Web Appendix for Dynamic Treatment Effects*

James J. Heckman
University of Chicago
and the American Bar Foundation

John Eric Humphries
University of Chicago

Gregory Veramendi
Arizona State University

February 1, 2016

*James Heckman: Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637; phone: 773-702-0634; fax: 773-702-8490; email: jjh@uchicago.edu. John Eric Humphries: Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637; phone: 773-980-6575; email: johneric@uchicago.edu. Gregory Veramendi: Arizona State University, 501 East Orange Street, CPCOM 412A, Tempe, AZ 85287-9801; phone: 480-965-0894; email: gregory.veramendi@asu.edu. This research was supported in part by the American Bar Foundation, the Pritzker Children's Initiative, NIH grants NICHD R37HD065072, NICHD R01HD54702, and NIA R24AG048081, an anonymous funder, Successful Pathways from School to Work, an initiative of the University of Chicago's Committee on Education funded by the Hymen Milgrom Supporting Organization, a grant from the Institute for New Economic Thinking (INET) supporting the Human Capital and Economic Opportunity Global Working Group (HCEO), an initiative of the Center for the Economics of Human Development (CEHD) and Becker Friedman Institute for Research in Economics (BFI). The views expressed in this paper are those of the authors and not necessarily those of the funders or persons named here or the official views of the National Institutes of Health.

Contents

A Web Appendix

A.1 The Matzkin Conditions

Consider a binary choice model, $D = \mathbf{1}[\psi(\boldsymbol{Z}) > V]$, where Z is observed and V is unobserved. Let ψ^* denote the true ψ and let F_V^* denote the the true cdf of V. Let $\mathcal{Z} \subseteq \mathbb{R}^K$ denote the support of $\boldsymbol{Z}$. Let $\mathcal{H}$ denote the set of monotone increasing functions from $\mathbb{R}$ into $[0,1]$. Assume:

(i) $\psi \in \Psi$, where Ψ is a set of real valued, continuous functions defined over $\mathcal{Z}$, which is also assumed to be the domain of definition of Ψ, and the true function is $\psi^* \in \Psi$. There exists a subset $\widetilde{\mathcal{Z}} \subseteq \mathcal{Z}$ such that (a) for all $\psi, \psi' \in \Psi$, and all $z \in \tilde{\mathcal{Z}}$, $\psi(\boldsymbol{z}) = \psi'(\boldsymbol{z})$, and (b) for all $\psi \in \Psi$ and all t in the range space of $\psi^*(\boldsymbol{z})$ for $z \in \mathcal{Z}$, there exists a $\tilde{\boldsymbol{z}} \in \tilde{\mathcal{Z}}$ such that $\psi(\tilde{\boldsymbol{z}}) = t$. In addition, ψ^* is strictly increasing in the K-th coordinate of Z.

(ii) $\boldsymbol{Z} \perp\!\!\!\perp V$.

(iii) The K-th component of Z possesses a Lebesgue density conditional on the other components of $\boldsymbol{Z}$.

(iv) F_V^* is strictly increasing on the support of $\psi^*(\boldsymbol{Z})$. Matzkin (1992) notes that if one assumes that V is absolutely continuous, and the other conditions hold, one can relax the condition that ψ^* is strictly increasing in one coordinate (listed in i) and the requirement in (iii).

Then (ψ^*, F_V^*) is identified within $\psi \times \mathcal{H}$, where F_V^* is identified on the support of $\psi^*(\boldsymbol{Z})$.

Matzkin establishes identifiability for the following alternative representations of functional forms that satisfy condition (i) for exact identification for $\psi(Z)$.

1. $\psi(\boldsymbol{Z}) = \boldsymbol{Z\gamma}$, $\|\boldsymbol{\gamma}\| = 1$ or $\gamma_1 = 1$.
2. $\psi(\boldsymbol{z})$ is homogeneous of degree one and attains a given value α at $\boldsymbol{z} = \boldsymbol{z}^*$ (e.g. cost functions).
3. The $\psi(\boldsymbol{Z})$ are least concave functions that attain common values at two points in their domain.
4. The $\psi(\boldsymbol{Z})$ are additively separable functions:
   (a) Functions additively separable into a continuous monotone increasing function and a continuous monotone increasing function which is concave and homogeneous of degree one;
   (b) Functions additively separable into the value of one variable and a continuous, monotone increasing function of the remaining variables;
   (c) A set of functions additively separable in each argument (see Matzkin, 1992, example 5, p.255).

A.2 Derivations

$$\text{Define } \; E[\boldsymbol{Y_j} D_j] = \int\limits_{\mathcal{Y}_j} \boldsymbol{y_j} \int\limits_{\underline{\eta}_0}^{\phi_0(\boldsymbol{Z})} \cdots \int\limits_{\underline{\eta}_{j-1}}^{\phi_{j-1}(\boldsymbol{Z})} \int\limits_{\phi_j(\boldsymbol{Z})}^{\overline{\eta}_j} f_{\boldsymbol{y_j},\eta_0,\ldots,\eta_j}(\boldsymbol{y_j},\eta_0,\ldots,\eta_j) \quad d\boldsymbol{y_j} d\eta_0 \ldots d\eta_j \quad \text{(A.1)}$$

For $\quad j \in \{1,\ldots,\overline{s}-1\}$,

For $\quad j = \overline{s}$:

$$E[\boldsymbol{Y_s} D_{\overline{s}}] = \int\limits_{\mathcal{Y}_j} \int\limits_{\underline{\eta}_0}^{\phi_0(\boldsymbol{Z})} \cdots \int\limits_{\overline{\eta}_{\overline{s}-1}}^{\phi_{\overline{s}-1}(\boldsymbol{Z})} \boldsymbol{y_{\overline{s}}} f_{\boldsymbol{y_{\overline{s}}},\eta_0,\ldots,\eta_{\overline{s}-1}}(\boldsymbol{y_{\overline{s}}},\eta_0,\ldots,\eta_{\overline{s}-1}) d\boldsymbol{y_{\overline{s}}} d\eta_0 \ldots \eta_{\overline{s}-1}$$

For $\quad j = 0$:

$$E[\boldsymbol{Y_0} D_0] = \int\limits_{\mathcal{Y}_0} \boldsymbol{y_0} \int\limits_{\phi_0(\boldsymbol{Z})}^{\overline{\eta}_0} f_{\boldsymbol{y_0},\eta_0}(\boldsymbol{y_0},\eta_0) d\boldsymbol{y_0} d\eta_0.$$

For $j \in \{1,\ldots,\overline{s}-1\}$:

$$\begin{aligned}\frac{\partial E[\boldsymbol{Y_j} D_j]}{\partial Z^n} =& \sum_{l=0}^{j-1} \frac{\partial \phi_l(\boldsymbol{Z})}{\partial Z^n} \int\limits_{\mathcal{Y}_j} \overbrace{\int\limits_{\underline{\eta}_0}^{\phi_0(\boldsymbol{Z})} \int\limits_{\underline{\eta}_{l-1}}^{\phi_{l-1}(\boldsymbol{Z})}}^{l \text{ fold}} \overbrace{\int\limits_{\underline{\eta}_{l+1}}^{\phi_{l+1}(\boldsymbol{Z})} \cdots \int\limits_{\eta_{j-1}}^{\phi_{j-1}(\boldsymbol{Z})} \int\limits_{\phi_j(\boldsymbol{Z})}^{\overline{\eta}_j}}^{j-(l+1) \text{ fold}} \cdot \\ & \boldsymbol{y_j} f_{\boldsymbol{y_j},\eta_0,\ldots,\eta_j}(\boldsymbol{y_j},\eta_0,\ldots,\phi_l(\boldsymbol{Z}),\eta_{l+1},\ldots,\eta_j) \quad d\boldsymbol{y_j} \underbrace{d\eta_0,\ldots,d\eta_{l-1}}_{l \text{ fold}} \underbrace{d\eta_{l+1},\ldots,d\eta_j}_{j-(l+1) \text{ fold}} \\ & - \frac{\partial \phi_j(\boldsymbol{Z})}{\partial Z^n} \int\limits_{\mathcal{Y}_j} \int\limits_{\underline{\eta}_0}^{\phi_0(\boldsymbol{Z})} \cdots \int\limits_{\underline{\eta}_{j-1}}^{\phi_{j-1}(\boldsymbol{Z})} \boldsymbol{y_j} f_{\boldsymbol{y_j},\eta_0,\ldots,\eta_j}(\boldsymbol{y_j},\eta_0,\eta_1,\ldots,\eta_{j-1},\phi_j(\boldsymbol{Z})) \quad d\boldsymbol{y_j} d\eta_0,\ldots,d\eta_{j-1}\end{aligned}$$

