Modeling Selection

Huan Deng

April 29, 2025

Outline

Introduction

The Roy Model
- Basics
- Parametric Identification of Roy Model
- Nonparametric Identification

Generalized Roy Model

Understanding Treatment Effect via the Generalized Roy Model
- Set-up
- LATE
- Marginal Treatment Effects (MTE)

Introduction

Motivation

- We have discussed identification and estimation of discrete choice models.
- Next, we will consider two canonical discrete choice models: the Roy model and BLP models.
- The Roy model is a workhorse model in labor economics and it deals with a key problem in causal identification: self-selection.
- The BLP model is a demand model with endogeneity, which is a workhorse model in empirical IO.
- Whatever field you choose, you will at least weakly benefit from understanding these two models.
- This lecture is about Roy Model, and we will cover BLP model in a separate lecture.
- This lecture closely follows French and Taber (2011), a chapter from the Handbook of Labor Economics. One of my favorite handbook chapters, highly recommended!

The Roy Model

The Roy Model

Basics

Motivation

- The classic **Roy Model** (Roy, 1951) studies how individuals select among different sectors (e.g., fishing vs. hunting) to maximize their payoff (often wage).
- This seminal framework underlies many *binary choice* scenarios in labor:
  - Work vs. not work (labor force participation),
  - Occupation choices,
  - Migration decisions,
  - Union vs. non-union,
  - marriage, fertility, etc.

Motivation

- Key **identification question**: How can we recover the *joint distribution of potential outcomes* (Y_f, Y_h) if we only observe Y_i in *one* occupation per individual?
- We introduce key assumptions, show parametric identification, and then present *nonparametric identification* results (Heckman and Honoré, 1990).

Basic Roy Model Setup

Two potential log wages:

$$Y_{fi} \quad \text{(log wage in fishing)} \quad \text{and} \quad Y_{hi} \quad \text{(log wage in hunting)}.$$

- $J_i \in \{f, h\}$ is the chosen occupation.
- **Outcome** we observe:

$$Y_i = \begin{cases} Y_{fi} & \text{if } J_i = f, \\ Y_{hi} & \text{if } J_i = h. \end{cases}$$

- The Roy model assumes $J_i = f$ if and only if $Y_{fi} > Y_{hi}$ (i.e. choose the occupation with higher payoff).

Basic Roy Model Setup

Parametric specification:

$$
\begin{aligned}
Y_{fi} &= X_{fi}'\gamma_{ff} + X_{0i}'\gamma_{0f} + \varepsilon_{fi},\\
Y_{hi} &= X_{hi}'\gamma_{hh} + X_{0i}'\gamma_{0h} + \varepsilon_{hi},\\
\begin{pmatrix} \varepsilon_{fi} \\ \varepsilon_{hi} \end{pmatrix} &\sim \mathcal{N}(\mathbf{0}, \Sigma).
\end{aligned}
$$

But more generally we examine *nonparametric* $g_f(\cdot)$, $g_h(\cdot)$ and error distributions.

Identification Challenges

What is the core selection problem?

- We *never* observe (Y_{fi}, Y_{hi}) jointly for the same person. Instead, only Y_{fi} if $J_i = f$, or Y_{hi} if $J_i = h$.
- So to understand how wages *would have* changed if a fisherman switched to hunting, we must impute Y_{hi} from the distribution. This is a *counterfactual* wage.

Conditions for identification:

- If we assume parametric forms (*e.g. linear, normal errors*) plus certain exclusion restrictions, we can identify $\gamma_{ff}, \gamma_{0f}, \gamma_{hh}, \gamma_{0h}$ and Σ.
- Without parametric assumptions, we typically need:
  - *Exclusion restrictions* or price/instrument variation that only affects one sector,
  - *Full support* or "identification at infinity" arguments.

The Roy Model

Parametric Identification of Roy Model

Step 1: Reduced-Form Probit

Choice equation: $J_i = f$ if $Y_{fi} > Y_{hi}$, i.e.

$$X_{fi}'\gamma_{ff} - X_{hi}'\gamma_{hh} + X_{0i}'(\gamma_{0f} - \gamma_{0h}) + (\varepsilon_{fi} - \varepsilon_{hi}) > 0.$$

Define

$$\sigma^* \;=\; \sqrt{\sigma_f^2 + \sigma_h^2 - 2\,\sigma_{fh}}, \quad \gamma^* \;=\; \frac{(\gamma_{ff},\, -\gamma_{hh},\, \gamma_{0f} - \gamma_{0h})}{\sigma^*}.$$

Then

$$\Pr(J_i = f \mid X_i) = \Phi\Big(\frac{X_{fi}'\gamma_{ff} - X_{hi}'\gamma_{hh} + X_{0i}'(\gamma_{0f} - \gamma_{0h})}{\sigma^*}\Big) = \Phi(X_i'\gamma^*),$$

which is a standard probit. We can estimate γ^* via Maximum Likelihood.
Note: We only get γ^* up to the scale σ^* here; σ^* is not identified from the choice equation alone.

Step 2: Heckman Two-Step for the Wage Equation

Heckman's (1979) two-step correction:

- We observe Y_{fi} *only* when $J_i = f$.
- The OLS regression of Y_{fi} on X_{fi}, X_{0i} would be *biased* due to sample selection ($J_i = f$ is not random).

