Using Observed Choices to Infer Agent's Information: Reconsidering the Importance of Borrowing Constraints, Uncertainty and Preferences in College Attendance *

Salvador Navarro

Abstract

I use economic theory and estimates of a semiparametrically identified structural model to analyze the role played by credit constraints, uncertainty and preferences in explaining college attendance. A methodology for inferring information available to the agent from individual choices is proposed and implemented. The test distinguishes which of the unobserved (to the analyst) components of future outcomes are known to the agent and which are unknown to both at a given stage of the life cycle. I use microdata on earnings, schooling and consumption to infer the agent's information set and estimate a model of college choice and consumption under uncertainty with equilibrium borrowing constraints. I estimate that 80% and 44% of the variances of college and high school earnings respectively are predictable by the agent. Moving to a no tuition economy increases college attendance from 48% to 50%. When people are allowed to smooth consumption, college increases to nearly 58%. General equilibrium effects not withstanding, credit constraints have a larger effect than previously suggested.

1 Introduction

Schooling, particularly college, is considered one of the main sources of human capital for the individual. Consequently, understanding why some individuals get a college education and some do not is, and has been for a while, an active area of research. While one may think that by now we would have a relatively good understanding of the determinants of college choice, we still do not have a clear picture. In part this is a consequence of the literature focusing mostly on whether a particular aspect of schooling can be explained by a particular cause. That is, different branches of the literature focus on estimating returns to schooling (e.g. Ashenfelter and Krueger (1994), Heckman and Vytlacil (1998), Card (2001)); ability and returns to education (e.g. Cawley, Heckman, Lochner, and Vytlacil (2000), Taber (2001), Belzil and

*I am particularly grateful to James Heckman for his continued support and assistance. I am indebted to Flavio Cunha, Lars Hansen, Paul Schrimpf and Robert Townsend for their help during the various stages of this paper. I also thank Alessandro Barbarino, Pedro Carneiro, Victor Chernozhukov, Jeremy Fox, Bruce Hansen, Kenneth Judd, John Kennan, Rasmus Lentz, Rosa Matzkin, Ronni Pavan, Alejandro Rodriguez and Ananth Seshadri for their helpful comments. Paul Schrimpf provided excellent research assistance. This work would not have been possible without financial support from CONACYT, Mexico; from the George G. Stigler and the Esther and T.W. Schultz fellowships at the University of Chicago; and from the Graduate School at the University of Wisconsin-Madison.
Address: University of Wisconsin-Madison, 1180 Observatory Drive. Madison, WI. 53706.
E-mail: snavarro@ssc.wisc.edu.

Hansen (2002)); the importance of parental income and, more generally, of borrowing constraints (e.g. Kane (1996), Carneiro and Heckman (2002), Cameron and Taber (2004), Brown, Scholz, and Seshadri (2007)); the impact of ability (e.g. Card (1995), Ashenfelter and Rouse (1998), Cawley, Heckman, and Vytlacil (2001)), etc. The few papers that model these determinants jointly (e.g. Cameron and Heckman (1998), Keane and Wolpin (2001), Cameron and Heckman (2001)) still focus mainly on the role played by a particular determinant.

In most cases, the literature ignores the role played by the uncertainty facing the agent and uses ex-post measures (e.g. earnings at age 40) to analyze the agent's schooling decision. When they account for uncertainty, they assume that the unobserved (to the analyst) variability and the uncertainty facing the agent essentially coincide.[1] One of the main contributions of this paper is to adapt the insight of Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) to develop and implement a methodology that distinguishes information unknown to the econometrician but forecastable by the agent and information unknown to both (*fundamental uncertainty*) at each period. The key to measuring uncertainty is to notice that individual choices reflect all the information known to the agent at a given time.[2] Responses of current decisions to future outcome innovations can be used to infer how much information the agent has. The semiparametric nature of the proposed test allows it to be used independently of the particular specification of the model as long as one considers families of models with the same determinants of choice.

In this paper, I use economic theory and estimates of a semiparametrically identified structural model to analyze the role played by uncertainty and its interaction with credit constraints and preferences in explaining college graduation. A structural model of schooling choice and consumption allocation under uncertainty in which borrowing constraints arise in equilibrium is estimated. Once a general model of schooling choice is estimated, a comparison of the importance of the (nonexclusive) determinants of college graduation can be carried out. The results I obtain do not single out any one particular determinant as the main reason why some people go to college and some don't. That this is the case, i.e. that all aspects of the problem play an important role, should not be surprising given the nature of the decision.

The key empirical results in the paper are:

1. At the time schooling decisions are made, earnings are predictable. In particular, the estimates of the model imply that 81% of the unexplained variance in college earnings is predictable by the agent at age 18. This fraction is 44% for high school earnings so college earnings are more predictable. In fact, while the total unobserved variance of college earnings is higher than that of high school (i.e the unobserved variance from the analyst's perspective), college earnings are less uncertain than high school earnings from the point of view of the agent. This is similar to the results obtained by Cunha, Heckman, and Navarro (2005), Guvenen (2007).

[1] For example, both Keane and Wolpin (2001), Cameron and Heckman (2001) estimate their dynamic models of schooling assuming what is known by the agent and what is known by the econometrician at each point in time. While they allow for unobserved heterogeneity, it is essentially treated as an initial condition. Once the econometrician conditions on the initial heterogeneity, the evolution of the information set of the agent is given. However, there is no prior reason why one should assume that the evolution of what is unknown to the analyst and what is unknown to the agent coincide.

[2] A similar idea motivates the work on the permanent income hypothesis of Flavin (1981) and Pistaferri (2001).

2. Once credit constraints are properly defined and relaxed, they play a much more important role than previously estimated in the literature. When people are allowed to smooth consumption perfectly (*i.e.,* when credit constraints and uncertainty are completely eliminated) college attendance only 5.5% of college graduates regret their choice while 21.7% of high school graduates would rather graduate college. That is, I estimate that college attendance would increase from 49% to nearly 58%. This result is not inconsistent with the evidence presented in Cameron and Heckman (2001), Keane and Wolpin (2001), Carneiro and Heckman (2002) and Cameron and Taber (2004) where credit constraints are found to be relatively unimportant. The analysis in these papers focuses on the inability of individuals to obtain funds to pay for tuition as the credit constraint. When I perform similar tuition reduction simulations I obtain results that are in line with the ones obtained in these papers. The large effect of credit constraints is in part due to the partial equilibrium nature of the exercise. However, it is still illustrative of the importance of credit constraints when one considers the full extent of the constraint and not only the lack of liquidity that maybe associated with tuition.

3. Individuals are relatively impatient. The estimated annual discount rate is 7% while the assumed riskless rate is 3%. This helps explain why, even if college earnings are higher than high school earnings late in life, individuals may choose not to pay the cost of lower earnings in the early part of their life and choose a high school education.

4. Agents may also have preferences over schooling beyond the consumption value of earnings which I capture via an additive "psychic" cost function. Ability is the main determinant of costs and, as such, plays a key role in determining schooling decisions. High ability individuals face very low costs while low ability individuals face large costs of attending college. This gives rise to schooling sorting by ability even though monetary gains do not differ considerably for high school and college graduates conditional on ability.

5. Schooling decisions are made by agents before all the relevant information about future outcomes has been revealed. Individual choices are made in an environment of uncertainty and agents base their decisions on their expectations and not on the realized outcomes observed by the econometrician. Expectations and realizations need not coincide. In particular, I estimate that eliminating uncertainty entirely – but keeping the credit constraints in place, 13% of high school graduates would instead choose to be college graduates and 16% of college graduates would regret their choice under uncertainty and pick high school instead. Aggregate college attendance is almost unchanged at 47.7% from 48.9% under uncertainty.

This paper contributes to the literature on schooling choice by explicitly looking at the role played by uncertainty as a determinant of schooling. The idea is closely related to the work of Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) in which a similar methodology is applied to extract agent's information at the schooling date. While Carneiro, Hansen, and Heckman (2003) assume no credit markets operate[3] and Cunha, Heckman, and Navarro (2005) assume an economy with perfect credit markets; I investigate an intermediate economy in which some credit markets operate and borrowing constraints arise in equilibrium.[4] This allows for a more general setting in which consumption is not equal to income every period nor is it necessary to assume that complete markets operate.

My work also contributes to the literature by modeling credit constraints as endogenous equilibrium phenomena. As opposed to Cameron and Taber (2004) who assume that constraints arise because of an exogenously imposed differential on the interest rate at which people borrow for consumption and that at which they borrow for education. Like Keane and Wolpin (2001), I assume that people face a constraint on the amount of money they can borrow. In my analysis, credit constraints arise in equilibrium as a consequence of repayment restrictions and uncertainty about individuals future income whereas in Keane and Wolpin (2001) the constraint is modeled as an exogenously specified function of human capital and age.

I focus on schooling, consumption and asset holdings. Keane and Wolpin (2001) also consider labor supply, marriage, residency with parents and the like. My simpler model is more easily interpretable and lets me focus on assumptions about the information structure and identification, topics they do not consider.[5] Keane and Wolpin (2001) and Gourinchas and Parker (2002) (among others) assume that shocks to outcomes are unobservable to the econometrician and uncertain to the agent. In my analysis, I infer the amount of uncertainty facing the agent from individual choices. I prove that the model in this paper is semiparametrically identified thus contributing to the growing literature on identification of dynamic discrete choice models.[6]

Finally, I also contribute to the literature on consumption inequality and partial insurance. By extending the insight developed by Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) and applying it to my model, I can identify what constitutes uncertainty at any stage in the life cycle.[7] A similar idea lies behind the test of the permanent income hypothesis in Flavin (1981). She picks a particular assumed ARMA (p, q) time series process for income and tests whether transitory income predicts consumption. Blundell and Preston (1998) and Blundell, Pistaferri, and Preston (2004) use the same idea to test for "partial insurance". As they acknowledge, their estimate of partial insurance combines the effects of information known to the agent but unknown to the econometrician and insurance. The methodology proposed in this paper can, in principle, distinguish between these two explanations.

The rest of the paper proceeds as follows. Section 2 presents a general version of the model of consumption allocation and schooling decisions that I use in the rest of the paper. The methodology used to infer the elements of the agent's information set is explained in section 3. In passing, I briefly sketch how semiparametric identification of the model is achieved. Appendix 1 presents a formal identification analysis. Section 4 describes the data that I use, my analysis of what is in the agent's information set and

[3] They also assume utility is logarithmic so their assumed coefficient of relative risk aversion equals 1.

[4] See Laitner (1992), Aiyagari (1994) and Gourinchas and Parker (2002).

[5] They do allow for variability to be different from uncertainty by allowing for unobserved types. This however, does not evolve over time. Conditional on the unobserved type which is given at the beginning of time, variability=uncertainty.

[6] See Taber (2000), Magnac and Thesmar (2002) and Heckman and Navarro (2007).

[7] Pistaferri (2001) uses a similar idea by looking at expected wages as measured through a survey and measured wages in a consumption analysis.

the empirical results from estimating the model using the right information are presented next. Section 5 concludes.

2 The Model

2.1 The Decision Process

Let $Y_{i,s,t}$ denote individual $i's$ income in schooling level s at time t, $A_{i,s,t}$ denote the assets he saves for the next period, $u\left(C_{i,s,t}\right)$ denote individual utility if the agent consumes $C_{i,s,t}$, $\tilde{u}\left(A_{i,s,T}\right)$ denote the utility in the terminal period when an individual reaches the last period with asset level $A_{i,s,T}$, and ρ be the discount rate. Let $\mathcal{I}_{i,t}$ be the information[8] available to the agent at time t, which is assumed to include all the past and current realizations of earnings as well as the (assumed constant) interest rate r and the asset stock. It may also contain some information about future earnings. Exactly how much is what I estimate in this paper.

Individuals live for $T+1$ periods and maximize expected lifetime utility in a world in which all risks arise from labor market risk and are idiosyncratic. At period $t=0$, for each fixed schooling level s (h=high school, c=college) and given the (expected) income sequence associated with each schooling s, agents select an optimal intertemporal consumption allocation rule. Agents can save and borrow as much as they want, subject to repayment constraints, via a single riskless asset A that pays a return r. Agents then make the schooling choice that maximizes expected utility.

More precisely, the agent's problem at period $t>0$ given schooling level s is to select how much of his available resources to consume and how much to transfer to the next period. The value function given information set $\mathcal{I}_{i,t}$ is

$$V_{i,s,t}\left(\mathcal{I}_{i,t}\right)=max_{A_{i,s,t}}u\left(C_{i,s,t}\right)+\frac{1}{1+\rho}E\left(V_{i,s,t+1}\left(\mathcal{I}_{i,t+1}\right)\mid\mathcal{I}_{i,t}\right) \tag{1}$$

$$s.t.\ C_{i,s,t}=Y_{i,s,t}+W_{i,s,t}+\left(1+r\right)A_{i,s,t-1}-A_{i,s,t},\ \ A_{i,s,0}=0,\ \ A_{i,s,T}\geq 0. \tag{2}$$

If the utility function satisfies standard conditions (*i.e.,* concavity and $lim_{C\rightarrow 0}u'\left(C\right)=\infty$), the restriction that the agent cannot die in debt ($A_{i,s,T}\geq 0$) imposes a borrowing constraint on the individual at every period. In this case, the minimum value that assets can take at any period t (*i.e.,* the maximum amount the agent can borrow) is

$$A_{i,s,t}^{MIN}=\frac{A_{i,s,t+1}^{MIN}-Y_{i,s,t+1}^{MIN}}{1+r}, \tag{3}$$

where $Y_{i,s,t}^{MIN}$ is the minimum certain value that income can take at time t and $W_{i,s,t}=\max\{Y_{i,s,t+1}^{MIN}-Y_{i,s,t},0\}$.

From the agent's perspective (*i.e.,* given his information at time t) the solution to the consumption allocation problem in equation (1) consists of a pair of time-schooling indexed functions: a policy function

[8] By information I mean the minimum sigma algebra generated by the random variables in $\mathcal{I}_{i,t}$. Since I associate the information set with a particular group of random variables $\mathcal{I}_{i,t}$ I use these concepts interchangeably.

that tells him how much to consume this period

$$C^*_{i,s,t} = \mathcal{C}_{s,t}\left(\mathcal{I}_{i,t}\right) \tag{4}$$

and the value function associated with it

$$V^*_{i,s,t} = \mathcal{V}_{s,t}\left(\mathcal{I}_{i,t}\right) \tag{5}$$

that gives the utility of consuming $C^*_{i,s,t}$ this period and then following his optimal rule for $\tau > t$.

Once the agent solves the consumption allocation problem and gets the value associated with each s at time $t = 1$ $\left(V^*_{i,s,1}\right)$ he uses it to select a schooling level. At period $t = 0$ the agent selects the schooling level s that maximizes his expected utility net of "psychic" costs. He will attend college if

$$E\left(\mathcal{V}_{c,1}\left(\mathcal{I}_{i,1}\right) - \mathcal{V}_{h,1}\left(\mathcal{I}_{i,1}\right) - Cost_i \mid \mathcal{I}_{i,0}\right) > 0. \tag{6}$$

2.2 Specification of the Model

Earnings for individual i at time t at schooling level s are written as[9]

$$lnY_{i,s,t} = \mu_{s,t}\left(X_{i,s,t}\right) + U_{i,s,t} \tag{7}$$

where $X_{i,s,t}$ represents variables that the econometrician observes and $U_{i,s,t}$ variables he cannot observe. I assume that the agent knows all of the variables in X at all times and that $U_{i,s,t}$ is revealed to him at period t. He may also know all or part of each $(U_{i,s,\tau}, \tau = t+1, ..., T)$ at time t. Uncertainty is thus associated with $\{U_{i,s,\tau}\}_{\tau=t+1}^{T}$. $U_{i,s,t}$ may also include measurement error in earnings. If this is the case, any estimate of uncertainty will be an upper bound since a fraction of the variance in $U_{i,s,t}$ will be due to the measurement error.

I write psychic costs in the schooling choice equation (6) as a function of variables Z that are observed by both the analyst and the agent. ζ represents variables not observed by the econometrician and may be (partially) known to the agent at $t = 0$. That is

$$Cost_i = \phi\left(Z_i\right) + \zeta_i. \tag{8}$$

The utility function is of the CRRA form

$$u\left(C\right) = \frac{C^{1-\psi}}{1-\psi}, \tag{9}$$

where $\psi \geq 0$ is the coefficient of relative risk aversion. Since I do not model retirement explicitly I assume

[9] Although not pursued in this paper, the separability assumption is not essential and can be relaxed using the analysis of Matzkin (2003) to analyze functions of the form $lnY_{i,s,t} = \mu_{s,t}\left(X_{i,s,t}, U_{i,s,t}\right)$.

that in the terminal period (i.e. the period after age 65) the utility function is given by

$$\tilde{u}\left(A_{i,s,t}\right) = b\frac{\left(\epsilon A_{i,s,T}\right)^{1-\chi}}{1-\chi} \tag{10}$$

where b weights the utility in the terminal period, ϵ represents mortality risk and I let the function have a different parameter of risk aversion.

