Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling
Author(s): James Heckman and Edward Vytlacil
Source: *The Journal of Human Resources*, Vol. 33, No. 4 (Autumn, 1998), pp. 974-987
Published by: University of Wisconsin Press
Stable URL: http://www.jstor.org/stable/146405
Accessed: 26-10-2015 16:31 UTC

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
Edward Vytlacil

ABSTRACT

This paper considers the use of instrumental variables to identify a correlated random coefficients model in which coefficients are correlated with (or stochastically dependent on) the regressors. A correlated random coefficients model is central to the human capital earnings model. Conditions are given under which instrumental variables identify the average rate of return. These conditions are applied to David Card's version of Gary Becker's Woytinsky lecture.

I. Introduction

A random coefficients model of the economic return to schooling has been an integral part of the human capital literature since the seminal research of Becker and Chiswick (1966), Chiswick (1974), Chiswick and Mincer (1972) and Mincer (1974). In its most stripped-down form, the model writes log earnings for person i with schooling level S_i as

$$(1)\quad \ln y_i = \alpha_{0i} + \alpha_{1i} S_i$$

James J. Heckman is a professor of economics at the University of Chicago and a Senior Fellow at the American Bar Foundation. Edward Vyltacil is a Sloan Fellow at the University of Chicago. *The essential identification results in this paper are presented in Heckman (1982, 1995), and have been widely circulated. This research is supported by NSF-SBR-97-09-873, NIH: R01-HD32058-01A1, and the American Bar Foundation. The authors thank Richard Blundell and Chris Taber for comments on a 1995 draft of this paper, and Kevin Murphy for insightful comments on the 1982 version. Data used in this article can be obtained from the authors from March 1999 through February 2002.*
[Submitted January 1998; accepted June 1998]

where the "rate of return" α_{1i} varies among persons as does the intercept, α_{0i}. Letting $\alpha_{0i} = \bar{\alpha}_0 + \varepsilon_{0i}$ and $\bar{\alpha}_{1i} = \bar{\alpha}_1 + \varepsilon_{1i}$ where $\bar{\alpha}_0$ and $\bar{\alpha}_1$ are the means of α_{0i} and α_{1i}, respectively, so that the means of ε_{0i} and ε_{1i} are zero, earnings equation (1) can also be written as

$$(2) \quad \ln y_i = \bar{\alpha}_0 + \bar{\alpha}_1 S_i + \{\varepsilon_{0i} + \varepsilon_{1i} S_i\}.$$

Equations (1) and (2) are the basis for a human capital analysis of wage inequality in which the variance of log earnings is decomposed into components due to the variance in S_i and components due to the variation in the rate of return (variance in α_{li}), the mean rate of return across regions or time ($\bar{\alpha}_1$), and mean schooling levels ($\bar{S}$).

Conventional in this literature is the assumption that the rate of return α_{1i} is uncorrelated with or is independent of S_i. This assumption is convenient but is not implied by economic theory. It is plausible that the return to schooling declines with the level of schooling. It is also plausible that there are unmeasured ability or motivational factors that affect the return to schooling and are also correlated with the level of schooling. Rosen (1977) discusses this problem in some detail within the context of hedonic models of schooling and earnings.[1]

This paper develops an instrumental variables estimator of $\bar{\alpha}_1$ for the case where α_{1i} is correlated with S_i. We examine the economic plausibility of the identifying assumptions by drawing on Card's (1995) recent random coefficient model of the returns to schooling, a model that is explicitly derived from economic theory. We also consider conditions under which it is possible to estimate the effect of schooling on the schooled in his model.

We develop a more general case than the one considered in the classical human capital literature. We also allow the rate of return to depend on observed characteristics as in Card and Krueger (1992) and Heckman, Layne-Farrar and Todd (1996).

II. Correlated Random Coefficient Models

Using a more general model, write the outcome for person i as

$$(3) \quad Y_i = X_i\beta_i \qquad \text{for observation } i,\ i = 1, \ldots, I.$$

X_i is a $1 \times K$ vector, β_i is a $K \times 1$ vector. The error term associated with the intercept is the usual error term in this model. In terms of Equation (1), $\alpha_{0i} = \beta_{0i}$ and $\alpha_{1i} = \beta_{1i}$ where β_{0i} is the intercept and β_{1i} is the slope coefficient.

In the standard random coefficients model as employed by Becker and Chiswick (1966), Chiswick (1974), and Mincer (1974), β_i is assumed to be a draw from a population with density $f(\beta)$ where β_i is assumed to be independent of (or at least uncorrelated with) X_i. Assuming that

$$(4) \quad \bar{\beta} = E(\beta_i) < \infty \text{ and that } \mathrm{Var}(\beta_i) = \Sigma_{\beta,\beta} < \infty \text{ in all components,}$$

1. A similar problem arises in analyses of the impact of unionism on relative wages. See Lewis (1963).

Equation (3) may be rewritten as

$$(5)\quad Y_i = X_i\beta_i = X_i\bar{\beta} + X_i(\beta_i - \bar{\beta}).$$

To simplify the notation, define $W_i = X_i(\beta_i - \bar{\beta})$. In the standard random coefficients model

$$(6a)\quad E(W_i|X_i) = 0$$

$$(6b)\quad \mathrm{Var}(W_i|X_i) = X_i' \Sigma_{\beta,\beta} X_i.$$

Under standard conditions, $\bar{\beta}$ is consistently estimated by ordinary least squares. There is heteroscedasticity in the errors that is readily solved using standard econometric methods (see, for example, Amemiya 1985).