For $\quad j = \overline{s}$:

$$\frac{\partial E[\boldsymbol{Y_{\overline{s}}} D_{\overline{s}}]}{\partial Z^n} = \sum_{l=0}^{\overline{s}} \frac{\partial \phi_l(\boldsymbol{Z})}{\partial Z^n} \int\limits_{\mathcal{Y}_{\overline{s}}} \int\limits_{\underline{\eta}_0}^{\phi_0(\boldsymbol{Z})} \cdots \int\limits_{\underline{\eta}_{\overline{s}-1}}^{\phi_{\overline{s}-1}(\boldsymbol{Z})} \boldsymbol{y_{\overline{s}}} f_{\boldsymbol{y_{\overline{s}}},\eta_0,\ldots,\eta_{\overline{s}-1}}(\boldsymbol{y_{\overline{s}}},\eta_0,\ldots,\eta_{\overline{s}-1}) d\boldsymbol{y_{\overline{s}}} d\eta_0 \ldots d\eta_{\overline{s}-1}$$

For $j = 0$:

$$\frac{\partial E[\boldsymbol{Y_0} D_0]}{\partial Z^n} = -\frac{\partial \phi_0(\boldsymbol{Z})}{\partial Z^n} \int\limits_{\mathcal{Y}_0} \boldsymbol{y_0} f_{\boldsymbol{y_0},\eta_0}(\boldsymbol{y_0},\phi_0(\boldsymbol{Z})) d\boldsymbol{y_0}.$$

Derivation of Expression (31)

For $l < j$,

$$\Omega(j,l) = \int_{\mathcal{Y}_j} \overbrace{\int_{\underline{\eta}_0}^{\phi_0(\boldsymbol{Z})} \dots \int_{\underline{\eta}_{l-1}}^{\phi_{l-1}(\boldsymbol{Z})}}^{l \text{ fold}} \overbrace{\int_{\underline{\eta}_{l+1}}^{\phi_{l+1}(\boldsymbol{Z})} \dots \int_{\phi_j(\boldsymbol{Z})}^{j-(l+1)}}^{j-(l+1) \text{ fold}} \cdot$$

$$\left\{ \boldsymbol{y_j} f_{\boldsymbol{y_j},\eta_0,\dots,\eta_j}(\boldsymbol{y_j}, \eta_0, \dots, \phi_l(\boldsymbol{Z}), \eta_{l+1}, \dots, \eta_j) d\boldsymbol{y_j} \underbrace{d\eta_0 \dots d\eta_{l-1}}_{l-\text{fold}}, \underbrace{d\eta_{l+1} \dots d\eta_j}_{j-(l+1) \text{ fold}} \right\}$$

and for $l = j$

$$\Omega(j,j) = -\int_{\mathcal{Y}_j} \int_{\underline{\eta}_0}^{\phi_0(\boldsymbol{Z})} \dots \int_{\underline{\eta}_{j-1}}^{\phi_{j-1}(\boldsymbol{Z})} \boldsymbol{y_j} f_{\boldsymbol{y_j},\eta_0,\dots,\eta_j}(\boldsymbol{y_j}, \eta_0, \eta_1, \dots, \phi_j(\boldsymbol{Z})) \quad d\boldsymbol{y_j} d\eta_0 \dots d\eta_{j-1}.$$

Explicit Expressions for $\tilde{\Omega}(j,l)$

For $l < j$

$$\tilde{\Omega}(j,l) =$$

$$\overbrace{\int_{\underline{\eta}_0}^{\phi_0(\boldsymbol{Z})} \dots \int_{\underline{\eta}_{l-1}}^{\phi_{l-1}(\boldsymbol{Z})}}^{l \text{ fold}} \dots \overbrace{\int_{\underline{\eta}_{l+1}}^{\phi_{l+1}(\boldsymbol{Z})} \dots \int_{\phi_j(\boldsymbol{Z})}^{\overline{\eta}_j}}^{j-(l+1) \text{ fold}} \cdot$$

$$f_{n_0}, \dots, \eta_j(\eta_{0_j}, \dots, \eta_{l-1}, \phi_l(\boldsymbol{Z}), \eta_{l+1}, \dots, \eta_j) \underbrace{d\eta_0 \dots d\eta_{l-1}}_{l-\text{fold}} \underbrace{d\eta_{l+1} \dots d\eta_j}_{j-(l+1)\text{fold}}$$

and for $l = j$

$$\tilde{\Omega}(j,j) = -\overbrace{\int_{\underline{\eta}_0}^{\phi_0(\boldsymbol{Z})} \dots \int_{\underline{\eta}_{j-1}}^{\phi_{j-1}(\boldsymbol{Z})}}^{j \text{ fold}} f_{\eta_0}, \dots, \eta_j(\eta_0, \dots, \eta_{j-1}, \phi_j(\boldsymbol{Z})) \quad d\eta_0 \dots d\eta_{j-1}.$$

B Description of the Data Used

This analysis uses the 1979 National Longitudinal Survey of Youth (NLSY79), a nationally representative sample of men and women born in the years 1957-64.The NLSY79 includes both a randomly chosen sample of 6,111 U.S. youth and a supplemental sample of 5,295 randomly chosen Black, Hispanic, and non-Black non-Hispanic economically disadvantaged youth. Both of these samples are drawn from the civilian population. In addition, there is a small sample of individuals (1,280) who were enrolled in the military in 1979. The respondents were first interviewed in 1979 when they were 14-22 years of age. The NLSY surveyed its participants annually from 1979 to 1992, and has surveyed them biennially since 1992. The NLSY measures a variety of later-life outcomes including labor market flows, asset and transfer income, and health outcomes. The survey measures many other aspects of the respondents' lives, such as scores on achievement tests, fertility, educational attainment, high school grades, and demographic information. This paper uses the core sample of males, which, after removing observations with missing covariates, contains 2242 individuals.[1]

B.1 Schooling Levels

We consider four different transitions and five final schooling levels. The transitions studied are (i) enrolled in high school deciding between graduating from high school and dropping out from high school (GEDs are treated as dropouts), (ii) high school graduates deciding whether or not to enroll in college, and (iii) college students deciding whether or not to earn a 4-year degree. Consequently, the final schooling levels are (I) high school dropout, (II) high school graduate, (III) some college and (IV) four-year college degree. Education at age 30 is treated as respondent's final schooling level.[2]

B.2 Measurement System

The cognitive and socioemotional factors in the model are identified from the joint estimation of the educational choices of agents as well as a supplemental measurement system of tests and other early-life outcomes. Sub-tests from the Armed Services Vocational Aptitude Battery (ASVAB) are used as measures of cognitive ability. Specifically, we consider the scores from Arithmetic Reasoning, Coding Speed, Paragraph Comprehension, World Knowledge, Math Knowledge, and Numerical Operations.[3]

[1]Respondents were dropped from the analysis if they did not have valid ASVAB scores, missed multiple rounds of interviews, had educational histories where true education could not be inferred, were missing control variables which could not be imputed, or had extreme and incomplete labor market histories. A number of imputations were made as necessary. Previous years' covariates were used when covariates where not available for a given year (such as region of residence). Responses from adjacent years were used for some outcomes when outcome variables were missing at the age of interest. Mother's education and father's education were imputed when missing.