Finally, I assume that consumption is measured with error:

$$\widehat{C}_{i,t} = C_{i,t} q\left(K_{i,t}, \xi_{i,t}\right) \tag{11}$$

where $C_{i,t}$ is "true" consumption, $\widehat{C}_{i,t}$ is measured consumption and $q()$ is multiplicative measurement error function that depends on observable variables $K_{i,t}$ and unobservable variables $\xi_{i,t}$. Measurement error in consumption is a natural assumption in this context since consumption is measured at the household level from where it is imputed to the individual.[10] $K_{i,t}$ includes variables that control for the household structure, $\xi_{i,t}$ is an unobserved term assumed to capture the rest.

3 Inferring the Agent's Information Set

The econometrician must know $\mathcal{I}_{i,t}$ to solve the model to develop estimating equations. Any conclusion extracted from the model relies crucially on the assumptions made about what constitutes the uncertainty facing the agent. It is thus important to develop a procedure to allow the analyst to separate the components of the agent's information set from what is unknown to him. I now turn my attention to this topic, sketching identification of the model in the process. Appendix 1 provides formal proof of identification.

3.1 Testing for Information Misspecification

I cast the problem of determining agent information sets as a testing problem[11]. In this section I deal with the test of whether a candidate information set is correctly specified in a general setting[12]. I develop a simple test of misspecification for a proposed information set that does not directly depend on the model being used.

For a given proposed information set, $\tilde{\mathcal{I}}_{i,t}$, from the model developed above it follows that $\ln \hat{C}_{i,s,t} = \ln \mathcal{C}_{s,t}\left(\tilde{\mathcal{I}}_{i,t}\right) + \ln q\left(K_{i,t}, \xi_{i,t}\right)$. That is, measured consumption equals the one predicted by the model via the policy function $\mathcal{C}_{s,t}$ plus measurement error. Notice, however, that this holds true for a whole class of models besides the model proposed. For a nonparametric function $g()$ of the proposed information set it is true that

$$\ln \hat{C}_{i,s,t} = g\left(\tilde{\mathcal{I}}_{i,t}\right) + \ln q\left(K_{i,t}, \xi_{i,t}\right). \tag{12}$$

The prediction that consumption will be a function of the state variables of the model (*i.e.* the information

[10] Imputation is done by dividing total consumption over the square root of the total number of members in the household.

[11] See Cunha, Heckman, and Navarro, 2005 where a version of this test for a perfect credit markets model of schooling choice is proposed.

[12] The particular implementation of the test used in this paper is shown in section 3.3.

set) is independent, for example, of the particular form of the utility function or of the earnings equations[13]. It is in this sense that we can work with a nonparametric function of $\tilde{\mathcal{I}}_{i,t}$ with the benefit that the solution of the dynamic program does not need to be computed and that the test will still be valid for a class of models predicting that consumption can be written as in (12).

The test is simple: we want to estimate the model (either solving the dynamic problem for $\mathcal{C}_{s,t}$ in (4) or using a nonparametric function - polynomials on the elements of $\tilde{\mathcal{I}}_{i,t}$ for example) using a candidate information set $\widetilde{\mathcal{I}}_{i,t}$. Equation (12) forms part of the contribution of individual i who selects schooling level s to the sample likelihood. Let $\pi_{\tau,t}$ be an auxiliary parameter. To define the proposed test, instead of basing the likelihood on equation (12) use

$$\ln \widehat{C}_{i,s,t} = g_{s,t}\left(\widetilde{\mathcal{I}}_{i,t}\right) + \ln q\left(K_{i,t}, \xi_{i,t}\right) + \sum_{\tau=t+1}^{T}\left[Y_{s,\tau} - E\left(Y_{s,\tau}|\tilde{\mathcal{I}}_{i,t}\right)\right]\pi_{\tau,t}.$$

By assumption, the predicted consumption $g_{s,t}\left(\tilde{\mathcal{I}}_{i,t}\right)$ will not depend on the earnings innovations in the last term since the agent integrates them out. The actual consumption decision, however, will be a function of the true agent's information set at t which may contain elements of $\sum_{\tau=t+1}^{T}\left[Y_{s,\tau} - E\left(Y_{s,\tau}|\tilde{\mathcal{I}}_{i,t}\right)\right]\pi_{\tau,t}$.

A test of which of the auxiliary parameters multiplying the earnings innovations $\{\pi_{\tau,t}\}_{\tau=t+1}^{T}$ equals zero is then a test of whether the proposed agent's information set at time t is correctly specified.[14] Notice that, as done in Cunha, Heckman, and Navarro, 2005, the schooling choice (or any other choice) may also be used when testing.

Given the many ways one can propose information sets $\tilde{\mathcal{I}}_{i,t}$, especially the elements of the set unobserved to the econometrician, the test may seem as a formidable proposition. The next sections present assumptions to make the test operational. The main intuition, that under a correct specification of the information set the information innovations should not predict current choices, remains regardless of the implementation.

3.2 The Factor Structure and the Arrival of Information

In order to separate unobserved (to the econometrician) variability from the uncertainty facing agents, it is useful to assume that the unobservables for agent i can be factor analyzed in the following way:

$$\begin{aligned} U_{i,s,t} &= \theta_i \alpha_{s,t} + \varepsilon_{i,s,t} \\ \zeta_i &= \theta_i \lambda + \omega_i \end{aligned} \tag{13}$$

where θ_i is a vector of mean zero mutually independent "factors", $\varepsilon_{i,s,t}$ and ω_i are also mean zero random variables called "uniquenesses". Uniquenesses, factors and measurement error for consumption, $\xi_{i,t}$, are all assumed mutually independent of each other for all schooling levels s and time periods t.

The equations in (13) are only a statistical decomposition and, by themselves, are not informative

[13] The idea that we do not necessarily need to solve the whole dynamic program to estimate some functions is not new. See Hotz and Miller (1993), Park (2004) and Carranza (2007) for recent examples.

[14] And it can in fact be considered as a form of Sims' test of causality (Sims (1972)).

about what is known to the agent at period t. I interpret elements of θ_i as permanent shocks that hit and influence earnings at different points in time. They provide a useful device for extracting components of uncertainty from future outcomes.

Assume that the factor structure is such that earnings in the first τ_1 periods are affected only by the first element of θ_i (so the loadings $\alpha_{s,t,l}$ for $l > 1$ and $t > \tau_1$ would all be zero). The next τ_2 periods of earnings are affected by the first two elements of θ_i, the next τ_3 periods by the first three and so on. These elements of θ_i would then be revealed to the agent through their effect on earnings. However, the agent might be able to forecast elements of θ_i that affect future earnings but do not affect past and currently observed outcomes. Let $\theta_i(t)$ denote those elements of θ_i that affect earnings at or before t and let $\bar{\theta}_i(t)$ denote those elements of θ_i that affect earnings after t. Break $\bar{\theta}_i(t)$ into two components, $\left(\bar{\theta}_i^k(t), \bar{\theta}_i^u(t)\right)$ where $\bar{\theta}_i^k(t)$ is known by the agent so it is in $\mathcal{I}_{i,t}$ and $\bar{\theta}_i^u(t)$ is unknown by the agent so it is not in $\mathcal{I}_{i,t}$.

The following assumptions are made about the arrival of information

(I-1) The information revelation process of the agent is such that he either knows element l of $\theta_i - \theta_{i,l}$ $(l = 1, \ldots L)$ $-$ or he does not. Revelation of information when it happens, is instantaneous.

(I-2) At period t, the agent observes his outcomes for the period and so he knows $\{\varepsilon_{i,s,\tau}\}_{\tau=1}^{t}$ and the elements of $\theta_i(t)$, that is those elements of θ_i that affect outcomes in that period (or in any previous periods). If $\theta_{i,l}$ affects outcomes at $\tau \leq t$, then it is known by the agent at time t.

(I-3) Agents have rational expectations so that the expectations they take and the mathematical expectation operator with respect to the actual distributions in the model coincide.

The rest of the information structure of the model is assumed to be such that the agent has knowledge of the parameters of the model ($e.g.$, $\rho, \psi, \mu(X), \phi(Z)$) as well as of the observables[15] X_i, X_i^M, K_i, Z_i and the uniqueness in the cost function ω_i. The econometrician never observes θ_i. By assumption, $\{\varepsilon_{i,s,\tau}\}_{\tau=t+1}^{T}$ is not part of the agent's information set $\mathcal{I}_{i,t}$.

3.3 Determining the Information Set Under the Factor Structure

In this section I cast the problem of determining agent information sets of section (3.1) in terms of the factor structure. Using the assumptions just made,[16] I redefine the test of section (3.1) to test whether a candidate information set ($i.e.$ a specification of $\bar{\theta}_i^k(t)$ and $\bar{\theta}_i^u(t)$) is correctly specified.

The test proposed in section (3.1) consists of including the income innovations relative to a proposed information set $\tilde{\mathcal{I}}_{i,t}$ and test whether they affect current choices. Given the factor structure and information assumptions **(I-1) - (I-3)**, the only source of earnings innovations "knowable" to the agent is given by $\bar{\theta}_i(t)$ (i.e. the factors that affect earnings only in periods after t). Armed with this intuition it is possible then to design a simple version of the test.

Start by assuming that the agents do not know the different elements of θ_i until they learn them when they hit earnings. That is, the model is estimated using a candidate information set $\widetilde{\mathcal{I}}_{i,t}$ that contains no elements of $\bar{\theta}_i(t)$ before time t. To define the proposed test, instead of basing the likelihood on equation

[15] This assumption can be relaxed by modeling the stochastic process that generates these variables. For example, Keane and Wolpin (1997) have one X, experience in each sector, the accumulation of which they model.

[16] The assumption that θ is independent of X, Z is also imposed.

(12) use

$$\ln \widehat{C}_{i,s,t} = g_{s,t}\left(\widetilde{\mathcal{I}}_{i,t}\right) + \ln q\left(K_{i,t}, \xi_{i,t}\right) + \overline{\theta}_i\left(t\right)\pi_t.$$

By assumption, the predicted consumption $g_{s,t}\left(\tilde{\mathcal{I}}_{i,t}\right)$ will not depend on $\overline{\theta}_i\left(t\right)$ since the agent integrates them out. The actual consumption decision, however, will be a function of the true agent's information set at t which may contain elements of $\overline{\theta}_i\left(t\right)$. In this case, the different elements of $\theta_i \bar{(t)}$

A test of whether those elements of π_t associated with coordinate l of $\overline{\theta}_i\left(t\right)$ equal zero is then a test of whether element l of $\overline{\theta}_i\left(t\right)$ belongs in the agent's information set at time t. That is, if the l^{th} element of $\overline{\theta}_i\left(t\right)$ is actually part of the agent's information set it will affect the consumption decision and the elements of π_t associated with it will be estimated to be different from zero. Notice that, since we assumed that the agent's do not know $\overline{\theta}_i\left(t\right)$ at t, the first time element l appears is the moment in which it becomes known to the agent. For example, if factor 3 affects earnings at time 5 for the first time but it shows that it affects consumption decisions at time 2 (i.e. its associated π_2 is estimated to be different from zero) we conclude that the agent knows factor 3 at time 2 even though it does not affect earnings until 3 periods later.

There is nothing special about the consumption decision. Any other decision variable that depends on future outcomes can also be used. The basic idea is that if element l of $\overline{\theta}_i\left(t\right)$ belongs in the information set of the agent at time t and the agent acts on it, it will affect the choices he makes at t. In particular it will affect the consumption decision and we can test for it. The same idea can be applied to agent schooling choices. Since the individual college decision is described by $E\left(\mathcal{V}_{c,1}\left(\mathcal{I}_{i,1}\right) - \mathcal{V}_{h,1}\left(\mathcal{I}_{i,1}\right) - Cost_i \mid \mathcal{I}_{i,0}\right) > 0$, estimating the model under $\widetilde{\mathcal{I}}_{i,0}$ (with no elements of $\overline{\theta}_i\left(0\right)$ contained in it) and using[17]

$$E\left(\mathcal{V}_{c,1}\left(\tilde{\mathcal{I}}_{i,1}\right) - \mathcal{V}_{h,1}\left(\tilde{\mathcal{I}}_{i,1}\right) - Cost_i \mid \widetilde{\mathcal{I}}_{i,0}\right) + \overline{\theta}_i\left(0\right)\pi_0 > 0$$

as the schooling choice rule allows us to test for $\pi_0 = 0$ as a test for misspecification.

3.4 Identification of the model

Formal semiparametric identification analysis of the factor model of equations (1) - (13) is established in Appendix 1. This section provides an intuitive sketch of the identification arguments used in the Appendix for the factor structure of earnings, as well as a discussion of the standard arguments under which preferences and measurement error in consumption are identified[18].

In this paper identification theory is used to understand what in principle can be recovered nonparametrically. I then use flexible parametric forms to obtain estimates of my high dimensional econometric model. The question of identifiability is a separate issue and should be judged independently of the choice of parametric forms for estimation purposes[19].

Earnings are identified by adapting a version of the argument in Carneiro, Hansen, and Heckman

[17] Notice that, same as before, we can use a nonparametric function of $\tilde{\mathcal{I}}_{i,1}$ instead of the actual solution to the dynamic problem.

[18] Identification of the parameters of the "psychic" cost function is established in appendix 1. See also Heckman and Navarro (2007).

[19] See Roehrig (1988), Heckman (2005) for more on this distinction

(2003) which I now sketch for the case in which θ_i is a scalar. I assume that the problem of selection (*i.e.,* we only observe college earnings for college graduates and high school earnings for high school graduates) is solved using the arguments in Appendix 1 which involve using variation in the Z_i to achieve limit sets. Without loss of generality, take the system of log-earnings equations for high school $\ln Y_{i,h,t} = \mu_{h,t}(X_{i,h,t}) + \theta_i \alpha_{h,t} + \varepsilon_{i,h,t},\ t \geq 1$.

First, notice that the factor θ_i has no natural scale, *i.e.,* $\theta_i \alpha = \kappa \theta_i \frac{\alpha}{\kappa}$ for any constant κ, so it needs to be set by a normalization as does the sign of the factor loading. Normalizing one loading takes care of both problems. Suppose that we normalize the loading in the first period so that $\alpha_{h,1} = 1$.

Assuming that X is independent of the error terms $\{U_{i,h,t}\}_{t=1}^{T}$ form the covariance matrix of high school log-earnings from the data. Solving the system of equations that comes from equating the data (left hand side) to the theoretical covariance predicted by the factor structure we obtain the factor loadings on the first factor from

$$\frac{cov\left(\ln Y_{i,h,t}, \ln Y_{i,h,t'} \mid X\right)}{cov\left(\ln Y_{i,h,1}, \ln Y_{i,h,t'} \mid X\right)} = \frac{\alpha_{h,t}\alpha_{h,t'}\sigma_\theta^2}{\alpha_{h,t'}\sigma_\theta^2} = \alpha_{h,t}, \qquad t \neq t'.$$

Given the loadings for all time periods, we can then recover the variance of the factor from the equations

$$\sigma_\theta^2 = \frac{cov\left(\ln Y_{i,h,1}, \ln Y_{i,h,t} \mid X\right)}{\alpha_{h,t}}, \qquad t = 1, ...T.$$

The variances of the uniquenesses for all t are recovered from $var\left(\ln Y_{i,h,t} \mid X\right) - \alpha_{h,t}^2\sigma_\theta^2 = \sigma_{\varepsilon_{h,t}}^2$, where we know the left hand side from the data and the preceding argument. If both θ_i and $\{\varepsilon_{i,h,t}\}_{t=1}^{T}$ are normally distributed their distributions are identified by recovering their variances and the assumption that their means are zero. Normality is not required to identify the distributions. Using Theorem 5 in Appendix 1, the distributions of both θ_i and $\{\varepsilon_{i,h,t}\}_{t=1}^{T}$ can be nonparametrically identified. Notice that, while we can form covariances for high school earnings over time, we can never form covariances of earnings across schooling levels since earnings are not observed on both schooling levels for anyone.

Given the normalizations we just made to the high school system of earnings, making a similar set of normalizations to the college system would amount to setting the sign (and magnitude) of the unobserved covariance of earnings between college and high school earnings. To see this, notice that the unobserved covariance of earnings in high school and college in period 1 is

$$cov\left(\ln Y_{i,h,1}, \ln Y_{i,c,1} \mid X\right) = \alpha_{c,1}\sigma_\theta^2 \tag{14}$$

so setting $\alpha_{c,1} = 1$ would impose a strong restriction that the covariance is positive and fixed by the variance determined in the high school system. Theorem 5 in the Appendix shows that restrictions of this nature do not need to be imposed. The system for college earnings is identified without additional normalizations if the distribution of θ_i is nonsymmetric.

If θ_i has a symmetric distribution suppose that, beside the data on earnings, consumption and schooling choices required for the rest of the model, a system of measurements that are not schooling dependent is available:

$$M_{i,j} = \mu_j^M \left(X_{i,j}^M\right) + \theta_i \alpha_j^M + \varepsilon_{i,j}^M, \qquad j = 1, ...J, \tag{15}$$

where $\varepsilon_i^M = \left(\varepsilon_{i,1}^M, .., \varepsilon_{i,J}^M\right)$ is a vector of mean zero mutually independent random variables. ε_i^M is assumed to be independent of θ_i and of $\varepsilon_{i,s,t}$ and ω_i for all t and s.

