A Education in South Africa

The South African education system includes primary (grades 1-7), secondary (grades 8-12) and post-secondary (grades 13 and above) levels. Schooling is compulsory for youth aged 7-15 or through completion of grade 9 (whichever comes first), but I treat this regulation as non-binding due to the presence of non-enrolled youth in this age range in the data. To enter post-secondary schooling, students must pass a nationally standardized "matric" exam at the completion of grade 12. Post-secondary education includes both academic and vocational programs (universities of technology, formerly Technikons). Additionally, students at the secondary level may enroll in technical colleges or vocational National Training Certificate (NTC) programs.[1] For simplicity, I do not distinguish between academic and vocational education in the model. Private schools serve less than 3% of the student population in South Africa, according to government statistics; in the Western Cape province the figure is 3.1% (Fiske and Ladd 2004).

The government subsidizes public education, but students must pay fees to attend and such fees vary considerably among schools. Public schools are self-governing and are free to set their own admissions policies and fees (subject to provincial government approval). Although admissions cannot discriminate based on race, test scores, or ability to pay fees, prevailing patterns of residential segregation serve to maintain quality differences among schools. Moreover, despite legal prohibitions, Fiske and Ladd (2004) conclude that "there is little doubt that many schools consider a family's likely ability to pay their fee when making admissions policy" (p. 143). Although low-income families may qualify for fee exemptions under a policy adopted in 1998, only 2.5% of primary school students and 3.7% of secondary school students receive the exemption (these figures rise to 4.1% and 5.7%, respectively, in historically white schools; Fiske and Ladd 2004). At the post-secondary level, the South African government offers subsidized loans to qualified students who pass a means test through the National Student Financial Aid Scheme (NSFAS), and individual institutions also offer financing. Private banks also offer student loans at market interest rates.

B Data Definitions

The data come from retrospective life history data collected in Wave 1 of CAPS, augmented with life events recorded in Waves 2-4.[2] The Wave 1 retrospective life histories record events by youth's

[1] See Appendix B for information on mapping NTC programs to grade levels.

[2] The Cape Area Panel Study Waves 1-2-3 were collected between 2002 and 2005 by the University of Cape Town and the University of Michigan, with funding provided by the US National Institute for Child Health and Human Development and the Andrew W. Mellon Foundation. Wave 4 was collected in 2006 by the University of Cape Town, University of Michigan and Princeton University. Major funding for Wave 4 was provided by the National Institute on Aging through a grant to Princeton University, in addition to funding provided by NICHD through the University

age, where age refers to the age at which the event occurred in the case of living arrangements and marriage, and to age at the beginning of the calendar year in the case of enrollment and progression through school, labor force participation, and pregnancy. I follow this convention in mapping Wave 2-4 responses to youth's age.

I make several sample restrictions. I keep only those observed until at least age 18. Those who report advancing two or more grades in a year, or without continuous information on enrollment, are dropped from the sample. I drop those who report entering school prior to age 4 or exiting school after age 24 (which effectively sets $T_d = 24$ as the decision horizon).[3] I also drop those whose educational histories, by the definitions below, place them with more than 16 years of completed schooling. The restriction on observing a person until at least age 18 accounts for more than two-thirds of those dropped from the sample.

Schooling level covers grades 1-16, with National Training Certificate (NTC) I, II and III mapped to grades 10, 11, and 12, respectively.[4] Students enrolled in university or university of technology programs that include grade 12 are considered enrolled in grade 12. Reporting successful completion of the grade level or reporting enrollment in a higher grade level in a subsequent year is considered passing the level for grades 1-12. Beginning at grade 13, reporting successful completion of the grade level or "no grade/continuing" are considered passing the level, up to a maximum of 16 years completed schooling. I make this distinction because "no grade/continuing" is the modal response for those enrolled in the post-secondary education sector, indicating that most youth are continuing in their programs of higher education, whereas "passing" reports at these levels drop considerably. Unfortunately, I am unable to determine whether students are making satisfactory progress towards degree completion. All other results while enrolled are considered failure. I define "dropout" as disenrollment following a year of enrollment, and "re-enrollment" as enrollment following a year of non-enrollment. Schooling histories in which levels regress with age are re-coded so that such regression cannot occur. Grades failed represent the accumulation of periods of enrollment in which the agent did not pass the grade, and therefore may include events such as withdrawal, illness or residential moves rather than outright academic failure. I top-code years enrolled, years not enrolled, grades failed, and work experience since age 12 at 8, 5, 3, and 4, respectively, which reduces the dimension of the state space while still accurately capturing more than 97% of the person-year observations in the sample for each variable.

of Michigan. CAPS is publicly available at `http://www.caps.uct.ac.za/`.

[3]Reporting school entry prior to age 4 is more likely to reflect measurement error than childhood precociousness, in my view. Few observations are available in the data after age 24 (less than 20% of the sample is observed beyond 24, due to starting ages in Wave 1 and attrition), raising concern that estimating the model with these enrollment choices will result in severe finite-sample bias.

[4]NTC conversion based on coding in CAPS, derived variable $w1h\_higrd$.

Labor force participation variables (i.e., work and search) and wages are conditional on non-enrollment at a given age, where reports of enrollment supersede reports of labor market participation. School fees are conditional on enrollment, and include total household expenditure on fees and other educational expenses in real rand per year (base year 2002). Wages are full-time annual equivalent based on 160 working hours per month (those reporting monthly hours above 160 are considered full-time and do not receive an adjustment). Wages and school fees are available only at the time of the interview, rather than as retrospective histories. To solve the model and estimate the enrollment utility parameters, I use predicted values of wages and school fees from the model, replacing predicted values below the minimum value observed in the sample with the first percentile from the data, in order to avoid non-positive predicted values and extreme outliers. Work experience includes only those periods of simultaneous work and non-enrollment; I exclude work experience while enrolled in school.[5]

Other covariates are largely self-explanatory. Ability quartiles refers to in-sample rank of age-adjusted score on the literacy and numeracy evaluation (LNE) administered to all CAPS respondents in Wave 1. Unfortunately, because this ability measure was taken in Wave 1, when the sampled youth were at least age 14, it is not predetermined with respect to enrollment choices in the model, which begins at age 12. I include it, however, because it is a measure common to all in the sample, and therefore helps to distinguish between the role of ability and human capital investment in labor market returns. To mitigate bias in the LNE score due to age differences in Wave 1, I adjust for age as follows: using the estimation sample, I regress the standardized literacy and numeracy evaluation (LNE) score on age and age squared at Wave 1 (when the test was administered) and get predicted residuals. I then sort observations into quartiles based on these residuals. Household income quintiles are derived from the distribution of household per capita income reported in Wave 1 of CAPS.[6] The pension-eligible grandparent indicator is set to one if the young adult reports living with a grandparent at the given age and there is at least one grandparent on the household roster who is of pension-eligible age (60 for females, 65 for males). The variable definition also makes note of changes in the household roster reported in Waves 2-4.

I calibrate the age-wage profile for years following the decision horizon (i.e., from periods $T_d + 1$ to T) using the 10% public use micro-sample of the 2001 South African Census. First, I define the estimation sample as native-born residents of urban areas in Western Cape province (which includes metropolitan Cape Town, from which CAPS respondents are drawn) who are ages 25-64, in the labor force, and classify themselves as one of the three major racial groups (white, black, or coloured).

[5] Work or search while enrolled peaks at 3.2% across all grade levels, and never exceeds 2% during grades 1-12.

[6] Due to non-response, 7% of the sample uses imputed values for household income, based on multiple imputation conducted by CAPS.

I also exclude self-employed and unpaid workers, leaving a sample of $n = 111,772$. I then predict employment and income by running logit and OLS regressions, respectively, using as controls race and gender dummies, years of schooling, a high school graduate dummy, race-schooling and race-high school graduate interactions, age and age squared.[7] I then create an expected income variable for each observation as the product of these predicted values (i.e., $\mathbb{E}(\widehat{\text{income}}) = \widehat{\Pr(\text{work})} \times \widehat{\text{income}}$), and regress expected income (in units of R10,000) on the same set of controls. I save the coefficients on age and age squared from this regression for use in the wage equation (2) of the structural model; the coefficients on age and age squared are .37 and -.005, respectively.[8] The macro environment variable is based on the South African employment/population ratio for 15-24 year olds, from the World Bank Africa Development Indicators.

All variables measured in monetary values used in this paper are in real South African rand per year (base year 2002), unless otherwise noted. The South African rand traded at 10.3 per US dollar in August, 2002 when CAPS Wave 1 began.

C Model estimation and the EM algorithm

This appendix describes additional details of the likelihood function, and how I estimate the model using conditional choice probability (CCP) estimation techniques developed by Hotz and Miller (1993) and Arcidiacono and Miller (2011).

C.1 Likelihood function: additional details

To estimate equation (5) by maximum likelihood, I use panel data to formulate the log likelihood function:

$$\begin{aligned}\ln L(\alpha_{e,m}) = \sum_{i=1}^{N}\sum_{t=1}^{T_d}\Bigg\{ & p_{i,t+1}\ln\left(\frac{\alpha_{e,m}}{\alpha_{e,m}+\tau_{e,i}}\pi_{0t,m}(\tilde{X}_{it}) + \frac{\tau_{e,i}}{\alpha_{e,m}+\tau_{e,i}}\pi_{it}\right) \\ & + (1-p_{i,t+1})\ln\left(1 - \left[\frac{\alpha_{e,m}}{\alpha_{e,m}+\tau_{e,i}}\pi_{0t,m}(\tilde{X}_{it}) + \frac{\tau_{e,i}}{\alpha_{e,m}+\tau_{e,i}}\pi_{it}\right]\right)\Bigg\} \qquad \text{(SA1)}\end{aligned}$$

[7] The included controls are the largest subset (other than age variables) of the controls used in the structural model that are available in the Census. Note that interactions of schooling and the high school graduate dummy are not included because the maximum years of schooling reported in the Census is 13. I use income rather than wages because the latter are unavailable in the Census. I adjust income to 2002 South African rand to be consistent with the base year of CAPS.

[8] Since I care only about the coefficients, inconsistent estimation of their standard errors due to the generated outcome variable is not problematic in this context.

I specify the prior mean $\pi_{0t,m}(\tilde{X}_{it})$ as a logistic transformation of a linear index with a type-specific intercept, as in equation (6), where the observable (to both agent and econometrician) state space $\tilde{X}_{it}$ is as specified in Table A1. The logistic functional form is chosen to ensure that the prior mean is constrained to the unit interval. I specify the prior precision as $\alpha_{e,m} = \exp(\upsilon_{e,m})$, where $\upsilon_{e,m}$ is a parameter to be estimated, in order to ensure that the precision is strictly positive. The likelihood function corresponding to equation (8) takes an analogous form.