Conditional expectation (if $J_i = f$):

$$E(Y_{fi} \mid J_i = f, X_i) = X_{fi}'\gamma_{ff} + X_{0i}'\gamma_{0f} \; + \; E(\varepsilon_{fi} \mid \varepsilon_{fi} - \varepsilon_{hi} > -X_i'\tilde{\gamma}^*),$$

where $\tilde{\gamma}^* \equiv (\gamma_{ff}, -\gamma_{hh}, \gamma_{0f} - \gamma_{0h})/\sigma^*$.
Inverse Mills Ratio (λ):

$$\lambda(\alpha) = \frac{\phi(\alpha)}{\Phi(\alpha)}, \quad \text{where } \phi \text{ and } \Phi \text{ are pdf \& cdf of the standard normal.}$$

We use $\lambda\left(X_i'\hat{\gamma}^*\right)$ to correct for the selection.

Heckman Two-Step (cont.)

$$E(\varepsilon_{fi} \mid J_i = f, X_i) = -\tau\,\lambda\big(X_i'\gamma^*\big), \quad \text{where} \quad \tau = \frac{\mathrm{Cov}(\varepsilon_{fi}, \varepsilon_{hi} - \varepsilon_{fi})}{\sigma^*}.$$

Hence

$$E(Y_{fi} \mid J_i = f, X_i) = X_{fi}'\gamma_{ff} + X_{0i}'\gamma_{0f} \; - \; \tau\,\lambda\big(X_i'\gamma^*\big).$$

Estimation:

1. Estimate the probit in Step 1, get $\hat{\gamma}^*$.
2. Construct $\lambda\big(X_i'\hat{\gamma}^*\big)$ for each individual who chose f.
3. Regress Y_{fi} on $(X_{fi}, X_{0i}, \lambda(\dots))$ using only $J_i = f$ sample.

Note: we don't need exclusion restrictions to identify the above equation in the parametric case, since λ is a nonlinear function. However, it is always better to have at least one variable in X_h that is excluded from the Y_f equation.

Step 3: Structural Probit

Goal: Identify σ^* and $\{\gamma_{hh}, \gamma_{0h}\}$, having already obtained $(\hat{\gamma}_{ff}, \hat{\gamma}_{0f}, \hat{\tau})$ from Step 2.

Structural Probit:

$$\Pr(J_i = f \mid X_i = x) = \Phi\left(\frac{1}{\sigma^*}\left(x_f'\hat{\gamma}_{ff} + x_0'\hat{\gamma}_{0f}\right) - x_h'\frac{\gamma_{hh}}{\sigma^*} - x_0'\frac{\gamma_{0h}}{\sigma^*}\right).$$

- That is, one just runs a probit of J_i on $(X_{fi}'\hat{\gamma}_{ff} + X_{0i}'\hat{\gamma}_{0f})$, X_{0i}, and X_{hi}, where $\hat{\gamma}_{ff}$ and $\hat{\gamma}_{0f}$ are our estimates of γ_{ff} and γ_{0f}.

Step 4: Identifying $\sigma_f^2, \sigma_h^2, \sigma_{fh}$

Finally we want: $(\sigma_f^2, \sigma_h^2, \sigma_{fh})$.
We already have:

$$\sigma^* = \sqrt{\sigma_f^2 + \sigma_h^2 - 2\sigma_{fh}}, \qquad \tau = \frac{\sigma_{fh} - \sigma_f^2}{\sigma^*}.$$

These give us two equations in three parameters. The third comes from the residual variance in the selection equation:

$$\text{var}(\varepsilon_{fi} \mid J_i = f, X_i = x) = \sigma_f^2 + \tau^2 \left[-\lambda(x'\gamma^*)x'\gamma^* - \lambda^2(x'\gamma^*) \right].$$

Let $i = 1, \ldots, N_f$ index those with $J_i = f$, and let $\hat{\varepsilon}_{fi}$ be the residual:

$$\hat{\varepsilon}_{fi} = Y_{fi} - X_{fi}'\hat{\gamma}_{ff} - X_{0i}'\hat{\gamma}_{0f}.$$

Then we estimate:

$$\hat{\sigma}_f^2 = \frac{1}{N_f} \sum_{i=1}^{N_f} \left(\hat{\varepsilon}_{fi} + \hat{\tau}\lambda(X_i'\hat{\gamma}^*)\right)^2 - \hat{\tau}^2 \left(-\lambda(X_i'\hat{\gamma}^*)X_i'\hat{\gamma}^* - \lambda^2(X_i'\hat{\gamma}^*) \right),$$

$$\hat{\sigma}_{fh} = \hat{\sigma}_f^2 + \hat{\tau}\hat{\sigma}^*, \qquad \hat{\sigma}_h^2 = \hat{\sigma}^{*2} - \hat{\sigma}_f^2 + 2\hat{\sigma}_{fh}.$$

The Roy Model

Nonparametric Identification

Model Setup

- Now consider a nonparametric specification:

$$Y_{fi} = g_f(X_{fi}, X_{0i}) + \varepsilon_{fi} \tag{1}$$

$$Y_{hi} = g_f(X_{hi}, X_{0i}) + \varepsilon_{hi} \tag{2}$$

- $J_i \in \{f, h\}$ is the chosen occupation.
- **Outcome** we observe:

$$Y_i = \begin{cases} Y_{fi} & \text{if } J_i = f, \\ Y_{hi} & \text{if } J_i = h. \end{cases}$$

- $J_i = f$ if and only if $Y_{fi} > Y_{hi}$ (i.e. choose the occupation with higher payoff).
- Next, we introduce the identification assumptions.