The measurement system provides an alternative way to identify the factor loadings and uniquenesses for college earnings without imposing further restrictions. To see why, notice that the loadings on the measurement system are identified from

$$\alpha_j^M = \frac{cov\left(\ln Y_{i,h,1}, M_{i,j} \mid X, X^M\right)}{\sigma_\theta^2}, \qquad j = 1, ..., J.$$

Take the covariance of college earnings with respect to a measurement equation: $cov\left(\ln Y_{i,c,t}, M_{i,j} \mid X, X^M\right) = \alpha_{c,t}\alpha_j^M \sigma_\theta^2$. Then, the loadings on the first factor for the college system are identified. By repeatedly applying Theorem 5 the distributions of $\left\{\varepsilon_{i,j}^M\right\}_{j=1}^J$ and $\{\varepsilon_{i,c,t}\}_{t=1}^T$ can also be nonparametrically identified. The argument can be extended to the multifactor case using the results in Appendix 1.

The risk aversion parameters, the discount factors and the measurement error in consumption are identified using relatively standard Euler equation arguments.[20] Since the arguments are well known, what follows is a simple sketch of how this is done assuming that we can write the measurement error in consumption as $q\left(K_{i,t}, \xi_{i,t}\right) = e^{K_{i,t}\delta + \xi_{i,t}}$ and $K_{i,t}, K_{i,t+1}$ are contained in $\mathcal{I}_{i,t}$.

Using the first order condition of equation (5) and using equations (9) and (11) it follows that

$$E\left(\frac{1+r}{1+\rho}\left(\frac{\widehat{C}_{i,t+1} q\left(K_{i,t}, \xi_{i,t}\right)}{\widehat{C}_{i,t} q\left(K_{i,t+1}, \xi_{i,t+1}\right)}\right)^{-\psi} - 1 \mid \mathcal{I}_{i,t}\right) = 0. \tag{16}$$

which is a standard consumption Euler equation except that it contains measurement error and so the standard argument of Hansen and Singleton (1983) cannot be applied directly. Instead, as noted by[21] Chioda (2004) one can take differences of two adjacent Euler equations to form a valid moment condition and identify ψ. With ψ in hand rewrite equation (16) as

$$\frac{1+r}{1+\rho}\left(\frac{\widehat{C}_{i,t+1} q\left(K_{i,t}, \xi_{i,t}\right)}{\widehat{C}_{i,t} q\left(K_{i,t+1}, \xi_{i,t+1}\right)}\right)^{-\psi} = \eta_{i,t} + 1$$

where $\eta_{i,t}$ is expectational error which is a function (among other things) of the elements of θ_i not contained in $\mathcal{I}_{i,t}$. Taking logs and a linear approximation of $\eta_{i,t} + 1$ around $\eta_{i,t} = 0$ we obtain

$$\ln \frac{\hat{C}_{i,t+1}}{\hat{C}_{i,t}} = \frac{1}{\psi} \ln\left(\frac{1+r}{1+\rho}\right) + \left(K_{i,t+1} - K_{i,t}\right)\delta + \left(\xi_{i,t+1} - \xi_{i,t} - \frac{\eta_{i,t}}{\psi}\right) \tag{17}$$

From the fact that the interest rate r is given it follows that we can identify the discount factor ρ and the

[20] See Hansen and Singleton (1983), Browning and Lusardi (1996) and Attanasio and Low (2004).
[21] See also Ventura (1994) for the parametric case.

observable effect of measurement error δ.

We next proceed to look at the Euler equation in the terminal period

$$-\left(\hat{C}_{i,T} q\left(K_{i,T}, \xi_{i,T}\right)^{-1}\right)^{-\psi} + \frac{B}{1+\rho} A_{i,T}^{-\chi} = 0 \tag{18}$$

where $B = bE\left(\epsilon^{-\chi} | \mathcal{I}_{i,T}\right)$ and there is no expectation around A_T since it is known at time T. For identification purposes it is helpful to rewrite equation (18) as

$$\ln \hat{C}_{i,T} = -\frac{1}{\psi} \ln\left(\frac{B}{1+\rho}\right) + \frac{\chi}{\psi} \ln\left(A_{i,T}\right) - K_{i,T}\delta - \xi_{i,T}. \tag{19}$$

Since ψ, and ρ are known from the argument above, B is identified and so is χ and the distribution of $\xi_{i,T}$.

Once identification of the risk aversion parameters ψ, χ and the discount factor ρ is secured, all the elements required to solve the consumption allocation problem of equation (1) are in place. To show that the distribution of the unobserved part of measurement error is identified remember that the assumption that $\xi_{i,t}$ is independent of θ_i and of $K_{i,t}$ is imposed. Then, since the left hand side of $\ln \widehat{C}_{i,t} - K_{i,t}\delta = \ln \mathcal{C}_t\left(\mathcal{I}_{i,t}\right) + \xi_{i,t}$ is known and so is the distribution of $\ln \mathcal{C}_t\left(\mathcal{I}_{i,t}\right)$ we can recover the distribution of $\xi_{i,t}$ by deconvolution.

4 Empirical Results

The model in this paper is estimated on a sample of white males who either graduated high school (and only high school) or are college graduates. Since no single dataset contains all the information required by my analysis, the sample used contains individuals from both the NLSY79 and PSID datasets pooled together. This requires a separate analysis of its own where the information not available in one dataset is integrated out against the appropriate distribution. In principle, this may invalidate identification theorems that assume access to all the required information. As shown in Appendix 2, where the procedure used to pool datasets is described, this is not the case and the missing information can be integrated out in a relatively straightforward fashion.

Table 1 presents a summary of the pooled dataset used to estimate the model.[22] The sample consists of a total of 2,986 white males born between 1923 and 1975, 1,097 from NLSY79 and 1,889 from PSID. In general, college graduates have higher present value of earnings, consumption and assets than high school graduates. They also have higher test scores, come from better family backgrounds, have fewer siblings and are more likely to live in a location where college tuition is lower and to have grown up in an urban area.

As a system of external measurements on ability I use five components from the ASVAB battery of tests. ASVAB tests are only available for the NLSY79 sample so the assumption that their distribution is the same in the PSID sample is imposed when integrating them out.[23] Each test is modeled as a function

[22] A complete description of the dataset is presented in Appendix 3 where it is shown that, in general, both datasets are roughly comparable in the years in which they overlap both in terms of earnings and of the covariates used in the analysis.

[23] See also Hansen, Heckman, and Mullen (2004) for an analysis of ASVAB tests and their relation to ability.

Table 1

Descriptive statistics of white male high school and college graduates from NLSY79 and PSID

Variable Name	Full Sample					High School					College				
	Obs	Mean	Std Dev	Min	Max	Obs	Mean	Std Dev	Min	Max	Obs	Mean	Std Dev	Min	Max
PV of Earnings 18-65*	5859	330.86	334.83	0.01	6327.05	2948	253.44	170.04	0.01	4897.43	2911	409.27	429.16	0.36	6327.05
South at age 14	17916	0.24	0.43	0.00	1.00	9150	0.25	0.43	0.00	1.00	8766	0.23	0.42	0.00	1.00
Urban at age 14	17916	0.78	0.42	0.00	1.00	9150	0.72	0.45	0.00	1.00	8766	0.83	0.38	0.00	1.00
Mother's Education	17916	4.20	1.59	1.00	9.00	9150	3.76	1.39	1.00	9.00	8766	4.66	1.67	1.00	9.00
Year of Birth	17916	52.49	11.18	23.00	75.00	9150	53.13	11.26	23.00	75.00	8766	51.82	11.06	23.00	67.00
PSID	17916	0.63	0.48	0.00	1.00	9150	0.61	0.49	0.00	1.00	8766	0.65	0.48	0.00	1.00
Schooling	17916	0.49	0.50	0.00	1.00	9150	0.00	0.00	0.00	0.00	8766	1.00	0.00	1.00	1.00
Number of Siblings	17916	3.05	2.17	0.00	18.00	9150	3.37	2.31	0.00	18.00	8766	2.71	1.96	0.00	15.00
Local Tuition at age 18*	17916	1.78	0.70	0.00	5.55	9150	1.83	0.73	0.00	5.55	8766	1.73	0.67	0.15	5.31
Local Tuition at age 19*	17916	1.78	0.68	0.00	5.67	9150	1.83	0.71	0.00	5.41	8766	1.73	0.65	0.16	5.67
Local Tuition at age 20*	17916	1.78	0.66	0.00	6.47	9150	1.82	0.69	0.00	5.41	8766	1.75	0.63	0.28	6.47
Local Tuition at age 21*	17916	1.80	0.67	0.00	6.66	9150	1.84	0.71	0.00	5.57	8766	1.76	0.64	0.27	6.66
Local Unemp at age 17: High School	17916	0.05	0.03	0.00	0.54	9150	0.05	0.03	0.00	0.54	8766	0.05	0.03	0.00	0.30
Local Wage at age 17: High School	17916	8.38	6.51	0.04	20.08	9150	9.04	6.34	0.04	20.08	8766	7.69	6.62	0.04	17.40
Local Unemp at age 17: College	17916	0.02	0.02	0.00	0.35	9150	0.02	0.02	0.00	0.35	8766	0.02	0.02	0.00	0.33
Local Wage at age 17: College	17916	12.51	9.73	0.07	28.15	9150	13.43	9.41	0.07	28.15	8766	11.55	9.96	0.08	26.99
ASVAB:															
Arithmetic Reasoning	6372	0.03	0.98	-2.63	1.31	3396	-0.46	0.92	-2.63	1.31	2976	0.58	0.72	-2.04	1.31
Paragraph Composition	6372	0.02	0.97	-4.08	1.09	3396	-0.43	1.04	-4.08	1.09	2976	0.52	0.54	-1.73	1.09
Word Knowledge	6372	0.03	0.98	-3.79	1.16	3396	-0.42	1.07	-3.79	1.16	2976	0.54	0.52	-1.81	1.16
Math Knowledge	6372	0.00	0.98	-3.05	2.65	3396	-0.35	0.93	-3.05	1.82	2976	0.40	0.86	-2.14	2.65
Coding Speed	6372	0.03	0.99	-2.47	1.43	3396	-0.59	0.78	-2.47	1.43	2976	0.73	0.68	-1.85	1.43
Grade Completed: 1980	6444	11.36	1.76	7.00	17.00	3414	10.95	1.29	7.00	12.00	3030	11.82	2.07	8.00	17.00
Enrolled: 1980	6444	0.68	0.47	0.00	1.00	3414	0.48	0.50	0.00	1.00	3030	0.91	0.28	0.00	1.00
Age: 1980	6582	19.18	2.17	16.00	23.00	3534	19.21	2.16	16.00	23.00	3048	19.13	2.19	16.00	23.00
Consumption*	1561	316.31	232.36	3.15	2215.00	783	251.04	142.86	3.15	1806.92	778	382.00	281.51	6.10	2215.00
Married	10302	0.70	0.42	0.00	1.00	5188	0.70	0.41	0.00	1.00	5114	0.69	0.42	0.00	1.00
Number of Children	10094	0.96	1.13	0.00	9.57	5101	1.01	1.15	0.00	8.00	4993	0.91	1.10	0.00	9.57
Age of Youngest Child	10033	2.92	4.00	0.00	20.00	5072	3.09	4.04	0.00	20.00	4961	2.74	3.95	0.00	19.67
Assets*	5113	117.06	347.73	0.00	9428.48	2596	68.07	150.06	0.00	3058.04	2517	167.58	466.29	0.00	9428.48

* Thousands of Dollars

of only the first factor

$$M_{i,j} = \beta_{0,j} + X_i^M \beta^M + \theta_{i,1} \alpha_j^M + \varepsilon_{i,j}^M,$$

where, to pin down the scale of $\theta_{i,1}$, the loading on the first test equation $\left(\alpha_1^M\right)$ is normalized to 1. This normalization associates higher levels of factor 1 with higher test scores, purged of the effect of family background and individual characteristics at test date, so I interpret $\theta_{i,1}$ as ability. In this interpretation, tests are assumed to be noisy proxies for ability which is given by $\theta_{i,1}$.

I assume that unemployed people have earnings of zero. Earnings for individuals who are missing are imputed only in the years in which there was no survey using an average of the earnings in the years immediately adjacent (i.e. the year before and the year after) to the missing year. If earnings are not available in either of these years they are left as missing. Missings are treated as random events.[24] Individual earnings life cycles are then simplified to six 8 year long periods.[25] This simplifying assumption

[24] See Fitzgerald, Gottschalk, and Moffitt (1998) for evidence that observed people in PSID have similar characteristics as those in the CPS so attrition is roughly random. See MaCurdy, Mroz, and Gritz (1998) for evidence that attrition is random in the NLSY79.

[25] Period 1 covers ages 18 to 25, period 2 ages 26 to 33, period 3 ages 34 to 41, period 4 ages 42 to 49, period 5 ages 50 to 57 and period 6 ages 58 to 65. There is an additional terminal period in the model which includes whatever happens after age 65. Utility in this terminal period is modeled as depending only on assets carried to that period so no additional data for the period is required.

is used to keep the computational complexity of the model manageable[26]. For each period, logearnings are calculated as the log of the present value of earnings for the period discounted at 3%.

For each schooling level $s = \{h, c\}$ and for each period of earnings $t = \{1, ..., 6\}$ I find that modeling $\ln Y_{i,s,t}$ as being generated by a three factor model:

$$lnY_{i,s,t} = X_i\beta_{s,t} + \theta_{i,1}\alpha_{s,t,1} + \theta_{i,2}\alpha_{s,t,2} + \theta_{i,3}\alpha_{s,t,3} + \varepsilon_{i,s,t}.$$

is enough to fit the data (see tests in section 4.1). To pin down the scale of $\theta_{i,2}$ its loading for the second period of high school earnings ($\alpha_{h,2,2}$) is normalized to one. Similarly for $\theta_{i,3}$ its loading for the fourth period of high school earnings ($\alpha_{h,4,3}$) is normalized to 1. To keep with the triangular restrictions needed to identify the factor structure I impose that the second factor hits earnings for the first time in the second period (so $\alpha_{s,1,2} = 0$ for $s = \{h, c\}$) and that the third factor hits earnings for the first time in period 4 (so $\alpha_{s,t,3} = 0$ for $s = \{h, c\}$, $t < 4$).

Table 2

List of Covariates

Variable Name	Earnings	Consumption	Test System	Cost Function
Period Dummies	Yes	Yes	No	No
Married	No	Yes	No	No
Number of Children	No	Yes	No	No
Age of Youngest Child	No	Yes	No	No
South at age 14	No	No	Yes	Yes
Urban at age 14	No	No	Yes	Yes
Mother's Education	No	No	Yes	Yes
Year of Birth	No	No	Yes	Yes
Grade Completed: 1980	No	No	Yes	No
Enrolled: 1980	No	No	Yes	No
Age: 1980	No	No	Yes	No
Number of Siblings	No	No	Yes	Yes
PSID	No	No	No	Yes
Local Unemp at age 17: High School	No	No	No	Yes
Local Wage at age 17: High School	No	No	No	Yes
Local Unemp at age 17: College	No	No	No	Yes
Local Wage at age 17: College	No	No	No	Yes

The cost function is also allowed to (in principle, since only factors in the agent's information set at the schooling decision age affect it) be a function of all three factors:

$$Cost_i = Z_i\gamma + \theta_{i,1}\lambda_1 + \theta_{i,2}\lambda_2 + \theta_{i,3}\lambda_3 + \omega_i.$$

The Z_i include variables that only affect the schooling decision like family background, local wages, local unemployment, etc. Table 2 shows the full set of covariates used for tests, logearnings, consumption and costs.

Each of the factors $\theta_{i,l}$ is allowed to follow a mixture of normals distribution $\theta_{i,l} \sim \sum_{j=1}^{J_l} \pi_{l,j} f\left(\theta_{i,l}; \mu_{l,j}, \sigma^2_{l,j}\right)$ where $f\left(x; \mu, \sigma^2\right)$ is a normal density with mean μ and variance σ^2. The uniquenesses $\varepsilon_{i,s,t}$ as well as the measurement error in consumption $\xi_{i,s,t}$ are also allowed to be distributed as mixtures of normals.[27] In

[26] Estimation of the model requires the solution of the dynamic programming problem every time a different parameter vector is tried. Furthermore, the evaluation of the likelihood itself requires calculating a multidimensional integral for each individual in the sample.

[27] Models with normal distributions were tested and failed to fit the data as well as the mixture case.

all cases mixtures with 2 elements are found to be adequate. The remaining distributions in the model are assumed to be normal.

The borrowing constraint is imposed so that it matches the lower limit of the assets distribution in the data every period. That is, in any given period $A^{MIN}_{i,s,t}$ in equation (3) is set equal to the lowest level of assets observed in the period. This automatically defines the $Y^{MIN}_{i,s,t}$ term in the same equation.

