Lecture Note 4: Willis, R. and S. Rosen (1979), "Education and Self-Selection," Journal of Political Economy, 87(5), part 2, S7-S36 ——Human Capital*

WanChuan Lin

10/25/2007

In this paper schooling choices is similar in spirit to the Roy model. The main empirical goal is to examine the effect of earnings expectation on educational choices. Specifically, paper's objective are:

(1). Estimate life time earning conditional on school choice, purged of selectivity bias; and

(2). Investigate how alternative earning prospect affect school choice (beyond the effect of family background and financial constraints).

The paper deals with a host of selectivity effects, sorting effects (related to "ability bias"), family effects and tastes.

There is an implicit assumption most of the literature on education that a person who is better at some activity, say c, is also better at any other activity h. Here there is no one measure of talent. Consequently, those who end up being lawyers, because this is what they are best at doing, might be very lousy plumbers.

1 Model

To keep things simple we have here two schooling choices:

*Based on Willis and Rosen (1978), "Education and Self-Selection"

A. attend college; or

B. not attend college.

Notation:

Y_{ij}: potential life time earnings for person i with schooling level j.

X_i: observed vector of talent and ability indicators.

τ_i: unobserved talent.

Z_i: observed family background variables.

ω_i: unobserved family components.

The potential earnings of person i with school level j is assumed to be a function of talent or ability. Hence we have that

$$Y_{ij} = y_j \left(X_i, \tau_i \right) \tag{1}$$

The present value (or value function) of the earning stream from choice j for i is assumed to depend on not only the earnings but also the family-background and taste variables.

$$V_{ij} = g \left(y_j, Z_i, \omega_i \right) \tag{2}$$

This kind of formulation is made to reflect the financial barriers and tastes to extending schooling. When there are n levels of schooling to choose, an individual will choose the one j that can give him the biggest present value. Therefore, we have that

$$V_{ij} = \max \left\{ V_{i1}, ..., V_{ij}, ..., V_{iJ} \right\}$$

Here, we should note that i's family background affect the potential life time earnings indirectly through impacting educational level not directly.

And the distribution of the unobservable are assumed as follows.

$$\left(\tau, \omega \right) \sim F \left(\tau, \omega \right)$$

and normality. Since we do not observe either τ or ω, we cannot estimate F, so we leave it unspecified.

Let

V_{ci} be the present value of earnings as a college graduate; and

V_{hi} be the present value of earnings as a high school graduate.

Note that c corresponds to *college*, while h corresponds to *high school.*

Hence, one would go to college if

$$V_{ci} > V_{hi}$$

Assume exponential growth in earnings.

For college graduates we have

$$W_{cit} = \begin{cases} 0 & \text{if } t \leq S, \\ \bar{y}_{ci} \exp\left\{g_{ci}\left(t - S\right)\right\} & \text{if } t > S, \end{cases}$$

where the notation we use here is as follows:

- S is number of years it takes to get a college degree;
- t represents age (measured as years since hight school graduation), so the $t - S$ is market experience for i choosing h;
- i is an individual;
- $\bar{y}_{ci}$ denotes the initial wage; and
- g_{ci} denotes the growth rate of wage.

For high school graduates we have:

$$W_{hit} = \bar{y}_{hi} \exp\left\{g_{hi} t\right\}$$

Assume an infinite horizon, a constant rate of discount for each person i, r_i, with $r_i > g_{ai}, g_{bi}$.

Thus, for college graduates we have

$$\begin{aligned} V_{ci} &= \int_0^\infty e^{-r_i t} W_{cit} dt \\ &= \int_S^\infty e^{-r_i t} \bar{y}_{ci} e^{g_{ci}(t-S)} dt \\ &= e^{-r_i S} \bar{y}_{ci} \int_S^\infty e^{(g_{ci} - r_i)t} dt \\ &= \frac{\overline{y}_{ci}}{r_i - g_{ci}} e^{-g_{ci} S} e^{-(r_i - g_{ci})S} \\ &= \frac{\overline{y}_{ci}}{r_i - g_{ci}} e^{-r_i s} \end{aligned}$$

while for hight school graduates we have

$$\begin{aligned} V_{hi} &= \int_0^\infty e^{-r_i t}\bar{y}_{hi}e^{g_{hi}t}dt \\ &= \int_0^\infty e^{-r_i t}\bar{y}_{hi}e^{g_{hi}t}dt \\ &= \bar{y}_{hi}\int_0^\infty e^{-(r_i - g_{bi})t}dt \\ &= \frac{e^{-r_i s}\bar{y}_{hi}}{r_i - g_{hi}} \end{aligned}$$

