Adding Uncertainty to a Roy Economy with Two Sectors

James J. Heckman
The University of Chicago

Nuffield College, Oxford University
This draft, August 7, 2005
$S$ denotes different sectors.

$S = 0$ denotes choice of the high school sector, and $S = 1$ denotes choice of the college sector.

$C$ reflects the cost associated with choosing the college sector.

\[
Y_1 = \sum_{t=0}^{T} \frac{Y_{1_{it}}}{(1 + r)^t},
\]

\[
Y_0 = \sum_{t=0}^{T} \frac{Y_{0_{it}}}{(1 + r)^t},
\]

$Y_1, Y_0$ and $C$ are \textit{ex post} realizations of cost and returns.
\( \mathcal{I}_0 \) denotes the information set of the agent at time period \( t = 0 \).

\[ S = \begin{cases} 
1, & \text{if } E(Y_1 - Y_0 - C \mid \mathcal{I}_0) \geq 0 \\
0, & \text{otherwise.} 
\end{cases} \]

**Essential Idea**

Suppose, contrary to what is possible, analyst observes \( Y_0, Y_1 \) and \( C \).

**Ideal data set**

Observe two different lifetimes

Construct \( Y_1 - Y_0 - C \) from *ex post* lifetime data.
Information set $\mathcal{I}_0$ of the agent. We seek to construct $E (Y_1 - Y_0 - C | \mathcal{I}_0)$. Suppose we assume we know the right information set ($\tilde{\mathcal{I}}_0 = \mathcal{I}_0$).

We then obtain

$$V_{\tilde{\mathcal{I}}_0} = (Y_1 - Y_0 - C) - E \left( Y_1 - Y_0 - C | \tilde{\mathcal{I}}_0 \right)$$

Our test is to determine if $S$ depends on $V_{\tilde{\mathcal{I}}_0}$.

Test for correct specification of $\mathcal{I}_0$: test if the coefficient on $V_{\tilde{\mathcal{I}}_0}$ in a discrete choice equation for $S$ is different from zero.

Search among candidate information sets $\tilde{\mathcal{I}}_0$ to determine which ones satisfy the requirement that the generated $V_{\tilde{\mathcal{I}}_0}$ does not predict $S$.

Procedure is in the form of a Sims (1972) version of a Wiener-Granger causality test.

It is also a test for misspecification of the information.
Components of $V_{\bar{I}_0}$ that do not predict $S$ are called intrinsic components of uncertainty.

Procedure as stated not practical.

We do not observe $Y_1$ and $Y_0$ together for anyone.
Specifics of the **BASIC STRATEGY** in complete market case.

Linear in parameters model:

\[
Y_{0it} = X_{it} \beta_{0t} + v_{0it}, \quad t = 0, \ldots, T \\
Y_{1it} = X_{it} \beta_{1t} + v_{1it} \\
C_i = Z_i \gamma + v_{iC}.
\]

\(\theta = (\theta_1, \theta_2, \ldots, \theta_L), \theta_i\) and \(\theta_j\) mutually independent random variables for \(i \neq j\).

\[
v_{0it} = \theta_i \alpha_{0t} + \varepsilon_{0it} \\
v_{1it} = \theta_i \alpha_{1t} + \varepsilon_{1it},
\]

where \(\alpha_{0t}\) and \(\alpha_{1t}\) are vectors.

\[
C_i = Z_i \gamma + \theta_i \alpha_C + \varepsilon_{iC}.
\]
Choice Equation:

\[ I = E \left( \sum_{t=0}^{T} \frac{(X_{it}\beta_{1t} + \theta_i\alpha_{1t} + \varepsilon_{1it}) - (X_{it}\beta_{0t} + \theta_i\alpha_{0t} + \varepsilon_{0it})}{(1 + r)^t} - (Z_i\gamma + \theta_i\alpha_C + \varepsilon_iC) \right| \mathcal{I}_0 \) 

\[ S = 1 \text{ if } I \geq 0; \ S = 0 \text{ otherwise.} \]

Let \( \odot \) denote the Hadamard product \((a \odot b = (a_1b_1, \ldots, a_Lb_L))\)
For candidate information set $\tilde{I}_0$,

$$
I = \sum_{t=0}^T \frac{E\left(X_{it} \mid \tilde{I}_0\right)}{(1 + r)^t} (\beta_{1t} - \beta_{0t}) + \sum_{t=0}^T \frac{[X_{it} - E\left(X_{it} \mid \tilde{I}_0\right)]}{(1 + r)^t} (\beta_{1t} - \beta_{0t}) \odot \Delta x
$$

$$
+ E(\theta_i \mid \tilde{I}_0) \left[ \sum_{t=0}^T \frac{(\alpha_{1t} - \alpha_{0t})}{(1 + r)^t} - \alpha_C \right] + [\theta_i - E(\theta_i \mid \tilde{I}_0)] \left\{ \left[ \sum_{t=0}^T \frac{(\alpha_{1t} - \alpha_{0t})}{(1 + r)^t} - \alpha_C \right] \odot \Delta \theta \right\}
$$

$$
+ \sum_{t=0}^T \frac{E(\varepsilon_{1it} - \varepsilon_{0it} \mid \tilde{I}_0)}{(1 + r)^t} + \sum_{t=0}^T \left[ (\varepsilon_{1it} - \varepsilon_{0it}) - E(\varepsilon_{1it} - \varepsilon_{0it} \mid \tilde{I}_0) \right] \Delta \varepsilon_t
$$

$$
- E\left(Z_i \mid \tilde{I}_0\right) \gamma - [Z_i - E\left(Z_i \mid \tilde{I}_0\right)] \odot \Delta Z - E\left(\varepsilon_i \mid \tilde{I}_0\right) - \varepsilon_i \odot \Delta \varepsilon
$$
A test of the validity of information set $\tilde{I}_0$ is that $\Delta_X = 0; \Delta_\theta = 0; \Delta_Z = 0; \Delta_{\varepsilon_C} = 0$ and $\Delta_{\varepsilon_t} = 0, \forall t$. Components associated with zero $\Delta$’s are the unforecastable elements.

- We test what components of the future income process unobservable to the econometrician enter the agents information set and are acted on by the agent at the time of his schooling decision and so are not components of uncertainty but rather components of heterogeneity.

- This procedure can be generalized so $\theta$ becomes $\theta_t$, a hidden state Markov process.
How is the Model Identified?

Problem: We need to construct counterfactuals. We only observe earnings in college for people who choose college and earnings in high school for people who choose high school. We can never form the covariance between college and high school in the raw data.

Solution: Extension of Factor Models to Nonlinear Settings (see Goldberger and Jöreskog, *MIMIC*, 1972; *LISREL*, Jöreskog, 1977 for linear versions)

Consider a simple example that motivates the main idea.
We observe

\[(Y_{1,1}, \ldots, Y_{1,T}) \text{ for } s = 1\]
\[(Y_{0,1}, \ldots, Y_{0,T}) \text{ for } s = 0\]

Let net utility of \( s = 1 \) be represented by the index.

\[I = \mu_I (X, Z) + U_I, \quad Z \text{ instruments}\]

\[Y_{1,t} = \mu_{1,t} (X) + U_{1,t} \quad t = 1, \ldots, T\]
\[Y_{0,t} = \mu_{0,t} (X) + U_{0,t} \quad t = 1, \ldots, T\]
We know $F(Y_{1,1}, \ldots, Y_{1,T} \mid S = 1, X, Z)$ and $F(Y_{0,1}, \ldots, Y_{0,T} \mid S = 0, X, Z)$.

Under conditions on regressors and support of $Z$ (Heckman (1990), Heckman and Smith (1998), Carneiro, Heckman and Hansen (2003)), we can identify from these distributions

$$\mu_{0,t}(X), \mu_{1,t}(X), \mu_I(X, Z) \quad \text{(up to scale)} \ t = 1, \ldots, T$$

and the joint distribution of

$$F(Y_{1,1}, \ldots, Y_{1,T}, I \mid X, Z)$$
$$F(Y_{0,1}, \ldots, Y_{0,T}, I \mid X, Z)$$

with the scale of $I$ not determined (must be normalized). $I^*$ is the normalized index.
Motivation on the Nonparametric Identification of The Joint Distribution of Outcomes and The Binary Choice Equation

Motivate why \( F(Y_0, I^* \mid X, Z) \) is identified. (Take simple case: two potential outcomes; \( T = 0 \) so one period model for simplicity only)

From Cosslett (1983), Manski (1988) and Matzkin (1992)

Can identify \( \frac{\mu_I(X, Z)}{\sigma_I} \) from \( \Pr(S = 1 \mid X, Z) = \Pr(\mu_I(X, Z) + U_I \geq 0 \mid X, Z) \)
(Support conditions and continuous regressors).

Can identify distribution of \( \frac{U_I}{\sigma_I} \).
From this information and

\[ F(Y_0 \mid S = 0, X, Z) = \Pr(Y_0 \leq y_0 \mid \mu_1(X, Z) + U_I \leq 0, X, Z) \]

Form

\[ F(Y_0 \mid S = 0, X, Z) \Pr(S = 0 \mid X, Z) = \Pr(Y_0 \leq y_0, I^* \leq 0 \mid X, Z) \]

Left hand side known.

Right hand side:

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

Since we know $\frac{\mu_I(X, Z)}{\sigma_I}$, 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

$$Y_0 = \mu_0(X) + U_0 \quad : \quad \text{can identify}$$

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

$X$ and $Z$ can be varied and $y_0$ is a number.

Trace out joint distribution of $\left( U_0, \frac{U_I}{\sigma_I} \right)$. 

14
Recover joint distribution of

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

Three key ingredients.

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

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

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

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

\[ F(Y_0 \mid S = 0, X, Z) \Pr(S = 0 \mid X, Z) \]

This is a function of \( \mu_0(X) \) and \( \frac{\mu_I(X, Z)}{\sigma_I} \) (Index sufficiency)

Varying the \( \mu_0(X) \) and \( \frac{\mu_I(X, Z)}{\sigma_I} \) traces out the distribution of \( U_0, \frac{U_I}{\sigma_I} \).