However, the economics of the model often suggest that Assumption 6a is false. Unobservables that determine β_i also determine the elements of X_i so that $E(W_i|X_i) \neq 0$. Ordinary least squares estimators of $\bar{\beta}$ are inconsistent. However, under the following conditions, it is possible to consistently estimate the components of $\bar{\beta}$ except for the overall intercept of the model.

Suppose that X_i can be expressed as a linear function of exogenous variables Z_{1i} and an error V_i:

$$X_i = Z_{1i}\pi + V_i, \qquad i = 1, \ldots, I$$

where Z_{1i} is a $1 \times K_1$ vector of variables and π is a $K_1 \times K$ matrix of coefficients. In addition, assume that

$$(7)\quad \beta_i = \Phi' Z_{2i}' + U_i$$

where Z_{2i} is a $1 \times K_2$ vector of variables, some of which may be in common with X_i and Z_{1i}, and Φ is a $K_2 \times K$ matrix of coefficients.

Array X_i, Z_{1i}, Z_{2i} into matrices $X(\mathrm{I} \times K)$, $Z_1(\mathrm{I} \times K_1)$ and $Z_2(\mathrm{I} \times K_2)$ respectively. Array U_i and V_i into $\mathrm{I} \times K$ matrices U and V respectively. In this notation, we assume that (X_i, Z_{1i}, Z_{2i}) are independent over i and that (U_i, V_i) are independent over i. Further assume that

$$(\mathrm{A1})\quad E(U_i|Z_{1i}, Z_{2i}) = 0$$

$$(\mathrm{A2})\quad E(V_i|Z_{1i}, Z_{2i}) = 0$$

$$(\mathrm{A3})\quad E(V_iU_i|Z_{1i}, Z_{2i}) = \Sigma_{u,v} \text{ is not a function of } Z_{1i}, Z_{2i}.$$

Conditional homoscedasticity of covariances Assumption A3 is controversial and is not satisfied in many discrete choice models of schooling and labor earnings as we discuss below. It does not require homoscedasticity of variances, however. Observe that as a consequence of these assumptions, π is identified from

$$E(X|Z_{1i}) = Z_{1i}\pi,$$

provided that the rank of population variance of Z_1, Σ_{z_1,z_1}, is K_1. For observation i we may write

$$(8)\quad Y_i = (Z_{1i}\pi + V_i)(\Phi' Z_{2i}' + U_i) = (Z_{1i}\pi)\cdot(\Phi' Z_{2i}') + (Z_{1i}\pi)U_i + V_i\Phi' Z_{2i}' + V_iU_i.$$

Rewriting the equation, we obtain the following regression model:

$$(9)\quad Y_i = (\pi_1' Z_{1i}' Z_{2i})\Phi_1 + (\pi_2' Z_{1i}' Z_{2i})\Phi_2 + \cdots + (\pi_K' Z_{1i}' Z_{2i})\Phi_K + \varepsilon_i$$

where π_ℓ and Φ_ℓ are the ℓth columns of π and Φ, respectively. Observe that as a consequence of assumptions (A1–A3), $E(\varepsilon_i | Z_{1i}, Z_{2i}) = E(V_i U_i)$, which does not depend on Z_{1i} or Z_{2i}. Define the KK_2-length row vector Z_i as

$$Z_i = (\pi_1' Z_{1i}' Z_{2i}, \pi_2' Z_{1i}' Z_{2i}, \ldots, \pi_K' Z_{1i}' Z_{2i}).$$

Z_i can be consistently estimated under the stated conditions since we can consistently estimate the π. Identification of Φ requires that the covariance matrix of Z_i be of rank $(K \cdot K_2)$:

$$(\text{A4})\quad \operatorname{rank} E(Z_i' Z_i) = K \cdot K_2.$$

The overall model intercept is not identified unless $E(V_i U_i) = 0$. The slope coefficients can be identified under the stated rank conditions.

The appendix presents the standard errors for a computationally simple two step estimator (that estimates the π in a first stage regression of X on Z_1 and substitutes the estimated π into (9)) and a more efficient GMM one step estimator. In the next section, we apply our analysis to Card's version of Becker's (1967) Woytinsky lecture.

III. An Application To Card's Model of Schooling and Earnings

In Card's (1995) model, the preferences of a person over income and schooling are:

$$U(y, S) = \ln y - \varphi(S) \quad \varphi'(S) > 0 \quad \text{and} \quad \varphi''(S) > 0.$$

The schooling-earnings relationship is $y = g(S)$. This is a hedonic model of schooling. The first order condition is

$$(10)\quad \frac{g'(S)}{g(S)} = \varphi'(S).$$

Linearize the model to obtain:

$$\frac{g'(S_i)}{g(S_i)} = \beta_i(S_i) = b_i - k_1 S_i \qquad k_1 \geq 0$$

$$\varphi'(S_i) = \delta_i(S_i) = r_i + k_2 S_i \qquad k_2 \geq 0.$$

Substituting into the first order condition, we obtain that the optimal level of schooling is $S_i = (b_i - r_i)/k$, where $k = k_1 + k_2$.

