
James Heckman
University of Chicago

This draft, February 9, 2012
Roy Model

\((Y_0, Y_1)\) potential outcomes

\(I^* = Y_1 - Y_0\) choice index

Observe \(Y_1\) if \(Y_1 \geq Y_0\).
Observe \(Y_0\) if \(Y_1 < Y_0\).

Cannot simultaneously observe \(Y_0\) and \(Y_1\).
Heuristically, we can conduct an identification analysis assuming we know

\[ I = \frac{I^*}{\sigma_{Y_1 - Y_0}} = \frac{Y_1 - Y_0}{\sigma_{Y_1 - Y_0}} \]

for each person where \( D = 1(I > 0) \).

Why?


Even though we do not ever observe \( I \), we observe \((Y_0, D)\) and \((Y_1, D)\). We never observe the full triple \((Y_0, Y_1, D)\) for anyone.
Under conditions specified in the literature, $F(Y_0, I|X, Z)$ and $F(Y_1, I|X, Z)$ are identified where

\[
\begin{align*}
Y_0 &= \mu_0(X) + U_0 \quad E(Y_0 | X) = \mu_0(X) \\
Y_1 &= \mu_1(X) + U_1 \quad E(Y_1 | X) = \mu_1(X) \\
I^* &= \mu_I(X, Z) + U_I \\
I &= \frac{\mu_I(X, Z)}{\sigma_U} + \frac{U_I}{\sigma_U}
\end{align*}
\]

Source: Heckman (1990), Heckman and Honoré (1990)

The key idea in these papers is “sufficient” variation in $Z$ holding $X$ fixed.
From the left-hand side of

$$\Pr(D = 1|X, Z) = \Pr(\mu_I(X, Z) + U_I \geq 0|X, Z),$$

we can identify the distribution of $\frac{U_I}{\sigma_{U_I}}$, as well as $\frac{\mu_I(X, Z)}{\sigma_{U_I}}$. This is true under normality or any assumed form for the distribution of $\frac{U_I}{\sigma_{U_I}}$. It is also true more generally. One does not have to assume the distribution of $U_I$ is known or that the functional form of $\mu_I(X, Z)$ is linear, e.g. $\mu_I(X, Z) = X\beta_I + Z\gamma_i$.

(See the conditions in the Matzkin (1992) paper and the survey in Matzkin, 2007, *Handbook of Econometrics.*)
This more general claim requires full support of $Z$ and restrictions on $\mu_I(X, Z)$. See the “Matzkin conditions” in Cunha, Heckman, and Navarro (2007, IER). A key condition is

$$\text{Support} \left( \frac{\mu_I(X, Z)}{\sigma_{U_I}} \right) \supseteq \text{Support} \left( \frac{U_I}{\sigma_{U_I}} \right)$$

and other regularity conditions.

Commonly it is assumed that for a fixed $X$

$$\text{Support} \left( \frac{\mu_I(X, Z)}{\sigma_{U_I}} \right) = (-\infty, \infty).$$

This is called “identification at infinity.” When we vary $Z$ (for each $X$) we trace out the full support of $\frac{U_I}{\sigma_{U_I}}$. 

Identify the Joint Distribution of \((Y_0, I)\)

We know the conditional distribution of \(Y_0\):

\[
F(Y_0 | D = 0, X, Z) = \Pr(Y_0 \leq y_0 | \mu_I(X, Z) + U_I \leq 0, X, Z)
\]

Multiply this by \(\Pr(D = 0 | X, Z)\):

\[
F(Y_0 | D = 0, X, Z) \Pr(D = 0 | X, Z) = \Pr(Y_0 \leq y_0, I^* \leq 0 | X, Z) \quad (*)
\]

We can follow the analysis of Heckman (1990), Heckman and Smith (1998), and Carneiro, Hansen, and Heckman (2003).
Left hand side of (*) is known from the data.

Right hand side:

$$\Pr \left( Y_0 \leq y_0, \frac{U_I}{\sigma_{U_I}} < -\frac{\mu_I(X, Z)}{\sigma_{U_I}} \mid X, Z \right)$$

Since we know $\frac{\mu_I(X, Z)}{\sigma_{U_I}}$ from the previous analysis, we can vary it for each fixed $X$. 
If $\mu_I(X, Z)$ gets small ($\mu_I(X, Z) \to -\infty$), recover the marginal distribution $Y$ and in this limit set we can identify the marginal distribution of

$$Y_0 = \mu_0(X) + U_0 \quad \therefore \text{can identify } \mu_0(X) \text{ in limit.}$$

(See Heckman, 1990, and Heckman and Vytlacil, 2007.)

More generally, we can form:

$$\Pr \left( U_0 \leq y_0 - \mu_0(X), \frac{U_I}{\sigma_{U_I}} \leq \frac{-\mu_I(X, Z)}{\sigma_{U_I}} \mid X, Z \right)$$

$X$ and $Z$ can be varied and $y_0$ is a number. We can trace out joint distribution of $\left( U_0, \frac{U_I}{\sigma_{U_I}} \right)$ by varying $(Y_0, Z)$ for each fixed $X$. 
Recall joint distribution of

\[ (Y_0, I) = \left( \mu_0(X) + U_0, \frac{\mu_I(X, Z) + U_I}{\sigma_{U_I}} \right). \]

Three key ingredients.

1. The independence of \((U_0, U_I)\) and \((X, Z)\).

2. The assumption that we can set \(\frac{\mu_I(X, Z)}{\sigma_{U_I}}\) to be very small (so we get the marginal distribution of \(Y_0\) and hence \(\mu_0(X)\)).

3. The assumption that \(\frac{\mu_I(X, Z)}{\sigma_{U_I}}\) can be varied independently of \(\mu_0(X)\).

Trace out the joint distribution of \(\left( U_0, \frac{U_I}{\sigma_{U_I}} \right)\). Result generalizes easily to the vector case. (Carneiro, Hansen, and Heckman, 2003, IER)
Another way to see this is to write:

$$F(Y_0 \mid D = 0, X, Z) \Pr(D = 0 \mid X, Z)$$

This is a function of $\mu_0(X)$ and $\frac{\mu_I(X, Z)}{\sigma_{U_I}}$ (Index sufficiency)
Varying the $\mu_0(X)$ and $\frac{\mu_I(X, Z)}{\sigma_{U_I}}$ traces out the distribution of $\left( U_0, \frac{U_I}{\sigma_{U_I}} \right)$.

This means effectively that we observe the pairs $\left( \frac{I}{\sigma_{U_I}}, Y_1 \right)$ and $\left( \frac{I}{\sigma_{U_I}}, Y_0 \right)$.

