

Supplement to “Wandering astray: Teenagers’ choices of schooling and crime”

(*Quantitative Economics*, Vol. 13, No. 2, May 2022, 387–424)

CHAO FU

Department of Economics, University of Wisconsin and NBER

NICOLÁS GRAU

Faculty of Economics and Business, University of Chile

JORGE RIVERA

Faculty of Economics and Business, University of Chile

1. MODEL SOLUTION: DETAILS

To solve our dynamic discrete choice model, which involves a very high-dimensional state space, we follow the method developed by Keane and Wolpin (1994). The key aspect of this approach consists of approximating the continuation payoff function via an integration and interpolation procedure. Given the continuation payoffs at any period and at any point of the state space, the dynamic problem is solved via backward induction. In the following, we describe in detail how we approximate the continuation payoff function in our model setting.

The state space contains two types of state variables $\Omega_{it} = \{\Omega_i^f, \Omega_{it}^{nf}\}$, those that are fixed throughout for the same teenager across time (Ω_i^f), and those evolve over time (Ω_{it}^{nf}). In particular, there are 612 possible combinations of variables in Ω_i^f : 3 levels of parental education, 3 levels of family income, 34 locations/municipalities, and 2 unobserved types ($612 = 3 \times 3 \times 34 \times 2$). Because these variables are discrete and do not have dynamics, that is, they are fixed from the beginning of the dynamic problem, we solve one dynamic problem for each possible combination of these variables. That is, we use the Keane and Wolpin (1994) Monte Carlo integration and interpolation method to solve 612 different dynamic programming problems.

For each of the 612 cases, we simulate and interpolate the other elements in the state space, which include: (1) the beliefs on ability (EA_{it-1} and VA_{it-1}), (2) the total number of past arrests and the total length of sentences received in the past ($J_{it-1,1}$, $J_{it-1,2}$), (3) the highest grade completed by the end of $t - 1$ (G_{it-1}), (4) whether or not one progressed

Chao Fu: cfu@ssc.wisc.edu

Nicolás Grau: ngrau@fen.uchile.cl

Jorge Rivera: jrivera@fen.uchile.cl

one grade between $t - 1$ and t (g_{it}), 5) whether or not one was enrolled last period (e_{it-1}), and 6) the characteristics of the school one attends. We use Monte Carlo integration to calculate the continuation payoffs at a subset of state variables in the support of Ω_{it}^{nf} , which are then used to interpolate the continuation payoffs for other possible values of Ω_{it}^{nf} via an OLS regression. In this OLS regression, the independent variables include all the dynamic state variables in Ω_{it}^{nf} , their squared terms (unless the variable is binary), and their interactions.

Specifically, for a given combination of parental education, family income group, location, and unobserved type (Ω_i^f), that is, one of the 612 cases, we conduct the following procedure to calculate the continuation payoff at each period t :

1. **(Simulation)** Given a particular Ω_i^f , for each period, we randomly draw $S = 500$ realizations of the state space vector in Ω_{it}^{nf} . We calculate the value function associated with each, resulting a matrix of 500 simulated value functions and state variable values $\{V_{it}(\Omega_i^f, \Omega_{it,s}^{nf}), \Omega_{it,s}^{nf}\}_{s=1}^{500}$ for a given Ω_i^f .
2. Let K_{nf} be the length of the vector Ω_{it}^{nf} , such that $\Omega_{it,s}^{nf} = \{\Omega_{it,s,m}^{nf}\}_{m=1}^{K_{nf}}$. Given the matrix $\{V_{it}(\Omega_i^f, \Omega_{it,s}^{nf}), \Omega_{it,s}^{nf}\}_{s=1}^{500}$ calculated in step 1, we run the following OLS regression:

$$V_{it}(\Omega_i^f, \Omega_{it,s}^{nf}) = \beta_0^t + \sum_{m=1}^{K_{nf}} \beta_m^t \Omega_{it,s,m}^{nf} + \sum_{m=1}^{K_{nf}} \beta_{K_{nf}+m}^t (\Omega_{it,s,m}^{nf})^2 + \sum_{m \neq m'} \beta_{m,m'}^t \Omega_{it,s,m}^{nf} \Omega_{it,s,m'}^{nf} + \varepsilon_s,$$

and obtain $\hat{\beta}^t$ as an estimate of the vector β^t , i.e., the time-specific OLS coefficient vector.

3. **(Interpolation)** Given any value of Ω_i^f and $\tilde{\Omega}_{it}^{nf}$, where the latter is a particular value of time-varying state variables that are not drawn in step 1, we approximate the continuation payoff as

$$\hat{V}_{it}(\Omega_i^f, \tilde{\Omega}_{it}^{nf}) = \hat{\beta}_0^t + \sum_{m=1}^{K_{nf}} \hat{\beta}_m^t \tilde{\Omega}_{it,m}^{nf} + \sum_{m=1}^{K_{nf}} \hat{\beta}_{K_{nf}+m}^t (\tilde{\Omega}_{it,m}^{nf})^2 + \sum_{m \neq m'} \hat{\beta}_{m,m'}^t \tilde{\Omega}_{it,m}^{nf} \tilde{\Omega}_{it,m'}^{nf}$$

2. ADDITIONAL TABLES

See Table B1.

TABLE B1. Municipalities characteristics.

	Municipalities included		Other municipalities	
Population	246,597	(101,932)	29,882	(33,541)
Average age	31.3	(2.2)	32.0	(1.8)
Adult education: Less than high school	0.32	(0.07)	0.47	(0.10)
Adult education: High school	0.40	(0.06)	0.38	(0.05)
Adult education: More than high school	0.27	(0.11)	0.16	(0.08)
Num. of crimes (per 1000 people per year)	62.4	(23.3)	51.4	(24.5)
Num. of arrests (per 1000 people per year)	11.7	(3.9)	7.8	(4.0)
Num. of municipalities		34		307*

Note: Rows 1–5: 2017 Census. Rows 6–7: Aggregate crime/arrest data from Ministry of Interior 2017. Cross-municipality standard deviations are in parentheses.

*There are 345 municipalities in Chile, info is missing for 4 very small municipalities.

REFERENCES

Keane, Michael and Kenneth Wolpin (1994), "The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence." *The Review of Economics and Statistics*, 76, 648–672. [1]

Co-editor Peter Arcidiacono handled this manuscript.

Manuscript received 14 September, 2020; final version accepted 1 September, 2021; available online 28 September, 2021.