Implication:

(1). As $\bar{y}_{hi}$ increases (holding all other parameters constant), the probability of going to college increases.

(2). As the growth rate of college graduate (g_{ci}) increase, college attendance increases.

(3). As r_i, the individual's interest rate increases, college attendance decreases.

In this paper the authors allow for heterogeneity in several parameters:

- interest rate, that is, r_i;
- Initial wages$(\bar{y}_{ci}, \bar{y}_{hi})$; and
- Wage growth rates(g_{ci}, g_{hi})

1.1 Selection rule

As indicated above, an individual will choose to go to college if

$$V_{ci} > V_{hi}$$

and an individual will choose not to go to college if

$$V_{hi} \geq V_{ci}$$

Let now

$$I_i = \log(V_{ci}/V_{hi})$$

Then we can rewrite this decision rule as: an individual will choose to attend college if

$$\begin{aligned} I_i &= \log V_{ci} - \log V_{hi} \\ &= \log(\bar{y}_{ci}) - \log(\bar{y}_{hi}) - r_i s - \log(r_i - g_{ci}) + \log(r_i - g_{hi}) > 0 \end{aligned}$$

They "cheat" at this point and make a simplifying assumption that (by Taylor expansion of the non-linear terms around their means, i.e., $(\bar{g}_c, \bar{g}_h, \bar{r})$ [1]) we get

$$
\begin{aligned}
I_i &= \log(V_{ci}) - \log(V_{hi}) \\
&\approx \log(\bar{y}_{ai}) - \log(\bar{y}_{bi}) + [-\bar{r}S - \log(\bar{r} - \bar{g}_a) + \log(\bar{r} - \bar{g}_b)] \\
&\quad + \frac{\partial I}{\partial g_a}(g_{ai} - \bar{g}_a) + \frac{\partial I}{\partial g_b}(g_{bi} - \bar{g}_b) + \frac{\partial I}{\partial r}(r_i - \bar{r}) \\
&= \alpha_0 + \alpha_1[\log(\bar{y}_{ci}) - \log(\bar{y}_{hi})] + \alpha_2 g_{ci} + \alpha_3 g_{hi} + \alpha_4 r_i \qquad (3)
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha_0 &= \ln\left(\frac{\bar{r} - \bar{g}_h}{\bar{r} - \bar{g}_c}\right) \\
\alpha_1 &= 1, \\
\alpha_2 &= \partial I/\partial g_c = 1/(\bar{r} - \bar{g}_c) > 0, \\
\alpha_3 &= \partial I/\partial g_h = -1/(\bar{r} - \bar{g}_h) < 0, \text{ and} \\
\alpha_4 &= -\left[s + \frac{(\bar{g}_c - \bar{g}_h)}{\bar{r} - \bar{g}_c)(\bar{r} - \bar{g}_h)}\right]
\end{aligned}
$$

Since there is personal heterogeneity, we can view V_a and V_b as random variables and V_{ai} and V_{bi} are realizations of them respectively.

$$
\begin{aligned}
\Pr(\text{choose } h) &= \Pr(V_a > V_b) = \Pr(I > 0), \qquad (4) \\
\Pr(\text{choose } c) &= \Pr(V_a \leq V_b) = \Pr(I \leq 0).
\end{aligned}
$$

Follow the nature of this problem put at the beginning of this paper, the potential earnings of a person of some schooling level is a function of ability and talents, whether observable or unobservable. So we can assume the structural earnings equation in the following way.

$$
\begin{aligned}
\log(\bar{y}_{ci}) &= X_i^{'}\beta_c + u_{1i}, \qquad (5) \\
g_{ci} &= X_i^{'}\gamma_c + u_{2i},
\end{aligned}
$$

if college *is* chosen, and

$$
\begin{aligned}
\log(\bar{y}_{hi}) &= X_i^{'}\beta_h + u_{3i}, \qquad (6) \\
g_{hi} &= X_i^{'}\gamma_h + u_{4i},
\end{aligned}
$$

if college is not chosen.