As is standard in dynamic discrete choice models following Rust (1987), the Type I Extreme Value assumption on the distribution of unobserved shocks allows for the following simplification the Bellman equation (9):

$$\begin{aligned}\mathbb{E}\left[\max_d\{V^d_{t+1}(S_{t+1},m|S_t,d_t,m)\}\right] &= \mathbb{E}\left[\ln\left(\sum_d \exp\left[V^d_{t+1}(X_{t+1},m|X_t,d_t,m)\right]\right)\right]+\kappa \\ &= \sum_X\left(\left[\ln\left(\sum_d \exp\left[V^d_{t+1}(X_{t+1},m|X_t,d_t,m)\right]\right)\right]+\kappa\right)\Pr(X_{t+1}|X_t,d_t,m)\end{aligned} \tag{SA2}$$

where $\kappa \cong .5772$ is Euler's constant. The first equality above results from the agent's expectations over the unobserved state variables ϵ, while the second incorporates the remaining expectation on transitions of the discretized observable state variables X. Substituting the above expression into (9) serves to calculate the agent's expected future value conditional on the current period's choice.

C.2 Conditional choice probability estimation of dynamic programming problem

To explain the estimation method I use, consider again the future value term (i.e., the $\mathbb{E}$[max] term) in (SA2) and manipulate the expression inside the natural log (for simplicity, I omit the constant):

$$\mathbb{E}\left[\ln\left(\sum_d \exp\left[V^d_{t+1}(X_{t+1},m|X_t,d_t,m)\right]\right)\right] = \tag{SA3}$$

$$\mathbb{E}\left[\ln\left(\frac{\exp\left[(V^j_{t+1}(X_{t+1},m)\right]}{\exp\left[(V^j_{t+1}(X_{t+1},m)\right]}\sum_d \exp\left[V^d_{t+1}(X_{t+1},m)\right]\right)|X_t,d_t,m\right] = \tag{SA4}$$

$$\mathbb{E}\left[(V^j_{t+1}(X_{t+1},m) - \ln\left(\Pr\left[d_{t+1}=0|X_{t+1},m\right]\right)\right] \tag{SA5}$$

The final line in (SA5) shows how, for any initial choice, the future value term in the Bellman equation can be expressed in terms of the value function of an arbitrary choice (labor market participation in this case) and its conditional choice probability. Arcidiacono and Miller (2011) observe that an important consequence of this property is that one may express the expected value of future utilities in terms of any sequence of future choices, regardless of whether such choices are optimal. Defining such a sequence strategically can then greatly reduce the computational burden of estimation in particular applications.

In my enrollment choice model, consider two choice sequences from an arbitrary starting time t through T_d: $\{1, 0, 0, \ldots, 0\}$ and $\{0, 0, 0, \ldots, 0\}$.[9] In other words, in one sequence an agent enrolls in period t and enters the labor market in all subsequent periods, while in the second sequence the agent never enrolls. Using (SA5) to construct the value functions for each sequence yields (again omitting the constant for simplicity):

$$\begin{aligned} v_1(S_t, m) &= U_e(S_t, m) + \sum_{\tau=t+1}^{T_d} \beta^{\tau-t} \left(\sum_X \{\mathbb{E}[w(X_\tau, m)] - \ln\left(\Pr\left[d_\tau = 0 | X_\tau, m\right]\right)\} \Pr(X_\tau | X_t, d_t = 1, m) \right) \\ v_2(S_t, m) &= \mathbb{E}[w(S_t, m)] + \sum_{\tau=t+1}^{T_d} \beta^{\tau-t} \left(\sum_X \{\mathbb{E}[w(X_\tau, m)] - \ln\left(\Pr\left[d_\tau = 0 | X_\tau, m\right]\right)\} \Pr(X_\tau | X_t, d_t = 0, m) \right) \end{aligned} \tag{SA6}$$

Although the value functions above are lengthy, they depend on a small number of objects: the flow enrollment utility, expected wages, conditional choice probabilities, and state transitions. All of these components except the enrollment utility function $U_e(S_t, m)$ may be calculated without having to solve the dynamic program. Because my assumptions on the choice-specific error terms lead to a dynamic binary logit, the flow enrollment utility parameters may be identified off of differences between these choice sequence-specific value functions.[10] Moreover, because the sequences are defined such that enrollment utility never appears in the future value terms, I need to solve the dynamic program just once rather than at each guess of the parameters. I can therefore work with the differenced value functions to form the likelihood:

$$\Pr(d_t = 1 | X_t, m) = \frac{1}{1 + \exp\left(v_2(X_t, m) - v_1(X_t, m)\right)} \tag{SA7}$$

[9] Arcidiacono and Miller (2010) (the working paper version) discuss this case.

[10] Note that this setup is not identical to an optimal stopping problem. Although agents stop enrollment from $t+1$ through T_d in the sequences I have chosen, re-enrollment is still permitted. Moreover, the value functions do not assume the behavior in each sequence is optimal; the presence of the conditional choice probabilities in the future value terms corrects for suboptimal choices (Arcidiacono and Ellickson 2011).

By calculating the future value terms in (SA7) outside the solution to the dynamic program, estimating this component of the likelihood is equivalent to estimating a static logit with differenced future values as an offset term. This reformulation greatly simplifies estimation and allows me to rely on standard estimation routines that use analytic derivatives rather than strictly numerical techniques. Such simplification is crucial to enjoy the advantages of the EM algorithm in recovering the type-specific parameters and latent type distribution, as described below.

C.3 Incorporating unobserved heterogeneity using the EM algorithm

Although estimation using the CCPs as described above can simplify calculation of the likelihood function for a given type, the inability of the researcher to observe an agent's type remains a challenge in estimation. To overcome this challenge, I use the expectation maximization (EM) algorithm to estimate the structural parameters and the distribution of unobserved types in the population. As shown by Arcidiacono and Jones (2003), the EM algorithm allows sequential estimation of separate components of the likelihood function in the presence of unobserved heterogeneity, eliminating the need to use full information maximum likelihood. The algorithm works by using Bayes' Rule to calculate an individual's probability of being a given type (the expectation step), then treating type as known when maximizing the type-specific likelihood (the maximization step) and using the individual type probabilities as weights in the overall likelihood function. The algorithm iterates between the expectation and maximization steps until convergence, with the population proportions of each type estimated using averages of the individual-specific type probabilities. I follow the approach of Arcidiacono and Miller (2011), who showed how the technique can be applied to dynamic discrete choice models using CCP estimation.

To illustrate the EM algorithm, suppose one has guesses of the type proportions $\tilde{\omega}$ and structural parameters $\tilde{\theta}$ for each type. Denote by $q_{im|s_0}$ individual i's probability of being of latent type m, conditional on initial schooling s_0. Using Bayes' Rule:

$$q_{im|s_0} = \frac{\tilde{\omega}_{m|s_0} l(\tilde{\theta}_m|\mathcal{X}_i)}{\sum_{m=1}^{M} \tilde{\omega}_{m|s_0} l(\tilde{\theta}_m|\mathcal{X}_i)} \tag{SA8}$$

where $l(\tilde{\theta}_m)$ is as in equation (15) and the conditioning is on the full set of covariates and outcomes $\mathcal{X}$ of agent i. Arcidiacono and Jones (2003) show that the overall likelihood can then be calculated for a sample of size N as:

$$L(\theta) = \sum_{i=1}^{N} \sum_{m=1}^{M} q_{im|s_0} \left(\sum_{k \in \{e,j,w,d\}} \ln l_k(\theta_m|\mathcal{X}_i) \times \mathbb{I}(\text{type} = m) \right) \tag{SA9}$$

In other words, using the individual-specific type probability q to weight the likelihood function instead of the population proportion ω restores additive separability to the components of the likelihood function. Working with (SA9) is equivalent to equation (16), but computationally faster and easier to maximize. Sample averages of $q_{im|s_0}$ across individuals are used to calculate the population proportions of each type.

Note that the dynamic programming problem is embedded in the likelihood component l_d, which uses the CCP estimation technique described above. Maximizing l_d relies on preliminary estimates of the structural parameters from the other likelihood components (the transition equations l_e and l_j, and the wage equation l_w) as inputs to estimate the remaining structural parameters, the enrollment utility parameters ϕ_b. I use the two-stage estimation routine described in Arcidiacono and Miller (2011) to estimate ϕ_b, which involves replacing l_d with a flexible logit for enrollment choices in (SA9) to estimate the CCPs. The converged parameter estimates and CCPs from this first stage become inputs to estimate ϕ_b in the second stage. Because estimation proceeds in stages, I bootstrap the entire procedure to account for the influence of sampling variability in first-stage parameter estimates on subsequent results when calculating standard errors.

To recapitulate, the first stage of the EM algorithm proceeds as follows:

1. *Expectation step.*
   (a) Use the previous iteration's structural parameter estimates $\tilde{\theta}$ and type probabilities $\tilde{\omega}$ to calculate the individual-specific type probabilities q according to (SA8).[11]
   (b) Update the population type probabilities by calculating $\tilde{\omega}_{m|s_0} = \frac{1}{N}\sum_{i=1}^{N} q_{im|s_0}$.
2. *Maximization step.*
   (a) Use the values of q calculated from step 1 to maximize (SA9), replacing the enrollment choice component l_d with a flexible logit.
   (b) Update $\tilde{\theta}$ with the maximized parameter values and return to the expectation step. Iterate until convergence.[12]

In the second stage, I use the first stage results to estimate the enrollment utility parameters ϕ_b by maximizing the likelihood component l_d as defined in (SA7), which is based on the dynamic programming problem of this paper. I set the discount factor β to 0.95.

[11] For the first iteration, I initialize $\tilde{\theta}$ with arbitrary values and use a uniform prior for $\tilde{\omega}$, i.e, $\tilde{\omega}_{m|s_0} = 1/M$ for all m. The routine converges to the same parameters regardless of starting values.

[12] Convergence criteria for the EM algorithm remain an unsettled issue. I follow Train (2008) and Arcidiacono and Miller (2011) in declaring convergence when the difference in likelihood values between iterations becomes sufficiently small.