Estimation of the model is done by maximum likelihood using a combination of simulated annealing, the Nelder-Meade simplex method and the Broyden-Fletcher-Goldfarb-Shanno variable metric algorithm to maximize the likelihood. The contribution of individual i who chooses schooling $S_i = s$

$$\int_\Theta \left[\begin{array}{c} \prod_{t=1}^5 f_{\varepsilon_{i,s,t}} \left(\ln Y_{i,s,t} - \mu_{s,t}\left(X_{i,s,t}\right) - \theta_i \alpha_{s,t} | \theta_i, X_{i,s,t}\right) f_{\varepsilon^M_{i,j}} \left(lnM_{i,j} - \mu^M_j \left(X^M_{i,j}\right) - \theta_i \alpha^M_j | \theta_i, X^M_{i,j}\right) \\ \prod_{t=1}^4 f_{\xi_{i,t}} \left(\ln \widehat{C}_{i,t} - \ln \mathcal{C}_{s,t}\left(\mathcal{I}_{i,t}\right) - K_{i,t}\pi | \theta_i, X_{i,s,t}, Z_i, K_{i,t}\right) Pr\left(S_i = s | X_{i,s,t}, Z_i, \theta_i\right) \end{array} \right] dF\left(\theta\right).$$

Evaluation of the likelihood requires that the econometrician solve the dynamic program (for a given proposed $\mathcal{I}_{i,t}$) in order to evaluate the schooling selection probability and the consumption policy function $\mathcal{C}_{s,t}$. Since the econometrician never observes any element of θ_i, he has to integrate against its distribution when evaluating the likelihood. In the model I estimate, the solution to the consumption allocation problem is approximated numerically using a piecewise linear approximant.

Table 3

Test for Information Set Misspecification

Additional Parameters	Schooling Choice	Consumption		
		Age 18 - 25	Age 26 - 33	Age 34 - 41
θ_1	*1.73*	-	-	-
Std. Error	*0.18*	-	-	-
θ_2	*-0.20*	*0.21*	-	-
Std. Error	*0.10*	*0.10*	-	-
θ_3	-0.51	-0.14	*-7.40*	-1.13
Std. Error	0.64	0.52	*3.69*	1.52

Let g($\mathcal{I}$) be the predicted choice as a function of the information set $\mathcal{I}$. The left out factors are added to the chocie function (after the agent integrates them out) and

4.1 Test of Misspecification

Table 3 presents the results of the proposed test of misspecification of the agent's information set using the auxiliary parameters, π, defined in section 3.3 for the elements of $\bar{\theta}_i(t)$. In terms of the empirical model described above, this entails testing whether $\theta_{i,3}$ affects the agent's choices before period 4 (when it hits earnings), $\theta_{i,2}$ before period 2 and whether $\theta_{i,1}$ affects schooling. As shown in the table, the hypothesis that the π associated with $\theta_{i,3}$ in the schooling equation is zero cannot be rejected at a 95% confidence, but we can reject it for the π associated with $\theta_{i,1}$ and $\theta_{i,2}$. That is, the test results imply that $\theta_{i,3}$ is not known at the time schooling decisions are made but $\theta_{i,1}$ and $\theta_{i,2}$ are known and so belong in the information set of the agent. The next column presents the test for the first period consumption decision. What we find is that $\theta_{i,2}$ belongs in the information set, supporting our finding from the previous line but that $\theta_{i,3}$ is not known in period 1. The third column shows that we reject the hypothesis that $\theta_{i,3}$ is not known at period 2.

In short, the results of the test show that both $\theta_{i,1}$ and $\theta_{i,2}$ affect schooling choices so $\theta_{i,2}$ is known before it hits earnings at period 2. They also show that $\theta_{i,3}$ becomes known in period 2 (before it hits earnings in period 4). As a consequence, all of the results presented in the next sections are based on estimates of the structural model using this information set. Tables A-2 to A-7 in the appendix present the parameter estimates for this model. In total the model has 163 parameters.

4.2 Model Fit to the Data

To validate the model estimates obtained under the information set chosen as a consequence of the test in the previous section a variety of checks of fit of predictions of the model versus their data counterparts are performed. First, the proportion of people who attend college in the data and the one predicted by the model are compared. Whereas 48.93% of the people in the sample are college graduates, the model predicts roughly 48.97% slightly above the actual number. When formal tests are performed, the null hypothesis of equality of predicted and actual proportions cannot be rejected.

Figures 1.1 through 1.6 present each period's distribution of logearnings in the data and the one predicted by the model for the overall sample. The model fits the data remarkably well not only for the overall sample but for the distributions conditional on choice too.[28]

Table 4

χ^2 goodness of fit test of equality between fitted and actual distributions

P-values

	Logearnings						Logconsumption	
	High School		College		Overall			
	T1	T2	T1	T2	T1	T2	T1	T2
Age 18-25	0.3392	0.2274	0.4369	0.1366	0.8072	0.7355	0.0066	0.1369
Age 25-33	0.1065	0.0775	0.0009	<0.0001	0.0680	0.0001	0.6597	0.1814
Age 34-41	0.7862	0.3370	0.1189	<0.0001	0.1023	<0.0001	0.0559	0.0507
Age 42-49	0.0522	0.0210	0.3326	0.2280	0.1644	0.1592	<0.0001	0.0108
Age 50-57	0.3697	0.0526	0.0001	0.0593	0.0546	0.0908	0.8034	0.7413
Age 58-65	0.1466	0.0709	0.4401	0.4215	0.0815	0.0005	0.0918	0.1575

*Equiprobable bins with aprox. 6 to 8 people per bin. P-values>0.05: cannot reject equality of distributions at 95% confidence. T1 uses data probabilities as denominator. T2 uses predicted probabilities as denominator.

Formal tests of equality of the data and fitted distributions are shown in the first six columns of table 4 for both when the χ^2 statistic is formed dividing over the probability in the data and when it is formed dividing over the predicted probability. For sixteen (or thirteen depending on the statistic used) out of the eighteen distributions analyzed the null hypothesis of equality cannot be rejected. Even in those cases in which equality is rejected, the predicted and actual distributions are actually very similar.

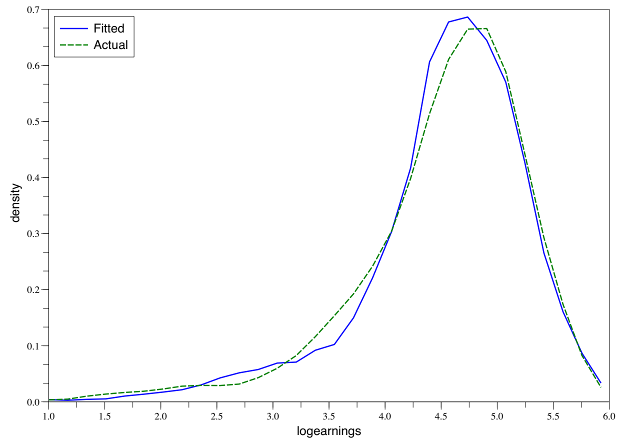
I also look at the fit of consumption in the data versus consumption predicted by the model.[29] Formal tests of equality are shown in the last two columns of table 4. As before, the fit is good. In 4 (or 5

[28] Fit graphs for the the high school and college samples are available at http://www.ssc.wisc.edu/~snavarro/reconsidering_graphs.htm.

[29] Fit graphs for consumption are available at http://www.ssc.wisc.edu/~snavarro/reconsidering_graphs.htm.

Figure 1.1

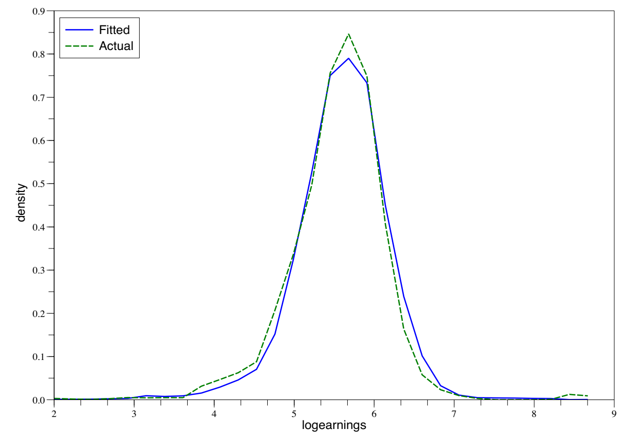
Densities of fitted and actual log present value of earnings for period 1 for overall sample



Log of present value of earnings from age 18 to 25 using an interest rate of 3%. Let (Y_0,Y_1) denote potential earnings in high school and college respectively. Let S=0 denote choice of high school and choice of college. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Let f(log(y)) denote the density function of observed earnings. The graph plots the density function f generated from the data (dashed line) against that predicted by the model (solid line). The plots are smoothed using a Gaussian kernel.

Figure 1.2

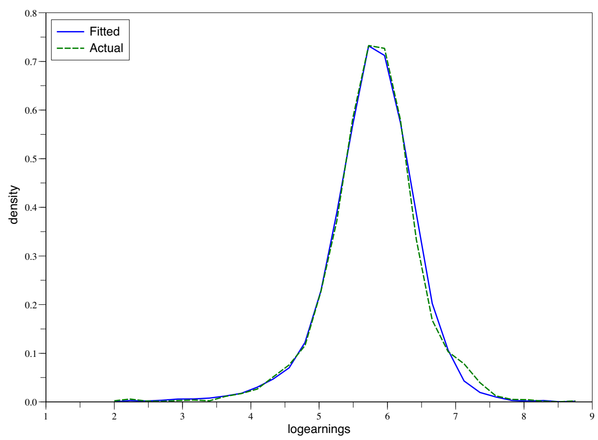
Densities of fitted and actual log present value of earnings for period 2 for overall sample



Log of present value of earnings from age 26 to 33 using an interest rate of 3%. Let (Y_0,Y_1) denote potential earnings in high school and college respectively. Let S=0 denote choice of high school and choice of college. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Let f(log(y)) denote the density function of observed earnings. The graph plots the density function f generated from the data (dashed line) against that predicted by the model (solid line). The plots are smoothed using a Gaussian kernel.

Figure 1.3

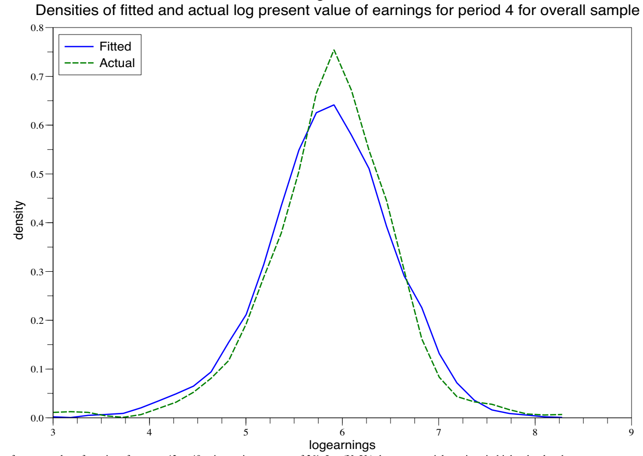
Densities of fitted and actual log present value of earnings for period 3 for overall sample



Log of present value of earnings from age 34 to 41 using an interest rate of 3%. Let (Y_0,Y_1) denote potential earnings in high school and college respectively. Let S=0 denote choice of high school and choice of college. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Let f(log(y)) denote the density function of observed earnings. The graph plots the density function f generated from the data (dashed line) against that predicted by the model (solid line). The plots are smoothed using a Gaussian kernel.

Figure 1.4

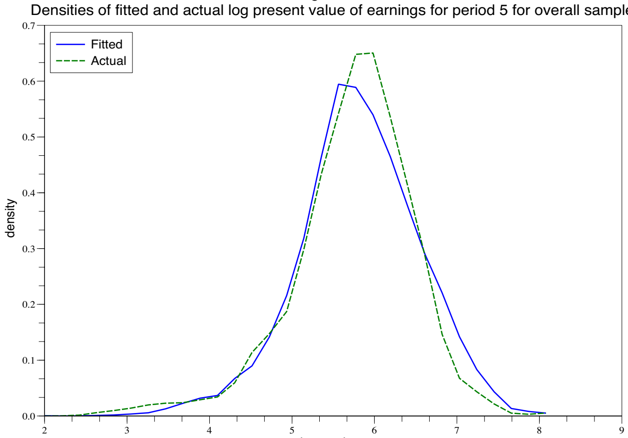
Densities of fitted and actual log present value of earnings for period 4 for overall sample



Log of present value of earnings from age 42 to 49 using an interest rate of 3%. Let (Y_0,Y_1) denote potential earnings in high school and college respectively. Let S=0 denote choice of high school and choice of college. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Let f(log(y)) denote the density function of observed earnings. The graph plots the density function f generated from the data (dashed line) against that predicted by the model (solid line). The plots are smoothed using a Gaussian kernel.

Figure 1.5

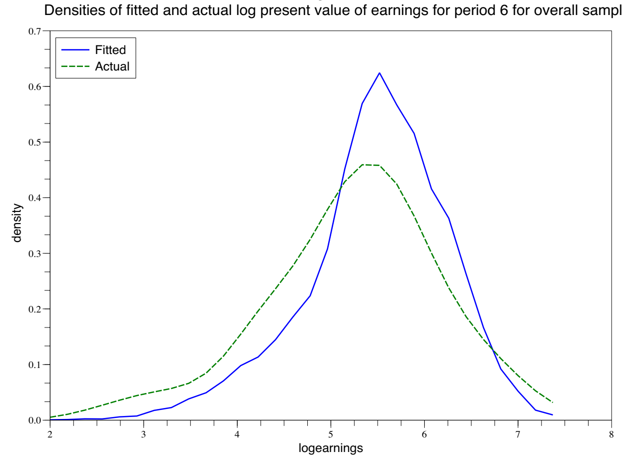
Densities of fitted and actual log present value of earnings for period 5 for overall sample



Log of present value of earnings from age 50 to 57 using an interest rate of 3%. Let (Y_0,Y_1) denote potential earnings in high school and college respectively. Let S=0 denote choice of high school and choice of college. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Let f(log(y)) denote the density function of observed earnings. The graph plots the density function f generated from the data (dashed line) against that predicted by the model (solid line). The plots are smoothed using a Gaussian kernel.

Figure 1.6

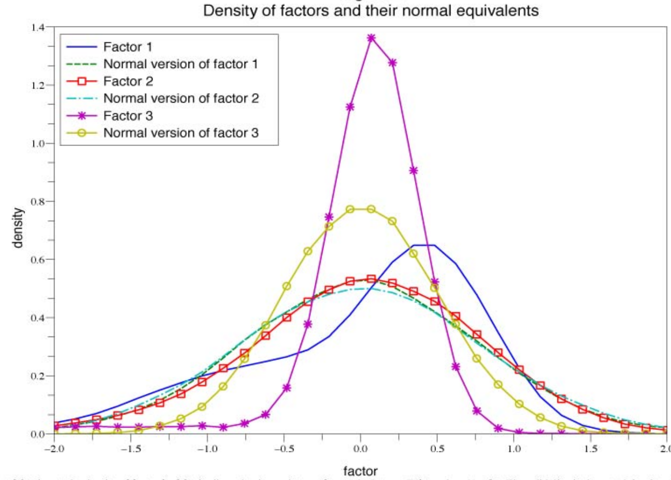
Densities of fitted and actual log present value of earnings for period 6 for overall sample



Log of present value of earnings from age 58 to 65 using an interest rate of 3%. Let (Y_0,Y_1) denote potential earnings in high school and college respectively. Let S=0 denote choice of high school and choice of college. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Let f(log(y)) denote the density function of observed earnings. The graph plots the density function f generated from the data (dashed line) against that predicted by the model (solid line). The plots are smoothed using a Gaussian kernel.

depending on the form of the statistic one looks at) out of the 6 periods the null hypothesis of equality cannot be rejected.

Figure 2
Density of factors and their normal equivalents

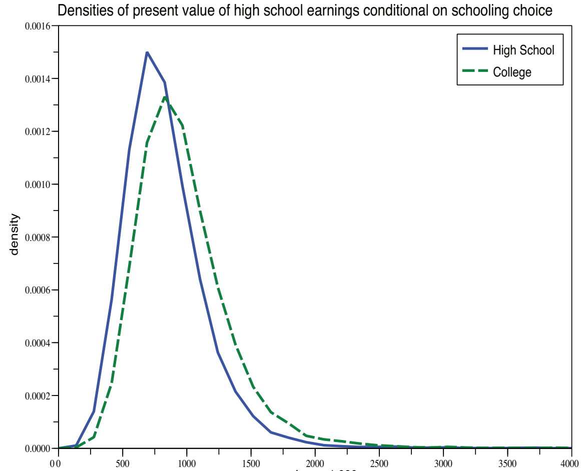


Let $f(\theta_i)$ denote the density of factor θ_i. $f(\theta_i)$ is allowed to be a mixture of normals. Let $m_i=E(\theta_i)$ and $v_i=Var(\theta_i)$. The solid line is the actual density of factor 1 while the dashed line is the density of a normal random variable with mean m_i and variance v_i. We proceed similarly for the other factors. The plots are smoothed using a Gaussian kernel.

Finally, in figure 2 I present evidence of the importance of allowing for the possibility of nonnormality on the factors. By comparing each factor with their normal equivalent, that is with a normal random variable with mean zero and variance equal to the estimated variance of the factor, it is clear that $\theta_{i,1}$ and $\theta_{i,3}$ are both highly nonnormal. Allowing for nonnormality is required to fit the data. This constitutes an extension to standard factor models which are based on normality assumptions.

