Supplementary Material

Supplement to "Dynamic skill accumulation, education policies, and the return to schooling"

(*Quantitative Economics*, Vol. 8, No. 3, November 2017, 895–927)

Christian Belzil
CREST, CNRS, Department of Economics, Ecole Polytechnique, Université de Paris-Saclay and IZA

Jorgen Hansen
Department of Economics, Concordia University, CIRANO, CIREQ, and IZA

Xingfei Liu
Department of Economics, University of Alberta and IZA

S.1. The data

The model has been estimated using a sample of white males from the 1979 youth cohort of the National Longitudinal Survey of Youth (NLSY). We restrict our sample to males from the core random sample who were 14–16 years old in 1979. We record information on education, wages, employment status, and hours of work for each individual from the time the individual is 16 up to age 30.

To capture differences in cognitive ability and to facilitate identification, we use information from AFQT scores, purged of possible effects of different educational attainment at the time of the test.

The wage information was obtained from self-reported pay at the main job and, unless reported as an hourly rate, converted to a rate per hour. Wages are adjusted for inflation using the consumer price index (CPI) and expressed in 1997 dollars. In Table S.1, we present observed and simulated average wage rates for each age between age 20 and 30. The simulated wage outcomes are similar to the observed ones for all ages, suggesting that the model is capable of generating wage outcomes that closely resemble the observed ones.

As mentioned above, we discretize hours of work by grouping actual observed hours into three categories: low, medium, and high intensity labor supply. The corresponding thresholds for the three categories are 2,000 hours per year and 2,500 hours per year, respectively. The observed distribution of these hours classes is shown in Table S.3 and it shows that there is substantial movement across the categories. In particular, the fraction of our sample that is working less than 2,000 hours per year is over 50 percent at

Christian Belzil: christian.belzil@gmail.com
Jorgen Hansen: Jorgen.Hansen@concordia.ca
Xingfei Liu: feiya66@gmail.com

DOI: 10.3982/QE321

Table S.1. Observed and simulated hourly wages.

Age	Observed	Simulated
20	9.292	8.803
	(3.987)	(3.656)
21	9.437	9.196
	(3.798)	(3.866)
22	10.066	9.657
	(4.278)	(4.115)
23	11.067	10.205
	(4.900)	(4.448)
24	12.485	10.840
	(8.116)	(4.871)
25	13.599	11.500
	(10.911)	(5.304)
26	14.651	12.130
	(13.322)	(5.703)
27	14.259	12.738
	(8.909)	(6.109)
28	15.065	13.425
	(12.250)	(6.537)
29	15.452	14.103
	(10.754)	(7.021)
30	16.101	14.798
	(13.250)	(7.471)

Note: Observed wages are conditional on being observed in the data, while simulated wages are conditional on positive simulated working hours. Standard deviations are given in parentheses.

Table S.2. Annual growth in observed hourly wages and earnings.

	Hourly Wages	Earnings
Education		
Less than high school	0.024	0.066
High school graduates	0.049	0.080
Some college	0.081	0.105
College graduates	0.098	0.147
AFQT (−1st dev)	0.026	0.091
AFQT (mean)	0.073	0.096
AFQT (+1st dev)	0.074	0.137

Note: The observed growth rates are conditional on observed quantities at both age 25 and age 30.

age 20 but only 19 percent at age 30. For the other labor supply classes, the opposite is observed and the fractions in medium and high intensity labor supply groups increase with 13 and 21 percentage points, respectively.

Table S.3. Observed annual hours of work by age and category.

