Career Pathways: Advanced Study or Immediate Work

Yin-Yun Li

December, 2024

1. Introduction

According to the Ministry of Interior in Taiwan, the proportion of the population with tertiary education has been increasing. In 2022, 42% of young adults aged between 25 and 64 completed the highest level of education. Some attribute this trend to higher education reforms in Taiwan since 1990s, which involved the extensive establishment of universities and made higher education more accessible to the younger generation. Others suggest that low tuition fees and school subsidies also play a role in shaping individuals' choices.

In Taiwan, compulsory education spans ages 6 to 15. Most students then choose to enroll in senior high school for an additional three years. After completing the 12-year national education program, they can decide whether to pursue advanced study or enter the labor market directly.

We usually think that going to colleges helps us develop occupation-specific skills and earn higher wages in the labor market. However, identifying the causality is empirically challenging. What factors affect a younger person's career choice? How is the cost of schooling measured? To what extent does policy reform matter?

To answer these questions, we construct a simple dynamic discrete choice model to analyze a single agent's decision based on human capital, education cost and heterogeneous shock.

2. Related Literature

Our model connects human capital and the return to education within a dynamic discrete choice framework. Human capital has been widely discussed in the literature, with education and experience being two of the most important forms of investment. Becker (1964) observed that education level tends to be positively correlated to earnings, while Mincer (1974) demonstrates years of experience rather than age, along with schooling, explains the wage differential.

To estimate the causal effect of returns to education, the reduced-form approach is the most common method. However, estimates from ordinary least squares (OLS) are often argued to be upward biased. To address the endogeneity issue, J. D. Angrist and Krueger (1991) employed the quarter of birth as an instrument to support the evidence that longer compulsory schooling results in higher wages, despite the insignificant difference between 2LSL and OLS estimators. Later, J. Angrist and Krueger (1992) used the draft lottery to isolate the variation in school enrollment and found that an additional year of schooling increases weekly earnings by 6.6 percent. The similar positive relation is also documented in Card (1993). All these studies suggest little evidence of ability bias in conventional OLS estimates. Card (1999, 2001) show that marginal returns to education are shaped more by supply-side factors—particularly the marginal cost of schooling—than by low ability.

Willis and Rosen (1979) adopted a structural model approach to support the hypothesis that expected earnings influence the decision to attend college. Positive selection at two levels of schooling echoes comparative advantage rather than ability bias. In the dynamic framework, Keane and Wolpin (1997) developed a dynamic human captial investment model to analyze schooling, work, and occupational decisions of the younger generation in the U.S. The model provides an excellent fit to observed data, including lifetime decisons and wage profiles. Concurrently, Belzil and Hansen (1997) proposed an optimal stopping model that incorporates learning costs, parental support, school ability, and labor market ability. The specification also captures non-wage benefits as a key determinants in career decisions.

In the context of Taiwan, Hsu (2021) used Heckman's two-stage selection model combined with a difference-in-differences strategy to validate the college expansion lowers the wages of university graduates. In contrast, Li (2009) leveraged the Panel Study of Family Dynamics (PSFD) with 2SLS, HT, and panel GMM to construct consistent estimators, finding an insignificant negative effect on returns as college enrollment rates increased. Chuang and Lai (2011) employed quantile regression to estimate heterogeneous returns to education across birth cohorts, but the underlying mechanisms remain unexplored due to data limitations.

To our knowledge, few studies have examined career paths and wage accumulations of young adults in Taiwan. We hope this basic human capital model can serve as a starting point for understanding the dynamic patterns and offering insights into education and labor market policy.

3. The Model

Let us start with a simple single-agent discrete choice model. Assume people only care about monetary factors. Each individual i faces a finite horizon problem from age $t = 18$ to $T = 30$.

The flow payoff function is:

$$U(X_{it}, \varepsilon_{it}, a_{it}) = \begin{cases} \log\left(1 + \beta_1 s_{it} + \beta_2 x_{it}\right) + \varepsilon_{it}(1) & \text{if } a_{it} = 1 \\ -C \cdot x_{it} + \varepsilon_{it}(0) & \text{if } a_{it} = 0, \end{cases} \tag{1}$$

where $U(.)$ is the flow utility, s represents years of tertiary education, and x denotes occupational experience. The term ε_{it} is an idiosyncratic shock across time and individuals, which is observed only by the agent. The opportunity cost of education, C, depends on the current level of occupational experience. That is, the cost of forgoing earnings to pursue further education rises as work experience accumulates.