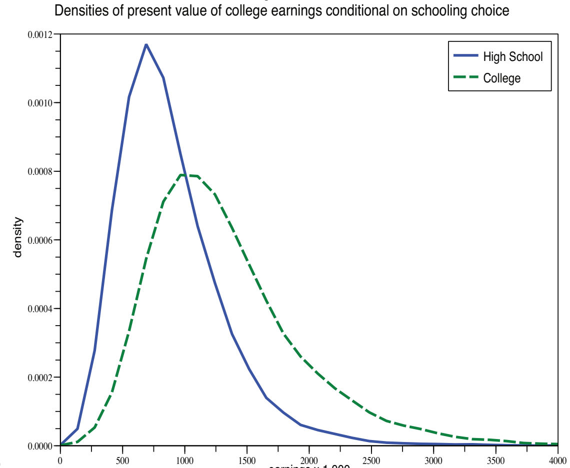
4.3 Counterfactual Analysis, Variability and Uncertainty

Figure 3.1
Densities of present value of high school earnings conditional on schooling choice



Log of present value of high school earnings from age 18 to 65 using an interest rate of 3%. Let Y_0 denote the present value of high school earnings for the period. The graph plots the density function $f(y_0|\text{choice=high school})$ (solid line) against $f(y_0|\text{choice=college})$ (dashed line). The plots are smoothed using a Gaussian kernel.

Figure 3.2
Densities of present value of college earnings conditional on schooling choice

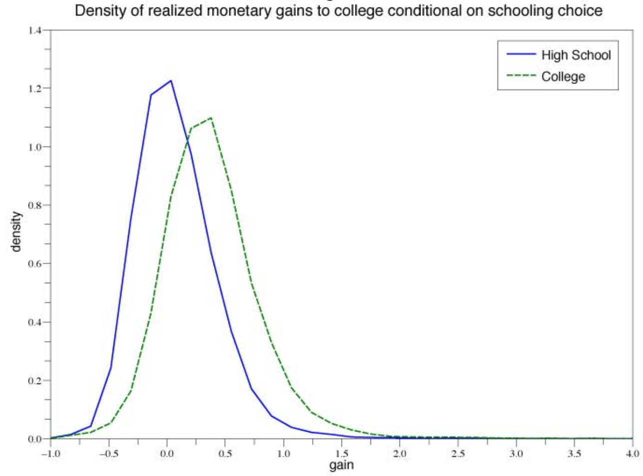


Log of present value of college earnings from age 18 to 65 using an interest rate of 3%. Let Y_1 denote the present value of college earnings for the period. The graph plots the density function $f(y_1|\text{choice=high school})$ (solid line) against $f(y_1|\text{choice=college})$ (dashed line). The plots are smoothed using a Gaussian kernel.

Figures 3.1 and 3.2 compare fitted and counterfactual distributions of earnings for each schooling level.

The figures show that people who actually graduate college would make more money than high school graduates in either sector. The difference, however, is much larger in the college sector so we conclude that comparative advantage is at work. That is, even though college graduates would make more money than high school graduates in either schooling level, they make comparatively more in the college sector so they tend to go to college. This, however, is not the whole story. In figure 3.3 I plot the percentage increase in lifetime earnings (*i.e.,* the ex-post realized gain) implied by the previous figures. Two features are worth noting, 1) a large proportion of high school graduates would have obtained positive gains if they had gone to college; and 2) a considerable fraction of college graduates get an ex-post negative gain.

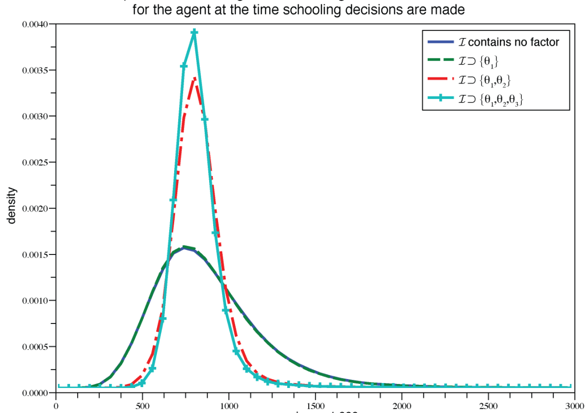
Figure 3.3
Density of realized monetary gains to college conditional on schooling choice



Let (Y_0, Y_1) denote the potential present value of earnings from age 18 to 65 (discounted using an interest rate of 3%) in the high school and college sectors, respectively. Define the realized monetary gain to college as $G=(Y_1-Y_0)/Y_0$. The solid line is the density of gains for people who choose high school, that is, f(g|choice=high school) and the dashed line f(g|choice=college). The plots are smoothed using a Gaussian kernel.

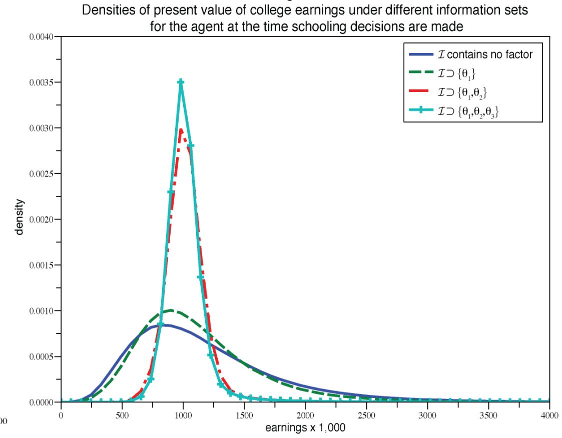
If people based their decisions only on monetary gains, and observed gains were the appropriate number to look at when explaining schooling decisions, one would be hard pressed to find an explanation to the significant proportion of people who attend college who actually obtain a negative ex-post monetary gain. A major advantage of writing a general model of schooling decision under uncertainty is that it allows me to distinguish between expected (by the agent) and observed (by the econometrician) outcomes.

Figure 4.1
Densities of present value of high school earnings under different information sets
for the agent at the time schooling decisions are made



Let Y_0 denote the present value of high school earnings from age 18 to 65 discounted at a 3% interest rate. Let $\mathcal{I}$ denote the agent's information set and $f(y_0|\mathcal{I})$ denote the density of the present value of earnings conditional on the information set $\mathcal{I}$. The graph shows $f(y_0|\mathcal{I})$ with $\mathcal{I}$ containing no factors, factor 1, factors 1 and 2, and all three factors. The plots are smoothed using a Gaussian kernel.

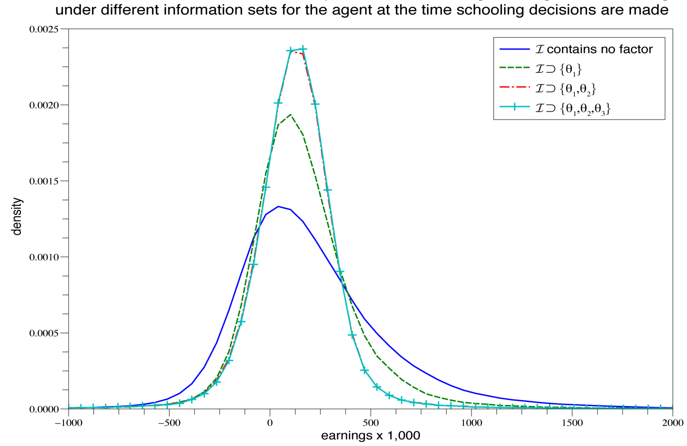
Figure 4.2
Densities of present value of college earnings under different information sets
for the agent at the time schooling decisions are made



Let Y_1 denote the present value of high school earnings from age 18 to 65 discounted at a 3% interest rate. Let $\mathcal{I}$ denote the agent's information set and $f(y_1|\mathcal{I})$ denote the density of the present value of earnings conditional on the information set $\mathcal{I}$. The graph shows $f(y_1|\mathcal{I})$ with $\mathcal{I}$ containing no factors, factor 1, factors 1 and 2, and all three factors. The plots are smoothed using a Gaussian kernel.

In order to get an idea of the difference between variability and uncertainty figures 4.1 and 4.2 present the distributions of present value of earnings under different assumed information sets for the agent at the time the schooling decision is made. Even though there is a lot of dispersion, not all of it is truly uncertain to the agent. The variability in college earnings is reduced to a much greater extent than the variability in high school earnings as we go from a state in which the agent knows no factor to one in which he knows $\theta_{i,1}$ and $\theta_{i,2}$ – the correctly specified information set. Figure 4.3 shows that gains to college are also predictable.

Figure 4.3
Densities of the difference between the present value of college and high school earnings under different information sets for the agent at the time schooling decisions are made



Let Y_0, Y_1 denote the present value of high school and college earnings from age 18 to 65 discounted at a 3% interest rate respectively. Define $G=Y_1-Y_0$. Let $\mathcal{I}$ denote the agent's information set and $f(g|\mathcal{I})$ denote the density of the present value of earnings conditional on the information set $\mathcal{I}$. The graph shows $f(g|\mathcal{I})$ with $\mathcal{I}$ containing no factors, factor 1, factors 1 and 2, and all three factors. The plots are smoothed using a Gaussian kernel.

This pattern is more explicitly analyzed in table 5.1. In particular, as shown in the table, under the estimated information set at time 0 (i.e. including $\theta_{i,1}$ and $\theta_{i,2}$) roughly 43% of what would otherwise be considered uncertainty in high school earnings is predictable by the agent at the time his schooling decision is made. In the same manner, the variance of the gains to college when the information set contains $\theta_{i,1}$ and $\theta_{i,2}$ is only 20% of the one we would obtain if we assumed the unobservables for the agent and the analyst coincide. That is, of the total observed variability in the present value of college earnings only 20% constitutes uncertainty for the individual, 57% for monetary gains.

Table 5.1

Agent's forecast* of the variance of the present value of earnings

Under different information sets at schooling choice date

		Var(Y_H)	*Var(Y_C)*	*Var(Y_C-Y_H)*
	Variance with $\mathcal{I}=\varnothing$	349205	402710	459648
$\mathcal{I}_1=\theta_1$	Variance	346736	285190	372504
	Fraction of the variance** with $\mathcal{I}=\varnothing$ explained by $\mathcal{I}_1$	0.71%	29.18%	18.96%
$\mathcal{I}_2=\{\theta_1,\theta_2\}$	Variance	193100	76074	260178
	Fraction of the variance with $\mathcal{I}=\varnothing$ explained by $\mathcal{I}_2$	44.70%	81.11%	43.40%
$\mathcal{I}_3=\{\theta_1,\theta_2,\theta_3\}$	Variance	187993	71090	258965
	Fraction of the variance with $\mathcal{I}=\varnothing$ explained by $\mathcal{I}_3$	46.17%	82.35%	43.66%

*Variance of the unpredictable component of earnings from age 18-65 as predicted at age 18.
**The variance of the unpredictable component of high school earnings with $\mathcal{I}_1=\theta_1$ is (1-0.007)*349205=325916

Establishing this difference is clearly important for interpretation of the results. For example, if one were to assume that the agent's unobservables and the analyst's unobservables coincide, one would conclude that college is a riskier state since the variance in the present value of college earnings is much higher than the variance of high school earnings. However, when the correctly specified information set is used instead, the variance of the uncertainty in present value of college earnings facing the agent is much lower than the variance for high school.

Table 5.2

Proportion of people who, after observing their realized outcomes (keeping credit constraints in place), regret their choice

Choice under Uncertainty	**Choice under Certainty**	
	High School Choose HS: 52.3%	College Choose Col: 47.7%
High School Choose HS: 51.1%	87.02%	12.98%
College Choose Col: 48.9%	16.03%	83.97%

In order to study the importance of uncertainty for schooling decisions I perform the following experiment. In tables 5.2 and 5.3 I eliminate uncertainty. That is, I let people base their decisions on realized earnings instead of on expectations, *keeping the credit constraints in effect.* I keep the constraints at the same level in order to get a picture of the importance of uncertainty alone even though the interpretation of the model implies that credit constraints arise as a consequence of uncertainty. Section 4.6 relaxes both.

Table 5.3

Average Annual Ex-post Gains and Equivalent Variations with and without Uncertainty (keeping credit constraints in place)

	Choice under Uncertainty		**Choice under Certainty**	
	High School	College	High School	College
Ex-post Gain[1]	2.15%	9.70%	0.48%	11.85%
Equivalent Variation[2]	-8.60%	15.20%	-9.95%	18.57%

[1]Let Y_1 be the present value of earnings in college and Y_0 in high school. The lifetime ex-post gain is defined as $G=(Y_1-Y_0)/Y_0$. The annual ex-post gain is simply G/4.

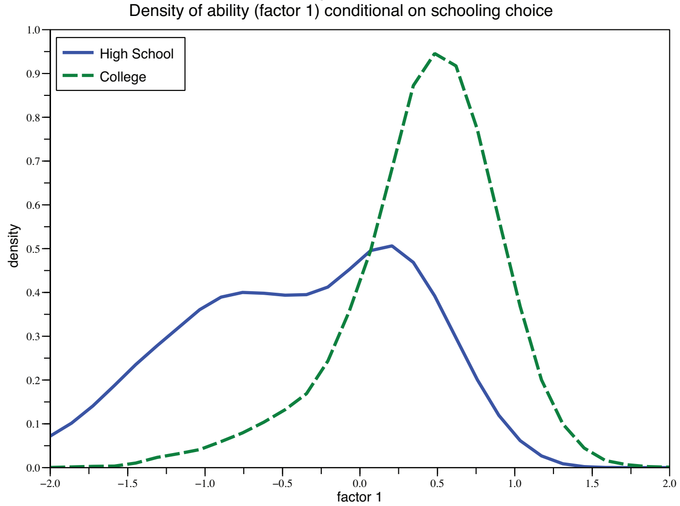
[2]The lifetime equivalent variation is defined as the proportion by which consumption (in each period) in high school needs to be changed so that the individual is indifferent between choosing high school and college. The annual equivalent variation is the lifetime equivalent variation divided by 4.

Table 5.2 shows the importance of looking at micro evidence when accounting for uncertainty in schooling decisions. If we look only at aggregate numbers the simulation shows a very small effect of eliminating uncertainty (keeping credit constraints in place) on college graduation. Allowing individuals to choose schooling using their realized outcomes instead of expectations decreases college graduation by only 1.2% from 48.9% to 47.7%. However, when we look at the micro evidence it becomes evident that the effect is much larger. Roughly 13% of the individuals who choose to stop at high school under uncertainty would choose to graduate college and roughly 16% of college graduates would regret their choice under uncertainty and would have stopped at high school. Notice that this matches exactly the patterns of figure 3.3: a proportion of high school graduates would have earned positive ex-post gains had they gone to college and a fraction of college graduates get negative ex-post gains and so may regret their decision.

As one would expect, sorting in terms of both ex-post gains increases. As shown in table 5.3, whereas the average ex-post gain to college for a college graduate who makes his decision using his expectations about future outcomes (*i.e.,* integrating out the unknown $\theta_{i,3}$, $\{\varepsilon_{i,h,t}, \varepsilon_{i,c,t}\}_{t=1}^{6}$) is 9.70% it would increase to 11.85% if he were allowed to choose based on actual realized earnings. The same experiment shows that the average ex-post gains for high school graduates would decrease from 2.15% to 0.48%.

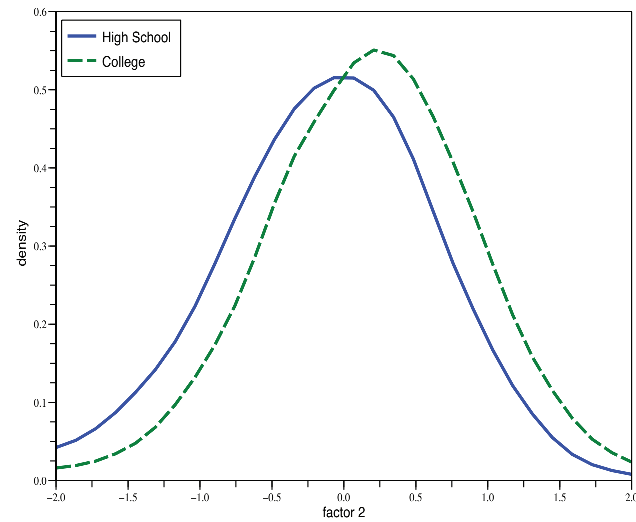
4.4 Sorting on Ability

Figure 5.1
Density of ability (factor 1) conditional on schooling choice



Let $f(\theta_1)$ denote the density function of factor θ_1. $f(\theta_1)$ is allowed to be a mixture of normals. The solid line plots the density of the factor conditional on choosing high school, that is, $f(\theta_1|\text{choice=high school})$. The dashed line plots the density of the factor conditional on choosing college, that is, $f(\theta_1|\text{choice=college})$. The plots are smoothed using a Gaussian kernel.

Figure 5.2
Density of factor 2 conditional on schooling choice



Let $f(\theta_2)$ denote the density function of factor θ_2. $f(\theta_2)$ is allowed to be a mixture of normals. The solid line plots the density of the factor conditional on choosing high school, that is, $f(\theta_2|\text{choice=high school})$. The dashed line plots the density of the factor conditional on choosing college, that is, $f(\theta_2|\text{choice=college})$. The plots are smoothed using a Gaussian kernel.