D Model Identification

My argument for identification of the model parameters relies on Magnac and Thesmar (2002, hereafter MT). MT, building on Rust (1994), show that dynamic discrete choice models are non-parametrically not identified, but that identification follows from a parametric assumption on the distribution of unobserved (to the econometrician) shocks and calibration of the discount factor.[13] The underlying assumptions of the structural model they consider are:

1. Flow utility functions are additively separable in observed (to the econometrician) states and unobserved shocks known only to the agent: $U_d(S_t) = u_d(X_t) + \epsilon_t^d$, where d denotes the agent's choice, to adopt the notation of the paper.

2. Agents' subjective expectations match the law of motion of the state space, i.e., $\mu(S_{t+1}|S_t, d_t) = \Pr(S_{t+1}|S_t, d_t)$, where $\mu(\cdot)$ represents agent beliefs.

3. Unobserved (to the econometrician) shocks are independent conditional on the observed state: $\Pr(S_{t+1}|S_t, d_t) = \Pr(S_{t+1}|X_t, d_t)$.

The model of this paper satisfies each of these assumptions. As MT note, these assumptions can also be consistent with learning: when the econometrician observes all state variables that persist over time, s/he can therefore model the learning process by using data on observed state transitions, as in Crawford and Shum (2005) and Stange (2012).[14]

MT show that parameterizing the cdf of unobserved shocks ϵ and the discount factor allows identification of the structural parameters. I assume that ϵ is distributed Type I extreme value and that the discount factor $\beta = 0.95$. MT also note that differences in choice-specific utility levels are identified by assuming knowledge of the utility level of a reference choice. I meet this criterion by specifying expected wages as the utility associated with non-enrollment, allowing me to identify option values.

Finally, MT show that the model remains identified under time-invariant preference heterogeneity that is known to the agent but unobserved to the econometrician. Arcidiacono and Miller

[13] Heckman and Navarro (2007, hereafter HN) provide identification results that relax assumptions made by MT. It would be useful, therefore, to appeal to HN's results to establish identification of my model and empirically implement it under the most general conditions. Unfortunately, my model departs from theirs in key features that prevent me from assuring its semiparametric identification. Most notably, theirs is an optimal stopping time model, in which school dropout is an absorbing state. In contrast, the option to re-enroll plays a central role in my model. As they write (HN, p. 380), "Our identification strategy in this paper relies on the non-recurrent nature of treatment. We leave the task of formulating and identifying a general recurrent state version of the model for another occasion." To my knowledge, identification of such a model under conditions as general as they consider has not been established.

[14] The synthetic cohort approach of Manski (1993), which he calls "social learning," can also be invoked here, though it is different from the rational expectations-type assumptions of the other cited papers.

(2011) showed how this unobserved heterogeneity could be extended to transition probabilities and recovered using conditional choice probability estimation as in Hotz and Miller (1993). I follow this approach to unobserved heterogeneity in the paper.

D.1 Identification of unobserved heterogeneity in prior precision

As noted in the paper, my model departs from previous dynamic discrete choice models employing Bayesian learning, such as Crawford and Shum (2005) and Stange (2012), by incorporating unobserved heterogeneity in prior precision. Identification of this novel feature therefore deserves further remarks. The prior precision in my model plays a role similar to a regression coefficient on past outcomes, helping to define the weight placed on these outcomes relative to the prior mean. When the prior precision is specific to time-invariant type, identification is analogous to including an interaction term between past outcomes and type.

Suppose for a moment that type were observable to the econometrician. If, conditional on the prior mean, past outcomes are better predictors of current outcomes for one type, then this type will place more weight on past outcomes, and the maximum likelihood estimator will deliver a lower estimated prior precision. When type is unobserved, the same principle applies, with improvements in model fit leading the estimation routine to convergence.

A simpler model with similar form may help illustrate these points. Consider a dynamic model in which agents choose whether to supply their labor (a binary choice only) in each period t, with D as an indicator for labor supply. Flow utility of labor supply is:

$$U(D_t = 1) = \gamma_{0,m} + \gamma_1 S_{t-1} + \gamma_2 D_{t-1} + \epsilon_1 \tag{SA10}$$

where $\gamma_{0,m}$ is the intercept for (time-invariant) type m and S is the (binary) state of the economy. Flow utility of non-participation is ϵ_0, where both ϵ_1 and ϵ_0 are distributed Type I Extreme Value.

Agents learn about the state of the economy over time in the same Bayesian process as the agents in my school-to-work model, i.e., their posterior prediction is:

$$\Pr(S_t = 1) = \omega_{t-1}\pi_0 + (1 - \omega_{t-1})\pi_{t-1} \tag{SA11}$$

where:

$$\pi_0 = \Lambda\left(\beta_{0,m} + \beta_1 S_{t-1}\right)$$
$$\pi_{t-1} = \frac{\sum_{\tau=0}^{t-1} S_\tau}{t-1}$$
$$\omega_{t-1} = \frac{\alpha_m}{\alpha_m + (t-1)}$$

with notation analogous to that above. That is, π_0 and π_{t-1} are the prior mean and observed outcomes of state variable S, with ω the weight on the prior mean; α_m is the prior precision for type m; and the logit cdf $\Lambda(\cdot)$ defines the functional form for the prior mean.

Agents also observe a wage during each period of labor force participation:

$$w_t = \delta_{0,m} + \delta_1 S_{t-1} + \delta_2 D_{t-1} + \epsilon_w \tag{SA12}$$

where ϵ_w is distributed $N(0, \sigma_w^2)$. Agents are of two types, with the proportion of type 1 agents equal to p.

Agents maximize discounted expected utility over a finite number of discrete periods, with the discount factor set to 0.95. When making decisions, agents consider flow utility plus the effect of their current choice on the stream of future utility, given the dependence of utility on their expectation about the state of the economy and their past decisions. The parameters to estimate are those of the utility function, the transition equation for the state of the economy, and the wage equation, plus the proportion of type 1 agents: $\theta = (\gamma_{0,m}, \gamma_1, \gamma_2, \beta_{0,m}, \beta_1, \alpha_m, \delta_{0,m}, \delta_1, \delta_2, p)$.[15]

Although simpler in form, the model retains the key features of the school to work model of the paper: Bayesian learning about a transitory binary state variable, with type-specific prior mean and prior precision; a type-specific wage; and type-specific utility. The model also allows for persistence in choices without relying on risk aversion, through the inclusion of the lagged labor supply indicator in the flow utility function.

I simulate data according to the model parameters, with each simulation having identical number of agents (3,374) and average decision periods per agent (10) as the panel used in the paper. I estimate the model in two ways, treating type as observed or unobserved to the econometrician (agents know their type throughout).[16] The estimation procedure is identical to that used in the paper. When estimating the model with type unobserved, I follow Arcidiacono and Miller (2011) and estimate the model using the EM algorithm.[17]

[15] I set the wage variance $\sigma_w^2 = 0.2$, but do not estimate this parameter.
[16] When type is known to the econometrician, it is unnecessary to estimate the type proportions.
[17] This version of the model can also make use of the one-period dependence property of the dynamic program,

Table SA1 presents results from estimating 100 bootstrap replications of the model. Column (1) lists the true parameters, while columns (2)-(3) shows parameter estimates when type is known and unknown to the econometrician, respectively. All estimates in column (2) match the true parameters. Estimates of the state transition parameters demonstrate that the Bayesian learning model is well formulated; parameters governing the prior mean and precision can be recovered using observed variation in the transitions of a binary state variable.

All estimates in column (3), in which agent type is unobserved to the econometrician, also match the true parameters. Note that I obtain consistent estimates of the type-specific components of both the prior mean and prior precision in the transition equation, with standard errors nearly identical to the case where I can distinguish agent types. I also correctly estimate the type proportions.

This model is admittedly simpler than the full model of the paper, and has only two agent types. Nonetheless, I view these results as an important "proof of concept" for the full model. The results show that type-specific prior precision can be consistently estimated by the procedure I use, even when type is unknown to the researcher and even in the absence of risk aversion, both of which were major concerns expressed by reviewers.

E Interpretation of structural parameter estimates

This appendix describes estimates of the structural parameters reported in Table A3.

E.1 Prior precision

I begin discussion with the prior precision parameters for academic ability and employability (α_e and α_j, respectively), which highlight the model estimates of agent learning. Recall that a useful interpretation of the prior precision is the equivalent number of outcomes an agent "observes" when forming beliefs, and that the precision may vary by time-invariant type. For instance, Type 1 agents form a prior belief as though they observed 4.9 grade outcomes at the outset of the model. This is sensible, given that the model begins at age 12, when students who began school on-time at age 6 will have accumulated 6 grade outcomes. By contrast, Type 2 agents have a prior precision for academic ability of 32.9, an amount greater than the number of school outcomes they would see in an academic career, reflecting high levels of certainty about academic ability. Standard errors for the estimates of α_e are large because of a small number of large outliers in some bootstrap replications. I interpret these results as a failure to reject that both types have a prior precision for passing sufficiently high to rule out gradual learning, although I will nonetheless treat the point

leading to additional computational gains.

estimates as valid in what follows.[18]

In contrast, the prior precision estimates for employability in column (2) are uniformly lower than their counterparts for academic ability, reflecting much less certainty. Here we observe less variation between types, with the equivalent of 3.5 job search outcomes for Type 1 and 4.3 for Type 2, with both parameters precisely estimated.

E.2 Prior means, wages, and enrollment utility estimates

The prior mean parameters for academic ability and employability (π_{0t} and ξ_{0t}, respectively) in columns (1) and (2) are mostly as expected. In the academic ability equation of column (1), the constant is a homogeneous intercept, and the Type 2-specific intercept represents a deviation from it (the Type 1 intercept is therefore equivalent to the constant). The Type 2 intercept demonstrates variation in average prior beliefs about academic ability among otherwise identical agents, and reflects permanent unobserved characteristics that influence passing (such as motivation or non-cognitive skills). Coefficients on observables capture how the prior mean for passing varies with these characteristics. Because these coefficients are estimated from passing outcomes in the data, they also correspond to variation in observed passing rates. I find that conditional on all other included covariates, blacks and females are significantly more likely to pass to the next grade level. The probability of passing is increasing in ability and for those who completed at least 6 years of schooling by age 12 (again, conditional on other included variables; unconditional probabilities may differ). The probability of passing declines with the level of schooling, though the separate slope and intercept terms for high school graduates are positive, reflecting the lower failure rates in post-secondary education reported in Table 2. Those living with a pension-eligible grandparent are less likely to pass, possibly reflecting a lower incentive for school performance as a result of intra-household transfers.