The law of motions for s and x evolves as follows:

$$\begin{aligned} x_{t+1} &= x_t + \mathbb{1}\{a_t = 1\}, \quad \forall t; \\ s_{t+1} &= s_t + \mathbb{1}\{a_t = 0\}, \quad s \leq \bar{s}, \ t \leq 30. \end{aligned} \tag{2}$$

Before the age of 30, if the agent chooses to work, occupational experience increases by 1 while years of schooling remain unchanged. Conversely, if the agent chooses to study, years of schooling increase by 1 while occupational experiences remain the same. Note that $\bar{s}$ denotes the maximum years of higher education, which we set to 12 for simplicity.

Our wage process follows the Mincer equation:

$$w_{it} = \beta_0 + \beta_1 s_{it} + \beta_2 x_{it} + \beta_3 x_{it}^2 + \xi_{it}, \tag{3}$$

The intercept term, β_0, represents the fixed skill endowment and the baseline wage offer with only a high school degree and no work experience. The coefficent of quadratic term, β_3, reflects the diminishing marginal returns of experience on earnings and is expected to be negative. The β_1, β_2 are the two main parameters of interest. To avoid identification issue, we normalize $\beta_0 = 1$ and $\beta_3 = 0$.

The error term, $\xi_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\xi^2)$, is the heterogeneity of productivity across individual. Since employers in the labor market cannot observe the error term, they pay each employee (i.e., agent) the average earnings conditional on personal state:

$$E[w_{it}|X_{it}] = X_{it}'\beta, \tag{4}$$

where $X_{it} = [1, s_{it}, x_{it}]$, $\beta = [1, \beta_1, \beta_2]'$. Agents' choice set consists of two alternatives: pursuing a job ($a_{it} = 1$) and keeping studying ($a_{it} = 0$). When an individual decides to work, he can always find a job and receive a wage offer $X_{it}'\beta$; once an agent decides to pursue further education, they incur an opportunity cost of $C \cdot x_{it}$, which may include forgone earnings or college tuition. We also assume there are no friction between the labor market and schooling.

Let $\delta \in (0, 1)$ be the discount factor. At each period t, agents aims to maximize the

lifetime expected utility by choosing $a_{it} \in \{0, 1\}$:

$$\max_{a_t \in \{0,1\}} \mathbb{E}\left(\sum_{\tau=0}^{T-t} \delta^\tau \{u(X_{t+\tau}, a_{t+\tau}) + \varepsilon_{t+\tau}(a_{t+\tau})\} \mid X_t, \varepsilon_t, a_t \right). \tag{5}$$

Note that we ignore the subscript i since every agent follows the same decision rule. Rewrite the maximization problem as Bellman equation:

$$V(X_t, \varepsilon_t) = \max_{a_t \in \{0,1\}} \left[u(X_t, a_t) + \varepsilon_t(a_t) + \delta \iint V(X_{t+1}, \varepsilon_{t+1}) dG(\varepsilon_{it}) dF(X_{t+1} \mid X_t, a_t) \right]. \tag{6}$$

3.1. Identification

The solution to this basic model involves in computing the optimal lifetime decision paths $\{a_t^*\}_{t=18}^T$ for each agent. Our goal is to recover the deep parameters $\{\beta_1, \beta_2, C, \delta, \sigma_\xi\}$. In the basic model, we only focus on the first three parameters and calibrate the discount factor at $\delta = 0.95$. Recall that agents in our model receive only average earnings. Although this assumption is somewhat unrealistic simplification, it allows us to sidestep the difficult task of estimating ability heterogeneity.

