Graduate Labor Lecture Notes #13 on Human Capital and Earnings Functions

ECON 792

October 29, 2007

Rosen and Willis (1979) examines the role of earnings expectations and family background in the schooling decision - specially, in the decision to go to college. They develop and estimate a model of the demand for schooling that takes accounts for earnings expectations and for heterogeneity in ability levels, in tastes, and in capacity to finance schooling investments. The model assumes that high school or college education prepare one for one of two occupational sectors, where schooling level and occupational sector are linked. As in the Roy (1951) and Heckman and Sedlacek (1985) models, their model imbeds the notion of latent abilities – that individuals may have talents that are not directly applied on their job. The model also allows for the possibility of comparative advantage.

Suppose we estimate the return to education by an earnings equation:

$$\ln y_i = \alpha + \beta ed_i + \gamma \exp_i + \gamma_2 \exp_i^2 + \varepsilon_i.$$

As noted in lecture #12, estimates of the return to education are potentially upward biased because of omitted ability. The R^2 values in earnings regressions usually fairly low (around 0.25) suggesting that there are lots of factors affecting earnings that are not explained by the model.

0.1 Model

Define the following notation:

$Y_{ij}-$ potential lifetime earnings in person i chooses education level j

X_i- observed talent or ability measures that affect the marginal return to schooling investment

τ_i- unobserved talent component

Z_i- observed family background and taste effects (includes measures of financial barriers) that affect the cost of funding schooling investments

ω_i- unobserved component of family background or taste

Also, let V_{ij} denote the value of schooling level j to person i. Utility is assumed to be a function of earnings, of taste shifters and of financial costs to schooling.

$$V_{ij} = g(Y_j(X_i, \tau_i), Z_i, \omega_i)$$

Individuals choose the schooling level that maximizes their lifetime utility:

$$max(V_{i1}, ..., V_{iJ}),$$

and the empirical work assumes two schooling levels: high school and college attendance. Assume a distribution on the unobservables:

$$(\tau, \omega)\tilde{}F(\tau, \omega).$$

Also, Willis and Rosen (1979) observe earnings only at two points in time, once just after entering the market and then again 20 years later. Their paper uses the NBER-Thorndike sample comprised of male WWII veterens who applied for the Army air corps). The econometric model is in part tailored to the data.

If person chooses to attend college (option A), earnings are given by:

$$\begin{aligned} y_{ai}(t) &= 0 \quad \text{for } 0 \leq t \leq s \\ &= \bar{y}_{ai} \exp(g_{ai}(t-s)) \quad \text{for s<t<}\infty \end{aligned}$$

where g_{ai} is the growth rate in earnings for option A and $t - s$ is experience (which assumes the individual works in every period after leaving school and ignores any direct costs of schooling).

If person chooses highschool (option B), get

$$y_{bi}(t) = \bar{y}_{bi} \exp(g_{bi}(t-s)) \text{ for 0<t<}\infty.$$

Assume that $(\bar{y}_{ai}, g_{ai}, \bar{y}_{bi}, g_{bi}, r_i)$ are randomly distributed in the population, where r_i is person i's discount rate.

$$\begin{aligned} V_{ai} &= \int_0^\infty y_{ai}(t) e^{-r_i t} dt \\ &= \int_s^\infty \bar{y}_{ai} e^{g_{ai} t - g_{ai} s - r_i t} dt \\ &= \bar{y}_{ai} e^{-g_{ai} s} \int_s^\infty e^{(g_{ai} - r_i)t} dt \\ &= \bar{y}_{ai} e^{-g_{ai} s} \left(\frac{e^{(g_{ai} - r_i)t}}{g_{ai} - r_i} \right) |_s^\infty \end{aligned}$$

Assume $g_{ai} - r_i < 0$, then

$$\begin{aligned} &= \bar{y}_{ai} e^{-g_{ai} s} \frac{e^{(g_{ai} - r_i)s}}{r_i - g_{ai}} \\ &= \frac{\bar{y}_{ai} e^{-r_i s}}{r_i - g_{ai}} \end{aligned}$$