[2]A negligible fraction of individuals change schooling levels after age 30.

[3]A subset of these tests are used to construct the Armed Forces Qualification Test (AFQT) score, which is commonly used as a measure of cognitive ability. AFQT scores are often interpreted as proxies for cognitive ability (Herrnstein and Murray, 1994). See the discussion in Almlund, Duckworth, Heckman, and Kautz (2011).

To identify the socioemotional factor, we use participation in minor risky or reckless activity in 1979 in the measurement system for the socioemotional endowment.[4] In order to identify the distribution of correlated factors, risky behavior is restricted to not load on the cognitive factor.

Many psychologists use a socioemotional taxonomy called the Big Five (John, Robins, and Pervin, 2008). This is an organizing framework that categorizes personality traits into 5 categories. The five traits are extraversion, agreeableness, conscientiousness, neuroticism, and openness. A growing body of work suggests that these traits and other socioemotional traits play key roles in academic success. Borghans, Golsteyn, Heckman, and Humphries (2011) and Almlund, Duckworth, Heckman, and Kautz (2011) show that the principal determinants of the grade point average are personality traits and not cognition. Similarly, Duckworth and Seligman (2005) find that self-discipline predicts GPA in 8th graders better than IQ. Duckworth, Quinn, and Tsukayama (2012) report three studies to show that self-control predicts grades earned in middle school better than IQ across racial and socioeconomic groups. Farsides and Woodfield (2003), Conard (2006), and Noftle and Robins (2007) find that Big 5 traits positively predict grades and academic success. These studies find predictive power after controlling for previous grades or test scores. In these studies, the benefits of personality traits are mediated through behaviors such as increased attendance or increased academic effort. A meta-analysis by Credé and Kuncel (2008) finds that study habits, skills, and attitudes have similar predictive power as standardized tests and previous grades in predicting college performance. They find that study skills are largely independent of high school GPA and standardized admissions tests, but have moderate correlations with personality traits.

The evidence that academic success (such as GPA) depends on cognitive ability, but also depends strongly on socioemotional traits such as conscientiousness, self-control, and self-discipline, motivates our identification strategy of including both a cognitive and socioemotional factor in 9th grade GPA. Much of the variance not explained through test scores has been shown to be related to socioemotional traits. Socioemotional skills are measured in part by their contribution towards 9th grade GPA in reading, social studies, science, and math.

GPA by grade and subject is constructed from high school transcript records. Up to 64 courses were recorded from school transcripts and included year taken, grade level taken, a class identification code, and the grade received. Using the class identification code, we identified all courses taken in either reading, social studies, science, or math in 9th grade and constructed subject level GPAs

As a robustness check for our measure of socioeconomic skills, we include five additional measures of adverse adolescent behavior to check our interpretation of the non-cognitive factor.[5] We consider violent behavior in 1979 (fighting at school or work and hitting or

[4]Preliminary data analysis suggested this measure was the least correlated with cognitive endowments among the available measures. This is a binary variable which is one if an agent answers yes to any of the following questions in 1980: "Taken something from the store without paying for it," "Purposely destroyed or damaged property that did not belong to you?," "Other than from a store, taken something that did not belong to you worth under $50?," and "Tried to get something by lying to a person about what you would do for him, that is, tried to con someone?"

[5]Gullone and Moore (2000) present a line of research which studies the relationship between personality traits and adolescent risk-behavior. Our five additional measures of early adverse behavior help demonstrate

threatening to hit someone), tried marijuana before age 15, daily smoking before age 15, regular drinking before age 15, and any intercourse before age 15. For violent behavior, we control for the potential effect of schooling. We estimate the cognitive and socioemotional distributions jointly with the educational choice system to account for the effect of schooling at the time of the measurement on measures of ability following the procedure developed in Hansen, Heckman, and Mullen (2004).

B.3 Control Variables

The variables used to control for observed characteristics depend on the timing and nature of the decision being made. In every outcome, measure, and educational choice, we control for race, broken home status, number of siblings, mother's education, father's education, and family income in 1979. We additionally control for region of residence and urban status at the time the relevant measure, decision, or outcome was determined.[6] For log wages at age 30, we additionally control for local economic conditions at age 30. When region of residence or urban status are not available for the age of a particular measure or outcome, the answer from previous or following surveys are used.

The educational choice models include additional choice-specific covariates. Following Carneiro, Heckman, and Vytlacil (2011), we control for both long run economic conditions, and contemporaneous deviations from those conditions. Controlling for the long-run local economic environment, local unemployment deviations capture contemporaneous economic shocks. The model for the choice to GED certify additionally controls for the difficulty of getting the GED within the state of residence in 1988.[7] The choices to enroll in college and graduate from college control for local 4-year college tuition at age 17 and 22 respectively.[8] When an instrument is missing for a particular age, the value from the previous or proceeding year is used.

The equation system for GPA controls for the variables used in all of our analyses, except for region dummies which are not available prior to 1979. The GPA model alternatively controls for urban status at age 14 and Southern residence at age 14. The ASVAB test scores models control for the standard controls, age, and age squared. As previously noted above, the ASVAB tests are estimated separately by education at the time of the test. Risky behavior in 1979 model controls for the standard controls, age and age squared. The risky behavior measure is also estimated by educational group, but due to data limitations pools high school graduates and those enrolled in college in 1979.

that our socioemotional factor is capturing traits that then explain these observed behaviors in an expected manner.

[6] Based on the data, we assume that high school, GED certification, and college enrollment decisions occur at age 17 while the choice to graduate from college is made at age 22.

[7] GED difficulty is proxied by the percent of high school graduates able to pass the test in one try given the state's chosen average and minimum score requirements.

[8] The cost of college, or the difficulty of earning a GED may affect an individual's choice to graduate from high school. In preliminary models, we found these "forward looking" variables to be statistically insignificant in the choice to graduate from high school and they are excluded from the high school choice.

B.4 Constructing the Data

As a baseline, our National Longitudinal Survey of Youth 1979 dataset uses the NLSY79 dataset used in Heckman, Stixrud, and Urzúa (2006) and Heckman (2001). Furthermore, we use instruments from Carneiro, Heckman, and Vytlacil (2011). We supplement this baseline dataset with grades from high school transcripts, risky behaviors at young ages, and later life outcomes that were not previously available, such as later earnings used to construct PV wages. Table B.1 provides an overview of how our base sample is constructed, and how many observations are lost at each point.

Table B.1: NLSY79 Data Set Construction and Effect of Deletions

Observations	Details
3,002	Core representative male NLSY population
2,975	require schooling defined (GED or HS) for 12 years completed
2,905	Not employed by military
2,763	Not enrolled in education at 30 years old
2,242	Require no missing education, covariates, ASVAB, Rosenberg, and, instruments (Heckman, Stixrud, and Urzúa (2006) sample)

B.4.1 Instrumental Variables

Using the factors estimated from our model, control variables, and decision-specific instruments, we estimate two-stage least squares for each educational choice. Instruments include long-run and current local unemployment rate, the difficulty of the GED exam, presence of a college in the county, and local college tuition. The IV estimates from the data are noisy with few instruments having statistically significant coefficients on educational choice. College tuition and local unemployment shocks have a statistically significant impact on the choice to enroll in college. Local unemployment shocks have a statistically significant impact on high school graduation.

