The Evolution of Earnings Risk in the US Economy

Flavio Cunha
University of Chicago

James Heckman
University of Chicago

August 4, 2005
This draft, August 7, 2005
1 The Model

Two periods \((T = 0, 1)\)
For \(t = 0, 1\), assume that \((Y_{0,t}, Y_{1,t})\)

\[
Y_{0,t} = X \beta_{0,t} + U_{0,t}, \quad (1a) \\
Y_{1,t} = X \beta_{1,t} + U_{1,t}, \quad (1b) \\
E(U_{0,t} \mid X) = E(U_{1,t} \mid X) = 0, E(Y_{0,t} \mid X) = X \beta_{0,t} (1c) \\
E(Y_{1,t} \mid X) = X \beta_{1,t} \quad (1d)
\]

If the interest rate is \(r\), \textit{ex post} gain for an individual who moves from \(S = 0\) to \(S = 1\) is \(\Delta = Y_{1,0} - Y_{0,0} + \frac{Y_{1,1} - Y_{0,1}}{1+r}\).
Under perfect certainty index $I$ is a net utility,

$$I = Y_{1,0} - Y_{0,0} + \frac{Y_{1,1} - Y_{0,1}}{1 + r} - C, \quad (2)$$

$C$ is the cost of participation in sector 1

$C = Z\gamma + U_C$

$Z$ are determinants of cost.

$$I = \mu_I(X, Z) + U_I. \quad (3)$$

Under perfect certainty,

$$\mu_I(X, Z) = X\beta_{1,0} - X\beta_{0,0} + \frac{X\beta_{1,1} - X\beta_{0,1}}{1 + r} - Z\gamma \quad (4)$$

$$U_I = U_{1,0} - U_{0,0} + \frac{U_{1,1} - U_{0,1}}{1 + r} - U_C. \quad (5)$$
Define $U_I$ as the error in the choice equation and similarly, $\mu_I(X, Z)$ may only be based on expectations of future $X$ and $Z$. 
1.1 Factor Models

A set of test scores:

\[ M_1 = X \beta_1^M + U_1^M \]
\[ M_2 = X \beta_2^M + U_2^M \]
\[ M_3 = X \beta_3^M + U_3^M \]
Decompose the error terms $U_k^M$ for $k = 1, 2, 3$ as:

\[
\begin{align*}
U_1^M &= \alpha_1^M \theta_1 + \varepsilon_1^M \\
U_2^M &= \alpha_2^M \theta_1 + \varepsilon_2^M \\
U_3^M &= \alpha_3^M \theta_1 + \varepsilon_3^M
\end{align*}
\]  

(7)

Assume $\theta_1$ is statistically independent of $X, Z$ and $(\varepsilon_0, \varepsilon_1, \varepsilon_I)$; $\theta_1 \sim N \left(0, \sigma_{\theta_1}^2\right)$. The $\varepsilon$'s are normally distributed and mutually independent with $E \left( \varepsilon_k^M \right) = 0$, $Var \left( \varepsilon_k^M \right) = \sigma_{\varepsilon_k}^2$. 

Can compute the covariances:

\[
\operatorname{Cov} \left( M_1 - X \beta_1^M, M_2 - X \beta_2^M \right) = \alpha_1^M \alpha_2^M \sigma_{\theta_1}^2 \tag{8}
\]

\[
\operatorname{Cov} \left( M_1 - X \beta_1^M, M_3 - X \beta_3^M \right) = \alpha_1^M \alpha_3^M \sigma_{\theta_1}^2 \tag{9}
\]

\[
\operatorname{Cov} \left( M_2 - X \beta_2^M, M_3 - X \beta_3^M \right) = \alpha_2^M \alpha_3^M \sigma_{\theta_1}^2 \tag{10}
\]

Normalize \( \alpha_1^M = 1 \).

