Estimation of Dynamic Discrete Choice Models by Maximum Likelihood and the Simulated Method of Moments

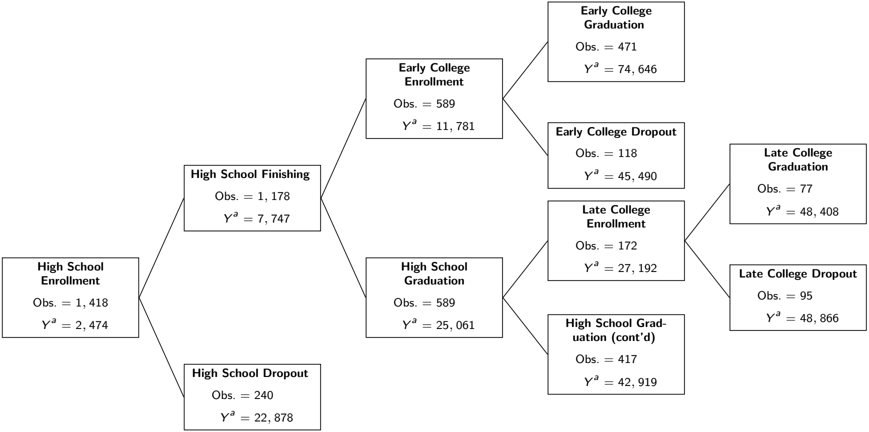
Philipp Eisenhauer, James J. Heckman, & Stefano Mosso
International Economic Review, 2015

Econ 312, Spring 2019

Structural Dynamic Discrete Choice Model of Schooling

Model

Figure 1: Decision Tree

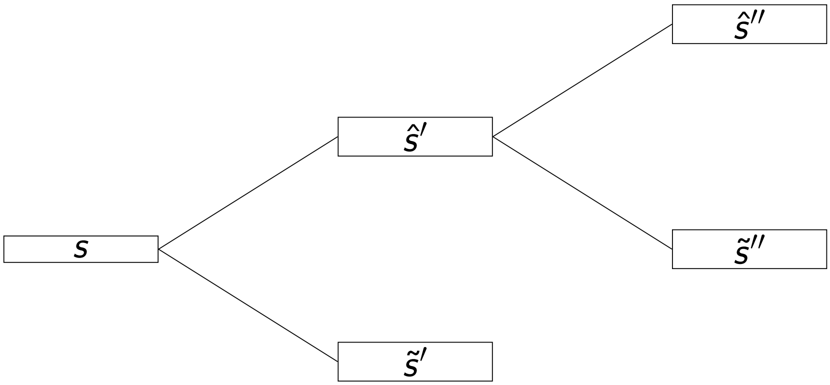


Notes: Y^a refers to average annual earnings in the state in 2005 dollars. Obs. refers to the number of observations in the state.

Setup

- Current state $s \in \mathcal{S} = \{s_1, \ldots, s_N\}$.
- $\mathcal{S}^v(s) \subseteq \mathcal{S}$: set of visited states.
- $\mathcal{S}^f(s) \subseteq \mathcal{S}$ the set of feasible states that can be reached from s.
- Choice set of the agent in state s:
  $\Omega(s) = \{s' \mid s' \in \mathcal{S}^f(s)\}$.
- Consider binary choices only, so $\Omega(s)$ has at most two elements.
- *Ex post*, the agent receives per period rewards
  $R(s') = Y(s') - C(s', s)$.
- Costs $C(s', s)$ associated with moving from state s to state s'.

Figure 2: Generic Decision Problem



Payoffs and Costs

$$Y(s) = \quad \mu_s(X(s)) + \theta'\alpha_s + \epsilon(s) \tag{1}$$

$$C(s', s) \quad = \quad \begin{cases} K_{\hat{s}',s}(Q(\hat{s}', s)) + \theta'\varphi_{\hat{s}',s} + \eta(\hat{s}', s) & \text{if} \quad s' = \hat{s}' \\ 0 & \text{if} \quad s' = \tilde{s}' \end{cases} \tag{2}$$

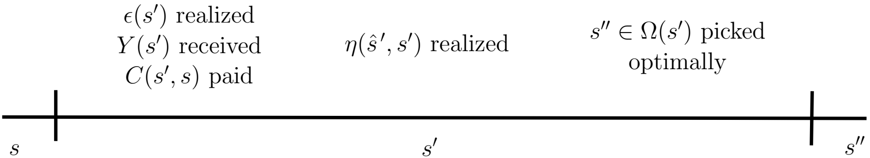
System of Measurement Equations

$$M(j) \quad = \quad X(j)'\kappa_j + \theta'\gamma_j + \nu(j) \qquad \forall \quad j \in M \tag{3}$$

θ is unobserved ability vector (cognitive and noncognitive)

Information

Timing



Information Set

$$\left.\begin{array}{ll} \text{for all } s \in \mathcal{S}^v(s) & \eta(\hat{s}', s); \epsilon(s) \\ \text{for } \hat{s}' \in \mathcal{S}^f(s) & \eta(\hat{s}', s) \\ \text{and for all } s & X(s); Q(\hat{s}', s); \theta \end{array}\right\} \in \mathcal{I}(s).$$

Value Function

$$V(s \,|\, \mathcal{I}(s)) = Y(s) + \max_{s' \in \Omega(s)} \left\{ \frac{1}{1+r} \Big(-\underbrace{C(s', s) + \mathbb{E}[V(s' \,|\, \mathcal{I}(s')) \,\big|\, \mathcal{I}(s)\,]}_{\text{Continuation value}} \Big) \right\}$$

Decision Rule

$$s' = \begin{cases} \hat{s}' & \text{if} \quad \mathbb{E}\left[V(\hat{s}') \,\middle|\, \mathcal{I}(s)\right] - C(\hat{s}', s) > \mathbb{E}\left[V(\tilde{s}') \,\middle|\, \mathcal{I}(s)\right] \\ \tilde{s}' & \text{otherwise} \end{cases}$$

Choice Probabilities

$$\Pr\Big(G(\hat{s}') = 1\Big) = F_{\eta(\hat{s}',s)}\bigg(\mathbb{E}\Big[V(\hat{s}') - V(\tilde{s}') \,\Big|\, \mathcal{I}(s)\Big] - (K_{\hat{s}',s}(Q(\hat{s}', s)) + \theta' \varphi_{\hat{s}',s})\bigg)$$

Ex Ante Net Return

$$NR(\hat{s}', \tilde{s}', s) = \frac{\mathbb{E}\left[V(\hat{s}') - V(\tilde{s}') \,\middle|\, \mathcal{I}(s)\right] - C(\hat{s}', s)}{\mathbb{E}\left[V(\tilde{s}') \,\middle|\, \mathcal{I}(s)\right]}$$

Ex Ante Gross Returns

$$GR(\hat{s}', \tilde{s}', s) = \frac{\mathbb{E}\left[\tilde{V}(\hat{s}') - \tilde{V}(\tilde{s}') \,\middle|\, \mathcal{I}(s)\right]}{\mathbb{E}\left[\tilde{V}(\tilde{s}') \,\middle|\, \mathcal{I}(s)\right]}$$

Option Value (Weisbrod, 1962)

$$OV(s', s) =$$

$$\frac{1}{1+r}\,\mathbb{E}\left[\underbrace{\max_{s''\in\Omega(s')}\left\{-C(s'',s') + \mathbb{E}\Big(V(s'')\Big)\right\}}_{\text{value of options arising from } s'} - \underbrace{\Big(V(\tilde{s}'')\Big)}_{\text{fallback value}}\;\Bigg|\;\mathcal{I}(s)\right]$$

Individual Contribution to Likelihood:

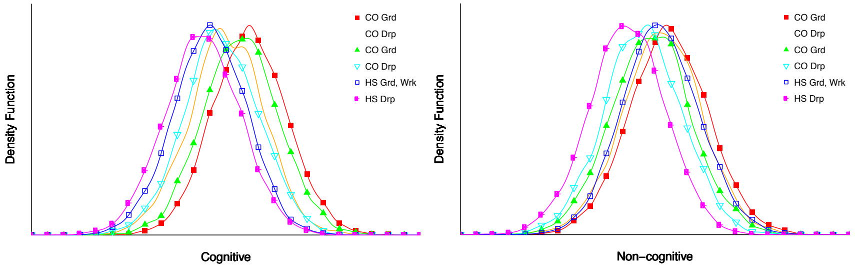
$$\mathcal{L} \int_{\underline{\Theta}} \left[\prod_{j \in M} \underbrace{f\left(M(j) \,\middle|\, D, \theta; \psi\right)}_{\text{Measurement}} \times \right. \tag{4}$$

$$\left. \prod_{s \in \mathcal{S}} \left\{ \underbrace{f\left(Y(s) \,\middle|\, D, \theta; \psi\right)}_{\text{Outcome}} \underbrace{\Pr\left(G(s) = 1 \,\middle|\, D, \theta; \psi\right)}_{\text{Transition}} \right\}^{\mathbb{1}\{s \in \Gamma\}} \right] dF(\theta) \tag{5}$$