This means effectively that we observe \( \left( \frac{I}{\sigma_I}, Y_1 \right), \left( \frac{I}{\sigma_I}, Y_0 \right) \)

We do not observe \( \left( \frac{I}{\sigma_I}, Y_0, Y_1 \right) \)
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 \(\theta\)

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

\[
E(\theta) = 0, \quad \text{and} \quad E(\theta^2) = \sigma_{\theta}^2.
\]
\[
E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_{I*}) = 0
\]
\[
Var(\varepsilon_0) = \sigma_{\varepsilon_0}^2, \quad Var(\varepsilon_1) = \sigma_{\varepsilon_1}^2
\]
\[
Var(\varepsilon_{I}) = \sigma_{\varepsilon_{I}}^2
\]

17
1 Factor Models: A Brief Digression

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

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

\[ \text{Cov}(R_1, R_2) = \alpha_1 \alpha_2 \sigma_{\theta}^2 \]
\[ \text{Cov}(R_1, R_3) = \alpha_1 \alpha_3 \sigma_{\theta}^2 \]
\[ \text{Cov}(R_2, R_3) = \alpha_2 \alpha_3 \sigma_{\theta}^2 \]

Normalize \( \alpha_1 = 1 \)

\[ \frac{\text{Cov}(R_2, R_3)}{\text{Cov}(R_1, R_2)} = \alpha_3 \]
::: We know $\sigma_\theta^2$ from $Cov(R_1, R_2)$. From $Cov(R_1, R_3)$ we know

$$\alpha_3, \alpha_4, \alpha_5.$$ 

Can get the variances of the $\varepsilon_i$ from variances of the $R_i$

$$Var(R_i) = \alpha_i^2 \sigma_\theta^2 + \sigma_{\varepsilon_i}^2.$$ 

If $T = 2$, all we can identify is $\alpha_1 \alpha_2 \sigma_\theta^2$.

If $\alpha_1 = 1$, $\sigma_\theta^2 = 1$, we identify $\alpha_2$. 

19
2 Factors:

\[ \theta_1 \perp \theta_2 \]

\[ \varepsilon_i \perp \varepsilon_j \ \forall i, j \]

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

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

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

\[ \text{Cov}(R_1, R_4) = \alpha_{41} \sigma_{\theta_1}^2, \quad \therefore \text{we get } \sigma_{\theta_1}^2 \quad \therefore \text{we get } \alpha = 1 \]

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

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

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

\[ \therefore \text{get } \sigma^2_{\theta_2} \text{ of “2” loadings.} \]
If we have dedicated measurements do not need a normalization on $R$.

\[
M_1 = \theta_1 + \varepsilon_{1M}
\]
\[
M_2 = \theta_2 + \varepsilon_{2M}
\]

\[
Cov (R_1, M) = \alpha_{11} \sigma^2_{\theta_1}
\]
\[
Cov (R_2, M) = \alpha_{21} \sigma^2_{\theta_1}
\]
\[
Cov (R_3, M) = \alpha_{31} \sigma^2_{\theta_1}
\]

\[
Cov (R_1, R_2) = \alpha_{11} \alpha_{12} \sigma^2_{\theta_1},
\]
\[
Cov (R_1, R_3) = \alpha_{11} \alpha_{13} \sigma^2_{\theta_1}, \therefore \alpha_{12} \sigma^2_{\theta_1},
\]

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

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

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

\[ E(\varepsilon) = 0 \]

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

\[ E(\theta) = 0 \]

\[ Var(R) = \Lambda \Sigma_\theta \Lambda' + D \quad \Sigma_\theta = E(\theta \theta') \]
The only source of information on $\Lambda$ and $\Sigma_\theta$ is from the covariances.

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

Each variance contributes one new parameter.

How many unique covariance terms do we have?

$$\frac{T(T-1)}{2}$$ This is the data.

We have $T$ uniquenesses; $TK$ elements of $\Lambda$.

$$\frac{K(K-1)}{2}$$ elements of $\Sigma_\theta$. $$\frac{K(K-1)}{2} + TK$$ parameters $(\Sigma_\theta, \Lambda)$. 
Observe that if we multiply \( \Lambda \) by an orthogonal matrix \( C \), \((CC'' = I)\), we have

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

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

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

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

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

(ii)

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

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

# of free parameters \hspace{1cm} data

“Ledermann Bound”

Can get more information by looking at higher order moments.
Recovering the Factor Loadings in the Roy Model (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.
Since
\[ \text{Cov} (U_1, U_0) = \alpha_1 \alpha_0 \sigma_{\theta}^2 \]
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 \( T > 0 \))

\[ \frac{\text{Cov} (Y_{1t'}, I^*)}{\text{Cov} (Y_{1t'}, Y_{1t})} = \alpha_{1t} \]
\[ \frac{\text{Cov} (Y_{0t'}, I^*)}{\text{Cov} (Y_{0t'}, Y_{0t})} = \alpha_{0t} \]
Crucial Idea of Identification

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

From \(S\) we can generate \(I^*\).

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 to Identify the Model

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

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

$$\text{Cov} (M, Y_0) = \text{Cov} (U_M, U_0) = \alpha_M \alpha_0 \sigma_\theta^2$$

$$\text{Cov} (M, Y_1) = \text{Cov} (U_M, U_1) = \alpha_M \alpha_1 \sigma_\theta^2$$

$$\text{Cov} (M, I^*) = \text{Cov} (U_M, U_{I^*}) = \alpha_M \alpha_{I^*} \sigma_\theta^2.$$ 

$\alpha_M = 1.$
Can form the ratios

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

Recover $\alpha_1$:
\[
\frac{Cov(U_1, U_{I^*})}{Cov(U_M, U_{I^*})} = \alpha_1
\]
\[
Cov(U_M, U_0) = \alpha_0 \sigma_\theta^2
\]

Can identify $\alpha_{I^*}$.
\[
Cov(U_M, U_{I^*}) = \alpha_{I^*} \sigma_\theta^2
\]
2 Intuition on Identification of the Normal Case Model

Generalized Roy versions of model:

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

(Measurement: A test score equation)

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

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

Cost

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

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

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

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

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

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

\[
\begin{align*}
Y^1_1 &= \mu^1_1(X) + U^1_1 \\
Y^1_2 &= \mu^1_2(X) + U^1_2 \\
Y^0_1 &= \mu^0_1(X) + U^0_1 \\
Y^0_2 &= \mu^0_2(X) + U^0_2
\end{align*}
\]

\( U^1_1, U^1_2 \text{ etc. match the error terms previously shown.} \)

\[
\begin{align*}
U^1_1 &= \theta_1\alpha^1_{1,1} + \theta_2\alpha^1_{2,1} + \varepsilon^1_1 \text{ etc.} \\
U_M &= \theta_1\alpha_{1,M} + \theta_2\alpha_{2,M} + \varepsilon_M
\end{align*}
\]

37
\[
E (Y_1^1 \mid X, Z, I > 0) = \mu_1^1 (X) + \frac{\text{Cov}(U_1^1, I)}{\text{Var}(I)} \lambda ()
\]

Using standard sample selection bias arguments we can identify beside the means,

\[
\mu_1^1 (X), \mu_2^1 (X), \mu_2^0 (X), \mu_2^0 (X), \text{the following parameters:}
\]

\[
\text{Cov} (U_1^1, U_2^1), \text{Var} (U_1^1), \text{Var} (U_2^1) \\
\text{Cov} (U_1^1, U_M), \text{Cov} (U_2^1, U_M), \text{Var} (U_M) \\
\text{Cov} (U_1^0, U_2^0), \text{Var} (U_1^0), \text{Var} (U_2^0) \\
\text{Cov} (U_1^0, U_M), \text{Cov} (U_2^0, U_M)
\]
Normal Case:

\[(\theta, \varepsilon) \perp (X, Z) : (\theta, \varepsilon) \text{ normal.}\]
Fact:

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

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

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

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

Why? $S = 1 \left[ X\beta > V - \theta \right]$. Rest follows from independence (between $V - \theta$, and $X$, and normality).
Unconditional Probability:

\[
\Pr(S = 1 \mid X, Z) = \Phi \left[ \frac{\mu_1^1(X) + \mu_2^1(X) - [\mu_1^0(X) + \mu_2^0(X)] - Z\gamma}{(\sigma_{\varepsilon}^2 + \alpha^2_{I,1}\sigma_{\theta_1}^2 + \alpha^2_{I,2}\sigma_{\theta_2}^2)^{1/2}} \right]
\]

Observe that if we know \(\mu_1^1(X), \mu_2^1(X), \mu_1^0(X), \mu_2^0(X)\) we know

\[
[\mu_1^1(X) + \mu_2^1(X)] - [\mu_1^0(X) + \mu_2^0(X)].
\]

If \(Z\gamma\) not perfectly collinear with this term (e.g. one or more elements of \(X\) not in \(Z\)) we can identify

\[
(\sigma_{\varepsilon}^2 + \alpha^2_{I,1}\sigma_{\theta_1}^2 + \alpha^2_{I,2}\sigma_{\theta_2}^2)^{1/2}
\]

\(\therefore\) we also identify \(\gamma\) (get absolute scale on costs).
Suppose agents do not know $\theta_2$ or the future $\varepsilon_1^1, \varepsilon_2^1, \varepsilon_1^0, \varepsilon_2^0$ but know $\varepsilon_c$ and $\theta_1$.

Then if what they know is set at mean zero, (they use rational expectations in a linear decision rule) and their mean forecast is the population mean,

$$\sigma_{\varepsilon_I}^2 = \sigma_{\varepsilon_c}^2$$

and $\alpha_{I,2} = 0$.

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 $\theta_2$.

\[
\begin{align*}
M_1^1 &= \theta_1 + \varepsilon_{1,M}^1 \\
M_2^1 &= \alpha_{2,M}^1 \theta_1 + \varepsilon_{1,M}^1 \\
M_3^1 &= \alpha_{3,M}^1 \theta_1 + \varepsilon_{1,M}^1 \\
\end{align*}
\right\} \text{ Cognitive Ability}
\]

\[
\begin{align*}
M_1^2 &= \theta_2 + \varepsilon_{1,M}^2 \\
M_2^2 &= \alpha_{2,M}^2 \theta_2 + \varepsilon_{2,M}^2 \\
M_3^2 &= \alpha_{3,M}^2 \theta_2 + \varepsilon_{2,M}^2 \\
\end{align*}
\right\} \text{ Noncognitive Ability}
\]

(See e.g. Heckman, Urzua and Stixrud, 2004)
Observe from $M^1$ system we get

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

From $M^2$ system we get

$$Var(\theta_2), \alpha_{2,M}^2, \alpha_{3,M}^2$$
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 \( \text{Var} \left( \varepsilon_1^1 \right), \text{Var} \left( \varepsilon_2^1 \right) \).