The sources of heterogeneity among persons in the model are: b_i, "ability" or market ability, the way S is transformed into earnings. (School quality may operate through the b_i for example). A second source of heterogeneity is r_i: the "opportunity cost'' (cost of schooling) or ''cost of funds.'' We integrate the first-order condition (10) to obtain the following structural model of earnings.

$$\text{(11)} \quad \ln y_i = a_i + b_i S_i - \frac{1}{2} k_1 S_i^2.$$

For simplicity we assume that $k_1 = 0$. Our methods apply to the more general case but the argument is more familiar for $k_1 = 0$.[2] In this framework $S_i = X_i$ of Section II and $a_i = \beta_{0i}$ and $b_i = \beta_{1i}$. We now use the analysis of Section II to address two questions:

(1) Can instrumental variables estimators recover anything interesting? and

(2) What parameters are identified?

Observe that S_i does *not* directly depend on the error term a_i. Of course, a_i may be stochastically dependent on (b_i, r_i). In the context of Card's model, we consider conditions under which one can identify $\bar{b}$, the mean ability in the population. First we consider the case where the marginal cost of funds, r_i, is observed.

A. $\mathbf{r_i}$ *is Observed*

Use the notation ''$\perp\!\!\!\perp$'' to denote statistical independence. Assume

(A5) $r_i \perp\!\!\!\perp (b_i, a_i)$.

Observing r_i implies that we observe b_i up to scale: $S_i = (b_i - r_i)/k$, so that $b_i = r_i + kS_i$ and $\bar{b} = E(b_i) = \bar{r} + kE(S_i)$.

r_i is a valid instrument for S_i under the assumption that $k_1 = 0$. In the notation of Section II, r_i is an element of Z_{1i}. Card's model excludes any regressors from Z_{2i}. Form

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

$$= \frac{\dfrac{1}{k}E[(\Delta r)(\Delta b)(\Delta b - \Delta r)] - \dfrac{\bar{b}}{k}\sigma_r^2}{-\dfrac{\sigma_r^2}{k}}$$

where $\Delta X = X - E(X)$. As a consequence of assumption (A5), $E[(\Delta r)(\Delta b)^2] = 0$ and $E[(\Delta r)^2 \Delta b] = 0$, so

$$\left[\frac{\text{Cov}(\ln y_i, r_i)}{\text{Cov}(S_i, r_i)}\right] = \bar{b}.$$

2. Heckman (1997, Section VII) presents a model closer to the original classical Rosen (1977)—Wicksell (1934, p. 178–83) ''tree cutting'' model of schooling which explicitly accounts for the earnings foregone while persons are in school.

Observe that $\bar{b}$ is not identified using r_i as an instrument if $b_i \not\!\perp\!\!\!\perp r_i$. In that case, $E[(\Delta r)(\Delta b)^2] \neq 0$ and $E[(\Delta r)^2(\Delta b)] \neq 0$. If r_i is known and we can write $r_i = L_i\gamma + M_i$ where $E(M_i|L_i) = 0$, then γ is identified, provided a rank condition is satisfied. In the notation of Section II, L_i is Z_{1i}. We require that L_i be at least mean independent of (M_i, b_i, a_i). From the schooling equation we can write $S_i = (b_i - L_i\gamma - M_i)/k$ and k is identified since we know γ.

Observe that we can estimate the distribution of b_i since

$$b_i = r_i + kS_i$$

and k is identified and (r_i, S_i) are known. This is true even if there are no instruments, ($\gamma = 0$), provided that $r_i \perp\!\!\!\perp (a_i, b_i)$. With the instruments that satisfy at least the mean independence condition, we can allow $r_i \not\!\perp\!\!\!\perp b_i$ and all parameters and distributions are still identified. The model is fully identified provided r_i is observed and $L_i \perp\!\!\!\perp (M_i, b_i, a_i)$.[3]

B. The Case where $\mathbf{r}_i$ 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 $L_i \perp\!\!\!\perp (M_i, a_i, b_i)$, then the analysis of Section II implies that $\bar{b}$ is identified. Recall that $\ln y_i = a_i + \bar{b}S_i + (b_i - \bar{b})\, S_i$. Substitute for S_i to get an expression of y_i in terms of L_i: $\ln y_i = a_i + b_i(b_i - L_i\gamma - M_i)/k$. We obtain the vector equation:

$$\mathrm{Cov}(\ln y_i, L_i) = \bar{b}\,\mathrm{Cov}(S_i, L_i)$$

so $\bar{b}$ is identified from the population moments. One can use the GMM formula presented in Hansen (1982, 1985) to construct an efficient estimator if there is more than one nonconstant element in L_i. Partition $\gamma = (\gamma_0, \gamma_1)$, where γ_0 is the intercept and γ_1 is the vector of slope coefficients. From the schooling equation, we have

$$\begin{aligned} S_i &= \frac{b_i - L_i\gamma_1 - M_i}{k} - \frac{\gamma_0}{k} \\ &= -L_i\frac{\gamma_1}{k} + \frac{b_i - M_i}{k} - \frac{\gamma_0}{k}. \end{aligned}$$

We can identify γ_1/k from the schooling equation. However, we cannot identify the distribution of b_i or r_i unless further assumptions are invoked; however, we *can* identify the mean of the b_i, $\bar{b}$. We cannot separately identify γ_0, γ_1 or k.