Age	Low Hours	Medium Hours	High Hours	Average Hours
16	0.931	0.000	0.069	960
17	0.851	0.095	0.054	1,131
18	0.651	0.280	0.069	1,527
19	0.614	0.269	0.117	1,685
20	0.531	0.339	0.130	1,770
21	0.474	0.372	0.153	1,856
22	0.411	0.393	0.196	1,933
23	0.337	0.429	0.234	2,071
24	0.298	0.476	0.226	2,140
25	0.264	0.487	0.249	2,198
26	0.242	0.493	0.264	2,211
27	0.217	0.491	0.292	2,241
28	0.210	0.483	0.306	2,256
29	0.215	0.482	0.303	2,271
30	0.190	0.472	0.338	2,320

Note: The low hours class corresponds to 1–1,999 hours per year, the medium hours class corresponds to 2,000–2,499 hours per year, and the high hours class corresponds 2,500 hours per year or more.

S.2. Estimation method

We now provide details about our estimation method. Recall that the elements of the vector of utility error terms $\{\varepsilon^s_{it}, \varepsilon^r_{it}, \varepsilon^h_{it}, \varepsilon^m_{it}, \varepsilon^l_{it}\}$ are assumed to be i.i.d. and to follow an extreme-value distribution. At each period t, the individual takes a decision based on the information set, which includes the random shocks and accumulated periods in each state:

$$\Omega_t = \left\{\varepsilon^s_t, \varepsilon^h_t, \varepsilon^m_t, \varepsilon^l_t, \varepsilon^w_t, S_t, R_t, L_t, M_t, H_t\right\}.$$

We model choices from age 16 onward over a total time horizon equal to 15 years (until age 30). For each possible choice, there is a specific value function

$$V^k_t(\Omega_t) = U^k_t + \beta EV_{t+1}(\Omega_{t+1} \mid d_{kt} = 1) \quad \text{for } k = s, r, l, m, h,$$

where

$$EV_{t+1}(\Omega_{t+1} \mid d_{kt} = 1) = E\max\left\{V^s_{t+1}(\cdot), V^r_{t+1}(\cdot), V^l_{t+1}(\cdot), V^m_{t+1}(\cdot), V^h_{t+1}(\cdot) \mid d_{kt} = 1\right\},$$

where β is the discount factor and where Ω_t is the set containing all state variables known by the agent at t.

Given these three computational issues, we decided to use the same approximation method as Sauer (2015), which is itself based on the method proposed by Geweke and Keane (2000).[1] We now describe the main steps undertaken toward estimation.

[1] One potential advantage of this approach is that it renders the allowance for correlated utility shocks possible without having to use a discretized distribution of the error terms. This is achieved in Sauer (2015).

- First, we set the initial value of ε^w_{t-1} (at age 16) to 0 for all individuals.

- We simulate the distribution of ε^w_{it} using 30 antithetic draws for each individual-period combination.

- Given our assumptions and defining $\bar{U}^k_t(\alpha)$ as the deterministic part of the per-period utility conditional on one realization (α) of ε^w_{it}, the expected maximum utility achievable in period t is equal to

$$EV_{\alpha,t} = \tau\left(\gamma + \ln\left\{\sum_{j=s,r,l,m,h} \exp\left(\frac{\overline{U}^j_t(\alpha) + \beta EV_{t+1}(\Omega_{t+1} \mid d_{jt} = 1)}{\tau}\right)\right\}\right),$$

where τ is the scale of the extreme-value distribution and γ is Euler's constant. The expected value function, EV_t, is obtained by averaging each $EV_{\alpha,t}$ over a total number of draws.

- Conditional on one realization (α), the choice probabilities, denoted $\Pr(d_{kt} = 1; \alpha)$, are expressed as

$$\Pr(d_{kt} = 1; \alpha) = \frac{\exp\big(\overline{U}^k_t(\alpha) + \beta EV_{t+1}(\Omega_{t+1} \mid d_{kt} = 1)\big)}{\sum_{j=s,r,l,m,h} \exp\big(\overline{U}^j_t(\alpha) + \beta EV_{t+1}(\Omega_{t+1} \mid d_{jt} = 1)\big)}$$

for $k = s, r, l, m, h$.