By similar reasoning, we get

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

\[ \text{Cov} (I, M^1_1) = \sigma^2_{\theta_1} \left[ \alpha^1_{1,1} + \alpha^1_{1,2} - (\alpha^0_{1,1} + \alpha^0_{1,2}) - \alpha_{1,C} \right] \]

∴ We can get \( \alpha_{1C} \) up to scale \( \sigma_I \), since we know everything else by the previous reasoning.

From

\[ \text{Cov} (I, M^2_1) = \sigma^2_{\theta_2} \left[ \alpha^1_{2,1} + \alpha^1_{2,2} - (\alpha^0_{2,1} + \alpha^0_{2,2}) - \alpha_{2,C} \right] \]

∴ we get \( \alpha_{2C} \) up to scale \( \sigma_I \).

From \( \Pr (S = 1 \mid X, Z) \), we can identify \( \sigma^2_{\varepsilon_I} \) using previous reasoning
Therefore we have that we can identify everything in the model if there is one $X$ not in $Z$ since we can identify the terms in the numerator.
But, can we test the model?

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

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

Apparently, we can test the null

$$H_0 : \Delta_{\theta_2} = 0$$

$\therefore$ apparently we can test if $\theta_2$ components enter or not.
The problem with this test is that if $\sigma_{\varepsilon c}^2 \neq 0$, we can always adjust its value to fit the model perfectly well. If we have a pure Roy model, the test is clean. A pure Roy model assumes $\sigma_{\varepsilon c}^2 = 0$.

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

Basic point: we don’t observe costs directly. \(\therefore\) we do not get a clean measurement on $\sigma_{\varepsilon c}^2$. We can identify $\sigma_I^2$ but the problem is that $\sigma_{\varepsilon c}^2$ can be adjusted.
Correct Test:

Form

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

we can compute the left hand side under the null. (with exclusion and normality). We identify all components of right hand side by a separate argument (from measurement systems).

Thus 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. (We do not use the information on \( \text{Cov} \left( \frac{I}{\sigma_I}, U_1^1 \right) \). Departures 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 $E(\theta_2 \mid I_0) = 0$. 
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.
Theorem 1 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 $f(\theta), f(v_1), \text{ and } f(v_2)$ 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

\[
\frac{I^* - \mu_{I^*}(X, Z)}{\alpha_{I^*}} = \theta + \frac{\varepsilon_{I^*}}{\alpha_{I^*}}
\]
\[
\frac{Y_0 - \mu_0(X)}{\alpha_0} = \theta + \frac{\varepsilon_0}{\alpha_0}
\]
\[
\frac{Y_1 - \mu_1(X)}{\alpha_1} = \theta + \frac{\varepsilon_1}{\alpha_1}
\]
\[ M - \mu_M(X) = \theta + \varepsilon_M \]
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 (Heckman, Matzkin, Navarro and Urzua, 2004)

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

\[
I = E_{\mathcal{I}_0} (V_1 (X, \theta, \varepsilon_{1,1}, a_0; \phi) - V_0 (X, \theta, \varepsilon_{0,1}, a_0; \phi) - Z'\gamma - \theta'\lambda - \varepsilon_{\text{cost}})
\]

- 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, 2004)
Present value of earnings from age 19 to 28 discounted using an interest rate of 3%. Let \((Y_0, Y_1)\) denote potential outcomes in high school and college sectors, respectively. Let \(S=0\) denote high school sector, and \(S=1\) denote college sector. Define observed earnings as \(Y=SY_1+(1-S)Y_0\). Finally, let \(f(y)\) denote the density function of observed earnings. Here we plot the density functions \(f\) generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 1.1
Densities of fitted and actual present value of earnings from age 19 to 28 for overall sample
Present value of earnings from age 29 to 38 discounted using an interest rate of 3%. Let \((Y_0, Y_1)\) denote potential outcomes in high school and college sectors, respectively. Let \(S=0\) denote high school sector, and \(S=1\) denote college sector. Define observed earnings as \(Y = SY_1 + (1-S)Y_0\). Finally, let \(f(y)\) denote the density function of observed earnings. Here we plot the density functions \(f\) generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 39 to 48 discounted using an interest rate of 3%. Let \((Y_0, Y_1)\) denote potential outcomes in high school and college sectors, respectively. Let \(S=0\) denote high school sector, and \(S=1\) denote college sector. Define observed earnings as \(Y=SY_1+(1-S)Y_0\). Finally, let \(f(y)\) denote the density function of observed earnings. Here we plot the density functions \(f\) generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 49 to 58 discounted using an interest rate of 3%. Let \((Y_0, Y_1)\) denote potential outcomes in high school and college sectors, respectively. Let \(S=0\) denote high school sector, and \(S=1\) denote college sector. Define observed earnings as \(Y=SY_1+(1-S)Y_0\). Finally, let \(f(y)\) denote the density function of observed earnings. Here we plot the density functions \(f\) generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 59 to 65 discounted using an interest rate of 3%. Let \((Y_0, Y_1)\) denote potential outcomes in high school and college sectors, respectively. Let \(S=0\) denote high school sector, and \(S=1\) denote college sector. Define observed earnings as \(Y=SY_1+(1-S)Y_0\). Finally, let \(f(y)\) denote the density function of observed earnings. Here we plot the density functions \(f\) generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Figure 2.1
Densities of fitted and actual present value of earnings from age 19 to 28 for people who choose to graduate high school

Present value of earnings from age 19 to 28 discounted using an interest rate of 3%. Earnings here are $Y_0$. Here we plot the density functions $f(y_0|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 29 to 38 discounted using an interest rate of 3%. Earnings here are $Y_0$. Here we plot the density functions $f(y_0|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 39 to 48 discounted using an interest rate of 3%. Earnings here are $Y_0$.

Here we plot the density functions $f(y_0 | S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 49 to 58 discounted using an interest rate of 3%. Earnings here are $Y_0$.

Here we plot the density functions $f(y_0|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 59 to 65 discounted using an interest rate of 3%. Earnings here are $Y_0$. Here we plot the density functions $f(y_0|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 19 to 28 discounted using an interest rate of 3%. This plot is for $Y_1$. Here we plot the density functions $f(y_1|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 29 to 38 discounted using an interest rate of 3%. This plot is for \( Y_1 \). Here we plot the density functions \( f(y_1 | S=1) \) generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Figure 3.3
Densities of fitted and actual present value of earnings from age 39 to 48 for people who choose to graduate college

Present value of earnings from age 39 to 48 discounted using an interest rate of 3%. This plot is for $Y_1$.
Here we plot the density functions $f(y_1|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 49 to 58 discounted using an interest rate of 3%. This plot is for $Y_1$. Here we plot the density functions $f(y_1|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Present value of earnings from age 59 to 65 discounted using an interest rate of 3%. This plot is for $Y_1$. Here we plot the density functions $f(y_1|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.
Let \( f(\theta_1) \) denote the probability density function of factor \( \theta_1 \). We allow \( f(\theta_1) \) to be a mixture of normals. Assume \( \mu_1 = \text{E}(\theta_1) \) and \( \sigma_1 = \text{Var}(\theta_1) \). Let \( \phi(\mu_1, \sigma_1) \) denote the density of a normal random variable with mean \( \mu_1 \) and variance \( \sigma_1 \). The solid curve is the actual density of factor \( \theta_1 \), \( f(\theta_1) \), while the dashed curve is the density of a normal random variable with mean \( \mu_1 \) and variance \( \sigma_1 \). We proceed similarly for factors 2 and 3.
Let $f(\theta_1)$ denote the probability density function of factor $\theta_1$. We allow $f(\theta_1)$ to be a mixture of normals. The solid line plots the density of factor 1 conditional on choosing the high school sector, that is, $f(\theta_1|\text{choice=high school})$. The dashed line plots the density of factor 1 conditional on choosing the college sector, that is, $f(\theta_1|\text{choice=college})$. 

Figure 5.1
Densities of "ability" (factor 1) by schooling level
Let $f(\theta_2)$ denote the probability density function of factor $\theta_2$. We allow $f(\theta_2)$ to be a mixture of normals. The solid line plots the density of factor 2 conditional on choosing the high school sector, that is, $f(\theta_2|\text{choice}=\text{high school})$. The dashed line plots the density of factor 2 conditional on choosing the college sector, that is, $f(\theta_2|\text{choice}=\text{college})$. 
Let $f(\theta_3)$ denote the probability density function of factor $\theta_3$. We allow $f(\theta_3)$ to be a mixture of normals. The solid line plots the density of factor 3 conditional on choosing the high school sector, that is, $f(\theta_3|\text{choice}=\text{high school})$. The dashed line plots the density of factor 3 conditional on choosing the college sector, that is, $f(\theta_3|\text{choice}=\text{college})$. 

Figure 5.3
Densities of factor 3 by schooling level
Let $Y_0$ denote the present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(Y_0)$ denote its density function. The dashed line plots the predicted $Y_0$ density conditional on choosing high school, that is, $f(Y_0|S=0)$, while the dashed line shows the counterfactual density function of $Y_0$ for those agents who are actually college graduates, that is, $f(Y_0|S=1)$. This assumes that the agent chooses schooling without knowing $\theta_3$ and $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...,T)$.
Let $Y_1$ denote the present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(Y_1)$ denote its density function. The dashed line plots the predicted $Y_1$ density conditional on choosing college, that is, $f(Y_1|S=1)$, while the solid line shows the counterfactual density function of $Y_1$ for those agents who are actually high school graduates, that is, $f(Y_1|S=0)$. This assumes that the agent chooses schooling without knowing $\theta_3$ and $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...,T)$.