C. The Effect of Treatment on the Treated in the Card Model

As noted by Heckman (1996, 1997) and Heckman and Smith (1998), the parameter "treatment on the treated" does not always address an economically interesting question. However, it is a widely used evaluation parameter and for this reason we

3. As we have stressed, the independence conditions are overly strong, but can be weakened to a mean independence assumption provided that we only seek to recover conditional means.

consider its identification in Card's model. The treated are those at a given level of schooling, for example, $S_i = s$ so that $(b_i - r_i)/k = s$. This defines a set of values for b_i and r_i that fall on the line: $b_i = sk + r_i$. These (b_i, r_i) pairs for each $S_i = s$ define the treated. The mean value of "treatment" is $E(b_i | b_i = sk + r_i)$. This expression informs us how much an extra unit of s raises ln y for those with a given $S_i = s$. Let $f_{b,r}$ be the joint density of (b, r), and let f_{b-r} be the density of $b - r$.[4] Then

$$E(b_i | b_i = sk + r_i) = \frac{\int_{b \geq r} b f_{b,r}(b, b - sk) db}{f_{b-r}(sk)}.$$

The requirement $b \geq r$ guarantees nonnegative schooling. Assuming that a_i is mean independent of (r_i, b_i), we may write

$$\begin{aligned} E(\ln y_i | S_i) &= E(a_i | S_i) + E(b_i S_i | S_i) \\ &= \mu_a + E(b_i | S_i) S_i \\ &= \mu_a + E(b_i | b_i = S_i k + r_i) S_i, \end{aligned}$$

where μ_a is not a function of S.

This informs us how much an exogenous increase in schooling raises earnings. Under the stated conditions we can estimate $E(b_i | S_i = s)$ for each level of s from a kernel regression of ln y_i on S_i, evaluated at the point $S_i = s$.[5] This is the effect of "treatment on the treated" at the level of s. In contrast to $\bar{b}$, no instrumental variable is needed to recover this parameter.

D. Adding Selection Bias to Card's Model

Selection bias can arise in two distinct ways in the Card model: through dependence between a_i and b_i and through dependence between a_i and r_i. In this more general case,

$$E(\ln y_i | S_i) = E(a_i | S_i) + E(b_i S_i | S_i) = E(a_i | S_i) + E(b_i | S_i) S_i$$

Suppose that there is a variable L_i that affects r_i but not b_i and is independent of (a_i, M_i), namely, $L_i \perp\!\!\!\perp (a_i, b_i, M_i)$ and $E(r_i | L_i)$ is a nontrivial function of L_i.[6] In the special case of a linear schooling model as in Section B,

$$\begin{aligned} E(\ln y_i | L_i) &= E(a_i | L_i) + E(b_i S_i | L_i) \\ &= \mu_a + E(b_i S_i | L_i) \\ &= \eta + \bar{b} E(S_i | L_i).^7 \end{aligned}$$

4. The density f_{b-r} can be derived from the density $f_{b,r}$ as follows: $f_{b-r}(t) = \int_r^\infty F(b, b - t) db$. Also, $f_{b-r}(t) = \int f_{b,r}(t + r, r) dr$.

5. Assuming that $E(b_i | S_i = s)$ contains no term of order s^{-1}, we can identify $E(b_i | S_i = s)$ by substracting the intercept off from a kernel regression and dividing by $s (> 0)$, or by taking differences at successive values of s and dividing by the change in s as the change gets small.

6. The independence condition can be weakened to a mean independence condition.

7. $\eta = \mu_a + (\sigma_b^2 / k) - (E(b_i M_i)/k)$, where $\sigma_b^2 = \text{VAR}(b_i)$.

Because we can identify $E(S_i|L_i)$ we can identify $\bar{b}$. Thus, under the stated conditions, the IV method identifies $\bar{b}$ when there is selection bias. In the general nonparametric case for the schooling equation, this argument breaks down because b_i determines S_i in a general way.

Consider next the parameter ''treatment on the treated'' ($E(b_i|S_i)$). Under our assumptions, we can construct

$$E(\ln y_i|S_i) = E(a_i|S_i) + E(b_iS_i|S_i)$$

and

$$E(\ln y_i|L_i) = \mu_a + E(b_iS_i|L_i) = \eta + \bar{b}E(S_i|L_i).$$

This allows us to recover $\bar{b}$, as before. As before, and for the same reason, this argument breaks down in the general case. Can we get any further information out of

$$E(\ln y_i|S_i, L_i) = E(a_i|S_i, L_i) + E(b_iS_i|S_i, L_i)$$

using the fact that $E(b_iS_i|S_i, L_i) = E(b_iS_i|S_i)$?

If b_i is not known at the time the schooling decision is made (so decisions are made on the basis of $E(b_i) = \bar{b}$, so that $S_i = (\bar{b} - r_i)/k$, and the shock $(b_i - \bar{b})$ is realized only after schooling is completed), then $E(b_i|S_i) = \bar{b}$ and so IV estimates ''treatment on the treated'' as before (see Heckman 1997) provided $E(a_i|S_i, L_i)$ does not depend on S_i or L_i. In the general case we cannot use the IV method to identify the parameter ''treatment on the treated'' in the presence of selection bias because of the stochastic dependence between a_i, and S_i.[8]

IV. Summary, Extensions and Qualifications

This paper considers the use of instrumental variables to identify the average response to a continuous treatment when the response to the treatment varies among persons exposed to the same treatment level and the treatment level is determined, at least in part, by the person-specific response to treatment. We apply our methods to identify various parameters of Card's model of earnings and schooling.