Ratio of (10) to (8) we get:

\[
\frac{\operatorname{Cov} \left( M_2 - X \beta_2^M, M_3 - X \beta_3^M \right)}{\operatorname{Cov} \left( M_1 - X \beta_1^M, M_2 - X \beta_2^M \right)} = \alpha_3^M
\]

Ratio of (10) to (9) allows us to recover \( \alpha_{21} \) since

\[
\frac{\operatorname{Cov} \left( M_2 - X \beta_2^M, M_3 - X \beta_3^M \right)}{\operatorname{Cov} \left( M_1 - X \beta_2^M, M_3 - X \beta_3^M \right)} = \alpha_2^M.
\]
Given the normalization of $\alpha_1^M = 1$, and identification of $\alpha_{21}$, we can identify $\sigma_{\theta_1}^2$ from:

$$
\sigma_{\theta_1}^2 = \frac{Cov \left( M_1 - X \beta_1^M, M_2 - X \beta_2^M \right)}{\alpha_2^M}
$$

$$
Var \left( M_k - X \beta_k^M \right) = \left( \alpha_k^M \right)^2 \sigma_{\theta_1}^2 + \sigma_{\varepsilon_k}^2 \text{ for } k = 1, 2, 3
$$

Therefore, we can identify the variances of uniquenesses $\varepsilon_k$ for $k = 1, 2, 3$. 
1.2 Using factor models to identify counterfactual distributions

Factor models cannot be applied directly to (1a) and (1b) because of the selection problem.

We only observe $Y_{0,t}, t = 0, 1$, choose $S = 0$ (and therefore have $I < 0$)

Only observe $Y_{1,t}, t = 0, 1$ for the agents that choose $S = 1$ (and have $I \geq 0$).

The selection problem prevents us from obtaining $\beta_{0,t}$ and $\beta_{1,t}$, $t = 0, 1$, from a simple nonparametric regression on $Y_s$ against $X$ for $s = 0, 1$.

First correct for selection in order to identify the parameters $\beta_{0,t}$ and $\beta_{1,t}, t = 0, 1$. 
Assume that all of the dependence across \( \{U_{0,t}, U_{1,t}\}_{t=0}^{1}, U_C \) is generated by two factors \( \theta_1 \) and \( \theta_2 \):

\[
U_{0,t} = \alpha_{0,t} \theta_1 + \delta_{0,t} \theta_2 + \varepsilon_{0,t},
\]

\[
U_{1,t} = \alpha_{1,t} \theta_1 + \delta_{1,t} \theta_2 + \varepsilon_{1,t},
\]

\[
U_C = \alpha_C \theta_1 + \delta_C \theta_2 + \varepsilon_C.
\]

(11)

Assume \( \theta_2 \) is statistically independent of \( X, Z, \{\varepsilon_{0,t}, \varepsilon_{1,t}\}_{t=0}^{1}, \varepsilon_I \) and that \( \theta_2 \sim N \left( 0, \sigma_{\theta_2}^2 \right) \).

The uniquenesses \( \varepsilon_{0,t}, \varepsilon_{1,t}, t = 0, 1 \), and \( \varepsilon_{I*} \) are normally distributed To show how one can recover the joint distribution of \( \{(Y_{0,t}, Y_{1,\tau})\}_{t,\tau=1}^{2} \) using factor models, break the argument into two parts.
1.3 Identification of the model under normality

Assumptions of normality of the factors $\theta_1$, $\theta_2$, and uniquenesses $\varepsilon_{0,t}$, $\varepsilon_{1,t}$ and $\varepsilon_{I^*}$ together with the linear representation in (11) imply that $\{U_{0,t}, U_{1,t}\}_{t=0}^1$ and $U_{I^*}$ are normally distributed random variables.

\[
E(Y_{0,t}|X, I < 0) = X\beta_{0,t} + E(U_{0,t}|X, I < 0) \tag{12a}
\]

\[
E(Y_{1,t}|X, I \geq 0) = X\beta_{1,t} + E(U_{1,t}|X, I \geq 0) \tag{12b}
\]
Note that:

\[ E(U_{0,t} \mid X, I < 0) = E(U_{0,t} \mid I < 0) = E(U_{0,t} \mid U_I < -\mu_I(X, Z)) \]

Note that:

\[
E(U_{0,t} \mid U_I < -\mu_I(X, Z)) = \sigma_I E \left( \frac{U_{0,t}}{\sigma_I} \middle| \frac{U_I}{\sigma_I} < -\frac{\mu_I(X, Z)}{\sigma_I} \right)
\]

\[
= \frac{\text{Cov}(U_{0,t}, U_I)}{\sigma_I} \lambda \left( \frac{\mu_I(X, Z)}{\sigma_I} \right)
\]

where:

\[
\lambda \left( \frac{\mu_I(X, Z)}{\sigma_I} \right) = E \left( \frac{U_I}{\sigma_I} \middle| \frac{U_I}{\sigma_I} < -\frac{\mu_I(X, Z)}{\sigma_I} \right)
\]