Figures 5.1 and 5.2 graph the densities of the factors 1 and 2 conditioning on schooling choice. Since $\theta_{i,3}$ is not known by the agent at the time the schooling decision is made, there is no selection based on it so the distribution does not change by schooling. There is strong evidence of selection in terms of ability (factor 1). The distribution of ability for college graduates is to the right of that for high school graduates. Individuals strongly sort in terms of ability even after controlling for family background and individual characteristics at test date. People with higher ability graduate college more than people with lower ability. The same holds true for factor 2. Since factor 2 is not tied to an external set of measurements it does not have a direct interpretation. However, since it is the factor that explains earnings the most, one could think of it as earnings ability. In other words, even after controlling for ability as measured by factor 1, college graduates tend to earn more money than high school graduates.

Table 6 shows further evidence of the importance of sorting in terms of ability. When the average realized gain to college by deciles of the ability ($\theta_{i,1}$) distribution is compared for people who choose college and people who choose high school the differences are minor. After conditioning on ability, college graduates obtain similar gains as high school graduates although uniformly larger. However, while almost

Table 6

Annual gains to college by deciles of the ability distribution

Ability decile	Choose high school		Choose college	
	Mean gain[1]	Proportion[2]	Mean gain[1]	Proportion[2]
1	-5.1%	18.5%	-3.9%	1.1%
2	-1.9%	16.7%	-1.5%	3.0%
3	1.1%	14.3%	1.6%	5.5%
4	3.2%	12.0%	4.5%	7.9%
5	5.0%	10.4%	6.4%	9.6%
6	6.1%	8.2%	7.9%	11.9%
7	7.6%	6.8%	9.3%	13.3%
8	8.6%	5.9%	11.4%	14.3%
9	10.6%	4.6%	13.1%	15.6%
10	13.4%	2.4%	16.8%	17.9%

[1] Let Y_C denote lifetime earnings in college and Y_H denote lifetime earnings in high school. The gain to college is $[(Y_C-Y_H)/Y_H]/4$

[2] Proportion of people who choose the schooling level indicated above and come from the ability decile to the left out of those who make the indicated choice. For example, 18.5% of those individuals who choose high school come from the first decile of the ability distribution.

61% of high school graduates come from the first four deciles of the ability distribution, only 16% of college graduates do. What accounts for the difference in figure 3.3 between high school and college graduates is in big part the different composition in terms of ability of the selected population which arise as a consequence of ability and its effects on preferences.

4.5 The Importance of Preferences

From the evidence presented so far, one thing should be clear: high school graduates and college graduates are not the same. There is a great deal of heterogeneity among people. Finding high returns to college for people who actually attend college (what is commonly called "treatment on the treated" in the evaluation literature) is not necessarily informative about how much people who choose not to attend college would make if they were to attend college.

Furthermore, not all of this variability is uncertain to the agent at the time he is making his decisions. He can actually forecast a considerable proportion of it. Even though this goes a long way to explain why individuals go to college, we are still facing the question of why some people would not take advantage of the 2.15% average gain they would get if they attended college.[30]

So far the fact that individual decisions are based on utility maximization and not on income comparisons has been ignored. The estimated risk aversion coefficient in the model is 0.81 (and 3.35 for the terminal period), within the numbers reported by Browning, Hansen, and Heckman (1999), so agents are in fact risk averse and do not care only about monetary returns.

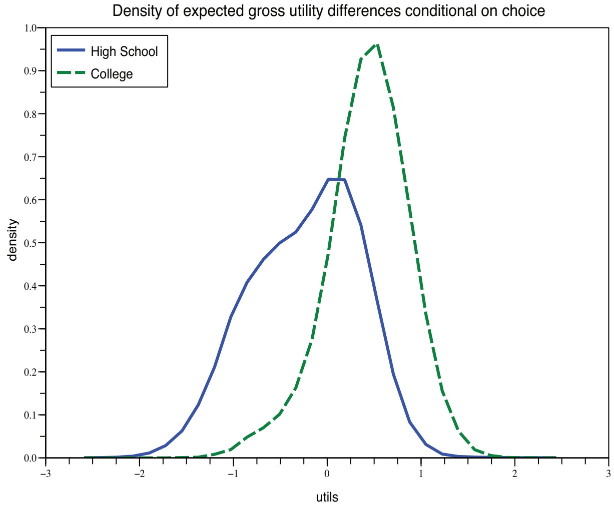
A number that better summarizes the gain that individuals obtain from their schooling decisions, the "equivalent variation", is presented in the second line of table 5.3. This number is calculated by solving for the fraction by which high school consumption needs to be multiplied every period for an individual to be indifferent between choosing high school or college. The fraction by which consumption needs to be changed is what I call the equivalent variation. The equivalent variation is the consumption value that an individual places on his schooling decision accounting for preferences and, depending on the case,

[30] Some people would argue we would also face the question of the "large" return to college for college graduates of 9.7%. Notice that this number roughly matches the number obtained from instrumental variables estimates. See Card (2001).

uncertainty. As shown in the table, high school graduates would need their consumption to be reduced by 8.6% every period for them to be indifferent. That is, even though high school graduates on average face a positive ex-post gain of going to college, once we account for the effect of preferences and uncertainty this is no longer the case. In the same manner, the seemingly large 9.7% ex-post gain college graduates obtain on average is increased to a 15.2% once preferences are accounted for.

Figure 6.1

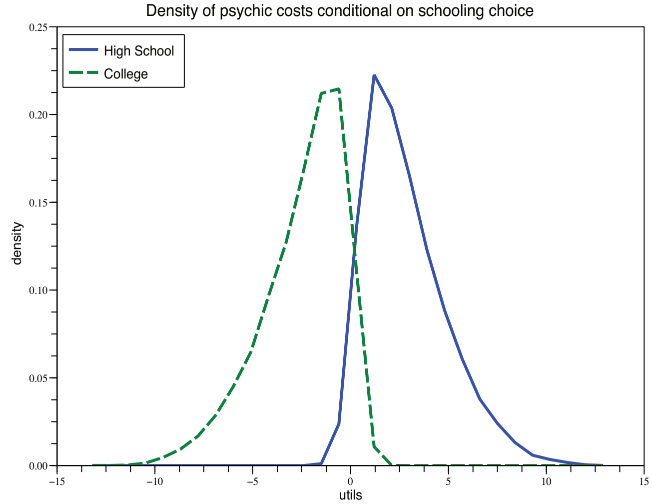
Density of expected gross utility differences conditional on choice



Let $V_{h,1}$ and $V_{c,1}$ denote the value functions for high school and college at period 1. Define the ex−ante gross utility difference, $D=E(V_{c,1}-V_{h,1}|\mathcal{I}_0)$ where the expectation is taken with respect to the information available at period 0. The solid line shows the density of D for agents who choose high school (i.e.

Figure 6.2

Density of psychic costs conditional on schooling choice



Let C denote the individual psychic costs associated with attending college. The solid line plots the density of the costs conditional on choosing high school, that is, f(C|choice=high school). The dashed line plots the density of costs conditional on choosing college, that is,

In figure 6.1 the the value function differences as perceived by the agent at the time the schooling decision is made is plotted. It is immediately apparent that, even though people who choose to go to college have a higher gross utility return to college than high school graduates, this is not enough to account for differences in college attendance. That is, we still find people who choose college with negative gross differences and people who choose high school with positive differences.

As shown in figure 6.2, the remaining part is captured by the difference in preferences for schooling (the "psychic" cost function). People with low psychic costs (*i.e.,* people who have a preference for school) are more likely to finish college. This preference is mostly driven by ability since people with higher ability tend to have lower psychic costs, but it is also driven by the difference in earnings captured by $\theta_{i,2}$.

4.6 Credit constraints, risk aversion and uncertainty

The final step in our analysis of schooling decisions is to account for credit constraints. For this purpose, I perform two different simulations. First, I follow the literature on the effect of borrowing constraints on schooling and look at the inability of people to pay for tuition. My first simulation consists on setting tuition to zero for everyone. The results are shown in the second line of table 7. Making college free for everyone effectively relaxes the credit constraint since it allows people to increase their consumption by the amount of money that they would otherwise dedicate to tuition. It also captures the effect of reducing the price of schooling so, the result of this exercise is an upper bound on the effect on relaxing the constraint via tuition. Roughly, there is an increase of 1.5% in college attendance. The result is consistent with the findings of Keane and Wolpin (2001), Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2006) where a similar experiment is performed and they all find a small effect.

In the second simulation, borrowing constraints are eliminated by allowing people to select using their realized earnings. In this case, we are back to a standard permanent income model. Notice that people are able to smooth consumption more effectively under this setting. The effects of eliminating the credit constraints and uncertainty with them are significantly different from the ones obtained from just eliminating tuition. From the third and fourth lines of table 7 we can see that college attendance decreases slightly when only uncertainty is eliminated, but in increases to roughly 58% when consumption smoothing is allowed. This is the case because only 5% of college graduates regret their choice while 21% of high school graduates do. As compared to the case in which only uncertainty is eliminated, by letting people smooth consumption by eliminating the borrowing constraints, we give them complete information and so to higher permanent incomes and their consumption at young ages (while in college) need not decrease as much.

Table 7

Percentage of people who choose college under different scenarios

Scenario	*Overall*
Original Economy	48.93%
Zero Tuition Economy	50.48%
Certainty with Credit Constraints	47.75%
Certainty without Credit Constraints	57.40%

Although these numbers clearly point to credit constraints playing an important role, they should be interpreted with caution. At least three of the assumptions I make contribute to these results. First, I am assuming that when people make their schooling decision they have no assets clearly making the constraint really important at early ages. Second, I am assuming that all risks are idiosyncratic. If some risks were aggregate (or in general not all risks were insurable) clearly, people would not be able to perfectly smooth consumption and the result would be less dramatic. The tuition subsidy example shows that this is the case by relaxing credit constraints but not completely eliminating them as would be the case of aggregate shocks[31]. Finally, as opposed to the case in which credit constraints are slightly relaxed via tuition, I now completely change the way the economy operates. General equilibrium effects would certainly dampen this response. [32]Even with the caveats just mentioned, the effect of credit constraints seems large enough to play an important role. If one were to cut the effect in half so that now college graduation only increases from 48.9% to 53.17% it is still almost three times the effect obtained by setting tuition to zero.

[31] Tuition simulations should also be carefully interpreted since they combine relaxing the constraint with a reduction in the price of a college education.

[32] See Heckman, Lochner, and Taber (1998a,b, 1999) for evidence on how important this general equilibrium effects can be.

5 Conclusion

This paper analyzes the role played by ability, uncertainty, preferences and credit constraints in explaining schooling choices. The conclusion of the paper is that there is no clear candidate for "the" explanation as to why some people got to college and some don't. This may not be a surprising result given the nature of the decision but it is by no means obvious a priori. As the results in this paper show, the college attendance decision is composed of many parts and all of them help explain the patterns in the data. Ability, preferences and uncertainty all play important roles. Once borrowing constraints are clearly defined and people are allowed to smooth consumption, as opposed to simply relaxing the constraint, credit constraints play a more important role than previously found.

To the extent that simpler models are easier to interpret and require weaker assumptions they are preferable. However, they cannot always answer the questions one maybe interested in. In very complex models, however, the relationship between results, assumptions and data is hard visualize. In an effort to show that the results in this paper are not simply a consequence of functional forms or distributional assumptions, I prove that the model I use is semiparametrically identified.

I build on the work Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) that identifies the uncertainty facing the agent. I find that agents can predict a considerable proportion of the variance of their future earnings as well as their gains at the time schooling decisions are made. The remaining uncertainty helps explain why some people do not go to college even though they would obtain positive returns and why some individuals attend college when their observed ex-post return is negative. Individuals make their decisions before all relevant information is revealed.

Ability is an important determinant of schooling choice. Individual ability helps explain college attendance mostly through the individual preference for school and not through its effects on earnings. That is, once I control for ability, the return to college is very similar for college graduates and for individuals with just a high school diploma. High ability individuals have lower psychic costs of attending college and it is mainly through this channel that ability affects college attendance.

I find little evidence of liquidity constraints when these are defined as individuals not being able to afford college. Moving to an economy with zero tuition increases college attendance by roughly 1.5%. Since this simulation also includes the effect of the reduction in the price for college this number is an upper bound on the effect of relaxing the constraint. However, when borrowing constraints are eliminated (as well as uncertainty) the effect is much larger and now college attendance increases by roughly 8%. Once credit constraints are defined in terms of consumption smoothing instead of liquidity constraints at a point in time they play a stronger role than previously found.

Appendix 1: Semiparametric Identification

The following assumptions are used throughout this section in order to prove semiparametric identification of all the elements of the model. For simplicity, delete the i subscript. Let $U_s = (U_{s,1}, ..., U_{s,T})$, $U = (U_h, U_c)$, $U^M = \left(U_1^M, ..., U_j^M\right)$, $\xi = (\xi_1, ..., \xi_T)$.

(A-1) U, U^M, W and ξ *have distributions that are absolutely continuous with respect to Lebesgue* *measure with support* $\mathbb{U}\times\mathbb{U}^{\mathbb{M}}\times\mathbb{W}\times\mathbb{Z}$ *that may be bounded or infinite. Variances are assumed to be finite. The cumulative distribution function of* W *is assumed to be strictly increasing over its full support.*[33]

(A-2) $\left(X, X^M, K, Z\right) \perp\!\!\!\perp \left(U, U^M, W, \xi\right)$ (Independence)

Semiparametric Identification of Measurements and Logearnings with and without a Factor Structure

Identification of the measurement system and logearnings equations is proved in theorems 5 and 5. Only the case in which the measurements are continuous is considered, but, as shown in Carneiro, Hansen, and Heckman (2003), the measurements could also be discrete or mixed discrete-continuous.[34]

Assume that the relevant parts of (A-1) and (A-2) hold (i.e., those for U^M, X^M), and that

(A-3) *Support* $\left(\mu_1^M\left(X^M\right), ..., \mu_J^M\left(X^M\right)\right) \supseteq$ *Support* $\left(U_1^M, ..., U_J^M\right)$.

Then, from data on $F\left(M \mid X^M\right)$ one can identify $\mu_j^M\left(X^M\right)$ over the support of X^M and the joint distribution of U^M: $F_{U^M}\left(u_M\right)$.

Proof. Identification of the mean functions over their support is trivial since we observe M_j for each X^M and can recover the marginal distribution of U_j^M. The intercepts are recovered from assumed zero mean of U_j^M. The joint follows immediately since $Pr\left(M < m \mid X^M\right) = F_{U^M}\left(m - \mu^M\left(X^M\right)\right)$ by assumption (A-2). Then, from (A-3) we can find an $X^M = x^M$ where $\mu^M\left(x^M\right) = k$ and k is a J dimensional vector. Let $m = k - \mu^M\left(x^M\right)$ so $Pr\left(M < m \mid X^M = x^M\right) = F\left(k\right)$. Since the point ω is arbitrary, we can vary it to identify the full joint distribution. ∎

Assume that the relevant elements of (A-1) and (A-2) (i.e., the joint conditions on X, Z, U, W) hold and that the following variation free condition holds:

(A-4) *Support*$\left(\phi\left(Z\right), \mu_s\left(X\right)\right) =$ *Support* $\left(\phi\left(Z\right)\right) \times$ *Support* $\left(\mu_s\left(X\right)\right)$. *(Variation free)*

Let $\mu_s\left(X\right) = \left(\mu_{s,1}\left(X\right), ..., \mu_{s,T}\left(X\right)\right)$. *Assume that Support*$\left(\phi\left(Z\right)\right) \supseteq$ *Support* $\left(W\right)$ *and Support* $\left(\mu_s\left(X\right)\right) \supseteq$ *Support* $\left(U_s\right)$. *Then, the mean functions* $\mu_{s,t}\left(X\right)$ *are identified on the support of* X. *Also, the joint distribution of* U_s *is nonparametrically identified for* $t = 1, ..., T$ *for each* $s = h, c$.

Proof. Under the conditions of the theorem, we can find limit sets $\mathcal{Z}^-$and a $\mathcal{Z}^+$such that $Pr\left(S = 1 \mid Z \in \mathcal{Z}^-\right) = 0$ and $Pr\left(S = 1 \mid Z \in \mathcal{Z}^+\right) = 1$ where we can still freely change the $\mu_s\left(X\right)$. In the limit sets, the conditions of Theorem *5* are satisfied and we can apply it for each system of schooling equations s. ∎

Further assume *Support* $\left(\mu_s\left(X\right), \mu^M\left(X^M\right)\right) \supseteq$ *Support* $\left(U_s, U^M\right)$. Identification of the joint distribution of $\left(U_s, U^M\right)$ for each s is straightforward since, in the limit set, we can form the left hand side of

$$\Pr\left(M \leq m, Y_s \leq y_s \mid X^M, X\right) = F_{U^M, U_s}\left(m - \mu^M\left(X^M\right), y_s - \mu_s\left(X\right)\right)$$

which we can trace by changing y_s and m. The intercepts of $\mu_s\left(X\right)$ and $\mu^M\left(X^M\right)$ are fixed from the assumption that the means (or medians) of U^M, U_s are zero.