Similarly, obtain

$$V_{bi} = \frac{\bar{y}_{bi}}{r_i - g_{bi}}$$

Define and index that describes the log of the relative lifetime utility of choosing schooling level A over B:

$$\begin{aligned} I_i &= \ln(V_{ai}/V_{bi}) = \ln(\frac{\bar{y}_{ai}}{\bar{y}_{bi}} \frac{r_i - g_{bi}}{r_i - g_{ai}} e^{-s r_i}) \\ &= \ln \bar{y}_{ai} - \ln \bar{y}_{bi} + \ln(r_i - g_{bi}) - \ln(r_i - g_{ai}) - s r_i \end{aligned}$$

Taylor expand the nonlinear terms around population means to get:

$$\ln(r_i - g_{ai}) = \bar{r}_i - \bar{g}_{ai} + \frac{1}{\bar{r}_i - \bar{g}_{ai}}(r_{ai} - \bar{r}_i) - \frac{1}{\bar{r}_i - \bar{g}_{ai}}(g_{ai} - \bar{g}_i)$$

Substitute into the expression for I to get

$$I_i = \alpha_0 + \alpha_1(\ln \bar{y}_{ai} - \ln \bar{y}_{bi}) + \alpha_2 g_{ai} + \alpha_3 g_{bi} + \alpha_4 r_i \qquad (*)$$

Next specify mean log earnings and mean growth rates as functions of

$$\begin{aligned}
\ln \bar{y}_{ai} &= X_i \beta_a + u_{1i} \\
g_{ai} &= X \gamma_a + u_{2i} \\
\ln \bar{y}_{bi} &= X_i \beta_b + u_{3i} \\
g_{bi} &= X \gamma_b + u_{4i} \\
r_i &= Z_i \delta + u_{5i}
\end{aligned}$$

where the variables on the LFS $(\bar{y}_{ai}, \bar{y}_{bi}, g_{ai}, g_{bi})$ are to be interpreted as the individual's expectation about future earnings at the time the decision to attend college is made.

Family background and taste effects are assumed to influence the schooling decision through the discount rate (Z). Substitute all the equations into (*) to get an estimating equation

$$\Pr(chooseA) = \Pr(\frac{\alpha_0 + \alpha_1(\ln \bar{y}_{ai} - \ln \bar{y}_{bi}) + \alpha_2 g_a + \alpha_3 g_b + \alpha_4 Z\delta}{\sigma} > \frac{u_{5i}}{\sigma})$$

or get reduced form equation

$$\begin{aligned}
I_i &= \alpha_0 + \alpha_1 X_i[(\beta_a - \beta_b) + \gamma_a - \gamma_b] + \alpha_4 Z\delta + \alpha_1(u_{1i} - u_3) + \alpha_2 u_2 + \alpha_3 u_4 + \alpha_4 u_5 \\
&= -W\pi - \varepsilon
\end{aligned}$$

with $W = [X, Z]$ and -$\varepsilon = \alpha_1(u_{1i} - u_3) + \alpha_2 u_2 + \alpha_3 u_4 + \alpha_4 u_5$.

$$\Pr(choose\ A|W) = \Pr(W\pi > \varepsilon) = F(\frac{W\pi}{\varepsilon_\varepsilon})$$

where F( ) is the standard normal cdf. The above equation is a probit function determining sample selection into college attendance (option A) and highschool only (option B). Note that the effect of Z elements that are also contained in X cannot be separately identified.

0.2 Estimation

The structural parameters of the model can be estimated in three steps. The earnings equations and the growth rate equations are interpreted as expected earnings and growth rates at the time of making the schooling decision, where it is assumed that expectations are unbiased with mean zero forecast errors that are iid.