We never observe the triple $\left( \frac{I}{\sigma_{U_I}}, Y_0, Y_1 \right)$. 
Use the intuition that we “know” \( I \). Actually we observe

\[
F(Y_0 \mid I < 0, X, Z)
\]

and

\[
F(Y_1 \mid I \geq 0, X, Z)
\]

and

\[
\Pr(I \geq 0 \mid X, Z)
\]

and can construct the joint distributions \( F(Y_0, I \mid X, Z) \) and \( F(Y_1, I \mid X, Z) \).
Roy Case

Armed with normality (or the nonparametric assumptions in Heckman and Honoré, 1990), we can estimate

\[
\text{Cov}(I, Y_1) = \frac{\text{Var}(Y_1) - \text{Cov}(Y_0, Y_1)}{\sigma_{Y_1}^2 + \sigma_{Y_0}^2 - 2\sigma_{Y_1,Y_0}} \\
\text{Cov}(I, Y_0) = -\frac{\text{Var}(Y_0) - \text{Cov}(Y_0, Y_1)}{\sigma_{Y_1}^2 + \sigma_{Y_0}^2 - 2\sigma_{Y_1,Y_0}}
\]

We know \(\text{Var} Y_1, \text{Var} Y_0\) (e.g. normal selection model or use limit sets)

\[\therefore \text{Cov}(Y_0, Y_1) \text{ is identified (actually over-identified)}.\]

This line of argument does not generalize if we add a cost component \((C)\) that is unobserved (or partly so).
The intuition is clear. In the Roy model the decision rule is generated solely by \((Y_1, Y_0)\). Knowing agent choices we observe the relative order (and magnitude) of \(Y_1\) and \(Y_0\).

Thus we get a second valuable piece of information from agent choices. This information is ignored in statistical approaches to program evaluation.

But does this analysis generalize?
Generalized Roy Model

Add cost

\[ I = Y_1 - Y_0 - C \]

and assume that we do not directly observe \( C \).

Observe \( Y_1 \mid I > 0 \),

Observe \( Y_0 \mid I < 0 \),

and

\[ I = \frac{Y_1 - Y_0 - C}{\sqrt{\text{Var}(Y_1 - Y_0 - C)}}. \]
We can identify $\text{Var} \ Y_1$ and can identify $\text{Var} \ Y_0$.

But we cannot directly identify $\text{Cov}(Y_0, Y_1)$ which measures comparative advantage in Willis-Rosen model.

Notice, however, we can determine if

$$E(Y_1 \mid I > 0) > E(Y_1)$$

$$E(Y_0 \mid I < 0) > E(Y_0)$$

(Are people who work in a sector above average for the sector?)
One way around this problem is to use panel data with a factor structure (two period version of Willis and Rosen).

\( Y_{j,t} \): earnings at section \( j \) at time \( t \).

Observed if a person chooses “1”

\[
\begin{align*}
Y_{1,1} &= \mu_{11}(X) + \alpha_{11}\theta + \varepsilon_{11} \\
Y_{1,2} &= \mu_{12}(X) + \alpha_{12}\theta + \varepsilon_{12}
\end{align*}
\]  

(1)

Observed if a person chooses “0”

\[
\begin{align*}
Y_{0,1} &= \mu_{01}(X) + \alpha_{01}\theta + \varepsilon_{01} \\
Y_{0,2} &= \mu_{02}(X) + \alpha_{02}\theta + \varepsilon_{02}
\end{align*}
\]  

(2)

Cost

\[
C = \mu_{C}(X, Z) + \alpha_{0C}\theta + \varepsilon_{C}
\]  

(3)
Assume that $\varepsilon_{11}, \varepsilon_{12}, \varepsilon_{01}, \varepsilon_{02}, \varepsilon_C$ are mutually independent and are all independent of $\theta$.

$\theta$ is assumed to be a scalar (for now).

To apply to educational choice suppose discount rate is zero.

Net present value of earnings: $I = \frac{Y_{1,1} + Y_{1,2} - Y_{0,1} - Y_{0,2} - C}{\sigma_{U_i}}$  \(4\)

Observe our heuristic goes through, we observe $I$ up to scale.
Before considering a general analysis, step back and consider identification in factor models

\[ Y_0 = \mu_0(X) + \alpha_0 \theta + \varepsilon_0 \]
\[ Y_1 = \mu_1(X) + \alpha_1 \theta + \varepsilon_1 \]
\[ l = \mu_l(X, Z) + \alpha_l \theta + \varepsilon_l \]
\[ \theta \perp (\varepsilon_1, \varepsilon_2, \varepsilon_l) \quad \text{Source of Identification} \]
Normalize $\alpha_0 = 1$ (Sets scale of factor)

Condition on $X$. Keep the conditioning implicit.

$$\text{Var}(Y_0) = \alpha_0^2 \sigma^2_\theta + \sigma^2_{\varepsilon_0}$$

$$\text{Var}(I, Y_1) = \alpha_1^2 \sigma^2_\theta + \sigma^2_{\varepsilon_1}$$

$$\text{Var}(I) = \alpha_I^2 \sigma^2_\theta + \sigma^2_{\varepsilon_I}$$

$$\text{Cov}(Y_0, Y_1) = \alpha_0 \alpha_1 \sigma^2_\theta$$

$$\text{Cov}(Y_0, I) = \alpha_0 \alpha_I \sigma^2_\theta$$

$$\text{Cov}(Y_1, I) = \alpha_1 \alpha_I \sigma^2_\theta$$
If $\alpha_0 = 1$ and $\alpha_1, \alpha_I \neq 0$,

$$\frac{\text{Cov}(Y_1, I)}{\text{Cov}(Y_1, Y_0)} = \frac{\alpha_1 \alpha_I \sigma^2_\theta}{\alpha_1 \sigma^2_\theta} = \alpha_I$$

If we know $\alpha_I$, we know $\sigma^2_\theta$ (from $\text{Cov}(Y_1, I)$)

$\therefore$ Know $\alpha_1$ and ALL of the $\sigma^2_{\varepsilon_j}$

ALL Cov identified

Can apply with modifications to panel data on the Roy model

(which is Willis-Rosen)
Consider a special case of the Willis and Rosen model:

Everyone who goes to high school gets zero earnings. College earnings are measured over two periods:

\[
\begin{align*}
Y_1 &= \mu_1(X) + \alpha_1 \theta + \varepsilon_1 \\
Y_2 &= \mu_2(X) + \alpha_2 \theta + \varepsilon_2 \\
C &= \mu_C(Z) + \alpha_C \theta + \varepsilon_C \\
I &= Y_1 + Y_2 - C \\
&= \mu_1(X) + \mu_2(X) - \mu_C(Z) + (\alpha_1 + \alpha_2 - \alpha_C) \theta + \varepsilon_1 + \varepsilon_2 - \varepsilon_C
\end{align*}
\]

Invoke independence among components:

\[
\sigma_I = \sqrt{(\alpha_1 + \alpha_2 - \alpha_C)^2 \sigma_\theta^2 + \sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 + \sigma_{\varepsilon_C}^2}
\]
We know $\mu_1(X), \mu_2(X)$
(Normal selection or nonparametric arguments)

\[ \therefore \text{can identify} \]
\[ \frac{\mu_1(X) + \mu_2(X) - \mu_C(Z)}{\sigma_1} \]

from choice equation.