When the unobservables are represented in terms of equations (13) and (15), the next theorems show that we can nonparametrically identify the distributions of the factors and the uniquenesses as well as the

[33] This assumption can easily be relaxed and is only made for convenience.

[34] In all cases, with additional assumptions, we can relax additive separability and identify functions of the form $y = \mu\left(X, U\right)$ by using the analysis in Matzkin (2003).

factor loadings. I first state a theorem that will be useful for this purpose.

Let Q_1 and Q_2 be two random variables that satisfy

$$\begin{aligned} Q_1 &= \theta + R_1 \\ Q_2 &= \theta + R_2 \end{aligned}$$

where $\theta, R_1$and R_2 are mutually independent with $E(\theta) < \infty, E(R_1) = 0, E(R_2) = 0$, the conditions of Fubini's theorem are satisfied for each random variable and they have nonvanishing (a.e.) characteristic functions. Then, the marginal densities of θ, R_1 and R_2 are identified.

Proof. See Kotlarski (1967). ■

Consider using only the information on measurements and earnings in each schooling state. For a given schooling level s we have a system of equations

$$\begin{aligned} M_1 &= \mu_1^M \left(X_1^M\right) + \theta\alpha_1^M + \varepsilon_1^M \\ &\vdots \\ M_J &= \mu_J^M \left(X_J^M\right) + \theta\alpha_J^M + \varepsilon_J^M \\ lnY_{s,1} &= \mu_{s,1}\left(X_{s,1}\right) + \theta\alpha_{s,1} + \varepsilon_{s,1} \\ &\vdots \\ lnY_{s,T} &= \mu_{s,T}\left(X_{s,T}\right) + \theta\alpha_{s,T} + \varepsilon_{s,T} \end{aligned} \tag{20}$$

The total number of equations is given by $J \cdot T$. Let L be the total number of factors. Sufficient conditions for identification will be: $L \leq 2 \cdot J \cdot T + 1$. That is, that the loadings are such that at least two equations depend only on the first factor, at least two depend only on the first and second factor, and so on up to the first $L-1$ factors and at least three equations depend on all L factors. Such an arrangement would be motivated by the assumptions about the arrival of information made in the text.

If we rearrange the equations in (20) and put the factor loadings in a matrix, it would have the form

$$\begin{array}{cccccc} \multicolumn{6}{c}{\text{Loadings for factor}} \\ \theta_1 & \theta_2 & \theta_3 & ... & \theta_{L-1} & \theta_L \\ \neq 0 & 0 & 0 & ... & 0 & 0 \\ \neq 0 & 0 & 0 & ... & 0 & 0 \\ \neq 0 & \neq 0 & 0 & ... & 0 & 0 \\ \neq 0 & \neq 0 & 0 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ \neq 0 & \neq 0 & \neq 0 & ... & \neq 0 & 0 \\ \neq 0 & \neq 0 & \neq 0 & ... & \neq 0 & 0 \\ \neq 0 & \neq 0 & \neq 0 & ... & \neq 0 & \neq 0 \\ \neq 0 & \neq 0 & \neq 0 & ... & \neq 0 & \neq 0 \\ \neq 0 & \neq 0 & \neq 0 & ... & \neq 0 & \neq 0 \end{array} \tag{21}$$

Now, assume that L_M factors enter the measurement equations and, without loss of generality, suppose we look at $s = h$.[35]

From the analysis in Theorems (5) and (5) we have data on $F\left(U^M, U_h \mid X, X^M\right)$ and assume that U^M, U_h has a factor structure representation as in (13). Then, the loadings $\left\{\alpha_j^M\right\}_{j=1}^J, \{\alpha_{h,t}\}_{t=1}^T$ are identified up to one normalization for each factor. The marginal distributions of $\{\theta_l\}_{l=1}^L, \left\{\varepsilon_j^M\right\}_{j=1}^J$ and $\{\varepsilon_{h,t}\}_{t=1}^T$ are nonparametrically identified as well.

Proof. To ease on notation I generically write the equations in (20) as $y_k = \sum_{l=1}^L \theta_l \delta_{k,l} + e_k$. So, if for example $\ln Y_{h,t}$ is the fifth equation, $y_5 = \ln Y_{h,t} - \mu_{h,t}(X)$, the loading on the third factor $\alpha_{h,t,3} = \delta_{5,3}$ and $\varepsilon_{h,t} = e_5$.

Notice that, since the factor has no natural scale, we need to set it (that is $\delta\theta = \kappa\delta\frac{\theta}{\kappa}$ for any constant κ). We also need to normalize the sign of the effect of the factor since, for example, having more of factor l and $\delta_{l,k} > 0$ is equivalent to having less of the factor and $\delta_{l,k} < 0$. To pin down sign and scale, we normalize one loading to one for each factor.

Start by taking the first two equations and any other equation $k > 2$ and normalize $\delta_{1,1} = 1$. Since we know their joint distribution we can identify the loadings on factor one for the second equation by forming

$$\frac{cov\left(y_k, y_2\right)}{cov\left(y_k, y_1\right)} = \delta_{2,1}$$

and the loading for the k^{th} equation from

$$\frac{cov\left(y_k, y_2\right)}{cov\left(y_1, y_2\right)} = \delta_{k,1}.$$

Since the choice of the k^{th} equation is arbitrary we can identify all of the loadings for factor one. With the loadings on hand, we can the take equation 1 and any equation k and form

$$y_1 = \theta_1 + e_1$$

$$\frac{y_k}{\delta_{k,1}} = \theta_1 + \sum_{l=2}^{L} \theta_l \frac{\delta_{k,l}}{\delta_{k,1}} + \frac{e_k}{\delta_{k,1}}.$$

Using Theorem 5 we can nonparametrically identify the distributions of θ_1, e_1 and $\sum_{l=2}^L \theta_l \frac{\delta_{l,k}}{\delta_{1,k}} + \frac{e_k}{\delta_{1,k}}$ (and of $\sum_{l=2}^L \theta_l \delta_{l,k} + e_k$).

We now take equations 3 and 4 and some arbitrary equation $k > 4$. Normalizing $\delta_{3,2} = 1$ we can identify the loadings on factor 2 on the remaining equations by forming

$$\frac{cov\left(y_k - \theta_1\delta_{k,1}, y_4 - \theta_1\delta_{4,1}\right)}{cov\left(y_3 - \theta_1\delta_{3,1}, y_4 - \theta_1\delta_{4,1}\right)} = \delta_{k,2}$$

[35] As before, if we are willing to make an additional normalization, we can identify the factor structure without the additive separability assumption using equations of the form $y = \mu\left(X, \theta'\alpha + \varepsilon\right)$ following the analysis in Heckman, Matzkin, Navarro, and Urzua (2004).

and take equations 3 and k

$$y_3 - \theta_1\delta_{3,1} = \theta_2 + e_3$$

$$\frac{y_k - \theta_1\delta_{k,1}}{\delta_{k,2}} = \theta_2 + \sum_{l=3}^{L}\theta_l\frac{\delta_{k,l}}{\delta_{k,2}} + \frac{e_k}{\delta_{k,2}}.$$

Apply theorem 5 and identify the distributions of θ_2,e_3 and $\sum_{l=3}^{L}\theta_l\frac{\delta_{k,l}}{\delta_{k,2}} + \frac{e_k}{\delta_{k,2}}$ nonparametrically. By proceeding sequentially we can identify all of the loadings and nonparametric distributions of the model for measurements and schooling $s = h$. ■

From the Theorem 5, we now have knowledge of the nonparametric distribution of the factors and uniquenesses as well as the loadings for measurements and schooling $s = h$. We next turn our attention to identification to the analogous system of equations in (20) for the college ($s = c$). Notice that, since the measurement equations do not depend on the schooling level chosen, they are identified from our previous argument and so are the distributions of the elements of θ.

Take any two equations from the college system and one of the equations that depend only on θ_1. First assume that at least one of the three equations corresponds to a measurement, say M_j. Then, from $cov\left(M_j, \ln Y_{c,t} \mid X^M, X\right) = \alpha_j^M\alpha_{c,t}\sigma_{\theta_1}^2$ we can identify $\alpha_{c,t}$ for any $t = 1, ..., T$. If, on the other hand, none of the equations is a measurement equation we can apply the next theorem that shows that if the distribution of θ is nonsymmetric, we can identify the factor loadings even if there are no measurements.

Take a generic system of three equations with a factor structure

$$y_1 = \theta_1\delta_{1,1} + e_1$$

$$y_2 = \theta_1\delta_{2,1} + e_2$$

$$y_3 = \theta_1\delta_{1,3} + \sum_{l=2}^{L}\theta_l\delta_{3,l} + e_3.$$

with θ a vector of mean zero mutually independent factors with known distribution of and e_1, e_2 and e_3 also mean zero independent of each other and of the factors θ. Then, if

(A-6) *$E\left(\theta_1^k\right) \neq 0$ for k an odd integer*

the loadings $\delta_{1,1}, \delta_{2,1}$ and $\delta_{3,1}$ are identified and so is the nonparametric distribution of the uniquenesses $\{e_j\}_{j=1}^2$ and of $\sum_{l=2}^{L}\theta_l\delta_{3,l} + e_3$.

Proof. I illustrate the case in which $k = 3$ for condition (A-6). The extension is obvious. Form

$$cov\left(y_1, y_2\right) = \delta_{1,1}\delta_{2,1}\sigma_{\theta_1}^2$$

$$cov\left(y_1, y_3\right) = \delta_{1,1}\delta_{3,1}\sigma_{\theta_1}^2$$

$$cov\left(y_2, y_3\right) = \delta_{2,1}\delta_{3,1}\sigma_{\theta_1}^2$$

and solve for

$$\delta_{1,1}^2 = \frac{cov\left(y_1, y_2\right) cov\left(y_1, y_3\right)}{cov\left(y_2, y_3\right)\sigma_{\theta_1}^2}.$$

Then from

$$E\left(y_1^2 y_j\right) = \delta_{1,1}^2 \delta_{1,j} E\left(\theta_1^3\right)$$

we identify $\delta_{j,1}$ for all $j > 1$. Going back to $cov\left(y_1, y_2\right)$ we recover $\delta_{1,1}$. The distributions of the uniquenesses follow by deconvolution.

Identification of the whole system for $s = c$ follows by applying the same logic of theorem 5 sequentially. ■

Identification of the Cost Function

Write the college attendance condition (6) as $E\left(\mathcal{V}_{c,1}\left(\mathcal{I}_1\right) - \mathcal{V}_{h,1}\left(\mathcal{I}_1\right) - \phi\left(Z\right) - \theta\lambda - \omega \mid \mathcal{I}_0\right) > 0$. Let θ^0 denote the elements of θ included in the agent's information set at the time the schooling decision is made and let λ^0 denote the subvector of λ associated with them. Define $E\left(\mathcal{V}_{c,1}^*\left(\mathcal{I}_1\right) - \mathcal{V}_{h,1}^*\left(\mathcal{I}_1\right) \mid \mathcal{I}_0\right) = \mu_V\left(X\right) + \tau\left(X, \theta^0\right)$ which is known since all of the elements of $E\left(\mathcal{V}_{c,1}^*\left(\mathcal{I}_1\right) - \mathcal{V}_{h,1}^*\left(\mathcal{I}_1\right) \mid \mathcal{I}_0\right)$ are known from our previous analysis. The econometrician has data on the left hand side of

$$\Pr\left(S = c | X, Z\right) = \Pr\left(\tau\left(X, \theta^0\right) - \theta^0\lambda^0 - \omega > \phi\left(Z\right) - \mu_V\left(X\right)\right).$$

Assume that the relevant elements of (A-1) hold. Change (A-2) so that the independence of Z and the error terms holds conditional on X.

Let Z^e be the elements of Z that are not a part of X (excluded) and further assume that we can define $\phi\left(z^e, x\right)$ for all pairs $\left(z^e, x\right)$ in the support of Z. As with all discrete choice problems the scale needs to be set. Assume that

(A-7) $var\left(\tau\left(\widetilde{x}, \theta^0\right) - \theta^0\lambda^0 - \omega\right) = 1$ for $X = \widetilde{x}$.

Then, if $\phi\left(Z\right)$ satisfies the requirements is a part of the Matzkin class of functions (Matzkin (1992), Heckman and Navarro (2007)), λ^0 and the nonparametric distribution of ω are identified up to normalization.

Proof. Define $\Upsilon\left(X, \theta^0\right) = \tau\left(X, \theta^0\right) - \theta^0\lambda^0 - \omega$ and fix $X = \widetilde{x}$. The observed probability that the agent chooses college (conditional on $X = \widetilde{x}$ and using A-7) is $\Pr\left(\Upsilon\left(\widetilde{x}, \theta^0\right) > \phi\left(Z^e, \widetilde{x}\right) - \mu_V\left(\widetilde{x}\right)\right)$ where $\mu_V\left(\widetilde{x}\right)$ is a known constant (conditional on $X = \widetilde{x}$). Using the (conditional on X) independence of $\theta^0\lambda_0 + \omega$ and Z^e we can then use the analysis of Matzkin (1992) to identify $\phi\left(Z^e, \widetilde{x}\right)$ and the distribution of $\Upsilon\left(\widetilde{x}, \theta^0\right)$, all of this conditional on $X = \widetilde{x}$, up to a normalization.

Next, still conditional on $X = \widetilde{x}$, we can form the joint distribution of measurements, logearnings and the choice index since know the left hand side

$$\Pr\left(U^M \leq m - \mu^M\left(X^M\right), U_s \leq y_s - \mu_s\left(\widetilde{x}\right) \mid X = \widetilde{x}, S = s, Z^e = z^e\right) \Pr\left(S = s \mid Z^e = z^e\right)$$

$$= \int_{-\infty}^{m - \mu^M\left(X^M\right)} \int_{-\infty}^{y_s - \mu_s\left(\widetilde{x}\right)} \int_{\phi\left(z^e, \widetilde{x}\right) - \mu_V\left(\widetilde{x}\right)}^{\infty} f\left(U^M, U_s, \Upsilon\left(\widetilde{x}, \theta^0\right)\right) d\Upsilon\left(\widetilde{x}, \theta^0\right) dU_s dU^M$$

and we can trace it by varying m, y, z_E. With the joint distribution in hand, form the covariance matrix and, following the reasoning of theorems 5 and 5 identify λ^0. To see how, take the covariance between the choice index and the first equation (i.e., the one that only depends on θ_1), say M_1 :

$$cov\left(\Upsilon\left(\widetilde{x},\theta^0\right),M_1\right)=cov\left(\tau\left(\widetilde{x},\theta^0\right),\theta_1\alpha^M_{1,1}\right)+\lambda^0_1\alpha^M_{1,1}\sigma^2_{\theta_1}$$

where everything but λ^0_1 is known so we can solve for it. Proceeding recursively we identify all of the elements of λ^0.

With λ^0 known, we can take $\Upsilon\left(\widetilde{x},\theta^0\right)$ and deconvolve the distribution of ω since it is independent of X by A-2 and of θ by the factor structure assumptions. Finally, by changing the value of $z^e,\widetilde{x}$ we can trace $\phi\left(Z^e,X\right)$ coordinate by coordinate since, for any value of X, the scale $var\left(\tau\left(X,\theta^0\right)-\theta^0\lambda_0-\omega\right)$ is a function of known elements. ■

Appendix 2: Pooling Datasets

A serious practical empirical problem plagues most life cycle analyses. It is a rare data set that includes the full life cycle earnings experiences of any person along with their test scores, measurements, schooling choices and background variables. Many data sets like the National Longitudinal Survey of Youth (NLSY79) have partial information up to some age. A few other data sets (*e.g.,* the Panel Survey on Income Dynamics, PSID) have full information on some life cycle variables but lack the full detail of the richer data which provide information only on truncated life cycles. As I show in this Appendix, the factor model setup used in this paper provides a natural framework for combining samples.

To fix ideas and motivate the empirical work, suppress the individual i subscripts and, following the text, write

$$\ln Y_{s,t}=X_t\beta_{s,t}+\theta_1\alpha_{s,t,1}+\theta_2\alpha_{s,t,2}+\varepsilon_{s,t}\qquad t=1,...,6. \tag{22}$$

An individual picks $S=1$ if $E\left(\mathcal{V}^*_{c,1}-\mathcal{V}^*_{h,1}-\phi\left(Z\right)-\theta_1\lambda_1-\theta_2\lambda_2-\omega\mid I_0\right)>0$, where Z may include elements in common with X. Following the notation in Appendix 1, I rewrite this in index form as

$$I=\mu_V\left(X\right)+\tau\left(X,\theta_0\right)-\phi\left(Z\right)-\theta_1\lambda_1-\theta_2\lambda_2-\omega>0.$$

Measured consumption is given by

$$\ln\widehat{C}_t=\ln\mathcal{C}_t\left(\mathcal{I}_t\right)+K_t\pi+\xi_t.$$

Finally, the external measurements are written as

$$M_j=X_M\beta_{M,j}+\theta_1\alpha_{M,j}+\varepsilon_{M,j},\qquad j=1,...,J,$$

where J is the number of test scores. For the case in which we have access to full life cycle data, the contribution to the likelihood of an individual who chooses $S=s$, is given by

$$\int_\Theta\prod_{t=1}^{6}\prod_{s=0}^{1}\left\{f\left(\ln Y_{s,t}|\theta,X_t\right)Pr\left(S=s|Z,\theta\right)\right\}^{1(S=s)}\prod_{t=1}^{4}f\left(\ln\widehat{C}_t|\theta,K\right)\prod_{j=1}^{J}f\left(M_j|\theta,X_M\right)dF\left(\theta\right).$$

Identification follows from the analysis in Appendix 1.

