Instrumental Variables Methods For The Correlated Random Coefficient Model: Estimating The Average Rate of Return to Schooling When the Return is Correlated With Schooling

James Heckman
University of Chicago

Edward Vytlacil
Stanford University

Fall, 1998
I. Introduction

\[ \ln y_i = \alpha_{0i} + \alpha_{1i} S_i \]

“rate of return” \( \alpha_{1i} \)

\[ \alpha_{0i} = \bar{\alpha}_0 + \varepsilon_{0i} \quad \text{and} \quad \bar{\alpha}_{1i} = \bar{\alpha}_1 + \varepsilon_{1i} \]

\[ \ln y_i = \bar{\alpha}_0 + \bar{\alpha}_{1i} S_i + \{ \varepsilon_{0i} + \varepsilon_{1i} S_i \} . \]
II. Correlated Random Coefficient Models

(3) \[ Y_i = X_i \beta_i \]

for observation \( i, i = 1, ..., I \).

\( X_i \) is \( 1 \times K \) vector, \( \beta_i \) is \( K \times 1 \) vector.

(1), \( \alpha_{0i} = \beta_{0i} \) and \( \alpha_{1i} = \beta_{1i} \)

(4) \[ \bar{\beta} = E(\beta_i) < \infty \quad \text{and} \]

\[ VAR(\beta_i) = \sum_{\beta,\beta} < \infty \]

(5) \[ Y_i = X_i \beta_i = X_i \bar{\beta} + X_i (\beta_i - \bar{\beta}) \]

(6a) \[ E(W_i \mid X_i) = 0 \]
(6b) \[ \text{VAR}(W_i | X_i) = X_i' \sum_{\beta, \beta} X_i. \]

\[ X_i = Z_{1i} \pi + V_i, \quad i = 1, \ldots, I \]

(7) \[ \beta_i = \Phi' Z_{2i}' + U_i \]

Array \( X_i, Z_{1i}, Z_{2i} \) into matrices

\[ X(I \times K), Z_1(I \times K_1) \text{ and } Z_2(I \times K_2) \]
(A-1) \[ E(U_i \mid Z_{1i}, Z_{2i}) = 0 \]

(A-2) \[ E(V_i \mid Z_{1i}, Z_{2i}) = 0 \]

(A-3) \[ E(V_i U_i \mid Z_{1i}, Z_{2i}) = \sum_{u,v} \]

is not a function of \( Z_{1i}, Z_{2i} \).

\[ E(X_i \mid Z_{1i}) = Z_{1i} \pi, \]

(8) \[ Y_i = (Z_{1i} \pi + V_i)(\Phi' Z'_{2i} + U_i) \]

\[ = (Z_{1i} \pi) \cdot (\Phi' Z'_{2i}) + (Z_{1i} \pi) U_i \]

\[ + V_i \Phi' Z'_{2i} + V_i U_i. \]
(9) \( Y_i = (\pi_1' Z_{1,i} Z_{2,i}) \Phi_1 + (\pi_2' Z_{1,i} Z_{2,i}) \Phi_2 \)

\[ + ... + (\pi_K' Z_{1,i} Z_{2,i}) \Phi_K + \varepsilon_i. \]
As a consequence of assumptions (A-1–A-3)

\[ E(\varepsilon_i \mid Z_{1i}, Z_{2i}) = E(V_i U_i) \]

\[ Z_i = (\pi'_1 Z'_{1i} Z_{2i}, \pi'_2 Z'_{1i} Z_{2i}, \ldots, \pi'_K Z'_{1i} Z_{2i}). \]

Identification of \( \Phi \)

\[ (A-4) \quad \text{rank } E(Z'_i Z_i) = K \cdot K_2. \]

Overall model intercept is not identified unless

\[ E(V_i U_i) = 0. \]
Card’s Model of Schooling and Earnings

\[ U(y, S) = \ln y(S) - \varphi(S) \quad \varphi'(S) > 0 \]

and

\[ \varphi''(S) > 0. \]

The schooling-earnings relationship is

\[ y = g(S). \]

\[ \frac{g'(S)}{g(S)} = \varphi'(S). \]

\[ \frac{g'(S_i)}{g(S_i)} = \beta_i(S_i) = b_i - k_1 S_i \quad k_1 \geq 0 \]

\[ \varphi'(S_i) = \delta_i(S_i) = r_i + k_2 S_i \quad k_2 \geq 0. \]

\[ S_i = \frac{(b_i - r_i)}{k}, \text{ where } k = k_1 + k_2. \]

\[ b_i : \text{“ability”} \]

\[ r_i : \text{“opportunity cost”} \]
(cost of schooling) or “cost of funds.” Structural model of earnings

\[ \ln y_i = a_i + b_i S_i - \frac{1}{2} k_1 S_i^2. \]
(1) Can instrumental variables estimators recover anything interesting?