- Observe role of unobservable θ (vector).
- Produces conditional independence.
- $(M(j) | Y(s) \perp\!\!\!\perp D || \theta)$

Empirical Results

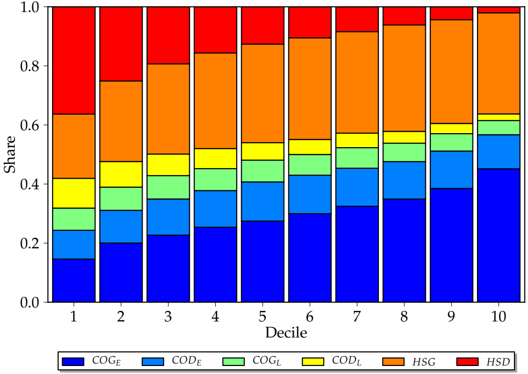
Figure 3: Ability Distributions by Terminal States



Simulate a sample of 50,000 agents based on the estimates of the model.

Ability Distributions by Final Education

Figure 4: Non-Cognitive Skills



Ability Distributions by Final Education

Figure 5: Cognitive Skills

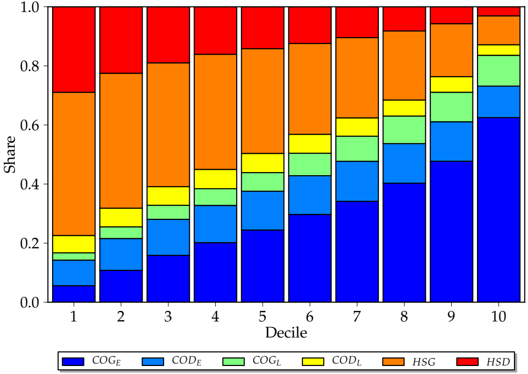
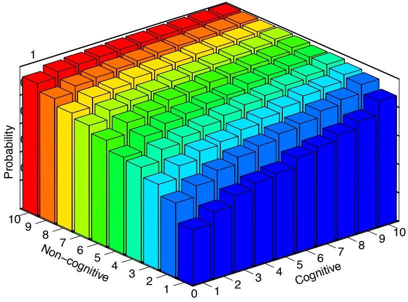
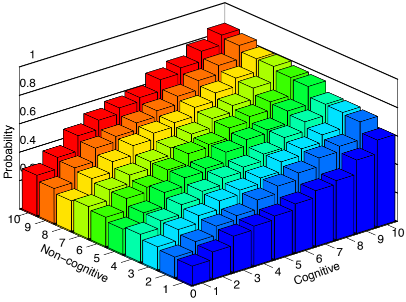


Figure 6: Transition Probabilities by Abilities



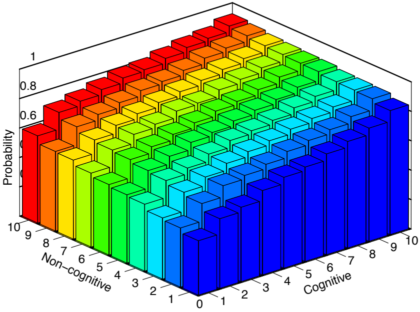
(a) High School Completion



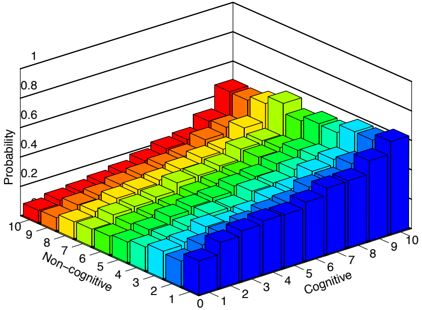
(b) Early College Enrollment

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state.

Figure 6: Transition Probabilities by Abilities (continued)



(a) Early College Graduation

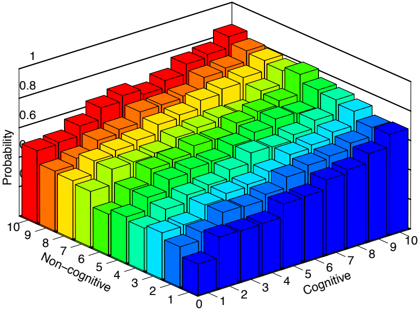


(b) Late College Enrollment

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state.

THE UNIVERSITY OF CHICAGO

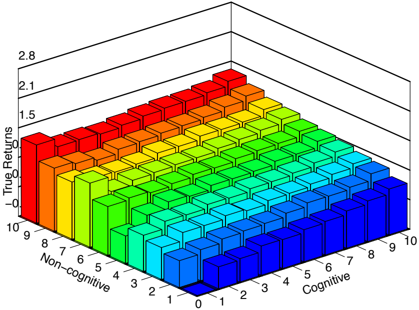
Figure 6: Transition Probabilities by Abilities (continued)



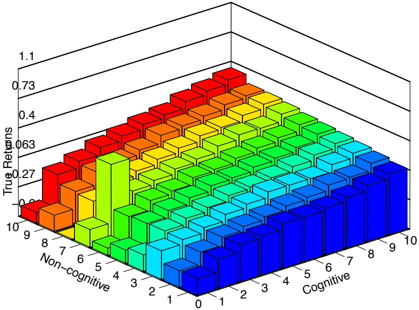
(a) Late College Graduation

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state.

Figure 7: Ex Ante Net Returns by Abilities



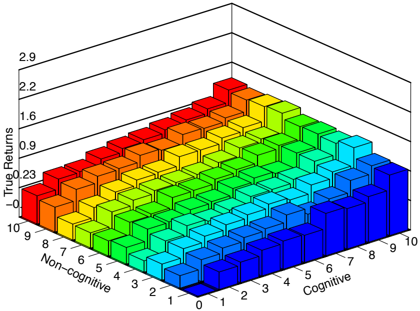
(a) High School Completion
$NR^a = 0.64$
$GR^a = 0.30$



(b) Early College Enrl.
$NR^a =$ -0.06
$GR^a = 0.17$

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state. Enrl. = Enrollment, Grad. = Graduation.

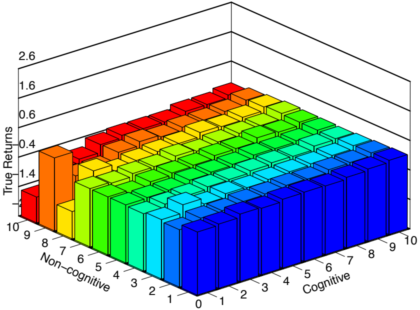
Figure 7: Ex Ante Net Returns by Abilities (continued)



(a) Early College Grad.

$NR^a = 0.57$

$GR^a = 0.89$



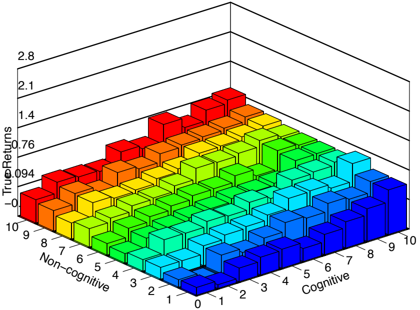
(b) Late College Enrl.

$NR^a =$ -0.23

$GR^a = 0.34$

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state. Enrl. = Enrollment, Grad. = Graduation.

Figure 7: Ex Ante Net Returns by Abilities (continued)



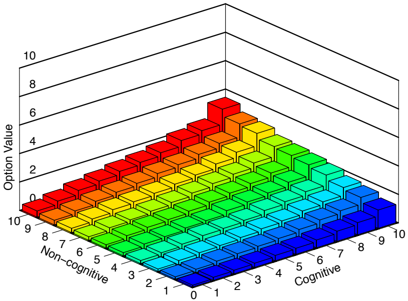
(a) Late College Grad.

