = 1).
\]
5.2 Intention to Treat and Substitution Bias

\( T = 1 \) for persons actually receiving training,

\( T = 0 \) otherwise.

\( T^* = 1 \) if control group members would have received training had they been offered otherwise.

\[
E(\Delta \mid X, D = 1, R = 1, T = 1) = E(\Delta \mid X, D = 1, T = 1)
\]
(Mean Impact of training on those members of the control group who actually receive it)

\[ E(\Delta \mid X, D = 1) \]

Only When \( \Delta_i = \Delta \ \forall i \) or else no selection on unobserved gain:

\[ (\Delta - E(\Delta)) \]

(5.A.6) \( E(Y \mid X, D = 1, R = 1, T = 0) = E(Y \mid X, D = 1, R = 0, T^* = 0). \)

Under (5.A.6),

52
\[
(5.1) \frac{E(Y | X,D=1,R=1) - E(Y | X,D=1,R=0)}{P(T=1 | X,D=1,R=1)}
\]

Let \( S = 1 \) denote control group members receiving substitute alternative sources and let \( S=0 \) for those who do not:

Assume

\[
(5.A.3') \quad Y_{1i} - Y_{0i} = Y_{2i} - Y_{0i} = \Delta_i \equiv \Delta
\]

for all \( i \).
Generalized version of 5.4:

\[ (5.4') \mathbb{E}(Y_1 - Y_0 \mid X, D=1, T=1, R=1) = \]
\[ \mathbb{E}(Y_2 - Y_0 \mid X, D = 1, S = 1, R = 0). \]

(Mean impact of training received by control group receiving it = mean impact of substitute training)

(5.3') implies (5.4').

(5.2)

\[ \frac{\mathbb{E}(Y \mid X, D=1, R=1) - \mathbb{E}(Y \mid X, D=1, R=0)}{\mathbb{P}(T=1 \mid X, D=1, R=1) - \mathbb{P}(S=1 \mid X, D=1, R=0)}. \]
5.3 Social Experiments in Practice

5.3.1 Two Important Social Experiments

(5.A.1) or (5.A.2a)-(5.A.2b)

5.3.2 The Practical Importance of Dropping Out and Substitution

5.3.3 Additional Problems Common to All Evaluations
6. Econometric Models of Outcomes and Program Participation

6.1 Uses of Economic Models

(1) They suggest lists of explanatory variables

(2) They sometimes suggest plausible “exclusion restrictions”

(3) They sometimes suggest specific functional forms of estimating equations motivated
by a priori theory or by cumulated empirical wisdom.
6.2. Prototypical Models of Earnings and Program Participation

One option for, $t = 1$ up through $t = k$:

$$Y_{0j} \quad j = 1, \ldots, k.$$  

After $k$, there are two potential outcomes

$$(Y_{0j}, Y_{1j}) \quad j = k + 1, \ldots, T$$

$T$ is the end of economic life.

(6.1)  

$$(D; Y_{0t}, t = 1, \ldots, k; (Y_{0t}, Y_{1t}),$$

$$t = k + 1, \ldots, T).$$
\[ Y_{0t} = D Y_{1t} + (1 - D) Y_{0t}. \]
6.3. *Expected Present Value of Earnings Maximization*

In period $k$, period $k$ is $I_k$.

c (direct costs)
\((D = 1)\) if

\[
(6.2) \mathbb{E} \left[ \sum_{j=1}^{T-k} \frac{Y_{1,k+j}}{(1+r)^j} - \sum_{j=0}^{T-k} \frac{Y_{0,k+j}}{(1+r)^j} \mid I_k \right] \geq 0,
\]

\(D = 0\) otherwise

### 6.3.1 Common Treatment Effect

\[Y_{1t} - Y_{0t} = \alpha_t, t > k, \alpha_t(X).\]

Ashenfelter and Card (1985) and Heckman and Robb (1985a, 1986a) develop it.

\[E(Y_{1t} - Y_{0t} \mid X, D = 1) = E(Y_{1t} - Y_{0t} \mid X)\]
(6.3) \( D = 1 \), if \( E \left( \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1 + r)^j} - c - Y_{0k} \mid I_k \right) \geq 0, \)

\( D = 0 \) otherwise.

