5.1 The Roy Model

$$\ln w_i^U = x_i'\beta^U + u_i^U, \quad u_i^U \sim N(0, \sigma_U^2)$$

$$\ln w_i^N = x_i'\beta^N + u_i^N, \quad u_i^N \sim N(0, \sigma_N^2)$$

$$U_i^* = \delta_0 + \delta_1(\ln w_i^U - \ln w_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i > 0$$

(a) Expected union/non-union wage differential for a randomly chosen individual with characteristics x_i?

$$\mathbb{E}(\ln w_i^U - \ln w_i^N \mid x_i) = \mathbb{E}[x_i'\beta^U + u_i^U - x_i'\beta^N - u_i^N \mid x_i] =$$

$$= x_i'(\beta^U - \beta^N) + \underbrace{\mathbb{E}(u_i^U - u_i^N \mid x_i)}_{\text{error terms are orthogonal to } x_i\text{, thus the whole term } = 0} = x_i'(\beta^U - \beta^N)$$

Expected wage differential for a union worker with x_i?

$$\mathbb{E}(\ln w_i^U - \ln w_i^N \mid x_i, U_i^* > 0) =$$

$$= x_i'(\beta^U - \beta^N) + \mathbb{E}(u_i^U - u_i^N \mid \delta_0 + \delta_1(\ln w_i^U - \ln w_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i > 0) =$$

$$= x_i'(\beta^U - \beta^N) + \mathbb{E}(u_i^U - u_i^N \mid \delta_0 + \delta_1(x_i'\beta^U + u_i^U - x_i'\beta^N - u_i^N) + x_i'\delta_2 + z_i'\delta_3 - v_i > 0) =$$

$$= x_i'(\beta^U - \beta^N) + \mathbb{E}(u_i^U - u_i^N \mid \underbrace{\delta_0 + \delta_1(x_i'\beta^U) - \delta_1(x_i'\beta^N) + x_i'\delta_2 + z_i'\delta_3}_{\equiv c} + \delta_1 u_i^U - \delta_1 u_i^N - v_i > 0) =$$

$$= x_i'(\beta^U - \beta^N) + \mathbb{E}(u_i^U - u_i^N \mid \delta_1(u_i^U - u_i^N) - v_i > -c) =$$

Let's regress $u_i^U - u_i^N$ on $\delta_1(u_i^U - u_i^N) - v_i$:

$$u_i^U - u_i^N = a_1 \cdot [\delta_1(u_i^U - u_i^N) - v_i] + \xi_i$$

$$a_1 = \frac{\text{cov}(u_i^U - u_i^N, \delta_1(u_i^U - u_i^N) - v_i)}{\text{var}(\delta_1(u_i^U - u_i^N) - v_i)} = \frac{\delta_1\sigma_U^2 + \delta_1\sigma_N^2}{\delta_1^2(\sigma_U^2 + \sigma_N^2) + \sigma_V^2}$$

$$\Leftrightarrow X_i'(\beta^u-\beta^N)+\mathbb{E}\left[a_1\left(\delta_1(u_i^u-u_i^N)-v_i\right)+\xi_i \mid \delta_1(u_i^u-u_i^N)-v_i>-c\right]=$$

$$= X_i'(\beta^u-\beta^N)+a_1\mathbb{E}\left[\delta_1(u_i^u-u_i^N)-v_i \mid \delta_1(u_i^u-u_i^N)-v_i>-c\right]+$$

$$+\underbrace{\mathbb{E}\left[\xi_i \mid \delta_1(u_i^u-u_i^N)-v_i>-c\right]}_{=0} =$$

$$= X_i'(\beta^u-\beta^N)+\frac{\delta_1(\sigma_u^2+\sigma_N^2)}{\delta_1^2(\sigma_u^2+\sigma_N^2)+\sigma_v^2}\mathbb{E}\left[\delta_1(u_i^u-u_i^N)-v_i \mid \delta_1(u_i^u-u_i^N)-v_i>-c\right]$$

(b) No, it's not possible because selection bias can arise for several reasons:

- participation in the union may be correlated with the error term for unionized workers, namely $\mathbb{E}(u_i^u / u_i^* = 1) \neq 0$
- non-participation in the union may be correlated with the error term for non-unionized workers, namely $\mathbb{E}(u_i^N / u_i^* = 0) \neq 0$
- both events happen simultaneously

(c) $(u_i^u, u_i^N, v_i) \sim N\left[\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}\sigma_u^2 & 0 & 0\\ 0 & \sigma_N^2 & 0\\ 0 & 0 & \sigma_v^2\end{pmatrix}\right]$

(d) In a first step a reduced-form selection equation is obtained after substituting the wage equations into the selection equation:

$$U_i^* = \delta_0 + \delta_1(X_i'\beta^U + u_i^U - X_i'\beta^N - u_i^N) + z_i'\delta_3 + v_i + x_i'\delta_2 =$$
$$= \gamma_0 + \gamma_1 w_i + \xi_i$$

w_i - all exogenous variables X_i, z_i

This equation is estimated by probit.

Union wage equation, conditional on union status is given by

$$\ln W_i^U = X_i'\beta^U + \sigma_{u\xi}\frac{\phi(\psi_i)}{\Phi(\psi_i)} + \eta_i^U$$

$$\ln W_i^N = X_i'\beta^N - \sigma_{n\xi}\frac{\phi(\psi_i)}{1-\Phi(\psi_i)} + \eta_i^n$$

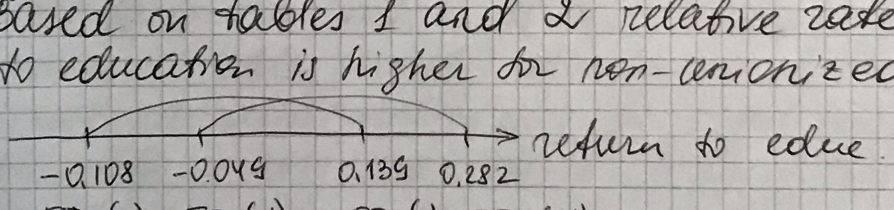
$\psi_i = \gamma_0 + \gamma_1 w_i$.

These two equations can be estimated by OLS using observations on the subsamples, where ψ_i is replaced by $\hat{\psi}_i = \hat{\gamma}_0 + \hat{\gamma}_1 w_i$

(e) We can use estimated values of $\ln\hat{W}_i^U$ and $\ln\hat{W}_i^N$ from the previous ex (d) and substitute them into $U_i^* = \delta_0 + \delta_1(\ln\hat{W}_i^U - \ln\hat{W}_i^N) + x_i\delta_2 + z_i'\delta_3 - v_i$

The latter equation is supposed to be estimated by probit

(f) Based on tables 1 and 2 relative rate of return to education is higher for non-unionized sector



$$\left|\frac{0.139+0.108}{-0.108}\right| = 2.28 \qquad \left|\frac{0.282+0.049}{0.049}\right| = 6.75$$

- Market experience contributes more to the union wage estimates. "The negative and significant coefficient of ME_2 confirms the strict concavity of the earning profile" (Lee)
- "Male operatives receive higher wages than female operatives" (Lee) "The wage difference between male and female operatives is larger in the unionized sector, so females may prefer the nonunionized sector" (Lee)

- "White operatives receive higher wages than nonwhite operatives." The black and white wage difference is less in the unionized sector than in the nonunionized sector. While there is evidence that black may be discriminated against entering labor unions, they do better in unionized sectors once they are members.
- ~~Higher~~ Health impediments are more restrictive for non-unionized workers, which means that people with health impediments get lower salary in nonunionized sectors

(g) Both selectivity coefficients are positive (because for the unionized wages the coef is for a variable $-f(\Psi_i)/F(\Psi_i)$ and for non-unionized for a variable $f(\Psi_i)/(1-F(\Psi_i))$
"The positive value results from the individual's selection of the relevant sector that pays him better than the average operatives with the same characteristics and under the same working circumstances" (Lee)

(h) Based on table 6 the "most powerful factor determining unionization status is the union-nonunion wage differential"

(i) "In table 7 the estimates give the net effects for the various factors on union status. A comparison of the estimates in table 6 and 7 shows some differences. The most educated workers will be in the nonunion sector. Experienced workers would tend to be in the unionized sector." (Lee)