As standard in two stage least squares, we estimate a first stage regression of D_j on θ, X_j, and Z_j and construct the linear projection $\hat{D}_j$. In the second stage we regress Y_j on D_j, θ, and X_j. X_j are control variables that include race, broken home status, number of siblings, mother's education, father's education, and family income in 1979.

C Specification of the Model

This section presents: (a) the specification and identification of our factor model (C.1), and (b) our likelihood (C.2).

C.1 Specification and Identification of the Factor Model

Let θ^C and θ^{SE} denote the levels of cognitive and socioemotional endowments and let $\boldsymbol{\theta} = (\theta^C, \theta^{SE})$. We allow θ^C and θ^{SE} to be correlated. Let $M^C_{m,s}$ be the m^{th} cognitive test score and $M^{C,SE}_{m,s}$ the m^{th} measure influenced by both cognitive and socioemotional endowments, all measured in state s. Parallel to the treatment of the index and outcome equations, we assume linear measurement systems:

$$M^C_{m,s} = \boldsymbol{X}^C_{m,s}\boldsymbol{\beta}^C_{m,s} + \theta^C\alpha^C_{m,s} + e^C_{m,s}, \tag{C.1}$$

$$M^{C,SE}_{m,s} = \boldsymbol{X}^{C,SE}_{m,s}\boldsymbol{\beta}^{C,SE}_{m,s} + \theta^C\tilde{\alpha}^C_{m,s} + \theta^{SE}\tilde{\alpha}^{SE}_{m,s} + e^{C,SE}_{m,s}. \tag{C.2}$$

The structure assumed in Equations (C.1) and (C.2) is identified even allowing for correlated factors, if we have one measure that is a determinant of cognitive endowments ($M^C_{m,s}$) and at least four measures that load on both cognitive ability and socio-emotional ability, and conventional normalizations are assumed.[9] In Heckman, Humphries, and Veramendi (2016), we test if additional unobservables beyond θ^C and θ^{SE} are required to capture the dependence between schooling and outcomes beyond that arising from observables. Our empirical estimates are essentially unchanged when we introduce a third factor to capture dependencies between schooling and outcomes not captured by the proxy factors. In the main text we report results from models that use measurements to proxy $\boldsymbol{\theta}$.

C.2 Likelihood

We estimate our model in two stages using maximum likelihood. The measurement system, the distribution of latent endowments, and the model of schooling decisions are estimated in the first stage. The outcome equations are estimated in the second stage using estimates from the first stage. We follow Hansen, Heckman, and Mullen (2004), and correct estimated factor distributions for the causal effect of choices on measurements by jointly estimating the choice and measurement equations in the first stage. The distribution of the latent factors is estimated only using data on educational choices and measurements. This allows us to interpret the factors as cognitive and socioemotional endowments. It links our estimates to an emerging literature on the economics of personality and psychological traits but the link is not strictly required if we only seek to control for selection in schooling choices and do not seek to identify the system of measurement equations presented in the text. We do not use the final outcome system to estimate the distribution of factors, thus avoiding tautologically strong predictions of outcomes from the system of estimated factors.

For convenience, we repeat the definitions from Section 2. Let $\mathcal{J}$ denote the set of possible terminal states. Let $D_j \in \mathcal{D}$ be the set of possible transition decisions that can be taken by the individual over the decision horizon. Let $\mathcal{S}$ denote the finite and bounded set of stopping states with $S = s$ if the agent stops at $s \in \mathcal{S}$. Define $\overline{s}$ as the highest attainable element in $\mathcal{S}$. $Q_j = 1$ indicates that an agent *gets to* decision node j. $Q_j = 0$ if the person never gets there. The history of nodes visited by an agent can be described by the collection of the Q_j such

[9] See, e.g., the discussion in Williams (2011) and Anderson and Rubin (1956). One of the factor loadings for both θ^C and θ^{SE} has to be normalized to set the scale of the factors. Nonparametric identification of the distribution of $\boldsymbol{\theta}$ is justified by an appeal to the results in Cunha, Heckman, and Schennach (2010).

that $Q_j = 1$. To ensure consistent notation, we define $Q_0 := 1$. $\boldsymbol{Y}_i$, $\boldsymbol{D}_i$, and $\boldsymbol{M}_i$ are vectors of individual i's outcomes, educational decisions and measurements of endowments, respectively. $\boldsymbol{Z}$ is a vector of observed determinants of decisions, $\boldsymbol{X}$ is a vector of observed determinants of outcomes, and $\boldsymbol{\theta}$ is the vector of unobserved endowments. The $\boldsymbol{Z}$ can include all variables in $\boldsymbol{X}$. When instrumental variable methods are used to identify components of the model, it is assumed that there are some variables in $\boldsymbol{Z}$ not in $\boldsymbol{X}$.

Although the decision structure of the ordered and unordered models are quite different, their likelihoods have a similar structure. The sets of possible transition decisions are, for the ordered model: $\mathcal{D}^{ordered} = \{D_0, ..., D_{\bar{s}-1}\}$; and for our unordered model: $\mathcal{D}^{unordered} = \{D_G, D_0, ..., D_{\bar{s}-1}\}$. Note that $Q_0 := 1$, so the likelihood will always evaluate the decision node D_0. Also note that the set of stopping states are $\mathcal{S}^{ordered} = \{0, ..., \bar{s}\}$ for the ordered model and $\mathcal{S}^{unordered} = \{G, 0, ..., \bar{s}\}$ for our unordered model. Hence the set $\{D_j | j \in \mathcal{S} \backslash \bar{s}\}$ is the set of possible transition nodes for both models ($\mathcal{D}$).

Assuming independence across individuals (denoted by i), the likelihood is:

$$\begin{aligned}\mathcal{L} &= \prod_i f(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{M}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i) \\ &= \prod_i \int f(\boldsymbol{Y}_i | \boldsymbol{D}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}) f(\boldsymbol{D}_i, \boldsymbol{M}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta},\end{aligned}$$

where $f(\cdot)$ denotes a probability density function. The last step is justified from the assumptions (A-1a) – (A-1g).

For the first stage, the sample likelihood is

$$\begin{aligned}\mathcal{L}^1 &= \prod_i \int_{\overline{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} f(\boldsymbol{D}_i, \boldsymbol{M}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) f_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}) \, d\overline{\boldsymbol{\theta}} \\ &= \prod_i \int_{\overline{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} \left[\prod_{j \in \mathcal{S} \backslash \{\bar{s}\}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_j)^{Q_{i,j}} \right] \\ &\qquad \times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_{m,s})^{H^m_{i,s}} \right] f_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_\theta) \, d\overline{\boldsymbol{\theta}}\end{aligned}$$

where we integrate over the distributions of the latent factors. H^m_s is an indicator for the level of the choice variable at the time the measurement m is taken and is equal to one if the individual had attained s at the time of the measurement and zero otherwise. Let $\mathcal{S}^M$ denote the set of possible states at the time of the measurement. The goal of the first stage is to secure estimates of $\boldsymbol{\gamma}_j$, $\boldsymbol{\gamma}_{m,s}$ and $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$, where $\boldsymbol{\gamma}_j$, $\boldsymbol{\gamma}_{m,s}$ and $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ are the parameters for the educational decision models, the measurement models and the factor distribution, respectively. We assume that the idiosyncratic shocks are mean zero normal variates.

We approximate the factor distribution using a mixture of normals.[10] We define the index,

[10]Mixtures of normals can be used to identify the true density nonparametrically, where the number of mixtures can be increased based on the size of the sample. For a discussion of sieve estimators, see Chen (2007).