Turning to estimates of the prior mean for employability ξ_{0t} in column (2), I again observe variation in beliefs by unobserved type. Most coefficients on observable characteristics have the expected sign. Despite their higher rates of grade advancement, blacks and females are significantly less likely to find work than coloureds and males. In addition, the probability of employment is increasing in schooling, with a higher intercept for high school graduates. Surprisingly, a number of measures of cognitive skill (ability quartiles 2-3, completion of at least 6 years of schooling at age 12, and post-secondary schooling) are associated with lower employment rates, possibly reflecting difficulties that youth have in signaling their ability credibly in a crowded labor market. Employment is quite sensitive to macro conditions, with a precisely estimated negative coefficient

[18] Removing bootstrap estimates of α_e above the 95th percentile reduces standard errors for Types 1 and 2 to 4.1 and 13.0, respectively, and leads to a rejection of equality between their prior precision at the 5% level.

on the bad macro environment dummy.

E.3 Enrollment utility, wage and school fee estimates

Table A3, column (3) presents parameter estimates from the enrollment utility function of the dynamic discrete choice model, i.e., the coefficients from equation (3).[19] The Type 2 intercept here demonstrates the presence of persistent variation in preferences for enrollment between types. Type 2 agents have enrollment utility R11,199 higher than Type 1, or about US$1,087.[20] The coefficient for black shows that blacks are (conditionally) more likely to enroll in school than coloureds, consistent with the shorter school careers of coloured youth. The large negative coefficient on high school graduation reflects the sharp increase in school dropout at this stage. The coefficients on recent and cumulative failure are negative and precisely estimated, representing strong evidence that youth dynamically update their enrollment behavior based on past schooling outcomes. The positive coefficient on enrollment last period shows that agents prefer continuity in their schooling careers and demonstrates the presence of a psychic switching cost for re-enrollment. However, coefficients on the interaction terms between enrollment last period and schooling show that this effect declines through the primary and secondary grades, turning negative at high school completion, and then rising again. The negative coefficients on years enrolled and years not enrolled demonstrate the tendency of youth to leave school as they age, conditional on all other characteristics. The interplay between the path dependence of youth enrollment decisions and their sensitivity to grade outcomes suggests that such decisions are complex and subject to several competing influences.

Because the school fee term in the enrollment utility function (3) accounts for the direct costs of schooling, the negative sign on the indicator for high school graduate may reflect several impediments to higher education, such as greater psychic costs, difficulty in obtaining admission, or credit constraints. The coefficients on the household income terms do not provide strong evidence in favor of credit constraints, however. The dummy variables for being from the lowest two household income quintiles are negative, as would be expected if household disadvantage is associated with greater dropout. The argument for credit constraints weakens when considering the interaction terms on high school graduate and these dummy variables. Conditional on high school graduation, youth from the poorest quintile of households are *more* likely to enroll in post-secondary education, while those from the second-poorest quintile are less likely, though only the latter correlation is sta-

[19] A note on the interpretation of coefficients from the enrollment equation in Table **??**: as is well known, logit coefficients such as these are identified only up to scale. Because the units of the labor market utility function (as well as the school fee term in the enrollment utility function) are in South African rand per year (ten thousands, base year 2002), so too are the units of the enrollment utility coefficients, provided the scale parameter of the i.i.d. Type I extreme value shocks is unity, as assumed.

[20] Based on an exchange rate of 10.3 rand per US dollar in August 2002, when CAPS Wave 1 began.

tistically significant. One interpretation of these results is that high school graduates from poorer households have already overcome large obstacles to academic success, and the same qualities that explain this attainment allow them to find ways to continue their studies.

Coefficients for the wage equation (parameter ϕ_w) presented in column (4) generally show the expected pattern with respect to demographic, schooling and ability variables, with a large wage return to higher education evident in the post-secondary schooling slope coefficient. A bad macro environment dampens wages, as expected.[21]

The school fee regression results are ($N = 2,644$, $R^2 = .28$):

$$\underset{(.08)}{\widehat{fee} = -.28} \quad \underset{(.02)}{-\ .08 black} \quad \underset{(.01)}{+\ .05 schooling} \quad \underset{(.05)}{+\ .67 HSG} \tag{SA13}$$

where the unit of observation is the person-year, and school fees are measured in tens of thousands of South African rand per year. Fees are lower for blacks, likely reflecting lower school quality. School fees increase with grade level, with a large and discontinuous jump in post-secondary schooling. The equation is estimated outside the structural model, with predicted fees inserted within the model.

E.4 Unobserved type proportions

The bottom of Table A3 presents estimates of the proportion of each type in the sample, conditional on initial schooling, showing that those with low initial schooling are more likely to be Type 2 (71% to 44%), while the reverse is true for those with high initial schooling. Estimating the model without unobserved heterogeneity (i.e., setting $M = 1$) reveals qualitatively similar coefficients on observable characteristics as the preferred model. Importantly, the relative certainty about academic ability versus employability remains, with the prior precision of academic ability estimated as 12.4 (standard error 1.4) and for employability 3.5 (0.2). However, ignoring unobserved heterogeneity fails to adequately capture the considerable heterogeneity in learning about academic ability discussed in Section 4.2. Full results described in robustness checks.

To give a sense of how this unobserved heterogeneity translates into differences in choices and outcomes, Table SA4 shows simulation results using the estimated structural parameters for each

[21] The coefficients that reverse sign between the prior mean for employment and wage equations could reflect differences in those characteristics' returns to job search, on the one hand, and productivity conditional on employment, on the other. For instance, those from the highest ability quartile may be unable to signal this quality credibly (conditional on schooling) when searching for a job, but are able to demonstrate it on the job and are rewarded with a higher wage.

type. Within type, agents with high initial schooling have higher attainment, enrollment, and lower dropout (panels [a]-[c]) than those with low initial schooling, as would be expected. Re-enrollment rates are similar between those with high and low initial schooling, suggesting that re-enrollees are not just those who began their schooling careers at a disadvantage. Interesting differences also emerge between types. Enrollment and dropout rates between Type 1 agents with high initial schooling and Type 2 agents with low initial schooling are similar, showing that unobserved heterogeneity is an important determinant of behavior. Both types re-enroll at similar rates regardless of initial schooling, suggesting the pervasiveness of re-enrollment among South African youth.

F Agent learning: individual examples

To understand agent learning further, it is also instructive to consider specific examples from the sample. Figure SA1, panel (a) shows how a coloured female with 6 years of schooling at age 12 learns about her academic ability. The gray diamonds show her schooling result for periods in which she enrolled, with a value of 1 (read off the right vertical axis) indicating pass and 0 indicating fail. The solid line running through the figure, which assumes the agent is Type 1, represents the update she makes to her posterior probability according to equation (5) after learning this result. Positive values (read off the left vertical axis) reflect an upward revision to her posterior prediction of passing the next grade based on this latest result, whereas negative values represent a downward revision, and zero indicates no change.[22] This student enrolled in school and passed at ages 12 and 13. Accordingly, she increased her posterior predicted pass rate in these periods, reflecting upward revisions to her beliefs about academic ability. She enrolled again at age 14 but failed, and revised her posterior downward in response. The pattern continues this way in subsequent periods of enrollment through age 17. From ages 18 onward she does not enroll in school and therefore makes no revision to her posterior prediction for grade advancement.

The dashed line in the same graph plots how the same agent learns under the alternative assumption that she is Type 2. The heterogeneity in learning about academic ability between types is evident in the figure. Type 1 agents update their posteriors for passing more dramatically than Type 2 agents when confronted with the same information, reflecting the lower precision they place on the prior. Figure SA1, panel (b) presents a graph analogous to panel (a) for a different

[22] This update to the posterior is normalized relative to any change in the prior mean, because the prior mean can also change between periods if state variables other than the cumulative pass rate change. In other words, the line in Figure SA1(a) plots $[\Pr(p_{t+1} = 1|\pi_t, m) - \Pr(p_t = 1|\pi_{t-1}, m)] - [\pi_{0t,m}(X_t) - \pi_{0t,m}(X_{t-1})]$, where all variables are as defined in (**??**). Although agents apply this updated posterior to the next period in which they enroll, I plot it as concurrent with the most recent result for ease of interpretation.

agent's learning about employability while on the labor market. As with learning about academic ability, we see in panel (b) how the agent incorporates new information to update his posterior prediction for obtaining employment, with (slight) variation in learning between types.[23]

To see how learning about academic ability and employability may interact with each other and influence dropout and re-enrollment, consider Figure SA1, panel (c). The figure shows the schooling and labor market results and corresponding updates to posterior predictions for a black female with 6 years of schooling at age 12 (the posterior predictions have been averaged across types for exposition). This young woman encounters a string of mixed results as a student from ages 12-18, and alters her posterior prediction for passing the next grade accordingly. She decides to drop out at age 19 but does not find work, leading her to revise downward her posterior probability of finding work in the future. With continued job search now less appealing, she re-enrolls in school at 20. She fails this school year, however, and returns to the labor market, this time finding work at age 21 and revising upwards her expectation of remaining employed. She remains in the labor market the following year before exiting the sample. Together, the panels of Figure SA1 portray a process of active learning by youth in response to new information about academic ability and employability, with important implications for their enrollment choices.[24]

G Additional robustness checks

This section describes a series of robustness checks of the model. Results from Table SA5 are discussed in the paper. Table SA6 shows estimates of the structural parameters for a model without unobserved heterogeneity. Coefficients on observable characteristics are similar to the preferred model. Importantly, the relative certainty about academic ability versus employability remains, with the prior precision of academic ability estimated as 12.4 (standard error 1.4) and for employability 3.5 (0.2). However, ignoring unobserved heterogeneity fails to adequately capture the considerable heterogeneity in learning about academic ability.

The baseline model treats transitions between high school and post-secondary education and transitions between schooling and labor force participation in the same framework. Although enrollment preferences, pass rates, and the returns to schooling may change discontinuously at high school completion, one might worry that school to work transitions for those with post-

[23]Posterior updates for employment do not exist for age 23 in panel (b) despite a labor market result because the person leaves the sample after this age, and therefore has no further results with which to apply his updated posterior prediction.

[24]Recall that agents will also dynamically update their wage expectations and enrollment utility in response to school and labor market outcomes. These adjustments are not shown in Figure SA1, panels (a)-(b) but will also be important influences on behavior.

secondary schooling are fundamentally different than for those without. To explore this possibility, I re-estimate the model using only observations from youth who never enrolled in post-secondary schooling.

