Structural Econometrics

Optimal Stopping-Model of Schooling

Background: *Belzil and Hansen (2002, Econometrica) and Belzil and Hansen (2007, Journal of Econometrics)*

- Model Structure: T periods to allocate (Finite Horizon) between two activities: work (full-time) vs education (with a maximum for schooling)

- Evevery individual starts in school and eventually works (limit to schooling)

- Assume to intermediate state (like no school-no work)

- Utility of attending school, $U_{it}^{s}(.)$:

$$\begin{aligned} U_{it}^{s}(.) &= X_i'\delta + \psi(S_{it}) + \upsilon_i^{\xi} + \varepsilon_{it}^{\xi} \\ &= \bar{U}_{it}^{s}(.) + \varepsilon_{it}^{\xi} \end{aligned}$$

$$\varepsilon_{it}^{\xi} \sim i.i.dN(0, \sigma_{\xi}^{2})$$

- Utility of Work is logarithm of wages

$$U_{it}^{w}(.) = \ln(w_{it}). = \bar{U}_{it}^{w}(.) + \varepsilon_{it}^{w} \text{ with } \varepsilon_{it}^{w} \sim i.i.dN(0, \sigma_{w}^{2})$$

$$\bar{U}_{it}^{w}(.) = \varphi_1(S_{it}) + \varphi_2.Exper_{it} + \varphi_3.Exper_{it}^{2} + \ \upsilon_i^{w}$$

- υ_i^w is innate market ability

- υ_i^ξ is innate taste for schooling (academic ability)

- υ_i^w and υ_i^ξ are correlated

- S_{max} is the maximum schooling achievable and $T = 65$

- **Key issues:**
  - Law of motion is deterministic:

$$S_{it} = \sum_{j=1}^{t-1} d_{it}$$

$$Exper_t = T - S_{it}$$

  - Post schooling human capital is exogenous (need to solve for S only)

$$\Omega_{it} = (S_{it}, \eta_{it}) \text{ and } \eta_{it} = (\varepsilon_{it}^{\xi}, \varepsilon_{it}^{w})$$

- However, because random shocks are i.i.d., we will omit them.

Bellman Equation: Value of Attending School:

$$V_t^s(S_t, \eta_t) = X_i' \delta + \psi(S_{it}) + \upsilon_i^{\xi} + \varepsilon_{it}^{\xi} +$$

$$\beta E(Max\{V_{t+1}^s(S_{t+1}), V_{t+1}^w(S_{t+1})\} \mid d_t = 1)$$

$$= \bar{U}_{it}^s + \varepsilon_{it}^{\xi} + \beta E V_{t+1}(. \mid d_t = 1)$$

$$= \bar{V}_t^s + \varepsilon_{it}^{\xi}$$

Define $\bar{V}_t^s$ as

$$\bar{V}_t^s = X_i' \delta + \psi(S_{it}) + \upsilon_i^{\xi} + \beta E V_{t+1}(.) \ .$$

Bellman Equation: Value of entering the labor market:

$$V_t^w(S_t, \eta_t) = \varphi_1(S_{it}) + \varphi_2.Exper_{it} + \varphi_3.Exper_{it}^2 + \ \upsilon_i^w + \varepsilon_{it}^w$$

$$+\beta E(V_{t+1}^w(S_{t+1} = S_t) \mid d_t = 0)$$

$$= \bar{V}_t^w(S_t, \eta_t) + \varepsilon_{it}^w$$

Define $\bar{V}_t^w(S_t, \eta_t)$ as

$$\bar{V}_t^w(S_t, \eta_t) = \varphi_1(S_{it}) + \varphi_2.Exper_{it} + \varphi_3.Exper_{it}^2 +$$

$$\upsilon_i^w + \beta E V_{t+1}^w(S_{t+1}, \eta_{t+1} \mid d_t = 0)$$

- . Two important thing to note:

- First, at entrance in the market, accumulated experience is 0), so

$$\bar{V}_t^w(S_t, \eta_t) = \varphi_1(S_{it}) + \ \upsilon_i^w + \beta EV_{t+1}^w(S_{t+1}, \eta_{t+1} \mid d_t = 0)$$

- Second, work is an "absorbing" state so no maximization beyond entrance in the market (EV_{t+1}^w will be a simple sum of discounted earnings)

$$EV_{t+1}^w(.) = \sum_{j=t+1}^{T} \beta^{j-(t+1)}(\varphi_1(S_j) + \varphi_2.Exper_j + \varphi_3.Exper_j^2 \ + \upsilon_i^w)$$

- Obviously, calculating $EV^w_{t+1}(.)$, the expected utility of working from $t+1$ until T, does not require any maximization.