If we have an exclusion restriction
(one or more variables in $X$ not in $Z$; this is not the usual IV exclusion) and $\mu_C(Z)$ and $(\mu_1(X) + \mu_2(X))$ are not collinear, we can identify $\sigma_1$.

This condition allows us to identify $\sigma_1$ using the known components of $\mu_1(X) + \mu_2(X)$ which are fixed by estimating the selection-corrected values $Y_1$ and $Y_2$.

\[ \therefore \text{Can identify } \mu_C(Z) \]
What else can be identified? 
(Use heuristic that we know \(I\))

By previous reasoning, set \(\alpha_1 = 1\)

\[\therefore \text{We know } \alpha_2, \alpha_1 + \alpha_2 - \alpha_C \quad \therefore \alpha_C\]

\[\therefore \text{We know } \sigma_\theta^2\]

\[\therefore \text{We know } \sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2 \quad \therefore \text{We know } \sigma_{\varepsilon_C}^2\]
An example of exclusion would be

$$\mu_{i,j}(X) = X\beta_{i,j}$$

and we know at least one component of $\beta_{i,j}$ does not have a corresponding variable in

$$\mu_c(X, Z) = X\beta_c + Z\gamma_c.$$ 

Thus the model is fully identified.
Willis and Rosen with a Factor Structure: Special Case

Apply the same reasoning to this case

1. We can identify \( F(Y_{1,1}, Y_{1,2}, I \mid X, Z) \) and 
   \( F(Y_{0,1}, Y_{0,2}, I \mid X, Z) \).

2. Thus, we can identify \( \mu_{11}(X), \mu_{12}(X), \mu_{01}(X), \mu_{02}(X) \) and
   \[
   \text{Cov}(Y_{1,1}, Y_{1,2}), \text{Cov}(Y_{1,1}, I), \text{Cov}(Y_{1,2}, I),
   \text{Cov}(Y_{0,1}, Y_{0,2}), \text{Cov}(Y_{0,1}, I), \text{Cov}(Y_{0,2}, I),
   \]

   \[
   \text{Var}(Y_{1,1}), \text{Var}(Y_{1,2}), \text{Var}(Y_{0,1}), \text{Var}(Y_{0,2})
   \]

3. We cannot identify the full distribution
   \[
   (Y_{0,1}, Y_{0,2}, Y_{1,1}, Y_{1,2}, I \mid X, Z)
   \]
From discrete choice analysis, we identify the combination of parameters:

\[
\frac{\mu_{11}(X) + \mu_{12}(X) - \mu_{0,1}(X) - \mu_{0,2}(X) - \mu_C(X, Z)}{\sigma_I}
\] (**)

where

\[
\sigma_I^2 = (\alpha_{11} + \alpha_{12} - \alpha_{01} - \alpha_{02} - \alpha_{0C})^2 \sigma_\theta^2 \\
+ \sigma_{\varepsilon_{11}}^2 + \sigma_{\varepsilon_{12}}^2 + \sigma_{\varepsilon_{01}}^2 + \sigma_{\varepsilon_{02}}^2 + \sigma_{\varepsilon_C}^2
\]
But we are missing all

\[ \text{Cov}(Y_{j,k}, Y_{\ell,m}) \quad j \neq \ell, \quad j, \ell \in \{0, 1\} \]

\[ k, m \in \{1, 2\} \]

\[ \text{Cov}(Y_{j,k}, C) \quad j, k \in \{1, 2\} \]

This creates a fundamental identification problem.

Even if we observe \( C \) we cannot, without additional assumptions, identify the patterns of comparative advantage (correlations of outcomes across schooling choices).
Invoke factor structure.
From what we do observe, we can identify

\[
\text{Cov}(Y_{1,1}, Y_{1,2}) = \alpha_{11}\alpha_{12}(\sigma^2_{\theta})
\]  
(C-1)

\[
\text{Cov}(Y_{1,1}, I) = \alpha_{11} \frac{(\alpha_{11} + \alpha_{12} - \alpha_{0,1} - \alpha_{0,2} - \alpha_{0,c})}{\sigma_I} (\sigma^2_{\theta})
\]  
(C-2)

\[
\text{Cov}(Y_{1,2}, I) = \alpha_{12} \frac{(\alpha_{11} + \alpha_{12} - \alpha_{0,1} - \alpha_{0,2} - \alpha_{0,c})}{\sigma_I} (\sigma^2_{\theta})
\]  
(C-3)

\[
\text{Cov}(Y_{0,1}, Y_{0,2}) = \alpha_{02}\alpha_{01}(\sigma^2_{\theta})
\]  
(C-4)

\[
\text{Cov}(Y_{0,1}, I) = \alpha_{01} \frac{(\alpha_{11} + \alpha_{12} - \alpha_{0,1} - \alpha_{0,2} - \alpha_{0,c})}{\sigma_I} (\sigma^2_{\theta})
\]  
(C-5)

\[
\text{Cov}(Y_{0,2}, I) = \alpha_{02} \frac{(\alpha_{11} + \alpha_{12} - \alpha_{0,1} - \alpha_{0,2} - \alpha_{0,c})}{\sigma_I} (\sigma^2_{\theta})
\]  
(C-6)
From our previous identification argument, we know

\[ \mu_{11}(X), \mu_{12}(X), \mu_{01}(X), \mu_{02}(X) \]

From our exclusion condition (we have some variables in \( X \) not in \( Z \)) and the assumed lack of collinearity we use \( (**) \) to identify \( \sigma_i \) and hence \( \mu_C(X, Z) \).

This requires that we be able to fix at least some known coordinates of \( X \).
Normalize $\alpha_{11} = 1$
(This assumes it is not zero; we cannot normalize a zero factor loading)

$$\frac{\text{Cov}(Y_{1,2}, I)}{\text{Cov}(Y_{1,1}, I)} = \alpha_{12}$$

∴ from (C-1) we get $\sigma_{\theta}^2$

∴ from either (C-2) or (C-3) we identify

$$\frac{\left(\alpha_{11} - \alpha_{12} - \alpha_{01} - \alpha_{02} - \alpha_{0C}\right)\sigma_{\theta}^2}{\sigma_I}$$

(***)

Divide into (C-5) and (C-6) respectively to get $\alpha_{01}$ and $\alpha_{02}$, respectively.
Given $\sigma^2_0$, $\sigma_I$ from (***), we can identify $\alpha_{0C}$.

Model strongly identified and testable.

Using the observations on $Y_{11}$, $Y_{12}$, $Y_{01}$, $Y_{02}$, we can identify $\sigma^2_{\varepsilon_{11}}$, $\sigma^2_{\varepsilon_{12}}$, $\sigma^2_{\varepsilon_01}$, $\sigma^2_{\varepsilon_02}$.

Using our knowledge of $\alpha_{11}$, $\alpha_{12}$, $\alpha_{01}$, $\alpha_{02}$, and $\alpha_{0C}$, from $\sigma^2_I$, we can identify $\sigma^2_{\varepsilon_C}$. 
Using Factor Analysis, Can Construct Joint Distributions of Counterfactuals

Example: One factor model.