Assumption 1

Assumption (On Error Distribution)

- $(\varepsilon_{fi}, \varepsilon_{hi})$ *is continuously distributed with distribution function* G *on* $\mathbb{R}^2$.
- *Independent of* X_i *(the observed covariates).*
- *The marginal distributions of* ε_{fi} *and* $(\varepsilon_{fi} - \varepsilon_{hi})$ *have medians equal to zero.*

Assumption 2

Assumption (On Support)

Full Support of g_f and g_h:

$$supp\big(g_f(X_{fi}, x_0),\ g_h(X_{hi}, x_0)\big) = \mathbb{R}^2 \quad \text{for all } x_0 \in supp(X_{0i}).$$

- *For any fixed value of $g_h(x_h, x_0)$, $g_f(x_f, x_0)$ spans the entire real line, and vice versa.*
- *Practical meaning: we can find covariates that make either occupation/treatment almost surely chosen.*
- *This is a strong requirement and underpins nonparametric identification.*

Assumptions 3 and 4

Assumption (On Decomposability of $\boldsymbol{X_i}$)

$$X_i = (X_{fi}, X_{hi}, X_{0i}) = (X_{fi}^c, X_{fi}^d,\ X_{hi}^c, X_{hi}^d,\ X_{0i}^c, X_{0i}^d),$$

- X^c = *continuously distributed parts (no point mass).*
- X^d = *discrete parts (all support points have positive mass).*

Assumption (On Continuity across $\boldsymbol{x^c}$)

- *For each discrete support point $(x_{fi}^d, x_{hi}^d, x_{0i}^d)$, the functions $g_f(\cdot)$ and $g_h(\cdot)$ vary almost surely continuously in x^c.*

Theorem on Identification (Heckman & Honoré (1990))

Statement:

- Suppose each individual i can choose $J_i \in \{f, h\}$, and we observe Y_i if $J_i = f$, along with X_i.
- These are generated by the Roy model.
- Under Assumptions 1–4, the functions g_f, g_h and the distribution G (of $(\varepsilon_{fi}, \varepsilon_{hi})$) are *identified* on a set X^* of measure 1.

Proof Sketch for Nonparametric Identification

1. Identification of the Choice Rule up to a Monotone Transform.

- $J_i = f$ iff
$$g_f(X_{fi}, X_{0i}) \; - \; g_h(X_{hi}, X_{0i}) \; > \; \varepsilon_{hi} - \varepsilon_{fi}.$$
- Using data only on choices, this model is only identified up to a monotonic transformation.
- Let F be the CDF of $\varepsilon_{hi} - \varepsilon_{fi}$. Define $\varepsilon_i \equiv F(\varepsilon_{hi} - \varepsilon_{fi}) \in [0,1]$.
- Then $J_i = f$ iff $\varepsilon_i < F\left[g_f(X_{fi}, X_{0i}) - g_h(X_{hi}, X_{0i})\right]$.
- Denote $g(x_i) = F\left[g_f(X_{fi}, X_{0i}) - g_h(X_{hi}, X_{0i})\right]$, then we have:

$$\begin{aligned}
\Pr(J_i = f \mid X_i = x) &= \Pr(g_f(x_f, x_0) - g_h(x_h, x_0) > \varepsilon_{hi} - \varepsilon_{fi}) && (3)\\
&= \Pr(F(g_f(x_f, x_0) - g_h(x_h, x_0)) > F(\varepsilon_{hi} - \varepsilon_{fi})) && (4)\\
&= \Pr(\varepsilon_i < g(x)) && (5)\\
&= g(x). && (6)
\end{aligned}$$

- Hence the reduced-form choice probability $\Pr(J_i = f | X_i)$ identifies that function *as a cdf in* ν_i. By normalizing $\varepsilon_i \sim \mathrm{Uniform}(0,1)$, we fix the monotonic transform.

Proof Sketch (cont.)

2. Identifying $g_f(\cdot)$ using "Identification at Infinity."

- Observe $\text{Med}(Y_{fi} \mid X_i = x, J_i = f) = g_f(x_f, x_0) + \text{Med}(\varepsilon_{fi} \mid \varepsilon_i < g(x))$
- If we pick x_h so that $g(x) \approx 1$ (i.e., everyone chooses f), then $\varepsilon_{fi} \mid \varepsilon_i < g(x)$ is essentially ε_{fi} unconditionally. Under $\text{Med}(\varepsilon_{fi}) = 0$, we get $g_f(x_f, x_0)$ directly.

3. Identifying $g_h(\cdot)$.

- For any (x_h, x_0), Assumption 2 guarantees that we can find x_f so that $g(x) = 0.5$.
- By the median property, $g_f(x_f, x_0) - g_h(x_h, x_0) = \text{Med}(\varepsilon_h - \varepsilon_f) = 0$, hence $g_h(x_h, x_0) = g_f(x_f, x_0)$.

Proof Sketch (cont.)

4. Distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$.

- With g_f and g_h known, from data we have:

$$\begin{aligned}\Pr(J_i = f, Y_{fi} < s \mid X_i = x) &= \Pr(g_h(x_h, x_0) + \varepsilon_{hi} \leq g_f(x_f, x_0) + \varepsilon_{fi},\ g_f(x_f, x_0) + \varepsilon_{fi} \leq s) \\ &= \Pr(\varepsilon_{hi} - \varepsilon_{fi} \leq g_f(x_f, x_0) - g_h(x_h, x_0),\ \varepsilon_{fi} \leq s - g_f(x_f, x_0)) \quad (7)\end{aligned}$$

- which is the cumulative distribution function of $(\varepsilon_{hi} - \varepsilon_{fi},\ \varepsilon_{fi})$ evaluated at the point $(g_f(x_f, x_0) - g_h(x_h, x_0),\ s - g_f(x_f, x_0))$.
- By varying the point of evaluation one can identify the joint distribution of $(\varepsilon_{hi} - \varepsilon_{fi},\ \varepsilon_{fi})$ from which one can derive the joint distribution of $(\varepsilon_{fi},\ \varepsilon_{hi})$.

