INTERNATIONAL ECONOMIC REVIEW

May 2015
Vol. 56, No. 2

ESTIMATION OF DYNAMIC DISCRETE CHOICE MODELS BY MAXIMUM LIKELIHOOD AND THE SIMULATED METHOD OF MOMENTS*

BY PHILIPP EISENHAUER, JAMES J. HECKMAN, AND STEFANO MOSSO[1]

The University of Chicago, U.S.A.; The University of Chicago Center for the Economics of Human Development and American Bar Foundation, U.S.A.; The University of Chicago, U.S.A.

We compare the performance of maximum likelihood (ML) and simulated method of moments (SMM) estimators for dynamic discrete choice models. We construct and estimate a simplified dynamic structural model of education that captures some basic features of educational choices in the United States in the 1980s and early 1990s. We use estimates from our model to simulate a synthetic data set and assess the ability of ML and SMM to recover the model parameters on this sample. We investigate the performance of alternative tuning parameters for SMM.

1. INTRODUCTION

Economic science uses economic theory to guide the interpretation of economic data and to shape policy. Kenneth Wolpin is a model economic scientist who integrates theory and data in a rigorous fashion. He summarizes his philosophy toward empirical research in Wolpin (2013). He is a major contributor to structural econometrics with particular emphasis on the study of dynamic discrete choice models. His contributions are both methodological and empirical. His methodological research focuses on promoting methods to increase the reliability of algorithms for structural estimation (Eckstein and Wolpin, 1989; Keane et al., 2011) and developing techniques to simplify their empirical implementation. His research on interpolation methods to solve dynamic discrete choice models with a large state space (Keane and Wolpin, 1994) is a prominent example. In his empirical contributions, he extensively applies theory-motivated methods to investigate many important issues such as educational

*Manuscript received August 2013; revised October 2014.

[1] We thank George Yates for numerous valuable comments, for excellent computational assistance in developing the ML estimator used in this article, and for assisting in the study of accuracy bounds for the computational algorithm. We thank Edward Sung and Jake Torcasso for their outstanding research assistance. We have benefited greatly from comments received from Chris Flinn, Kenneth Judd, Michael Keane, Bernard Salanié, Petra Todd, and Stefan Wild. We thank the editor and anonymous referees for their valuable comments. This research was supported in part by the American Bar Foundation, the Pritzker Children's Initiative, the Buffett Early Childhood Fund, NICHD R37HD065072, R01HD054702, the Human Capital and Economic Opportunity Global Working Group—an initiative of the Becker Friedman Institute for Research in Economics—funded by the Institute for New Economic Thinking (INET), and an anonymous funder. The views expressed in this article are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Institutes of Health. Philipp Eisenhauer thanks Prof. Wolfgang Franz and the Centre for European Economic Research (ZEW Mannheim) for their support. We are grateful to George Yates for his help in preparing the appendix. For further information or questions and suggestions, please contact us at info@policy-lab.org.

Please address correspondence to: James J. Heckman, Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637. Phone: 773-702 0634. E-mail: *jjh@uchicago.edu*.

attainment (Keane and Wolpin, 1997; Eckstein and Wolpin, 1999), the role of credit constraints in educational attainment (Keane and Wolpin, 2001), and labor market dynamics (Lee and Wolpin, 2006, 2010).

This article contributes to the literature on estimating dynamic discrete choice models. It investigates the empirical performance of widely used versions of simulated method of moments (SMM), a computationally tractable method for estimating complex structural models. SMM estimates parameters by fitting a vector of empirical moments to their theoretical counterparts simulated from a structural model (McFadden, 1989). We compare its performance against standard maximum likelihood (ML) estimation.[2]

We estimate a deliberately simplified dynamic discrete choice model of schooling based on a sample of white males from the National Longitudinal Survey of Youth of 1979 (NLSY79) using ML. Our model is more restrictive compared to standard dynamic discrete choice models (Keane and Wolpin, 1997, 2001) with respect to the number of choices and the timing of decisions and outcomes. We restrict agents to binary choices at a sequence of decision nodes. This allows us to evaluate the likelihood analytically, without the need for any simulation or interpolation (Keane, 1994), which provides a clean comparison of ML against simulation-based estimation methods such as SMM. Using the estimates of model parameters, we simulate a synthetic data set. In a series of Monte Carlo studies, we compare estimates based on our precisely calculated ML with those from widely used, computationally tractable versions of SMM. Because our synthetic sample is derived from real data, our analysis provides useful lessons on the performance of SMM for the estimation of structural models.[3]

SMM has been used to estimate models of job search (Flinn and Mabli, 2008), educational and occupational choices (Adda et al., 2011, 2013), household choices (Flinn and Del Boca, 2012), stochastic volatility models (Andersen et al., 2002; Raknerud and Skare, 2012), and dynamic stochastic general equilibrium models (Ruge-Murcia, 2012). SMM can be used for any model, however complex or difficult to compute the likelihood, as long as it is possible to simulate it. Under conditions presented in the literature, the SMM estimator is consistent and asymptotically normal (Gouriéroux and Monfort, 1997). If the score vector for SMM happens to be correctly specified, then SMM is asymptotically efficient (Gallant and Tauchen, 1996; Gouriéroux et al., 1993).[4]

Implementing any estimation strategy requires numerous choices. In the case of SMM, users have discretion in selecting (a) the moments used in estimation, (b) the number of replications used to compute the simulated moments, (c) the moment weighting matrix, and (d) the algorithm used for optimization. It is unclear how such choices affect the performance of the SMM estimator and how they depend on the structure of the model estimated. We propose diagnostic tools to test their validity.

We suggest a Monte Carlo procedure that allows SMM users to gain confidence for their particular implementation of the algorithm. We present a new optimization algorithm for solving derivative-free nonlinear least-squares problems that is well suited for conventional SMM implementations. A benchmarking exercise demonstrates significant speed improvements

[2] Alternative estimation methods have been proposed to overcome the rigidities and complexities of ML estimation. Most require the analyst to characterize the likelihood function but simplify its computation. One of the most popular methods, simulated ML (SML), substitutes the exact likelihood function with a simulated one. An example is the Hajivassiliou-Geweke-Keane (HGK-SML) estimator (Geweke, 1989; Hajivassiliou and McFadden, 1998; Keane, 1994) used for multinomial probit estimation. Approximations of the dynamic programming problem have often been combined with SML in models with a large state space (Keane and Wolpin, 1994). Another popular method is the conditional choice probabilities (CCP) algorithm first proposed by Hotz and Miller (1993) and recently extended to allow for unobserved heterogeneity (Arcidiacono and Miller, 2011). Using CCP, a consistent estimator of the model parameters can be derived without the need of the full solution of the dynamic programming problem. The CCP method, however, restricts the flexibility of the estimable models by imposing assumptions that limit the expectation formation of agents and restrict the stationarity of the environment. SMM is a more general alternative to ML estimation.

[3] As in Skrainka (2012), we use simulation experiments in realistic settings to investigate the finite sample behavior of widely used estimators.

[4] See Nickl and Pötscher (2010) and Gach and Pötscher (2011) for recent additional results.

compared to the algorithms commonly used in the literature. Combining state-of-the-art optimization methods with parallel computing allows analysts to perform our proposed Monte Carlo exercise even in computation-intensive models.

We present our schooling model in Section 2. Section 3 presents baseline results. Section 4 outlines our Monte Carlo study and compares the performance of ML and SMM estimation. Section 5 concludes.

2. DYNAMIC MODEL OF EDUCATIONAL CHOICES

This section presents a computationally tractable dynamic discrete choice model of education and establishes conditions under which it is identified. We specify a model with a simple state space by assuming that agents move from one schooling state to the next. Agents are assumed to have two choices at each decision node. The value of each state is determined by its immediate rewards and costs and by the expected future value of all feasible states made available by a choice. Agents have private information on their own type and form expectations about future states with respect to their current information set.

Our simple specification comes at the expense of a less realistic empirical analysis of the dynamics of schooling choices compared to those of Keane and Wolpin (1997, 2001) and Johnson (2013). We restrict agents to binary choices, and our model is based on educational states. We make these assumptions because they allow us to evaluate the likelihood without any simulation, which provides a clean comparison of ML estimation with simulation-based alternatives.

2.1. *Setup.* Given the current state $s \in \mathcal{S} = \{s_1, \ldots, s_N\}$, let $\mathcal{S}^v(s) \subseteq \mathcal{S}$ denote the set of visited states and $\mathcal{S}^f(s) \subseteq \mathcal{S}$ the set of feasible states that can be reached from s. We collect the choice set of the agent in state s in $\Omega(s) = \{s' \mid s' \in \mathcal{S}^f(s)\}$. We consider binary choices only, so $\Omega(s)$ has at most two elements. Ex post, the agent receives per period rewards $R(s') = Y(s') - C(s', s)$ defined as the difference between per period earnings, $Y(s')$, and the costs $C(s', s)$ associated with moving from state s to state s'. The costs combine monetary expenses such as tuition and psychic costs (e.g., Cunha et al., 2005). We can only identify the differences in the costs for two alternative states. We thus normalize the cost of one of the exits to zero. In the subsequent analysis it is useful to explicitly distinguish between the nonzero ($\hat{s}'$) and zero cost ($\tilde{s}'$) exits from s. We collect the subset of states with a costly exit in S^c. We assume earnings and costs are separable functions of observed covariates $X(s) \in \mathcal{X}$ for earnings and $Q(\hat{s}', s) \in \mathcal{Q}$ for costs. There is a stochastic component $(U_Y(s), U_C(\hat{s}', s))$ to each of them. Earnings are expressed as

$$Y(s) = \mu_s(X(s)) + U_Y(s). \tag{1}$$

The costs of going from state s to state $\hat{s}'$ are defined by

$$C(\hat{s}', s) = K_{\hat{s}',s}(Q(\hat{s}', s)) + U_C(\hat{s}', s). \tag{2}$$

Some variables in $Q(\hat{s}', s)$ and $X(s')$ might be the same. Their distinct elements constitute the exclusion restrictions.

We assume a factor structure on the unobservables by postulating that a low-dimensional vector of latent factors θ is the sole source of dependency among the unobservables of the model (Hansen et al., 2004; Cunha et al., 2005):

$$U_Y(s) = \theta'\alpha_s + \epsilon(s) \qquad U_C(\hat{s}', s) = \theta'\varphi_{\hat{s}',s} + \eta(\hat{s}', s).$$

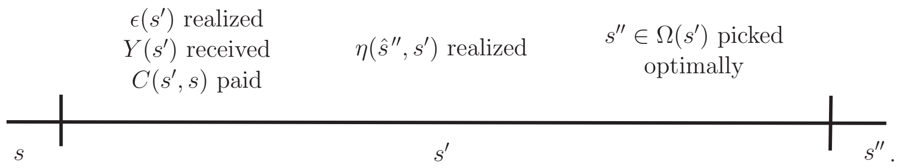
The individual-specific factors θ are known to agents but unknown to the econometrician, whereas the idiosyncratic shocks $\epsilon(s)$ and $\eta(\hat{s}', s)$ are unknown to the econometrician and only known by the agents at different stages of the decision process. The idiosyncratic shocks are independent but not identically distributed. We thus generalize the i.i.d. innovation assumption in Keane and Wolpin (1997). The impact of these traits on earnings and costs is given by the factor loadings $(\alpha_s, \varphi_{\hat{s}',s})$. We allow for unobservable correlations in outcomes and choices across states through θ and the loadings vary by states.