\(\mu_I(X, Z)\) is as defined in equation (4), and \(U_I\) defined in equation (5).
\[ Y_{0,t} = \mathbf{X}\beta_{0,t} + \frac{\text{Cov}(U_{0,t}, U_I)}{\sigma_I} \lambda \left( -\frac{\mu_I(X, Z)}{\sigma_I} \right) + V_{0,t} \]

then identify the mean function \( \beta_{0,t} \).

Can construct the residual:

\[ U_{0,t} = Y_{0,t} - \mathbf{X}\beta_{0,t} \]

Proceed similarly for those who choose college and obtain \( U_{1,t} \) corrected for selection. From discrete choice models, recover \( \mu_I(X, Z) \) up to scale.
Compute the covariances (recall that we normalized $\alpha_1^M = 1$):

\[
Cov \left( M_1 - \mathbf{X} \beta_1^M, Y_{0,t} - \mathbf{X} \beta_{0,t} \right) = \alpha_{0,t} \sigma_{\theta_1}^2 
\]

\[
Cov \left( M_1 - \mathbf{X} \beta_1^M, Y_{1,t} - \mathbf{X} \beta_{1,t} \right) = \alpha_{1,t} \sigma_{\theta_1}^2 
\]

\[
Cov \left( M_1 - \mathbf{X} \beta_1^M, I - \mu_I (\mathbf{X}, \mathbf{Z}) \right) = \alpha_I \sigma_{\theta_1}^2 
\]

The left-hand side of equations (13) and (14) are known.

From the measurement equations, we have already identified $\sigma_{\theta_1}^2$.

Under these conditions, we can identify $\alpha_{1,t}$, $\alpha_{0,t}$, and $\alpha_I$ from the known covariances above.
Next, the following covariances:

\[ \text{Cov} \left( Y_{0,1} - X \beta_{0,1}, I - \mu_I(X, Z) \right) - \alpha_{0,1} \alpha_I \sigma_{\theta_1}^2 = \delta_{0,1} \delta_I \sigma_{\theta_2}^2 \]  
(15)

\[ \text{Cov} \left( Y_{0,2} - X \beta_{0,2}, I - \mu_I(X, Z) \right) - \alpha_{0,2} \alpha_I \sigma_{\theta_1}^2 = \delta_{0,2} \delta_I \sigma_{\theta_2}^2 \]  
(16)

\[ \text{Cov} \left( Y_{0,1} - X \beta_{0,1}, Y_{0,2} - X \beta_{0,2} \right) - \alpha_{0,1} \alpha_{0,2} \sigma_{\theta_1}^2 = \delta_{0,1} \delta_{0,2} \sigma_{\theta_2}^2 \text{ for } t = 1, 2 \]  
(17)

Note that we have introduced a second factor. Impose \( \delta_{0,1} = 1 \).

From the ratio of (16) to (17) we get:

\[ \frac{\text{Cov} \left( Y_{0,2} - X \beta_{0,2}, I - \mu_I(X, Z) \right) - \alpha_{0,2} \alpha_I \sigma_{\theta_1}^2}{\text{Cov} \left( Y_{0,1} - X \beta_{0,1}, Y_{0,2} - X \beta_{0,2} \right) - \alpha_{0,1} \alpha_{0,2} \sigma_{\theta_1}^2} = \delta_I \]

Can identify \( \delta_{0,2} \). Using (15) we can solve for \( \sigma_{\theta_2}^2 \).
Can recover the variance of the uniquenesses $\varepsilon_{s,t}$ for $s = 0, 1$, $t = 1, 2$ from:

$$Var \left( Y_{s,t} - \mathbf{X} \beta_{s,t} \right) = \alpha_{s,t}^2 \sigma_{\theta_1}^2 + \delta_{s,t}^2 \sigma_{\theta_2}^2 + \sigma_{\varepsilon_{s,t}}^2$$

where the only unknown term is $\sigma_{\varepsilon_{s,t}}^2$.