Assume that all of the dependence across \((U_0, U_1, U_{I*})\) is generated by a scalar factor \(\theta\)

\[
U_0 = \theta \alpha_0 + \varepsilon_0 \\
U_1 = \theta \alpha_1 + \varepsilon_1 \\
U_{I*} = \theta \alpha_{I*} + \varepsilon_{I*}.
\]

\[
E(\theta) = 0, \quad \text{and} \quad E(\theta^2) = \sigma_\theta^2. \\
E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_{I*}) = 0 \\
Var(\varepsilon_0) = \sigma_{\varepsilon_0}^2, \ Var(\varepsilon_1) = \sigma_{\varepsilon_1}^2 \\
Var(\varepsilon_{I}) = \sigma_{\varepsilon_{I}}^2
\]
1. Factor Models: A Brief Digression

\[ E(\theta) = 0; \quad E(\varepsilon_i) = 0; \quad i = 1, \ldots, 5 \]

\[ R_1 = \alpha_1 \theta + \varepsilon_1, \quad R_2 = \alpha_2 \theta + \varepsilon_2, \quad R_3 = \alpha_3 \theta + \varepsilon_3, \]
\[ R_4 = \alpha_4 \theta + \varepsilon_4, \quad R_5 = \alpha_5 \theta + \varepsilon_5, \quad \varepsilon_i \perp \perp \varepsilon_j, \quad i \neq j \]

\[
\begin{align*}
\text{Cov} (R_1, R_2) & = \alpha_1 \alpha_2 \sigma^2_\theta \\
\text{Cov} (R_1, R_3) & = \alpha_1 \alpha_3 \sigma^2_\theta \\
\text{Cov} (R_2, R_3) & = \alpha_2 \alpha_3 \sigma^2_\theta 
\end{align*}
\]

Normalize \( \alpha_1 = 1 \)

\[
\frac{\text{Cov} (R_2, R_3)}{\text{Cov} (R_1, R_2)} = \alpha_3
\]
\[ \therefore \text{We know } \sigma^2_\theta \text{ from } \text{Cov}(R_1, R_2). \text{ From } \text{Cov}(R_1, R_3) \text{ we know } \alpha_3, \alpha_4, \alpha_5. \]

Can get the variances of the \( \varepsilon_i \) from variances of the \( R_i \)

\[ \text{Var}(R_i) = \alpha_i^2 \sigma^2_\theta + \sigma^2_{\varepsilon_i}. \]

If \( T = 2 \), all we can identify is \( \alpha_1 \alpha_2 \sigma^2_\theta. \)

If \( \alpha_1 = 1, \sigma^2_\theta = 1 \), we identify \( \alpha_2. \) Otherwise model is fundamentally underidentified.
2 Factors: (some examples)

\[ \theta_1 \perp \theta_2 \]

\[ \varepsilon_i \perp \varepsilon_j \quad \forall i \neq j \]

\[
\begin{align*}
R_1 &= \alpha_{11} \theta_1 + (0) \theta_2 + \varepsilon_1 \\
R_2 &= \alpha_{21} \theta_1 + (0) \theta_2 + \varepsilon_2 \\
R_3 &= \alpha_{31} \theta_1 + \alpha_{32} \theta_2 + \varepsilon_3 \\
R_4 &= \alpha_{41} \theta_1 + \alpha_{42} \theta_2 + \varepsilon_4 \\
R_5 &= \alpha_{51} \theta_1 + \alpha_{52} \theta_2 + \varepsilon_5
\end{align*}
\]

Let \( \alpha_{11} = 1, \alpha_{32} = 1 \). (Set scale)
\[ \text{Cov} (R_1, R_2) = \alpha_{21} \sigma_{\theta_1}^2 \]
\[ \text{Cov} (R_1, R_3) = \alpha_{31} \sigma_{\theta_1}^2 \]
\[ \text{Cov} (R_2, R_3) = \alpha_{21} \alpha_{31} \sigma_{\theta_1}^2 \]

Form ratio of \[ \frac{\text{Cov} (R_2, R_3)}{\text{Cov} (R_1, R_2)} = \alpha_{31}, \quad \therefore \text{we identify } \alpha_{31}, \alpha_{21}, \sigma_{\theta_1}^2, \]

as before.

\[ \text{Cov} (R_1, R_4) = \alpha_{41} \sigma_{\theta_1}^2, \quad \therefore \text{since we know } \sigma_{\theta_1}^2 \text{ we get } \alpha_{41}. \]

\[ \vdots \]

\[ \text{Cov} (R_1, R_k) = \alpha_{k1} \sigma_{\theta_1}^2 \]

\[ \therefore \text{we identify } \alpha_{k1} \text{ for all } k \text{ and } \sigma_{\theta_1}^2. \]
\[
\begin{align*}
\text{Cov} (R_3, R_4) - \alpha_3 \alpha_4 \sigma^2_{\theta_1} &= \alpha_4 \sigma^2_{\theta_2} \\
\text{Cov} (R_3, R_5) - \alpha_3 \alpha_5 \sigma^2_{\theta_1} &= \alpha_5 \sigma^2_{\theta_2} \\
\text{Cov} (R_4, R_5) - \alpha_4 \alpha_5 \sigma^2_{\theta_1} &= \alpha_5 \alpha_4 \sigma^2_{\theta_2},
\end{align*}
\]

By same logic,

\[
\frac{\text{Cov} (R_4, R_5) - \alpha_4 \alpha_5 \sigma^2_{\theta_1}}{\text{Cov} (R_3, R_4) - \alpha_3 \alpha_4 \sigma^2_{\theta_1}} = \alpha_5
\]

∴ get \(\sigma^2_{\theta_2}\) of “2” loadings.
If we have dedicated measurements on each factor do not need a normalization on the factors of $R$. Dedicated measurements set the scales and make factor models interpretable:

$$
M_1 = \theta_1 + \varepsilon_{1M}
$$

$$
M_2 = \theta_2 + \varepsilon_{2M}
$$

$$
\text{Cov} (R_1, M) = \alpha_{11} \sigma^2_{\theta_1}
$$

$$
\text{Cov} (R_2, M) = \alpha_{21} \sigma^2_{\theta_1}
$$

$$
\text{Cov} (R_3, M) = \alpha_{31} \sigma^2_{\theta_1}
$$

$$
\text{Cov} (R_1, R_2) = \alpha_{11} \alpha_{12} \sigma^2_{\theta_1},
$$

$$
\text{Cov} (R_1, R_3) = \alpha_{11} \alpha_{13} \sigma^2_{\theta_1}, \quad \therefore \alpha_{12} \sigma^2_{\theta_1},
$$

$\therefore$ We can get $\alpha_{12}, \sigma^2_{\theta_1}$ and the other factors.
General Case

\[ R_{T \times 1} = M_{T \times 1} + \Lambda_{T \times K} \theta_{K \times 1} + \varepsilon_{T \times 1} \]

\( \theta \) are factors, \( \varepsilon \) uniquenesses

\[ E(\varepsilon) = 0 \]

\[ \text{Var}(\varepsilon\varepsilon') = D = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{\varepsilon_2}^2 & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma_{\varepsilon_T}^2 \end{pmatrix} \]

\[ E(\theta) = 0 \]

\[ \text{Var}(R) = \Lambda \Sigma_\theta \Lambda' + D \quad \Sigma_\theta = E(\theta\theta') \]
The only source of information on $\Lambda$ and $\Sigma_\theta$ is from the covariances. (Each variance is “contaminated” by a uniqueness.)