and

(2) What parameters are identified?
A. $r_i$ is observed

\( \perp\perp \) denotes statistical independence.

\[ r_i \perp\perp (b_i, a_i). \]

Observing \( r_i \) implies that we observe \( b_i \) up to scale:

\[ S_i = \frac{(b_i - r_i)}{k}, \]  
so that \( b_i = r_i + kS_i \) and

\[ \bar{b} = E(b_i) = \bar{r} + kE(S_i). \]
\[
\frac{\text{COV}(\ln y_i, r_i)}{\text{COV}(S_i, r_i)} = \frac{[\bar{a} - a_i + (\bar{b} - \bar{b})(S_i - \bar{S}) + \bar{b}S_i + \bar{b}S - \bar{b}\bar{S}]}{E \left\{ \left[ \frac{b_i - r_i}{k} \right] [r_i - \bar{r}] \right\}}
\]

\[
= \frac{1}{k} E[(\Delta r)(\Delta b)(\Delta b - \Delta r)] - \frac{\bar{b}}{k} \sigma^2_r \left( \frac{\sigma^2_r}{k} \right)
\]

(A-5) \( E[(\Delta r)(\Delta b)^2] = 0 \) and \( E[(\Delta r)^2\Delta b] = 0 \),

\[
\left[ \frac{\text{COV}(\ln y_i, r_i)}{\text{COV}(S_i, r_i)} \right] = \bar{b}.
\]
$\bar{b}$ is not identified using $r_i$ as an instrument if $b_i \not\perp r_i$. In that case, $E[(\Delta r)(\Delta b)^2] \neq 0$ and

$E[(\Delta r)^2(\Delta b)] \neq 0$. Suppose we write

$$r_i = L_i \gamma + M_i$$

$$S_i = (b_i - L_i \gamma - M_i)/k$$ and $k$ is identified since we know $\gamma$.

Can estimate the distribution of $b_i$:

$$b_i = r_i + k S_i$$

$k$ is identified and $(r_i, S_i)$ are known.
True even if $\gamma = 0$ provided that
\[ r_i \perp \perp (a_i, b_i). \]

With instruments for $r$, can let $r_i \perp \perp b_i$. The model is fully identified $L_i \perp \perp (M_i, b_i, a_i)$. 
B. The Case where $r_i$ is not observed

If $r_i$ is not observed and so cannot be used as an instrument, but

$$r_i = L_i \gamma + M_i$$

and $L_i \perp\!\!\!\!\!\!\!\!\!\perp (M_i, a_i, b_i)$

$$\ln y_i = a_i + \bar{b}S_i + (b_i - \bar{b})S_i.$$  
$$\ln y_i = a_i + b_i (b_i - L_i \gamma - M_i)/k.$$  
$\text{COV}(\ln y_i, L_i) = \bar{b}\text{COV}(S_i, L_i)$

$$\gamma = (\gamma_0, \gamma_1),$$
\[
S_i = \frac{b_i - L_i \gamma_1 - M_i}{k} - \frac{\gamma_0}{k}
\]

\[
S_i = -L_i \frac{\gamma_1}{k} + \frac{b_i - M_i}{k} - \frac{\gamma_0}{k}
\]
C. The Effect of Treatment on the Treated in the Card Model

These \((b_i, r_i)\) pairs for each \(S_i = s\) define the treated. \(E(b_i \mid b_i = sk + r_i)\).

\[
E(b_i | b_i = sk + r_i) = \frac{\int b f_{b,r}(b, b - sk)db}{f_{b-r}(sk)_{b \geq r}}.
\]

Assuming that \(a_i\) is mean independent of \((r_i, b_i)\)
\[ E(\ln y_i \mid S_i) = E(a_i \mid S_i) + E(b_iS_i \mid S_i) = \mu_a + E(b_i \mid S_i)S_i = \mu_a + E(b_i \mid b_i = S_i k + r_i)S_i, \]

\( \mu_a \) is not a function of \( S \).
D. Adding Selection Bias to Card’s Model

\[ E(\ln y_i \mid S_i) = E(a_i \mid S_i) + E(b_i S_i \mid S_i) \]

\[ = E(a_i \mid S_i) + E(b_i \mid S_i) S_i. \]

\[ L_i \perp \perp (a_i, b_i, M_i) \text{ and } E(r_i \mid L_i) \text{ is linear.} \]

