Estimating an Optimal Stopping Model of Schooling

1 Data

- The Data come from *Génération 98*, a large scale survey conducted in France by Céreq.
- It provides detailed information on the socio-demographic background and employment characteristics of young individuals who left school in the year 1998 and were interrogated in early 2001.
- Re-interviews have been conducted for half of the sample in 2003.
- The aim of *Génération 98* is to document many aspects of early labor market transitions. In particular, *Génération 98* provides information on spells of employment, unemployment, and training experienced between school completion (labor market entrance) and the date of the survey.
- Because *Génération 98* is a national survey of those who left the educational system at a particular point in time (1998), all individuals faced the same labor market conditions after 1998.

2 Variables

- Father's occupation : Farmer_F, trade_F, Exec_F, ...
- Mother's occupation : Farmer_M, Trade_M, Exec_M ...
- Late_6 or repeat (1 if has repeated grade in primary school), noted as *rep*
- Urban=1 if lives in urban area
- Ethnic origin
  - ORFR: both parents born in France
  - ORoecd: at laest one born in OECD
  - Orother: at least one born in non-oecd
- male=1 if male

- Foreign born=1 if not born in France
- Montly Earnings at entrance
  - Wage_first (0 month)
  - full_first
- Montly Earnings at second interview (36 months)
  - Employed_36 (equal to 1 if employed, 0 if not)
  - Full_36 (equal to 1 if full time)
  - Wage_36
  - Experience_36
- Montly Earnings at third interview (60 months)
  - Employed_60 (equal to 1 if employed, 0 if not)
  - Full_60 (equal to 1 if full time)
  - Wage_60
  - Experience_60

3 A simple Model Structure

- **Two states for $\mathbf{t = 1, 2, ...T}$ periods to allocate (T=50 years) :**
  - **School (every individual starts there)**
  - **Full Time work**
- Schooling is constructed as follows:
  - Assume $s = 1, 2, S$ grade levels with a standard number of years for duration of grade s :
    * $s = 1$: Lycee (t=1), 1 year
    * $s = 2$: Lycee (CAP/BEP), 1 year
    * $s = 3$: Lycee (bac), 1 year
    * $s = 4$: Enseignement sup (sans diplome), 1 year
    * $s = 5$: Bac+2, 1 year
    * $s = 6$: Bac+3: 1 year

    * $s = 7$: Bac+4, 1 year
    * $s = 8$: Bac+5 (or more) 1 year

- For each grade level, s,
  - Assume the "normal" age level of entry in the market (A_s). In other words, and individual who completes grade s, must enter the labor market at age A_s (in case he does not go further) with starting values $A(s = 1) = 16$.
  - Grade-level-specific total (maximum) experience, $T = 50 - A_s$.
  - In general, at any date/age t, accumulated experience is $t - A_s$.
- Only choices while schooling (no come back to school). No uncertainty in the labor market.

The Utility Equations

Utility of Schooling for individual i:

$$U_i^s = \; X_i' \delta_x + \delta_s + \delta_L \cdot \mathbb{1}(rep) + \varepsilon_i^s$$

$$\text{with } \varepsilon_i^s \sim i.i.dN(0, \sigma = 1)$$

Utility of working for schooling s at time t (**no random term**) :

$$U_i^w(s,t) = \ln(w_i) = \varphi_1 + \varphi_s \cdot \mathbb{1}(s) + \varphi_2.(t - A_s) + \varphi_3.(t - A_s)^2 + \; \phi_4 \cdot \mathbb{1}(rep)$$

In last period, the utility of work is:

$$U_i^w(s,50) = \ln(w_i) = \varphi_1 + \varphi_s \cdot \mathbb{1}(s) + \varphi_2.(50 - A_s) + \varphi_3.(50 - A_s)^2 + \; \phi_4 \cdot \mathbb{1}(rep)$$

Timing

1. Individuals are in school at $t = 1, s = 1$
2. At the end of each period, they receive draws (randomness) and choose between working or staying in school at the beginning of next period.
3. Individuals leave school at age A(s) and stay in work until $T = 50$ (no ε's while working). They are forced to leave if $s = 8$.