Following Rust (1987), we assume the unobservable component ε_{it} follows Type I Extreme Value (T1EV) distribution, and we can define an integrated value function in the state of X_t before the realization of random shock:

$$\bar{V}(X_t) = \int_\varepsilon V(X_t, \varepsilon_t) g(\varepsilon_t) d\varepsilon_t = \mathbb{E}_\varepsilon \left[V(X_t, \varepsilon_t) \right]. \tag{7}$$

By Assumption (A.2), the integrated value function can be derived as:

$$\begin{aligned}
\mathbb{E}_\varepsilon \left[V(X_t, \varepsilon_t) \right] &= \mathbb{E}_\varepsilon \left[\max_{a_t} \{ \varepsilon(a_t) + u(X_t, a_t) + \delta \mathbb{E}(V(X_{t+1}, \varepsilon_{t+1})) \} \right] \\
&= \ln \left\{ \sum_{a_t} \exp \left(u(X_t, a_t) + \mathbb{E}_{X_{t+1}}[\bar{V}(X_{t+1}) | X_t, a_t] \right) \right\} + \gamma,
\end{aligned} \tag{8}$$

where γ is Euler-Mascheroni constant.

Note that in equation (8), we can define the choice-specific value function which does not depend on ε_t:

$$\begin{aligned}
v(X_t, a_t) &\equiv u(X_t, a_t) + \delta \int \bar{V}(X_{t+1}) df(X_{t+1} \mid X_t, a_t) dX_{t+1} \\
&= u(X_t, a_t) + \delta \mathbb{E}_{X_{t+1}}[\bar{V}(X_{t+1}) | X_t, a_t].
\end{aligned} \tag{9}$$

Last, we can derive the conditional choice probability (CCP) of choosing action a_t given

state X_t:

$$\Pr(a_t = j \mid s_t, x_t) = \Pr(v_j + \varepsilon(j) > v_k + \varepsilon(k), k \neq j \mid s_t, x_t)$$
$$= \frac{\exp(v_j)}{\exp(v_1) + \exp(v_0)}, \quad j, k \in \{0, 1\}, \tag{10}$$

where $s \in [0, 12]$ and $x \in [0, 12]$.

3.2. Data Generation Process

Individuals decision process begins at age 18, after graduating from senior high school. In each period, they draw two independent random shocks from the T1EV distribution and compute the lifetime expected reward associated with each option. After comparing the two alternatives, agents choose the action that yields the higher utility and update the state variables according to (2).

For simplicity, we assume that after the age of 30, all agents exclusively choose to work. At this stage in life, schooling no longer increases job earnings, and the option to study is no longer available. Consequently, the decision to work becomes an absorbing state, as agents remain in this state for the remainder of their lives, with occupational experiences increasing by one each period.

Table (1) documents an overview of the descriptive statistics from the simulated data. At the aggregated level, most individuals choose to work, while only a few opt to pursue higher education, with the maximum years of schooling reaching 9.

Figure (1) shows the probability of choosing higher education versus entering the labor market by age. Monotonic patterns are observed for both choices: individuals are more likely to attend college at younger ages, and they gradually enter the workforce after completing the studies.

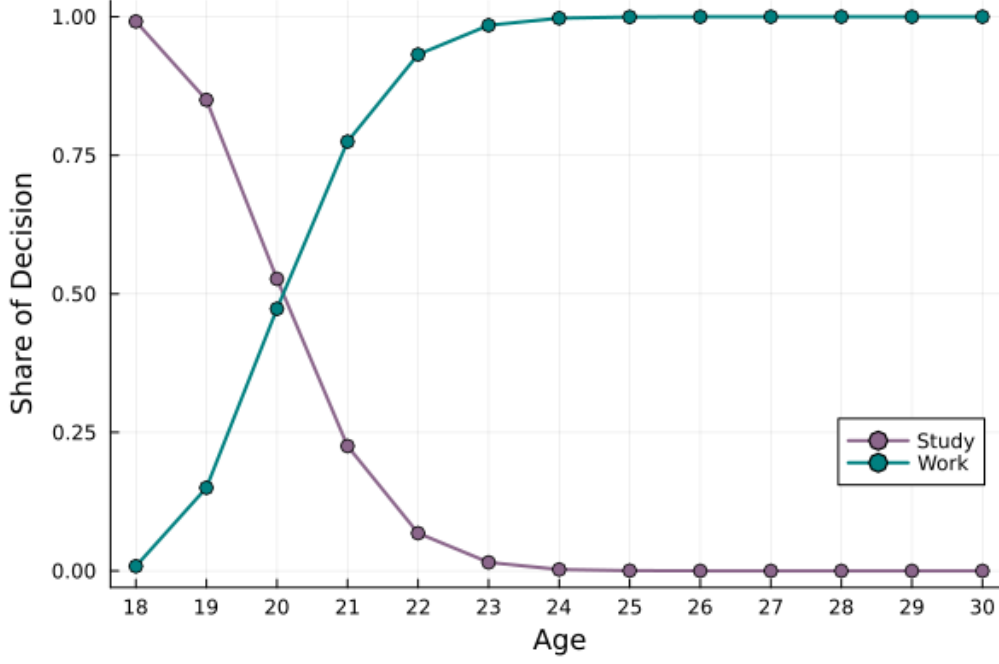|	mean	min	median	max
years of higher education	2.2474	0	2	9
work experiences	3.7526	0	3	12
% choosing to work	79.823%	-	-	-