Following Carneiro et al. (2003), Cunha et al. (2010), and Heckman et al. (2013), we assume access to a J-dimensional vector of individual measures M (such as test scores or behavioral indicators) proxying individual factors θ. We use the measures as noisy signals of the factors θ:

$$M(j) = \mu_j(X(j)) + \theta'\gamma_j + \nu(j) \quad \text{for } j = 1, \ldots, J. \tag{3}$$

We assume that $\epsilon(s)$, $\eta(\hat{s}', s)$, and $\nu(j)$ are mutually independent for all $j, s, \hat{s}'$. In measurement system (3), we interpret the unobserved factors as individual-specific traits.

We assume that agents are risk neutral and maximize discounted lifetime rewards when making their educational choices. When an agent makes his educational choice to proceed from state s to s', he knows the stochastic component of the transition $\eta(\hat{s}', s)$ but not future earnings $\epsilon(s')$. The assumed timing of the arrival of information is as follows:



The agents know $X(s)$, $Q(\hat{s}', s)$, and θ for all s. Under this timeline, we define $\mathcal{I}(s)$ as the information set of the agent in state s by specifying all components known in the state:

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

The agents in state s know the costs associated with a transition to any feasible state s'. We assume that the agent uses the distributions for the earnings shocks $\epsilon(s)$, denoted by $F_{E,s}(\epsilon(s))$, and for the transition costs shock $\eta(\hat{s}', s)$, denoted by $F_{H,\hat{s}',s}(\eta(\hat{s}', s))$, to form expectations about future states. The distributions of the shocks can vary across states.[5]

We define the agent's value function at state s, given the available information in s, recursively as

$$V(s) = Y(s) + \max_{s' \in \Omega(s)} \left\{ \frac{1}{1+r} \left(-C(s', s) + \mathbb{E}[V(s') \,|\, \mathcal{I}(s)\,] \right) \right\}. \tag{4}$$

[5] We differ from Keane and Wolpin (1997) in our specification of the distribution of the unobserved components. In their specification, agents have different initial conditions for each state variable. The distribution of initial conditions is multinomial with five components. They assume that there are only four types (values) of initial conditions in the population. Serial dependence is induced through the persistence of the initial conditions as determinants of current state variables. In addition, at each age, the agent receives five shocks associated with the rewards of each choice. The shocks are joint normally distributed, and serially uncorrelated, and they are assumed to be i.i.d. over time. In our model, we allow for state dependence in the distribution of the unobservables by letting earnings and cost shocks be drawn from normal distributions with different variances at each state and at each transition. Moreover, we allow unobserved portions of cost and return functions to be contemporaneously and serially correlated through their common dependence on the factors θ. Our θ are normally distributed so we have a continuum of types. The Keane and Wolpin (1997) specification of persistent heterogeneity is a version of a factor model in which all factor loadings are implicitly determined (through Bellman iterations) by the parameters of the deterministic portions of cost and return functions and the distribution functions of unobserved variables and the sample distribution of observables. In our approach, the factor loadings are specified independently of the parameters of the deterministic portions of the cost and return functions and the sample distribution of observed variables.

For future reference, we define the continuation value of state s as the second term on the right-hand side of (4):

$$CV(s) = \max_{s' \in \Omega(s)} \left\{ \frac{1}{1+r} \big(-C(s', s) + \mathbb{E}[V(s') \,|\, \mathcal{I}(s)\,] \big) \right\}. \tag{5}$$

The agent's policy function determines the optimal transitions. An agent in s chooses his next feasible state s' according to the following rule:

$$s' = \begin{cases} \hat{s}' & \text{if } \mathbb{E}\big[V(\hat{s}') \,\big|\, \mathcal{I}(s)\big] - C(\hat{s}', s) > \mathbb{E}\big[V(\tilde{s}') \,\big|\, \mathcal{I}(s)\big] \\ \tilde{s}' & \text{otherwise.} \end{cases} \tag{6}$$

We now define the returns to schooling and the concept of the option value.

2.2. *Returns to Education.* We define the ex ante and ex post net returns to schooling. The net return (NR) to schooling includes per period earnings and costs associated with each educational choice and the option value of future opportunities (discussed in the next subsection). The ex ante net returns are defined before the unobservable components of future earnings are realized. They depend on agents' expectations and determine their choices. Standard methods for computing rates of return such as Mincer coefficients or internal rates of returns ignore costs and option values of future opportunities. They are only interpretable for terminal choices and ex post realized earnings streams.[6] We define the ex ante net return of $\hat{s}'$ over $\tilde{s}'$ for an agent currently in state s as

$$\frac{\mathbb{E}\big[V(\hat{s}') - V(\tilde{s}') \,\big|\, \mathcal{I}(s)\big] - C(\hat{s}', s)}{\mathbb{E}\big[V(\tilde{s}') \,\big|\, \mathcal{I}(s)\big]} = NR^a(\hat{s}', \tilde{s}', s). \tag{7}$$

We also define the ex ante gross return (GR), which includes all future earnings but omits all costs related to educational choices. Define the gross value of a state s recursively as

$$\tilde{V}(s) = Y(s) + \frac{1}{1+r} \big(\mathbb{E}[\tilde{V}(s') \,|\, \mathcal{I}(s)\,] \big),$$

where state $s' \in \Omega(s)$ maximizes the discounted future rewards according to the policy function defined in Equation (6). Although agents do not base their educational choices upon the gross returns, they are important, as they are defined in terms of earnings streams only and are the focus of much applied work reporting rates of return. We define the ex ante gross return of $\hat{s}'$ over $\tilde{s}'$ for an agent in s as

$$\frac{\mathbb{E}\big[\tilde{V}(\hat{s}') - \tilde{V}(\tilde{s}') \,\big|\, \mathcal{I}(s)\big]}{\mathbb{E}\big[\tilde{V}(\tilde{s}') \,\big|\, \mathcal{I}(s)\big]} = GR^a(\hat{s}', \tilde{s}', s). \tag{8}$$

We formulate the net and gross ex post returns in the same way but use the value functions that include the realizations of the earnings shock. The ex post returns can be used to evaluate an agent's regret of his educational choice.

2.3. *Option Values of Schooling.* Consider a high school enrollee who is contemplating whether to either graduate or drop out. Part of his evaluation of the benefits of high school graduation is the option to start college. From the perspective of state s, the option value of s'

[6] See Heckman et al. (2006a) for a discussion of conventional methods for estimating rates of return and their economic interpretation.

is defined as the difference between the value of taking the optimal choice when moving from s' and the fallback value of the zero cost exit $\tilde{s}''$. The zero cost exit is usually associated with maintaining the current education level, for example, remaining a high school graduate and not enrolling in college. Then the option value[7] of feasible state s' from the perspective of s is

$$\begin{aligned} OV(s', s) &= \frac{1}{1+r}\,\mathbb{E}\left[\mathbb{E}\Big[\max_{s''\in\Omega(s')}\Big\{V(s'') - C(s'', s')\Big\}\Big|\mathcal{I}(s')\Big] - \mathbb{E}\Big[V(\tilde{s}'')\,\big|\,\mathcal{I}(s')\Big]\,\Big|\,\mathcal{I}(s)\right] \\ &= \frac{1}{1+r}\,\mathbb{E}\left[\underbrace{\max_{s''\in\Omega(s')}\{V(s'') - C(s'', s')\}}_{\text{value of options arising from } s'} - \underbrace{V(\tilde{s}'')}_{\text{fallback value}}\,\Big|\,\mathcal{I}(s)\right]. \end{aligned} \tag{9}$$

We define the option value contribution $OVC(s', s) = \frac{OV(s',s)}{\mathbb{E}[V(s')\,|\,\mathcal{I}(s)]}$ as the relative share of the option value in the overall value of a state. This component is not reported in standard calculations of the Mincer return or the internal rate of return.

2.4. *Identification.* Our model is semiparametrically identified using a straightforward extension of the arguments in Heckman and Navarro (2007). The main arguments of the proof, presented in Web Appendix A (http://heckman.uchicago.edu/MLvsSMM>>, consist of using (a) a limit set argument to identify the joint distribution of earnings and measurements free of selection ("identification at infinity"), (b) the measurement system on the factor structure that facilitates identification of the joint distribution of factors, (c) the choice structure and exclusion restrictions to identify the distribution of costs in the final choice equation, and (d) backward induction to identify relevant distributions in all states showing that the future value function acts as an exclusion restriction in current choices. We can identify all of the parameters of the model including the discount rate.

3. BASELINE ESTIMATES

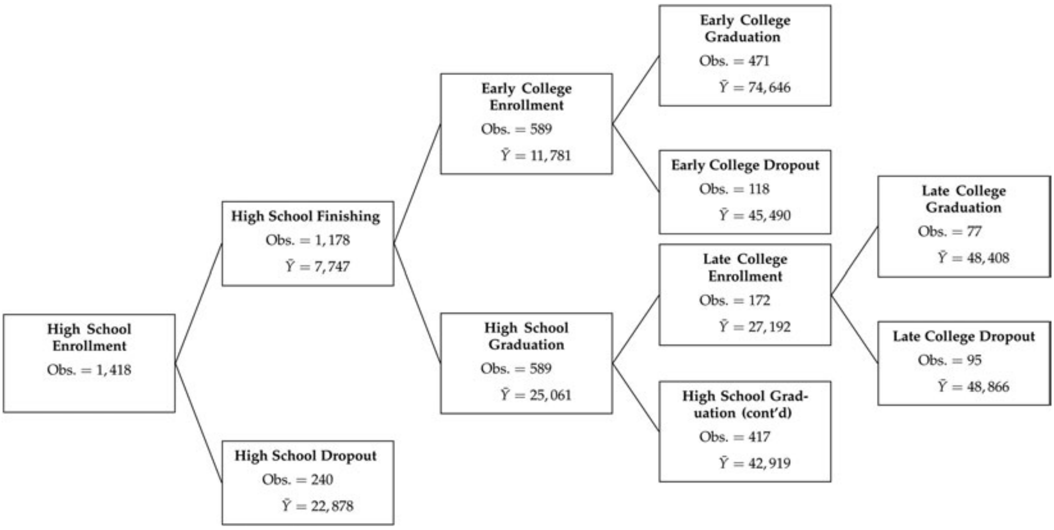
We fit the model on a sample of 1,418 white males from the NLSY79 using ML estimation.[8] Figure 1 shows the decision tree for our model.

All agents start in high school and decide to either drop out or finish. If they finish high school, they can enroll in college immediately or remain high school graduates with the option to enroll in college later or not at all. Conditional on early or late college enrollment, agents can either graduate or drop out. At each decision node, we designate the lower transition to be the zero cost exit. Our addition of the distinction between late and early enrollment is the only place in the model where time is introduced. We do this to improve the fit of the model and incorporate an important feature of the data on education.