---

[1] Taylor expransion to linearization: if function $f(\cdot)$ has two dependent variables, for example, the Taylor expransion around some point $(\overline{x}, \overline{y})$ is

$$f(x, y) \approx f(\overline{x}, \overline{y}) + (x - \overline{x}) f_1 + (y - \overline{y}) f_2$$

where $f_1 = \frac{\partial f(x,y)}{\partial x}|_{(\overline{x},\overline{y})}$, and $f_2 = \frac{\partial f(x,y)}{\partial y}|_{(\overline{x},\overline{y})}$

If h is not chosen, where X_i is a set of measured characteristics that influence a person's lifetime earnings potentials, and u_{1i}, u_{2i}, u_{3i}, and u_{4i} denote permanent person-specific unobserved components reflecting unmeasured factors influencing earnings potential.

Also, assume that

$$r_i = Z_i^{'}\delta + u_{5i}, \tag{7}$$

where Z_i denote another vector of observed variables that influence the schooling decision through their effect on the individual's discount rate r_i, and u_{5i} is a permanent unobserved component influencing financial barriers to school choice.

family background $\rightarrow$ (r_i) schooling level $\rightarrow$ potential life time earnings

The above three formulae $(5), (6)$ and (7), specify the general form of the model given by (1) and (2).

We also assume that the joint distribution of the error terms is normal, i.e.,

$$u_i = (u_{1i,}u_{2i,}u_{3i,}u_{4i,}u_{5i})' \sim N(0, \Sigma)$$

(3),$(5), (6)$ and (7) give the structural model.

Then by substituting $(5) - (7)$ into(3), we have the reduced form below.

$$\begin{aligned} I_i &= \log(V_{ci}) - \log(V_{hi}) \\ &\approx \alpha_0 + X_i^{'}\left[\alpha_1\left(\beta_c - \beta_h\right) + \alpha_2\gamma_c + \alpha_3\gamma_h\right] + \alpha_4 Z_i^{'}\delta \\ &\quad +\alpha_1\left(u_{1i} - u_{3i}\right) + \alpha_2 u_{2i} + \alpha_3 u_{4i} + \alpha_5 u_{5i} \\ &= W_i^{'}\pi - \epsilon_i \end{aligned}$$

where

$$\begin{aligned} W_i' &= (1, X_i, Z_i), \\ -\varepsilon_i &= \alpha_1\left(u_{1i} - u_{3i}\right) + \alpha_2 u_{2i} + \alpha_3 u_{4i} + \alpha_5 u_{5i} \end{aligned}$$

and let σ_ε be the standard deviation of ε_i. Then, the reduced form of probability of choosing h is

$$\begin{aligned} \Pr(\text{choose } h) &= \Pr(I > 0) \\ &= \Pr(W\pi > \epsilon) \\ &= \Pr\left(\frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right) \\ &= F\left(\frac{W\pi}{\sigma_\varepsilon}\right), \end{aligned} \tag{8}$$

where F is the cdf of $\frac{\epsilon}{\sigma_\varepsilon}$. Since $(u_{1,}u_{2,}u_{3,}u_{4,}u_5)$ is jointly normal with zero mean and $\epsilon = -\left[\alpha_1\left(u_1 - u_3\right) + \alpha_2 u_2 + \alpha_3 u_4 + \alpha_5 u_5\right]$, we have that ϵ is normal with zero mean. Hence, $\frac{\epsilon}{\sigma_\varepsilon}$ is standard normal. So F is the standard normal cdf.