Schooling model as before.

\[ E(\ln y_i \mid L_i) = E(a_i \mid L_i) + E(b_i S_i \mid L_i) \]

\[ L_i \]
\[ \eta = \mu_a + \frac{\sigma_b^2}{k} - \frac{E(b_i M_i)}{k}, \text{ where} \]

\[ \sigma_b^2 = \text{VAR}(b_i). \]

\( \bar{b} \) is identified even with selection bias.
Consider Treatment on The Treated

\[ E(\ln y_i \mid S_i) = E(a_i \mid S_i) + E(b_iS_i \mid S_i) \]

and

\[ E(\ln y_i \mid L_i) = \mu_a + E(b_iS_i \mid L_i) = \]

\[ \eta + \bar{b}E(S_i \mid L_i) \]

if and only \( \bar{b} \) is in the agent information set
Any Further Information?

\[
E(\ln y_i \mid S_i, L_i) =
\]

\[
E(a_i \mid S_i, L_i) + E(b_iS_i \mid S_i, L_i)
\]

\[
E(b_iS_i \mid S_i, L_i) = E(b_iS_i \mid S_i)
\]

If \( b_i \) not known when \( S_i \) determined and

we use mean value:

\[
S_i = \frac{\bar{b} - r_i}{k}.
\]
Adding Instruments for $b$

Take $k_1 = 0$.

\[ r_i = L_i \gamma + M_i \]

\[ b_i = N_i \delta + O_i \]

Assume initially $L_i$ and $N_i$ are disjoint.

\[(L_i, N_i) \perp \perp (M_i, O_i, a_i),\]

\[ E(M) = E(O) = 0. \]
Schooling equation:

\[ S_i = \frac{1}{k} \left( (N_i \delta - L_i \gamma) + (O_i - M_i) \right) \]

\[ \gamma = (\gamma_0, \gamma_1), \quad \delta = (\delta_0, \delta_1), \]

Identify \( \frac{\gamma_1}{k} \) and \( \frac{\delta_1}{k} \) from schooling equation.
From earnings equation:

\[ E(\ln y_i | S_i = s, L_i = l, N_i = n) \]

\[ = E(a_i | S_i = s, L_i = l, N_i = n) \]

\[ + E(b_i | S = s, L = l, N = n)s \]

\[ = E(a_i | S_i = s) + n\delta s + \]

\[ E(O_i | k_s = (n\delta - l\gamma) + O_i - M_i))s \]

\[ = E(a_i | S_i = s) + n\delta s \]

\[ + E(O_i | O_i = k_s - (n\delta - l\gamma) + M_i)s \]
Likewise:

\[
E(\ln y_i|S_i = s, L_i = l, N_i = n) = E(a_i|S_i = s) + l\gamma s + ks^2 \\
+ E(M_i|M_i = -ks + (n\delta - l\gamma) + O_i)s.
\]

Let \(Z_i = (L_i, N_i)\), and consider two evaluation points \((z, z')\) s.t. \(n\delta_1 - l\gamma_1 = n'\delta_1 - l'\gamma_1\).

(Can fix this from the schooling equation)
\[
E(\ln y_i|S_i = s, Z_i = z) - E(\ln y_i|S_i = s, Z_i = z')
\]
\[
= (l - l')\gamma_1 s = (n - n')\delta_1 s
\]

identify \(\gamma_1\) and \(\delta_1\).
Now consider

\[
k \left[ E(\ln y_i | Z_i = z) - E(\ln y_i | Z_i = z') \right]
\]

\[
= k \left[ E(b_i S_i | Z_i = z) - E(b_i S_i | Z_i = z') \right]
\]

\[
= E \left( (n\delta + O_i)(n\delta - l\gamma + O_i - M_i) \right)
\]

\[
- E \left( (n'\delta + O_i)(n'\delta - l'\gamma + O_i - M_i) \right)
\]

\[
= E \left( (n\delta + O_i)(n\delta - l\gamma + O_i - M_i) \right)
\]

\[
- E \left( (n'\delta + O_i)(n'\delta - l'\gamma + O_i - M_i) \right)
\]

\[
= (n\delta_1)^2 - (n'\delta_1)^2 + 2\delta_0(n - n')\delta_1
\]

\[
n\delta_1 l\gamma_1 - n'\delta_1 l'\gamma_1 + \delta_0(l - l')\gamma_1
\]

28
\[ + \gamma_0 (n - n') \delta_1. \]
We have knowledge of $\delta_1$ and $\gamma_1$ from above, we know $(n\delta_1)^2 - (n'\delta_1)^2$ and $n\delta_1 l\gamma_1 - n'\delta_1 l'\gamma_1$:

$$2\delta_0(n-n')\delta_1 + \delta_0(l-l')\gamma_1 + \gamma_0(n-n')\delta_1.$$

By evaluating this expression at $(z, z')$ s.t. $n = n'$, $(l \neq l')$, we identify $\delta_0$. Given the identification of $\delta_0$, we can use $(Z, Z')$ such that $n \neq n'$ to identify $\gamma_0$. We identify the distribution of $O - M$. 