Generalized Roy Model

Why a Generalized Roy Model?

- **Classic Roy Model:** Individuals choose between two sectors (f and h) solely on the basis of expected wage ($Y_{fi} > Y_{hi}$).
- **Generalized Roy Model:** Allows non-pecuniary factors to also enter the decision:

$$U_{fi} = Y_{fi} + \phi_f(Z_i, X_{0i}) + \nu_{fi}, \quad U_{hi} = Y_{hi} + \phi_h(Z_i, X_{0i}) + \nu_{hi}.$$

  - Z_i affects tastes or utility only, but not wages. - X_{fi}, X_{hi}, X_{0i} can shift sector-specific wages.
- Observed data: $(J_i, Y_i, X_{fi}, X_{hi}, X_{0i}, Z_i)$ with $Y_i = Y_{fi}$ if $J_i = f$ and $Y_i = Y_{hi}$ if $J_i = h$.
- **Goal:** Nonparametrically identify the wage functions g_f, g_h, the preference functions ϕ_f, ϕ_h, and the joint distribution of unobserved components $(\varepsilon_{fi}, \varepsilon_{hi}, \nu_{fi}, \nu_{hi})$.

Wage Equations and Utility

Wage Equations:

$$Y_{fi} = g_f(X_{fi}, X_{0i}) + \varepsilon_{fi}, \quad Y_{hi} = g_h(X_{hi}, X_{0i}) + \varepsilon_{hi}.$$

- X_{fi}: variables that enter fishing wage only.
- X_{hi}: variables that enter hunting wage only.
- X_{0i}: variables that may shift *both* sectors' wages.
- $(\varepsilon_{fi}, \varepsilon_{hi})$: unobserved skill or productivity components.

Preference (Non-Pecuniary) Components:

$$\phi_f(Z_i, X_{0i}), \quad \phi_h(Z_i, X_{0i}), \quad \nu_{fi}, \nu_{hi} \text{ as unobserved taste shocks.}$$

Utility:

$$U_{fi} = Y_{fi} + \phi_f(Z_i, X_{0i}) + \nu_{fi}, \quad U_{hi} = Y_{hi} + \phi_h(Z_i, X_{0i}) + \nu_{hi}.$$

Choice and Observables

- **Choice:** $J_i = f$ if $U_{fi} > U_{hi}$, i.e.

  $$g_f(X_{fi}, X_{0i}) - g_h(X_{hi}, X_{0i}) + \phi_f(Z_i, X_{0i}) - \phi_h(Z_i, X_{0i}) > (\varepsilon_{hi} + \nu_{hi}) - (\varepsilon_{fi} + \nu_{fi})$$

- **Observed wage:**

  $$Y_i = \begin{cases} Y_{fi}, & \text{if } J_i = f, \\ Y_{hi}, & \text{if } J_i = h. \end{cases}$$

- **Data Gaps:** We never observe Y_{hi} for those who pick f (and vice versa). Non-pecuniary utilities are also never directly observed.

Identification Question: Under what conditions can we recover

$$g_f,\ g_h,\ \phi_f,\ \phi_h,\ \text{ and } F(\varepsilon_{fi}, \varepsilon_{hi}, \nu_{fi}, \nu_{hi})$$

from observed (J_i, Y_i, X_i, Z_i)?

Reduced Form and the Uniform Normalization

Denote F^* as the distribution function of the composite error

$$\varepsilon_{h_i} + \nu_{h_i} - (\varepsilon_{f_i} + \nu_{f_i}),$$

we introduce

$$\nu_i = F^*\left(\varepsilon_{h_i} + \nu_{h_i} - \varepsilon_{f_i} - \nu_{f_i}\right),$$

and set ν_i to be uniformly distributed on $[0, 1]$. We define:

$$\phi(Z_i, X_i) \;=\; F^*\big(g_f(X_{f_i}, X_{0i}) + \phi_f(Z_i, X_{0i}) - g_h(X_{h_i}, X_{0i}) - \phi_h(Z_i, X_{0i})\big),$$

Now we can show that: $J_i = f$ iff $\phi(Z_i, X_i) > \nu_i$

List of Assumptions

Assumption 1 (Error distributions):

- $(\nu_i, \varepsilon_{f_i}, \varepsilon_{h_i})$ is continuously distributed and independent of (Z_i, X_i).
- ν_i is uniform on $[0, 1]$.
- $\text{med}(\varepsilon_{f_i}) = 0$ and $\text{med}(\varepsilon_{h_i}) = 0$.

Assumption 2 (Support condition for ϕ):

- The support of $\phi(Z_i, x)$ is $[0, 1]$ for all $x \in \text{supp}(X_i)$.

Assumption 3 (Discrete and continuous parts):

- $(Z_i, X_i) = (Z_i^c, Z_i^d, X_f^c, X_f^d, X_h^c, X_h^d, X_0^c, X_0^d)$ with certain components continuous (no point mass) and others discrete (all support points have positive mass).

Assumption 4 (Continuity in arguments):

- g_f, g_h, ϕ_f, ϕ_h are almost surely continuous across the relevant supports.

Key Identification Result

Statement: Under Assumptions 1–4,

- ϕ, g_f, g_h,
- the joint distribution of $(\nu_i, \varepsilon_{f_i})$, and the joint distribution of $(\nu_i, \varepsilon_{h_i})$

are identified almost everywhere from the joint distribution of (J_i, Y_i) and observables (Z_i, X_i).