Table 1: Summary Statistics (N = 30000)

Figure 1: Age Profile of Study and Work Decisions

4. Estimation

4.1. NFXP

Given the simulated data, we use nested fixed point (NFXP) algorithm to back up the parameters of the model. The algorithm consists of two loops. We begin with an initial set of arbitrary parameters, $\hat{\theta}_0 = \{\hat{\beta}_1, \hat{\beta}_2, \hat{C}\}$, and solve $\bar{V}(X_t; \hat{\theta}_0)$ in (8) via value function iteration, which helps us construct the CCP in (10). In the outer loop, we use maximum likelihood estimation to update the parameters to $\hat{\theta}_1$. The process proceeds iteratively until convergence. The likelihood function is:

$$\mathcal{L}(\theta \mid X_{it}, a_{it}) = \prod_{n=1}^{N} \prod_{t=18}^{30} \prod_{j=1}^{2} \Pr(a_{it} = j \mid X_{it})^{\mathbf{1}\{a_{it}=j\}}. \tag{11}$$

4.2. Preliminary Results

Table (2) summarizes the predetermined parameters and the estimation results.

5. Policy Experiment

In this section, we use the estimated model to conduct a policy experiment. Suppose the government implements a subsidy program of the fixed amount $S = 1.5$ to encourage individuals

		predetermined value	estimates
Schooling Effect	β_1	3.6	3.674
Work Experiences Effect	β_2	1.79	1.821
Cost of Schooling	C	2.58	2.561

Table 2: Estimates of Model Parameters

to pursue higher education. The flow utility function is modified as follows:

$$U(X_{it}, \varepsilon_{it}, a_{it}) = \begin{cases} \log(X_{it}\beta) + \varepsilon_{it}(1) & \text{if } a_{it} = 1 \\ S - C \cdot x_{it} + \varepsilon_{it}(0) & \text{if } a_{it} = 0, \end{cases} \tag{12}$$

and individuals solve the Bellman equation in (5) with this new flow utility function.

Figure (2) shows that the subsidy policy provides a strong incentive for individuals to pursue higher education, leading to a higher probability of completing master's or PhD degrees. The monotonic patterns remain: once individuals enter the labor market, they are unikely to return to school.

Figure (3) compares the lifetime utility, as characterized by wage profiles, under the estimated model and the counterfactual scenario. With the subsidy, more individuals attain tertiary education and accumulate additional years of schooling. This leads to an upward shift in the wage profile since schooling effect in our model is larger than the experience effect ($\beta_1 > \beta_2$). However, the policy's effectiveness should be discussed with caution, as this basic model does not account for the government's budget and the potential social cost.