Solution

It is now easy to use backward recursions

- Start at $\tilde{t}$ with $(S_{\text{max}} - 1)$ years of schooling already accumulated:

- This is the last date at which we need to compute a schooling choice probability because

$$S_{\tilde{t}} = S_{\text{max}} - 1$$

- Either
  - $d_{\tilde{t}} = 1$(obtain S$_{\text{max}}$ years of schooling and next period start working for $EV^w_{\tilde{t}+1}(S_{\text{max}}, Exper = 0.)$
  - $d_{\tilde{t}} = 0$(obtain S$_{\text{max}}$ -1 years of schooling, start working in $\tilde{t}$, and next period work for $EV^w_{\tilde{t}+1}(S_{\text{max}} - 1, Exper = 1.)$
  - Note that both $EV^w_{\tilde{t}+1}(.)$ is easily computed

- Write down the probabilistic statement (Prob $d_{\tilde{t}} = 1$) as

$$\mathsf{Pr}(d_{\tilde{t}} = 1) = \mathsf{Pr}(V^s_{i\tilde{t}} - V^w_{i\tilde{t}} > 0) =$$

$$= \mathsf{Pr}\{(X_i'\delta + \psi(S_{\mathsf{max}}) + \upsilon_i^{\xi} + \varepsilon_{it}^{\xi} + \beta EV^w_{\tilde{t}+1}(S_{\mathsf{max}}, Exper = 0.)$$

$$> \varphi_1(S_{\mathsf{max}} - 1) + \ \upsilon_i^w + \varepsilon_{it}^w + EV^w_{\tilde{t}+1}(S_{\mathsf{max}} - 1, Exper = 1.)\}$$

$$= \mathsf{Pr}(\varepsilon^s_{it} - \varepsilon^w_{it} > \bar{V}^w_{i\tilde{t}} - \bar{V}^s_{i\tilde{t}})$$

which is simply a non-linear probit equation (conditional on υ_i^w and υ_i^s).

- Given $\mathsf{d}_{\tilde{t}-1} = 1$,

$$
\begin{aligned}
EV_{\tilde{t}}(.) &= pr(d_{\tilde{t}} = 1) \cdot E(V^s_{i\tilde{t}} \mid V^s_{i\tilde{t}} > V^w_{i\tilde{t}}) + \\
pr(d_{\tilde{t}} &= 0) \cdot E(V^w_{i\tilde{t}} \mid V^w_{i\tilde{t}} > V^s_{i\tilde{t}})
\end{aligned}
$$

- Take one component: say $E(V^s_{i\tilde{t}} \mid V^s_{i\tilde{t}} > V^w_{i\tilde{t}})$

$$
\begin{aligned}
E(V^s_{i\tilde{t}} &\mid V^s_{i\tilde{t}} > V^w_{i\tilde{t}}) \\
&= X_i'\delta + \psi(.) + \upsilon_i^{\xi} + \beta EV^w_{\tilde{t}+1}(S_{\mathsf{max}}, Exp = 0) \\
+E\{\varepsilon^s_{it} &\mid \varepsilon^s_{it} - \varepsilon^w_{it} > EV^w_{\tilde{t}+1}(S_{\mathsf{max}} - 1, Exp = 1) \\
-\beta EV^w_{\tilde{t}+1}(S_{\mathsf{max}}, Exper &= 0.)\}
\end{aligned}
$$

- Evaluating $E(\varepsilon_{it}^s \mid \varepsilon_{it}^s - \varepsilon_{it}^w > ......)$ is easily done using standard truncated normal distribution expression (selectivity literature).

- **Digression: Truncated Normal Distributions**

$Z \sim N(\mu, \sigma^2)$, c is a constant (truncation point), $\alpha = \frac{(c-\mu)}{\sigma}$

$$E(Z \quad | \quad Z > c) = \mu + \sigma \cdot \frac{\phi(\alpha)}{(1 - \Phi(\alpha))}$$

$$E(Z \quad | \quad Z < c) = \mu - \sigma \cdot \frac{\phi(\alpha)}{\Phi(\alpha)}$$

where the expression, $\frac{\phi(\alpha)}{(1-\Phi(\alpha))}$, is often referred to as a **Mill's ratio**.