Figure 6.2
Densities of ex post present value of counterfactual and fitted earnings from age 19 to 65 in the college sector

Let $Y_1$ denote the present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(Y_1)$ denote its density function. The dashed line plots the predicted $Y_1$ density conditional on choosing college, that is, $f(Y_1|S=1)$, while the solid line shows the counterfactual density function of $Y_1$ for those agents who are actually high school graduates, that is, $f(Y_1|S=0)$. This assumes that the agent chooses schooling without knowing $\theta_3$ and $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...,T)$.
Let $\varepsilon = (\varepsilon_{0,t}, \varepsilon_{1,t}, t=0, \ldots, T)$. Let $E_{\theta_3, \varepsilon}(Y_0)$ denote the ex ante present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(E_{\theta_3, \varepsilon}(Y_0))$ denote its density function. The solid curve plots the predicted $Y_0$ density conditional on choosing high school, that is, $f(E_{\theta_3, \varepsilon}(Y_0)|S=0)$, while the dashed line shows the counterfactual density function of $E_{\theta_3, \varepsilon}(Y_0)$ for those agents who are actually college graduates, that is, $f(E_{\theta_3, \varepsilon}(Y_0)|S=1)$. This assumes that the agent chooses schooling without knowing $\theta_3$ and $\varepsilon$. 

Figure 6.3

Densities of ex ante present value of counterfactual and fitted earnings from age 19 to 65 in the high school sector

Thousands of Dollars

Let $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0, \ldots, T)$. Let $E_{\theta_3, \varepsilon}(Y_0)$ denote the ex ante present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(E_{\theta_3, \varepsilon}(Y_0))$ denote its density function. The solid curve plots the predicted $Y_0$ density conditional on choosing high school, that is, $f(E_{\theta_3, \varepsilon}(Y_0)|S=0)$, while the dashed line shows the counterfactual density function of $E_{\theta_3, \varepsilon}(Y_0)$ for those agents who are actually college graduates, that is, $f(E_{\theta_3, \varepsilon}(Y_0)|S=1)$. This assumes that the agent chooses schooling without knowing $\theta_3$ and $\varepsilon$. 

79
Let $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...,T)$. Let $E_{\theta,\varepsilon}(Y_1)$ denote the ex ante present value of earnings from age 19 to 65 in the college sector (discounted at a 3\% interest rate). Let $f(E_{\theta,\varepsilon}(Y_1))$ denote its density function. The solid curve plots the counterfactual $Y_1$ density conditional on choosing high school, that is, $f(E_{\theta,\varepsilon}(Y_1)|S=0)$, while the dashed line shows the predicted density function of $E_{\theta,\varepsilon}(Y_1)$ for those agents who are actually college graduates, that is, $f(E_{\theta,\varepsilon}(Y_1)|S=1)$. This assumes that the agent chooses schooling without knowing $\theta_3$ and $\varepsilon$. 

Figure 6.4
Densities of ex ante present value of counterfactual and fitted earnings from age 19 to 65 in the college sector
Let $Y_0$ denote the present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(Y_0)$ denote its density function. The solid curve plots the predicted $Y_0$ density conditional on choosing high school, that is, $f(Y_0|S=0)$, while the dashed line shows the counterfactual density function of $Y_0$ for those agents who are actually college graduates, that is, $f(Y_0|S=1)$. This assumes that the agent chooses schooling with complete knowledge of future earnings.
Let $Y_1$ denote the present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(Y_1)$ denote its density function. The solid curve plots the counterfactual $Y_1$ density conditional on choosing high school, that is, $f(Y_1|S=0)$, while the dashed line shows the predicted density function of $Y_1$ for those agents who are actually college graduates, that is, $f(Y_1|S=1)$. This assumes that the agent chooses schooling with complete knowledge of future earnings.

Figure 6.6
Densities of present value of counterfactual and fitted earnings from age 19 to 65 assuming perfect certainty in the college sector

Let $Y_1$ denote the present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(Y_1)$ denote its density function. The solid curve plots the counterfactual $Y_1$ density conditional on choosing high school, that is, $f(Y_1|S=0)$, while the dashed line shows the predicted density function of $Y_1$ for those agents who are actually college graduates, that is, $f(Y_1|S=1)$. This assumes that the agent chooses schooling with complete knowledge of future earnings.
Let $Y_0, Y_1$ denote the present value of earnings in the high school and college sectors, respectively. Define ex post returns to college as the ratio $R = (Y_1 - Y_0) / Y_0$. Let $f(r)$ denote the density function of the random variable $R$. The solid line is the density of ex post returns to college for high school graduates, that is, $f(r|S=0)$. The dashed line is the density of ex post returns to college for college graduates, that is, $f(r|S=1)$. This assumes that the agent chooses schooling without knowing $\theta_3$ and $\varepsilon = (\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,\ldots,T)$.

Figure 7.1
Densities of ex post returns to college by level of schooling chosen

Fraction of the Base State

High School
College
Let $\varepsilon = (\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...,T)$. Let $Y_0, Y_1$ denote the present value of earnings in the high school and college sectors, respectively. Define ex ante returns to college as the ratio $E_{\theta, \varepsilon}(R) = E_{\theta, \varepsilon}((Y_1 - Y_0)/Y_0)$. Let $f(r)$ denote the density function of the random variable $E_{\theta, \varepsilon}(R)$. The solid line is the density of ex post returns to college for high school graduates, that is $f(r|S=0)$. The dashed line is the density of ex post returns to college for college graduates, that is, $f(r|S=1)$. This assumes that the agent chooses schooling without knowing $\theta, \varepsilon$. 

Figure 7.2
Densities of ex ante returns to college by level of schooling chosen

Let $\varepsilon = (\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...,T)$. Let $Y_0, Y_1$ denote the present value of earnings in the high school and college sectors, respectively. Define ex ante returns to college as the ratio $E_{\theta, \varepsilon}(R) = E_{\theta, \varepsilon}((Y_1 - Y_0)/Y_0)$. Let $f(r)$ denote the density function of the random variable $E_{\theta, \varepsilon}(R)$. The solid line is the density of ex post returns to college for high school graduates, that is $f(r|S=0)$. The dashed line is the density of ex post returns to college for college graduates, that is, $f(r|S=1)$. This assumes that the agent chooses schooling without knowing $\theta, \varepsilon$. 

84
Let $Y_0, Y_1$ denote the present value of earnings in the high school and college sectors, respectively (discounted at a 3% interest rate). Define returns to college as the ratio $R = (Y_1 - Y_0)/Y_0$. Let $f(r)$ denote the density function of the random variable $R$. The solid line is the density of returns to college for high school graduates, that is $f(r|S=0)$. The dashed line is the density of returns to college for college graduates, that is, $f(r|S=1)$. This assumes that the agent chooses schooling with complete knowledge of future earnings.
Let $C$ denote the monetary value of psychic costs. Let $f(c)$ denote the density function of psychic costs in monetary terms. The dashed line shows the density of psychic costs for high school graduates, that is, $f(c|S=0)$. The dotted line shows the density of psychic costs for college graduates, that is, $f(c|S=1)$. The solid line is the unconditional density of the monetary value of psychic costs, $f(c)$.

Figure 8
Densities of monetary value of psychic cost
both overall and by schooling level

Let $C$ denote the monetary value of psychic costs. Let $f(c)$ denote the density function of psychic costs in monetary terms. The dashed line shows the density of psychic costs for high school graduates, that is, $f(c|S=0)$. The dotted line shows the density of psychic costs for college graduates, that is, $f(c|S=1)$. The solid line is the unconditional density of the monetary value of psychic costs, $f(c)$. 
Let $\Theta$ denote the agent's information set. Let $Y_0$ denote the present value of earnings in the high school sector (discounted at a 3% interest rate). Let $f(y_0|\Theta)$ denote the density of the present value of earnings in high school conditioned on the information set $\Theta$. Then:
The solid line plots $f(y_0|\Theta)$ under no information, i.e. $\Theta=\emptyset$.
The dashed line plots $f(y_0|\Theta)$ when only factor 1 is in the information set, i.e. $\Theta=(\theta_1)$.
The dashed-dotted line plots $f(y_0|\Theta)$ when factors 1 and 2 are in the information set, i.e. $\Theta=(\theta_1,\theta_2)$.
The crossed line plots $f(y_0|\Theta)$ when all factors are in the information set, i.e. $\Theta=(\theta_1,\theta_2,\theta_3)$.
The $X$ are put at the mean and are assumed to be known. The $\theta_i$ when known, are set at their mean of zero.
Figure 9.2
Densities of present value of college earnings under different information sets for the agent calculated for the entire population irregardless of schooling choice

Let \( \Theta \) denote the agent's information set. Let \( Y_1 \) denote the present value of earnings in the college sector (discounted at a 3\% interest rate). Let \( f(y_1|\Theta) \) denote the density of the present value of earnings in high school conditioned on the information set \( \Theta \). Then:

- The solid line plots \( f(y_1|\Theta) \) under no information, i.e. \( \Theta=\emptyset \).
- The dashed line plots \( f(y_1|\Theta) \) when only factor 1 is in the information set, i.e. \( \Theta=(\theta_1) \).
- The dashed-dotted line plots \( f(y_1|\Theta) \) when factors 1 and 2 are in the information set, i.e. \( \Theta=(\theta_1,\theta_2) \).
- The crossed line plots \( f(y_1|\Theta) \) when all factors are in the information set, i.e. \( \Theta=(\theta_1,\theta_2,\theta_3) \).

The X are put at the mean and are assumed to be known. The \( \theta \), when known, are set at their mean of zero.
Figure 9.3
Densities of returns college vs high school under different information sets for the agent calculated for the entire population irregardless of schooling choice

Let $\Theta$ denote the agent's information set. Let $Y_0, Y_1$ denote the present value of earnings in the high school and college sectors, respectively (discounted at a 3% interest rate). Let $D=Y_0-Y_1$ be the difference of the present value of earnings in the college and high school sector. $f(d|\Theta)$ denote the density of the difference of present value of earnings conditioned on the information set $\Theta$. Then:

The solid line plots $f(d|\Theta)$ under no information, i.e. $\Theta=\emptyset$.
The dashed line plots $f(d|\Theta)$ when only factor 1 is in the information set, i.e. $\Theta=(\theta_1)$.
The dashed-dotted line plots $f(d|\Theta)$ when factors 1 and 2 are in the information set, i.e. $\Theta=(\theta_1,\theta_2)$.
The crossed line plots $f(d|\Theta)$ when all factors are in the information set, i.e. $\Theta=(\theta_1,\theta_2,\theta_3)$.