Associated with each variance of $R_i$ is a $\sigma^2_{\varepsilon_i}$.

Each uniqueness variance contributes one new parameter.

How many unique covariance terms do we have?

$$\frac{T(T - 1)}{2}$$
We have $T$ uniquenesses; $TK$ elements of $\Lambda$.

\[
\frac{K(K-1)}{2} \text{ elements of } \Sigma_\theta.
\]

\[
\frac{K(K-1)}{2} + TK \text{ parameters } (\Sigma_\theta, \Lambda).
\]

Need this many covariances to identify model “Ledermann Bound”:

\[
\frac{T(T-1)}{2} \geq TK + \frac{K(K-1)}{2}
\]
Lack of Identification Up to Rotation

Observe that if we multiply $\Lambda$ by an orthogonal matrix $C$, $(CC' = I)$, we obtain

$$Var (R) = \Lambda C \left[ C' \Sigma_\theta C \right] C' \Lambda' + D$$

$C$ is a “rotation”. Cannot separate $\Lambda C$ from $\Lambda$.

Model not identified against orthogonal transformations in the general case.
Some common assumptions:

(i) $\theta_i \perp \theta_j$, $\forall \ i \neq j$

$$
\Sigma_\theta = \begin{pmatrix}
\sigma^2_{\theta_1} & 0 & \ldots & 0 \\
0 & \sigma^2_{\theta_2} & 0 & \vdots \\
\vdots & 0 & \ddots & \vdots \\
0 & \ldots & 0 & \sigma^2_{\theta_K}
\end{pmatrix}
$$
joined with

(ii)

\[ \Lambda = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & \ldots & 0 \\
\alpha_{21} & 0 & 0 & 0 & \ldots & 0 \\
\alpha_{31} & 1 & 0 & 0 & \ldots & 0 \\
\alpha_{41} & \alpha_{42} & 0 & 0 & \ldots & 0 \\
\alpha_{51} & \alpha_{52} & 1 & 0 & \ldots & 0 \\
\alpha_{61} & \alpha_{62} & \alpha_{63} & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & 1 & \vdots
\end{pmatrix} \]
We know that we can identify of the $\Lambda$, $\Sigma_\theta$ parameters.

$$\frac{K(K-1)}{2} + TK \leq \frac{T(T-1)}{2}$$

"Ledermann Bound"

Can get more information by looking at higher order moments.
(See, e.g. Bonhomme and Robin, 2004)
Recovering the Factor Loadings (Go back to simple case)
The Case when there is information only on $Y_0$ for $I < 0$ and $Y_1$ for $I > 0$

Can identify $F(U_0, U_{I^*})$ and $F(U_1, U_{I^*})$, $\therefore$ can identify

$$\text{Cov}(U_0, U_{I^*}) = \alpha_0 \alpha_{I^*} \sigma^2_\theta$$
$$\text{Cov}(U_1, U_{I^*}) = \alpha_1 \alpha_{I^*} \sigma^2_\theta.$$ 

Scale of the unobserved $I$ is normalized

$$\sigma^2_\theta = 1 \quad \alpha \theta = k \alpha \frac{\theta}{k}$$
Normalize some $\alpha_j$ to one.

$$\alpha_{I^*} = 1$$

Identify $\alpha_1$ and $\alpha_0$ from the known covariances above.

4 parameters: $\alpha_0$, $\alpha_{I^*}$, $\alpha_1$, $\sigma^2_\theta$. 2 equations.

Normalize like crazy: $\alpha_{I^*} = 1$, $\alpha_1 = 1$ \therefore $\sigma^2_\theta$ \therefore $\alpha_1$.

Can make alternative normalizations.
Since

$$\text{Cov} (U_1, U_0) = \alpha_1 \alpha_0 \sigma^2_\theta$$

we can identify covariance between $Y_1$ and $Y_0$

Do not observe the pair $(Y_1, Y_0)$

Access to more observations (say from panel data $T > 0$)

$$\frac{\text{Cov} (Y_{1t'}, Y_{1t})}{\text{Cov} (Y_{1t'}, I^*)} = \alpha_{1t}$$

$$\frac{\text{Cov} (Y_{0t'}, Y_{0t})}{\text{Cov} (Y_{0t'}, I^*)} = \alpha_{0t}$$
Then, if $T \geq 2$, given say $\alpha_{11} = 1$, we can identify $\sigma^2_{\theta}$, $\alpha_{1j}$, $j \geq 1$.

Knowing $\sigma^2_{\theta}$, can identify $\alpha_{0t} \forall t$, $\therefore$ we can identify $\alpha_{I*}$ up to scale $\sigma_I$.

Key idea: Get more identification with more data.

$\therefore$ We have many more equations than parameters.
Crucial Idea of Identification in Roy Models with and without the Factor Structure

We never observe \((Y_1, Y_0)\) as a pair, both \(Y_0\) and \(Y_1\) are linked to \(D\) through the choice equation.

From distribution of \(D\) we can generate the distribution of \(I^*\)

We essentially observe \((Y_0, I^*)\) and \((Y_1, I^*)\).

The common dependence of \(Y_0\) and \(Y_1\) on \(I^*\) secures identification of the joint distribution of \(Y_0, Y_1, I^*\).
Adding a Measurement Equation Helps

(Go back to the case on slide 49.)

Suppose we have a measurement for $\theta$ observed whether $D = 1$ or $D = 0$

Measured ability $M$ is

$$M = \mu_M(X) + U_M.$$ 

Assume that

$$U_M = \alpha_M \theta + \varepsilon_M.$$
We assume $\alpha_M \neq 0$. Can form

\[
\begin{align*}
\operatorname{Cov}(M, Y_0) &= \operatorname{Cov}(U_M, U_0) = \alpha_M \alpha_0 \sigma_\theta^2 \\
\operatorname{Cov}(M, Y_1) &= \operatorname{Cov}(U_M, U_1) = \alpha_M \alpha_1 \sigma_\theta^2 \\
\operatorname{Cov}(M, I^*) &= \operatorname{Cov}(U_M, U_{I^*}) = \alpha_M \alpha_{I^*} \sigma_\theta^2
\end{align*}
\]

\[\begin{aligned} \left\{ \text{3 new equations} \right. \end{aligned}\]

$\alpha_M = 1.$
Can form the ratios

Identify $\alpha_0$:
$$\frac{\text{Cov} (U_0, U_{I^*})}{\text{Cov} (U_M, U_{I^*})} = \alpha_0$$