In the first step, estimate the previously described reduced form probit model. In the second step, estimate mean earnings using the control function to control for sample selectivity:

$$\begin{aligned}
E(\ln y_{ai}|I &> 0) = X\beta_a + E(u_{1i}|I > 0) \\
&= X\beta_a + \frac{cov(u_{1i},\varepsilon)}{var(\varepsilon)} E(\frac{\varepsilon}{\sigma_e}|\frac{\varepsilon}{\sigma_e} > \frac{-W_i\pi}{\sigma_\varepsilon}) \\
&= X\beta_a + \frac{cov(u_{1i},\varepsilon)}{var(\varepsilon)} \frac{\phi(\frac{-W_i\pi}{\sigma_\varepsilon})}{1-\Phi(\frac{-W_i\pi}{\sigma_\varepsilon})}
\end{aligned}$$

Use a similar control function approach to estimate the parameters of

$$\begin{aligned}
g_a &= X\gamma_a + u \\
\ln y_b &= X\beta_b + u_3 \\
g_b &= X\gamma_b + u,
\end{aligned}$$

where g_a and g_b are individual-specific observed growth rates. This estimation will yield estimates of $\hat{\beta}_a, \hat{\beta}_b, \hat{\gamma}_a, \hat{\gamma}_b$.

In the third step, form the predicted values of $\ln(\frac{\bar{y}_{ai}}{\bar{y}_{bi}}), \hat{g}_{ai}, \hat{g}_{bi}$ for each person and estimate the structural probit model to get

$$\frac{\alpha_1}{\sigma_\varepsilon}, \frac{\alpha_2}{\sigma_\varepsilon}, \frac{\alpha_3}{\sigma_\varepsilon}, \frac{\alpha_4\delta}{\sigma_\varepsilon} \text{ and } \frac{1}{\sigma_\varepsilon}$$

The estimator for $\frac{1}{\sigma_\varepsilon}$ comes from the estimated coefficient associated with $\ln(\frac{\bar{y}_{ai}}{\bar{y}_{bi}})$.

The model could also be specified using the observed level of earnings at time $\bar{t}$ instead of initial earnings. The earnings equation in that case is:

$$\begin{aligned}
\ln y_a(\bar{t}) &= X(\beta_a + \gamma_a\bar{t}) + u_1 + tu_2 \\
\ln y_b(\bar{t}) &= X(\beta_b + \gamma_b\bar{t}) + u_3 + tu_4
\end{aligned}$$

Rosen and Willis (1979) estimate the model both ways as a way of checking the validity of the model.

0.3 Empirical Results

The model was estimated on 3,611 respondents to the NBER-Thorndike-Hagen survey of 1968-71. The sample consists of WWII veterans who applied for the army air corps. The dataset contains extensive information on family background and ability.

Table 1 contains the descriptive statistics. More than 75% attended some college, which is attributable in part to the GI bill. Mean and variance of earnings is smaller for high school group. Earnings growth also smaller for high school group. HS group tends to score lower in math and reading but higher in mechanical ability. The HS group also tends to come from larger families and to have higher birth order.