1.2 Selection bias

Now we solve for the selection bias.

$$\begin{aligned} & E\left[\ln\left(\bar{y}_c\right)|I>0\right] \\ = \quad & X\beta_c + E\left[u_1|I>0\right] \\ = \quad & X\beta_c + E\left[u_1|\frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right]. \end{aligned}$$

Also,

$$u_1 = \frac{\sigma_{1\epsilon}}{\sigma_\epsilon^2}\epsilon + \xi,$$

where ξ is normal with zero mean and independent of ϵ. Hence,

$$\begin{aligned} E\left[u_1|\frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] &= E\left[\frac{\sigma_{1\epsilon}}{\sigma_\epsilon^2}\epsilon + \xi|\frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] \\ &= E\left[\frac{\sigma_{1\epsilon}}{\sigma_\epsilon^2}\epsilon|\frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] + E\left[\xi|\frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] \\ &= E\left[\frac{\sigma_{1\epsilon}}{\sigma_\epsilon^2}\epsilon|\frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] + E\left[\xi\right] \\ &= \frac{\sigma_{1\epsilon}}{\sigma_\epsilon}E\left[\frac{\epsilon}{\sigma_\varepsilon}|\frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] + 0. \end{aligned}$$

Since $\frac{\epsilon}{\sigma_\varepsilon}$ is standard normal, we have that

$$E\left[\frac{\epsilon}{\sigma_\varepsilon}|\frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] = -f\left(\frac{W\pi}{\sigma_\varepsilon}\right)/F\left(\frac{W\pi}{\sigma_\varepsilon}\right) \doteq \lambda_a\left(\frac{W\pi}{\sigma_\varepsilon}\right),$$

where f is the pdf of the standard normal distribution[2].

[2] Theorem for Moments of the Truncated Normal Distribution:
If $x \sim N\left[\mu, \sigma^2\right]$ and a is a constant, then

$$\begin{aligned} E\left[x|\text{truncation}\right] &= \mu + \sigma\lambda\left(\alpha\right) \\ Var\left[x|\text{truncation}\right] &= \sigma^2\left[1 - \delta\left(\alpha\right)\right] \end{aligned}$$

where $\alpha = \left(a - \mu\right)/\sigma$, $\phi\left(\alpha\right)$ is the standard normal density and

$$\begin{aligned} \lambda\left(\alpha\right) &= \phi\left(\alpha\right)/\left[1 - \Phi\left(\alpha\right)\right] \quad \text{if truncation is } x > a \\ \lambda\left(\alpha\right) &= -\phi\left(\alpha\right)/\Phi\left(\alpha\right) \quad \text{if truncation is } x < a \end{aligned}$$

and

$$\delta\left(\alpha\right) = \lambda\left(\alpha\right)\left[\lambda\left(\alpha\right) - \alpha\right]$$

(Ref: Theorem 22.2, on Page 759, of Green's book, Fifth Edition.)

Therefore,

$$\begin{aligned}
E\left[\ln\left(\bar{y}_c\right)|I>0\right] &= X\beta_c + \frac{\sigma_{1\epsilon}}{\sigma_\epsilon} E\left[\frac{\epsilon}{\sigma_\varepsilon} | \frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] \\
&= X\beta_c + \frac{\sigma_{1\epsilon}}{\sigma_\epsilon}\left(-f\left(\frac{W\pi}{\sigma_\varepsilon}\right)/F\left(\frac{W\pi}{\sigma_\varepsilon}\right)\right) \\
&= X\beta_c + \sigma_1\rho_1\lambda_c,
\end{aligned}$$

where

$$\begin{aligned}
\rho_1 &= \rho\left(\frac{u_1}{\sigma_1}, \frac{\epsilon}{\sigma_\varepsilon}\right) = \frac{\sigma_{1\varepsilon}}{\sigma_1\sigma_\varepsilon}\text{, and} \\
\lambda_c &= -\frac{f\left(\frac{W\pi}{\sigma_\varepsilon}\right)}{F\left(\frac{W\pi}{\sigma_\varepsilon}\right)}.
\end{aligned}$$

The above formula can also be written as

$$E\left[\ln\left(\bar{y}_c\right)|I>0\right] = X\beta_c + \frac{\sigma_{1\epsilon}}{\sigma_\epsilon}\lambda_c. \tag{9}$$

Similarly, for the individuals not choosing college,

$$E\left[\ln\left(\bar{y}_h\right)|I\leq 0\right] = X\beta_h + \frac{\sigma_{3\epsilon}}{\sigma_\epsilon}\lambda_h. \tag{10}$$

where $\lambda_h = \frac{f\left(\frac{W\pi}{\sigma_\varepsilon}\right)}{1-F\left(\frac{W\pi}{\sigma_\varepsilon}\right)}$.