Recover $\alpha_1$:
$$\frac{\text{Cov} (U_1, U_{I^*})}{\text{Cov} (U_M, U_{I^*})} = \alpha_1$$

$$\text{Cov} (U_M, U_0) = \alpha_0 \sigma^2_{\theta}$$

$\therefore$ know $\sigma^2_{\theta}$

Can identify $\alpha_{I^*}$.

$$\text{Cov} (U_M, U_{I^*}) = \alpha_{I^*} \sigma^2_{\theta}$$
Intuition on Identification of the Normal Version Model with a Test Score

Generalized Roy versions of model:

\[ M = \mu (X) + \theta_1 \alpha_{1,M} + \theta_2 \alpha_{2,M} + \varepsilon_M \]

Assume \( \theta_1 \perp \perp \theta_2 \)

(Measurement: A test score equation)

\[
\begin{align*}
Y_1^1 &= \mu_1^1 (X) + \theta_1 \alpha_{1,1}^1 + \theta_2 \alpha_{2,1}^1 + \varepsilon_1^1 \\
Y_2^1 &= \mu_2^1 (X) + \theta_1 \alpha_{1,2}^1 + \theta_2 \alpha_{2,2}^1 + \varepsilon_2^1 \\
Y_1^0 &= \mu_1^0 (X) + \theta_1 \alpha_{1,1}^0 + \theta_2 \alpha_{2,1}^0 + \varepsilon_1^0 \\
Y_2^0 &= \mu_2^0 (X) + \theta_1 \alpha_{1,2}^0 + \theta_2 \alpha_{2,2}^0 + \varepsilon_2^0
\end{align*}
\]

\{ \text{College earnings} \}

\{ \text{High School earnings} \}

Cost

\[ C = Z \gamma + \theta_1 \alpha_{1C} + \theta_2 \alpha_{2C} + \varepsilon_C \]
Decision Rule Under Perfect Certainty:
(Assume $r = 0$)

\[ I = \mu_1^1(X) + \mu_2^1(X) + \theta_1 (\alpha_{1,1}^1 + \alpha_{1,2}^1) \]
\[ + \theta_2 (\alpha_{2,1}^1 + \alpha_{2,2}^1) + \varepsilon_1 + \varepsilon_2 \]
\[ - \left[ \mu_1^0(X) + \mu_2^0(X) + \theta_1 (\alpha_{1,1}^0 + \alpha_{1,2}^0) \right] \]
\[ + \theta_2 (\alpha_{2,1}^0 + \alpha_{2,2}^0) + \varepsilon_1^0 + \varepsilon_2^0 \]
\[ - \mathbf{Z} \gamma - \theta_1 \alpha_{1C} - \theta_2 \alpha_{2C} - \varepsilon \mathbf{C} \]
\[ = \mu_1^1(X) + \mu_2^1(X) - \left[ \mu_1^0(X) + \mu_2^0(X) + \mathbf{Z} \gamma \right] \]
\[ + \theta_1 \left[ (\alpha_{1,1}^1 + \alpha_{1,2}^1) - (\alpha_{1,1}^0 + \alpha_{1,2}^0) - \alpha_{1C} \right] \]
\[ + \theta_2 \left[ (\alpha_{2,1}^1 + \alpha_{2,2}^1) - (\alpha_{2,1}^0 + \alpha_{2,2}^0) - \alpha_{2C} \right] \]
\[ + (\varepsilon_1 + \varepsilon_2^1) - (\varepsilon_1^0 + \varepsilon_2^0) - \varepsilon \mathbf{C} \]
In Reduced Form

\[ I = \varphi(X, Z) + \alpha_{I,1}\theta_1 + \alpha_{I,2}\theta_2 + \varepsilon_I. \]

Set \( U_I = \alpha_{I,1}\theta_1 + \alpha_{I,2}\theta_2 + \varepsilon_I. \)

\[ \therefore \text{we can write} \]

\[ Y_1^1 = \mu_1^1(X) + U_1^1 \]
\[ Y_2^1 = \mu_2^1(X) + U_2^1 \]
\[ Y_1^0 = \mu_1^0(X) + U_1^0 \]
\[ Y_2^0 = \mu_2^0(X) + U_2^0 \]

\( U_1^1, U_2^1 \) etc. match the error terms previously shown.

\[ U_1^1 = \theta_1\alpha_{1,1}^1 + \theta_2\alpha_{2,1}^1 + \varepsilon_1^1 \]

etc.

\[ U_M = \theta_1\alpha_{1,M} + \theta_2\alpha_{2,M} + \varepsilon_M \]
\[
E \left( Y_1 \mid X, Z, I > 0 \right) = \mu_1(X) + \frac{\text{Cov} \left( U_1, I \right)}{\text{Var} \left( I \right)} \lambda()
\]

Using standard sample selection bias arguments we can identify in addition to the means,

\[
\mu_1(X), \mu_2(X), \mu_0(X), \mu_2(X)
\]

the following parameters:

\[
\text{Cov} \left( U_1, U_2 \right), \text{Var} \left( U_1 \right), \text{Var} \left( U_2 \right) \\
\text{Cov} \left( U_1, U_M \right), \text{Cov} \left( U_2, U_M \right), \text{Var} \left( U_M \right) \\
\text{Cov} \left( U_0, U_0 \right), \text{Var} \left( U_1 \right), \text{Var} \left( U_2 \right) \\
\text{Cov} \left( U_0, U_M \right), \text{Cov} \left( U_0, U_M \right)
\]
Normal Case:

\[(\theta, \varepsilon) \perp \perp (X, Z)\]

\[(\theta, \varepsilon) \text{ normal.}\]

\[
\Pr(D = 1 \mid X, Z, \theta_1, \theta_2) = \Phi \left[ \frac{1}{\sigma_{\varepsilon_l}} \left[ \begin{array}{c} \mu_1^1(X) + \mu_2^1(X) - [\mu_1^0(X) + \mu_2^0(X)] \\ -Z\gamma + \theta_1\alpha_{l,1} + \theta_2\alpha_{l,2} \end{array} \right] \right]
\]

\[
\sigma_{\varepsilon_l}^2 = \text{Var}(\varepsilon_l) = \left[ \sigma_{\varepsilon_c}^2 + \text{Var}(\varepsilon_1^1) + \text{Var}(\varepsilon_2^1) + \text{Var}(\varepsilon_1^0) + \text{Var}(\varepsilon_2^0) \right]
\]
Fact:

If $D = 1 [X\beta + \theta > V]$, $X \perp \perp (\theta, V)$

$\theta, V$ are normal, $\theta \perp \perp V$, $E(\theta) = 0$, $E(V) = 0$

$$Pr(D = 1 | X, \theta) = \Phi \left( \frac{X\beta + \theta}{\sigma_V} \right)$$

$$Pr(D = 1 | X) = \Phi \left( \frac{X\beta}{(\sigma_V^2 + \sigma_\theta^2)^{\frac{1}{2}}} \right)$$