The X are put at the mean and are assumed to be known. The $\theta$ when known are set at their mean of zero.
### Table 1

**Estimated Ex Post Returns to Schooling on Schooling and Their Effects on Schooling Choice using OLS and IV To Estimate The Ex Post Returns To Schooling**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (OLS)</th>
<th>Std. Error (OLS)</th>
<th>Coefficient (IV)</th>
<th>Std. Error (IV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>School (High School vs. College)</td>
<td>0.2735</td>
<td>0.0344</td>
<td>0.2573</td>
<td>0.0451</td>
</tr>
<tr>
<td>School*ASVAB</td>
<td>0.0279</td>
<td>0.0063</td>
<td>0.0153</td>
<td>0.0083</td>
</tr>
</tbody>
</table>

**Instrumental Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>School</td>
<td>0.2573</td>
<td>0.0451</td>
</tr>
<tr>
<td>School*ASVAB</td>
<td>0.0153</td>
<td>0.0083</td>
</tr>
</tbody>
</table>

**Schooling Choice Probit Equation Using OLS Results**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_{School} + b_{School*ASVAB}*ASVAB$</td>
<td>12.6244</td>
<td>0.7284</td>
</tr>
<tr>
<td>Marginal Effect</td>
<td>4.8333</td>
<td>0.2654</td>
</tr>
</tbody>
</table>

**Using IV Coefficients**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_{School} + b_{School*ASVAB}*ASVAB$</td>
<td>22.9150</td>
<td>1.3221</td>
</tr>
<tr>
<td>Marginal Effect</td>
<td>8.7731</td>
<td>0.4817</td>
</tr>
</tbody>
</table>

---

1 Includes controls for Mincer experience (age - years of schooling - 6), experience squared, cohort dummies, and ASVAB scores.

2 We use parental education, family income, broken home, number of siblings, distance to college, local tuition, cohort dummies, South at age 14 and urban at age 14 to instrument for schooling. We then interact the instrument value of school with ASVAB scores.

3 We use the predicted return to school to test whether future earnings affect current schooling choices. We include controls for family background, cohort dummies, distance to college, and local tuition.
<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Full Sample</th>
<th>High School Sample</th>
<th>College Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>Mean</td>
<td>Std. Dev</td>
</tr>
<tr>
<td>Asvab AR*</td>
<td>1362</td>
<td>0.72</td>
<td>0.95</td>
</tr>
<tr>
<td>Asvab PC*</td>
<td>1362</td>
<td>0.42</td>
<td>0.80</td>
</tr>
<tr>
<td>Asvab WK*</td>
<td>1362</td>
<td>0.52</td>
<td>0.72</td>
</tr>
<tr>
<td>Asvab MK*</td>
<td>1362</td>
<td>0.62</td>
<td>0.85</td>
</tr>
<tr>
<td>Asvab CS*</td>
<td>1362</td>
<td>0.21</td>
<td>0.85</td>
</tr>
<tr>
<td>Urban at age 14</td>
<td>3695</td>
<td>0.79</td>
<td>0.40</td>
</tr>
<tr>
<td>Parents Divorced</td>
<td>3695</td>
<td>0.15</td>
<td>0.36</td>
</tr>
<tr>
<td>Number of Siblings</td>
<td>3695</td>
<td>2.86</td>
<td>1.96</td>
</tr>
<tr>
<td>Father's Education</td>
<td>3695</td>
<td>4.31</td>
<td>1.94</td>
</tr>
<tr>
<td>Mother's Education</td>
<td>3695</td>
<td>4.21</td>
<td>1.55</td>
</tr>
<tr>
<td>Born between 1906 and 1915</td>
<td>3695</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Born between 1916 and 1925</td>
<td>3695</td>
<td>0.04</td>
<td>0.19</td>
</tr>
<tr>
<td>Born between 1926 and 1935</td>
<td>3695</td>
<td>0.07</td>
<td>0.25</td>
</tr>
<tr>
<td>Born between 1936 and 1945</td>
<td>3695</td>
<td>0.09</td>
<td>0.29</td>
</tr>
<tr>
<td>Born between 1946 and 1955</td>
<td>3695</td>
<td>0.20</td>
<td>0.40</td>
</tr>
<tr>
<td>Born between 1956 and 1965</td>
<td>3695</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>Born between 1966 and 1975</td>
<td>3695</td>
<td>0.04</td>
<td>0.21</td>
</tr>
<tr>
<td>Education</td>
<td>3695</td>
<td>1.47</td>
<td>0.50</td>
</tr>
<tr>
<td>Age in 1980</td>
<td>3695</td>
<td>26.87</td>
<td>12.32</td>
</tr>
<tr>
<td>Grade Completed 1980</td>
<td>1362</td>
<td>12.06</td>
<td>1.66</td>
</tr>
<tr>
<td>Enrolled in 1980</td>
<td>1362</td>
<td>0.57</td>
<td>0.50</td>
</tr>
<tr>
<td>PV of Earnings**</td>
<td>7152</td>
<td>2.38</td>
<td>1.64</td>
</tr>
<tr>
<td>Tuition at age 17</td>
<td>3695</td>
<td>1.80</td>
<td>0.72</td>
</tr>
</tbody>
</table>

*Note:  AR=Arithmetic Reasoning  
PC=Paragraph Composition  
WK= Word Knowledge  
MK=Math Knowledge  
CS=Coding Speed  
**In thousands of Dollars
<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Cost Function</th>
<th>Test System</th>
<th>Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban at age 14</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Parents Divorced</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Number of Siblings</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Father's Education</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Mother's Education</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Born between 1906 and 1915</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Born between 1916 and 1925</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Born between 1926 and 1935</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Born between 1936 and 1945</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Born between 1946 and 1955</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Born between 1956 and 1965</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Born between 1966 and 1975</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Age in 1980</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Grade Completed 1980</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Enrolled in 1980</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Tuition at age 17</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
### Table 2.3
Estimated Coefficients in Schooling Choice Equation

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.2504</td>
<td>0.3587</td>
</tr>
<tr>
<td>Mother's Education</td>
<td>0.2250</td>
<td>0.0274</td>
</tr>
<tr>
<td>Father's Education</td>
<td>0.3386</td>
<td>0.0246</td>
</tr>
<tr>
<td>Parents Divorced</td>
<td>-0.1976</td>
<td>0.0845</td>
</tr>
<tr>
<td>Number of Siblings</td>
<td>-0.1012</td>
<td>0.0163</td>
</tr>
<tr>
<td>Urban Residence at age 14</td>
<td>0.1998</td>
<td>0.0755</td>
</tr>
<tr>
<td>Dummy birth 1916-1925</td>
<td>0.6076</td>
<td>0.3582</td>
</tr>
<tr>
<td>Dummy birth 1926-1935</td>
<td>0.5553</td>
<td>0.3471</td>
</tr>
<tr>
<td>Dummy birth 1936-1945</td>
<td>0.7050</td>
<td>0.3417</td>
</tr>
<tr>
<td>Dummy birth 1946-1955</td>
<td>0.4160</td>
<td>0.3355</td>
</tr>
<tr>
<td>Dummy birth 1956-1965</td>
<td>-0.2064</td>
<td>0.3346</td>
</tr>
<tr>
<td>Dummy birth 1966-1975</td>
<td>-1.4159</td>
<td>0.3703</td>
</tr>
<tr>
<td>Tuition at 4-year college</td>
<td>-0.0953</td>
<td>0.0447</td>
</tr>
<tr>
<td>Loading Factor 1</td>
<td>1.3523</td>
<td>0.1315</td>
</tr>
<tr>
<td>Loading Factor 2</td>
<td>0.4785</td>
<td>0.1335</td>
</tr>
<tr>
<td>Loading Factor 3</td>
<td>-0.0624</td>
<td>0.1274</td>
</tr>
</tbody>
</table>
Table 2.4
Estimated Coefficients for High School Earnings Equation

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Period Zero</th>
<th>Period One</th>
<th>Period Two</th>
<th>Period Three</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Dev</td>
<td>Mean</td>
<td>Std Dev</td>
<td>Mean</td>
</tr>
<tr>
<td>Dummy birth 1916-1925</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dummy birth 1926-1935</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dummy birth 1936-1945</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.1779</td>
</tr>
<tr>
<td>Dummy birth 1946-1955</td>
<td>-</td>
<td>-</td>
<td>-0.7107</td>
<td>0.0637</td>
<td>-0.2936</td>
</tr>
<tr>
<td>Dummy birth 1956-1965</td>
<td>-0.6730</td>
<td>0.0960</td>
<td>-0.2360</td>
<td>0.2267</td>
<td>-</td>
</tr>
<tr>
<td>Constant</td>
<td>2.6276</td>
<td>0.0658</td>
<td>2.4021</td>
<td>0.0935</td>
<td>1.8880</td>
</tr>
<tr>
<td>Loading Factor 1</td>
<td>0.1636</td>
<td>0.0433</td>
<td>0.1059</td>
<td>0.0485</td>
<td>0.0164</td>
</tr>
<tr>
<td>Loading Factor 2</td>
<td>-1.2138</td>
<td>0.0903</td>
<td>-1.6282</td>
<td>0.1142</td>
<td>-1.4415</td>
</tr>
<tr>
<td>Loading Factor 3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2428</td>
</tr>
</tbody>
</table>