Step 1: Identification of the Choice Model (ϕ)

What We Observe

$\Pr(J_i = f \mid Z_i = z, X_i = x)$ is directly estimable from the data (sample f

Key Normalization

$$\nu_i \sim \text{Uniform}(0,1).$$

This implies

$$\Pr\big(J_i = f \mid Z_i = z, X_i = x\big) = \Pr\big(\nu_i \leq \phi(z,x)\big) = \phi(z,x).$$

The function $\phi(z,x)$ is **identified** by the conditional probability of choosing occupation f.

Step 2: Identification of g_f and g_h

Earnings in Occupation f:

$$Y_i \;=\; g_f(X_{f_i}, X_{0i}) + \varepsilon_{f_i} \quad \text{if } J_i = f.$$

Key trick: Look at $\mathrm{Med}(Y_i \mid J_i = f, Z_i = z, X_i = x)$ as $\phi(z,x) \to 1$.

$$\lim_{\phi(z,x)\to 1} \mathrm{Med}\big(Y_i \mid J_i = f, Z_i = z, X_i = x\big) \;=\; g_f(x_f, x_0) + \lim_{\phi(z,x)\to 1} \mathrm{Med}\big(\varepsilon_{f_i} \mid \nu_i \le \phi(z,x)\big).$$

Since ν_i is uniform and independent, as $\phi(z,x) \to 1$, we are conditioning on almost everyone in occupation f. The median of ε_{f_i} (given these assumptions) goes to the unconditional median, which is assumed to be 0. Hence:

$$\lim_{\phi(z,x)\to 1} \mathrm{Med}(Y_i \mid J_i = f, Z_i = z, X_i = x) \;=\; g_f(x_f, x_0).$$

Thus we can **identify** g_f. Similarly, we can also **identify** g_h

Step 3: Joint Distribution $(\nu_i, \varepsilon_{f_i})$ and $(\nu_i, \varepsilon_{h_i})$

Idea: We know the occupation choice depends on $\nu_i \leq \phi(z, x)$.

$$\Pr(J_i = f, Y_f^i \leq y \mid Z_i = z, X_i = x) \;=\; \Pr\big(\nu_i \leq \phi(z, x),\; g_f(x_f, x_0) + \varepsilon_{f_i} \leq y\big).$$

Hence:

$$\Pr(J_i = f, Y_f^i \leq y \mid Z_i = z, X_i = x) \;=\; G_{\nu,\varepsilon_f}\big(\,\phi(z, x),\; y - g_f(x_f, x_0)\big),$$

where G_{ν,ε_f} is the CDF of $(\nu_i, \varepsilon_{f_i})$.

Conclusion of Step 4

From variation in $\phi(z, x)$ and the distribution of Y_f^i, we **identify the joint distribution** G_{ν,ε_f}. A similar argument identifies G_{ν,ε_h}.

Additional Assumptions for Identification

To recover the non-pecuniary component difference $\phi_f - \phi_h$ fully, we introduce:

Assumption 5 (Median restriction):

$$\text{med}\Big(\varepsilon_{h_i} + \nu_{h_i} - \varepsilon_{f_i} - \nu_{f_i}\Big) \;=\; 0.$$

Assumption 6 (Full support for wage differences):

$g_f(X_{f_i}, x_0) - g_h(X_{h_i}, x_0)$ has full real support for all $(z, x_0) \in \text{supp}(Z_i, X_{0i})$

These allow us to pin down how much of the utility difference is due to $\phi_f - \phi_h$ versus earnings differences.

Identification of Non-Pecuniary Difference $\phi_f - \phi_h$

Sketch of the argument:

$$0.5 = \Pr(J_i = f \mid Z_i = z, X_i = x) = \Pr\Big(\nu_i \leq g_f - g_h + \phi_f - \phi_h\Big).$$

So,

$$g_f(x_f, x_0) - g_h(x_h, x_0) + \big(\phi_f(z, x_0) - \phi_h(z, x_0)\big) = \text{median of } \nu_i,$$

which is zero by assumption. Thus

$$\phi_f(z, x_0) - \phi_h(z, x_0) \;=\; g_h(x_h, x_0) - g_f(x_f, x_0).$$

This pins down the non-pecuniary difference $\phi_f - \phi_h$. Furthermore, F^* can also be identified.

What is *Not* Identified

- The *joint distribution* of $\left(\varepsilon_{f_i}, \varepsilon_{h_i}\right)$ is **not** identified.
- Even with parametric assumptions (e.g. bivariate normal), we cannot recover $\mathrm{Cov}(\varepsilon_f, \varepsilon_h)$ from choice and wage data alone, because each person's wage is only observed in one chosen occupation.

Understanding Treatment Effect via the Generalized Roy Model

Understanding Treatment Effect via the Generalized Roy Model

Set-up

Treatment Effects Setup

Defining a Treatment Effect for individual i:

$$\pi_i = Y_{f_i} - Y_{h_i},$$

where Y_{f_i} is the outcome under "treatment" (f) and Y_{h_i} the outcome otherwise (h).

Key Roy-type Insight:

- Individuals self-select into f vs. h. Observing Y_{f_i} only when f is chosen (and Y_{h_i} only when h is chosen) complicates identification.
- Need an "exclusion restriction" (instrument) Z_i that shifts the choice probability without directly altering outcomes.
- Since the joint distribution of $\left(\varepsilon_{f_i}, \varepsilon_{h_i}\right)$ is not identified, we can't identify the distribution of π_i

Constant vs. Heterogeneous Effects

A special case:

$$g_f(X_{f_i}, X_{0_i}) \;=\; g_h(X_{h_i}, X_{0_i}) + \pi_0, \quad \varepsilon_{f_i} = \varepsilon_{h_i}.$$

- Then $\pi_i = \pi_0$ (constant).
- Often unrealistic; real-world treatments have *heterogeneous* impacts.