ℓ, for each mixture, where $f_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\gamma_\theta}) = \sum_\ell \rho_\ell f^\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\gamma}^\ell_{\boldsymbol{\theta}})$. The weights for each mixture are ρ_ℓ and they must satisfy $\sum_\ell \rho_\ell = 1$. $f^\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\gamma}^\ell_{\boldsymbol{\theta}})$ is the PDF for mixture ℓ. Since the mean of the overall factor distribution is not identified, we also require that $E[\boldsymbol{\theta}] = \mathbf{0}$ which places constraints on the mixture parameters $\boldsymbol{\gamma}^\ell_{\boldsymbol{\theta}}$. The log-likelihood can be rewritten as

$$\begin{aligned}
\log \mathcal{L}^1 = \sum_i \log \int_{\overline{\boldsymbol{\theta}}\in\boldsymbol{\Theta}} & \left[\prod_{j\in\mathcal{S}\setminus\overline{s}} f(\boldsymbol{D}_{i,j}|\boldsymbol{Z}_{i,j},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}_j)^{Q_{i,j}}\right] \\
& \times \left[\prod_{m=1}^{N_M}\prod_{s\in\mathcal{S}^M} f(\boldsymbol{M}_{i,m,s}|\boldsymbol{X}_{i,m,s},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}_{m,s})^{H^m_{i,s}}\right] \times \left[\sum_\ell \rho_\ell f^\ell_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}^\ell_{\boldsymbol{\theta}})\right] d\overline{\boldsymbol{\theta}} \\
= \sum_i \log \Bigg\{ \sum_\ell \rho_\ell \int_{\overline{\boldsymbol{\theta}}\in\boldsymbol{\Theta}} & \left[\prod_{j\in\mathcal{S}\setminus\overline{s}} f(\boldsymbol{D}_{i,j}|\boldsymbol{Z}_{i,j},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}_j)^{Q_{i,j}}\right] \\
& \left[\prod_{m=1}^{N_M}\prod_{s\in\mathcal{S}^M} f(\boldsymbol{M}_{i,m,s}|\boldsymbol{X}_{i,m,s},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}_{m,s})^{H^m_{i,s}}\right] f^\ell_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}};\boldsymbol{\gamma}^\ell_{\boldsymbol{\theta}})d\overline{\boldsymbol{\theta}} \Bigg\}.
\end{aligned}$$

We use Gauss-Hermite quadrature to numerically evaluate the integral. Although there are a number of ways to numerically evaluate an integral, one advantage of Gaussian quadrature is that it gives analytical expressions for the integral. Analytical expressions for the gradient and hessian can then be calculated which allows for the use of efficient second-order optimization routines. Since the models are very smooth, a second-order optimization strategy leads to faster convergence. Given that we are using a mixture of normals, $f^\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\gamma}^\ell_{\boldsymbol{\theta}}) = \phi(\boldsymbol{\theta};\boldsymbol{\mu}^\ell_{\boldsymbol{\theta}},\boldsymbol{\sigma}^\ell_{\boldsymbol{\theta}})$ is a multivariate normal, where we assume for now that the components are independent. This assumption can easily be relaxed, but keeping it simplifies notation. The Gauss-Hermite quadrature rule is $\int f(v)e^{-v^2}dv = \sum_n \lambda_n f(v_n)$, where the weights, λ_n, and nodes, v_n, are defined by the quadrature rule depending on the number of points used (Judd, 1998).[11] Applying the Gauss-Hermite rule and making a change of variables ($\overline{\boldsymbol{\theta}} = \sqrt{2}\boldsymbol{\sigma}^\ell_{\boldsymbol{\theta}} \circ \boldsymbol{v}_n + \boldsymbol{\mu}^\ell_{\boldsymbol{\theta}}$)[12], we can rewrite the likelihood as

$$\begin{aligned}
\log \mathcal{L}^1 = \sum_i \log \Bigg\{ \sum_\ell \rho_\ell \sum_{n1} \lambda_{n1} \sum_{n2} \lambda_{n2} & \left[\prod_{j\in\mathcal{S}\setminus\overline{s}} f(\boldsymbol{D}_{i,j}|\boldsymbol{Z}_{i,j},\boldsymbol{\theta}=\sqrt{2}\boldsymbol{\sigma}^\ell_{\boldsymbol{\theta}}\circ\boldsymbol{v}_n+\boldsymbol{\mu}^\ell_{\boldsymbol{\theta}};\boldsymbol{\gamma}_j)^{Q_{i,j}}\right] \\
& \times \left[\prod_{m=1}^{N_M}\prod_{s\in\mathcal{S}^M} f(\boldsymbol{M}_{i,m,s}|\boldsymbol{X}_{i,m,s},\boldsymbol{\theta}=\sqrt{2}\boldsymbol{\sigma}^\ell_{\boldsymbol{\theta}}\circ\boldsymbol{v}_n+\boldsymbol{\mu}^\ell_{\boldsymbol{\theta}};\boldsymbol{\gamma}_{m,s})^{H^m_{i,s}}\right] \Bigg\}
\end{aligned}$$

where $\boldsymbol{v_n} = \begin{pmatrix} v_{n1}, & v_{n2} \end{pmatrix}$ represents the vector of nodes. Multivariate normal variables with correlated components can be rewritten as the sum of independent standard normal variables and then one can use the same procedure.

[11] We use 16 quadrature points. Using 32 point did not substantively change any of our results.

[12] $\circ$ is the Hadamard or entrywise product.

The goal of the first stage is then to maximize $\log \mathcal{L}^1$ and obtain estimates $\hat{\boldsymbol{\gamma}}_j$, $\hat{\boldsymbol{\gamma}}_{m,s}$, $\hat{\boldsymbol{\sigma}}^{\ell}_{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\mu}}^{\ell}_{\boldsymbol{\theta}}$, and $\hat{\rho}_\ell$ for $j \in J^{MS}$. If a density $f(\cdot)$ cannot be calculated either because of missing data or because that model does not apply to individual i[13], then $f(\cdot) = 1$.

One can think of the inner brackets as the PDF of $\boldsymbol{\theta}$ for each individual i. This is useful in two respects. First, we can now predict the factorscores $(\hat{\boldsymbol{\theta}}_i)$ via maximum likelihood where the likelihood for each individual i is

$$\mathcal{L}^{\theta}_i = \left[\prod_{j \in \mathcal{S} \setminus \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{\theta}_i; \hat{\boldsymbol{\gamma}}_j)^{Q_{i,j}}\right] \times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta}_i; \hat{\boldsymbol{\gamma}}_{m,s})^{H^m_{i,s}}\right]$$

Secondly, we can correct for measurement error in the outcome equations by integrating over the PDF of the latent factor. The likelihood for the outcome equations is

$$\begin{aligned}\log \mathcal{L}^2_k = \sum_i \log \Bigg\{ \sum_\ell \rho_\ell \sum_{n1} \lambda_{n1} \sum_{n2} \lambda_{n2} & \left[\prod_{j \in \mathcal{S} \setminus \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{\theta} = \sqrt{2}\hat{\boldsymbol{\sigma}}^{\ell}_{\boldsymbol{\theta}} \circ \boldsymbol{v}_n + \hat{\boldsymbol{\mu}}^{\ell}_{\boldsymbol{\theta}}; \hat{\boldsymbol{\gamma}}_j)^{Q_{i,j}}\right] \\ & \times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \sqrt{2}\hat{\boldsymbol{\sigma}}^{\ell}_{\boldsymbol{\theta}} \circ \boldsymbol{v}_n + \hat{\boldsymbol{\mu}}^{\ell}_{\boldsymbol{\theta}}; \hat{\boldsymbol{\gamma}}_{m,s})^{H^m_{i,s}}\right] \\ & \times \left[\prod_{s \in \mathcal{S}} f(\boldsymbol{Y}^k_{i,s} | \boldsymbol{X}_{i,k,s}, \boldsymbol{\theta} = \sqrt{2}\hat{\boldsymbol{\sigma}}^{\ell}_{\boldsymbol{\theta}} \circ \boldsymbol{v}_n + \hat{\boldsymbol{\mu}}^{\ell}_{\boldsymbol{\theta}}; \boldsymbol{\gamma}_{s,k})^{H_{i,s}}\right] \Bigg\}.\end{aligned}$$

where $H_{i,s}$ is an indicator for the highest level of schooling attained by individual i. The goal of the second stage is to maximize $\log \mathcal{L}^2_k$ and obtain estimates $\hat{\boldsymbol{\gamma}}_{s,k}$. Since outcomes $(\boldsymbol{Y}^k_s)$ are independent from the first stage outcomes conditional on $\boldsymbol{X}, \boldsymbol{\theta}$ and we impose no cross-equation restrictions, we obtain consistent estimates of the parameters for the adult outcomes. Standard errors and confidence intervals are calculated by estimating two hundred bootstrap samples for the combined stages.