$NR^a = 0.15$

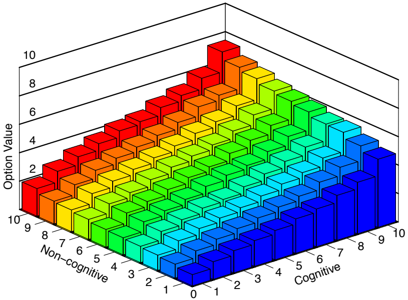
$GR^a = 0.33$

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state. Enrl. = Enrollment, Grad. = Graduation.

Figure 8: Option Values by Abilities



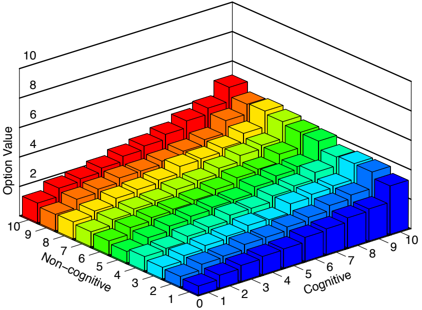
(a) High School Completion
$OV \quad = 0.99$
$OVC = 0.10$



(b) Early College Enrollment
$OV \quad = 3.33$
$OVC = 0.30$

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state. In units of $100,000.

Figure 8: Option Values by Abilities (continued)



(a) Late College Enrollment

$OV \quad = 2.19$

v$OVC = 0.19$

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state. In units of $100,000.