ATE (Average Treatment Effect):

$$\text{ATE} = \mathbb{E}(\pi_i) = \mathbb{E}\big(Y_{f_i}\big) - \mathbb{E}\big(Y_{h_i}\big).$$

- Though we can't identify the distribution of π_i or summary statistics in general, we have some hope of identifying ATE.
- Identification typically requires observing (or inferring) $\mathbb{E}(Y_{f_i})$ and $\mathbb{E}(Y_{h_i})$ under exogenous variation.

Identification of ATE

- Assume Y_{fi} and Y_{hi} have finite expectation.
- Abstracting away from X_i, let $\varphi(z)$ be the probability of choosing occupation f given $Z_i = z$.
- We focus on the limits $\varphi(z) \to 1$ (all choose f) and $\varphi(z) \to 0$ (all choose h):

$$\begin{aligned} &\lim_{\varphi(z)\to 1} \mathbb{E}(Y_{fi} \mid Z_i = z, J_i = f) - \lim_{\varphi(z)\to 0} \mathbb{E}(Y_{hi} \mid Z_i = z, J_i = h) \\ &= \lim_{\varphi(z)\to 1} \mathbb{E}(Y_{fi} \mid \nu_i \le \varphi(z)) - \lim_{\varphi(z)\to 0} \mathbb{E}(Y_{hi} \mid \nu_i > \varphi(z)) \\ &= \mathbb{E}(Y_{fi}) - \mathbb{E}(Y_{hi}). \end{aligned} \tag{8}$$

Why It Matters:

- Allows us to pin down $E(Y_{fi}) - E(Y_{hi})$ by taking the difference.
- Requires that $\varphi(Z_i)$ actually attains values near 0 and 1 in the data.
- Hence called *identification at infinity*: we need "extreme" groups who *always* select occupation f or h.

Implications for Empirical Work

- We need variables Z_i that shift $\varphi(Z_i)$ close to 0 or 1 to identify the average wage outcomes.
- **No Variation Near 0 or 1** $\implies$ Incomplete identification:
  - If $\varphi(Z_i) < \varphi^u < 1$, we never see enough people in the "fisherman" occupation to learn about $E(Y_{fi})$.
  - If $\varphi(Z_i) > \varphi^\ell > 0$, we never see enough people in "hunter" occupation to learn about $E(Y_{hi})$.
- **Conceptual Message:** Finding instruments (or experimental variation) that produce near-corner solutions ($\varphi \approx 0$ or 1) is crucial to identify ATE.

Understanding Treatment Effect via the Generalized Roy Model

LATE

Roy Model Notation for LATE

- Individuals choose $J_i = f$ if
$$\varphi(Z_i) > \nu_i,$$
else $J_i = h$. Here, ν_i is unobservable heterogeneity uniformly distributed on $[0, 1]$.
- Z_i is a binary instrument that shifts $\varphi(Z_i)$, the probability of choosing f.
- Observed outcome:
$$Y_i = \begin{cases} Y_{f_i}, & \text{if } J_i = f, \\ Y_{h_i}, & \text{if } J_i = h. \end{cases}$$

Objective: Derive *Local Average Treatment Effect* for those who *switch* occupation/treatment status as Z_i goes from 0 to 1.

Choice Groups

- We can observe choices under z:

$$\Pr(J_i = f \mid Z_i = 0) \; = \; \Pr(\nu_i \leq \varphi(0)) = \phi(0),$$

$$\Pr(J_i = f \mid Z_i = 1) \; = \; \Pr(\nu_i \leq \varphi(1) = \phi(1).$$

- Without loss of generality we will assume $\phi(1) > \phi(0)$, $\phi(0)$ and $\phi(1)$ separate $[0, 1]$ into three distinct groups:
  - **Never-takers**: $\nu_i > \varphi(1)$ (choose h either way).
  - **Always-takers**: $\nu_i \leq \varphi(0)$ (choose f either way).
  - **Compliers**: $\varphi(0) < \nu_i \leq \varphi(1)$ (switch if Z_i changes).
- The threshold-crossing property of the selection model implies the monotonicity assumption in LATE.

Expressing LATE

- We can identify below 6 objects from data:

$$\Pr(J_i = f \mid Z_i = 0) = \phi(0) \tag{9}$$

$$\Pr(J_i = f \mid Z_i = 1) = \phi(1) \tag{10}$$

$$\mathbb{E}(Y_i \mid Z_i = 0,\ J_i = f) = \mathbb{E}\big(Y_{f_i} \mid \nu_i \leq \varphi(0)\big) \tag{11}$$

$$\mathbb{E}(Y_i \mid Z_i = 0,\ J_i = h) = \mathbb{E}\big(Y_{h_i} \mid \nu_i > \varphi(0)\big) \tag{12}$$

$$\mathbb{E}(Y_i \mid Z_i = 1,\ J_i = f) = \mathbb{E}\big(Y_{f_i} \mid \nu_i \leq \varphi(1)\big) \tag{13}$$

$$\mathbb{E}(Y_i \mid Z_i = 1,\ J_i = h) = \mathbb{E}\big(Y_{h_i} \mid \nu_i > \varphi(1)\big) \tag{14}$$

- Now we want to express LATE using the above identified objects and prove:

$$\mathbb{E}(\pi_i \mid \varphi(0) < \nu_i \leq \varphi(1)) = \frac{\mathbb{E}(Y_i \mid Z_i = 1) - \mathbb{E}(Y_i \mid Z_i = 0)}{\Pr(J_i = f \mid Z_i = 1) - \Pr(J_i = f \mid Z_i = 0)} \tag{15}$$

- the left-hand side is LATE, and the right hand-side is the Wald estimator we have seen before.