For every state s, agents work in the labor market and receive earnings $Y(s)$. When agents pursue higher education by transitioning to the costly state $\hat{s}'$, they incur cost $C(\hat{s}', s)$. Agents face uncertainty about components of future earnings and costs when determining the ex ante value of each state $V(s)$ given the information available to them. As noted in Section 2, we assume that the agent knows his type and all past, present, and future covariates including local labor market conditions. His expectations about the distributions of all future shocks are assumed to be consistent with their actual realizations.

Following Carneiro et al. (2003) and Heckman et al. (2006b), we assume that the agent's type vector θ is summarized by cognitive and noncognitive abilities. We use the scores on the Armed Services Vocational Aptitude Battery (ASVAB) as noisy measures on cognitive abilities. For

[7] Weisbrod (1962) was the first to analyze option values in the context of schooling and human capital accumulation.
[8] See Web Appendix B and Bureau of Labor Statistics (2001) for a description of our sample and the NLSY79.



NOTES: $\bar{Y}$ refers to average annual earnings in the state in 2005 \$. Obs. refers to the number of observations in the state.

FIGURE 1

DECISION TREE

noncognitive skills, we rely on Rotter scores (administered in 1979), on the Rosenberg scores (administered in 1980), and on indicators of risky behaviors such as drug and alcohol use.

In a state s, we assign each agent a duration $D(s)$ based on the number of periods spent in that state. For an agent who spent four years in college, the duration of the college enrollment state will be four. We set the duration for an agent's counterfactual state to the median duration among the agents who actually visit that state. Let $Y(t, s)$ denote the observed earnings in the NLSY79 at time t for an agent in state s. We collapse all $Y(t, s)$ within state s into one discounted average,

$$Y(s) = \frac{\sum_{t=1}^{D(s)} \left(\frac{1}{1+r}\right)^{t-1} Y(t, s)}{\sum_{t=1}^{D(s)} \left(\frac{1}{1+r}\right)^{t-1}}.$$

We do the same for time-varying covariates in $X(s)$ and $Q(\hat{s}', s)$. This setup differs from standard dynamic discrete choice models as the timing of earnings within each state does not matter. We do not estimate the discount factor r and instead set $r = 0.04$.[9]

We discuss the construction of our sample in Web Appendix B. The NLSY79 only has data up to approximately age 45. We extend the duration of the terminal states up to age 65 using parameters estimated on the available sample to project earnings in unobserved years. The high school enrollment state characterizes initial conditions in our model. We assume earnings and costs are functions of standard individual characteristics and local economic conditions.[10]

[9] Heckman and Navarro (2007) and Web Appendix A present conditions under which r is identified.

[10] In each state, earnings depend on the number of children in the household, parental education (as the maximum between the mother's and father's education), indicators for the presence of a baby (child less than three years old) in the household, marriage status, urban residence at age 14, the region of residence (North East, North Central, South, and West), hourly wage, and unemployment levels in the state of residence for the relevant age group (we use two age groups, younger than 30 years old or older). For the cost equations, we exclude the indicator for marriage and the regional dummies, adding instead an indicator for whether the family is intact or not, the number of siblings, and state-level tuition for public two- and four-year colleges for the transitions to college enrollment states. The state representing the conclusion of high school is estimated using only an intercept, the two factors, and an unobservable component. All transition and outcome equations also include the cognitive and noncognitive factor and an idiosyncratic unobserved component.

TABLE 1
CROSS-SECTION MODEL FIT

State	Average Earnings		State Frequencies	
	Observed	ML	Observed	ML
High school finishing	0.77	0.78	0.83	0.86
High school dropout	2.29	2.57	0.17	0.14
Early college enrollment	1.18	1.40	0.42	0.40
High school graduation	2.51	2.48	0.42	0.45
Early college graduation	7.47	6.77	0.33	0.29
Early college dropout	4.55	3.84	0.08	0.11
Late college enrollment	2.72	2.54	0.12	0.14
High school graduation (cont'd)	4.29	3.83	0.29	0.32
Late college graduation	4.84	6.16	0.05	0.08
Late college dropout	4.89	4.95	0.07	0.06

NOTES: Earnings are discounted using the within-state duration and measured in units of $10,000. Statistics are calculated on the NLSY79 sample and for ML based on 50,000 simulated agents using the parameter estimates. State frequencies are unconditional.

Figure 1 presents the average annual earnings and the number of observations by state. Earnings are low during the year of graduation ($7,747). High school graduates earn $42,919, which is almost twice as much as high school dropouts ($22,878). Our distinction between early and late college enrollment is important. Early enrollees earn much less while in college ($11,781) compared with late enrollees ($27,192). Also, early college graduation boosts average annual earnings to $74,646 compared with only $48,408 for late graduation. In the case of late college enrollment, the difference in earnings among graduates and dropouts is minor: $48,408 compared with $48,866. This explains why, in our sample, the number of late college dropouts (95) is actually larger than of late college graduates (77). For the case of early enrollment (589), the vast majority are graduates (471). The Mincer coefficient is 0.116.[11]

3.1. *Model Fit.* Table 1 shows the fit of the model estimated by ML for model fit statistics that are typically used in the literature. Average earnings and state frequencies are well fit by our model. Small discrepancies show up for terminal states. Terminal states are populated by very few agents, which requires us to constrain the outcome and cost parameters of terminal college states to be the same for early and late transitions.

Comparing the fit of the model with cross-section moments is a weak criterion for a dynamic model. A more exacting criterion is to predict sequences of educational choices (Heckman, 1981). We follow Heckman and Walker (1990) and Heckman (1984) and use χ^2 goodness-of-fit tests to examine our model's performance. In Table 2, we report the p-value of a joint test of the relative share of agents for each state for all realizations of selected covariates.[12] For most cells the fit is good.

One exception (at a 5% significance level) is *Parental Education*, where we fail to fit the observed patterns for early college enrollment and early college graduation. For *Broken Home*, we overpredict the relative share of individuals from a broken home among high school dropouts. For all other variables and states, the p-values indicate that the model is consistent with the data. Because tests within covariates across all states are not independent, we use a Bonferroni test to evaluate the joint hypothesis that the predicted covariate distributions fit at each state. The test is based on the maximum χ^2 statistic over all states for each covariate. A 5% Bonferroni test is passed by all covariates besides *Parental Education*. Here, the poor prediction for early college graduates leads to an overall rejection.

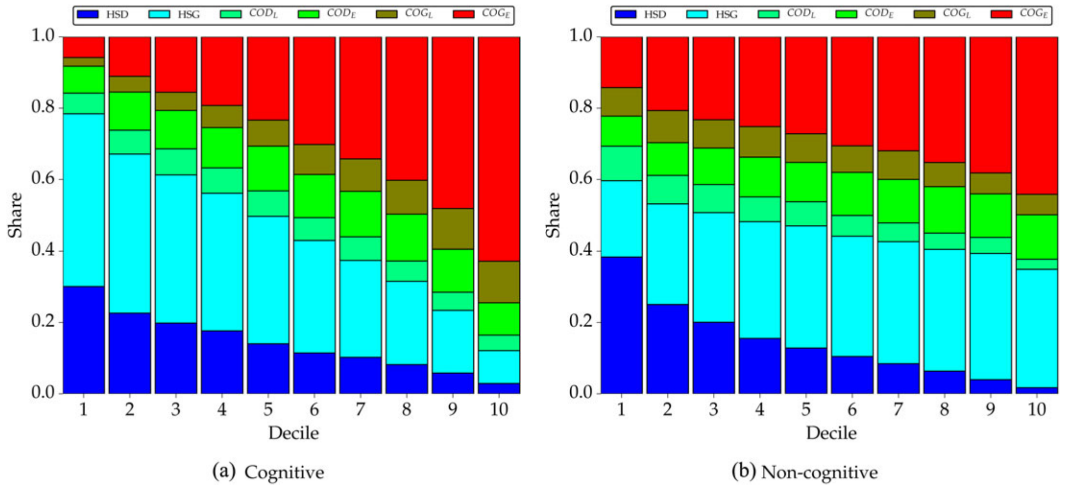
[11] Web Appendix C presents additional descriptive statistics and estimates of conventional internal rates of return.

[12] In the χ^2 test, the predicted covariate distributions depends on estimated parameters. We do not adjust the test statistic to account for parameter estimation error as suggested by Heckman (1984) because the adjustments are usually slight (Heckman and Walker, 1990).

TABLE 2

CONDITIONAL MODEL FIT (P VALUES)

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



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

FIGURE 2

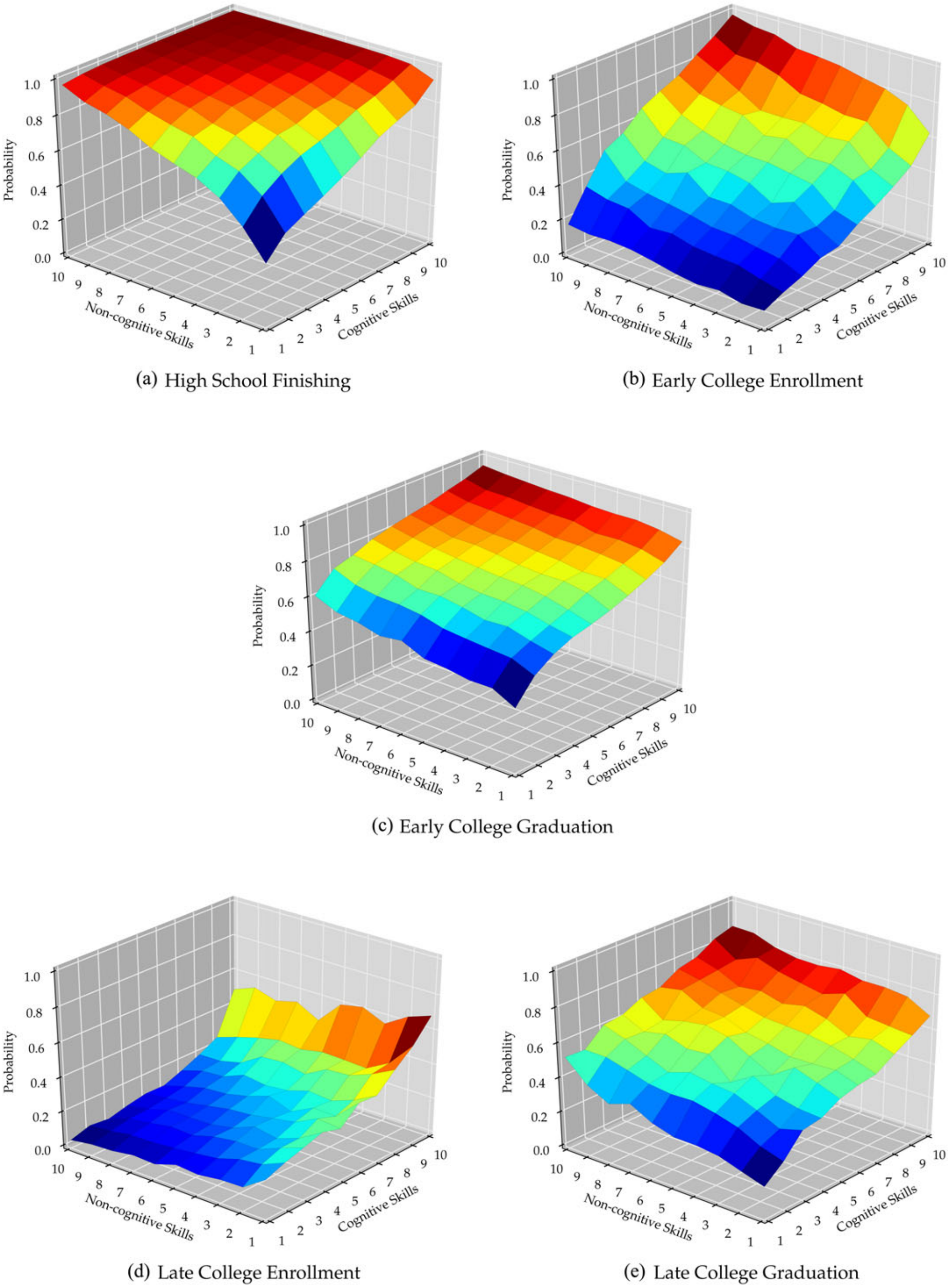
ABILITY DISTRIBUTIONS BY FINAL EDUCATION