Why? $D = 1 [X\beta > V - \theta]$. Rest follows from independence (between $V - \theta$, and $X$, and normality).
Unconditional Probability: (Not conditioning on $\theta_1$ and $\theta_2$)

$$\Pr(D = 1 \mid X, Z) = \Phi \left[ \frac{\mu_1^1(X) + \mu_2^1(X) - [\mu_1^0(X) + \mu_2^0(X)] - Z \gamma}{\left( \sigma_{\varepsilon_1}^2 + \alpha_{l,1}^2 \sigma_{\theta_1}^2 + \alpha_{l,2}^2 \sigma_{\theta_2}^2 \right)^{1/2}} \right]$$

$\uparrow \quad \uparrow$

with dedicated measurements we know $\sigma_{\theta_1}^2$ and $\sigma_{\theta_2}^2$. 
Observe that if we know $\mu^1_1(X), \mu^1_2(X), \mu^0_1(X), \mu^0_2(X)$ we know

$$\left[\mu^1_1(X) + \mu^1_2(X)\right] - \left[\mu^0_1(X) + \mu^0_2(X)\right].$$

If $Z \gamma$ not perfectly collinear with this term (e.g. one or more elements of $X$ not in $Z$) and we have exclusion (elements of $\mu^1_1(X) + \mu^1_2(X) - [\mu^0_1(X) + \mu^0_2(X)]$ that are known so we can fix), we can identify

$$\left(\sigma^2_{\varepsilon_i} + \alpha^2_{i,1} \sigma^2_{\theta_1} + \alpha^2_{i,2} \sigma^2_{\theta_2}\right)^{\frac{1}{2}}$$

(This is the reasoning used earlier)

$\therefore$ we identify $\gamma$ (get absolute scale on costs).
Testing What is in the Agent’s Information Set

Suppose agents do not know $\theta_2$ or the future $\epsilon_1^1, \epsilon_2^1, \epsilon_1^0, \epsilon_2^0$ but know $\epsilon_c$ and $\theta_1$.

Then if what they don’t know is set at mean zero, (they use rational expectations in a linear decision rule) and their mean forecast is the population mean, $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_c}^2$ and $\alpha_{l,2} = 0$.

What can we identify?
Is the model testable?
What information do we have about covariances?

Suppose we have two dedicated measurement systems for $\theta_1$ and $\theta_2$.

\[
\begin{align*}
M^1_1 &= \theta_1 + \varepsilon^{1}_{1,M} \\
M^1_2 &= \alpha^{1}_{2,M} \theta_1 + \varepsilon^{1}_{2,M} \\
M^1_3 &= \alpha^{1}_{3,M} \theta_1 + \varepsilon^{1}_{3,M}
\end{align*}
\] 
\begin{align*}
\text{Cognitive Ability}
\end{align*}

\[
\begin{align*}
M^2_1 &= \theta_2 + \varepsilon^{2}_{1,M} \\
M^2_2 &= \alpha^{2}_{2,M} \theta_2 + \varepsilon^{2}_{2,M} \\
M^2_3 &= \alpha^{2}_{3,M} \theta_2 + \varepsilon^{2}_{3,M}
\end{align*}
\] 
\begin{align*}
\text{Noncognitive Ability}
\end{align*}

(See e.g. Heckman, Urzua and Stixrud, 2006)
Observe from $M^1$ system we get

$$\text{Var} (\theta_1), \alpha_{2,M}^1, \alpha_{3,M}^1$$

From $M^2$ system we get

$$\text{Var} (\theta_2), \alpha_{2,M}^2, \alpha_{3,M}^2$$
Then

\[ \text{Cov} \left( U_1^1, M_1^1 \right) = \alpha_{1,1}^1 \sigma_{\theta_1}^2 \]
\[ \text{Cov} \left( U_2^1, M_1^1 \right) = \alpha_{1,2}^1 \sigma_{\theta_1}^2 \]

\[ \therefore \text{we get all of the factor loadings in } Y^1 \text{ on } \theta_1. \]

Using \( M_1^2 \) we get \( \alpha_{2,1}^1, \alpha_{2,2}^1 \) and we get variances of uniquenesses \( \text{Var} (\varepsilon_1^1), \text{Var} (\varepsilon_2^1) \).

By similar reasoning, we get

\[ \alpha_{1,1}^0, \alpha_{2,1}^0, \alpha_{1,2}^0, \alpha_{2,2}^0 \]
\[ \text{Var} (\varepsilon_1^0), \text{Var} (\varepsilon_2^1) \]
Observe from

\[
\text{Cov} \left( I, M_1^1 \right) = \frac{\sigma^2_{\theta_1}}{\sigma_I} \left[ \alpha_{1,1}^1 + \alpha_{1,2}^1 - \left( \alpha_{1,1}^0 + \alpha_{1,2}^0 \right) - \alpha_{1,C} \right]
\]

\[
\therefore \text{We can get } \alpha_{1C} \text{ up to scale } \sigma_I, \text{ since we know everything else by the previous reasoning.}
\]

But we have \( \sigma_I \). \( \therefore \) we have \( \alpha_{1C} \)

From

\[
\text{Cov} \left( I, M_1^2 \right) = \frac{\sigma^2_{\theta_2}}{\sigma_I} \left[ \alpha_{2,1}^1 + \alpha_{2,2}^1 - \left( \alpha_{2,1}^0 + \alpha_{2,2}^0 \right) - \alpha_{2,C} \right]
\]

\[
\therefore \text{ we get } \alpha_{2C} \text{ up to scale } \sigma_I.
\]

But we know \( \sigma_I \). \( \therefore \) know \( \alpha_{2C} \).

From \( \text{Pr} \left( D = 1 \mid X, Z \right) \), we can identify \( \sigma_I \) using previous reasoning.
Therefore we can identify everything in the model if there is one $X$ not in $Z$ and we satisfy the absence of collinearity assumption since we can identify the terms in the numerator, since we know $\sigma_1$. 
But, can we test the model?

In the previous notation, we have that for a test of whether $\theta_2$ belongs in the model

$$\Pr(D = 1 \mid X, Z) = \Phi \left[ \frac{\mu^1_1(X) + \mu^1_2(X) - [\mu^0_1(X) + \mu^0_2(X)] - Z\gamma}{\left(\sigma^2_{\varepsilon} + \alpha^2_{l,1}\sigma^2_{\theta_1} + \alpha^2_{l,2}\sigma^2_{\theta_2}\Delta_{\theta_2}\right)^{\frac{1}{2}}} \right]$$

Apparently, we can test the null

$$H_0 : \Delta_{\theta_2} = 0$$ so we can test if $\theta_2$ components enter or not
The problem with this test is that if $\sigma^2_{\varepsilon_c} \neq 0$, we can always adjust its value to fit the model to rationalize the component arising from $\sigma^2_{\theta_2}$. If we have a pure Roy model, the test is clean. A pure Roy model assumes $\sigma^2_{\varepsilon_c} = 0$ so this identification issue is moot.