Variable	High School (Group B) Mean	SD	More than High School (Group A) Mean	SD
Father's ED	8.671	2.966	10.26	3.623
Father's ED^2	83.99	55.53	118.4	78.09
DK ED	.0999	...	.0464	...
Manager	.3628	...	.4954	...
Clerk	.1239	...	.1450	...
Foreman	.2238	...	.1695	...
Unskilled	.1492	...	.0819	...
Farmer	.1062	...	.0720	...
DK job	.0177	...	.0124	...
Catholic	.2933	...	.2138	...
Jew	.0405	...	.0617	...
Old sibs	1.143	1.634	.9035	1.383
Young sibs	.9381	1.486	.8138	1.266
Mother works:				
Full 5	.0468	...	.0486	...
Part 5	.0392	...	.0504	...
None 5	.7168	...	.7507	...
Full 14	.0822	...	.0936	...
Part 14	.0708	...	.0851	...
None 14	.6384	...	.6713	...
H.S. shop	.2592	...	.0908	...
Read	20.57	10.17	24.06	11.63
NR read	.0291	...	.0128	...
Mech	59.24	18.27	58.88	18.96
NR mech	.0025	...	0	...
Math	18.13	11.82	28.94	17.17
NR math	.0683	...	.0188	...
Dext	50.04	9.359	50.68	9.811
NR dext	0	...	.0071	...
Exp	29.33	2.439	24.54	2.907
Exp^2	866.1	147.1	610.4	147.4
S13–15	...	...	.3106	...
S16	...	...	.3993	...
S20	...	...	.0823	...
Year 48	46.62	1.584	48.05	1.869
Year 69	69.11	.3691	69.08	.3437
$\ln \bar{y}$	8.635	.4107	8.526	.3871
$\ln y(t)$	9.326	.4573	9.639	.4904
g	.0309	.0251	.0535	.0283
λ_a	−1.2870	.2873	−.3193	.2256
λ_b	.4666	.3763	1.605	.5212
No. observations	791		2820	

Table 1: Descriptive Statistics

Table 2 shows estimates from the reduced form probit model that described the relationship between college going and the family background and ability measures. The math score has a strong positive effect and the mechanical score a strong negative effect on the college going decision.

Variable	Reduced Form (16) Coefficient	t	Structure (26) Coefficient	t	Structure (29) Coefficient	t
Constant	.0485	.20	.1512	.22	.1030	.17
Background:						
Father's ED	−.0145	−.41	−.0168	−.54	−.0152	−.49
Father's ED²	.0037	2.05	.0038	2.26	.0037	2.26
DK ED	−.4059	−3.96	−.3924	−2.79	−.4001	−2.91
Manager	.1897	2.17	.1825	2.13	.1871	2.21
Clerk	.0556	.54	.0561	.59	.0554	.59
Foreman	.0182	.19	.0210	.23	.0200	.22
Unskilled	−.0910	−.85	−.0948	−.89	−.0928	−.87
Farmer	−.2039	−2.12	−.2256	−2.27	−.2094	−2.14
DK job	−.0413	−.19	−.0629	−.29	−.0609	−.28
Catholic	−.1144	−1.91	−.0982	−1.51	−.1083	−1.66
Jew	−.0293	−.23	.0143	.12	−.0158	−.14
Old sibs	−.0162	−.93	−.0162	−.93	−.0161	−.93
Young sibs	.0122	.63	.0096	.49	.0112	.57
Mother works:						
Full 5	.1039	.66	.1168	.81	.1104	.76
Part 5	.2179	1.42	.2106	1.52	.2156	1.56
None 5	.0655	.63	.0677	.65	.0661	.64
Full 14	.2898	2.29	.2884	2.30	.2888	2.33
Part 14	.2709	2.20	.2768	2.02	.2693	2.03
None 14	.1980	1.91	.1990	1.92	.1966	1.92
H.S. shop	−.4411	−6.14	−.4397	−3.74	−.4379	−3.90
Ability:						
Read	.0047	1.67	...	...	...	...
NR read	−.2575	−1.41	...	...	...	...
Mech	−.0070	−4.29	...	...	...	...
NR mech	−3.0236	−1.04	...	...	...	...
Math	.0244	12.34	...	...	...	...
NR math	−.7539	−5.75	...	...	...	...
Dext	.0019	.72	...	...	...	...
NR dext	2.2797	.47	...	...	...	...
Earnings:						
$\ln(\bar{y}_a/\bar{y}_b)$	...	...	5.1486	2.25	...	...
g_a	...	...	138.3850	1.83	7.6632	.11
g_b	...	...	−44.2697	−1.28	71.8981	2.34
$\ln y_a(t)/y_b(t)$	...	...	...	...	5.1501	2.57
Observations		3611		3611		3611
Limit observations		791		791		791
Nonlimit observations		2820		2820		2820
−2 ln (likelihood ratio)		579.5		568.8		576.6
χ^2 degree freedom		28		23		23