3.2. *Economic Implications.* We now present the economic implications of our baseline results. We first discuss the impact of unobserved abilities on educational choices and earnings and then turn to the role of psychic costs and option values for the net returns to schooling. We conclude with a counterfactual policy evaluation.

3.2.1. *Impact of abilities.* Figure 2 shows the share of agents in each of the final states by deciles of the overall factor distribution. The distributions of abilities differ substantially across schooling outcomes. Early college graduates (COE_E) are strong in cognitive and noncognitive abilities. High school dropouts (HSD) are weak in both. High school graduates who never enroll in college (HSG) are weak in cognitive abilities but quite strong in noncognitive abilities.

Figure 3 shows the transition probabilities to each state by factor deciles. Higher cognitive skills increase the likelihood of continued educational achievement for all choices. The effect of noncognitive abilities is mixed. Although they clearly increase the likelihood of finishing high school, higher noncognitive skills decrease the probability of late college enrollment (conditional on working after high school graduation). Delay of college enrollment is associated with lower levels of noncognitive skills.

3.2.2. *Returns to education.* Figure 4 presents the ex ante net return to schooling by factor deciles. The effect of latent skills on returns differs by state. The return of finishing high school



(a) High School Finishing

(b) Early College Enrollment

(c) Early College Graduation

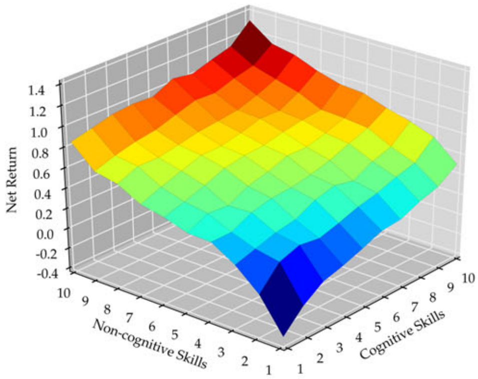
(d) Late College Enrollment

(e) 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 who actually visit the relevant decision state.

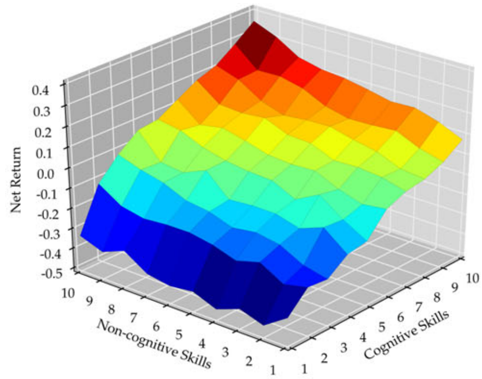
FIGURE 3

TRANSITION PROBABILITIES BY ABILITIES



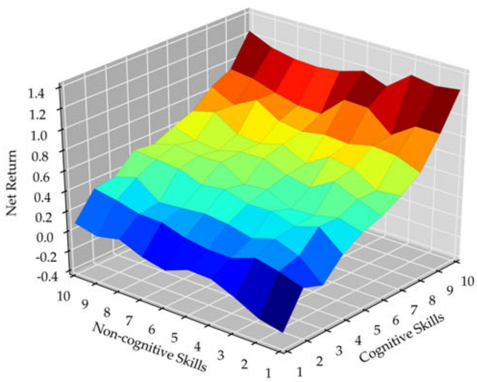
(a)

High School Finishing
$NR^a = 0.64$
$GR^a = 0.27$



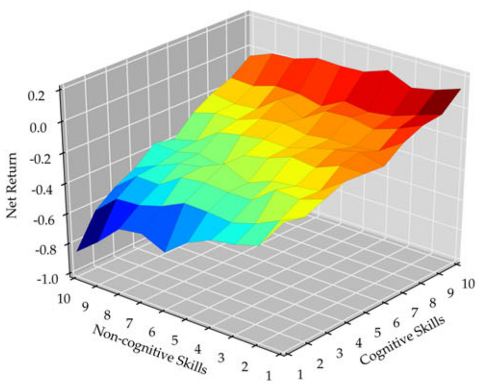
(b)

Early College Enrollment
$NR^a = -0.03$
$GR^a = 0.14$



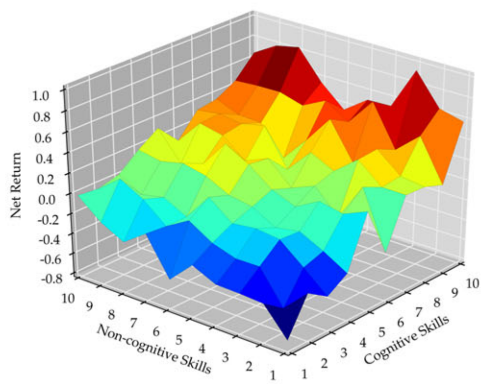
(c)

Early College Graduation
$NR^a = 0.50$
$GR^a = 0.75$



(d)

Late College Enrollment
$NR^a = -0.21$
$GR^a = 0.29$



(e)

Late College Graduation
$NR^a = 0.10$
$GR^a = 0.24$

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

FIGURE 4

EX ANTE NET RETURNS BY ABILITIES

TABLE 3
COSTS

State	Mean	2nd Decile	5th Decile	8th Decile
High school finishing	−2.38**	−5.52***	−2.40**	0.79*
Early college enrollment	2.73	−0.65	2.69	6.10
Early college graduation	1.82	−3.88	1.89	7.61
Late college enrollment	5.53**	1.72	5.48**	9.37**
Late college graduation	1.13	−4.72	1.35	7.32

NOTES: We simulate a sample of 50,000 agents based on the estimates of the model. We condition on the agents who actually visit the relevant decision state. Costs are in units of $100,000. We determine the accuracy of our estimates using the simulation approach proposed by Krinsky and Robb (1986, 1990) with 1,200 replications. Level of significance: ***1%, **5%, *10%.

is strongly affected by the noncognitive factor. Usually the effect of cognitive skills is more pronounced. Nevertheless, our estimates show evidence of strong complementarity between abilities and schooling for most states. Figure 4 also presents median returns. The median net return for early college enrollment is around zero, and the return of delayed enrollment is even negative (−21%). College dropouts pay the cost of college without benefiting from the much larger returns of graduating. The returns from graduating late (10%) are much smaller than for those graduating early (50%). We report the difference between net and gross returns in Figure 4.

Psychic costs are crucial determinants of net returns. For example, the median gross return for early and late college enrollment is positive, whereas the median net return is negative in both cases. As only agents with a positive net return choose to continue their education, this follows directly from our estimates (and the data) as more than half of the agents who are faced with the decision to enroll in college refrain from doing so.

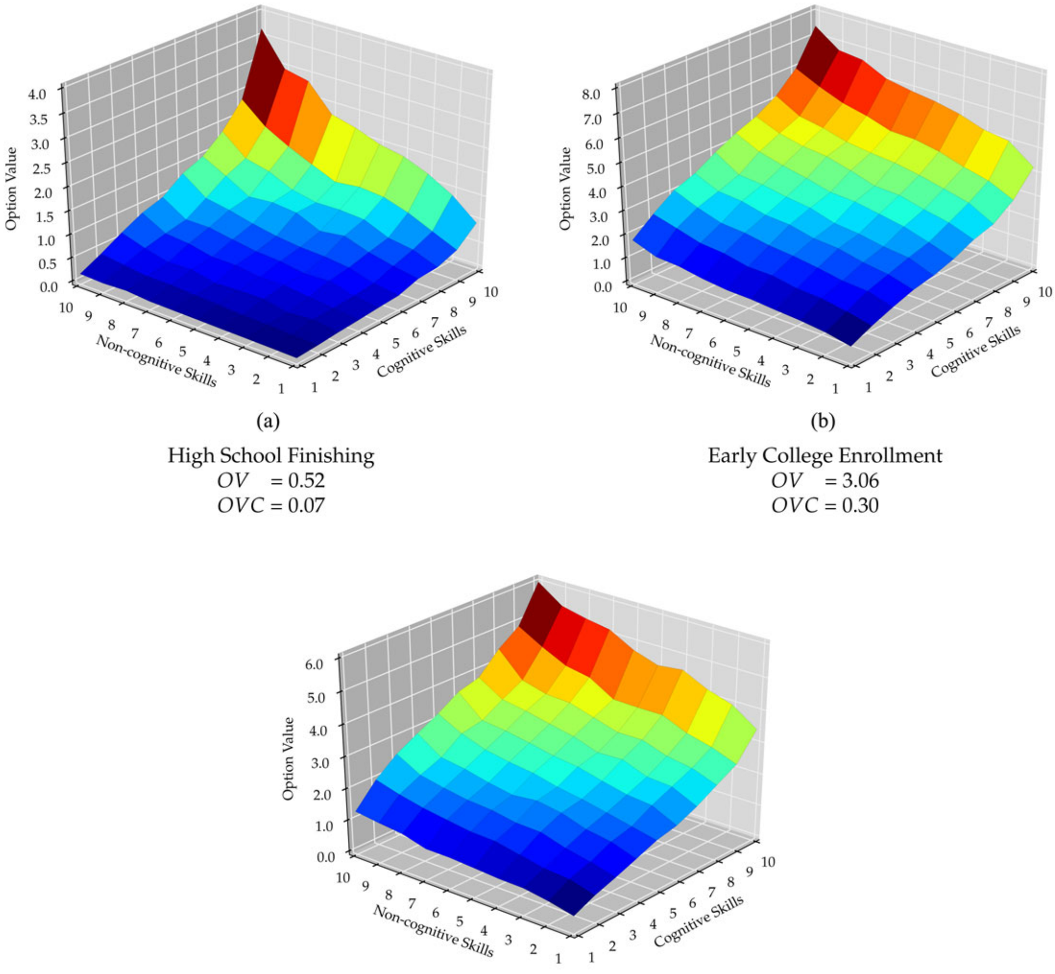
We estimate the overall costs associated with each educational choice. Our estimated costs combine monetary expenses such as tuition and psychic costs. Table 3 reports the average costs associated with each transition. It reports the second, fifth, and eighth decile of their distribution to document their substantial heterogeneity. Costs are key components of the net returns; ignoring them results in strongly biased estimates. The largest costs are associated with early and late college enrollment. These are the only states with psychic as well as monetary costs from tuition. Enrolling early costs the equivalent of $273,000 compared with $553,000 for late enrollment. At least 20% of agents have negative schooling costs in most states. They experience psychic benefits. For high school graduation, even the average cost is negative. Psychic costs play a dominant role in explaining schooling decisions. This is an unsatisfactory feature of the models in the literature; see, for example, Cunha et al. (2005) and Abbott et al. (2013).