Notice, however, that we can tolerate $\gamma \neq 0$ so long as $\sigma^2_{\varepsilon_c} = 0$. Thus we can depart from the Roy model somewhat.

Basic point: we don’t observe costs directly. \because we do not get a clean measurement on $\sigma^2_{\varepsilon_c}$. We can identify $\sigma^2_I$ but the problem is that $\sigma^2_{\varepsilon_c}$ can be adjusted to fit the data with or without $\sigma^2_{\theta_2}$. 


A Better Test:

Form

\[
\text{Cov} \left( \frac{I}{\sigma_I}, U_1^1 \right) = \frac{\sigma_{\theta_1}^2}{\sigma_I} \alpha_{1,1}^1 \left[ \alpha_{1,1}^1 + \alpha_{1,2}^1 - (\alpha_{1,1}^0 + \alpha_{1,2}^0) - \alpha_{1,c} \right] \\
+ \Delta_{\theta_2} \sigma_{\theta_2}^2 \alpha_{2,1}^1 \left[ \alpha_{2,1}^1 + \alpha_{2,2}^1 - (\alpha_{2,1}^0 + \alpha_{2,2}^0) - \alpha_{2,c} \right]
\]

we can compute the left hand side under the null. (with exclusion, lack of collinearity and normality).

Under the null that \( \Delta_{\theta_2} = 0 \), we can identify \( \sigma_{\varepsilon c}^2 \).

\[\therefore\text{we construct a test under null:}\]

\[
\text{Cov} \left( \frac{I}{\sigma_I}, U_1^1 \right) - \frac{\sigma_{\theta_1}^2}{\sigma_I} \alpha_{1,1}^1 \left[ \alpha_{1,1}^1 + \alpha_{1,2}^1 - (\alpha_{1,1}^0 + \alpha_{1,2}^0) - \alpha_{1,c} \right] = 0
\]

We know both terms under the null. Departures from null are evidence that agents know \( \theta_2 \).
It is assumed that if the agent knows $\theta_1$ but not $\theta_2$, he sets

$$E(\theta_2) = 0.$$ 

This is justified by linearity of the criterion and rational expectations, assuming $E(\theta_2 | I_0) = 0$. 
Then we can test among models by deciding

- Which model fits the data better?

Average effect (we estimate the average probability):

\[
\int \Pr (S = 1 \mid X, Z, \theta_1, \Delta \theta_2, \theta_2) f (\theta_1) f (\theta_2) d\theta.
\]

(we test \( \Delta \theta_2 = 0 \))

This is what is done in the Hicks lecture.

Don’t need normality.
If we can extract factor scores from measurements, we can test directly if $\theta_1$ and $\theta_2$ enter directly into the choice equation.

In this way we do not have to assume the agent sets $\theta_2 = 0$.

It is really a test of $\alpha_2 = 0$. 
Recovering the Distributions Nonparametrically

Theorem

Suppose that we have two random variables $T_1$ and $T_2$ that satisfy:

\[
T_1 = \theta + v_1 \\
T_2 = \theta + v_2
\]

with $\theta, v_1, v_2$ mutually statistically independent, $E(\theta) < \infty$, $E(v_1) = E(v_2) = 0$, that the conditions for Fubini’s theorem are satisfied for each random variable, and the random variables possess nonvanishing (a.e.) characteristic functions, then the densities $f(\theta), f(v_1), f(v_2)$ are identified.

Proof.

See Kotlarski (1967).
\[ I^* = \mu_{I^*}(X, Z) + \alpha_{I^*} \theta + \varepsilon_{I^*} \]
\[ Y_0 = \mu_0(X) + \alpha_0 \theta + \varepsilon_0 \]
\[ Y_1 = \mu_1(X) + \alpha_1 \theta + \varepsilon_1 \]
\[ M = \mu_M(X) + \theta + \varepsilon_M. \]

System can be rewritten as

\[ \frac{I^* - \mu_{I^*}(X, Z)}{\alpha_{I^*}} = \theta + \frac{\varepsilon_{I^*}}{\alpha_{I^*}} \]
\[ \frac{Y_0 - \mu_0(X)}{\alpha_0} = \theta + \frac{\varepsilon_0}{\alpha_0} \]
\[ \frac{Y_1 - \mu_1(X)}{\alpha_1} = \theta + \frac{\varepsilon_1}{\alpha_1} \]
\[ \frac{M - \mu_M(X)}{\alpha_1} = \theta + \varepsilon_M \]
Applying Kotlarski’s theorem, identify the densities of
\( \theta, \frac{\varepsilon_{I^*}}{\alpha_{I^*}}, \frac{\varepsilon_0}{\alpha_0}, \frac{\varepsilon_1}{\alpha_1}, \varepsilon_M. \)

We know \( \alpha_{I^*}, \alpha_0 \) and \( \alpha_1. \) Can identify the densities of 
\( \theta, \varepsilon_{I^*}, \varepsilon_0, \varepsilon_1, \varepsilon_M. \) Recover the joint distribution of \( (Y_1, Y_0) \)

\[
F(Y_1, Y_0 \mid X) = \int F(Y_1, Y_0 \mid \theta, X) \, dF(\theta).
\]

\( F(\theta) \) is known

\[
F(Y_1, Y_0 \mid \theta, X) = F(Y_1 \mid \theta, X) \, F(Y_0 \mid \theta, X).
\]

\( F(Y_1 \mid \theta, X) \) and \( F(Y_0 \mid \theta, X) \) identified

\[
F(Y_1 \mid \theta, X, S = 1) = F(Y_1 \mid \theta, X)
\]

\[
F(Y_0 \mid \theta, X, S = 0) = F(Y_0 \mid \theta, X).
\]

Can identify the number of factors generating dependence among the \( Y_1, Y_0, C, S \) and \( M. \)
• Crucial idea: even though we never observe \((y_1, y_0)\) as a pair, both \(y_0\) and \(y_1\) are linked to \(S\) through the choice equation (I) or a measurement equation (M).

• Can extend to nonseparable models (Cunha, Heckman, and Schennach, 2010)

• For the other market structures the decision rule is no longer linear (solution to a dynamic programming problem). That is

\[
I = E_{I_0} (V_1 (X, \theta, \varepsilon_{1,1}, a_0; \phi) - V_0 (X, \theta, \varepsilon_{0,1}, a_0; \phi) - Z' \gamma - \theta' \lambda - \varepsilon_{\text{cost}})
\]

• The argument still goes through using external measurements like the test equations instead of the choice equation as common identifying relationships.
Alternatively, we can also identify the factor loadings using nonsymmetric $\theta$ (nonlinear factor analysis).

Can fit model, determine the number of factors and generate counterfactuals.

$\phi$ is identified. Obvious if we use consumption data. Also true without consumption. Under large support conditions, factors and uniquenesses nonparametrically identified are means. We maintain separability.

Can extend to multiple periods and multiple schooling levels (Heckman and Navarro, 2007)