D Counterfactual Educational Attainment

Table D.1 shows how many individuals who stopped education as a dropout or a high school graduate would have continued on to additional education if forced to make their terminal educational transition. The first row shows the proportion of high school dropouts who would have enrolled in college and graduated from college if forced to graduate from high school. The second rows shows the proportion of terminal high school graduates who would have graduated from college if forced to enroll in college. The third row shows the proportion of high school graduates who would have earned a GED if they were forced to drop out of high school. All estimates are generated by simulations from our main model.

[13]For example, the individual i is a high school dropout and the model corresponds to the graduate college decision.

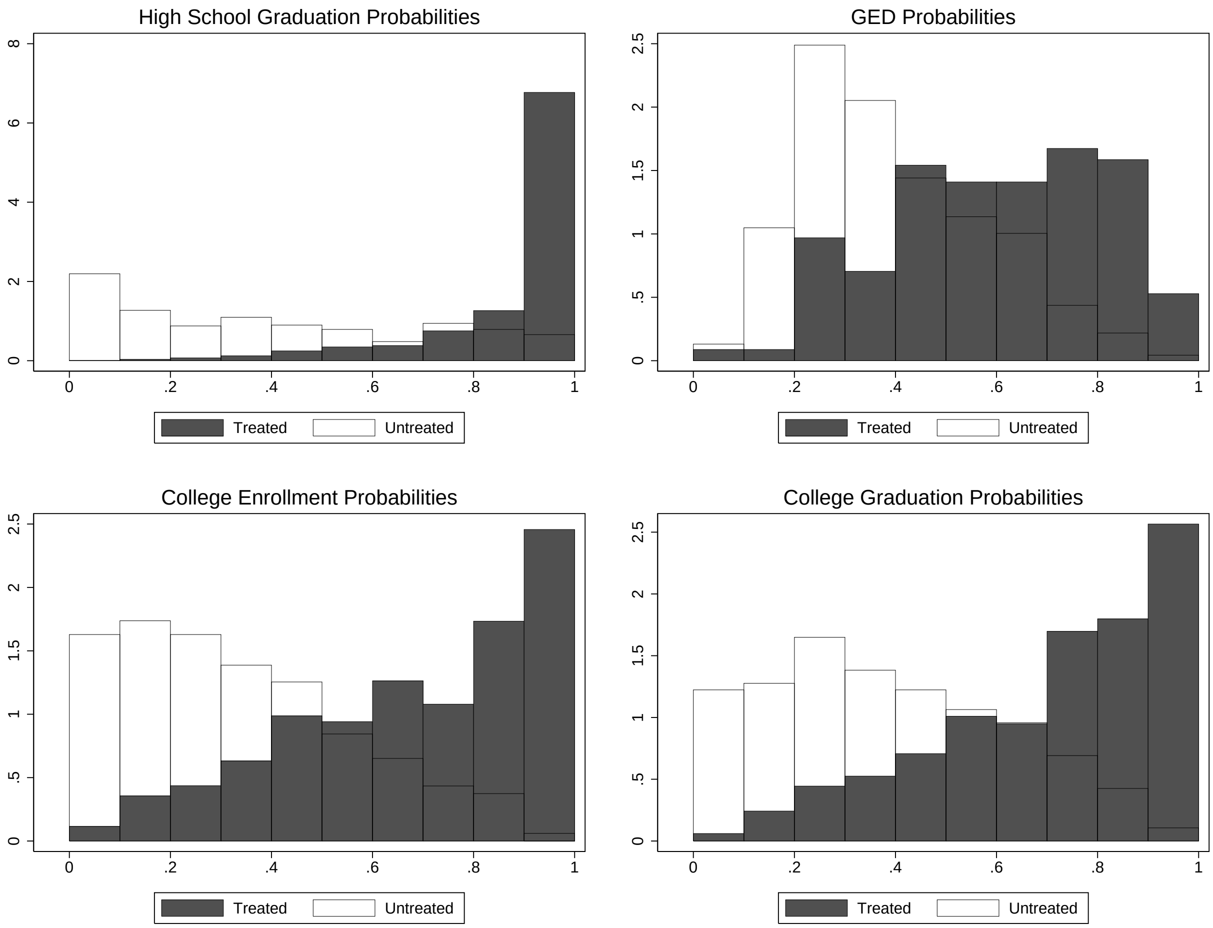
Table D.1: Counterfactual Educational Attainment

	Coll. Enroll.	Coll. Grad.	GED
Dropouts	0.180	0.038	.
Terminal HS Graduates	.	0.279	.
High School Graduates	.	.	0.739

Notes: The first row of the table shows the estimated proportion of individuals who would have enrolled in college if they had not dropped out of high school and the proportion who would have enrolled and completed college if they had not dropped out of high school. The second row shows the estimated proportion of terminal high school graduates who would have graduated from 4-year college if they had enrolled. The third row estimates the proportion of high school graduates who would have earned a GED if forced to not graduate from high school.

E Properties of the Estimated Model and Some Simulations from It

Figure E.1: Supports of $Pr(D_j = 0|Z, X, Q_j = 1)$ at Each Decision Node



Notes: Each plot is for the population who reaches that decision node in the data. “Treated” are those who choose to complete the reported level of schooling, while “Untreated” are those who choose to not complete the reported level of schooling (but reach the decision node). Probabilities are estimated by a probit model that controls for race, broken home status, number of siblings, mother’s education, father’s education, and family income in 1979.

F Forward-Looking Educational Decisions

Table F.1 below shows the coefficients on the various instruments in our model, where we allow tuition and presence of a local college to affect high school graduation decisions. In a forward looking model, instruments that affect future decisions, and are anticipated, should also affect earlier decisions. We find that local tuition and presence of a college in the county have small and statistically insignificant impacts when included in the high school graduation decision, although they affect college enrollment decisions.

Table F.1: Forward Looking Instruments in Educational Decisions

Variable	D_0: Graduate HS vs. Drop out of HS		D_G: **GED** vs. **HS Dropout**		D_1: Enroll College vs. **HS Graduate**		D_2: **4-year College Degree** vs. **Some College**	
	β	StdErr.	β	StdErr.	β	StdErr.	β	StdErr.
4-year College in County	**0.046**	**0.087**	**0.105**	**0.139**	0.152	0.080	0.060	0.106
Tuition (age 17)	**0.065**	**0.065**	**-0.012**	**0.100**	-0.255	0.062		
Long-run Local Unemployment	-10.295	4.812	-2.785	7.330	-6.051	4.547	-5.395	4.924
Local Unemployment (age 17)	1.918	2.448	5.775	3.825	4.511	2.330		
Local Unemployment (age 22)							-0.044	1.908
GED exam HS pass rate			0.002	0.069				
Tuition (age 22)							-0.015	0.086
N	2242		522		1720		891	

Notes: This table shows the loadings from the various instruments in each educational decision when the model allows future values of variables to enter stage-specific choices. We present a general model for all transitions. However, the variables associated with the bolded numbers are deleted in the model producing the estimates in this paper. Other controls include race, broken home status, number of siblings, mother's education, father's education, and family income in 1979 and are not shown.