Supplementary Material
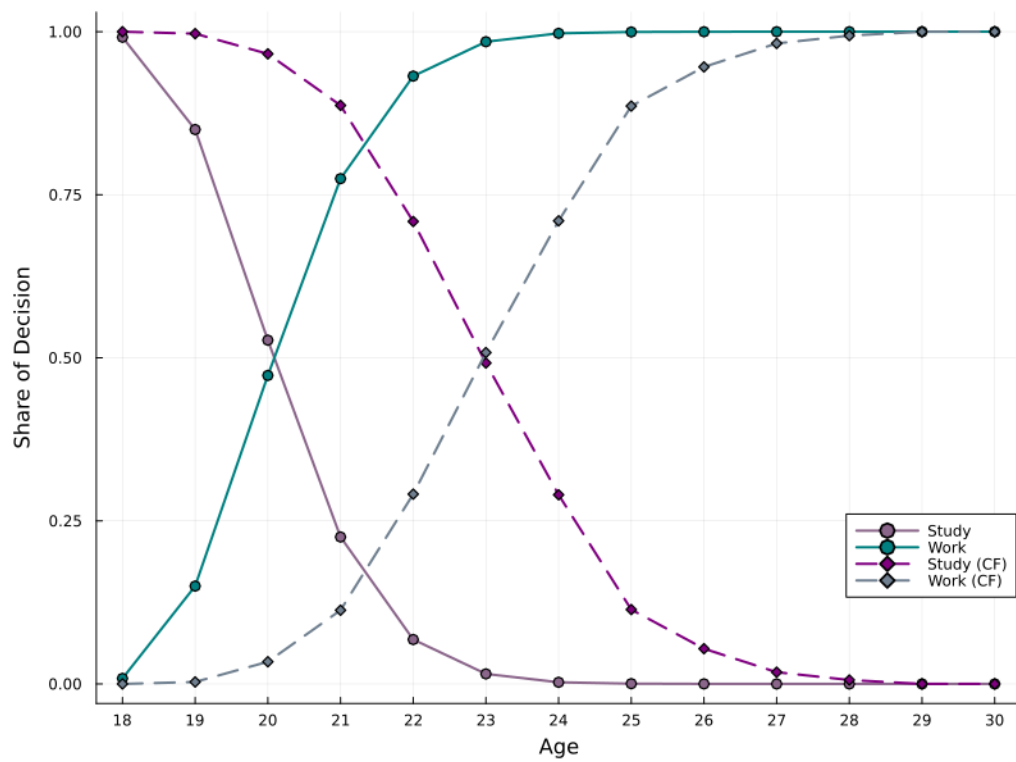
The replication code is available on GitHub.

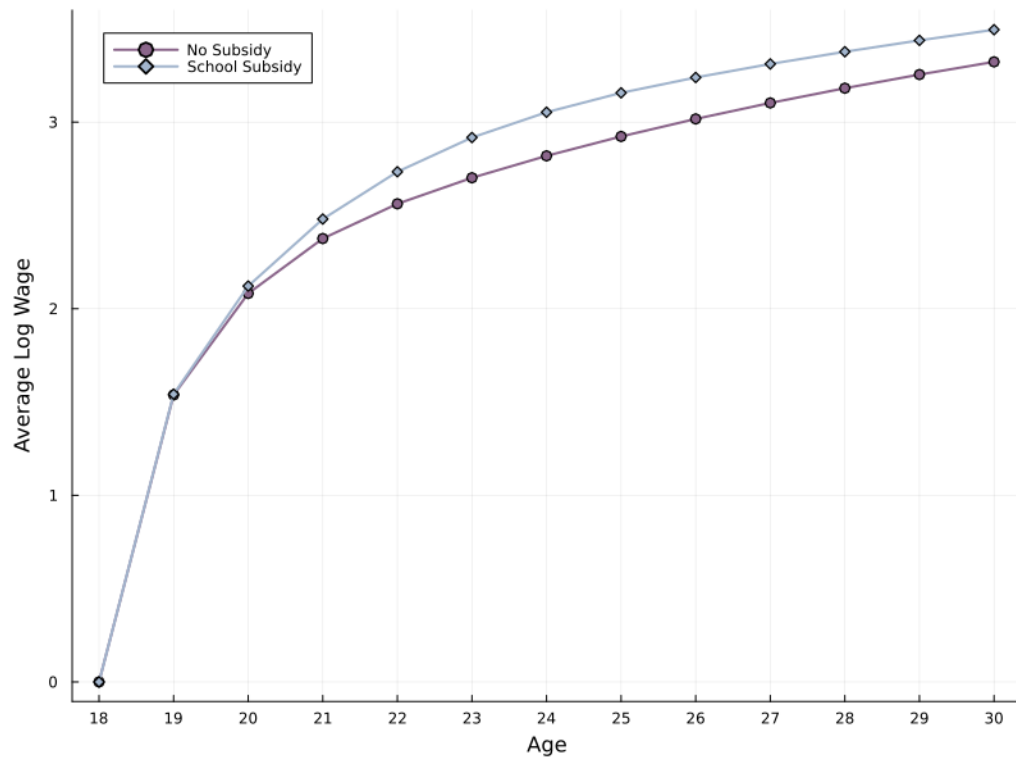Figure 2: Age Profile of Study and Work Decisions: Baseline v.s. Counterfactual