Figure 9: Choice Probability, Early College Enrollment

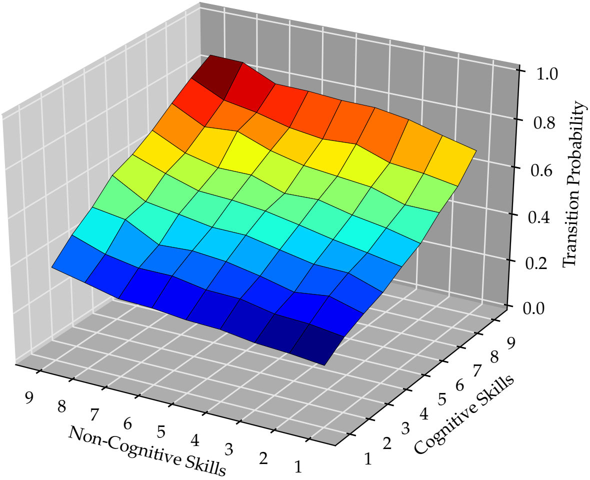


Figure 10: Gross Return, Early College Enrollment

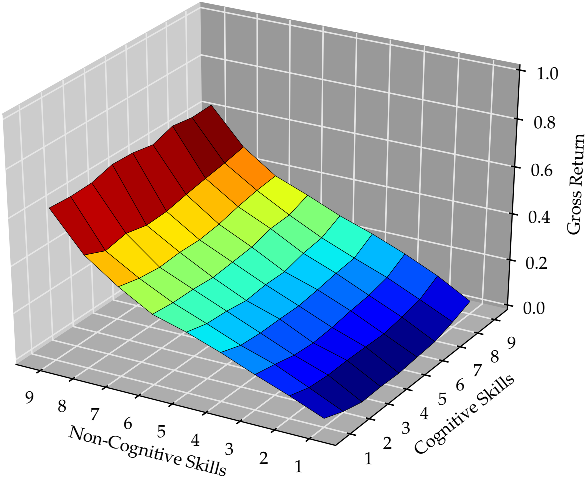


Figure 11: Net Return, Early College Enrollment

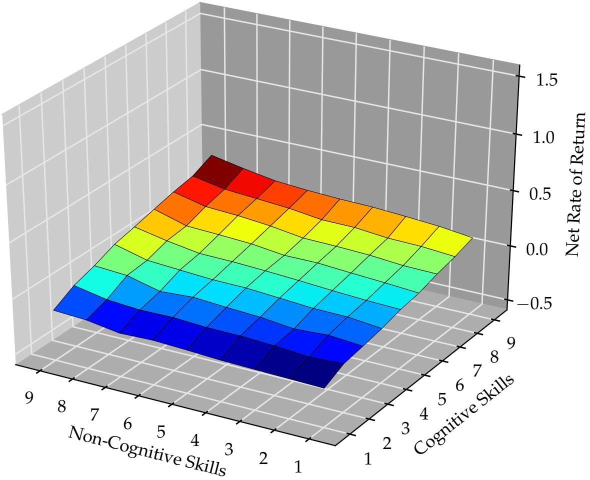


Figure 12: Schooling Attainment by Cognitive Skills

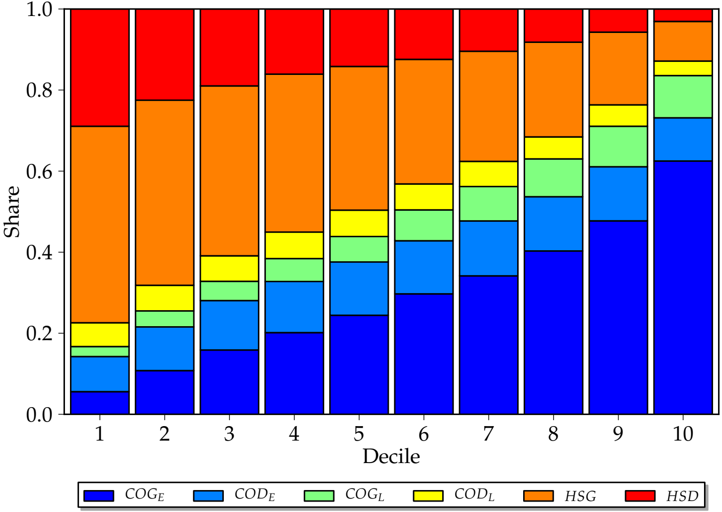


Figure 13: Schooling Attainment by Non-Cognitive Skills

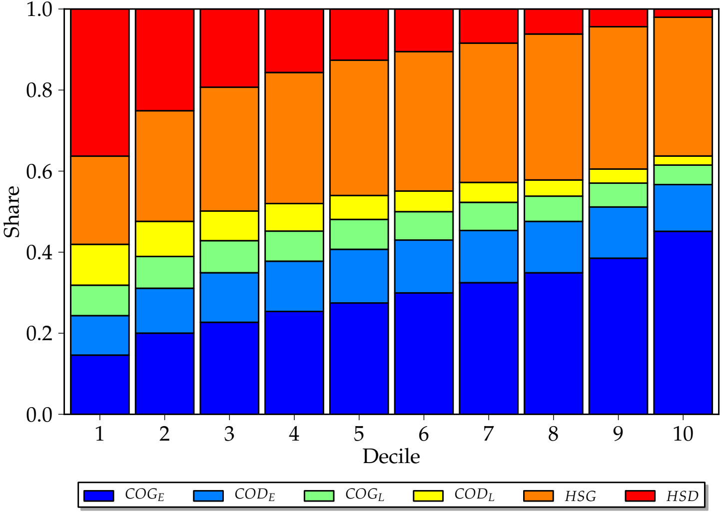


Figure 14: Net Returns (ex ante), High School Graduation

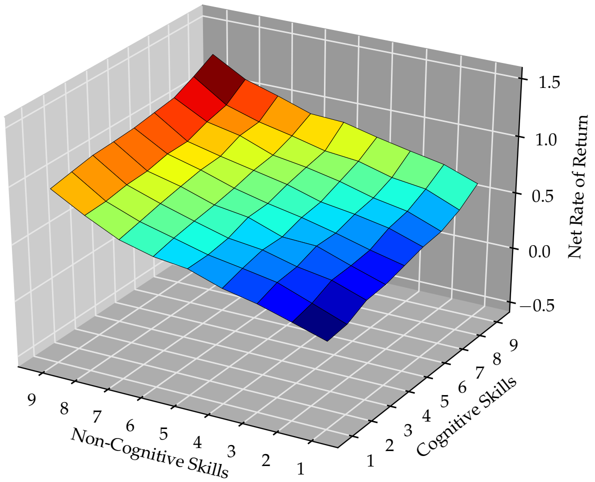


Figure 15: Net Returns (ex ante), Early College Enrollment

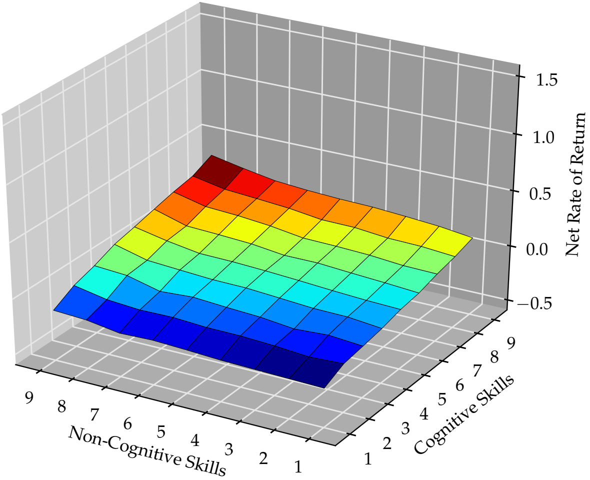


Figure 16: Net Returns (ex ante), Early College Graduation

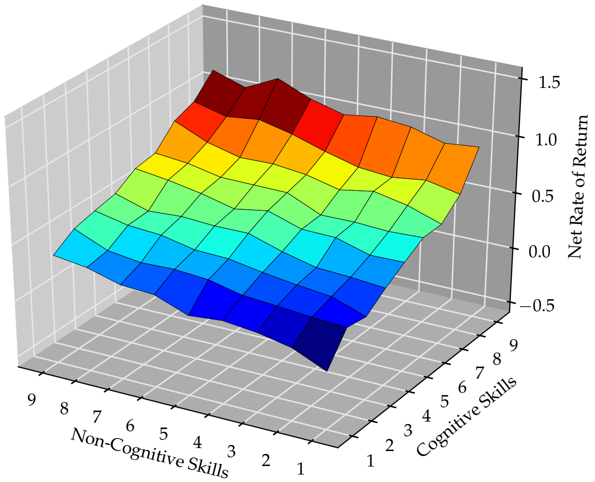


Figure 17: Net Returns (ex ante), Late College Enrollment

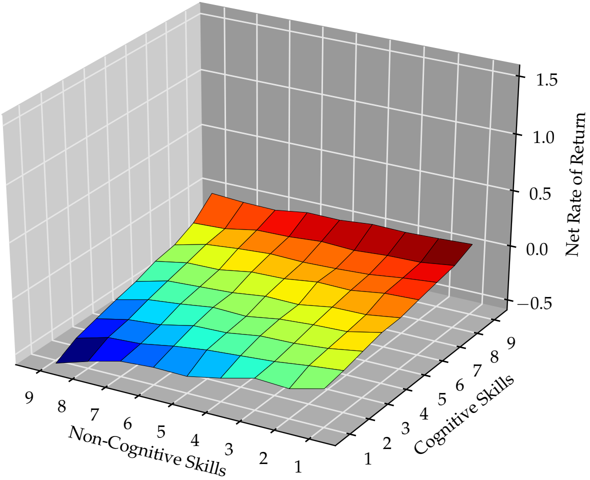


Figure 18: Net Returns (ex ante), Late College Graduation

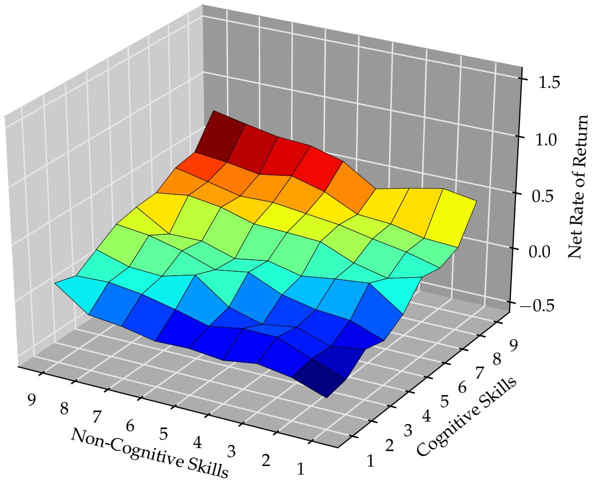


Figure 19: Option Values, High School Graduation

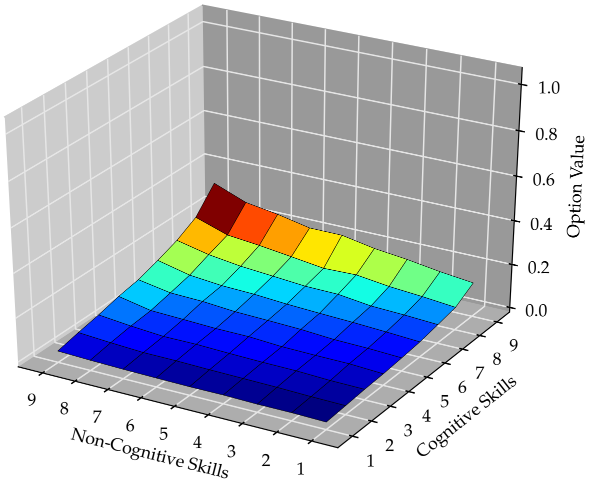


Figure 20: Option Values, Early College Enrollment

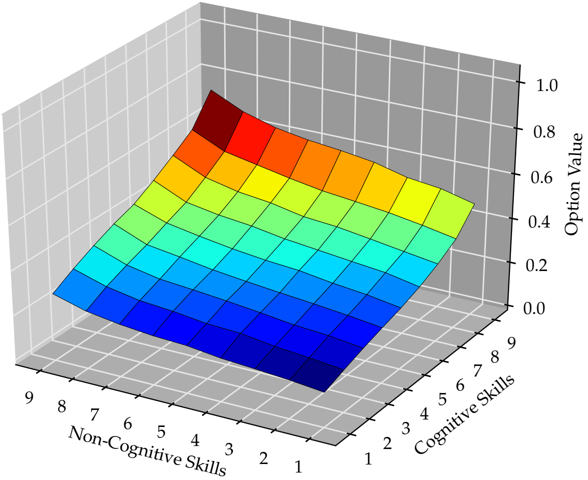


Figure 21: Option Values, Late College Graduation

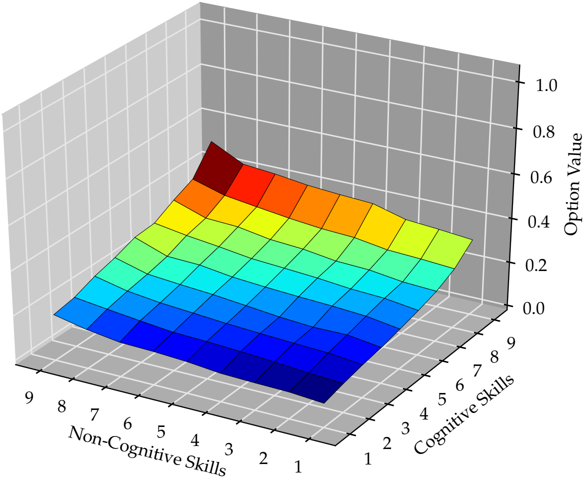


Figure 22: Choice Probability, High School Graduation

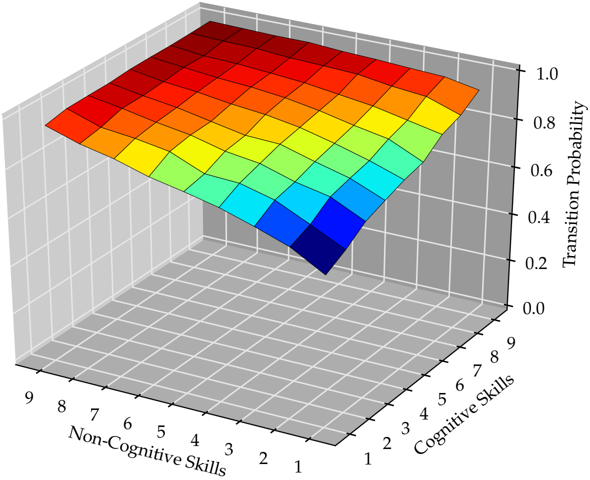


Figure 23: Choice Probability, Early College Enrollment

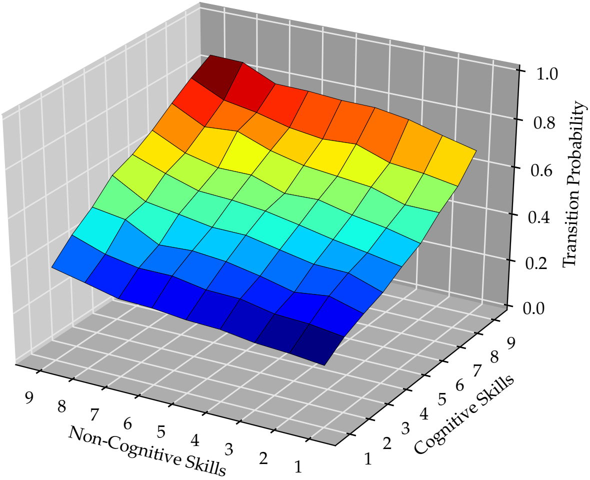


Figure 24: Choice Probability, Early College Graduation

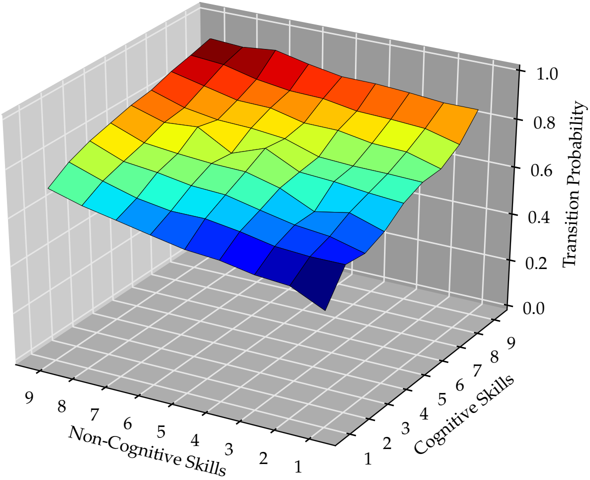


Figure 25: Choice Probability, Late College Enrollment

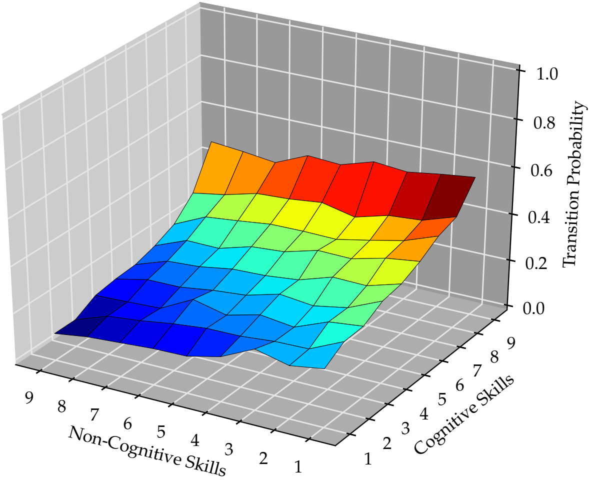


Figure 26: Choice Probability, Late College Graduation

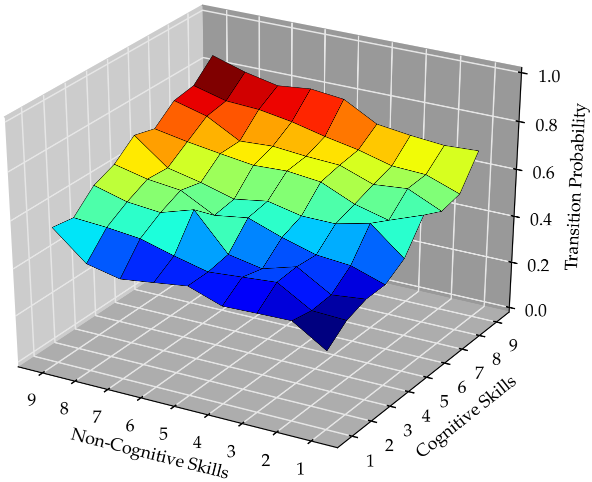


Table 1: Cross Section Model Fit

State	Average Earnings		State Frequencies	
	Observed	ML	Observed	ML
High School Graduates	4.29	3.84	0.30	0.32
High School Dropouts	2.29	2.59	0.17	0.14
Early College Graduates	6.73	7.46	0.29	0.29
Early College Dropouts	4.55	3.87	0.12	0.12
Late College Graduates	4.84	6.22	0.06	0.07
Late College Dropouts	4.89	4.88	0.06	0.06

Table 2: Conditional Model Fit

State	Number of Children	Baby in Household	Parental Education	Broken Home
High School Dropout	0.77	0.26	0.37	0.03
High School Finishing	0.88	0.73	0.55	0.35
High School Graduation	0.91	0.94	0.65	0.91
High School Graduation (cont'd)	0.95	0.33	0.40	0.85
Early College Enrollment	0.46	0.54	0.01	0.15
Early College Graduation	0.06	0.86	0.00	0.14
Early College Dropout	0.33	0.27	0.54	0.75
Late College Enrollment	0.80	0.23	0.90	0.60
Late College Graduation	0.90	0.39	0.90	0.60
Late College Dropout	0.89	0.42	0.91	0.76

Table 3: Internal Rates of Return

		All	
High School Graduation	vs.	High School Dropout	215%
Early College Graduation	vs.	Early College Dropout	24%
Early College Graduation	vs.	High School Graduation (cont'd)	19%
Late College Dropout	vs.	High School Graduation (cont'd)	10%
Late College Graduation	vs.	High School Graduation (cont'd)	17%
Late College Dropout	vs.	High School Graduation (cont'd)	16%

Notes: The calculation is based on 1,407 individuals in the observed data.