4 Bellman Equations

Value of Attending Grade s:

$$V^s(s) \quad = \quad X_i'\delta_x + \delta_s \cdot \mathbb{1}(s) + \delta_L \cdot \mathbb{1}(rep) + \varepsilon_i^t + \beta \mathbb{E}(\max\{V^s(s+1), V^w(s, A_s)\})$$

Value of entering the labor market with grade s, at age A_s : sum over all remaining periods without randomness :

$$\begin{aligned} V^w(s, A_s) &= \varphi_1 + \varphi_s \cdot \mathbb{1}(s) + \quad \phi_4 \cdot \mathbb{1}(rep) \\ &\quad + \beta\{\varphi_1 + \varphi_s \cdot \mathbb{1}(s) + \varphi_2.(t - A_s) + \varphi_3.(t - A_s)^2 + \ \phi_4 \cdot \mathbb{1}(rep) \\ &\quad +... + \\ &\quad + \beta^{(T-A_s-1)}\{\varphi_1 + \varphi_s \cdot \mathbb{1}(s) + \varphi_2.(T - A_s) + \varphi_3.(T - A_s)^2 + \ \phi_4 \cdot \mathbb{1}(rep) \\ &= \sum_{j=A_s}^{T} \beta^{j-A_s}\{\varphi_1 + \varphi_s \cdot I(s) + \varphi_2.(j - A_s) + \varphi_3.(j - A_s)^2 + \ \phi_4 \cdot \mathbb{1}(rep)\} \end{aligned}$$

5 Solution : Backwards induction

Because work is an absorbing state without randomness, one only needs to study schooling choices.

- Start from $s = 8$: it is clear that the value function of entering grade 8 (and automatically starting full-time work at age A_8 since grade 8 is terminal), $V^s(8)$, has a trivial structure, since no Emax(.) needs to be taken:

  $$V^s(8) = X_i'\delta_x + \delta_8 + \delta_L \cdot \mathbb{1}(rep) + \varepsilon_i^8 + V^w(8, A_8)$$

  The future utility associated to entering grade 8, $V^w(8, A_8)$, is known (sum of life cycle earnings in logs from age A_8 to age 60).

- When contemplating entering grade 8 (end of period 7), the choice requires a comparison between $V^s(8)$ and $V^w(7, A_7)$. So, the choice probability is:

  $$\begin{aligned} \mathbb{P}(d_8 &= 1) = \mathbb{P}(V^s(8) > V^w(7, A_7)) \\ &= \mathbb{P}(X_i'\delta_x + \delta_8 + \delta_L \cdot \mathbb{1}(rep) + \varepsilon_i^8 + V^w(8, A_8) > V^w(7, A_7)) \\ &= \mathbb{P}\{\varepsilon_i^8 > V^w(7, A_7) - V^w(8, A_8) - X_i'\delta_x - \delta_8 - \delta_L \cdot \mathbb{1}(rep)\} \\ &= 1 - \Phi\{V^w(7, A_7) - V^w(8, A_8) - X_i'\delta_x - \delta_8 - \delta_L \cdot \mathbb{1}(rep)\} \end{aligned}$$

Obviously, the choice of not attending grade 8 (stopping at grade 7) is :

$$\begin{aligned}\mathbb{P}(d_8 &= 0) = \mathbb{P}(V^s(8) \leq V^w(7, A_7)) \\ &= \Phi\{V^w(7, A_7) - V^w(8, A_8) - X_i'\delta_x - \delta_8 - \delta_L \cdot \mathbb{1}(rep)\}\end{aligned}$$

- When we will solve the problem recursively, the first non-trivial exercise will be when contemplating grade 7 :
  - You will need to compute $EMax\{V^s(s+1), V^w(s)\}$ or, in other words, $EMax\{V^s(8), V^w(7, A_7)\}$
  - To reduce notation burden, denote $EMax\{V^s(8), V^w(7, A_7)\}$ by $EV(8)$ since this choice is made at the beginning of period 8,

$$\begin{aligned}EV(8) &= \mathbb{P}(d_8 = 1) \cdot \mathbb{E}(V^s(8) \mid V^s(8) > V^w(7, A_7)) \\ &+ \mathbb{P}(d_8 = 0) \cdot \mathbb{E}(V^w(7, A_7) \mid V^w(7, A_7) > V^s(8)) \\ &= \mathbb{P}(d_8 = 1) \cdot \mathbb{E}(V^s(8) \mid V^s(8) > V^w(7, A_7)) \\ &+ (1 - \mathbb{P}(d_8 = 1)) \cdot \mathbb{E}(V^w(7, A_7))\end{aligned}$$

- Note that
  - $V^w(7, A_7)$ is non stochastic, so $\mathbb{E}(V^w(7, A_7)) = V^w(7, A_7)$ and is already known.
  - we only need to evaluate $\mathbb{E}(V^s(8) \mid V^s(8) > V^w(7, A_7))$
  - Formally, this is :

$$\begin{aligned}\mathbb{E}\{X_i'\delta_x + \delta_s \cdot I(8) + \delta_L \cdot \mathbb{1}(rep) + \varepsilon_i^8 + V^w(8, A_8) \quad &| \\ X_i'\delta_x + \delta_s \cdot I(8) + \delta_L \cdot \mathbb{1}(rep) + \varepsilon_i^8 + V^w(8, A_8) \quad &> \quad V^w(7, A_7)\}\end{aligned}$$