Proof

- We derive $\mathbb{E}(Y_i \mid Z_i = 1)$ as:

$$\begin{aligned}
\mathbb{E}(Y_i \mid Z_i = 1) &= \mathbb{E}(Y_i \mid Z_i = 1,\, J_i = f)\varphi(1) + \mathbb{E}(Y_i \mid Z_i = 1,\, J_i = h)(1 - \varphi(1)) \\
&= \mathbb{E}\big(Y_{f_i} \mid \nu_i \leq \varphi(1)\big)\varphi(1) + \mathbb{E}\big(Y_{h_i} \mid \nu_i > \varphi(1)\big)(1 - \varphi(1)) \\
&= \mathbb{E}\big(Y_{f_i} \mid \nu_i \leq \varphi(0)\big)\varphi(0) + \\
&\quad \mathbb{E}\big(Y_{f_i} \mid \varphi(0) < \nu_i \leq \varphi(1)\big)(\varphi(1) - \varphi(0)) + \\
&\quad \mathbb{E}\big(Y_{h_i} \mid \nu_i > \varphi(1)\big)(1 - \varphi(1)) \qquad (16)
\end{aligned}$$

- Similarly, we can derive $\mathbb{E}(Y_i \mid Z_i = 0)$ as:

$$\begin{aligned}
\mathbb{E}(Y_i \mid Z_i = 0) = {} & \mathbb{E}\big(Y_{f_i} \mid \nu_i \leq \varphi(0)\big)\varphi(0) + \\
& \mathbb{E}\big(Y_{h_i} \mid \varphi(0) < \nu_i \leq \varphi(1)\big)(\varphi(1) - \varphi(0)) + \\
& \mathbb{E}\big(Y_{h_i} \mid \nu_i > \varphi(1)\big)(1 - \varphi(1)) \qquad (17)
\end{aligned}$$

- Substitute the above two equations into the Wald estimator, we are done.

Equivalence Theorem by Vytlacil (2002)

- So far we have shown that the selection model implies LATE. Does the reverse hold?
- Vytlacil (2002): "The assumption of an unobserved index crossing a threshold that defines the selection model is equivalent to the independence and monotonicity assumptions at the center of the LATE approach. In particular, the selection model assumptions imply the LATE assumptions, and given the LATE assumptions, there always exists a selection model that rationalizes the observed and counterfactual data."
- Thus the selection model is not more restrictive than the LATE model.
- The threshold-crossing model has a behavioral interpretation as the monotonicity condition.

Understanding Treatment Effect via the Generalized Roy Model

Marginal Treatment Effects (MTE)

Definition

Definition: Following Heckman and Vytlacil, define

$$\Delta^{MTE}(x,\nu) \;=\; E\big(\pi_i \,\big|\, X_i = x,\ \nu_i = \nu\big),$$

where π_i is individual-level treatment effect.

Interpretation

- $\Delta^{MTE}(x,\nu) =$ the treatment effect for individuals with covariates x and latent index ν.
- The MTE is a useful definition because it uses the selection model to partition the population based on all unobservable and observable determinants of their outcomes.
- Observable treatment heterogeneity is captured by X_i while unobservable heterogeneity is captured by ν_i.

Identifying MTE

- Recall that in the Generalized Roy Model, the choice variable is:
$$J_i = f \quad \Leftrightarrow \quad \varphi(Z_i, X_i) \ > \ \nu_i.$$
- Consider a pair of values (z', x) and (z^h, x) in the support,
$$\varphi(z', x) \ < \ \varphi(z^h, x),$$
and individuals with $\nu_i \in [\varphi(z', x), \varphi(z^h, x))$.
- We can identify MTE by identifying the marginal treatment response:
$$\begin{aligned}
&E\left(Y_{fi} \mid \varphi(z^\ell, x) \leq \nu_i < \varphi(z^h, x),\ X_i = x\right) \\
&= \frac{E(Y_{fi} \mid (Z_i, X_i) = (z^h, x),\ J_i = f) \cdot \Pr(J_i = f \mid (Z_i, X_i) = (z^h, x))}{\Pr(J_i = f \mid (Z_i, X_i) = (z^h, x)) - \Pr(J_i = f \mid (Z_i, X_i) = (z^\ell, x))} \\
&\quad - \frac{E(Y_{fi} \mid (Z_i, X_i) = (z^\ell, x),\ J_i = f) \cdot \Pr(J_i = f \mid (Z_i, X_i) = (z^\ell, x))}{\Pr(J_i = f \mid (Z_i, X_i) = (z^h, x)) - \Pr(J_i = f \mid (Z_i, X_i) = (z^\ell, x))}
\end{aligned} \tag{18}$$

MTE and Treatment Effects

- From the previous (identified) equation, we have:

$$\lim_{\substack{\varphi(z^{\ell},x)\uparrow\nu \\ \varphi(z^{h},x)\downarrow\nu}} E\left(Y_{fi} \mid \varphi(z^{\ell},x) \leq \nu_i < \varphi(z^{h},x),\ X_i = x\right) = E(Y_{fi} \mid \nu_i = \nu, X_i = x) \tag{19}$$

- Once we identify $E(Y_{fi} \mid \nu_i = \nu, X_i = x)$ and similarly $E(Y_{hi} \mid \nu_i = \nu, X_i = x)$, we can form:

$$\Delta^{MTE}(x,\nu) \ = \ E\big(Y_{fi} \mid \nu_i = \nu, X_i = x\big) - E\big(Y_{hi} \mid \nu_i = \nu, X_i = x\big). \tag{5.15}$$

- With MTE, we can construct various causal parameters such as ATE, ATT, LATE.
- See Mogstad and Torgovitsky (2018) and Mogstad and Torgovitsky (2024) for more details.