The analysis of this paper extends readily to the case where the response to treatment and the treatment itself are nonlinear functions of the instruments Z_{1i} and Z_{2i}, respectively, provided that the error terms in both the response and treatment functions are additively separable and A3 is satisfied. Thus, assume the following model:

$$X_i = X(Z_{1i}; \pi) + V_i, \quad i = 1, \ldots, I$$

8. It is instructive to compare the Rosen-Wicksell model of schooling analyzed in Heckman (1997; Section VII) with the Card model. In Heckman's model, the term comparable to the a_i in Card's framework enters the schooling equation so there is always selection bias (as defined here) and a nonparametric estimator of $\ln y$ on S does not recover the parameter treatment on the treated. The fundamental difference is that in his model, earnings foregone are a cost of schooling. In Card's models they are not. Under Heckman's conditions, IV does not estimate the term comparable to $\bar{b}$, but instead estimates a ratio which is economically difficult to interpret. In his notation, IV estimates $E(\alpha_1/(\alpha_1 - r - \beta)|\beta > \alpha_1 - r > 0)/E(1/(\alpha_1 - r - \beta)|\beta > \alpha_1 - r > 0)$.

where there are K_π π parameters, and

$$\beta_i = \beta(Z_{2i}; \Phi) + U_i, \quad i = 1, \ldots, I$$

for K_Φ Φ parameters. X_i and β_i are general nonlinear functions of Z_{1i} and Z_{2i}, respectively, and are assumed to be continuously differentiable in terms of the parameters of the model. The model is identified, subject to a rank condition, if $E(V_i U_i | Z_{1i}, Z_{2i})$ does not depend on Z_{1i} or Z_{2i}. Define

$$Z_i = \left(\frac{\partial X(Z_{1i}; \pi)}{\partial \pi} \beta(Z_{2i}; \Phi), X(Z_{1i}; \pi) \frac{\partial \beta(Z_{2i}; \Phi)}{\partial \Phi} \right).$$

The required rank condition is that $E(Z_i Z_i')$ has rank $K_\pi \cdot K_\Phi$ when evaluated at the true parameter values π^0 and Φ^0. It is also possible to extend the analysis to a more general semiparametric setting.

Assumption A3 is a crucial identifying assumption that is not satisfied in many discrete choice models. In general, both the treatment response and treatment variables are *nonseparable* functions of both observables and unobservables. This makes it difficult to satisfy A3. For example, consider an extension of the Roy model, as developed in Heckman and Honoré (1990):

$$y_0 = \mu_0(\tau_0) + \varepsilon_0$$
$$y_1 = \mu_1(\tau_1) + \varepsilon_1$$

$(\varepsilon_0, \varepsilon_1) \perp\!\!\!\perp (\tau_0, \tau_1, w)$; $D = 1$ if $y_1 \geq y_0 + \varphi(w)$; $D = 0$ otherwise. The outcome equation is

$$\text{(12)} \quad y = Dy_1 + (1 - D)y_0$$
$$= \mu_0(\tau_0) + (\mu_1(\tau_1) - \mu_0(\tau_0) + \varepsilon_1 - \varepsilon_0)D + \varepsilon_0.$$

In the notation of Section II, D plays the role of X, $\mu_1(\tau_1) - \mu_0(\tau_0) + \varepsilon_1 - \varepsilon_0$ plays the role of β, and w plays the role of the Z_1. Then

$$E(y | \tau_0, \tau_1, w) = \mu_0(\tau_0) + (\mu_1(\tau_1) - \mu_0(\tau_0) + E(\varepsilon_1 - \varepsilon_0 | \tau_0, \tau_1, w, D = 1)) \times E(D = 1 | \tau_0, \tau_1, w).$$

For reasons presented in Heckman (1997), varying w while fixing τ_0 and τ_1 does not in general identify $\mu_1(\tau_1) - \mu_0(\tau_0)$ $(= \bar{\beta}$ in the 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.

An alternative way to make the same point is to write the conditional expectation of D as

$$D = E(D | \tau_0, \tau_1, w) + \nu$$

and substitute into Equation (12) to obtain

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

ν plays the role of V and $\varepsilon_1 - \varepsilon_0$ plays the role of U. Now, by construction, the third and fourth terms of the final expression vanish upon taking conditional expectations with respect to τ_0, τ_1, and w. The residual ν depends on $(\varepsilon_1 - \varepsilon_0)$ and it can be shown that

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

If $\varepsilon_1 - \varepsilon_0$ enters the agent's information set, this term does not vanish and in general depends on τ_0, τ, and w. The method of IV requires covariance homoscedasticity assumption A3, and A3 is artificial in many choice theoretic economic models of schooling. Using alternative identification methods, Heckman and Smith (1997, 1998) present conditions for identification of the entire distribution of treatment effects in the general case of dichotomous treatments when this homoscedasticity assumption is not satisfied.