For the growth rate of the wages of people choosing college or not, we have that

$$\begin{aligned}
E\left[g_c|I>0\right] &= E\left[X\gamma_c + u_2|I>0\right] \\
&= X\gamma_c + E\left[u_2|I>0\right] \\
&= X\gamma_c + E\left[\frac{\sigma_{2\epsilon}}{\sigma_\epsilon^2}\epsilon + \eta | \frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] \\
&= X\gamma_c + \frac{\sigma_{2\epsilon}}{\sigma_\epsilon} E\left[\frac{\epsilon}{\sigma_\varepsilon} | \frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] + E\left[\eta | \frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] \\
&= X\gamma_c + \frac{\sigma_{2\epsilon}}{\sigma_\epsilon} E\left[\frac{\epsilon}{\sigma_\varepsilon} | \frac{W\pi}{\sigma_\varepsilon} > \frac{\epsilon}{\sigma_\varepsilon}\right] + E\left[\eta\right] \\
&= X\gamma_c + \frac{\sigma_{2\epsilon}}{\sigma_\epsilon}\lambda_c. \qquad (11)
\end{aligned}$$

where η is normal with zero mean and independent of ϵ, and $\sigma_{2\epsilon} = cov\left(u_2, \epsilon\right)$.

And

$$\begin{aligned} E\left[g_h|I\leq 0\right] &= E\left[X\gamma_h+u_4|I\leq 0\right] \\ &= X\gamma_h+E\left[u_4|\frac{W\pi}{\sigma_\varepsilon}\leq\frac{\varepsilon}{\sigma_\varepsilon}\right] \\ &= X\gamma_h+E\left[\frac{\sigma_{4\epsilon}}{\sigma_\epsilon^2}\epsilon+\omega|\frac{W\pi}{\sigma_\varepsilon}\leq\frac{\varepsilon}{\sigma_\varepsilon}\right] \\ &= X\gamma_h+\frac{\sigma_{4\epsilon}}{\sigma_\epsilon}E\left[\frac{\varepsilon}{\sigma_\varepsilon}|\frac{W\pi}{\sigma_\varepsilon}\leq\frac{\varepsilon}{\sigma_\varepsilon}\right]+E\left[\omega|\frac{W\pi}{\sigma_\varepsilon}\leq\frac{\varepsilon}{\sigma_\varepsilon}\right] \\ &= X\gamma_h+\frac{\sigma_{4\epsilon}}{\sigma_\epsilon}E\left[\frac{\varepsilon}{\sigma_\varepsilon}|\frac{W\pi}{\sigma_\varepsilon}\leq\frac{\varepsilon}{\sigma_\varepsilon}\right]+E\left[\omega\right] \\ &= X\gamma_h+\frac{\sigma_{4\epsilon}}{\sigma_\epsilon}\lambda_h, \end{aligned} \tag{12}$$

and

$$\begin{aligned} \sigma_{k\epsilon} &= Cov\left(u_k,\epsilon\right) \\ &= Cov\left(u_k,-\left[\alpha_1\left(u_1-u_3\right)+\alpha_2u_2+\alpha_3u_4+\alpha_5u_5\right]\right) \\ &= -\alpha_1Cov\left(u_k,u_1-u_3\right)-\alpha_2Cov\left(u_k,u_2\right)-\alpha_3Cov\left(u_k,u_4\right)-\alpha_5Cov\left(u_k,u_5\right) \\ &= -\left[\alpha_1\left(\sigma_{1k}-\sigma_{3k}\right)+\alpha_2\sigma_{2k}+\alpha_3\sigma_{4k}+\alpha_5\sigma_{5k}\right], \end{aligned}$$

for $k=1,2,3,4$.