Figure 3: Wage Profile: Baseline v.s. Counterfactual

A. Rust Assumptions

We follow the standard assumptions in Rust (1987) to solve the dynamic discrete choice model.

Assumption A.1 (Additive Separability)

The one-period utility function is additively separable:

$$U(X_{it}, \varepsilon_{it}, a_{it}; \beta) = u(X_{it}, a_{it}; \beta) + \varepsilon_{it}(a_{it}). \tag{13}$$

Assumption A.2 (IID T1EV shocks)

The unobserved components are i.i.d. across alternatives, agents and time with T1EV CDF $G(.)$; The unobserved shocks are contemporarily independent of X_{it}:

$$\varepsilon_{it}(a_{it}) \overset{iid}{\sim} G_{\varepsilon}(\varepsilon_{it})$$
$$X_{it} \perp \varepsilon_{it}. \tag{14}$$

Assumption A.3 (Conditional Independence of Future States; CI-X)

The future state $\{X_{it+1}, \varepsilon_{it+1}\}$ depends only on the current state and action:

$$P(X_{it+1}, \varepsilon_{it+1} \mid X_{it}, a_{it}, \varepsilon_{it}) = P(X_{it+1}, \varepsilon_{it+1} \mid X_{it}, a_{it})$$
$$= F_X(X_{it+1} \mid X_{it}, a_{it}) G(\varepsilon_{it+1}) \tag{15}$$

Note that in the second equality we use the Assumption (A.2).

Assumption A.4 (Discrete Support of X_{it}; DIS)

The state variables X_{it} (years of schooling s & occupational experience x) have a discrete and finite support:

$$X_{it} \in \{X^{(1)}, X^{(2)}, \ldots, X^{(|X|)}\}, \quad |X| < \infty. \tag{16}$$

That is,

$$s_{it}, x_{it} \in \{0, 1, 2, \ldots, 12\}. \tag{17}$$

References

Angrist, J., & Krueger, A. B. (1992). *Estimating the payoff to schooling using the vietnam-era draft lottery*. National bureau of economic research Cambridge, Mass., USA.

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The quarterly journal of economics*, *106*(4), 979–1014.

Becker, G. (1964). *Human capital: A theoretical and empirical analysis with special reference to education* (Tech. Rep.). National Bureau of Economic Research, Inc.

Belzil, C., & Hansen, J. (1997). Estimating the returns to education from a non-stationary dynamic programming model.

Card, D. (1993). *Using geographic variation in college proximity to estimate the return to schooling*. National Bureau of Economic Research Cambridge, Mass., USA.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics/Elsevier*.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, *69*(5), 1127–1160.

Chuang, Y.-C., & Lai, W.-W. (2011). A cohort analysis of returns to education with heterogeneous ability: The case of taiwan. *Taiwan Economic Review*, *39*(1), 81–113.

Hsu, M. (2021). Higher education expansion policy in taiwan: Effects on the wages of university graduates and the overeducated. *Taiwan Economic Forecast & Policy*, *51*(2).

Keane, M. P., & Wolpin, K. I. (1997). The career decisions of young men. *Journal of political Economy*, *105*(3), 473–522.

Li, J. (2009). *The returns to education in taiwan: With consideration of the impact of higher education expansion* (Master's Thesis). National Tsing Hua University, Department of Economics.

Mincer, J. (1974). Schooling, experience, and earnings. human behavior & social institutions no. 2.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, 999–1033.

Willis, R. J., & Rosen, S. (1979). Education and self-selection. *Journal of political Economy*, *87*(5, Part 2), S7–S36.