Table 2.5
Estimated Coefficients for College Earnings Equation

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Period Zero</th>
<th>Period One</th>
<th>Period Two</th>
<th>Period Three</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Dev</td>
<td>Mean</td>
<td>Std Dev</td>
<td>Mean</td>
</tr>
<tr>
<td>Dummy birth 1916-1925</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dummy birth 1926-1935</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dummy birth 1936-1945</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0059</td>
</tr>
<tr>
<td>Dummy birth 1946-1955</td>
<td>-</td>
<td>-</td>
<td>-0.1944</td>
<td>0.1262</td>
<td>-0.0512</td>
</tr>
<tr>
<td>Dummy birth 1956-1965</td>
<td>-0.7375</td>
<td>0.0686</td>
<td>-0.2340</td>
<td>0.1182</td>
<td>-0.1081</td>
</tr>
<tr>
<td>Dummy birth 1966-1975</td>
<td>-0.3459</td>
<td>0.1736</td>
<td>1.3144</td>
<td>0.7365</td>
<td>-</td>
</tr>
<tr>
<td>Constant</td>
<td>2.2802</td>
<td>0.0670</td>
<td>3.5270</td>
<td>0.1191</td>
<td>3.1859</td>
</tr>
<tr>
<td>Loading Factor 1</td>
<td>0.2225</td>
<td>0.0853</td>
<td>0.3137</td>
<td>0.1296</td>
<td>-0.2870</td>
</tr>
<tr>
<td>Loading Factor 2</td>
<td>1.0000</td>
<td>0.0000</td>
<td>2.3887</td>
<td>0.1573</td>
<td>2.3194</td>
</tr>
<tr>
<td>Loading Factor 3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
### Table 2.5
Estimated Coefficients of Test Equations

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Arithmetic Reasoning</th>
<th></th>
<th>Composition</th>
<th></th>
<th>Word Knowledge</th>
<th></th>
<th>Math Knowledge</th>
<th></th>
<th>Coding Speed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>Mean: -1.1198</td>
<td>Std Dev: 0.2256</td>
<td>Mean: -1.0262</td>
<td>Std Dev: 0.1719</td>
<td>Mean: -0.5180</td>
<td>Std Dev: 0.2032</td>
<td>Mean: -1.4751</td>
<td>Std Dev: 0.2265</td>
<td>Mean: -1.2706</td>
<td>Std Dev: 0.2281</td>
</tr>
<tr>
<td>Mother's Education</td>
<td>Mean: 0.0735</td>
<td>Std Dev: 0.0177</td>
<td>Mean: 0.0529</td>
<td>Std Dev: 0.0136</td>
<td>Mean: 0.0614</td>
<td>Std Dev: 0.0158</td>
<td>Mean: 0.0469</td>
<td>Std Dev: 0.0178</td>
<td>Mean: 0.0561</td>
<td>Std Dev: 0.0175</td>
</tr>
<tr>
<td>Father's Education</td>
<td>Mean: 0.0494</td>
<td>Std Dev: 0.0136</td>
<td>Mean: 0.0593</td>
<td>Std Dev: 0.0105</td>
<td>Mean: 0.0461</td>
<td>Std Dev: 0.0121</td>
<td>Mean: 0.0168</td>
<td>Std Dev: 0.0139</td>
<td>Mean: 0.0870</td>
<td>Std Dev: 0.0135</td>
</tr>
<tr>
<td>Family Income in 1979</td>
<td>Mean: 0.0008</td>
<td>Std Dev: 0.0015</td>
<td>Mean: 0.0009</td>
<td>Std Dev: 0.0012</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0014</td>
<td>Mean: 0.0038</td>
<td>Std Dev: 0.0016</td>
<td>Mean: 0.0021</td>
<td>Std Dev: 0.0015</td>
</tr>
<tr>
<td>Parents Divorced</td>
<td>Mean: -0.0584</td>
<td>Std Dev: 0.0564</td>
<td>Mean: -0.0514</td>
<td>Std Dev: 0.0440</td>
<td>Mean: -0.0947</td>
<td>Std Dev: 0.0508</td>
<td>Mean: 0.0458</td>
<td>Std Dev: 0.0569</td>
<td>Mean: -0.0138</td>
<td>Std Dev: 0.0560</td>
</tr>
<tr>
<td>Number of Siblings</td>
<td>Mean: -0.0193</td>
<td>Std Dev: 0.0111</td>
<td>Mean: -0.0397</td>
<td>Std Dev: 0.0086</td>
<td>Mean: -0.0143</td>
<td>Std Dev: 0.0099</td>
<td>Mean: -0.0273</td>
<td>Std Dev: 0.0115</td>
<td>Mean: -0.0313</td>
<td>Std Dev: 0.0110</td>
</tr>
<tr>
<td>South Residence at age 14</td>
<td>Mean: -0.1278</td>
<td>Std Dev: 0.0463</td>
<td>Mean: -0.0906</td>
<td>Std Dev: 0.0358</td>
<td>Mean: -0.0064</td>
<td>Std Dev: 0.0423</td>
<td>Mean: -0.1418</td>
<td>Std Dev: 0.0475</td>
<td>Mean: -0.1365</td>
<td>Std Dev: 0.0464</td>
</tr>
<tr>
<td>Urban Residence at age 14</td>
<td>Mean: 0.0640</td>
<td>Std Dev: 0.0461</td>
<td>Mean: -0.0243</td>
<td>Std Dev: 0.0361</td>
<td>Mean: 0.0117</td>
<td>Std Dev: 0.0422</td>
<td>Mean: 0.0258</td>
<td>Std Dev: 0.0468</td>
<td>Mean: 0.0529</td>
<td>Std Dev: 0.0466</td>
</tr>
<tr>
<td>Enrolled at School at Test Date</td>
<td>Mean: 0.0646</td>
<td>Std Dev: 0.0528</td>
<td>Mean: -0.0036</td>
<td>Std Dev: 0.0403</td>
<td>Mean: -0.0515</td>
<td>Std Dev: 0.0471</td>
<td>Mean: 0.0074</td>
<td>Std Dev: 0.0527</td>
<td>Mean: 0.3122</td>
<td>Std Dev: 0.0529</td>
</tr>
<tr>
<td>Age at Test Date</td>
<td>Mean: 0.0096</td>
<td>Std Dev: 0.0164</td>
<td>Mean: 0.0237</td>
<td>Std Dev: 0.0128</td>
<td>Mean: -0.0170</td>
<td>Std Dev: 0.0148</td>
<td>Mean: 0.0048</td>
<td>Std Dev: 0.0165</td>
<td>Mean: -0.0510</td>
<td>Std Dev: 0.0166</td>
</tr>
<tr>
<td>Highest Grade Completed at Test Date</td>
<td>Mean: 0.0911</td>
<td>Std Dev: 0.0198</td>
<td>Mean: 0.0604</td>
<td>Std Dev: 0.0155</td>
<td>Mean: 0.0721</td>
<td>Std Dev: 0.0179</td>
<td>Mean: 0.1082</td>
<td>Std Dev: 0.0201</td>
<td>Mean: 0.1732</td>
<td>Std Dev: 0.0198</td>
</tr>
<tr>
<td>Loading Factor 1</td>
<td>Mean: 1.0000</td>
<td>Std Dev: 0.0000</td>
<td>Mean: 0.6801</td>
<td>Std Dev: 0.0321</td>
<td>Mean: 0.8069</td>
<td>Std Dev: 0.0377</td>
<td>Mean: 0.5648</td>
<td>Std Dev: 0.0319</td>
<td>Mean: 0.9562</td>
<td>Std Dev: 0.0293</td>
</tr>
<tr>
<td>Loading Factor 2</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0000</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0000</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0000</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0000</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0000</td>
</tr>
<tr>
<td>Loading Factor 3</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0000</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0000</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0000</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0000</td>
<td>Mean: 0.0000</td>
<td>Std Dev: 0.0000</td>
</tr>
</tbody>
</table>
### Table 3a

**Goodness of Fit Tests: Predicted Earnings Densities vs. Actual Densities**

**The Three-Factor Model**

<table>
<thead>
<tr>
<th></th>
<th>High School</th>
<th>College</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Period 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-Square Stat</td>
<td>91.9681</td>
<td>74.2503</td>
<td>204.3823</td>
</tr>
<tr>
<td>Critical Value*</td>
<td>107.5217</td>
<td>82.5287</td>
<td>178.4854</td>
</tr>
<tr>
<td><strong>Period 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-Square Stat</td>
<td>86.6649</td>
<td>107.6417</td>
<td>207.6152</td>
</tr>
<tr>
<td>Critical Value*</td>
<td>116.5110</td>
<td>116.5110</td>
<td>218.8205</td>
</tr>
<tr>
<td><strong>Period 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-Square Stat</td>
<td>26.2658</td>
<td>45.5301</td>
<td>106.5721</td>
</tr>
<tr>
<td>Critical Value*</td>
<td>43.7730</td>
<td>55.7585</td>
<td>91.6702</td>
</tr>
<tr>
<td><strong>Period 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-Square Stat</td>
<td>35.3846</td>
<td>29.7218</td>
<td>55.5758</td>
</tr>
<tr>
<td>Critical Value*</td>
<td>31.4104</td>
<td>30.1435</td>
<td>55.7585</td>
</tr>
<tr>
<td><strong>Period 5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-Square Stat</td>
<td>23.2193</td>
<td>14.9131</td>
<td>41.8657</td>
</tr>
<tr>
<td>Critical Value*</td>
<td>23.6848</td>
<td>16.9190</td>
<td>35.1725</td>
</tr>
</tbody>
</table>

* 95% Confidence, equiprobable bins with aprox. 15 people per bin
<table>
<thead>
<tr>
<th>Period</th>
<th>$\chi^2$ Statistic</th>
<th>High School</th>
<th>College</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Critical Value*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period 1</td>
<td></td>
<td>109.5702</td>
<td>132.3027</td>
<td>267.4894</td>
</tr>
<tr>
<td></td>
<td></td>
<td>107.5217</td>
<td>82.5287</td>
<td>178.4854</td>
</tr>
<tr>
<td>Period 2</td>
<td></td>
<td>104.1649</td>
<td>150.5556</td>
<td>247.6732</td>
</tr>
<tr>
<td></td>
<td></td>
<td>116.5110</td>
<td>116.5110</td>
<td>218.8205</td>
</tr>
<tr>
<td>Period 3</td>
<td></td>
<td>40.7028</td>
<td>61.7322</td>
<td>114.1692</td>
</tr>
<tr>
<td></td>
<td></td>
<td>43.7730</td>
<td>55.7585</td>
<td>91.6702</td>
</tr>
<tr>
<td>Period 4</td>
<td></td>
<td>39.7253</td>
<td>47.5559</td>
<td>64.2503</td>
</tr>
<tr>
<td></td>
<td></td>
<td>31.4104</td>
<td>30.1435</td>
<td>55.7585</td>
</tr>
<tr>
<td>Period 5</td>
<td></td>
<td>18.3217</td>
<td>26.5855</td>
<td>40.4078</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23.6848</td>
<td>16.9190</td>
<td>35.1725</td>
</tr>
</tbody>
</table>