- To facilitate estimation of complicated dynamic discrete choice models, we follow Sauer (2015) and adopt a solution method that borrows from Geweke and Keane (2000), who have proposed to replace the future component of the value function by a flexible polynomial in state variables. Their approach is particularly well suited to frameworks where the econometrician has access to data on choices and outcomes. Geweke and Keane (2000) actually show from various numerical applications to artificial data that specifying the future component as a flexible polynomial (i) has negligible effects of the estimated values of the parameters of the payoff functions and (ii) the misspecified rule inferred from the data is itself very close to the actual optimal rule. As is done in Sauer (2015), we adjust the Geweke–Keane solution approach so as to incorporate more model structure in our estimation strategy.[2] At any period t, the future component of the intertemporal utility, $EV_{t+1}(\Omega_{t+1} \mid d_{kt} = 1, \Omega_t)$, is represented by the expression

$$EV_{t+1}(\Omega_{t+1} \mid d_{kt} = 1, \Omega_t) = E \max_k \big\{U^k_{t+1}(\Omega_{t+1}) + \digamma\big(\Omega_{t+2}(\Omega_{t+1}, d_{kt+1})\big)\big\},$$

where $\digamma(\Omega_{t+2}(\cdot))$ is a flexible polynomial in state variables reflecting all potential choices in $t+1$. Our approach differs precisely from the approach suggested by Geweke and Keane (2000) in that the imbedded polynomial of the state space intervenes in $t+2$ as opposed to directly in $t+1$. This allows us to incorporate more model structure than in the Geweke–Keane method.

[2]Compared to Sauer (2015), our model contains a smaller number of potential choices but it is estimated over a much longer period and also incorporates a richer heterogeneity distribution.

- In practice, including state variables in levels is problematic because some of the utilities (labor supply especially) also depend on accumulated periods in each state. For this reason, we use a polynomial approximation that incorporates squared terms as well as interactions. The polynomial equation used in the paper is

$$\begin{aligned} F(\cdot) = {} & \varrho_1 \cdot S_{i,t+2}^2 + \varrho_2 \cdot R_{i,t+2}^2 + \varrho_3 \cdot L_{i,t+2}^2 + \varrho_4 \cdot M_{i,t+2}^2 + \varrho_5 \cdot H_{i,t+2}^2 + \varrho_6 \cdot S_{i,t+2} \cdot R_{i,t+2} \\ & + \varrho_7 \cdot S_{i,t+2} \cdot L_{i,t+2} + \varrho_8 \cdot S_{i,t+2} \cdot M_{i,t+2} + \varrho_9 \cdot S_{i,t+2} \cdot H_{i,t+2} + \varrho_{10} \cdot R_{i,t+2} \cdot L_{i,t+2} \\ & + \varrho_{11} \cdot R_{i,t+2} \cdot M_{i,t+2} + \varrho_{12} \cdot R_{i,t+2} \cdot H_{i,t+2} + \varrho_{13} \cdot L_{i,t+2} \cdot M_{i,t+2} \\ & + \varrho_{14} \cdot L_{i,t+2} \cdot H_{i,t+2} + \varrho_{15} \cdot M_{i,t+2} \cdot H_{i,t+2}, \end{aligned}$$

where the ϱs are parameters to be estimated.

For each individual i at date t, there is a vector of observed outcomes $O_{it} = \{d_{ist}, d_{irt}, d_{ilt}, d_{imt}, d_{iht}, w_{it}^O\}$, where w_{it}^O denotes observed wage outcome. To estimate the model, we set M to four and normalize τ to 1.

The likelihood function for individual i is given by

$$L_i(\cdot) = \sum_{m=1}^{M} \prod_{t=1}^{T} \Pr(O_{it} \mid \mathit{type}\ m) \Pr(\mathit{type}\ m).$$

The total likelihood is the product of each $L_i(\cdot)$ over 1,199 individuals. Structural parameters are obtained by maximizing the logarithm of the likelihood function using Fortran routines.

Co-editors Orazio Attanasio and Petra E. Todd handled this manuscript.

Manuscript received 29 October, 2012; final version accepted 28 December, 2016; available online 26 January, 2017.