30
Shift $\ell$ holding $n$ Constant

\[
E(\ln y_i | S_i = s, L_i = l, N_i = n)
\]

\[-E(\ln y_i | S_i = s, L_i = l', N_i = n)\]

\[= E(O_i | O_i - M_i = k s - (n \delta - l \gamma)) s\]

\[-E(O_i | O_i - M_i = k s - (n \delta - l' \gamma)) s.\]
Further Results on Identification

\[ E(O_i | O_i - M_i = c) - E(O_i | O_i - M_i = c') \]

for a range of \((c, c')\) values that depend on the support of the \(L\) instrument. Likewise, can identify

\[ E(M_i | O_i - M_i = c) - E(M_i | O_i - M_i = c') \]

for a range of \((c, c')\) values that depend on the support of the \(N\) instrument.
Can thus identify $\gamma, \delta$, the distribution of $O - M$, and

$$E(O_i | O_i - M_i = c) - E(M_i | O_i - M_i = c').$$

This allows us to identify the averages

$$E(b_i) = E(N)\delta, \text{ and } E(r_i) = E(L_i)\gamma.$$

Does not imply identification of the distribution of $(O, M)$, and thus does not imply identification of the treatment on the treated parameters unless $\alpha_i$ is independent of $S_i$. 

33
Cannot separate $E(\alpha_i \mid S_i = s)$ from $E(O_i \mid S = s)s$, and thus cannot identify treatment on the treated.
Control Function Estimator

\[ \ell n y_i = a_i + b_i S_i \]

\[ = \bar{a}_i + \bar{b} S_i + [(b_i - \bar{b}) S_i + a_i - \bar{a}] \]

Control Function (Developed in Selection Bias Literature)

\[ E((b_i - \bar{b}) S_i + (a_i - \bar{a}) \mid S_i, Z_i) \]

\[ = (a_i - \bar{a} \mid S_i, Z_i) + E(b_i - \bar{b} \mid S_i, Z_i) S_i \]

Need An Exclusion Restriction (Z).
LATE and LIV Parameters In the Card Model

\[ \text{LATE}(z, z') = \frac{E(\ln y_1 | Z_i = z) - E(\ln y_1 | Z_i = z')} {E(S | Z_i = z') - E(S | Z_i = z)} \]

\[ = \frac{E(b_i S_i | Z_i = z) - E(b_i S_i | Z_i = z')} {E(S | Z_i = z') - E(S | Z_i = z)} \]

\[ = \frac{(n\delta)^2 - (n'\delta)^2 + n\delta l'\gamma - n'\delta l\gamma} {(n - n')\delta + (l - l')\gamma} . \]
If $n = n'$, then

$$\text{LATE}(z, z') = n\delta = E(b|N = n)$$

If $l = l'$, this equals

$$\text{LATE}(z, z') = l\gamma + \frac{(n\delta)^2 - (n'\delta)^2}{(n - n')\delta}$$

$$= l\gamma + 2n\delta - (n - n')\delta$$

$$= l\gamma + 2n\delta - (n - n')\delta$$

$$= l\gamma + n\delta + n'\delta$$

$$= E(r|L = l) + E(b|N = n) + E(b|N = n')$$
If $n = n'$, and $l' \to l$, then

$$\text{LATE}(z, z') \to n\delta$$

If $l = l'$, and $n' \to n$, then

$$\text{LATE}(z, z') \to l\gamma + 2n\delta$$

$$= E(r|L = l) + 2E(b|N = n).$$
V. Distribution of (O-M) Identified. Not of (O,M). Some Variances Are