If the \( \alpha_{k+j} \) are constant in all periods and \( T \)
is large (\( T \to \infty \)) the criterion simplifies to

(6.4) \( D = 1 \) if \( E \left( \frac{\alpha}{r} - c - Y_{0k} \mid I_k \right) \geq 0, \)

\( D = 0 \) otherwise.

Expected future rewards \( E \left[ \left( \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1 + r)^j} \right) \mid I_k \right] \).
\[
\Pr(D = 1) = \Pr \left( \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1 + r)^j} > c + Y_{0k} \right).
\]

\[
\Pr(D = 1) = \Pr \left( \frac{\alpha}{r} \geq c + Y_{0k} \right).
\]

\(D = 1\) if and only if \(IN > 0\).

\[
IN = H(X) - V.
\]

\[
H(X) = \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1 + r)^j}, \text{ and } V = c + Y_{0k}.
\]

(6.5) \(\Pr(D=1|X) = \Pr(V < H(X)|X)\).

\[
\Pr(D = 1 \mid X) = \Pr(V < H(X))
\]
(6.6) \quad \Pr(D = 1 \mid X) = \Pr(V < H(X)) = \Phi \left( \frac{H(X) - \mu_1}{\sigma_V} \right)

\Pr(D = 1 \mid X) = \frac{\exp(H(X))}{1 + \exp(H(X))}.

\Pr(D = 1 \mid X) \text{ based on }

\text{IN} = Y_1 - Y_0 = (\mu_1(X) - \mu_0(X)) + (U_1 - U_0).

X \text{ is independent of } U_1 - U_0,

\left[ \text{Var}(U_1 - U_0) \right]^{1/2}
Cross Section Estimator:

Under what conditions is it plausible to assume that

\[(4.\text{A.3}) E(Y_{0t} | D=1) = E(Y_{0t} | D = 0), t > k \]

i.e.

\[(4.\text{A.3}) E(Y_{0t} | \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1 + r)^j} - c - Y_{0k} \geq 0) =

E(Y_{0t} | \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1 + r)^j} - c - Y_{0k} < 0), t > k.\]
Matching Assumptions:

\[(4.A.3')E(Y_{0t} \mid D=1,X)=E(Y_{0t} \mid D=0,X)\].

\[
E(Y_{0t} \mid D = 1, Y_{0k}) = \\
E(Y_{0t} \mid \sum_{j=1}^{T-k} \frac{\alpha_{k+j}}{(1 + r)_j} - Y_{0k} \geq c, Y_{0k})
\]

\[= E(Y_{0t} \mid Y_{0k}) = E(Y_{0t} \mid D = 0, Y_{0k}).\]

\[X = Y_{0k}.\]

\[E(Y_{0t} \mid D=1,(c,Y_{0k}))) = E(Y_{0t} \mid c,Y_{0k})\]

and

\[E(Y_{0t} \mid D = 0,(c,Y_{0k}))) = E(Y_{0t} \mid c,Y_{0k})\]
Need

\[ 0 < Pr(D = 1 \mid X) < 1 \]

(Some People in Both States)
Difference in Difference

Assumption (4.A.1). If in a time homogeneous environment,

\begin{equation}
Y_{0t} = \beta_t + \varphi + U_{0t} \quad \text{for all } t
\end{equation}

and \( E(U_{0t} \mid \varphi) = 0 \) \ for all \( t \)

then \( \varphi \) is “permanent income” and the \( U_{0t} \) are “transitory deviations” around it.

\begin{equation}
(6.3) \text{ for } t > k > t', \text{ we have}
\end{equation}
\[ E(Y_{0t} - Y_{0t'} | D = 1) = \alpha_t + \beta_t - \beta_{t'}, \]

since \( E(U_{0t} | D=1) - E(U_{0t'} | D=1) = 0. \)

(6.8) \[ E(U_{0t} - U_{0t'} | D = 1) = 0 \]

\( U_{0t} \) are serially correlated, then (4.A.1) will generally not be satisfied.