Ex ante and ex post returns do not necessarily agree because agents cannot predict their future earnings. Decisions that are optimal for an agent ex ante might be suboptimal ex post. For this reason, we calculate the percentage of agents experiencing regret, that is, those agents for whom the ex post and ex ante returns do not agree in sign.[13] A substantial share of late college enrollees (34%) regret the decision to graduate. For finishing high school, the share is much smaller (4%). However, 24% of high school dropouts regret their decision.

3.2.3. *Option values of schooling.* Our structural model allows us to calculate the option values of educational choices.[14] We defined the option value in Equation (9) as the difference in the value associated with the optimal continuation of choices versus the fallback value. Figure 5 shows the option values conditional on the deciles of the factor distributions, their median

[13] See Web Appendix C for additional results on ex post returns and regret.

[14] Other models taking into account option values have been proposed by Comay et al. (1973), Cunha et al. (2007), Heckman et al. (2014a), and Trachter (2014). See also Cameron and Heckman (1993).



(a)
High School Finishing
$OV = 0.52$
$OVC = 0.07$

(b)
Early College Enrollment
$OV = 3.06$
$OVC = 0.30$

(c)
Late College Enrollment
$OV = 1.87$
$OVC = 0.17$

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

FIGURE 5

OPTION VALUES BY ABILITIES

(OV), and their contribution to the total value of each state (OVC). The option values make a sizable contribution to the overall value of the states and vary by abilities.

Early college enrollment has the highest option value, as graduation yields a large gain in earnings compared to dropping out. As the net returns to college graduation increase in cognitive and noncognitive abilities, so does the option value of college enrollment.

3.2.4. *Policy analysis.* Counterfactual policy analysis is one of the main motivations for the estimation of dynamic structural models (Wolpin, 2013). We investigate the impact of a 50% reduction in tuition cost on agents' college-going decisions. We simulate 50,000 agents from our model and compare their educational choices under the baseline regime and the policy alternative. Agents are forward-looking, and due to the sequential decision tree, reducing tuition for college attendance already increases high school graduation rates by one percentage point as its option value increases. Overall college enrollment increases by roughly 10 percentage, points as many high school graduates now decide to enroll in college. The increase is evenly split between early and late enrollment. However, there are considerable differences in graduation rates among those induced to enter into college depending on the time of enrollment. About half of the new early enrollees will eventually graduate, whereas only a quarter of the late enrollees will do so as well.

4. COMPARISON OF ML AND SMM

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) to (3) are linear-in-parameters. Recall that S^c denotes the subset of states with a costly exit.

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

All unobservables of the model are normally distributed:

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

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).

4.1. *The 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 1 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. Then the likelihood for observation i is given by

$$(10) \quad \int_{\underline{\Theta}} \left[\prod_{j\in M} \underbrace{f\big(M(j) \mid D, \theta; \psi\big)}_{\text{Measurement}} \prod_{s\in\mathcal{S}} \left\{ \underbrace{f\big(Y(s) \mid D, \theta; \psi\big)}_{\text{Outcome}} \underbrace{\Pr\big(G(s) = 1 \mid D, \theta; \psi\big)}_{\text{Transition}} \right\}^{\mathbb{1}\{s\in\Gamma\}} \right] dF(\theta),$$

where $\underline{\Theta}$ is the support of θ. After taking the logarithm of Equation (10) 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

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

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

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

Finally, the continuation value of s is

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

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 Web Appendices D and E. We use Gaussian quadrature to evaluate the integrals of the model.[15] We maximize the sample log likelihood using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Press et al., 1992).

4.2. *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. The criterion function takes the following form:

$$\Lambda(\psi) = \big[\check{f} - \hat{f}(\psi)\big]' W^{-1} \big[\check{f} - \hat{f}(\psi)\big], \tag{11}$$

[15] See Judd and Skrainka (2011) for a comparison of alternative integration strategies.

where $\check{f}$ represents a vector of moments computed on the observed data and $\hat{f}(\psi)$ denotes an average vector of moments calculated from R simulated data sets and W is a positive definite weighting matrix. We 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 in Equation (11).

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). 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. (2014b) as an auxiliary model to motivate our moment choice.[16] Heckman et al. (2014b) 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. Overall, we start with a total of 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

[16] If the weighting matrices are appropriately chosen and the auxiliary model is correctly specified, then both approaches are asymptotically equivalent to ML (Gouriéroux et al., 1993). The EMM approach requires analytical derivatives for the auxiliary model, which is a very time-consuming, error-prone, and tedious task for large and complex models. For this reason, the EMM approach is not commonly used to estimate dynamic discrete choice models but widely applied to fit stochastic volatility models. In the latter case, several tractable auxiliary models such as ARCH and GARCH are readily available (Andersen et al., 1999). See Carrasco and Florens (2002) for an accessible comparison of EMM to other simulation-based methods and additional references.

TABLE 4
CROSS-SECTION MODEL FIT

	Average Earnings			State Frequencies		
State	Observed	ML	SMM	Observed	ML	SMM
High school finishing	0.78	0.76	0.78	0.86	0.84	0.86
High school dropout	2.50	2.53	2.54	0.14	0.16	0.14
Early college enrollment	1.45	1.42	1.45	0.41	0.41	0.41
High school graduation	2.46	2.44	2.45	0.45	0.43	0.45
Early college graduation	6.81	6.99	6.67	0.29	0.30	0.29
Early college dropout	3.91	3.97	4.02	0.12	0.11	0.12
Late college enrollment	2.51	2.55	2.52	0.13	0.13	0.13
High school graduation (cont'd)	3.88	3.83	3.79	0.32	0.30	0.32
Late college graduation	6.03	6.21	6.19	0.07	0.07	0.07
Late college dropout	5.10	4.89	5.05	0.06	0.06	0.06

	ML	SMM
RMSE	0.05058	0.05748

NOTES: Earnings are discounted using the within-state duration and measured in units of \$10,000. Statistics calculated for ML and SMM approaches based on 50,000 simulated agents using the parameter estimates. RMSE = root-mean-square error.

of a standard nonlinear least-squares problem in our optimization. Due to our choice of the weighting matrix, we can rewrite Equation (11) 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.

Our criterion is not a smooth function of the model parameters. Small changes in the structural parameters cause some simulated agents to change their educational choices, resulting in discrete jumps in our set of moments (Smith and Keane, 2004). Thus we cannot use gradient-based methods for optimization and rely on derivative-free alternatives instead. Moré and Wild (2009) show that model-based solvers perform better than standard derivative-free direct search solvers used in the existing literature; see Adda et al. (2011, 2013) and Del Boca et al. (2014) for applications of derivative-free direct search solvers. From the class of model-based solvers, we choose the Practical Optimization Using No Derivatives for Sums of Squares (POUNDerS) algorithm (Munson et al., 2012). POUNDerS exploits the special structure of the nonlinear least-squares problem within a derivative-free trust-region framework and forms a smooth approximation model of the objective function to converge to a minimum.[17]

4.3. *Results.* We compare ML and SMM estimation to learn whether our version of SMM is a good substitute for ML. First, we compare basic model fit statistics. Second, we study the estimates for the returns to education and perform a counterfactual policy exercise. Finally, we explore alternative choices for the set of moments, weighting matrix, number of replications, and optimization algorithm.

4.3.1. *Model fit.* Table 4 shows the average annual earnings for each state and the conditional state frequencies. Overall, both estimation approaches fit these aggregate statistics quite well. The model fit for the average earnings among late college graduates and late college dropouts

[17] See Nocedal and Wright (2006) for a discussion of the nonlinear least-squares problem and Kortelainen et al. (2010) for a detailed description of the underlying mechanics of POUNDerS.

TABLE 5
CONDITIONAL MODEL FIT (p VALUES)

State	Number of Children		Baby in Household		Parental Education		Broken Home	
	SMM	ML	SMM	ML	SMM	ML	SMM	ML
High school dropout	0.90	0.75	0.83	0.95	0.26	0.56	0.65	0.62
High school finishing	0.99	0.99	0.99	0.99	0.96	0.99	0.97	0.89
High school graduation	0.84	0.72	0.99	0.99	0.86	0.99	0.40	0.50
High school graduation (cont'd)	0.67	0.90	0.98	0.98	0.90	0.98	0.37	0.42
Early college enrollment	0.04	0.49	0.98	0.97	0.87	0.94	0.39	0.58
Early college graduation	0.63	0.91	0.81	0.86	0.07	0.06	0.89	0.58
Early college dropout	0.42	0.72	0.99	0.99	0.86	0.99	0.40	0.50
Late college enrollment	0.27	0.62	0.99	0.94	0.14	0.25	0.84	0.96
Late college graduation	0.56	0.11	0.97	0.99	0.07	0.06	0.72	0.97
Late college dropout	0.71	0.77	0.62	0.17	0.08	0.89	0.45	0.89

TABLE 6
ECONOMIC IMPLICATIONS

State	Gross Return			Net Return		
	True	ML	SMM	True	ML	SMM
High school finishing	28%	35%	33%	66%	62%	138%
Early college enrollment	14%	18%	18%	−2%	−1%	−4%
Early college graduation	71%	75%	61%	48%	48%	93%
Late college enrollment	28%	30%	29%	−23%	−21%	−58%
Late college graduation	22%	22%	16%	9%	6%	24%
	ML			SMM		
RMSE	0.03416			0.29775		

NOTES: Statistics calculated for ML and SMM approaches based on 50,000 simulated agents using the parameter estimates. RMSE = root-mean-square error calculated in units of 100%.

is slightly worse than for the other states, as the agent count in those states is low. This problem affects the SMM estimates more than ML. The state frequencies are matched very well in both cases.

We report the root-mean-square error (RMSE) based on the difference between the simulated and observed statistics. There are only minor discrepancies for both estimation approaches. Nevertheless, they are slightly smaller for the ML results.