Table 4: Net Returns

State	All	Treated	Untreated
High School Finishing	64%	80%	-39%
Early College Enrollment	-6%	30%	-38%
Early College Graduation	57%	103%	-59%
Late College Enrollment	-23%	31%	-45%
Late College Graduation	15%	79%	-61%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

Table 5: Gross Returns

State	All	Treated	Untreated
High School Finishing	30%	32%	16%
Early College Enrollment	17%	23%	13%
Early College Graduation	89%	102%	57%
Late College Enrollment	34%	43%	30%
Late College Graduation	33%	48%	15%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

Table 6: Regret

State	All	Treated	Untreated
High School Finishing	7%	4%	24%
Early College Enrollment	15%	28%	2%
Early College Graduation	29%	33%	19%
Late College Enrollment	21%	27%	19%
Late College Graduation	27%	34%	18%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

Table 7: Option Value Contribution

State	All	Treated	Untreated
High School Finishing	10%	11%	5%
Early College Enrollment	30%	37%	24%
Late College Enrollment	19%	25%	16%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

Table 8: Psychic Costs

State	Mean	2^{nd} Decile	5^{th} Decile	8^{th} Decile
High School Finishing	-2.39	-5.55	-2.40	0.79
Early College Enrollment	2.74	-0.64	2.70	6.09
Early College Graduation	1.78	-3.98	1.86	7.63
Late College Enrollment	5.53	1.75	5.48	9.33
Late College Graduation	1.29	-4.79	1.45	7.40

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. We condition on the agents that actually visit the relevant decision state. Costs are in units of $100,000.

- Table 9 reports the relative size of the psychic costs compared to the total ex ante monetary value of the target state for each transition.
- Please note that the psychic costs for “High School Finishing” are negative on average and that the focus on the average masks considerable heterogeneity.

Table 9: Psychic Costs

State	Mean
High School Finishing	—
Early College Enrollment	23%
Early College Graduation	12%
Late College Enrollment	47%
Late College Graduation	10%

Notes: We simulate a sample of 50,000 individuals based on the estimates of the model. We condition on the agents who actually visit the relevant decision state.

Comparison of ML and SMM

- Using simulated data

- We use the baseline estimates of our structural parameters to simulate a synthetic sample of 5,000 agents.
- This sample captures important aspects of our original data such as model complexity and sizable unobserved variation in agent behaviors.
- We disregard our knowledge about the true structural parameters and estimate the model on the synthetic sample by ML and SMM to compare their performance in recovering the true structural objects.

- We first describe the implementation of both estimation procedures.
- Then we compare their within-sample model fit and assess the accuracy of the estimated returns to education and policy predictions.
- Finally, we explore the sensitivity of our SMM results to alternative tuning parameters such as choice of the moments, number of replications, weighting matrix, and optimization algorithm.