Appendix A

Following Newey and McFadden (1994), we derive the standard errors for a two step estimation procedure and for a general GMM method procedure for the model presented in Section II.

A1. Two-Stage Estimator

Because we identify the overall intercept plus $E(V_i U_i)$, but neither term separately, in this appendix we *normalize* $E(V_i U_i) = 0$. Let $\tilde{\Phi}$ be Φ represented as a KK_2 column vector: $\tilde{\Phi} = (\Phi_1', \Phi_2', \ldots, \Phi_K')'$. Let Z be the $I \times KK_2$ matrix of Z_i stacked over individuals. Let X_i^k and V_i^k be the kth elements of X_i and V_i, respectively, and let X^k and V^k be the $I \times 1$ vectors of X_i^k and V_i^k stacked over individuals. Let $\tilde{X} = (X^{1'}, X^{2'}, \ldots, X^{K'})'$, $\tilde{\pi} = (\pi_1', \pi_2', \ldots, \pi_K')'$, and $\tilde{V} = (V^{1'}, V^{2'}, \ldots, V^{K'})'$. Let $\tilde{Z}_1$ be the $IK \times KK_1$ block diagonal matrix given as $\tilde{Z}_1 = \operatorname{diag}(Z_1, \ldots, Z_1)$. Let $Z(\tilde{\pi})$ be the Z matrix evaluated at $\tilde{\pi}$. Let $\Sigma_{\tilde{V}|Z} = E(\tilde{V}\tilde{V}'|Z_1, Z_2)$. Let $\theta = (\tilde{\pi}', \Phi')'$. We assume the following:

1. $(Y_i, X_i, Z_{1i}, Z_{2i})$ are i.i.d.
2. plim $(1/I)\ \tilde{Z}_1'\tilde{Z}_1 = \Sigma_{\tilde{Z}_1, \tilde{Z}_1}$, a positive definite matrix.
3. plim $(1/I)\ \tilde{Z}_1'(\Sigma_{\tilde{V}|Z})\tilde{Z}_1 = R$, a positive definite matrix.
4. plim $(1/I)\ Z(\tilde{\pi})'Z(\tilde{\pi})|_{\theta=\theta_0} = \Sigma_{Z,Z}$, a positive definite matrix.

In addition, we assume the other standard conditions invoked in GMM estimation (see Theorems 2.6 and 3.4 of Newey and McFadden, 1994), so that GMM and two stage estimation are consistent.

We define the two-stage estimator, $\tilde{\Phi}^{2SLS}$, as follows. π is estimated in the first stage as a seemingly unrelated regression. Let $\tilde{\pi}^{SUR}$ denote the resulting estimate of $\tilde{\pi}$. In the second stage, we estimate $\tilde{\Phi}^{2SLS}$ by an OLS regression of Y on $Z(\tilde{\pi}^{SUR})$.

The first stage regression can be rewritten as

(1) $\tilde{X} = \tilde{Z}_1\tilde{\pi} + \tilde{V}$

Let $\tilde{\pi}^{SUR}$ denote the SUR estimate of $\tilde{\pi}$. We immediately have

$$\sqrt{I}(\tilde{\pi}^{SUR} - \tilde{\pi}_0) \xrightarrow{d} N(0, \Sigma_{\tilde{Z}_1,\tilde{Z}_1}^{-1} R \, \Sigma_{\tilde{Z}_1,\tilde{Z}_1}^{-1}).$$

In the second stage, we estimate $\tilde{\Phi}^{2SLS}$ by an OLS regression of Y on $Z(\tilde{\pi}^{SUR})$. Let $\theta^{2SLS} = (\tilde{\pi}^{SUR'}, \tilde{\Phi}^{2SLS'})'$, and let $\theta_0 = (\tilde{\pi}_0', \tilde{\Phi}_0')'$. The $2SLS$ estimate of $\tilde{\Phi}$ evaluated at $\tilde{\pi}^{SUR'}$ is a GMM solution to:

(2) $(Z(\tilde{\pi})'Y - Z(\tilde{\pi})'Z(\tilde{\pi})\tilde{\Phi})|_{\tilde{\Phi}^{2SLS}} = 0.$

The proof of the consistency of $\tilde{\Phi}^{2SLS}$ is immediate. To derive the asymptotic distribution of $\tilde{\Phi}^{2SLS}$, we follow Newey and McFadden (1984) and obtain that $\sqrt{I}(\tilde{\Phi}^{2SLS} - \tilde{\Phi}_0)$ is asymptotically normal with feasible variance

$$\widehat{\mathrm{Var}}(\sqrt{I}(\tilde{\Phi}^{2SLS} - \tilde{\Phi}_0)) = \hat{G}_{\tilde{\Phi}}^{-1}\left[\frac{1}{I}\sum_{i=1}^{I}(\hat{g}_i + \hat{G}_{\tilde{\pi}}\psi_i)(\hat{g}_i + \hat{G}_{\tilde{\pi}}\psi_i)'\right]\hat{G}_{\tilde{\Phi}}^{-1}$$