We apply χ^2 goodness-of-fit tests (Heckman, 1984; Heckman and Walker, 1990) to the estimated and actual probabilities. In Table 5, we report the p-value of a joint test of the relative share of agents within each state conditional on all possible realizations of selected covariates.[18]

Overall, the level of p-values is high. For ML estimation, all p-values indicate that our model is consistent with the data at the 5% significance level. In the case of SMM, we do not pass the test conditional on *Number of Children* only among early college enrollees. Because tests within covariates across all states are not independent, we use a Bonferroni test to evaluate the joint hypothesis that the predicted covariate distributions fit at each state. The test is based on the maximum χ^2 statistic over all states for each covariate. We pass a 5% Bonferroni test for all covariates and both estimation approaches.

4.3.2. *Economic implications.* Table 6 presents the median ex ante gross returns $GR^a(\hat{s}', \tilde{s}', s)$ and net returns $NR^a(\hat{s}', \tilde{s}', s)$ of pursuing a higher education by transitioning from s to $\hat{s}'$. Both

[18] In the χ^2 test, the predicted conditional distributions depends on estimated parameters. We do not adjust the test statistic to account for parameter estimation error as suggested by Heckman (1984) because the adjustments are usually slight (Heckman and Walker, 1990).

TABLE 7
STANDARD DEVIATIONS OF RANDOM COST SHOCKS

	$\hat{\sigma}_{\eta(\hat{s}',s)}$		
State	True	ML	SMM
High school finishing	0.27	0.24	0.61
Early college enrollment	0.20	0.19	0.47
Early college graduation	0.61	0.60	1.30
Late college enrollment	0.22	0.20	0.56
Late college graduation	0.61	0.60	1.30
RMSE		0.016	0.496

NOTE: RMSE = root-mean-square error.

types of returns capture all current and future earnings. However, they differ with regard to their inclusion of current and future costs. Their systematic parts are included in the calculation of the $NR^a(\hat{s}', \tilde{s}', s)$ but not the $GR^a(\hat{s}', \tilde{s}', s)$, as we discussed in Subsection 2.2.

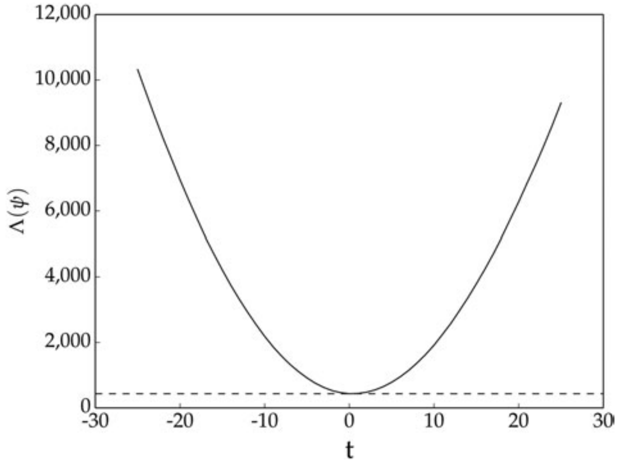
The estimates for the gross returns $GR^a(\hat{s}', \tilde{s}', s)$ are similar for the two approaches and close to their true values. However, for the net returns $NR^a(\hat{s}', \tilde{s}', s)$, only the ML results are close to the truth. The SMM results are off by up to a factor of 2. For example, the true net return of finishing high school is 66%, whereas SMM estimates 138%. The RMSE is roughly one order of magnitude larger for SMM than ML estimation. This difference is solely driven by the discrepancies in the net returns.

Table 7 sheds light on the poor performance of our SMM approach in the estimation of the net returns. In contrast to the gross returns, these returns include the current costs and the systematic part of all future costs of educational choices. SMM is unable to detect the systematic differences in the cost faced by agents. We overestimate the variance of the unobserved component determining choices $\sigma_{\eta(\hat{s}',s)}$. Too much of the agents' decisions is attributed to random cost shocks and not their systematic differences. This translates into an excess net return, as we underestimate the cost associated with future educational choices. Despite encouraging values for model fit criteria, SMM fails to accurately estimate the net return to educational choices.

We also explore the impact of a 50% reduction in tuition cost on agents' college-going decisions. We simulate 50,000 agents from our model and compare their educational choices under the baseline regime and the policy alternative using the results from the two estimation approaches. Based on the ML results, all policy predictions line up with the underlying truth. This is only partly true for the SMM estimation, where the predicted graduation rate for those induced to enroll in college late is too optimistic. Only a quarter will actually graduate, whereas the SMM results forecast about half. The SMM's failure to distinguish between the systematic and unsystematic cost components driving educational choices translates into (partly) flawed policy conclusions as well.

We now investigate the poor performance of our application of SMM and start with some evidence that we indeed recover a global minimum of our criterion function. Figure 6 shows the value of $\Lambda(\psi)$ around our SMM estimates as we perturb all parameters in a random direction in t increments. All perturbations increase the discrepancies between the observed and simulated sample. However, $\Lambda(\hat{\psi})$ is not zero because of remaining differences between estimated and true structural parameters. Even if we set $\hat{\psi} = \psi^*$, then $\Lambda(\psi^*)$ evaluates at 434 (horizontal dashed line) due to the random variation in agents' behaviors and state experiences. The moments provide noisy information about the data-generating process due to the random components. The more variation due to unobservables, the less information is contained in the data. As it turns out, the value of the criterion function evaluated at our estimates $\hat{\psi}$ is actually slightly smaller than $\Lambda(\psi^*)$.

Next we consider alternative choices for (a) set of moments $\hat{f}(\psi)$, (b) number of replications R, (c) weighting matrix W, and (d) optimization algorithm.



NOTE: Investigation using estimation sample of 5,000 agents with 30 replications.

FIGURE 6

CRITERION FUNCTION

4.3.3. *Set of moments.* We use the sequential schooling model of Heckman et al. (2014b) to inform our choice of moment conditions in the spirit of EMM estimation (Gallant and Tauchen, 1996). For our baseline, we match a number of conditional moments such as parameters of OLS regressions for within-state earnings and LP models characterizing the state transitions. We explicitly include determinants of future costs and earnings among the regressors in the LP models to capture the dynamics of agents' educational choices. We add aggregate statistics of the data such as average earnings and their standard deviations as well as state frequencies. In Table 8, we study the effects of using alternative sets of moment conditions. In particular, we specify a cross-sectional version in which we do not include future outcome covariates in the models of educational choice. We also study three alternative sets of dynamic moments. We increase their number from 440 up to 868, adding moments that provide additional information about the observed agent transitions. We thereby hope to improve the estimation of the systematic differences in the psychic cost of educational choices. We add a dynamic Probit model for each transition (Alt. A) and correlations of state outcomes and each covariate $(Y(s), X(s))$, between outcomes over time $(Y(s), Y(s'))$, and correlations of choice indicators with current cost covariates $(G(s'), Q(\hat{s}', s))$ (Alt. B).

We also report the value of the criterion function at the true structural parameters $\Lambda(\psi^*)$. Its difference from zero is solely driven by the presence of the random disturbances u_r. The final values of our criterion function are always below $\Lambda(\psi^*)$, which gives us further confidence that we have attained a global minimum in those cases.

We show the implications of alternative moments for the estimated median ex ante gross and net returns to education in Table 9. Once dynamic moments are included in the criterion function, the effect of adding even more is rather small. The estimates for the gross and net returns are all very similar. However, when using only cross-sectional moments for the criterion function, the performance of SMM deteriorates and its ability to recover the net returns is undermined further.

We assess the information content of selected moments $\hat{f}_i$ and investigate the effect of perturbations around $\hat{\psi}$. In Figure 7, we perturb the intercept in the structural earnings equation for early college graduates in t increments. This has a direct effect on average earnings in that state (Figure 8a). However, agents are forward-looking, and these changes also affect moments associated with earlier decisions such as finishing high school (Figure 8b). This is true even though the immediate benefits of doing so (Figure 8c) are unaffected. Agents change their

TABLE 8
SET OF MOMENTS

	Cross-Section Moments	Dynamic (Panel) Moments		
Sets	Base	Base	Alt. A	Alt. B
Outcome models				
Means	√	√	√	√
Standard deviations	√	√	√	√
Ordinary least squares	√	√	√	√
Correlations				√
Choice models				
State frequencies	√	√	√	√
Linear probability				
Cross section	√			
Dynamic		√	√	√
Probit				
Dynamic			√	√
Correlations				√
Overall statistics				
Number of moments	222	440	690	868
Number of replications	50	50	50	50
Weighting matrix	diagonal variance matrix			
Algorithm	POUNDerS			
Quality of fit measures				
$\Lambda(\hat{\psi})$	130.69	383.49	666.57	798.33
$\Lambda(\psi^*)$	222.12	434.07	685.94	847.64

NOTES: Alt. = Alternative.

early educational choices due to the increase in the option value of finishing high school, which includes the expected future value of potentially graduating from college.

4.3.4. *Number of replications.* For a given set of structural parameters, we create multiple simulated data sets from which we calculate the moments. Averaging over those moments, we reduce the effect of random components determining agents' choices and state experiences. In Figure 8, we show the value of the criterion function at the true structural parameters ψ^* for different numbers of replications R. The difference from zero is solely driven by the random components determining agents' choices and outcomes. If the model is simulated only once, then $\Lambda(\psi^*)$ takes value 825. Initially, increases in R result in a large drop of $\Lambda(\psi^*)$. However, this effect levels off after more than 20 replications. Afterwards, the value of $\Lambda(\psi^*)$ oscillates around 435. In a finite sample, differences between $\check{f}$ and $\hat{f}(\psi^*)$ remain even for a very large number of replications. Although the random values of $(\epsilon(s), \eta(\hat{s}', s))$ wash out in the simulated moments, their particular realizations remain relevant in the finite observed data.[19] For our baseline estimates we set $R = 30$. Further increases do not change model fit or economic implications.

4.3.5. *Weighting matrix.* Our optimization algorithm is only guaranteed to converge to local minimizers. Figure 9 plots the surface of our criterion function around ψ^* for two alternative choices of W given the true values of u_r. Thus, $\check{f} = \hat{f}(\psi^*)$ and $\Lambda(\psi^*)$ evaluates initially to zero

[19] See Kristensen and Salanié (2013) for a comprehensive statistical analysis of estimation methods, where the objective function is approximated through simulation or discretization.

TABLE 9

ROBUSTNESS OF ECONOMIC IMPLICATIONS OF ALTERNATIVE IMPLEMENTATIONS OF SMM

State	True	Cross-Section Moments	Dynamic (Panel) Moments		
		Base	Base	Alt. A	Alt. B
Gross return					
High school finishing	28%	38%	33%	34%	35%
Early college enrollment	14%	18%	18%	19%	19%
Early college graduation	71%	74%	61%	67%	61%
Late college enrollment	28%	17%	29%	25%	26%
Late college graduation	22%	18%	16%	19%	14%
Net return					
High school finishing	66%	154%	138%	137%	137%
Early college enrollment	−2%	−4%	−4%	−4%	−4%
Early college graduation	48%	89%	93%	94%	93%
Late college enrollment	−23%	−72%	−58%	−55%	−56%
Late college graduation	9%	16%	24%	22%	24%

NOTES: Statistics calculated for SMM based on 50,000 simulated agents using the parameter estimates.