Table SA7 shows results from the sample with no post-secondary schooling. Results are qualitatively similar to those from the full sample. As in the baseline model, types incorporate new information about academic ability at different rates. This difference is even more pronounced than in the baseline model, with Type 2 agents effectively not considering new information from grade outcomes at all. Learning about employability is similar between types and occurs at a similar rate as the baseline model, however. As in the full sample, youth who never enroll in post-secondary schooling are more likely to drop out of school as a function of recent grade failure. This result suggests that the process of dynamic updating of the expected relative returns to enrollment versus labor force participation is similar for these youth. The coefficient on the indicator for being enrolled last period in the enrollment utility function, representing the switching cost of re-enrollment, is also similar to the baseline model.

Tables SA8 and SA9 present results of the model when including whites in the sample and when removing exclusion restrictions across equations, respectively. Again, results are qualitatively similar to the baseline model. One potential form of model misspecification is if more recent work experience affects state transitions and utility differently from previous work experience. I therefore re-estimate the model by including an indicator for work experience in the most recent period of non-enrollment separately from previous work experience.[25] Doing so allows agents to experience job gain/loss as an immediate shock to bargaining power that alters wages, enrollment utility, or both. Parameter estimates appear in Table SA10. Recent work experience has a larger effect on both wages and enrollment utility than previous experience, as might be expected, though the distinction is precise only for enrollment utility. Other parameter estimates, including those governing learning, change little from the baseline model.

A final form of model misspecification I consider concerns the coefficient on expected wages, which I normalize to 1 in the paper. Estimating this coefficient directly does not alter other parameter estimates by much, as reported in Table A6, column (6) of the paper. Table SA14 shows the option value of re-enrollment calculated under various normalizations of the expected wage coefficient (denoted ρ). Estimating the wage coefficient directly alters option values by only a small amount. Even under an expected wage coefficient of 0.5, meaning the flow of expected wages is

[25] The updating process for employability does not make this distinction; conditional on an individual's previous employment rate during periods of non-enrollment, the particular employment sequence does not matter for posterior beliefs. Although a limitation, I maintain this structure because it is consistent with the maintained hypothesis of Bayesian updating. I discuss the implications for distinguishing learning versus persistence in outcomes elsewhere in this document.

valued at just one half of its nominal amount, this option value changes proportionally little.

H Post-secondary schooling subsidy

Given the sharp increase in school fees for post-secondary education, I consider how a subsidy for post-secondary education would affect enrollment. Specifically, I simulate a 25% reduction in school fees for post-secondary education by modifying the school fee term in the enrollment utility function (3) to reflect the subsidy. The post-secondary fee subsidy has small effects on the proportion of the sample enrolling in post-secondary school by age 22. Table SA15 shows that in the full sample, the subsidy increases post-secondary enrollment by 0.3 percentage points, an amount which is statistically indistinguishable from zero. The subsidy also has negligible and statistically insignificant effects on youth from the first two quintiles of household incomes, the groups most likely to face credit constraints. The results suggest that fees do not explain much about post-secondary enrollment patterns. Extending the subsidy to eliminate all school fees for primary and secondary school has similarly small effects on the schooling distribution, with statistically insignificant increases in the schooling distribution at age 20.[26]

References

Arcidiacono, Peter and John Bailey Jones, "Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm," *Econometrica*, May 2003, *71* (3).

____ and Paul B. Ellickson, "Practical Methods for Estimation of Dynamic Discrete Choice Models," *Annual Review of Economics*, 2011, *3* (1), 363–394.

____ and Robert A. Miller, "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Unpublished working paper*, 2010.

____ and ____ , "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, November 2011, *79* (6), 1823–1867.

Borkum, Evan, "Can Eliminating School Fees in Poor Districts Boost Enrollment? Evidence from South Africa," *Economic Development and Cultural Change*, January 2012, *60* (2), 359–398.

Crawford, Gregory S. and Matthew Shum, "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, July 2005, *73* (4).

[26] Results not shown but available upon request. These results are consistent with other studies (Borkum 2012, Garlick 2012), who find that the limited introduction of fee-free schooling in South Africa had modest effects on enrollment of around 1-2 percentage points at the secondary level. This policy was enacted in 2007, after the end of the sample period I use in this paper, so I am unable to take advantage of this policy variation.

Fiske, Edward B. and Helen F. Ladd, *Elusive Equity: Education Reform in Post Apartheid South Africa*, Brookings Institution Press, July 2004.

Garlick, Robert, "How price sensitive is secondary school enrollment? Evidence from nationwide tuition fee reforms in South Africa," 2012.

Heckman, James J. and Salvador Navarro, "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, February 2007, *136* (2), 341–96.

Hotz, V. Joseph and Robert A. Miller, "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, July 1993, *60* (3).

Magnac, Thierry and David Thesmar, "Identifying Dynamic Discrete Decision Processes," *Econometrica*, March 2002, *70* (2), 801–816.

Manski, Charles F., "Dynamic Choice in Social Settings: Learnings from the Experiences of Others," *Journal of Econometrics*, July 1993, *58* (1-2), 121–136.

Rust, John, "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 1987, *55* (5).

——, "Structural Estimation of Markov Decision Processes," in Robert F. Engle and Daniel L. McFadden, eds., *Handbook of econometrics. Volume 4.*, Handbooks in Economics, vol. 2., 1994, pp. 3081–3143.

Stange, Kevin M., "An Empirical Investigation of the Option Value of College Enrollment," *American Economic Journal: Applied Economics*, January 2012, *4* (1), 49–84.

Train, K. E., "EM algorithms for nonparametric estimation of mixing distributions," *Journal of Choice Modelling*, 2008, *1* (1), 40–69.

Table SA1: Monte Carlo simulation results

	True	Estimated	
		type	
		observed	unobserved
	(1)	(2)	(3)
State transition			
Prior mean			
type 1	-0.7	-0.701	-0.698
		(0.045)	(0.045)
type 2	-2	-1.999	-1.998
		(0.026)	(0.026)
lagged state	1	0.991	0.991
		(0.045)	(0.045)
Prior precision			
type 1	3	3.000	2.993
		(0.246)	(0.248)
type 2	10	10.097	10.083
		(0.686)	(0.692)
Wage			
type 1	1	1.000	1.005
		(0.011)	(0.011)
type 2	2	1.999	1.998
		(0.007)	(0.007)
lagged state	0.4	0.401	0.399
		(0.008)	(0.009)
lagged choice	0.1	0.100	0.098
		(0.008)	(0.008)
Utility function			
type 1	1	0.943	0.987
		(0.063)	(0.063)
type 2	-1	-1.061	-1.058
		(0.030)	(0.030)
lagged state	0.5	0.518	0.517
		(0.029)	(0.030)
lagged choice	1	1.005	1.002
		(0.028)	(0.028)
Type proportions			
Pr(type=1)	0.2	N/A	0.199
			(0.007)
N		3,374	3,374
T		10	10
bootstrap replications		100	100

Table SA2: Panel balance

		Proportion censored at age			
	N	18	20	22	24
full sample	3,374	0.00	0.26	0.57	0.82
black	1,714	0.00	0.26	0.55	0.81
coloured	1,660	0.00	0.27	0.57	0.82

Cells show number of observations (N) or percent of sample with missing enrollment information by age. Survey weights used in calculation.

Table SA3: Schooling, by re-enrollment history

	Re-enrollment history		
	Never	At least once	At least once before completing grade 12
Completed schooling			
<9 years	24.2	4.5	6.6
9-11 years	43.7	42.8	62.2
12 years	24.3	28.6	26.4
>12 years	7.8	24.1	4.9
Grades failed	1.4	1.4	1.8
Schooling after re-enrollment	0.0	1.5	1.7

Table SA4: Choices and outcomes, by type

Type	type 1		type 2	
Initial schooling	low	high	low	high
Completed schooling				
panel (a): attainment at age 20				
<9 years	0.38	0.10	0.24	0.02
$9-11$ years	0.47	0.39	0.67	0.45
12 years	0.13	0.33	0.08	0.25
>12 years	0.01	0.19	0.01	0.28
panel (b): enrollment				
<9 years	0.53	0.74	0.77	0.94
$9-11$ years	0.27	0.39	0.46	0.55
12 years	0.04	0.10	0.10	0.26
>12 years	0.09	0.34	0.28	0.55
panel (c): dropout				
<9 years	0.14	0.06	0.06	0.01
$9-11$ years	0.39	0.24	0.25	0.16
12 years	0.87	0.65	0.70	0.39
>12 years	0.73	0.34	0.45	0.22
panel (d): re-enrollment				
<9 years	0.08	0.10	0.10	0.17
$9-11$ years	0.06	0.07	0.07	0.08
12 years	0.01	0.02	0.02	0.03
>12 years	0.02	0.05	0.04	0.07

Table shows proportion of sample with given schooling attainment at age 20, enrollment, dropout,and re-enrollment (panels a-d, respectively) in model simulation, by type and initial schooling. "Dropout" is defined as non-enrollment following a period of enrollment. In simulation, 30 simulated histories are generated for each observation, with 15 each for each unobserved type.