Assume \( a_i \perp \perp (b_i, r_i) \)

\[ \omega_i = \ell n y_i - (\bar{a} + \bar{b} S_i) \]

\[ b_i = N_i \delta + O_i \]

\[ r_i = L_i \gamma + M_i \]

\[ \omega_i = (a_i - \bar{a}) + \frac{(N_i \delta - L_i \gamma)}{k} O_i \]

\[ + \frac{(O_i - M_i)}{k} O_i. \]
\[
V\text{ar}(\omega_i) = V\text{ar}(a) + V\text{ar}\left(\frac{[O - M]O}{k}\right) \\
+ COV\left(\frac{[O][O - M]}{k}, O\right)\left(\frac{N_i\delta - L_i\gamma}{k}\right) \\
+ \left[\frac{(N_i\delta - L_i\gamma)}{k}\right]^2 \sigma^2_O.
\]

**Run Regression**

\[
\tilde{S} = \frac{N_i\delta - L_i\gamma}{k}
\]

\[
V\text{ar}(\omega_i) = q_0 + \tilde{S} \left( COV\left[\frac{(O)(O-M)}{k}, O\right] + \right) \\
\tilde{S}^2 \sigma^2_0.
\]

Can identify \( \sigma^2_0\).
Examining The Identifying Assumptions

Assumption (A-3) is a crucial identifying assumption Roy model. Consider Heckman and Honoré (1990):

\[ y_0 = \mu_0(\tau_0) + \varepsilon_0 \]
\[ y_1 = \mu_1(\tau_1) + \varepsilon_1 \]
\[ (\varepsilon_0, \varepsilon_1) \perp \perp (\tau_0, \tau_1, w); \]
\[ D = 1 \text{ if } y_1 \geq y_0 + \varphi(w) \]

\[ D = 0 \quad \text{otherwise.} \]

(12) \[ y = Dy_1 + (1 - D)y_0 = \mu_0(\tau_0) + \]

\[ (\mu_1(\tau_1) - \mu_0(\tau_0)) \]

\[ + \varepsilon_1 - \varepsilon_0)D + \varepsilon_0. \]

\( D \) plays the role of \( X \)
\[ \mu_1(\tau_1) - \mu_0(\tau_0) + \varepsilon_1 - \varepsilon_0 \] plays the role of \( \beta \), \( w \) plays the role of the \( Z \).
\[ E(y|\tau_0, \tau_1, w) = \mu_0(\tau_0) \]

\[ + \left\{ \mu_1(\tau_1) - \mu_0(\tau_0) + E(\varepsilon_1 - \varepsilon_0|\tau_0, \tau_1, w, D = 1) \right\} \]

\[ \cdot E(D = 1|\tau_0, \tau_1, w). \]
Varying $w$ while fixing $\tau_0$ and $\tau_1$ does not in general identify $\mu_1(\tau_1) - \mu_0(\tau_0)$ ($= \bar{\beta}$ in the previous notation of Section II) unless $\varepsilon_1 - \varepsilon_0$ does not enter the decision maker’s information set at the time the decision about $D$ is made.
Alternative Way to Make Point

\[ D = E(D|\tau_0, \tau_1, w) + \nu. \]
\[ y = \mu_0(\tau_0) + (\mu_1(\tau_1) - \mu_0(\tau_0) + \varepsilon_1 - \varepsilon_0) \cdot \\
\quad (E(D|\tau_0, \tau_1, w) + \nu) + \varepsilon_0 \\
= \mu_0(\tau_0) + [\mu_1(\tau_1) - \mu_0(\tau_0)]E(D|\tau_0, \tau_1, w) \\
\quad + (\varepsilon_1 - \varepsilon_0)E(D|\tau_0, \tau_1, w) \\
\quad + [\mu_1(\tau_1) - \mu_0(\tau_0)]\nu + (\varepsilon_1 - \varepsilon_0)\nu + \\
\varepsilon_0. \]

\( \nu \) plays the role of \( V \) and \( \varepsilon_1 - \varepsilon_0 \) plays the role of \( U \).
\[ E[(\varepsilon_1 - \varepsilon_0)\nu|\tau_0, \tau_1, w] \]

\[ = E(\varepsilon_1 - \varepsilon_0|D = 1, \tau_0, \tau_1, w)E(D = 1|\tau_0, \tau_1, w). \]

If \( \varepsilon_1 - \varepsilon_0 \) enters the agent’s information set, this term does not vanish and in general depends on \( \tau_0, \tau, \) and \( w. \)
Training As A Form of Job Search

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

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

In the unemployed state, a person receives a nonmarket benefit, $N$.

“Gittens Index”
\[ V_{ju} = \]
\[
N - c_j + \frac{\lambda_j}{1 + r E_j} \max\{V_{je}, V_{ju}\} + \\
\frac{(1 - \lambda_j)}{1 + r} V_{ju}.
\]
Value of Nonmarket time