  - Recall that $\varepsilon_i^8 \sim N(\mu, \sigma^2)$, $\mu = 0, \sigma^2 = 1$
  - And we have $\mathbb{E}(\varepsilon_i^8 \mid \varepsilon_i^8 > c) = \mu + \sigma\frac{\phi((c-\mu)/\sigma)}{1-\Phi((c-\mu)/\sigma)}$ where $c = V^w(7, A_7) - V^w(8, A_8) - X_i'\delta_x - \delta_8 - \delta_L \cdot \mathbb{1}(rep)$
  - So $\mathbb{E}(V^s(8) \mid V^s(8) > V^w(7, A_7)$) is equal to :

$$\begin{aligned}X_i'\delta_x + \delta_8 + \delta_L \cdot \mathbb{1}(rep) + V^w(8, A_8) \\ +\mathbb{E}(\varepsilon_i^8 \mid \varepsilon_i^8 > V^w(7, A_7) - V^w(8, A_8) - X_i'\delta_x - \delta_8 - \delta_L \cdot \mathbb{1}(rep)\} \\ = X_i'\delta_x + \delta_8 + \delta_L \cdot \mathbb{1}(rep) + V^w(8, A_8) + \frac{\phi(c)}{1 - \Phi(c)}\end{aligned}$$

- Keep $EV(8)$ in memory
- Proceed recursively until you reach the initial period (the minimum schooling).
- In the preceding choice (from period 6 to 7), we get
  - Compare $V^s(7)$ with $V^w(6, A_6)$
  - where $V^s(7)$ is simply

$$\begin{aligned} V^s(7) &= X_i'\delta_x + \delta_7 + \delta_L \cdot \mathbb{1}(rep) + \varepsilon_i^7 + \mathbb{E}(\max\{V^s(8), V^w(7, A_7)\}) \\ &= X_i'\delta_x + \delta_7 + \delta_L \cdot \mathbb{1}(rep) + \varepsilon_i^7 + EV(8) \end{aligned}$$

    where $EV(8)$ is now given
  - Compute
    * $\mathbb{P}(d_7 = 1)$
    * $\mathbb{P}(d_7 = 0)$
    * $\mathbb{E}(V^s(7) \mid V^s(7) > V^w(6, A_6)$
    * $EV(7)$ and keep it
- Compare $V^s(6)$ with $V^w(5, A_5)$
  - etc.....

Estimation is achieved by maximum likelihood

- The probability of having reached grade level $\tilde{s}$, is

$$\begin{aligned} \mathcal{L} &= \mathbb{P}[(d_1 = 1), (d_2 = 1)....(d_{\tilde{s}} = 1), (d_{\tilde{s}+1} = 0)] \\ &= \mathbb{P}(d_1 = 1) \cdot \mathbb{P}(d_2 = 1)....\mathbb{P}(d_{\tilde{s}} = 1) \cdot \mathbb{P}(d_{\tilde{s}+1} = 0) \end{aligned}$$

  except for the probability of having reached grade level 8, which is

$$\begin{aligned} \mathcal{L} &= \mathbb{P}[(d_1 = 1), (d_2 = 1)....(d_8 = 1)] \\ &= \mathbb{P}(d_1 = 1) \cdot \mathbb{P}(d_2 = 1).... \cdot \mathbb{P}(d_8 = 1) \end{aligned}$$

6 Estimation/Programming

Before estimation of the structural model, we need the following:

Step 0 : Open the dataset, create useful variables (indicators).

Step 1: Transform Monthly earnings on a yearly basis (multiply by 12)

Step 2:Estimate independently $\varphi_s \cdot I(s) + \varphi_2.(t - A_s) + \varphi_3.(t - A_s)^2 + \phi_4 \cdot \mathbb{1}(rep)$ using data on transformed yearly earnings provided in the data. You can use random effect model, since we need return to schooling or simple OLS.

Step 3: Using these estimates, form the expression for $V^w(s, A_s)$ for $s = 1, 2, ..8$

1. Define the functions $A(s)$ and $V^w(s, A_s)$ with estimated coefficients
2. Solve the recursive program for period 8 (Write $\mathbb{P}(d_8 = 1)$ and $EV8$).
3. Do the same for all periods from 7 to 1
4. Write and optimize the likelihood

One important but not so easy point is to pick *good* **starting values**. Here are some hints :

- Start with 0 values for X's and try to fit estimated probabilities with real hazard rates playing on starting values for the coefficients δ_s. Just compare hazard functions for observed choices and compare to estimated probabilities.
- If optimization routine gives you some errors, try to check if there is no numerical problems (such as degenerated probabilities in the V functions, meaning, probabilities always equalling 0 or 1). The idea is that to optimize, starting values do not have to be corner values (e.g. recall that the "numerical support" of a normal law is [-3,3] : outside, values are too close to zero).
- For X's begin with 0 and if no problems, keep it. If new numerical errors, do the same as previously.
- You can work on subsample to optimize faster or test your program on few individuals.