where

$$\begin{aligned}
\psi_i &= [\psi_i^{1'}, \ldots, \psi_i^{K'}]' \\
\psi_i^{\ell} &= (Z_1'Z_1)^{-1}Z_{1i}'\hat{V}_i^{\ell} \\
\hat{G}_{\tilde{\pi}} &= \hat{A} + \hat{B} = \frac{\partial}{\partial\tilde{\pi}'}[Z(\tilde{\pi})'(Y - Z(\tilde{\pi}))\tilde{\Phi}] \\
\hat{g}_i &= Z_i'\varepsilon_i \\
\hat{G}_{\tilde{\Phi}} &= \hat{\Sigma}_{Z,Z}
\end{aligned}$$

with

$$\begin{aligned}
\hat{\varepsilon}_i &= Y_i - Z_i(\tilde{\pi}^{SUR})\tilde{\Phi}^{2SLS} \\
\hat{V}_i^{\ell} &= X_i^{\ell} - Z_{1i}'\pi_{\ell}^{SUR} \\
\hat{A} &= \frac{1}{I}\,\mathrm{diag}(\Sigma_i\, y_i Z_{2i}'Z_{1i}) = \frac{\partial[Z(\tilde{\pi})'Y]}{\partial\tilde{\pi}'}
\end{aligned}$$

a $KK_2 \times KK_1$ block diagonal matrix and $\hat{B} = (\partial[Z(\tilde{\pi})'Z(\tilde{\pi})\tilde{\Phi}]/\partial\tilde{\pi}')$ a $KK_2 \times KK_1$ matrix that can be partitioned into $K_2 \times K_1$ blocks with the (ℓ, m) block given as

$$\sum_{i=1}^{I}\frac{\partial(A_{i\ell}'A_{im}\Phi_m)}{\partial\pi_m'}$$

where $A_{i\ell} = \pi_\ell' Z_{1i}'Z_{2i}$. If $\ell \neq m$, the (ℓ, m) block is given by

$$\sum_{i=1}^{I}(A_{i\ell}'\Phi_m'Z_{2i}'Z_{1i})$$

If $\ell = m$, the (ℓ, m) block is given by

$$\sum_{i=1}^{I} [A'_{i\ell}\Phi'_{\ell}Z'_{2i}Z_{1i} + Z'_{2i}Z_{1i}A_{i\ell}\Phi_{\ell}].$$

A2. The Efficient GMM Estimator

We now stack the moment restrictions and estimate the parameters in a single step. Let $\rho(Z_{1i}, Z_{2i}, \theta_0) = (V_i, \varepsilon_i)'$. We have that

$$(4) \quad E(\rho(Z_{1i}, Z_{2i}, \theta_0) | Z_{1i}, Z_{2i}) = 0.$$

Let $\Upsilon(Z_{1i}, Z_{2i}) = E(\rho(Z_{1i}, Z_{2i}, \theta_0)\rho(Z_{1i}, Z_{2i}, \theta_0)' | Z_{1i}, Z_{2i})$, which is given by:

$$(5) \quad \Upsilon(Z_{1i}, Z_{2i}) = \begin{pmatrix} E(V'_i V_i | Z_{1i}, Z_{2i}) & E(V'_i \varepsilon_i | Z_{1i}, Z_{2i}) \\ E(\varepsilon_i V_i | Z_{1i}, Z_{2i}) & E(\varepsilon_i^2 | Z_{1i}, Z_{2i}) \end{pmatrix}.$$

Let $D(Z_{1i}, Z_{2i}) = E([\partial\rho(Z_{1i}, Z_{2i}, \theta_0)/\partial\theta'_0] | Z_{1i}, Z_{2i})$, a $(K + 1) \times (KK_2 + KK_1)$ matrix, so that we have

$$(6) \quad D(Z_{1i}, Z_{2i}) = \begin{pmatrix} -\Pi_i & 0 \\ -\zeta_i & -Z_i(\theta_0) \end{pmatrix}$$

where Π_i is a $K \times KK_1$ given as $\text{diag}(Z_{1i})$, and ζ_i is a KK_1 row vector given as $(\Phi'_1 Z'_{2i}Z_{1i}, \ldots, \Phi'_K Z'_{2i}Z_{1i})$. We thus have that the optimal instrument is $D(Z_{1i}, Z_{2i})'$ $\Upsilon(Z_{1i}, Z_{2i})^{-1}$ (see Hansen 1982, 1985). If we knew $D'(Z_{1i}, Z_{2i})\Upsilon(Z_{1i}, Z_{2i})^{-1}$, we could estimate θ by GMM:

$$(7) \quad \hat{\theta} = \arg\min \sum_{i=1}^{I} \rho(Z_{1i}, Z_{2i}, \theta)'\Upsilon(Z_{1i}, Z_{2i})^{-1}D(Z_{1i}, Z_{2i})$$
$$\times \left[\sum_{i=1}^{I} D(Z_{1i}, Z_{2i})'\Upsilon(Z_{1i}, Z_{2i})^{-1}D(Z_{1i}, Z_{2i})\right]^{-1}$$
$$\times \sum_{i=1}^{I} D(Z_{1i}, Z_{2i})'\Upsilon(Z_{1i}, Z_{2i})^{-1}\rho(Z_{1i}, Z_{2i}, \theta)$$

with the resulting $\hat{\theta}$ covariance matrix being $\{E[D(Z_{1i}, Z_{2i})'\Upsilon(Z_{1i}, Z_{2i})^{-1}D(Z_{1i}, Z_{2i})]\}^{-1}$. While we don't know $D(Z_{1i}, Z_{2i})'\Upsilon(Z_{1i}, Z_{2i})^{-1}$, we can consistently estimate it in a first step and plug the estimated values into the above GMM problem to obtain asymptotically efficient estimates under the stated conditions in Hansen (1982, 1985), or Newey and McFadden (1994).