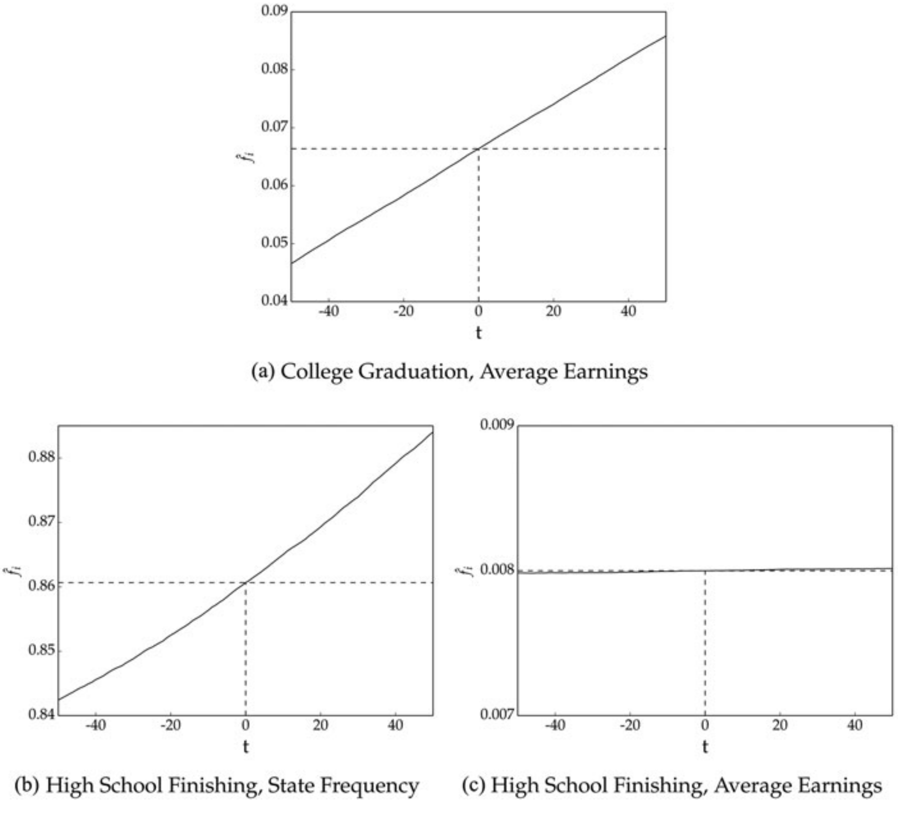
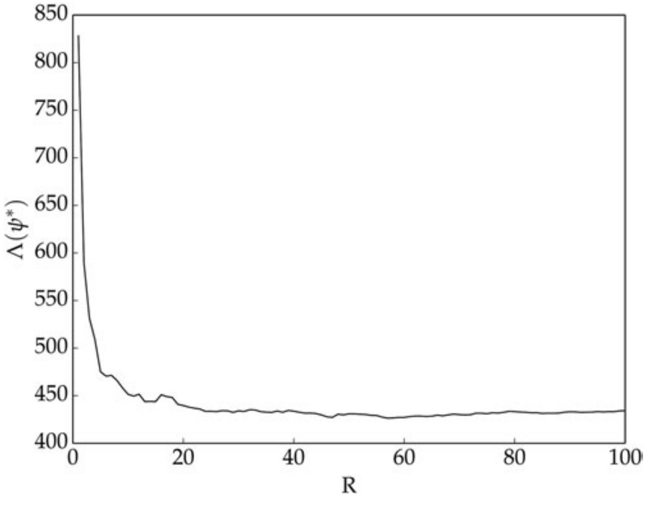


FIGURE 7

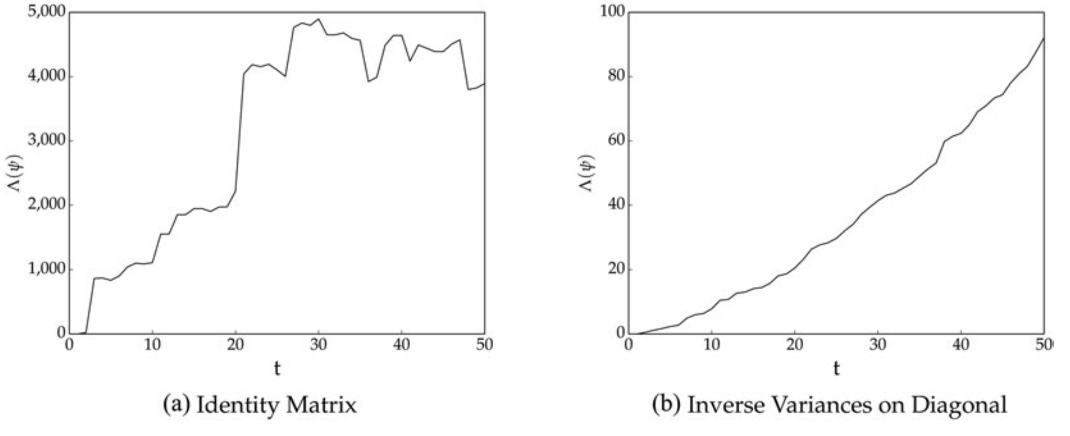
PARAMETER PERTURBATIONS, OUTCOME



NOTE: Investigation using estimation sample of 5,000 agents with varying number of replications.

FIGURE 8

ROLE OF REPLICATIONS



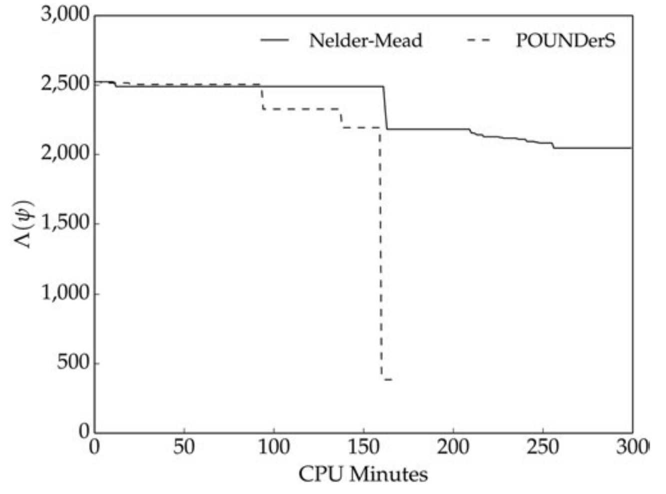
NOTE: Investigation using estimation sample of 5,000 agents with one replication and alternative weighting matrices.

FIGURE 9

ALTERNATIVE WEIGHTING MATRICES

regardless of the weighting matrix used. Then we perturb all the structural parameters in a random direction in t increments. We show the surface of $\Lambda(\psi)$ when either the identity matrix (Figure 10a) or the diagonal matrix with the variances of the moments (Figure 10b) is used. Choosing the identity matrix for W results in multiple local minima, whereas using the variances smoothes the overall surface of the criterion function.

4.3.6. *Optimization algorithm.* Because we repeat the SMM estimation many times for our Monte Carlo study, we benefit from a fast optimization algorithm. In Figure 10, we compare the performance of POUNDerS to the standard Nelder–Mead algorithm (Nelder and Mead, 1965) applied by Del Boca et al. (2014) and French and Jones (2011) among others. We perturb our estimates $\hat{\psi}$ and run the two algorithms as implemented in the Toolkit for Advanced Optimization (TAO; Munson et al., 2012) to investigate their relative performance. Following



NOTE: Investigation using estimation sample of 5,000 agents with 30 replications and all tuning parameters of the algorithms set to their default values.

FIGURE 10

OPTIMIZATION ALGORITHMS

Moré and Wild (2009) the solvers are tested using their default options.[20] Both algorithms are derivative-free but differ in their search strategy and how they exploit the structure of the criterion function. Nelder–Mead applies a direct search method, whereas POUNDerS forms an approximation model within a trust region, which exploits the special structure of our nonlinear least-squares problem. We show a minute-by-minute account of the criterion function $\Lambda(\psi)$ over five hours.

The POUNDerS algorithm attains a lower bound of $\Lambda(\psi) \approx 385$ after about two and a half hours and terminates. With the Nelder–Mead algorithm, the criterion function still takes a value of $\Lambda(\psi) \approx 2,050$ after five hours. Even after 36 hours, the Nelder–Mead solution $\Lambda(\psi) \approx 1{,}126$ is still about three times as large as the POUNDerS solution.

We are unable to improve the SMM results by using alternative tuning parameters. Our discussion cautions that inspection of model fit statistics alone does not guarantee accurate economic implications. For our model, large unobserved variation in educational choices translates into a noisy criterion function, which leaves SMM unable to recover the true returns to education. The structural variances of the unobserved cost shocks are poorly estimated; we are unable to correctly distinguish between the systematic and unsystematic cost components of educational choices.

Our results do not discredit SMM as a useful tool for the estimation of complex economic models. Our results are highly model dependent, but our diagnostics are not. We now outline a Monte Carlo exercise that allows SMM users to build confidence in their particular implementation in any applied setting.

4.3.7. *Monte Carlo exercise to gain confidence in an SMM algorithm.* Let $\mathcal{M}(\psi)$ denote the structural model parametrized by ψ, which is fit to the observed data D^{obs} to produce an estimated set of parameters $\hat{\psi}^{obs}$ using SMM.

[20] We are aware that performance can change for other choices. However, our practical experience throughout this project lines up with the results from this stylized presentation. We illustrate the relative performance of the two algorithms using a single processor only. Both algorithms allow parallel implementations as well (Lee and Wiswall, 2007; Munson et al., 2012).

Step 1: Simulate a synthetic sample D^{syn} from $\mathcal{M}(\hat{\psi}^{obs})$ using the estimated results.
Step 2: Fit $\mathcal{M}(\psi)$ on the synthetic sample D^{syn} using SMM to produce $\hat{\psi}^{syn}$.
Step 3: Compare $\hat{\psi}^{obs}$ to the results from the synthetic sample $\hat{\psi}^{syn}$.

Using the initial estimates as the parametrization for the Monte Carlo exercise ensures that important features of the data-generating process, in our case the large unobserved variability in agent behaviors, are accounted for. In Step 2, it is crucial to follow the same estimation approach applied to the original data as closely as possible, for example, choice of starting values. Combining the application of fast state-of-the-art optimization algorithms with parallel computing allows for such an analysis even in computation-intensive models.

This exercise showcases the performance of the estimator in a favorable setting as the model is correctly specified. If the structural parameters $\hat{\psi}^{syn}$ are successfully recovered, this is encouraging but does not provide a definite proof of the performance in the observed data. A failure, however, offers reason for concern. The algorithm might be improved, for example, by varying the set of moments used to test SMM.

5. CONCLUSION

We compare the performance of SMM and ML estimation in dynamic discrete choice models. We estimate a simplified dynamic model of educational choices, which emphasizes the role of unobserved heterogeneity, psychic costs, and option values for the net returns to schooling. The primary value of the model comes as input to the simulation study that is the core of this article.