We observe either $\{Y_{0,t}\}$ or $\{Y_{1,t}\}$, never obtain from the data the covariance:

$$Cov \left[ Y_{0,t} - \mathbf{X} \beta_{0,t}, Y_{1,\tau} - \mathbf{X} \beta_{1,\tau} \right] = \alpha_{0,t} \alpha_{1,\tau} \sigma_{\theta_1}^2 + \delta_{0,t} \delta_{1,\tau} \sigma_{\theta_2}^2$$

(18)
because factor structure we know we can compute the right hand side of (18) for $t, \tau = 1, 2$.

Can identify $F \left( Y_{0,t}, Y_{1,\tau} \right)$ for $t, \tau = 1, 2$. 
2 Distinguishing between Heterogeneity and Uncertainty

In the literature on earnings dynamics (e.g., Lillard and Willis, 1978), it is common to estimate an earnings equation of the sort

\[ y_t = \mathbf{X} \beta + \delta S + v_t, \]  

(19)

where \( y_t, \mathbf{X}, S, v_t \) denote, respectively, earnings, observable characteristics.

For example,

\[ v_t = \phi + \varepsilon_t. \]  

(20)
\( \phi \) is a person-specific fixed effect. The error term \( \varepsilon_t \) is generally assumed to be serially correlated, say \( \varepsilon_t = \rho \varepsilon_{t-1} + \eta_t \) where \( \eta_t \) is an independently and identically distributed innovation with mean zero.
Let $\mathcal{I}$ denote the information set of the agent at the time the schooling choice.

The decision rule governing sectoral choice is, in the population,

$$S = 1 \text{ if } E \left( \sum_{t=0}^{1} \frac{Y_{1,t} - Y_{0,t}}{(1 + r)^t} - C \mid \mathcal{I} \right) \geq 0; \ S = 0, \text{ otherwise.}$$

$$I = E \left( \sum_{t=0}^{1} \frac{X(\beta_{1,t} - \beta_{0,t}) + \theta_1(\alpha_{1,t} - \alpha_{0,t}) + \theta_2(\delta_{1,t} - \delta_{0,t}) + \varepsilon_{1,t} - \varepsilon_{0,t}}{(1+r)^t} \mid \mathcal{I} \right) - (Z\gamma + \theta_1\alpha_C + \theta_2\delta_C + \varepsilon_C)$$

$$S = 1 \text{ if } I_i \geq 0; \ S_i = 0 \text{ otherwise.}$$
• For the sake of simplicity, assume that $X$, $Z$ and $\theta_1$ are in the information set of the agent, and interest rate is zero ($r = 0$).

• Assume it is also known that the error terms $\varepsilon_{s,t}$ for $s, t = 0, 1$ are not in the information set of the agent.

We want to test whether $\theta_2$ is also in the information set of the agent at the time of the schooling choice. If $\theta_2 \in \mathcal{I}$, then it reflects heterogeneity across agents (since agents know and act on it). If $\theta_2 \notin \mathcal{I}$, then it reflects either uncertainty or some information that agents know, but don’t act on it.
Suppose that $\theta_1 \in \mathcal{I}$, but $\theta_2 \notin \mathcal{I}$. By definition, the index is

$$I = E(Y_{1,1} + Y_{1,2} - Y_{0,1} - Y_{0,2} - C|\mathcal{I}) = \mu_I(X, Z) + \alpha_I \theta_1.$$ 

Under the null, $\theta_2$ is not in the information set of the agent. Consequently, we have that:

$$Cov (I - \mu_I(X, Z), Y_{1,1} - X \beta_{1,t}) = \alpha_I \alpha_{1,1} \sigma_{\theta_1}^2$$

Let $\Delta_{\theta_2}$ be such that:

$$Cov (I - \mu_I(X, Z), Y_{1,1} - \mu_1(X)) - \alpha_I \alpha_{1,1} \sigma_{\theta_1}^2 - \Delta_{\theta_2} \alpha_I \delta_{1,1} \sigma_{\theta_1}^2 = 0$$

Then, we reject the null, and conclude that agents know and act on the information contained in factor 2, $\theta_2$, if we reject that $\Delta_{\theta_2} = 0$. 
Idea of our test is thus very simple: the components of future earnings that are forecastable are captured by the factors that are known by the agents when they make their educational choices. The predictable factors are estimated with a nonzero loading in the choice equation.