- We assume the same functional forms and distributions of unobservables for ML and SMM.
- Measurement, outcome, and cost equations (1)–(3) are linear-in-parameters.
- Recall that S^c denotes the subset of states with a costly exit.

$$
\begin{array}{lcll}
M(j) & = & X(j)'\kappa_j + \theta'\gamma_j + \nu(j) & \forall \quad j \in M \\
Y(s) & = & X(s)'\beta_s + \theta'\alpha_s + \epsilon(s) & \forall \quad s \in \mathcal{S} \\
C(\hat{s}', s) & = & Q(\hat{s}', s)'\delta_{\hat{s}', s} + \theta'\varphi_{\hat{s}', s} + \eta(\hat{s}', s) & \forall \quad s \in \mathcal{S}^c
\end{array}
$$

- All unobservables of the model are normally distributed in simulation:

$$\begin{array}{rcllrcll} \eta(\hat{s}', s) & \sim & \mathcal{N}(0, \sigma_{\eta(\hat{s}',s)}) & \forall \quad s \in \mathcal{S}^c & \epsilon(s) & \sim & \mathcal{N}(0, \sigma_{\epsilon(s)}) & \forall \\ \theta & \sim & \mathcal{N}(0, \sigma_\theta) & \forall \quad \theta \in \Theta & \nu(j) & \sim & \mathcal{N}(0, \sigma_{\nu(j)}) & \forall \end{array}$$

- The unobservables $(\epsilon(s), \eta(\hat{s}', s), \nu(j))$ are independent across states and measures.
- The two factors θ are independently distributed.
- This still allows for unobservable correlations in outcomes and choices through the factor components θ (Cunha et al., 2005).

ML Approach

- We now describe the likelihood function, its implementation, and the optimization procedure.
- For each agent we define an indicator function $G(s)$ that takes value one if the agent visits state s. Let $\psi \in \Psi$ denote a vector of structural parameters and Γ the subset of states visited by agent i.
- We collect in $D = \{\{X(j)\}_{j \in M}, \{X(s), Q(\hat{s}', s)\}_{s \in S}\}$ all observed agent characteristics.

- After taking the logarithm of equation (4) and summing across all agents, we obtain the sample log likelihood.
- Let $\phi_\sigma(\cdot)$ denote the probability density function and $\Phi_\sigma(\cdot)$ the cumulative distribution function of a normal distribution with mean zero and variance σ.
- The density functions for measurement and earning equations take a standard form conditional on the factors and other relevant observables:

$$\begin{aligned} f\Big(M(j) \mid \theta, X(j)\Big) &= \phi_{\sigma_{\nu(j)}}\Big(M(j) - X(j)'\kappa_j - \theta'\gamma_j\Big) \quad \forall \quad j \in M \\ f\Big(Y(s) \mid \theta, X(s)\Big) &= \phi_{\sigma_{\epsilon(s)}}\Big(Y(s) - X(s)'\beta_s - \theta'\alpha_s\Big) \quad \forall \quad s \in \mathcal{S}. \end{aligned}$$

- The derivation of the transition probabilities has to account for forward-looking agents who make their educational choices based on the current costs and expectations of future rewards.
- Agents know the full cost of the next transition and the systematic parts of all future earnings and costs $(X(s)'\beta_s, Q(\hat{s}', s)'\delta_{\hat{s}',s})$.
- They do not know the values of future random shocks.

- Agents at state s decide whether to transition to the costly state $\hat{s}'$ or the no-cost alternative $\tilde{s}'$.
- Their *ex ante* valuations $T(s')$ incorporate expected earnings and costs, and the continuation value $CV(s')$ from future opportunities.
- Given our functional form assumptions, the *ex ante* value of state s' is:

$$T(s') = \begin{cases} X'(\hat{s}')\beta_{\hat{s}'} + \theta'\alpha_{\hat{s}'} - Q(\hat{s}', s)'\delta_{\hat{s}',s} - \theta'\varphi_{\hat{s}',s} + CV(\hat{s}') \\ X'(\tilde{s}')\beta_{\tilde{s}'} + \theta'\alpha_{\tilde{s}'} + CV(\tilde{s}') \end{cases}$$

- The *ex ante* state evaluations and distributional assumptions characterize the transition probabilities:

$$\Pr\Big(G(s') = 1 \;\Big|\; D, \theta; \psi\Big) \quad = \quad \begin{cases} \Phi_{\sigma_{\eta(\hat{s}',s)}}\left(T(\hat{s}') - T(\tilde{s}')\right) & \text{if} \quad s' = \hat{s}' \\ 1 - \Phi_{\sigma_{\eta(\hat{s}',s)}}\left(T(\hat{s}') - T(\tilde{s}')\right) & \text{if} \quad s' = \tilde{s}'. \end{cases}$$

- Finally, the continuation value of s is:

$$\begin{aligned} CV(s) = & \left[\Phi_{\sigma_{\eta(\hat{s}',s)}}\left(T(\hat{s}') - T(\tilde{s}')\right)\right] \times \int_{-\infty}^{T(\hat{s}')-T(\tilde{s}')} \left[T(\hat{s}') - \eta\right] \frac{\phi_{\sigma_{\eta(\hat{s}',s)}}(\eta)}{\Phi_{\sigma_{\eta(\hat{s}',s)}}(T(\hat{s}') - T(\tilde{s}'))} d\eta \\ + & \left[1 - \Phi_{\sigma_{\eta(\hat{s}',s)}}\left(T(\hat{s}') - T(\tilde{s}')\right)\right] \times T(\tilde{s}'), \end{aligned}$$

where we integrate over the conditional distribution of $\eta(\hat{s}', s)$ as the agent chooses the costly transition to $\hat{s}'$ only if $T(\hat{s}') - \eta(\hat{s}', s) > T(\tilde{s}')$.

- We compare ML against SMM for statistical and numerical reasons.
- ML estimation is fully efficient as it achieves the Cramér-Rao lower bound.
- The numerical precision of the overall likelihood function is very high with accuracy up to 15 decimal places.
- This guarantees at least three digits of accuracy for all estimated model parameters.

- We discuss the numerical properties of the likelihood and bounds on approximation error in the web appendix.
- We use Gaussian quadrature to evaluate the integrals of the model.
- We maximize the sample log likelihood using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Press et al., 1992).

The SMM Approach

- We present the basic idea of the SMM approach and the details of the criterion function.
- Then we discuss the choice of tuning parameters.
- The goal in the SMM approach is to choose a set of structural parameters ψ to minimize the weighted distance between *selected moments* from the observed sample and a sample simulated from a structural model.

- Define $\hat{f}(\psi)$ as:

$$\hat{f}(\psi) = \frac{1}{R} \sum_{r=1}^{R} \hat{f}_r(u_r; \psi).$$

- The simulation of the model involves the repeated sampling of the unobserved components $u_r = \{\{\epsilon(s), \eta(\hat{s}', s)\}_{s \in S}\}$ determining agents' outcomes and choices.
- We repeat the simulation R times for fixed ψ to obtain an average vector of moments.
- $\hat{f}_r(u_r; \psi)$ is the set of moments from a single simulated sample.

- We solve the model through backward induction and simulate 5,000 educational careers to compute each single set of moments.
- We keep the conditioning on exogenous agent characteristics implicit.

- We account for θ by estimating a vector of factor scores based on M that proxy the latent skills for each participant (Bartlett, 1937).
- The scores are subsequently treated as ordinary regressors in the estimation of the auxiliary models.
- We use the true factors in the simulation steps, assuring that SMM and ML are correctly specified.

- The random components u_r are drawn at the beginning of the estimation procedure and remain fixed throughout.
- This avoids chatter in the simulation for alternative ψ, where changes in the criterion function could be due to either ψ or u_r (McFadden, 1989).
- To implement our criterion function it is necessary to choose a set of moments, the number of replications, a weighting matrix, and an optimization algorithm.
- Later, we investigate the sensitivity of our results to these choices.

- We select our set of moments in the spirit of the efficient method of moments (EMM), which provides a systematic approach to generate moment conditions for the generalized method of moments (GMM) estimator (Gallant and Tauchen, 1996).

Link to Appendix

- Overall, we start with a total 440 moments to estimate 138 free structural parameters.

- We set the number of replications R to 30 and thus simulate a total of 150,000 educational careers for each evaluation of the criterion function.
- The weighting matrix W is a matrix with the variances of the moments on the diagonal and zero otherwise.
- We determine the latter by resampling the observed data 200 times.
- We exploit that our criterion function has the form of a standard nonlinear least-squares problem in our optimization.
- Due to our choice of the weighting matrix, we can rewrite as:

$$\Lambda(\psi) = \sum_{i=1}^{I} \left(\frac{\check{f}_i - \hat{f}_i(\psi)}{\hat{\sigma}_i} \right)^2,$$

where I is the total number of moments, f_i denotes moment i, and $\hat{\sigma}_i$ its bootstrapped standard deviation.