If \( t \) and \( t' \) are symmetrically aligned

(6.9) \[ E(U_{0t} | c + \beta_t + U_{0k}) = E(U_{0t'} | c + \beta_{k} + U_{0k}), \]

Some evidence of non-stationary wage growth
presented by Farber and Gibbons (1996),

\[(6.10) \quad Y_{0t} = \beta_t + \eta + \sum_{j=0}^{t} \nu_j, \]

\[E(\eta) = 0.\]

\[E(U_{0t} | c + \beta_t + U_{0k}) = \frac{\sigma_{\eta}^2 + k\sigma_{\nu}^2}{\sigma_c^2 + \sigma_{\eta}^2 + k\sigma_{\nu}^2} (c + U_{0k} - E(c))\]

and \[E(U_{0t'} | c + \beta_t + U_{0k}) = \frac{\sigma_{\eta}^2 + t'\sigma_{\nu}^2}{\sigma_c^2 + \sigma_{\eta}^2 + t\sigma_{\nu}^2} (c + U_{0k} - E(c)).\]

\[Y_{0t} = \mu_{0t}(X) + \eta + U_{0t},\]

where
\[ U_{0t} = \sum_{j=1}^{k} \rho_{0j} U_{0,t-j} + \sum_{j=1}^{m} m_{0j} \nu_{t-j}, \]

There may also be “time effects,” so that

\[ \beta_t \neq \beta_{t'}, \]
6.3.2 A Separable Representation

\[ Y_{0t} = g_{0t}(X) + U_{0t} \]
\[ Y_{1t} = g_{1t}(X) + U_{1t}. \]

(6.11a) \[ E(Y_{0t} \mid D = 0, X) = g_{0t}(X) \]
\[ + E(U_{0t} \mid D = 0, X). \]

(6.11b) \[ E(Y_{1t} \mid D = 1, X) = g_{1t}(X) \]
\[ + E(U_{1t} \mid D = 1, X). \]

(P.1) The random utility representation (6.5) is valid.
\[(P.2) \quad (U_{0t}, U_{1t}, V) \perp \perp X,\]

(" \perp \perp " denotes stochastic independence)

\[(P.3)\text{ the distribution of } V, F(V) \text{ is strictly increasing in } V.\]

\[(6.12a) E(U_{0t} \mid D=1, X) = K_{0t}(Pr(D=1 \mid X)).\]

and

\[(6.12b) E(U_{1t} \mid D=1, X) = K_{1t}(Pr(D=1 \mid X)).\]

Assume that \( U_{0t}, V \) are jointly continuous random variables, with density \( f(U_{0t}, V \mid X). \)
From (P.2)

\[ f(U_0t, V \mid X) = f(U_0t, V). \]

Thus

\[
E(U_0t \mid X, D = 1) = \int_{-\infty}^{\infty} U_0t \int_{-\infty}^{\infty} f(U_0t, V) dU_0t dV \\
\hspace{1cm} \underbrace{\int_{-\infty}^{\infty} f(V) dV}_{H(X)} \hspace{1cm} \underbrace{\int_{-\infty}^{\infty} H(X)}_{H(X)}
\]

Now \( \Pr(D=1|X) = \int_{-\infty}^{\infty} f(V) dV. \) Inverting, we obtain
\[ H(X) = F_V^{-1}(\Pr(D = 1 \mid X)). \]

Thus

\[
E(U_{0t} \mid X, D=1) \quad \infty 
\int U_{0t} 
\int f(U_{0t}, V) dV dU_{0t} 
\]

\[
= \frac{-\infty}{-\infty} \int \frac{f(U_{0t}, V) dV dU_{0t}}{\Pr(D=1 \mid X)} 
\]

\[ \text{def } K_{0t}(\Pr(D = 1 \mid X)). \]

6.3.3. Variable Treatment Effect

\((Y_{0t}, Y_{1t})\) a pair of random variables:

\[ \alpha_t = Y_{1t} - Y_{0t}, \quad t > k \]
Treatment on Treated Is the Same as ATE

When

\[ E(\alpha_t \mid D = 1, I_k) = E(\alpha_t \mid D = 0, I_k) \]

\[ = E(\alpha_t \mid I_k). \]

If

\[ E[E(\alpha_t \mid I_k) \mid X, D = 1] = \]

\[ E[E(\alpha_t \mid I_k) \mid X, D = 0], \]

\[ E[E(\alpha_t \mid I_k, D = 1) \mid X, D = 1] = \]

\[ E[E(\alpha_t \mid I_k, D = 0) \mid X, D = 0]. \]
As long as the ex-post objective expectation of the subjective expectations is the same,

\[ E(\alpha_t \mid I_k, D = 1) = E(\alpha_t \mid I_k, D = 0) = \bar{\alpha}_t \]

\[ = E(\alpha_t \mid I_k) \]