References

Amemiya, T. 1985. *Advanced Econometrics.* Cambridge: Harvard University Press.

Becker, G. 1975. ''Human Capital and The Personal Distribution of Income: An Analytical Approach'' (W. S. Woytinsky Lecture). In *Human Capital*, second edition, New York: Columbia University Press.

Becker, G., and B. Chiswick. 1966. ''Education and The Distribution of Earnings.'' *American Economic Review* 56:358–69.

Card, D. 1995. ''Earnings, Ability and Schooling Revisited.'' In *Research In Labor Economics*, vol. 14, ed. S. Polachek. Greenwich, Conn.: JAI Press.

Card, D., and A. Krueger. 1992. ''Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States.'' *Journal of Political Economy* 100(1):1–40.

Chiswick, B. 1974. *Income Inequality: Regional Analyses Within A Human Capital Framework.* New York: Columbia University Press for NBER.

Chiswick, B., and J. Mincer. 1972. ''Time Series Changes In Personal Income Inequality.'' *Journal of Political Economy* 80(3):S34–S66.

Hansen, L. P. 1982. ''Large Sample Properties of Generalized Method of Moments Estimators.'' *Econometrica* 50(4):1029–54.

———. 1985, ''A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moment Estimators.'' *Journal of Econometrics* 30(1–2):203–38.

Heckman, J. 1982. ''The Empirical Content of Alternative Models of Labor Earnings,'' unpublished manuscript, University of Chicago, Department of Economics.

———. 1995. ''Notes on Earnings, Ability and Schooling Revisited,'' unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. 1996. ''Randomization as an Instrumental Variables Estimator.'' *Review of Economics and Statistics* 78(2):336–41.

———. 1997. ''Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations.'' *Journal of Human Resources* 32(3):441–62.

Heckman, J., and B. Honoré. 1990. ''The Empirical Content of the Roy Model.'' *Econometrica* 58(5):1121–49.

Heckman, J., A. Layne-Farrar, and P. Todd. 1996. ''Human Capital Pricing Equations with an Application To Estimating The Effect of Schooling Quality on Earnings.'' *Review of Economics and Statistics* 78(4):562–610.

Heckman, J., and J. Smith. 1997. ''Making The Most Out of Social Experiments: The Intrinsic Uncertainty in Evidence From Randomized Trials With An Application to the National JTPA Experiment.'' *Review of Economic Studies* 64:487–535.

———. 1998. ''Evaluating the Welfare State.'' Forthcoming in *Frisch Centennial*, ed. S. Strom. Econometric Society Monograph Series. Cambridge: Cambridge University Press.

Lewis, H. G. 1963. *Unionism and Relative Wages.* Chicago: University of Chicago Press.

Mincer, J. 1974. *Schooling Experience and Earnings.* New York, Columbia University Press for NBER.

Newey, W., and D. McFadden. 1994. ''Large Sample Estimation and Hypothesis Testing,'' in *Handbook of Econometrics*, Vol. IV, ed. R. F. Engle and D. L. McFadden, Amsterdam, North Holland.

Rosen, S. 1977. ''Human Capital: A Survey of Empirical Research.'' In *Research in Labor Economics*, vol. 1, ed. R. Ehrenberg. Greenwich, Conn.: JAI Press.

Wicksell, K. 1934. *Lectures on Political Economy: Volume 1. General Theory*, translated by E. Classen. London: Rutledge and Kegan Paul, Ltd.

Note Added in Proof

After this paper was submitted, refereed and accepted, David Card pointed out to us a related paper by Jeff Wooldridge (1997) who comments on an earlier paper by John Garen (1984). Wooldridge considers a special case of the model presented here in which, in our notation, there is only one correlated random coefficient, our A1 and A2 holds but the conditional mean of the error is linear in V_i: $E(U_i|Z_{1i}, Z_{2i}, V_i) = kV_i$ where U_i and V_i are scalar. As a consequence of this assumption, instead of our conditional covariance homoscedasticity Assumption A3, he invokes a conditional variance homoscedasticity assumption: $E(V_i^2|Z_{1i}, Z_{2i}) = E(V_i^2)$ which is not required in our approach. However, the calculations of the standard errors simplify in this case. An alternative set of assumptions presented by Wooldridge exploits homoscedasticity for U_i: $E(U_i^2|Z_{1i}, Z_{2i}) = E(U_i^2)$. His linearity of conditional expectations and homoscedasticity assumptions are not required for the model presented in this paper. Like Heckman (1982, 1995), Wooldridge demonstrates under his stated conditions it is possible to identify the average treatment effect slope coefficients without identifying the model intercept.

Additional References

Garen, J. 1984. "The Returns to Schooling: A Selectivity Bias Approach With A Continuous Choice Variable," *Econometrica*, 52:1199–1218.

Wooldridge, J. 1997. "On Two Stage Least Squares Estimation of the Average Treatment Effect in A Random Coefficient Model," *Economics Letters* 56:129–133.