Appendix

- Gallant and Tauchen (1996) propose using the expectation under the structural model of the score from an auxiliary model as the vector of moment conditions.
- We do not directly implement EMM but follow a Wald approach instead, as we do not minimize the score of an auxiliary model but a quadratic form in the difference between the moments on the simulated and observed data.
- Nevertheless, we draw on the recent work by Heckman et al. (2014) as an auxiliary model to motivate our moment choice.

- Heckman et al. (2014) develop a sequential schooling model that is a halfway house between a reduced form treatment effect model and a fully formulated dynamic discrete choice model such as ours.
- They approximate the underlying dynamics of the agents' schooling decisions by including observable determinants of future benefits and costs as regressors in current choice.
- We follow their example and specify these dynamic versions of Linear Probability (LP) models for each transition.

- In addition, we include mean and standard deviation of within state earnings and the parameters of Ordinary Least Squares (OLS) regressions of earnings on covariates to capture the within state benefits to educational choices.
- We add state frequencies as well.

Return to Main Text

Web Appendix

Identification

- We establish that our model is semi-parametrically identified.
- Our estimated model of schooling restricts agents to binomial choices at each decision node and there is no role for time.
- However, we provide identification results for a broader class of models.
- We allow for multinomial choices and introduce time $t \in \mathcal{T} = \{1, \ldots, T\}$.
- The model in the paper is a special case of our more general analysis. In this more flexible model, earnings functions are specified by:

$$Y(t,s) = \mu_{t,s}(X(t,s)) + \theta' \alpha_{t,s} + \epsilon(t,s),$$

let $p(t,s) = \theta' \alpha_{t,s} + \epsilon(t,s)$.

- The costs functions are specified by:

$$C(t, s', s) = K_{t,s',s}(Q(t, s', s)) + \theta' \varphi_{t,s',s} + \eta(t, s', s),$$

  let $w(t, s', s) = \theta' \varphi_{t,s',s} + \eta(t, s', s)$.
- Finally, the measurement functions are specified by:

$$M(j) = \mu_j(X(j)) + \theta' \gamma_j + \nu(j),$$

- Let $e(j) = \theta' \gamma_j + \nu(j)$.

- The observed components are determined by covariates $X(t,s) \in \mathcal{X}(t,s)$ for earnings, $Q(t,s',s) \in \mathcal{Q}(t,s',s)$ for costs, and $X(j) \in \mathcal{X}(j)$ for measurements.
- We show that all functions $\mu_{t,s}(X(t,s))$, $K_{t,s',s}(Q(t,s',s))$, $\mu_j(X(j))$ and all distributions $F_{P(t,s)}(p(t,s))$ of unobservables for outcome equations, all distributions $F_{W(t,s',s)}(w(t,s',s))$ of the unobservables in all costly exits from each state, and all distributions $F_{E(j)}(e(j))$ of the unobservables in all measurement equations are identified for any t, s', s, and j.
- We extend the results from Heckman and Navarro (2007) to a context of recurring states and multinomial transitions.
- To simplify notation we remove individual subscripts and consider vectors of individual observations indexed over t and s.
- Variables without arguments refer to any t, s, j, and i.

- Define
$U(t', \omega \,|\, \mathcal{I}(t,s)) = -K_{t',\omega,s}(Q(t',\omega,s)) + \mathbb{E}[V(t',\omega) \,|\, \mathcal{I}(t,s)]$
and consider the difference:

$$\Delta[t',\omega \big| \mathcal{I}(t,s)] = (U(t',\omega \,|\, \mathcal{I}(t,s)) - w(t',\omega,s)) - \max_{\substack{\sigma \in \Omega(t,s) \\ \sigma \neq \omega}} (U(t',\sigma \,|\, \mathcal{I}(t,s)) - w(t',\sigma,s)),$$

  such that state ω is picked whenever $\Delta[t',\omega \big| \mathcal{I}(t,s)] > 0$.
- This condition defines a partition in the space of the unobservables such that state ω is selected.

Theorem 1

Assume that:

1. $P, W,$ and E are continuous random variables with mean zero, finite variance, and support $Supp(P) \times Supp(W) \times Supp(E)$. Assume that the cumulative distribution function of W is strictly increasing over its full support for any t and s.
2. $X, Q \perp\!\!\!\perp (P, W, E)$ for all t and s.
3. $Supp(\mu(X), \mu_j(X), U(Q)) = Supp(\mu(X)) \times Supp(\mu_j(X)) \times Supp(U(Q))$.
4. $Supp(-W) \subseteq Supp(U(Q))$ for any t and s.

Then $\mu_{t,s}(X(t,s))$ *is identified for any* t *and* s*,* $\mu_j(X(j))$ *is identified for all* j*, and the joint distribution* $F_{P(t,s),E(j)}(p(t,s), e(j))$ *is identified for any* t, s, j*.*

Proof.

Conditions (iii) and (iv) guarantee that there exist sets $\bar{\mathcal{Q}}(t, s', s)$ such that

$$\lim_{Q(t,s',s) \to \bar{\mathcal{Q}}(t,s',s)} \mathrm{P}(\Delta[t', s'] > 0) = 1.$$

In the limit sets, we can form:

$$Pr[p(t,s) < Y(t,s) - \mu_{t,s}(X(t,s)), e(j) < M(j) - \mu_j(X(j)) \,|\, X(j) = x(j),\ X(t,s) = x(t,s)] =$$
$$= F_{P(t,s),E(j)}(Y(t,s) - \mu_{t,s}(x(t,s)), M(j) - \mu_j(x(j))),$$

and then we can trace out the whole distribution $F_{P(t,s),E(j)}(p(t,s), e(j))$ by independently varying the points of evaluation. $\square$ $\square$

- Whenever the limit set condition is not satisfied in the analyzed sample, then identification relies either on the assumption that in large samples such limit sets exist, or it is conditional on a subset and only bounds for model parameters can be recovered.
- Notice that the plausibility of these conditions depends on the postulated model.
- In particular, the richer the specification for the set of feasible future states $\mathcal{S}^f(t,s)$ and the finer the time partition for the model, the harder it is to have this condition satisfied in the data.

- Fewer observations will populate each state in any given finite sample.
- Given the above theorem, which mimics Theorem 4 in Heckman and Navarro (2007), we can identify the joint distribution of outcomes across different states s and times t using factor analysis as described in the aforementioned paper.
- Factor analysis also allows to identify the factor loadings $(\alpha_{t,s}, \gamma_j)$ and to separately identify the marginal distributions of the factors θ and the marginal distribution of the idiosyncratic shocks $\epsilon(t, s)$ and $\nu(j)$ for any t, s, and j.
- Note that the measurement system is not needed for identification of the factor distributions if the state space is sufficiently large (the number of states plus the number of transitions is greater than $2N + 1$ when N is the number of factors).
- However, it increases efficiency and aids in the interpretation of the factors, e.g., as cognitive and non-cognitive abilities.

Theorem 2

Assume that:

1. (i) *Conditions (i) to (iv) of Theorem 1 are satisfied.*
2. (ii) $K_{t,s}(Q(t,s',s))$ *is a continuous function for any* t *and any* s*.*
3. (iii) $Q(t,s',s) \in \mathcal{Q}$*, a common set over* t *and* s*.*
4. (iv) *For each transition remaining in the current state is always a costless option. For an agent in state* s *in* t*:* $K_{t',s',s}(Q(t',s',s)) + w(t',s',s) = 0$ *if* $s' = s$*.*
5. (v) *For all alternatives* $\omega \in \Omega(t,s)$ *there exist a coordinate of* $Q(t',\omega,s)$ *that possesses an everywhere positive Lebesgue density conditional on the other coordinates and it is such that* $K_{t',\omega,s}(Q(t',\omega,s))$ *is strictly increasing in this coordinate.*
6. (vi) $U(t',\omega \mid \mathcal{I}(t,s))$ *belongs to the class of Matzkin (1993) functions according to her Lemmas 3 and 4.*

Then we identify the function $K_{t,\omega,s}(Q(t,\omega,s))$*, the marginal distribution of the unobservable portion of the cost functions* $F_{W(t,\omega,s)}(w(t,\omega,s))$*, and exploiting the factor structure representations, the factor loadings* $\varphi_{t,\omega,s}$ *and marginal distribution of the idiosyncratic shocks in the costs functions* $F_{H(t,\omega,s)}(\eta(t,\omega,s))$ *for all transitions.*

Proof.