Linking MTR to Target Parameters (from Mogstad and Torgovitsky (2024))

TABLE 5 **Marginal treatment response weights for common target parameters.**

		MTR weights	
Target parameter	**Expression**	$\omega\,(1 \mid u, z, x)$	$\omega\,(0 \mid u, z, x)$
Average treated outcome	$\mathbb{E}[Y_i(1)]$	1	0
Average untreated outcome	$\mathbb{E}[Y_i(0)]$	0	1
Average treatment effect (ATE)	$\mathbb{E}[Y_i(1) - Y_i(0)]$	1	-1
Conditional ATE	$\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i \in \mathscr{X}]$	$\frac{\mathbb{1}[x \in \mathscr{X}]}{\mathbb{P}[X_i \in \mathscr{X}]}$	$-\omega(1 \mid u, z, x)$
Average treatment on the treated (ATT)	$\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$	$\frac{\mathbb{1}[u \le p(z,x)]}{\mathbb{P}[D_i = 1]}$	$-\omega(1 \mid u, z, x)$
Average treatment on the untreated (ATU)	$\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 0]$	$\frac{\mathbb{1}[u > p(z,x)]}{\mathbb{P}[D_i = 0]}$	$-\omega(1 \mid u, z, x)$
Generalization of the LATE to $U_i \in [\underline{u}, \bar{u}]$	$\mathbb{E}[Y_i(1) - Y_i(0) \mid U_i \in [\underline{u}, \bar{u}]]$	$\frac{\mathbb{1}[\underline{u} < u \le \bar{u}]}{\bar{u} - \underline{u}}$	$-\omega(1 \mid u, z, x)$
Average selection on treatment effects	$\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1] - \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 0]$	$\frac{\mathbb{1}[u \le p(z,x)]}{\mathbb{P}[D_i = 1]} - \frac{\mathbb{1}[u > p(z,x)]}{\mathbb{P}[D_i = 0]}$	$-\omega(1 \mid u, z, x)$
Average selection bias	$\mathbb{E}[Y_i(0) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0]$	$\frac{\mathbb{1}[u \le p(z,x)]}{\mathbb{P}[D_i = 1]} - \frac{\mathbb{1}[u > p(z,x)]}{\mathbb{P}[D_i = 0]}$	0
Policy relevant treatment effect (PRTE)	$\frac{\mathbb{E}[Y_i^\circ] - \mathbb{E}[Y_i]}{\mathbb{E}[D_i^\circ] - \mathbb{E}[D_i]}$	$\frac{\mathbb{P}[p^\circ(X_i, Z_i^\circ) \ge u] - \mathbb{P}[p(X_i, Z_i) \ge u]}{\mathbb{E}[p^\circ(X_i, Z_i^\circ)] - \mathbb{E}[p(X_i, Z_i)]}$	$-\omega(1 \mid u, x)$

The weights show how to produce the specified target parameter through the formula $\text{target parameter} = \mathbb{E}\left[\int_0^1 \mathrm{MTR}(1 \mid u, X_i)\,\omega(1 \mid u, Z_i, X_i)\,\mathrm{d}u - \int_0^1 \mathrm{MTR}(0 \mid u, X_i)\,\omega(0 \mid u, Z_i, X_i)\,\mathrm{d}u\right]$.

Local Instrumental Variables

- Heckman and Vytlacil (1999, 2001, 2005) suggest procedures to estimate the marginal treatment effect.
- We show that:

$$\Delta^{MTE}(x,\nu) = \frac{\partial}{\partial \nu} E\big(Y_i \mid X_i = x,\ \varphi(X_i, Z_i) = \nu\big). \qquad (20)$$

- Proof:

$$\begin{aligned}
\frac{\partial E(Y_i|X_i = x, \varphi(X_i, Z_i) = v)}{\partial v} &= \frac{\partial[E(Y_{fi}|X_i = x, v_i \leq v)\,\Pr(v_i \leq v) + E(Y_{hi}|X_i = x, v_i > v)\,\Pr(v_i > v)]}{\partial v} \\
&= \frac{\partial\left[\int_0^v E(Y_{fi}|v_i = \omega, X_i = x)d\omega + \int_v^1 E(Y_{hi}|v_i = \omega, X_i = x)d\omega\right]}{\partial v} \\
&= E(Y_{fi}|v_i = v, X_i = x) - E(Y_{hi}|v_i = v, X_i = x) \\
&= \Delta^{\mathrm{MTE}}(x, v). \qquad (21)
\end{aligned}$$

Local Instrumental Variables

- Step 1: Estimate $\varphi(x,z) = \Pr(J_i = f \mid X_i = x, Z_i = z)$. - This is the "first-stage" that maps instruments (x,z) to ν-values.
- Step 2: Regress Y_i on $\hat{\varphi}(X_i, Z_i)$ nonparametrically:

$$Y_i \ = \ f\big(\hat{\varphi}(X_i, Z_i)\big) + \epsilon_i .$$

- Step 3: $\Delta^{MTE}(x,\nu)$ is essentially the derivative of $f(\nu)$.