Table SA5: Robustness checks: enrollment parameter (ϕ_b) estimates

	(1) $T = 52$	(2) $T = 65$	(3) $T = 80$	(4) $\beta = .5$	(5) $\beta = .9$	(6) no wage-age profile	(7) wage coeff. estimated
constant	-7.41	-7.40	-7.40	-5.05	-7.06	-5.81	-7.38
	(0.30)	(0.30)	(0.30)	(0.27)	(0.30)	(0.36)	(0.30)
Type 2	1.12	1.12	1.12	1.29	1.15	-1.88	1.11
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
black	0.82	0.83	0.83	0.83	0.76	0.88	0.86
	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
female	-0.02	-0.01	-0.02	-0.06	-0.05	-0.05	0.00
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
$\mathbb{I}$(schooling $\geq$ 6, age 12)	-0.09	-0.10	-0.10	-0.06	-0.07	-0.06	-0.12
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
ability quartile 2	0.00	0.00	0.00	0.08	0.00	0.00	0.01
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
ability quartile 3	-0.01	-0.02	-0.02	0.15	0.01	-0.01	-0.03
	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)
ability quartile 4	0.16	0.14	0.15	0.55	0.26	0.11	0.09
	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)
HH income quintile 1	-0.27	-0.27	-0.27	-0.32	-0.27	-0.33	-0.28
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
HH income quintile 2	-0.19	-0.19	-0.19	-0.26	-0.20	-0.28	-0.19
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
schooling	0.84	0.84	0.84	0.70	0.83	0.86	0.83
	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
high school graduate	-2.82	-2.87	-2.85	-1.73	-2.55	-3.10	-2.97
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.14)	(0.13)
post-secondary schooling	0.17	0.20	0.19	0.17	0.09	0.17	0.31
	(0.11)	(0.11)	(0.11)	(0.10)	(0.10)	(0.11)	(0.12)
schooling*d_{-1}	-0.54	-0.54	-0.54	-0.53	-0.53	-0.54	-0.54
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
high school graduate*d_{-1}	-0.36	-0.36	-0.36	-0.22	-0.35	-0.42	-0.34
	(0.14)	(0.14)	(0.14)	(0.13)	(0.14)	(0.15)	(0.15)
post-secondary schooling*d_{-1}	0.73	0.73	0.73	0.71	0.70	0.76	0.76
	(0.14)	(0.14)	(0.14)	(0.13)	(0.14)	(0.15)	(0.15)
HH income quintile 1*HSG	0.02	0.04	0.03	-0.14	-0.03	0.08	0.07
	(0.12)	(0.12)	(0.12)	(0.11)	(0.12)	(0.12)	(0.13)
HH income quintile 1*HSG	-0.22	-0.22	-0.22	-0.33	-0.25	-0.17	-0.21
	(0.12)	(0.13)	(0.13)	(0.12)	(0.12)	(0.13)	(0.13)
failed last grade enrolled	-2.48	-2.48	-2.48	-2.06	-2.40	-2.50	-2.47
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
previous grade failures	-0.12	-0.11	-0.12	-0.15	-0.13	-0.22	-0.10
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
enrolled last period (d_{-1})	6.69	6.69	6.69	6.59	6.62	6.66	6.71
	(0.31)	(0.31)	(0.31)	(0.28)	(0.30)	(0.36)	(0.31)
work experience	-0.55	-0.57	-0.56	-0.50	-0.52	-0.59	-0.59
	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)
years enrolled	-0.35	-0.35	-0.35	-0.51	-0.38	-0.35	-0.33
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)
years not enrolled	-0.44	-0.44	-0.44	-0.50	-0.45	-0.48	-0.43
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
pension-eligible grandparent	-0.11	-0.10	-0.10	-0.19	-0.13	-0.08	-0.09
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)
E(wage)							1.25
							(0.11)
N	3,374	3,374	3,374	3,374	3,374	3,374	3,374
$\ln L$	-734,683.5	-734,084.7	-734257.0	-721,350.0	-732,051.7	-750,789.5	-733,581.1

Table shows enrollment utility parameter (ϕ_b) estimates under modifications to model, as indicated by column. T refers to time horizon of model; β refers to discount factor; "no wage-age profile" refers to absence of wage growth with age after decision horizon T_d ends; and "wage coefficient estimated" allows coefficient on expected wages in utility function to differ from 1. Model with $T = 80$ assumes pension receipts of R18,000/year from ages 61-80. All estimates use survey and type-specific probability weights estimated from EM algorithm of Stage 1 of model. All standard errors robust and based on enrollment likelihood function, but may underestimate true standard errors due to sampling variability in Stage 1 parameter estimates used as inputs to model.

Table SA6: Structural parameter estimates: no unobserved heterogeneity

	academic ability		employability		enrollment		wages	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
prior precision	α_e		α_j					
	12.40	(1.44)	3.54	(0.22)				
prior mean/utility	π_{0t}		ξ_{0t}		ϕ_b		ϕ_w	
constant	4.44	(0.15)	-2.85	(0.24)	-6.15	(0.55)	0.96	(0.25)
black	0.25	(0.06)	-0.98	(0.07)	0.65	(0.05)	-0.68	(0.07)
female	0.35	(0.06)	-0.44	(0.07)	0.01	(0.03)	-0.60	(0.08)
$\mathbb{I}$(schooling ≥ 6, age 12)	0.85	(0.06)	-0.24	(0.08)	-0.22	(0.05)	0.22	(0.09)
ability quartile 2	0.51	(0.07)	-0.18	(0.10)	-0.02	(0.04)	-0.01	(0.08)
ability quartile 3	0.97	(0.09)	-0.15	(0.10)	-0.01	(0.04)	0.07	(0.09)
ability quartile 4	1.92	(0.10)	-0.01	(0.11)	0.13	(0.07)	0.67	(0.15)
HH income quintile 1					-0.23	(0.05)		
HH income quintile 2					-0.17	(0.04)		
schooling	-0.56	(0.02)	0.31	(0.03)	0.69	(0.06)	0.15	(0.03)
high school graduate	1.62	(0.21)	0.15	(0.11)	-2.74	(0.31)	0.00	(0.13)
post-secondary schooling	1.27	(0.29)	-0.19	(0.10)	-0.02	(0.06)	1.06	(0.22)
schooling*d_{-1}					-0.46	(0.04)		
high school graduate*d_{-1}					-0.45	(0.16)		
post-secondary schooling*d_{-1}					0.70	(0.14)		
HH income quintile 1*HSG					0.10	(0.12)		
HH income quintile 2*HSG					-0.20	(0.14)		
failed last grade enrolled					-2.12	(0.09)		
previous grade failures					0.00	(0.04)		
enrolled last period (d_{-1})					5.91	(0.46)		
work experience					-0.47	(0.09)	0.09	(0.04)
years enrolled					-0.20	(0.02)		
years not enrolled					-0.46	(0.03)	0.08	(0.03)
pension-eligible grandparent	-0.17	(0.10)	-0.02	(0.10)	-0.02	(0.06)	0.05	(0.11)
$\mathbb{I}$(bad macro environment)			-1.76	(0.10)			-0.83	(0.06)
N	3,374							
$\ln L$	-766,359							

Table shows parameter estimates for prior mean and prior precision in equation (5) for grade advancement and equation (8) for employment; and for enrollment utility and wages (equations (3) and (2), respectively), in R10,000/year. Model does not include unobserved heterogeneity ($M = 1$). Log likelihood value of second stage of sequential maximum likelihood estimation shown. Standard errors calculated from 100 bootstrap replications.

Table SA7: Structural parameter estimates: no post-secondary schooling

	academic ability		employability		enrollment		wages	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
prior precision	α_e		α_j					
Type 1	0.51	(0.10)	3.38	(0.48)				
Type 2	59.90	(25.67)	3.80	(0.36)				
prior mean/utility	π_{0t}		ξ_{0t}		ϕ_b		ϕ_w	
constant	5.12	(0.23)	-3.51	(0.22)	-3.66	(0.39)	0.25	(0.18)
Type 2	-0.71	(0.21)	0.59	(0.09)	1.53	(0.05)	0.43	(0.06)
black	0.25	(0.05)	-0.95	(0.07)	1.18	(0.06)	-0.63	(0.05)
female	0.28	(0.05)	-0.45	(0.06)	0.02	(0.05)	-0.66	(0.06)
$\mathbb{I}(\text{schooling} \geq 6,\ \text{age } 12)$	0.47	(0.06)	-0.15	(0.08)	-0.25	(0.08)	0.23	(0.06)
ability quartile 2	0.44	(0.06)	-0.16	(0.09)	0.10	(0.07)	-0.02	(0.06)
ability quartile 3	0.76	(0.07)	-0.14	(0.09)	0.16	(0.07)	0.09	(0.07)
ability quartile 4	1.44	(0.09)	-0.02	(0.11)	0.14	(0.09)	0.52	(0.11)
HH income quintile 1					-0.34	(0.07)		
HH income quintile 2					-0.27	(0.06)		
schooling	-0.54	(0.02)	0.32	(0.02)	0.55	(0.04)	0.17	(0.02)
schooling*d_{-1}					-0.57	(0.04)		
failed last grade enrolled					-2.71	(0.08)		
previous grade failures					0.05	(0.06)		
enrolled last period (d)					6.36	(0.38)		
work experience					-2.07	(0.29)	0.07	(0.03)
years enrolled					-0.55	(0.03)		
years not enrolled					-1.39	(0.06)	0.14	(0.02)
pension-eligible grandparent	-0.13	(0.08)	-0.04	(0.09)	-0.02	(0.09)	0.07	(0.08)
$\mathbb{I}(\text{bad macro environment})$			-1.46	(0.08)			-0.66	(0.06)
Type proportions	$\omega_{m\|s_0}$							
Type 1								
low initial schooling	0.13	(0.24)						
high initial schooling	0.23	(0.29)						
Type 2								
low initial schooling	0.87	(0.24)						
high initial schooling	0.77	(0.29)						
N	3,374							
$\ln L$	-593,910.1							

Table shows parameter estimates for prior mean and prior precision in equation (5) for grade advancement and equation (8) for employment; and for enrollment utility and wages (equations (3) and (2), respectively), in R10,000/year. Sample excludes agents who ever enrolled in post-secondary schooling. Estimates of proportion of each unobserved type also shown. Log likelihood value from Stage 2 of sequential maximum likelihood estimation reported. Standard errors are asymptotic results from sequential maximum likelihood estimation, and may underestimate true standard errors.