G Does a Strong Instrument Identify Economically Meaningful Parameters?

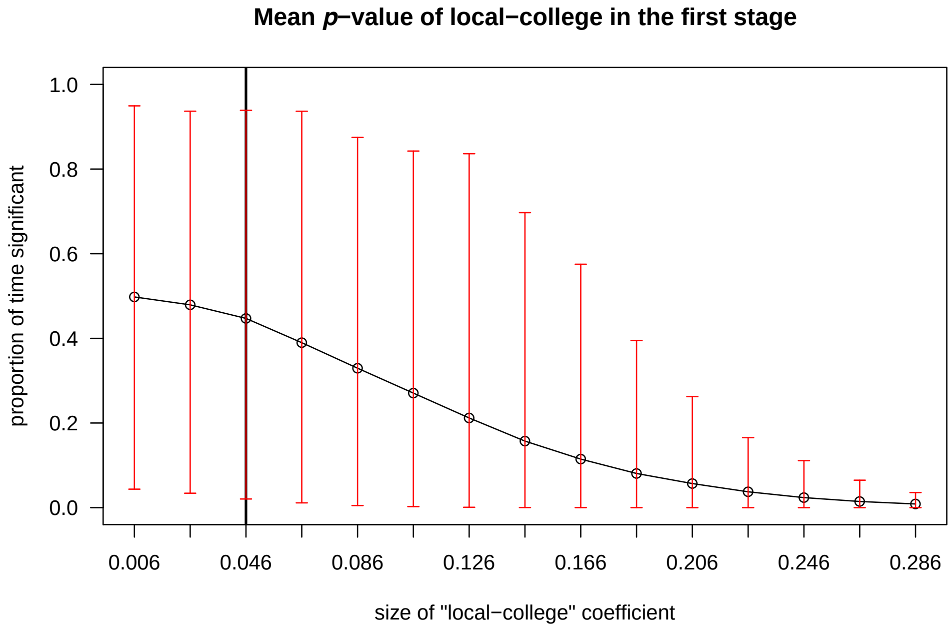
Using simulation, this section explores the relationship of the sampling variance of the IV estimator and the strength of the instrument as measured by its predictive power in the first stage. We also explore the ability of generated instruments to recover economically interpretable parameters when instruments are strong. Using our baseline model as a point of departure, we create a series of counterfactual simulations. Across the simulations we vary the size of the coefficient on "presence of a college in county" variable in the first stage (this is the only model parameter that we vary). Running IV on these simulated data sets, we explore the properties of IV estimators. In a second stage, we simulate outcomes using the simulated first stage to predict choices generated by the instrument. *Ceteris paribus* higher (in absolute value) coefficients have greater predictive power. We apply IV to the model using the choices and outcomes so generated. This exercise gives us a measure of the ability of IV methods to recover economically interpretable parameters.

G.1 First Stage Analysis

Starting with a baseline coefficient of 0.46[14], we simulate new data with local-college coefficient starting at 0.006 and increasing to 0.286 by increments of 0.02. Figure G.1 plots the average p-value as well as the p-values for the 5th and 95th percentile from 400 simulated data sets of 2242 observations for each of the 15 simulated coefficient values. The X-axis shows the assumed value of the coefficient on local-college while the Y-axis shows the mean p-value across the 400 samples as well as the 5th and 95th percentile values. For each counterfactual value, we draw 400 simulated data sets of 2242 observations and, for each, we run a probit with covariates that match those in the true data generating process. Using the simulated data sets, The solid line corresponds to the insignificant point-estimate we find for the coefficient on local-college in the high school graduation decision from Table F.1. Overall, we find that the estimated value of 0.046 for "presence of a college in county" gives a very weak instrument, but that, as expected, samples with a larger coefficient produce stronger instruments and display much lower p-values.

[14] This estimate is taken from Table F.1, which shows coefficients from our model when local-college and local college tuition enter the high school graduation decision

Figure G.1: IV Simulation: first stage



NOTES: Figure shows the mean p-value (and 5th and 95th percentile) from 400 simulations of 2242 individuals for 15 different counterfactual values of the college-in-county instrument. The IV regression includes a dummy indicating whether or not the individual graduates from high school or not, cognitive and non-cognitive factors, race indicators, broken home indicators, parents' education, region indicators, an urban residence indicator, and average local unemployment rates. The high school graduation dummy is instrumented with the counter-factual college-tuition dummy and local unemployment rates at age 17. The solid line indicates the (statistically insignificant) estimate on local-college reported in Table F.1.

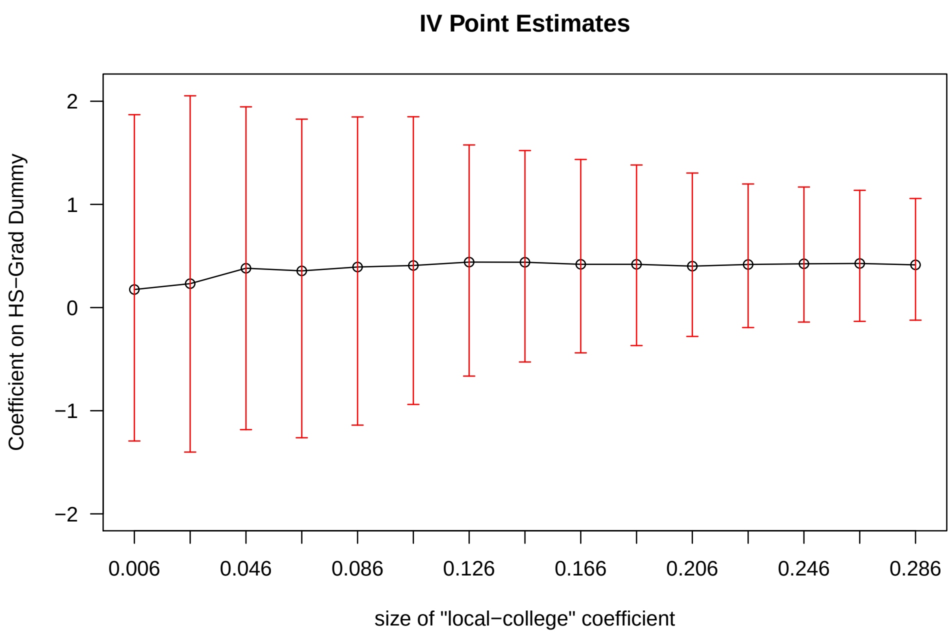
G.2 IV Estimates

This section reports results from the second stage of two stage least squares analysis of IV for our simulated datasets. For each counter-factual value of the local-college instrument we simulate 400 datasets with 2242 observations and use TSLS to estimate treatment effects in each data set. The TSLS regression includes a dummy indicating if the individual graduates from high school or not, cognitive and non-cognitive factors, race indicators, broken home indicators, parents' education, region indicators, an urban residence indicator, and average local unemployment rates. Two stage least squares is used to predict the model generated high school graduation indicator dummy using tuition, measures of ability and local unemployment rates at age 17. The high school graduation dummy is instrumented with the counter-factual college-tuition dummy and local unemployment rates at age 17.

Figure G.2 shows the average coefficient on the high school graduation dummy for each counterfactual value of the instrument in the 2SLS regression. 95% confidence intervals are presented. Given different samples of TSLS, IV estimates vary wildly, even when the instrument is strong, ranging from large positive returns to large negative returns.