Obviously, λ_c is always negative; λ_h is always positive. So, positive selection bias if $\frac{\sigma_{j\varepsilon}}{\sigma_\varepsilon}<0$, $j=1,2$ and $\frac{\sigma_{j\varepsilon}}{\sigma_\varepsilon}>0$, $j=3,4$. Positive bias in both c and h implies comparative advantage.

2 Estimations

Now we can estimate this model in three steps.

2.1 Step 1

In this step we get reduced form probit of college attendance on W_i. Note that

$$\begin{aligned} \Pr\left(College|W_i\right) &= \Pr\left(W_i^{'}\pi>\varepsilon_i\right) \\ &= \Pr\left(\frac{\epsilon_i}{\sigma_\epsilon}<\frac{W_i\pi}{\sigma_\epsilon}\right) \\ &= F\left(\frac{W_i\pi}{\sigma_\epsilon}\right) \end{aligned}$$

That is we have a simple probit model for the choice of college attendance. This gives estimate $\widehat{\pi/\sigma_\varepsilon}$ of π/σ_ε.

2.2 Step 2

Assume that we have data on initial wages and wages T periods later. We can estimate earnings equations using Heckman's two step procedure.

Corresponding to (9) , (10) , (11) and (12), we can consider the regression applied to observed data , that is

$$\begin{aligned}
\log \bar{y}_c &= X'\beta_c + \beta_c^*\lambda_c + \eta_1 \\
g_c &= X'\gamma_c + \gamma_c^*\lambda_c + \eta_2 \\
\log \bar{y}_h &= X'\beta_h + \beta_h^*\lambda_h + \eta_3 \\
g_h &= X'\gamma_h + \gamma_h^*\lambda_h + \eta_4
\end{aligned} \tag{13}$$

Note that by the above specification β_c^* estimates $\sigma_{1\varepsilon}/\sigma_\varepsilon$, etc.

π/σ_ε is unknown, but we can form estimates $\hat{\lambda}_{ci}$ and $\hat{\lambda}_{hi}$ for λ_{ci} and λ_{hi} using the estimates $\widehat{\pi/\sigma_\varepsilon}$ for π/σ_ε from step 1.

Now we can run the regressions in (13) using the estimates $\hat{\lambda}_{ci}$ and $\hat{\lambda}_{hi}$ in place of λ_{ci} and λ_{hi} respectively.

2.3 Step 3

Finally, we go back to the "Structural probit", that is,

$$\begin{aligned}
\Pr(\text{College}|W_i) &= \Pr(\log(V_{ci}) - \log(V_{hi}) > 0|W_i) \\
&= \Pr\left(\frac{\alpha_0 + \alpha_1[\log(\bar{y}_{ci}) - \log(\bar{y}_{hi})] + \alpha_2 g_{ci} + \alpha_3 g_{hi} + \alpha_4 Z_i'\delta}{\sigma_\epsilon} > \frac{\varepsilon_i}{\sigma_\epsilon}\right) \\
&= \Pr\left(\frac{\alpha_0 + \alpha_1\left[X_i'\beta_c - X_i'\beta_h\right] + \alpha_2 X_i'\gamma_c + \alpha_3 X_i'\gamma_h + \alpha_4 Z_i'\delta}{\sigma_\epsilon} > \frac{\varepsilon_i}{\sigma_\epsilon}\right)
\end{aligned}$$

Clearly, one can see that there is a scale problem. That is, we can always multiply by any positive constant without changing the sign. And σ_ε is the the only parameter we don't know now. A common normalization is $\text{Var}(\epsilon_i) = 1$.

Then one can get consistent estimates of $\alpha_1, \alpha_2, \alpha_3$,and α_4 by running a probit of college attendance on $\left[X_i'\hat{\beta}_c - X_i'\hat{\beta}_h\right], X_i'\hat{\gamma}_c, X_i'\hat{\gamma}_h$, and $Z_i'\delta$.

Note that the standard errors must be corrected to account for the fact that one of the regressor is only an estimate of the true regressor.