Table SA8: Structural parameter estimates: sample includes whites

	academic ability		employability		enrollment		wages	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
prior precision	α_e		α_j					
Type 1	11.44	(1.74)	2.71	(0.22)				
Type 2	14.70	(2.59)	2.10	(0.17)				
prior mean/utility	π_{0t}		ξ_{0t}		ϕ_b		ϕ_w	
constant	5.84	(0.20)	-4.41	(0.32)	-6.93	(0.44)	3.61	(0.47)
Type 2	-0.04	(0.07)	-0.07	(0.08)	-0.08	(0.04)	-0.06	(0.04)
black	-0.11	(0.15)	-1.30	(0.17)	0.43	(0.11)	-3.55	(0.39)
coloured	-0.53	(0.14)	0.03	(0.16)	-0.37	(0.11)	-2.86	(0.40)
female	0.36	(0.05)	-0.46	(0.07)	0.04	(0.04)	-0.43	(0.08)
$\mathbb{I}(\text{schooling} \geq 6,\ \text{age } 12)$	1.06	(0.06)	-0.40	(0.08)	-0.27	(0.07)	0.04	(0.05)
ability quartile 2	0.74	(0.06)	-0.21	(0.09)	-0.05	(0.05)	0.00	(0.05)
ability quartile 3	1.55	(0.08)	0.00	(0.10)	-0.05	(0.06)	0.19	(0.06)
ability quartile 4	2.44	(0.10)	-0.04	(0.13)	0.32	(0.08)	0.50	(0.09)
HH income quintile 1					-0.33	(0.06)	0.15	(0.02)
HH income quintile 2					-0.21	(0.06)	-0.15	(0.11)
schooling	-0.67	(0.02)	0.48	(0.03)	0.85	(0.05)	1.79	(0.50)
high school graduate	1.65	(0.19)	-0.03	(0.10)	-2.56	(0.06)	0.01	(0.07)
post-secondary schooling	1.08	(0.22)	-0.40	(0.09)	-0.06	(0.05)	-0.68	(0.09)
schooling*d_{-1}					-0.19	(0.16)		
high school graduate*d_{-1}					0.62	(0.17)		
post-secondary schooling*d_{-1}					0.13	(0.14)		
HH income quintile 1*HSG					0.13	(0.14)		
HH income quintile 2*HSG					-0.22	(0.13)		
failed last grade enrolled					-2.56	(0.06)		
previous grade failures					-0.06	(0.05)		
enrolled last period (d_{-1})					6.69	(0.42)		
work experience					-1.10	(0.14)	0.11	(0.04)
years enrolled					-0.25	(0.03)		
years not enrolled					-0.58	(0.04)	0.11	(0.02)
pension-eligible grandparent	-0.14	(0.09)	0.19	(0.09)	-0.06	(0.07)	0.01	(0.07)
$\mathbb{I}(\text{bad macro environment})$			-2.18	(0.09)			-0.68	(0.09)
Type proportions	$\omega_{m\|s_0}$							
Type 1								
low initial schooling	0.58	(0.41)						
high initial schooling	0.62	(0.40)						
Type 2								
low initial schooling	0.45	(0.40)						
high initial schooling	0.47	(0.40)						
N	3,374							
$\ln L$	-771,708.5							

Table shows parameter estimates for prior mean and prior precision in equation (5) for grade advancement and equation (8) for employment; and for enrollment utility and wages (equations (3) and (2), respectively), in R10,000/year. Estimates of proportion of each unobserved type also shown. Sample includes whites, but otherwise identical to main estimation sample. Log likelihood value from Stage 1 of sequential maximum likelihood estimation. Standard errors are asymptotic results from sequential maximum likelihood estimation, and may underestimate true standard errors.

Table SA9: Structural parameter estimates: no exclusion restrictions

	academic ability		employability		enrollment		wages	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
prior precision	α_e		α_j					
Type 1	94.94	(64.26)	3.92	(0.37)				
Type 2	39.92	(10.00)	3.97	(0.39)				
prior mean/utility	π_{0t}		ξ_{0t}		ϕ_b		ϕ_w	
constant	3.96	(0.64)	-4.20	(0.37)	-6.89	(0.31)	1.02	(0.24)
Type 2	-1.26	(0.05)	0.24	(0.07)	1.47	(0.05)	1.23	(0.10)
black	0.49	(0.05)	-0.91	(0.07)	0.87	(0.05)	-0.45	(0.06)
female	0.35	(0.04)	-0.41	(0.06)	-0.04	(0.04)	-0.60	(0.05)
$\mathbb{I}(\text{schooling} \geq 6,\ \text{age } 12)$	0.36	(0.05)	-0.41	(0.09)	-0.01	(0.07)	0.61	(0.09)
ability quartile 2	0.42	(0.05)	-0.21	(0.08)	-0.01	(0.06)	-0.03	(0.06)
ability quartile 3	0.83	(0.06)	-0.18	(0.09)	-0.10	(0.06)	-0.03	(0.07)
ability quartile 4	1.65	(0.08)	-0.03	(0.10)	0.09	(0.07)	0.55	(0.09)
HH income quintile 1	-0.50	(0.06)	-0.19	(0.09)	-0.22	(0.06)	-0.29	(0.07)
HH income quintile 2	-0.39	(0.06)	-0.25	(0.09)	-0.16	(0.06)	-0.26	(0.06)
schooling	-0.28	(0.06)	0.48	(0.05)	0.80	(0.04)	0.09	(0.03)
high school graduate	1.46	(0.32)	0.12	(0.14)	-2.91	(0.13)	-0.07	(0.15)
post-secondary schooling	0.23	(0.37)	0.11	(0.15)	0.05	(0.11)	0.88	(0.15)
schooling*d_{-1}	-0.15	(0.06)	-0.12	(0.05)	-0.45	(0.03)	-0.03	(0.05)
high school graduate*d_{-1}	0.40	(0.35)	0.79	(0.18)	-0.72	(0.16)	-0.45	(0.22)
post-secondary schooling*d_{-1}	1.48	(0.52)	-0.49	(0.18)	0.84	(0.15)	0.18	(0.27)
HH income quintile 1*HSG	-0.09	(0.31)	-0.28	(0.16)	-0.05	(0.13)	-0.37	(0.12)
HH income quintile 2*HSG	0.90	(0.44)	-0.02	(0.16)	-0.31	(0.14)	-0.37	(0.14)
failed last grade enrolled			0.76	(0.12)	-2.75	(0.12)	-0.53	(0.13)
previous grade failures			0.49	(0.10)	-0.35	(0.11)	-0.39	(0.09)
enrolled last period (d_{-1})	1.08	(0.64)	1.28	(0.43)	6.05	(0.32)	0.38	(0.49)
grades passed since age 12			0.21	(0.10)	0.03	(0.11)	0.07	(0.08)
work experience	0.77	(0.24)			-0.32	(0.08)	0.05	(0.03)
years enrolled			-0.37	(0.10)	-0.29	(0.11)	0.00	(0.08)
years not enrolled	-0.78	(0.08)			-0.61	(0.04)	0.11	(0.02)
pension-eligible grandparent	-0.04	(0.07)	-0.04	(0.08)	0.00	(0.07)	0.09	(0.07)
$\mathbb{I}(\text{bad macro environment})$	-1.00	(0.06)	-1.72	(0.08)	-1.69	(0.05)	-0.80	(0.06)
Type proportions	$\omega_{m\|s_0}$							
Type 1								
low initial schooling	0.34	(0.33)						
high initial schooling	0.58	(0.36)						
Type 2								
low initial schooling	0.66	(0.33)						
high initial schooling	0.42	(0.36)						
N	3,374							
$\ln L$	-747,722.8							

Table shows parameter estimates for prior mean and prior precision in equation (5) for grade advancement and equation (8) for employment; and for enrollment utility and wages (equations (3) and (2), respectively), in R10,000/year. Estimates of proportion of each unobserved type also shown. Model has no cross-equation exclusion restrictions except those necessary to avoid perfect collinearity. Log likelihood value from Stage 1 of sequential maximum likelihood estimation. Standard errors are asymptotic results from sequential maximum likelihood estimation, and may underestimate true standard errors.

Table SA10: Parameter estimates with “worked last period” included

	academic ability		employability		enrollment		wages	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
prior precision	α_e		α_j					
Type 1	4.75	(0.54)	3.46	(0.31)				
Type 2	34.12	(8.12)	4.45	(0.42)				
prior mean/utility	π_{0t}		ξ_{0t}		ϕ_b		ϕ_w	
constant	4.74	(0.12)	-3.38	(0.21)	-6.95	(0.29)	0.11	(0.18)
Type 2	-0.49	(0.06)	0.81	(0.06)	1.07	(0.04)	0.83	(0.07)
black	0.26	(0.05)	-0.98	(0.06)	0.80	(0.05)	-0.70	(0.05)
female	0.36	(0.04)	-0.44	(0.06)	-0.02	(0.04)	-0.62	(0.05)
$\mathbb{I}$(schooling ≥ 6, age 12)	0.57	(0.05)	-0.01	(0.07)	-0.11	(0.06)	0.41	(0.06)
ability quartile 2	0.48	(0.05)	-0.17	(0.08)	0.00	(0.05)	-0.01	(0.06)
ability quartile 3	0.93	(0.06)	-0.16	(0.08)	-0.02	(0.06)	0.07	(0.07)
ability quartile 4	1.88	(0.08)	0.00	(0.10)	0.14	(0.07)	0.70	(0.10)
HH income quintile 1					-0.26	(0.05)		
HH income quintile 2					-0.19	(0.05)		
schooling (s)	-0.53	(0.02)	0.30	(0.02)	0.80	(0.03)	0.16	(0.02)
high school graduate (HSG)	1.70	(0.15)	0.28	(0.09)	-2.66	(0.12)	0.16	(0.09)
post-secondary schooling (s_{ps})	1.30	(0.23)	-0.23	(0.08)	0.18	(0.11)	1.03	(0.13)
schooling*d_{-1}					-0.51	(0.03)		
high school graduate*d_{-1}					-0.47	(0.14)		
post-secondary schooling*d_{-1}					0.76	(0.14)		
HH income quintile 1*HSG					0.02	(0.12)		
HH income quintile 2*HSG					-0.23	(0.12)		
failed last grade enrolled					-2.43	(0.05)		
previous grade failures					-0.10	(0.04)		
enrolled last period (d_{-1})					6.26	(0.30)		
worked last non-enrolled period					-1.11	(0.11)	0.06	(0.06)
previous work experience					-0.12	(0.08)	0.01	(0.03)
years enrolled					-0.49	(0.03)		
years not enrolled					-0.51	(0.03)	0.17	(0.02)
pension-eligible grandparent	-0.15	(0.07)	-0.06	(0.08)	-0.10	(0.06)	0.03	(0.08)
$\mathbb{I}$(bad macro environment)			-1.79	(0.08)			-0.73	(0.06)
Type proportions	$\omega_{m\|s_0}$							
Type 1								
low initial schooling	0.28	(0.28)						
high initial schooling	0.56	(0.32)						
Type 2								
low initial schooling	0.72	(0.28)						
high initial schooling	0.44	(0.32)						
N	3,374							
$\ln L$	-731,654.2							

Standard errors not bootstrapped, and may be underestimated.