Consider all final transitions. We define transitions to be final when they lead to final states. A state s is defined as final if $\Omega(t,s) = \{s\}$ for all t. No choice is left to the agent but to remain in the current state. Recall that remaining in the current state involves no costs. For any final state $\omega \in \Omega(t,s)$ we have:

$$\begin{aligned} U(t', \omega \mid \mathcal{I}(t,s)) &= -K_{t',\omega,s}(Q(t',\omega,s)) + \mathbb{E}[V(t',\omega) \mid \mathcal{I}(t,s)] \\ &= -K_{t',\omega,s}(Q(t',\omega,s)) + \mathbb{E}[(\mu_{t',\omega}(X(t',\omega)) + p(t',\omega)) \mid \mathcal{I}(t,s)] \\ &= -K_{t',\omega,s}(Q(t',\omega,s)) + \mu_{t',\omega}(X(t',\omega)) + \mathbb{E}[p(t,\omega) \mid \Delta[t',\omega \mid \mathcal{I}(t,s)] > 0,\ \mathcal{I}(t,s)] \\ &= -K_{t',\omega,s}(Q(t',\omega,s)) + \mu_{t',\omega}(X(t',\omega)) + \theta' \alpha_{t,\omega}. \end{aligned}$$

Notice that $\mu_{t',\omega}(X(t',\omega)) + \theta'\alpha_{t,\omega}$ is known by Theorem 1 and due to the factor structure assumption. Thus we can identify the cost equation $K_{t',\omega,s}(Q(t',\omega,s))$. Imposing restrictions on the generality of the cost function $K_{t',\omega,s}(Q(t',\omega,s))$ is necessary such that $U(t',\omega \mid \mathcal{I}(t,s))$ satisfies (ii), (v), and (iv). Standard arguments from Matzkin (1993) guarantee identification of the function $K_{t,s',s}(Q(t,s',s))$. We do not have to worry about the fact that only differences in utilities are identified in her setup as by (iii), we always have an alternative which implies zero costs. We can also identify the distribution $F_{W(t',\omega,s)}(w(t',\omega,s))$ for any final states. Exploiting the factor structure we can then identify the joint distribution $F_{W(t',\omega,s),P(t',\omega,s),E(j)}(w(t',\omega,s), p(t',\omega,s), e(j))$ for all final transitions and by isolating the dependency between unobservables, we identify the marginal distribution $F_{H(t,\omega,s)}(\eta(t,\omega,s))$ for each final transition. Once these are obtained, by backward induction all expected value functions are identified and therefore all $K_{t,s',s}(Q(t,s',s))$ and $F_{W(t',\omega,s),P(t',\omega,s),E(j)}(w(t',\omega,s), p(t',\omega,s), e(j))$ for any transition and all marginal distributions $F_{H(t,\omega,s)}(\eta(t,\omega,s))$ for any transition are identified. Note that linearity does not fulfill the necessary conditions and only allows for identification up to scale. We therefore need to consider the case separately where the scale of the cost function is not identified. □ □

Theorem 3

Assume that:

1. *Conditions of Theorem 1 and 2 are satisfied, but for the fact that the scale of $K_{t,s',s}(Q(t,s',s))$ is not identified as when it is linear.*
2. *(a) In any final state, $\mathcal{X}(t,s)\backslash\mathcal{Q}(t,s's)$ is not empty and $\mu_{t,s}(X(t,s))$ has an additive component which depends only on variables in $\mathcal{X}(t,s)\backslash\mathcal{Q}(t,s's)$. Alternatively, (b) there is a coordinate of the vector $Q(t,s',s)$ such that $K_{t,s',s}(Q(t,s's))$ is additively separable in that coordinate and it has a known coefficient.*

Then the scale of $K_{t,s',s}(Q(t,s',s))$ is determined.

Proof.

Assumption (ii.a) guarantees that there is a component which can be identified in the outcome equations by the limit sets argument and that can be independently varied from other elements in $U(t, s)$. Applying (ii.b) implies that the scale is known. Notice that the expected value function has an equivalent role as one of the variables in the set defined by (ii.a) for any non final transition, provided that the discount rate is known. Otherwise, if the discount rate is not known and therefore appears as a coefficient in front of $U(t', \omega)$ for future accessible states, we require exclusion restrictions of the type in (ii) in at least one non final transition to identify it. □ □

Following the analysis of Heckman and Navarro (2007), we can identify the discount rate under the same conditions given there.

Data Description

- Our baseline data is the NLSY79 (Bureau of Labor Statistics, 2001).
- We restrict our sample to white males only.
- We construct longitudinal schooling histories by compiling all information on school attendance, including self-reports and the high school survey.
- We then check the compatibility of all the information for each individual within and across time.
- In the presence of contradictions, we review all information for the questionable observation and try to identify the source of the error and correct it.
- If impossible, we drop the observation. Finally, we impose the structure of our decision tree on the agents' educational histories.
- We ignore any form of adult education.

- We use the following set of observables: annual earnings, current geographic location, small child in household, number of siblings, mother's and father's education, dummy variables for marriage status, intact families in 1979, south at age 14, and urban area at age 14.
- We impute missing values.
- When dealing with time constant covariates, imputation is straightforward.
- If information on time varying covariates is missing for only a few years, we use a three year moving average for continuous covariates and the last value for discrete variables.
- Otherwise the agent is dropped from our sample.
- If annual earnings are missing for a limited time only, we impute them using a three year moving average.

- We use tuition data for two- and four-year colleges from the Integrated Postsecondary Education Data System (IPEDS).
- We carefully construct state averages.
- We ensure comparability of the tuition data over time and address the change in the definitions in 1986.
- We only use tuition from public universities.
- We construct local economic conditions such as hourly wages and unemployment using the Current Population Survey (CPS) data by state, level of education, ethnicity, and gender.
- We merge all datasets using the NLSY Geocode Data.

Rates of Return, Option Values, and Regret

- Table 10 presents internal rates of return for selected comparisons of schooling levels.
- For definition of this traditional concept, see Heckman et al. (2006).
- We compare the recorded earnings streams until age 45.
- We therefore consider earnings in all states up to the one in the first column.
- Missing earnings are set to zero, unless during high school enrollment.
- There we impute a three year moving average.

Table 10: Internal Rates of Return

All			
High School Graduation	vs.	High School Dropout	215%
Early College Graduation	vs.	Early College Dropout	24%
Early College Graduation	vs.	High School Graduation (cont'd)	19%
Late College Dropout	vs.	High School Graduation (cont'd)	10%
Late College Graduation	vs.	High School Graduation (cont'd)	17%
Late College Dropout	vs.	High School Graduation (cont'd)	16%

Notes: The calculation is based on 1,407 individuals in the observed data.

- The Mincer rate of return is 11.6%.

- Table 11 reports the median *ex ante* net returns to education by treatment status.
- We condition on agents that actually visit the relevant decision state.
- The treated choose the transition to the state in the first column.

Table 11: Net Returns

State	All	Treated	Untreated
High School Finishing	64%	75%	-27%
Early College Enrollment	-3%	24%	-28%
Early College Graduation	50%	82%	-44%
Late College Enrollment	-21%	22%	-38%
Late College Graduation	10%	62%	-51%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

- Table 12 reports the average *ex ante* gross returns to education by treatment status.
- We condition on agents that actually visit the relevant decision state.
- The treated choose the transition to the state in the first column.

Table 12: Gross Returns

State	All	Treated	Untreated
High School Finishing	27%	29%	16%
Early College Enrollment	14%	20%	8%
Early College Graduation	75%	84%	49%
Late College Enrollment	29%	28%	29%
Late College Graduation	24%	36%	9%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

- Table 13 shows the percentage of agents experiencing regret, i.e., those agents for which the *ex post* and *ex ante* returns do not agree in sign.
- We condition on agents that actually visit the relevant decision state.
- The treated choose the transition to the state in the first column.

Table 13: Regret

State	All	Treated	Untreated
High School Finishing	7%	4%	24%
Early College Enrollment	15%	28%	2%
Early College Graduation	29%	33%	19%
Late College Enrollment	21%	27%	19%
Late College Graduation	27%	34%	18%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

c

- Table 14 reports the option value contribution, i.e., the relative share of the option value in the overall value of each state.
- We condition on agents that actually visit the relevant decision state.
- The treated choose the transition to the state in the first column.

Table 14: Option Value Contribution

State	All	Treated	Untreated
High School Finishing	7%	8%	2%
Early College Enrollment	30%	37%	23%
Late College Enrollment	17%	24%	15%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.