Ex post surprise \((\alpha_t - \bar{\alpha}_t)\) does not determine \(D\):

\[ E(\alpha_t - \bar{\alpha}_t \mid X, D = 1) = 0. \]

So

\[ E(Y_{1t} - Y_{0t} \mid X, D = 1) = \bar{\alpha}_t. \]

77
Absence of bias for one parameter are different from

\[ E(U_{1t} - U_{0t} \mid X, D=1) = \]

\[ E(\Delta_t \mid X, D=1) - E(\Delta_t \mid X). \]

avoid bias for both mean parameters is if

\[ E(U_{1t} - U_{0t} \mid X, D = 1) = 0. \]
6.3.4 Imperfect Credit Markets

\[(6.13)\quad D = 1 \text{ if }\]
\[
E \left[ \sum_{j=1}^{T-k} \beta^j \{ G(Y_{1,k+j} + \eta_{k+j}) - G(Y_{0,k+j} + \eta_{k+j}) \} \right] 
+ G(\eta_k - c_k) - G(Y_{0k} + \eta_k) \mid I_k 
> 0; \text{ otherwise}\]

Difference in Differences

Parameter is not identified:

\[E(U_{0t} \mid X, D = 1) \neq 0\]

even though \( E(U_{0t'} \mid X, D = 1) = 0, \)

so \( E(U_{0t} - U_{0t'} \mid X, D = 1) \neq 0. \)
Training As A Form of Job Search

Program \( j \), wage offers arrive from a distribution \( F_j \) at rate \( \lambda_j \). Persons pay \( c_j \) to sample from \( F_j \).

At any point in time, persons pick the search option with the highest expected return.

In the unemployed state, a person receives a nonmarket benefit, \( N \).

"Gittens Index" \( V_{ju} = \)
\[ N - c_j + \frac{\lambda_j}{1 + r} E_j \max[V_{je}; V_{ju}] + \frac{(1 - \lambda_j)}{1 + r} V_{ju}. \]

Value of Nonmarket time
\[ V_{je} = Y_j + \frac{(1 - \sigma_{je})}{1 + r} V_{je} \]

\[ + \frac{\sigma_{je}}{1 + r} E_j \left[ \max(V_N, V_{ju}) \right] \]

\[ V_{je} = Y_j + \frac{(1 - \sigma_{je})}{1 + r} V_{je} + \frac{\sigma_{je}}{1 + r} V_{ju} \]

so

\[ V_{je} = \frac{\sigma_{je}}{r + \sigma_{je}} V_{ju} + \frac{(1 + r)Y_j}{r + \sigma_{je}}. \]
\[ V_{ju} = \frac{(1+r)(N-c_j) + \lambda_j E_{j}(V_{je} \mid V_{je} > V_{ju})}{r + \lambda_j \Pr(Y_j > V_{ju}(r/(1+r))) \cdot \Pr(Y_j > V_{ju}(r/(1+r)))} \]

The optimal search strategy is

\[ \hat{j} = \arg \max_j \{V_{ju}\} \]

\[ V_{ju} > V_N \text{ for at least one } j. \quad T_j = t_j \]
\[
\Pr(T_j \geq t_j) = [1 - \lambda_j(1 - F_j(V_{ju}(r/(1+r))))]^t_j
\]

\[
1 - \lambda_j(1 - F_j(V_{ju}(r/1 + r)), (1 - \lambda_j)
\]

\[
(\lambda_j F_j(V_{ju}(r/1 + r)).
\]
The Role of Program Eligibility Rules

In Determining Participation

\[(6.15) \quad \frac{\alpha}{r} - Y_{ik} - c_i > 0 \text{ and the eligibility rules } Y_{i,k-1} < K\]

Administrative Discretion and the Efficiency and Equity of Training Provision
The Conflict Between The Economic Approach to Program Evaluation And The Modern Approach to Social Experiments

Under ideal conditions, social experiments identify \( E(Y_1 - Y_0|X, D = 1) \).

Social experiments balance bias, do not eliminate the \( U_0 \) and \( D \) or \( U_1 \) and \( D \).

\[
f(y_0|X, D = 1) \quad \text{and} \quad f(y_1|X, D = 1).
\]

\[
E(Y_0|X, D = 1) = g_0(X) + E(U_0|X, D = 1)
\]
and

\[ E(Y_1|X, D = 1) = g_1(X) + E(U_1|X, D = 1). \]