We estimate our model on a sample of white males from the NLSY79. We discuss its implications for schooling decisions and present estimates of option values by cognitive and noncognitive factors. Given our estimates, we simulate a synthetic sample, creating a realistic setting to compare ML and SMM estimation. Our model allows for ML estimation without the need for any simulation in the likelihood function, which provides a clean comparison of ML against simulation-based estimation methods such as SMM. ML and SMM pass standard model fit tests. However, although the ML estimates are close to the true structural objects of interest, our version of SMM fails to recover the true net returns to education and policy effects. The SMM is unable to distinguish between systematic and unsystematic cost components driving educational choices.

We investigate alternative tuning parameters for implementing our SMM procedure. We specify alternative sets of moment conditions and show how the benefit of additional moments depends on the unique information they provide. Moments that capture the dynamics of agent behavior are crucial for getting reliable estimates of dynamic models. A large replication count in the simulation step reduces the effect of random noise in the measurement of the criterion function. An appropriate choice of the weighting matrix smoothes the surface of the criterion function and reduces the risk of local minima. Based on our analysis, we recommend that more exacting model specification tests and Monte Carlo evidence be provided to verify the performance of SMM in any application.

REFERENCES

ABBOTT, B., G. GALLIPOLI, C. MEGHIR, AND G. L. VIOLANTE, "Education Policy and Intergenerational Transfers in Equilibrium," Working Paper 18782, National Bureau of Economic Research, February 2013.

ADDA, J., C. DUSTMANN, C. MEGHIR, AND J.-M. ROBIN, "Career Progression, Economic Downturns, and Skills," Working Paper 18832, National Bureau of Economic Research, February 2013.

———, ———, AND K. STEVENS, "The Career Costs of Children," Working Paper 6201, Institute for the Study of Labor, December 2011.

ANDERSEN, T. G., L. BENZONI, AND J. LUND, "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance* 57 (2002), 1239–84.

———, H.-J. CHUNG, AND B. E. SORENSEN, "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Econometrics* 91 (1999), 61–87.

ARCIDIACONO, P., AND R. A. MILLER, "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica* 79 (2011), 1823–67.

BARTLETT, M. S., "The Statistical Conception of Mental Factors," *British Journal of Psychology* 1 (1937), 97–104.

BUREAU OF LABOR STATISTICS, *NLS Hanbook 2001: The National Longitudinal Surveys* (Washington, DC: U.S. Department of Labor, 2001).

CAMERON, S. V. AND J. J. HECKMAN, "The Nonequivalence of High School Equivalents," *Journal of Labor Economics* 11 (1993), 1–47.

CARNEIRO, P., K. HANSEN, AND J. J. HECKMAN, "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review* 44 (2003), 361–422.

CARRASCO, M., AND J.-P. FLORENS, "Simulation-Based Method of Moments and Efficiency," *Journal of Business & Economics and Statistics* 20 (2002), 482–92.

COMAY, Y., A. MELNIK, AND M. A. POLLATSCHEK, "The Option Value of Education and the Optimal Path for Investment in Human Capital," *International Economic Review* 14 (1973), 421–35.

CUNHA, F., J. J. HECKMAN, AND S. NAVARRO, "Separating Uncertainty from Heterogeneity in Life Cycle Earnings," *Oxford Economic Papers* 57 (2005), 191–261.

———, ———, AND ———, "The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds," *International Economic Review* 48 (2007), 1273–309.

———, ———, AND S. SCHENNACH, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica* 78 (2010), 883–931.

DEL BOCA, D., C. FLINN, AND M. WISWALL, "Household Choices and Child Development," *Review of Economic Studies* (2014), 137–85.

ECKSTEIN, Z., AND K. I. WOLPIN, "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey," *The Journal of Human Resources* 24 (1989), 562–98.

———, AND ———, "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica* 67 (1999), 1295–339.

FLINN, C., AND D. DEL BOCA, "Endogenous Household Interaction," *Journal of Econometrics* 166 (2012), 49–65.

———, AND J. MABLI, "On-the-Job Search, Minimum Wages, and Labor Market Outcomes in an Equilibrium Bargaining Framework," Working Paper 91, Carlo Alberto Notebooks, December 2008.

FRENCH, E., AND J. B. JONES, "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica* 79 (2011), 693–732.

GACH, F., AND B. M. PÖTSCHER, "Nonparametric Maximum Likelihood Density Estimation and Simulation-Based Minimum Distance Estimators," *Mathematical Methods of Statistics* 20 (2011), 288–326.

GALLANT, A. R., AND G. TAUCHEN, "Which Moments to Match?" *Econometric Theory* 12 (1996), 657–81.

GEWEKE, J. F., "Bayesian Inference in Econometric Models Using Monte-Carlo Integration," *Econometrica* 57 (1989), 1317–39.

GOURIÉROUX, C., AND A. MONFORT, *Simulation-Based Econometric Methods* (New York, NY: Oxford University Press, 1997).

———, ———, AND E. M. RENAULT, "Indirect Inference," *Journal of Applied Econometrics* 8 (1993), S85–S118.

HAJIVASSILIOU, V., AND D. L. MCFADDEN, "The Method of Simulated Scores for the Estimation of LDV Models," *Econometrica* 66 (1998), 863–96.

HANSEN, K., J. J. HECKMAN, AND K. MULLEN, "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics* 121 (2004), 39–89.

HECKMAN, J. J., "Heterogeneity and State Dependence," in S. Rosen, ed., *Studies in Labor Markets* (Chicago, IL: University of Chicago Press, 1981), 91–140.

———, "The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated from Microdata," *Econometrica* 56 (1984), 1543–7.

———, AND S. NAVARRO, "Dynamic Discrete Choice and Dynamic Treatment Effects,"*Journal of Econometrics* 136 (2007), 341–96.

———, AND J. R. WALKER, "The Relationship Between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data," *Econometrica* 58 (1990), 1411–41.

———, J. E. HUMPHRIES, AND T. KAUTZ, eds., *The Myth of Achievement Tests: The GED and the Role of Character in American Life* (Chicago, IL: University of Chicago Press, 2014a).

———, ———, AND G. VERAMENDI, "Education, Health and Wages," No. W19971, National Bureau of Economic Research, March 2014b.

———, L. J. LOCHNER, AND P. E. TODD, "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in E. A. Hanushek and F. Welch, eds., *Handbook of the Economics of Education*, volume 1 (Amsterdam, Netherlands: Elsevier, 2006a), 307–458.

———, R. Pinto, and P. A. Savelyev, "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review* 103 (2013), 2052–86.

———, J. Stixrud, and S. Urzúa, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics* 24 (2006b), 411–82.

Hotz, V. J., and R. A. Miller, "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies* 60 (1993), 497–529.

Johnson, M. T., "Borrowing Constraints, College Enrollment, and Delayed Entry," *Journal of Labor Economics* 31 (2013), 669–725.

Judd, K. L., and B. Skrainka, "High Performance Quadrature Rules: How Numerical Integration Affects a Popular Model of Product Differentiation," CeMMAP Working Papers, November 2011.

Keane, M., and K. I. Wolpin, "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation: Monte Carlo Evidence," *Review of Economics and Statistics* 76 (1994), 648–72.

———, P. E. Todd, and K. I. Wolpin, "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, volume 4A (Amsterdam, Netherlands: Elsevier, 2011), 331–461.

Keane, M. P., "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica* 62 (1994), 95–116.

———, and K. I. Wolpin, "The Career Decisions of Young Men," *Journal of Political Economy* 105 (1997), 473–522.

———, and ———, "The Effects of Parental Transfers and Borrowing Constraints on Education," *International Economic Review* 42 (2001), 1051–103.

Kortelainen, M., T. Lesinski, J. Moré, W. Nazarewicz, J. Sarich, N. Schunck, M. V. Stoitsov, and S. Wild, "Nuclear Energy Density Optimization," *Physical Review C* 82 (2010), 024313.

Krinsky, I., and L. A. Robb, "On Approximating the Statistical Properties of Elasticities," *The Review of Economics and Statistics* 68 (1986), 715–19.

———, and ———, "On Approximating the Statistical Properties of Elasticities: A Correction," *The Review of Economics and Statistics* 72 (1990), 189–90.

Kristensen, D., and B. Salanié, "Higher-Order Properties of Approximate Estimators," CeMMAP Working Paper No. CWP45/13, Centre for Microdata Methods and Practice, 2013.

Lee, D., and M. Wiswall, "A Parallel Implementation of the Simplex Function Minimization Routine," *Computational Economics* 30 (2007), 171–87.

———, and K. I. Wolpin, "Intersectoral Labor Mobility and the Growth of the Service Sector," *Econometrica* 74 (2006), 1–46.

———, and ———, "Accounting for Wage and Employment Changes in the U.S. from 1968-2000: A Dynamic Model of Labor Market Equilibrium," *Journal of Econometrics* 156 (2010), 68–85.

McFadden, D. L., "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica* 57 (1989), 995–1026.

Moré, J. J., and S. M. Wild, "Benchmarking Derivative-Free Optimization Algorithms," *SIAM J. Optimization* 20 (2009), 172–91.

Munson, T., J. Sarich, S. Wild, S. Benson, and L. C. McInnes, *TAO 2.0 Users Manual*, (Argonne, IL: Mathematics and Computer Science Division, Argonne National Laboratory, July 2012).

Nelder, J. A., and R. Mead, "A Simplex Method for Function Minimization," *The Computer Journal* 4 (1965), 308–13.

Nickl, R., and B. M. Pötscher, "Efficient Simulation-Based Minimum Distance Estimation and Indirect Inference," *Mathematical Methods of Statistics* 19 (2010), 327–64.

Nocedal, J., and S. J. Wright, *Numerical Optimization* (New York, NY: Springer, 2006).

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Fortran 77: The Art of Scientific Computing* (New York, NY: Cambridge University Press, 1992).

Raknerud, A. and O. Skare, "Indirect Inference Methods for Stochastic Volatility Models Based on Non-Gaussian Ornstein-Uhlenbeck Processes," *Computational Statistics and Data Analysis* 56 (2012), 3260–75.

Ruge-Murcia, F., "Estimating Nonlinear DSGE Models by the Simulated Method of Moments: With an Application to Business Cycles," *Journal of Economic Dynamics and Control* 36 (2012), 914–38.

Skrainka, B. S., "A Large Scale Study of the Small Sample Performance of Random Coefficient Models of Demand," SSRN Working Paper 1942627, April 2012.

Smith, A. A., Jr., and M. Keane, "Generalized Indirect Inference for Discrete Choice Models," Econometric Society 2004 North American Winter Meeting, 512, August 2004.

Trachter, N., "Stepping Stone and Option Value in a Model of Postsecondary Education," Working Paper 14-03, Federal Reserve Bank of Richmond, February 2014.

Weisbrod, B. A., "Education and Investment in Human Capital," *Journal of Political Economy* 70 (1962), 106–23.

Wolpin, K. I., *The Limits of Inference without Theory* (Cambridge, MA: MIT Press, 2013).