Table SA11: Structural parameter estimates, omitting initial schooling

	academic ability		employability		enrollment		wages	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
prior precision	α_e		α_j					
Type 1	365.36	(740.05)	3.43	(0.33)				
Type 2	41.19	(10.75)	4.17	(0.40)				
prior mean/utility	π_{0t}		ξ_{0t}		ϕ_b		ϕ_w	
constant	5.57	(0.14)	-3.57	(0.23)	-7.64	(0.33)	0.09	(0.19)
Type 2	-1.64	(0.05)	0.71	(0.06)	1.27	(0.04)	0.52	(0.06)
black	0.18	(0.04)	-0.98	(0.06)	0.84	(0.05)	-0.78	(0.05)
female	0.40	(0.04)	-0.45	(0.06)	-0.03	(0.04)	-0.59	(0.05)
ability quartile 2	0.97	(0.06)	-0.20	(0.08)	-0.03	(0.05)	0.00	(0.06)
ability quartile 3	1.96	(0.08)	-0.19	(0.09)	-0.09	(0.06)	0.10	(0.07)
ability quartile 4	-0.48	(0.02)	-0.08	(0.10)	-0.01	(0.07)	0.71	(0.10)
HH income quintile 1					-0.28	(0.06)		
HH income quintile 2					-0.20	(0.05)		
schooling (s)	-0.48	(0.02)	0.33	(0.03)	0.84	(0.04)	0.20	(0.02)
high school graduate (HSG)	1.66	(0.14)	0.27	(0.09)	-0.21	(0.04)	0.11	(0.09)
post-secondary schooling (s_{ps})	1.28	(0.22)	-0.23	(0.08)	0.22	(0.11)	1.02	(0.12)
schooling*d_{-1}					-0.52	(0.03)		
high school graduate*d_{-1}					-0.41	(0.15)		
post-secondary schooling*d_{-1}					0.70	(0.14)		
HH income quintile 1*HSG					0.01	(0.12)		
HH income quintile 2*HSG					-0.24	(0.13)		
failed last grade enrolled					-2.58	(0.05)		
previous grade failures					-0.21	(0.04)		
enrolled last period (d_−1)					6.46	(0.34)		
work experience					-0.56	(0.08)	0.05	(0.03)
years enrolled					-0.32	(0.02)		
years not enrolled					-0.43	(0.03)	0.13	(0.02)
pension-eligible grandparent	-0.12	(0.07)	-0.05	(0.08)	-0.08	(0.06)	0.05	(0.08)
‘ $\mathbb{I}$(bad macro environment)			-1.77	(0.08)			-0.77	(0.06)
Type proportions			$\omega_{m\|s_0}$					
Type 1								
low initial schooling				0.24	(0.28)			
high initial schooling				0.58	(0.35)			
Type 2								
low initial schooling				0.76	(0.28)			
high initial schooling				0.42	(0.35)			
N	3,374							
$\ln L$	-735,666.8							

Table shows parameter estimates for prior mean and prior precision in equation (5) for grade advancement and equation (8) for employment; and for enrollment utility and wages (equations (3) and (2), respectively), in R10,000/year. Estimates of proportion of each unobserved type also shown. Log likelihood value from Stage 2 of sequential maximum likelihood estimation. Standard errors not bootstrapped, and may be underestimated. Standard errors of precision parameters calculated using Delta Method.

Table SA12: Dropout and re-enrollment in response to household shocks

	(1) dropout	(2) dropout	(3) re-enroll	(4) re-enroll
household shock	-0.001		-0.020	
	(0.031)		(0.010)**	
household shock($a-1$)		-0.038		0.001
		(0.093)		(0.017)
failed last grade enrolled	0.270	0.135		
	(0.053)***	(0.122)		
cumulated grades failed	0.169	0.323		
	(0.051)***	(0.134)**		
worked last period of non-enrollment			-0.027	-0.009
			(0.017)	(0.021)
work experience			-0.002	-0.036
			(0.013)	(0.049)
Observations	4085	2576	4568	3081
R-squared	0.79	0.83	0.79	0.9

Sample is person-years for which household shock defined. All regressions include individual, age and schooling fixed effects. "Household shock" refers to death; serious illness or injury; job loss; business failure or bankruptcy; abandonment or divorce; theft, fire or property damage; or other major shock. Contemporaneous and lagged household shock coefficients not included in same regression because they are not separately identified, due to lack of observations with both variables for the same individual in the sample. Standard errors clustered by individual.

Table SA13: Self-reported reasons for dropout

Reason for dropout	proportion
economic	0.15
academic/behavioral	0.03
health (self)	0.07
family	0.03
pregnancy or baby	0.02
pregnancy or baby (females only)	0.04
other/don't know/no response	0.73

Cells report survey-weighted fraction of observations in person-year panel among students who transition from enrollment to non-enrollment.uts (reasons for dropout and can't afford school). Reasons for non-enrollment classified as follows: "economic" includes work, search, or can't afford school; "academic/behavioral" includes passed matric or completed all offered schooling, failed, expelled, school isn't important, or imprisoned; "family" includes married, pregnant, moved or caring for ill relative.

Table SA14: Option value of re-enrollment under alternative expected wage normalizations

normalization	$\rho = 1$	$\rho = 0.75$	$\rho = 0.5$	ρ estimated
age=12	326,884	312,930	299,631	341,334
schooling=6	233,544	222,358	212,939	243,879

Table shows option value (South African rand) of value of labor market participation under alternative normalizations of expected wage coefficient (ρ). Option value calculated as difference between value function for entering labor force with re-enrollment permitted before completing high school minus the analogous value function with re-enrollment prohibited before completing high school, based on simulation using age 12 state as initial condition. In simulation, 30 simulated histories are generated for each observation, with unobserved type assigned randomly based on individual-specific type probability. Survey weights and individual-specific type probability weights applied to calculation.

Table SA15: Enrollment in post-secondary schooling by age 22 under post-secondary fee subsidy

	Predicted (no subsidy) (1)	Predicted (25% subsidy) (2)	Difference (2)-(1)
full sample	17.4	17.7	0.3
	(0.1)	(0.1)	(0.2)
household income quintile			
first (bottom)	11.2	11.4	0.2
	(0.2)	(0.2)	(0.3)
second	10.1	10.3	0.2
	(0.2)	(0.2)	(0.3)

Table shows percentage of sample enrolling in post-secondary education by age 22, by indicated characteristic, based on model simulation. In simulation, 30 simulated histories are generated for each observation, with unobserved type assigned randomly based on individual-specific type probability. Column (1) shows simulation results of unrestricted model. Column (1) shows simulation results of unrestricted model. Column (2) shows simulation results in which post-secondary fees reduced by 25%. Column (3) shows difference is means, Column (2) minus Column (1). All results calculated using survey weights. All results calculated using survey weights.

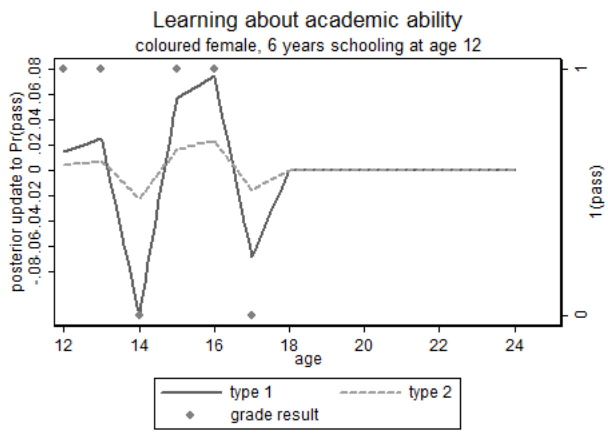
Table SA16: Option value of re-enrollment before secondary school completion, setting enrollment utility to zero

completed schooling	option value amount	option value % of LFP	age	option value amount	option value % of LFP
0	412,177	37.3	12	667,771	47.2
1	585,439	51.8	13	623,234	44.5
2	572,364	48.5	14	572,548	41.4
3	581,837	47.4	15	515,339	37.6
4	591,800	46.0	16	450,996	33.2
5	588,831	44.2	17	372,610	27.8
6	569,681	41.9	18	290,909	22.2
7	511,619	37.6	19	232,924	18.0
8	437,816	32.2	20	188,050	14.7
9	364,018	26.7	21	153,422	12.1
10	289,592	21.0	22	119,180	9.4
11	215,788	15.4	23	82,785	6.6

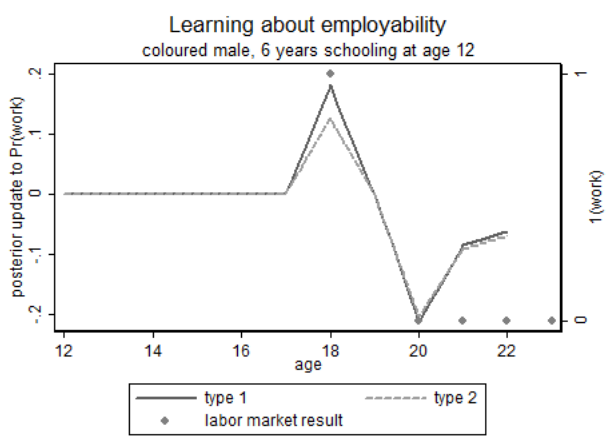
Table shows option value of re-enrollment before high school completion in South African rand and as proportion of overall value of labor market participation, averaged over individuals with years of completed schooling or age indicated in table. Option value calculated as difference between value function for entering labor force with re-enrollment permitted before completing high school, excluding the flow utility of enrollment, minus the analogous value function with re-enrollment prohibited before completing high school. Survey weights and individual-specific type probability weights applied to calculation.

Figure SA1: Model estimates: learning examples

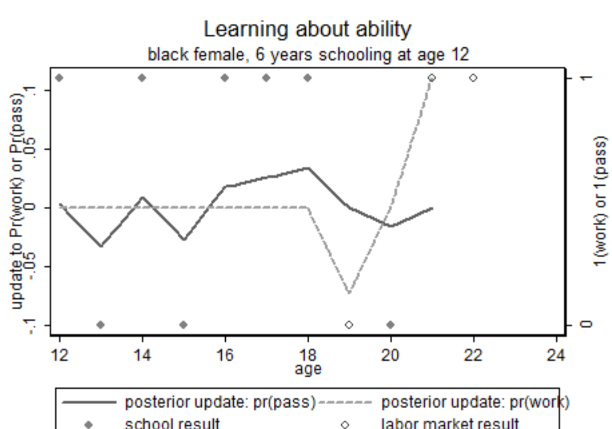
(a)



(b)



(c)



Line graphs in this figure plot the change in the posterior prediction, normalized by the change in prior mean. In other words, the lines in panels (a)-(b) plot $[\Pr(p_{t+1} = 1|\pi_t, m) - \Pr(p_t = 1|\pi_{t-1}, m)] - [\pi_{0,m}(X_t) - \pi_{0,m}(X_{t-1})]$, where all variables are as defined in (**??**); and analogously for the posterior for employment in panels (c)-(d). Although agents apply this updated posterior to the next period in which they enroll or enter the labor market, the update is plotted as concurrent with the most recent result for ease of interpretation.