- In our example:
  - $\mu = 0$ (since the $\varepsilon' s$ have mean 0),
  - σ is the standard -deviation of $\varepsilon_{it}^{s} - \varepsilon_{it}^{w}$,
  - c is $EV_{\tilde{t}+1}^{w}(S_{\mathsf{max}} - 1, Exper = 1.) - \beta EV_{\tilde{t}+1}^{w}(S_{\mathsf{max}}, Exper = 0.)$
- Then, evaluate $E(V_{i\tilde{t}}^{w} \mid V_{i\tilde{t}}^{w} > V_{i\tilde{t}}^{s}) = \ldots\ldots$
- In the end, you have

$$pr(d_{\tilde{t}} = 1) \cdot E(V_{i\tilde{t}}^{s} \mid V_{i\tilde{t}}^{s} > V_{i\tilde{t}}^{w}) + pr(d_{\tilde{t}} = 0) \cdot E(V_{i\tilde{t}}^{w} \mid V_{i\tilde{t}}^{w} > V_{i\tilde{t}}^{s})$$

- Going to $\tilde{t}-1$,

$$V^s_{\tilde{t}-1} = X_i'\delta + \psi(S_{\mathsf{max}} - 1) + \upsilon_i^{\xi} + \varepsilon_{it}^{\xi} + \beta EV_{\tilde{t}}(. \mid d_{\tilde{t}-1} = 1)$$

$$V^w_{\tilde{t}-1} = \varphi_1(S_{\mathsf{max}} - 2) + \ \upsilon_i^{w} + \varepsilon_{it}^{w} + EV^w_{\tilde{t}+1}(S_{\mathsf{max}} - 2, experience = 1.)$$

- Proceed recursively until you reach the initial period (say the minimum schooling).

Estimation is achieved by maximum likelihood.

- Take an individual with τ years of schooling. This individual would have wages observed from $\tau + 1$ to T

- The Likelihood has 3 parts:

- the probability of having spent at most τ years in school, which can be easily derived from the structure of the model

$$L_1 = Pr[(d_0 = 1), (d_1 = 1)....(d_\tau = 1)]$$

- the probability of entering the labor market in year $\tau + 1$, at observed wage $w_{\tau+1}$, which can be factored as the product of a normal conditional probability times a marginal.

$$L_2 = \mathsf{Pr}(d_{\tau+1} = 0, w_{\tau+1}) = \mathsf{Pr}(d_{\tau+1} = 0 \mid w_{\tau+1}) \cdot \mathsf{Pr}(w_{\tau+1})$$

- the density of observed wages and employment rates from $\tau + 2$ until 1990.

$$L_3 = Pr\left(\{w_{\tau+2}\} \cdot .... \mathsf{Pr}\{w_{1990}\}\right)$$

The likelihood function, conditional on $\vartheta_j = \left(\upsilon^{\xi}, \upsilon^{w},\right)_j$, is given by

$$L_i(\vartheta_j) = L_{1i}(\vartheta_j) \cdot L_{2i}(\vartheta_j) \cdot L_{3i}(\vartheta_j)$$

$$\mathrm{Pr}(\upsilon^{\xi} = \upsilon_j^{\xi}, \upsilon^{w} = \upsilon_j^{w}) = p_j$$

with,

$$p_j = \frac{\exp(p_j^0)}{\sum_{k=1}^{K=6} \exp(p_k^0)}$$

with $p_6^0 = 0$.

The unconditional contribution to the log likelihood, for individual i, is therefore given by

$$\log L_i = \log \sum_{j=1}^{K=6} p_j \cdot L_i(. \mid \vartheta_j)$$

where each p_j represents the population proportion of type ϑ_j.

Estimation provides parameter estimates for all the parameters, including $\psi(S_{it})$, $\upsilon^{\xi}, \upsilon^{w}$ and p_j.

NOTES:

- The number of types should go to N (number of individuals) but this is not feasible.
- The number of types is not a standard parameter (no obvious asymptotic results since you search over a space that has no openness property)
- One approach: Start from large number, and reduce it (the number of types) by likelihood ratio test
- Another approach (to remove arbitrariness): use penalized likelihood approaches using Bayesian criterion.

Model Simulation:

- First, set a population size (say 10,000 individuals)

- Out of these 10,000 individuals, the population proportions will be given by $\hat{p}_j$

- Generate 10,000 variates from a uniform distribution, and assign a type to each 10,000 individuals

- for each individual, generate a random shock for each state and each period and solve the problem

- you end-up with 10,000 wage-schooling histories

- Note: you know schooling choices, employment but also υ_i^{ξ} and υ_i^{w}

- you can evaluate correlation between schooling and $\upsilon_i^{\xi}, \upsilon_i^{w}$

- you can simulate what would happen if you change $\varphi_1(S_{it})$

- Belzil and Hansen (2002) use a dynamic programming model to
  - investigate the shape of the wage schooling relationship
  - evaluate ability bias as defined by the correlation between schooling achievement and the individual-specific intercept term of the wage equation.
- The model is implemented on a panel of white males taken from the NLSY covering the years 1979 to 1990.
- BH (2002) deal with the *Convexity of the Wage-Schooling Relationship and the Ability Bias*. $\varphi_1(S_{it})$ is a non-linear function of S_{it}