* 95% Confidence, equiprobable bins with aprox. 15 people per bin
Table 4.1: Ex-Post Conditional Distributions (College Earnings Conditional on High School Earnings)
Pr(d_i < Yc < d_i+1 | d_j < Yh < d_j+1) where d_i is the i-th decile of the College Lifetime Ex-Post Earnings Distribution and d_j is the j-th decile of the High School Ex-Post Lifetime Earnings Distribution

Correlation(Yc, Yh) = -0.3899

<table>
<thead>
<tr>
<th>High School</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0035</td>
<td>0.0109</td>
<td>0.0240</td>
<td>0.0326</td>
<td>0.0524</td>
<td>0.7538</td>
<td>0.1137</td>
<td>0.1557</td>
<td>0.2511</td>
<td>0.2808</td>
</tr>
<tr>
<td>2</td>
<td>0.0098</td>
<td>0.0244</td>
<td>0.0419</td>
<td>0.0631</td>
<td>0.0894</td>
<td>0.1122</td>
<td>0.1391</td>
<td>0.1747</td>
<td>0.2048</td>
<td>0.1407</td>
</tr>
<tr>
<td>3</td>
<td>0.0160</td>
<td>0.0466</td>
<td>0.0741</td>
<td>0.0877</td>
<td>0.1041</td>
<td>0.1213</td>
<td>0.1441</td>
<td>0.1549</td>
<td>0.1581</td>
<td>0.0931</td>
</tr>
<tr>
<td>4</td>
<td>0.0236</td>
<td>0.0603</td>
<td>0.0911</td>
<td>0.1062</td>
<td>0.1220</td>
<td>0.1298</td>
<td>0.1348</td>
<td>0.1372</td>
<td>0.1266</td>
<td>0.0683</td>
</tr>
<tr>
<td>5</td>
<td>0.0439</td>
<td>0.0848</td>
<td>0.1108</td>
<td>0.1227</td>
<td>0.1303</td>
<td>0.1309</td>
<td>0.1211</td>
<td>0.1139</td>
<td>0.0928</td>
<td>0.0489</td>
</tr>
<tr>
<td>6</td>
<td>0.0627</td>
<td>0.1074</td>
<td>0.1214</td>
<td>0.1304</td>
<td>0.1330</td>
<td>0.1218</td>
<td>0.1168</td>
<td>0.0954</td>
<td>0.0695</td>
<td>0.0415</td>
</tr>
<tr>
<td>7</td>
<td>0.0963</td>
<td>0.1256</td>
<td>0.1340</td>
<td>0.1334</td>
<td>0.1200</td>
<td>0.1200</td>
<td>0.0937</td>
<td>0.0784</td>
<td>0.0554</td>
<td>0.0433</td>
</tr>
<tr>
<td>8</td>
<td>0.1378</td>
<td>0.1659</td>
<td>0.1529</td>
<td>0.1396</td>
<td>0.1114</td>
<td>0.0925</td>
<td>0.0740</td>
<td>0.0561</td>
<td>0.0296</td>
<td>0.0402</td>
</tr>
<tr>
<td>9</td>
<td>0.1939</td>
<td>0.1970</td>
<td>0.1498</td>
<td>0.1180</td>
<td>0.1002</td>
<td>0.0771</td>
<td>0.0534</td>
<td>0.0362</td>
<td>0.0200</td>
<td>0.0543</td>
</tr>
<tr>
<td>10</td>
<td>0.3354</td>
<td>0.1983</td>
<td>0.1167</td>
<td>0.0812</td>
<td>0.0515</td>
<td>0.0351</td>
<td>0.0266</td>
<td>0.0152</td>
<td>0.0130</td>
<td>0.1271</td>
</tr>
</tbody>
</table>
Table 4.2: Ex-Ante Conditional Distribution (College Earnings Conditional on High School Earnings)

\[ \Pr(d_i < Yc < d_i + 1 \mid d_j < Yh < d_j + 1) \] where \( d_i \) is the \( i \)th decile of the College Lifetime Ex-Ante Earnings Distribution and \( d_j \) is the \( j \)th decile of the High School Ex-Ante Lifetime Earnings Distribution

Individual expects out \( \theta_3 \) and \( \varepsilon_{s,t} \) for \( t=0, \ldots, 4 \), which are unknown by the agent at the time of the schooling choice.

Correlation\( (Y_C, Y_H) = -0.69936591 \)

<table>
<thead>
<tr>
<th>High School</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0002</td>
<td>0.0079</td>
<td>0.0108</td>
<td>0.0226</td>
<td>0.0421</td>
<td>0.0594</td>
<td>0.0909</td>
<td>0.1447</td>
<td>0.2236</td>
<td>0.3978</td>
</tr>
<tr>
<td>2</td>
<td>0.0044</td>
<td>0.0180</td>
<td>0.0286</td>
<td>0.0530</td>
<td>0.0720</td>
<td>0.1010</td>
<td>0.1362</td>
<td>0.1686</td>
<td>0.2114</td>
<td>0.2068</td>
</tr>
<tr>
<td>3</td>
<td>0.0106</td>
<td>0.0362</td>
<td>0.0578</td>
<td>0.0786</td>
<td>0.1062</td>
<td>0.1152</td>
<td>0.1498</td>
<td>0.1618</td>
<td>0.1692</td>
<td>0.1146</td>
</tr>
<tr>
<td>4</td>
<td>0.0200</td>
<td>0.0546</td>
<td>0.0786</td>
<td>0.1024</td>
<td>0.1204</td>
<td>0.1266</td>
<td>0.1376</td>
<td>0.1406</td>
<td>0.1290</td>
<td>0.0902</td>
</tr>
<tr>
<td>5</td>
<td>0.0390</td>
<td>0.0740</td>
<td>0.1004</td>
<td>0.1130</td>
<td>0.1291</td>
<td>0.1387</td>
<td>0.1295</td>
<td>0.1206</td>
<td>0.1010</td>
<td>0.0546</td>
</tr>
<tr>
<td>6</td>
<td>0.0454</td>
<td>0.1017</td>
<td>0.1253</td>
<td>0.1353</td>
<td>0.1333</td>
<td>0.1323</td>
<td>0.1189</td>
<td>0.1011</td>
<td>0.0754</td>
<td>0.0314</td>
</tr>
<tr>
<td>7</td>
<td>0.0873</td>
<td>0.1299</td>
<td>0.1437</td>
<td>0.1451</td>
<td>0.1299</td>
<td>0.1199</td>
<td>0.0965</td>
<td>0.0777</td>
<td>0.0519</td>
<td>0.0180</td>
</tr>
<tr>
<td>8</td>
<td>0.1336</td>
<td>0.1603</td>
<td>0.1613</td>
<td>0.1431</td>
<td>0.1160</td>
<td>0.0974</td>
<td>0.0793</td>
<td>0.0589</td>
<td>0.0389</td>
<td>0.0112</td>
</tr>
<tr>
<td>9</td>
<td>0.2063</td>
<td>0.2016</td>
<td>0.1651</td>
<td>0.1293</td>
<td>0.1056</td>
<td>0.0840</td>
<td>0.0540</td>
<td>0.0317</td>
<td>0.0155</td>
<td>0.0068</td>
</tr>
<tr>
<td>10</td>
<td>0.4123</td>
<td>0.2318</td>
<td>0.1393</td>
<td>0.0868</td>
<td>0.0556</td>
<td>0.0365</td>
<td>0.0210</td>
<td>0.0110</td>
<td>0.0049</td>
<td>0.0006</td>
</tr>
</tbody>
</table>
Table 5.1
Average present value of earnings\(^1\) for high school graduates
Fitted and Counterfactual\(^2\)
White males from NLSY79

<table>
<thead>
<tr>
<th></th>
<th>High School (Fitted)</th>
<th>College (counterfactual)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Present Value of Earnings</td>
<td>605.92</td>
<td>969.34</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>13.719</td>
<td>67.164</td>
</tr>
<tr>
<td>Random(^3)</td>
<td>573.53</td>
<td>987.21</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>15.799</td>
<td>47.132</td>
</tr>
</tbody>
</table>

Average returns\(^4\) to college for high school graduates

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average returns</td>
<td>1.17</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>0.1350</td>
</tr>
</tbody>
</table>

\(^1\) Thousands of dollars. Discounted using a 3% interest rate.
\(^2\) The counterfactual is constructed using the estimated college outcome equation applied to the population of persons selecting high school.

\(^3\) It defines the result of taking a person at random from the population regardless of his schooling choice.

\(^4\) As a fraction of the base state, i.e. \(\frac{PV\text{earnings(Col)}-PV\text{earnings(HS)}}{PV\text{earnings(HS)}}\).
### Table 5.2
Average present value of earnings\(^1\) for college graduates

Fitted and Counterfactual\(^2\)

<table>
<thead>
<tr>
<th></th>
<th>High School (Counterfactual)</th>
<th>College (fitted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Present Value of Earnings</td>
<td>536.43</td>
<td>1007.64</td>
</tr>
<tr>
<td>(Std. \ Err.)</td>
<td>26.187</td>
<td>35.113</td>
</tr>
<tr>
<td>Random(^3)</td>
<td>573.53</td>
<td>987.21</td>
</tr>
<tr>
<td>(Std. \ Err.)</td>
<td>15.799</td>
<td>47.132</td>
</tr>
</tbody>
</table>

Average returns\(^4\) to college for college graduates

| Average returns     | 1.33 |
| \(Std. \ Err.\)    | 0.0958 |

\(^1\) Thousands of dollars. Discounted using a 3% interest rate.

\(^2\) The counterfactual is constructed using the estimated high school outcome equation applied to the population of persons selecting college.

\(^3\) It defines the result of taking a person at random from the population regardless of his schooling choice.

\(^4\) As a fraction of the base state, ie \((\text{PVearnings}(\text{Col}) - \text{PVearnings}(\text{HS}))/\text{PVearnings}(\text{HS})\).
Table 5.3

Average present value of earnings\(^1\) for population of persons indifferent between high school and college

Conditional on education level

White males from NLSY79

<table>
<thead>
<tr>
<th></th>
<th>High School</th>
<th>College</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Present Value of Earnings</td>
<td>571.33</td>
<td>975.16</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>37.066</td>
<td>70.557</td>
</tr>
</tbody>
</table>

Average returns\(^3\) to college for people indifferent between high school and college

High School vs Some College

<table>
<thead>
<tr>
<th>Average returns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.26</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>0.3691</td>
</tr>
</tbody>
</table>

\(^1\) Thousands of dollars. Discounted using a 3% interest rate.