Table 2: Probit Estimates

The structural estimates of the parameters of the earnings and growth equations (corrected for sample selectivity) are shown in Table 3 (t-statistics in parentheses). Willis and Rosen (1979) do not include actual labor market experience because they want to get a measure of expected earnings at the time of making the schooling decision. The ability measure that has the largest effect on initial earnings is math score for college attendees. Ability indicators are more important for earnings growth and for later earnings. Effect of mechanical ability is negative in all cases.

The estimated coefficients on the selectivity correction term show no selectivity bias for initial earnings of high school graduates but positive bias for growth rates for high school earnings. The estimated coefficients also indicate indicate positive bias bias for initial college attendee earnings and negative bias for earnings growth rates. There is no evidence of selectivity bias for late earnings. This pattern provides support for comparative advantage in the labor market.

	Dependent Variable					
Regressor	$\ln \bar{y}_a$ (1)	$\ln \bar{y}_b$ (2)	g_a (3)	g_b (4)	$\ln y_a(\bar{t})$ (5)	$\ln y_b(\bar{t})$ (6)
Constant	8.7124	2.8901	.1261	.2517	10.3370	7.5328
	(16.51)	(1.37)	(3.90)	(2.11)	(5.52)	(2.08)
Read	.0009	−.0019	.0001	.0003	.0027	.0057
	(1.21)	(−1.17)	(1.11)	(3.20)	(2.80)	(3.28)
NR read	.0791	.0506	−.0034	−.0046	.0033	−.0402
	(1.24)	(.58)	(−.76)	(−.89)	(.04)	(−.42)
Mech	−.0002	−.0005	−.0001	−.0001	−.0021	−.0017
	(−.48)	(−.54)	(−2.16)	(−1.13)	(−3.59)	(−1.73)
NR mech	...	.1969	...	.0002	...	.2196
		(.69)		(.01)		(.68)
Math	.0015	−.0013	.0001	−.0000	.0030	−.0019
	(2.02)	(.74)	(1.18)	(−.20)	(3.31)	(−1.00)
NR math	−.1087	.0562	.0015	.0006	−.0877	.0712
	(−1.94)	(.83)	(.38)	(.15)	(−1.24)	(.96)
Dext	.0008	−.0019	−.0000	.0003	.0002	.0036
	(1.03)	(−1.21)	(−.78)	(2.77)	(.16)	(2.19)
NR dext	.0751	...	−.0004	...	.1466	...
	(.28)		(−.02)		(.43)	
Exp	−.0523	.4260	−.0028	−.0154	−.0129	.0776
	(−1.49)	(3.10)	(−1.11)	(−1.93)	(−.29)	(.53)
Exp^2	.0015	−.0067	.0000	.0002	−.0000	−.0012
	(2.22)	(−2.95)	(.21)	(1.82)	(−.01)	(−.49)
Year 48	−.0020	−.0156	...	...	...	...
	(−.48)	(−1.72)				
Year 69	...	...	...	...	−.0067	.0039
					(−.26)	(.09)
S13–15	.1288	...	−.0062	...	.0168	...
	(5.15)		(−3.49)		(.52)	
S16	.0760	...	.0026	...	.1095	...
	(3.82)		(1.79)		(4.26)	
S20	.1318	...	.0049	...	.2560	...
	(4.10)		(2.13)		(6.15)	
λ_a	−.1069	...	.0058	...	.0206	...
	(−3.21)		(2.45)		(.49)	
λ_b	...	−.0558	...	.0118	...	.2267
		(−.66)		(2.39)		(2.48)
R^2	.0750	.0439	.1578	.0513	.0740	.0358

Table 3: Structural estimates

Turning back to Table 2, see the structural probit estimtes, which are remarkably similar whether the model is based on initial earnings or late earnings.