Comparing IV from a strong instrument and the estimated treatment effects Even with a strong instrument and simulated data, the IV estimates widely vary for a sample of 2242 observations, ranging from more than 1 to nearly -1. Averaging across the simulations with a local-college coefficient of 0.286, IV estimates are on average 0.41, which is far from the treatment effects found in the first column of Table 1 in the main paper (which estimates an AMTE and ATE of 0.09). IV does not produce interpretable or policy relevant estimates within a dynamic framework where agents make multiple educational decisions and the returns to skills and characteristics vary by schooling-level. Notably, TSLS estimates for the college enrollment decision (where we have the best instruments and find a weakly statistically significant point estimate) also produces treatment effects much larger than our estimate ATE or AMTE (with the TSLS estimator having a point estimate of 0.51 compared to an ATE of 0.13 and an AMTE of 0.10).

Figure G.2: IV Simulation: second stage



NOTES: Figure shows the average returns to graduating from high school estimated from two-stage least squares on simulated data. Each circle represents the mean returns based on IV run on 400 samples of 2242 observations. The IV regression includes our standard control variables and the cognitive and non-cognitive factors. The red bars show the 95% confidence range. The X-axis shows the various counterfactual values of the local-college instrument used in the first stage.

H Sensitivity to Estimating Outcomes for GEDs Separately From High School Dropouts

The recent literature (Heckman, Humphries, and Kautz, 2014) shows that controlling for ability, GEDs earn the same as dropouts. In this section, we compare estimates that constrain GEDs and dropouts to be characterized by a common behavioral equation and those that estimate separate equations for each group. In a first model, dropouts can earn a GED and have their own outcome functions. In a second model, this option is removed and GEDs are assumed to be the same as dropouts. The different models have different economic interpretations. In the first model, dropouts who earn a GED have (slightly) higher prices of their skills in the labor-market. The second model imposes the requirement that dropouts face the same skill-prices as dropouts regardless of whether or not they have earned a GED. Overall, these two models produce similar treatment effects, with the point-estimates for the model without the GED option being slightly higher than the point-estimates of the treatment effects for the model that includes the GED option.

In an intermediate step, we decompose the total change into the change resulting from constraining to equality the earnings functions of GEDs and Dropouts, and the effect of eliminating the GED option. We decompose the total change in the estimated treatment effects into changes due to constraining the earnings function when GEDs and Dropouts are pooled (but the GED option remains as a second choice), and the total effect from closing down the GED option.

Table H.1 shows the estimated dynamic treatment effect for graduating from high school in the model of this paper (first row) and in a model where the GED choice is eliminated and GEDs and dropouts are pooled in estimation (third row). The second row shows an intermediate step which constrains the earnings functions of dropouts to be those of GEDs, but where the GED choice option is still available. In this intermediate step,restrictions to equality the earnings functions of dropouts and GEDs (but still allowing individuals to earn a GED) leads to slightly smaller treatment effects. This happens because dropouts are assigned the (slightly) higher wages of GEDs. Closing down the GED option entirely results in slightly larger treatment effects as an option is removed from the dropout choice set, and this effect dominates the price effect just discussed.

The table shows four different treatment effects: “ATE” is the average treatment effect as defined in equation (10) of Section 3. “TT” is the average treatment on the treated, or the gains for those who choose to graduate high school. “TUT” is the average effect of treatment on the untreated, or the gains for those who chose not to graduate from high school. “AMTE” is the average marginal treatment effect as defined in equation (14), or the average gains for those nearly indifferent to graduating from high school or not. All treatment effects include continuation values.

Table H.1: Dynamic Treatment Effects for High School Graduation (log wages age 30): including (row 1) and excluding (row 3) to choice to get a GED

	ATE	TT	TUT	AMTE
High School Graduation				
with GED:	0.09	0.09	0.10	0.10
with GED (constrained):	0.08	0.08	0.09	0.09
without GED:	0.12	0.12	0.11	0.11

Notes: This table compares the dynamic treatment effects of graduating from high school for a model that includes the choice to earn a GED, and a model where dropping out of high school is a terminal state. The first row shows treatment effect for the model that includes the GED. The second row constrains the wage equation for dropouts to be the same as the wage equation from the model that pools dropouts and GEDs (row 3) but includes the option to earn a GED as in row 1. The third row removes the GED option and estimates dropout outcomes pooling dropouts and GEDs. The dynamic treatment effects shown here are inclusive of continuation value as defined in Section 3 of the paper. "ATE" is the average treatment effect as defined in Section 3. "TT" is the average treatment on the treated, or the gains for those who choose to graduate high school. "TUT" is the average treatment on the untreated, or the gains for those who chose not to graduate from high school. "AMTE" is the average marginal treatment effect as defined in Section 3.

References

Almlund, M., A. Duckworth, J. J. Heckman, and T. Kautz (2011): "Personality Psychology and Economics," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Wößmann, vol. 4, pp. 1–181. Elsevier, Amsterdam.

Anderson, T. W., and H. Rubin (1956): "Statistical Inference in Factor Analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, vol. 5, pp. 111–150. University of California Press, Berkeley, CA.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2011): "Identification Problems in Personality Psychology," *Personality and Individual Differences*, 51(3: Special Issue on Personality and Economics), 315–320.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101(6), 2754–2781.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6B. Elsevier, Amsterdam.

Conard, M. A. (2006): "Aptitude is not enough: How personality and behavior predict academic performance," *Journal of Research in Personality*, 40(3), 339–346.

Credé, M., and N. R. Kuncel (2008): "Study Habits, Skills, and Attitudes: The Third Pillar Supporting Collegiate Academic Performance," *Prospectives on Psychological Science*, 3(6), 425–453.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–931.

Duckworth, A. L., P. D. Quinn, and E. Tsukayama (2012): "What No Child Left Behind Leaves Behind: The Roles of IQ and Self-Control in Predicting Standardized Achievement Test Scores and Report Card Grades," *Journal of Educational Psychology*, 104(2), 439–451.

Duckworth, A. L., and M. E. P. Seligman (2005): "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents," *Psychological Science*, 16(12), 939–944.

Farsides, T., and R. Woodfield (2003): "Individual differences and undergraduate academic success: The roles of personality, intelligence, and application," *Personality and Individual Differences*, 34(7), 1225–1243.

Gullone, E., and S. Moore (2000): "Adolescent risk-taking and the five-factor model of personality," *Journal of Adolescence*, 23(4), 393–407.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004): "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121(1–2), 39–98.

Heckman, J. J. (2001): “Econometrics and Empirical Economics,” *Journal of Econometrics*, 100(1), 3–6.

Heckman, J. J., J. E. Humphries, and T. Kautz (eds.) (2014): *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. University of Chicago Press, Chicago.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016): “The Causal Effects of Education on Earnings and Health,” Under revision, *Journal of Political Economy*.

Heckman, J. J., J. Stixrud, and S. Urzúa (2006): “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” *Journal of Labor Economics*, 24(3), 411–482.

Herrnstein, R. J., and C. A. Murray (1994): *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press, New York.

John, O. P., R. W. Robins, and L. A. Pervin (eds.) (2008): *Handbook of Personality: Theory and Research*. Guilford Press, New York, NY, third edn.

Judd, K. L. (1998): *Numerical Methods in Economics*. MIT Press, Cambridge, MA.

Matzkin, R. L. (1992): “Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models,” *Econometrica*, 60(2), 239–270.

Noftle, E. E., and R. W. Robins (2007): “Personality Predictors of Academic Outcomes: Big Five Correlates of GPA and SAT Scores,” *Personality Processes and Individual Differences*, 93(1), 116–139.

Williams, B. (2011): “Notes on Identification of Factor Models,” Unpublished manuscript, University of Chicago, Department of Economics.