\(^2\) It defines the result of taking a person at random from the population regardless of his schooling choice.

\(^3\) As a fraction of the base state, ie (PVearnings(Col)-PVearnings(HS))/PVearnings(HS).
Table 5.4
Average ex-post, ex-ante and perfect certainty returns
White males from NLSY79

<table>
<thead>
<tr>
<th></th>
<th>For People who choose High School</th>
<th>For People who choose College</th>
<th>For People Indifferent Between High School and College</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ex-post(^2)</td>
<td>ex-ante(^3)</td>
<td>perfect certainty(^4)</td>
</tr>
<tr>
<td>Average</td>
<td>1.1594</td>
<td>1.1594</td>
<td>0.9337</td>
</tr>
<tr>
<td>Std. Err.</td>
<td>0.1362</td>
<td>0.1362</td>
<td>0.1154</td>
</tr>
</tbody>
</table>

1 Let \( Y_0, Y_1 \) denote the present value of earnings in high school and college, respectively. The return to college \( R \) is defined as

\[
R = \left( \frac{Y_1 - Y_0}{Y_0} \right)
\]

2 Let \( I \) denote the schooling choice index. Let \( \Theta_0 \) denote the information set of the agent at the time of the schooling choice. Let \( R \) denote the return to college. The ex-post mean return to college for a high-school graduate is \( E (R \mid E_0(I) < 0) \), where \( E_0(I) = E(I \mid \Theta_0) \). Similarly, the ex-post mean return to college for a college graduate is \( E (R \mid E_0(I) \geq 0) \).

3 Let \( I \) denote the schooling index. Let \( \Theta_0 \) denote the information set of the agent at the time of the schooling choice. Let \( R \) denote the return to college. The ex-ante mean return to college for a high-school graduate is \( E (E_0(R) \mid E_0(I) < 0) \). Similarly, the ex-ante mean return to college for a college graduate is \( E (E_0(R) \mid E_0(I) \geq 0) \). The ex-ante mean return to an agent just indifferent between college and high-school is \( E (R \mid E_0(I) = 0) \).

4 Let \( I \) denote the schooling index. Let \( R \) denote the return to college. The return to college under perfect certainty for a high-school graduate is \( E (R \mid I < 0) \). Note that now the agent makes his schooling choice under perfect certainty (that is why we condition on \( I \)). Similarly, the return to college under perfect certainty for a college graduate is \( E (R \mid I \geq 0) \). The return to college under perfect certainty for an agent just indifferent between college and high-school is \( E (R \mid I = 0) \).
Table 6.1
Agent’s Forecast Variance of Present Value of Earnings* Under Different Information Sets
(fraction of the variance explained by $\Theta$)**
The Calculation is for the Entire Population Irregardless of Schooling Choice.

<table>
<thead>
<tr>
<th>Variance when $\Theta$</th>
<th>$\text{Var}(Y_c)$</th>
<th>$\text{Var}(Y_h)$</th>
<th>$\text{Var}(Y_c-Y_h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>156402.14</td>
<td>73827.89</td>
<td>267796.38</td>
</tr>
<tr>
<td>$\Theta = {\theta_1}$</td>
<td>0.95%</td>
<td>0.27%</td>
<td>0.44%</td>
</tr>
<tr>
<td>$\Theta = {\theta_1, \theta_2}$</td>
<td>29.10%</td>
<td>29.43%</td>
<td>47.42%</td>
</tr>
<tr>
<td>$\Theta = {\theta_1, \theta_2, \theta_3}$</td>
<td>68.03%</td>
<td>32.27%</td>
<td>62.65%</td>
</tr>
</tbody>
</table>

*We use an interest rate of 3% to calculate the present value of earnings.
+Variance of the unpredictable component of earnings between age 19 and 65 as predicted at age 19.
**The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is $(1-0.0095)*156402.14$
Table 6.2
Agent’s Forecast Variance of Period Zero Earnings*
Under Different Information Sets
(fraction of the variance explained by $\Theta$)**
The Calculation is for the Entire Population Irregardless of Schooling Choice.

<table>
<thead>
<tr>
<th></th>
<th>$\text{Var}(Y_c)$</th>
<th>$\text{Var}(Y_h)$</th>
<th>$\text{Var}(Y_c-Y_h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>For lifetime: +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance when $\Theta = \emptyset$</td>
<td>13086.24</td>
<td>14303.35</td>
<td>33910.17</td>
</tr>
<tr>
<td>$\Theta = {\theta_1}$</td>
<td>1.90%</td>
<td>0.91%</td>
<td>0.05%</td>
</tr>
<tr>
<td>$\Theta = {\theta_1, \theta_2}$</td>
<td>23.58%</td>
<td>30.08%</td>
<td>41.02%</td>
</tr>
</tbody>
</table>

* We use an interest rate of 3% to calculate the present value of earnings.

+ Variance of the unpredictable component of earnings between age 19 and 28 as predicted at age 19.

**The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is $(1-0.0190) \times 13086.24$
Table 6.3
Agent’s Forecast Variance of Period One Earnings*
Under Different Information Sets
(fraction of the variance explained by Θ)**
The Calculation is for the Entire Population Irregardless of Schooling Choice.

<table>
<thead>
<tr>
<th></th>
<th>Var((Y_c))</th>
<th>Var ((Y_h))</th>
<th>Var((Y_c-Y_h))</th>
</tr>
</thead>
<tbody>
<tr>
<td>For lifetime:†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance when Θ = (\emptyset)</td>
<td>26618.64</td>
<td>17545.90</td>
<td>65804.89</td>
</tr>
<tr>
<td>Θ = {θ₁}</td>
<td>1.90%</td>
<td>0.31%</td>
<td>0.34%</td>
</tr>
<tr>
<td>Θ = {θ₁, θ₂}</td>
<td>62.43%</td>
<td>43.00%</td>
<td>69.60%</td>
</tr>
</tbody>
</table>

*We use an interest rate of 3% to calculate the present value of earnings.
†Variance of the unpredictable component of earnings between age 29 and 38 as predicted at age 19.
**So we would say that the variance of the unpredictable component of period 1 college earnings \(Θ = \{θ₁\}\) is \((1-0.0190)\times26618.64\)
Table 6.4
Agent’s Forecast Variance of Period Two Earnings*
Under Different Information Sets
(fraction of the variance explained by $\Theta$)**
The Calculation is for the Entire Population Irregardless of Schooling Choice.

<table>
<thead>
<tr>
<th>Information Set</th>
<th>$\text{Var}(Y_c)$</th>
<th>$\text{Var}(Y_h)$</th>
<th>$\text{Var}(Y_c-Y_h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Theta = \emptyset$</td>
<td>40406.20</td>
<td>16716.50</td>
<td>68918.36</td>
</tr>
<tr>
<td>$\Theta = {\theta_1}$</td>
<td>0.95%</td>
<td>0.00%</td>
<td>0.63%</td>
</tr>
<tr>
<td>$\Theta = {\theta_1, \theta_2}$</td>
<td>38.66%</td>
<td>35.02%</td>
<td>58.63%</td>
</tr>
<tr>
<td>$\Theta = {\theta_1, \theta_2, \theta_3}$</td>
<td>75.25%</td>
<td>40.17%</td>
<td>70.98%</td>
</tr>
</tbody>
</table>

*We use an interest rate of 3% to calculate the present value of earnings.

†Variance of the unpredictable component of earnings between age 39 and 48 as predicted at age 19.

**The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is $(1-0.0095)\times40406.20$
Table 6.5
Agent’s Forecast Variance of Period Three Earnings* Under Different Information Sets
(fraction of the variance explained by Θ)**
The Calculation is for the Entire Population Irregardless of Schooling Choice.

<table>
<thead>
<tr>
<th></th>
<th>Var(Yc)</th>
<th>Var(Yh)</th>
<th>Var(Yc-Yh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>For lifetime:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance when Θ = ∅</td>
<td>53194.23</td>
<td>14605.29</td>
<td>66926.12</td>
</tr>
<tr>
<td>Θ = {θ₁}</td>
<td>0.65%</td>
<td>0.08%</td>
<td>0.73%</td>
</tr>
<tr>
<td>Θ = {θ₁, θ₂}</td>
<td>16.18%</td>
<td>24.55%</td>
<td>34.65%</td>
</tr>
<tr>
<td>Θ = {θ₁, θ₂, θ₃}</td>
<td>81.20%</td>
<td>31.53%</td>
<td>70.11%</td>
</tr>
</tbody>
</table>

*We use an interest rate of 3% to calculate the present value of earnings.
+Variance of the unpredictable component of earnings between age 49 and 58 as predicted at age 19.
**The variance of the unpredictable component of period 1 college earnings Θ = {θ₁} is (1-0.0065)*53194.23
Table 6.6
Agent’s Forecast Variance of Period Four of Earnings*
Under Different Information Sets
(fraction of the variance explained by $\Theta$)**
The Calculation is for the Entire Population Irregardless of Schooling Choice.

For lifetime: +

<table>
<thead>
<tr>
<th>$\Theta$</th>
<th>$\text{Var}(Y_c)$</th>
<th>$\text{Var}(Y_h)$</th>
<th>$\text{Var}(Y_c - Y_h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varnothing$</td>
<td>23096.81</td>
<td>10656.83</td>
<td>32236.82</td>
</tr>
<tr>
<td>${\theta_1}$</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>${\theta_1, \theta_2}$</td>
<td>6.84%</td>
<td>4.10%</td>
<td>11.41%</td>
</tr>
<tr>
<td>${\theta_1, \theta_2, \theta_3}$</td>
<td>56.70%</td>
<td>6.16%</td>
<td>37.95%</td>
</tr>
</tbody>
</table>

*We use an interest rate of 3% to calculate the present value of earnings.
+ Variance of the unpredictable component of earnings between age 59 and 65 as predicted at age 19.
**The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is $(1-0.00) \times 23096.81$