Now, suppose that we only have access to a sample A in which we only get to observe some of the variables at early stages of the life cycle. In particular, assume that sample A does not include observations on $\{Y_{s,t}\}_{t=4}^{6}$ as is the case with the NLSY79.

The contribution to the likelihood of an individual who chooses, for example, $S = 1$ is

$$\int_{\Theta} \left[\begin{array}{c} \prod_{t=1}^{3} \{f(\ln Y_{s,t}|\theta, X_t)\} Pr(I > 0|Z, \theta) \prod_{t=1}^{3} f\left(ln\widehat{C}_t|\theta, K\right) \\ \prod_{j=1}^{J} f(M_j|\theta, X_M) \left\{\prod_{t=4}^{6} \int f(\ln Y_{s,t}|\theta, X_t) dF(\ln Y_{s,t})\right\} \end{array} \right] dF(\theta)$$

$$= \int_{\Theta} \prod_{t=1}^{3} \{f(\ln Y_{s,t}|\theta, X_t)\} Pr(I > 0|Z, \theta) \prod_{t=1}^{3} f\left(\ln \widehat{C}_t|\theta, K\right) \prod_{j=1}^{J} f(M_j|\theta, X_M) dF(\theta) \tag{23}$$

We integrate out data for the periods in which we do not observe earnings. In the same way if, for example, consumption was not observed for this individual in period 3, we could simply integrate it out.

Notice that we can still identify the factor structure on earnings (for periods 1 through 3), the consumption measurement error parameters and distributions for the periods observed, risk aversion parameter if at least two periods of consumption observations are available and the measurement equations using the argument on Appendix 1. The latent index and the consumption information for the periods in which no consumption observations are available are lost since we need to identify all of the components of the solution to the dynamic program to form the schooling decision.

Now, suppose that we have access to a second independent sample B that is generated by the same process that generates sample A.[36] In this second sample, we do not observe $\{M_k\}_{k=1}^{K}$ but we do observe earnings, consumption and schooling choices (and X and Z) for all time periods. For an individual with $S = 1$ sampled from B, his contribution to the likelihood would be given by

$$\int_{\Theta} \prod_{t=1}^{6} \{f(\ln Y_{s,t}|\theta, X_t)\} Pr(I > 0|Z, \theta) \prod_{t=1}^{4} f\left(\ln \widehat{C}_t|\theta, K\right) \prod_{j=1}^{J} \int f(M_j|\theta, X_M) dF(M_j) dF(\theta)$$

$$= \int_{\Theta} \prod_{t=1}^{6} \{f(\ln Y_{s,t}|\theta, X_t)\} Pr(I > 0|Z, \theta) \prod_{t=1}^{4} f\left(\ln \widehat{C}_t|\theta, K\right) dF(\theta). \tag{24}$$

From this sample alone we cannot recover the loadings or the marginal distributions of $\theta_1, \theta_2, \theta_3, \left\{\{\varepsilon_{s,t}\}_{s=1}^{2}\right\}_{t=1}^{6}$ and the cost function, without additional assumptions.[37]

Suppose we combine both samples (so that a person's contribution to the likelihood is given by (23) if an individual comes from sample A and is given by (24) if he comes from sample B). In this case, we would be able to recover all of the elements of the model. To see why, notice that from sample A alone all

[36] By this I mean that the parameters and distributions of the implied random variables of both samples are the same.

[37] It is clear we will never recover any of the parameters of the measurements. If we changed our normalizations on the rest of the system however, so that now θ_2 did not enter earnings at $t = 1, 2$ for example, we could recover all of the remaining parameters of the model. Alternative normalizations would produce identification, but inconsistency in the models fit across samples.

that was left to recover in terms of earnings were the parameters and distributions for earnings in $t > 3$. Now, in sample B we can form

$$\begin{aligned} Cov\left(Y_{s,t}, Y_{s,1}\right) &= \alpha_{s,t,1}\alpha_{s,1,1}\sigma^2_{\theta_1} \\ Cov\left(Y_{s,t}, Y_{s,2}\right) &= \alpha_{s,t,1}\alpha_{s,2,1}\sigma^2_{\theta_1} + \alpha_{s,t,2}\alpha_{s,2,2}\sigma^2_{\theta_2}, \qquad t = \{4,5,6\}, \end{aligned}$$

where all parameters except $\alpha_{s,t,1}$ and $\alpha_{s,t,2}$ are known from our analysis of sample A. It is straightforward to show that we can solve for the unknowns $\alpha_{s,t,1}$and $\alpha_{s,t,2}$ for $t > 3$ and $s = 0, 1$. Identification of the parameters of the third factor follows then immediately from our analysis of Appendix 1. Since we have identified all of the parameters of the earnings equations, we can solve for the parameters of the cost equation by solving the dynamic program. More generally, we can obtain more efficient estimates for the over-identified parameters by pooling samples.

This procedure abstracts from cohort effects on the coefficients and factor loadings, and cohort effects on the distributions of θ. With additional structure (*e.g.*, additivity), I can identify such effects, but I acknowledge that general cohort effects can dramatically bias estimates based on pooling the data.

Appendix 3: Data

I use data on white males from NLSY79 and pool it with data for white males that are household heads from PSID. In the original NLSY79 sample there are 2439 white males. Out of this, 1334 have either a high school degree (and high school only) or a college degree. I then try to recover earnings for as many individuals as possible. First, individual earnings are formed by taking total income from wages and salary in the past calendar year directly from the NLSY deflated to year 2000 prices using the Consumer Price Index reported by the Bureau of Labor Statistics. Then, because the NLSY survey was not administered in 1995, 1997 and 1999 so earnings for any individual in these years are not observed I impute earnings in these years by taking the average of the earnings in the immediately adjacent years if available. Unemployed individuals have zero earnings on those years and are flagged as having missing earnings otherwise.

The PSID sample is a little more problematic since attrition is much more common in that survey than in the NLSY79.[38] Many of the characteristics required for my analysis, in particular for the analysis of test scores, are not available. The PSID, however, allows me to analyze people of all ages something that cannot be done with the NLSY79.[39] Earnings in the PSID sample are obtained by using the annual labor earnings variable. As with NLSY79, I impute earnings for the years in which there was no survey using an average of the two immediately adjacent years if possible.

The individual life (ages 18 to 65) is then simplified into 7 periods: $t = 0$ (schooling choice decision, right before 18), $t = 1$ (18-25), $t = 2$ (26-33), $t = 3$ (34-41), $t = 4$ (42-49), $t = 5$ (50-57) and $t = 6$ (58-65) where earnings for each period are defined as the present value of earnings for the ages included in the

[38] But see Fitzgerald, Gottschalk, and Moffitt (1998) for evidence that observed people in PSID have similar characteristics as those in the CPS so attrition is roughly random.

[39] NLSY79 interviews people born between 1957 and 1964 so it is only possible to follow all cohorts up to age 40 and the older cohort up to age 47.

Table A-1

NLSY79 and PSID for people born between 1957 and 1964

Variable Name	NLSY79		PSID	
	Mean	*Std. Err.*	*Mean*	*Std. Err.*
South at age 14	0.22	0.42	0.25	0.44
Urban at age 14	0.75	0.44	0.88	0.33
Mother's Education	4.43	1.45	4.57	1.59
Year of Birth	60.82	2.17	60.02	2.21
Schooling	0.46	0.50	0.45	0.50
Number of Siblings	2.82	1.84	2.92	1.86
Local Tuition at age 18*	2.09	0.80	2.02	0.56
Local Tuition at age 19*	2.04	0.75	1.95	0.59
Local Tuition at age 20*	1.99	0.70	1.82	0.56
Local Tuition at age 21*	2.01	0.71	1.84	0.61
Local Unemp at age 17: High School	0.07	0.02	0.06	0.02
Local Wage at age 17: High School	13.65	1.50	12.90	1.18
Local Unemp at age 17: College	0.03	0.01	0.03	0.01
Local Wage at age 17: College	20.46	2.29	19.07	1.75
Married	0.40	0.44	0.80	0.34
Number of Children	0.53	0.95	1.01	0.98
Age of Youngest Child	0.87	2.10	1.54	1.92

* Thousands of Dollars

period discounted using an interest rate of 3%. If the present value of earnings cannot be formed then the individual is flagged as having missing present value of earnings for the period.

The procedure to get consumption data in both samples is fairly similar in principle. Household consumption at time period t is defined as the difference between available resources -household income plus assets available at the beginning of the period- minus the discounted assets available the next period. [40] In order to recover more consumption observations, if assets are not observed at the beginning of the period as required I use assets either one year before or after it (discounted appropriately). Imputing household consumption for the PSID sample is slightly more problematic since questions about wealth were only asked in 1984, 1989, 1994, 1999 and 2001 but it is done in the same way as with NLSY79. To impute individual consumption I divide household consumption by the square root of the number of members in the household. I try to correct for the measurement error introduced by this procedure in the estimation as explained in the text.

Tuition between 1972 and 2000 is defined as the average in state tuition in colleges in the county of residence. If there is no college in the county then average tuition in the state is taken instead. For years prior to 1972, national tuition trends are used to project county tuition backwards keeping the average observed structure of the period 1972-1977. That is, a regression of the difference between national tuition and county tuition between 1972 and 1977 against county dummies is run. The predicted value is the average difference between national and county tuition which can then be added to the national tuition observed in the years previous to 1972.

To perform a quick check of whether the NLSY79 and PSID background variables are comparable,

[40] See Browning and Leth-Petersen (2003) for evidence that, at least for the case of Denmark where very detailed data on consumption is available, this procedure works well.

table A-1 shows the means in both datasets for the 1957-1964 cohort.[41] No significant differences in either means or standard deviations for the comparable cohort on both datasets can be seen with the exception of the consumption measurement error variables. Table 1 in the text contains a summary of all variables in the pooled data set.

NLSY specific variables

In 1980, NLSY respondents were administered a battery of ten achievement tests referred to as the Armed Forces Vocational Aptitude Battery (ASVAB) (See Cawley, Conneely, Heckman, and Vytlacil (1997) for a complete description). The math and verbal components of the ASVAB can be aggregated into the Armed Forces Qualification Test (AFQT) scores.[42] Many studies have used the overall AFQT score as a control variable, arguing that this is a measure of scholastic ability. In this paper, the interpretation that AFQT is an imperfect proxy for scholastic ability is taken and the factor structure is used to capture this. Potential aggregation bias is avoided by using each of the components of the ASVAB score as a separate measure.

[41] I look at these variables in particular because the questions for the PSID sample refer to the period when "the individual grew up" whereas the NLSY79 variables refer to a particular age.

[42] Implemented in 1950, the AFQT score is used by the U.S. Army to screen draftees.

Appendix 4: Parameter Estimates

Table A-2

Estimated Parameters for Test Equations

	Arithmetic Reasoning		Word Knowledge		Paragraph Completion		Coding Speed		Mathematics Knowledge	
Variable	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*
Constant	-0.99	0.19	-1.00	0.19	-0.98	0.19	-1.00	0.19	-0.98	0.19
Lived in South at age 14	-0.08	0.04	-0.08	0.04	-0.08	0.04	-0.08	0.04	-0.08	0.04
Lived in Urban at age 14	0.06	0.04	0.06	0.04	0.06	0.04	0.06	0.04	0.06	0.04
Mother's Education	0.09	0.01	0.09	0.01	0.09	0.01	0.09	0.01	0.09	0.01
Number of Siblings	-0.03	0.01	-0.03	0.01	-0.03	0.01	-0.03	0.01	-0.03	0.01
Highest grade at test date	0.13	0.02	0.13	0.02	0.13	0.02	0.13	0.02	0.13	0.02
Enrolled at test date	0.03	0.04	0.03	0.04	0.03	0.04	0.03	0.04	0.03	0.04
Age at test date	-0.04	0.02	-0.04	0.02	-0.04	0.02	-0.04	0.02	-0.04	0.02
θ_1	1.00	-	0.87	0.03	0.93	0.03	0.61	0.04	0.96	0.04
Variance	0.24	0.01	0.33	0.02	0.33	0.02	0.61	0.03	0.24	0.02

Table A-3

Estimated Parameters for High School Earnings Equation

	Age 18-25		Age 26-33		Age 34-41		Age 42-49		Age 50-57		Age 58-65	
Variable	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*
Constant	9.50	0.00	9.50	0.00	9.50	0.00	9.50	0.00	9.50	0.00	9.50	0.00
Age 26-33	1.46	0.04	1.46	0.04	1.46	0.04	1.46	0.04	1.46	0.04	1.46	0.04
Age 34-41	1.78	0.03	1.78	0.03	1.78	0.03	1.78	0.03	1.78	0.03	1.78	0.03
Age 42-49	1.63	0.00	1.63	0.00	1.63	0.00	1.63	0.00	1.63	0.00	1.63	0.00
Age 50-57	1.63	0.00	1.63	0.00	1.63	0.00	1.63	0.00	1.63	0.00	1.63	0.00
Age 58-65	0.79	0.00	0.79	0.00	0.79	0.00	0.79	0.00	0.79	0.00	0.79	0.00
θ_1	0.43	0.08	0.09	0.04	0.06	0.01	-0.12	0.00	0.20	0.00	-0.22	0.00
θ_2	0.00	-	1.00	-	1.16	0.03	1.08	0.00	0.90	0.00	0.10	0.00
θ_3	0.00	-	0.00	-	0.00	-	1.00	-	1.26	0.00	0.95	0.00
Mixture Parameters:												
Mean 1	0.39	0.06	0.03	0.03	0.05	0.01	0.03	0.00	0.18	0.00	0.41	0.00
Mean 2	-1.77	-	-0.88	-	-0.64	-	-0.24	-	-0.72	-	-0.65	-
Variance 1	0.74	0.12	0.17	0.02	0.08	0.01	0.09	0.01	0.12	0.00	0.31	0.00
Variance 2	5.12	1.47	14.99	2.00	7.19	0.65	0.81	0.08	1.93	0.00	2.16	0.00
Mixture Weight 1	0.82	0.08	0.96	0.01	0.93	0.02	0.88	0.00	0.80	0.00	0.62	0.00
Mixture Weight 2	0.18	-	0.04	-	0.07	-	0.12	-	0.20	-	0.38	-

Table A-4

Estimated Parameters for College Earnings Equation

	Age 18-25		Age 26-33		Age 34-41		Age 42-49		Age 50-57		Age 58-65	
Variable	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*
Constant	8.64	0.00	8.64	0.00	8.64	0.00	8.64	0.00	8.64	0.00	8.64	0.00
Age 26-33	2.78	0.03	2.78	0.03	2.78	0.03	2.78	0.03	2.78	0.03	2.78	0.03
Age 34-41	3.03	0.01	3.03	0.01	3.03	0.01	3.03	0.01	3.03	0.01	3.03	0.01
Age 42-49	3.05	0.00	3.05	0.00	3.05	0.00	3.05	0.00	3.05	0.00	3.05	0.00
Age 50-57	2.85	0.00	2.85	0.00	2.85	0.00	2.85	0.00	2.85	0.00	2.85	0.00
Age 58-65	2.36	0.00	2.36	0.00	2.36	0.00	2.36	0.00	2.36	0.00	2.36	0.00
θ_1	0.17	0.10	0.21	0.04	0.47	0.00	1.12	0.00	1.50	0.00	1.14	0.00
θ_2	0.00	-	0.93	0.03	1.25	0.01	1.27	0.00	1.28	0.00	0.81	0.00
θ_3	0.00	-	0.00	-	0.00	-	0.37	0.00	1.12	0.00	1.19	0.00
Mixture Parameters:												
Mean 1	0.46	0.06	0.05	0.02	0.06	0.00	0.06	0.00	0.22	0.00	0.45	0.00
Mean 2	-1.46	-	-0.60	-	-1.11	-	-0.41	-	-3.22	-	-1.39	-
Variance 1	0.79	0.13	0.20	0.02	0.10	0.00	0.07	0.00	0.10	0.00	0.25	0.00
Variance 2	3.30	0.71	5.72	0.50	7.40	0.02	2.13	0.00	3.73	0.00	2.36	0.00
Mixture Weight 1	0.76	0.09	0.92	0.01	0.95	0.00	0.88	0.00	0.94	0.00	0.76	0.00
Mixture Weight 2	0.24	-	0.08	-	0.05	-	0.12	-	0.06	-	0.24	-

Table A-5

Estimated Parameters for Consumption Measurement Error

	Age 18-25		Age 26-33		Age 34-41		Age 42-49		Age 50-57		Age 58-65	
Variable	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*
Constant	0.54	0.04	0.54	0.04	0.54	0.04	0.54	0.04	0.54	0.04	0.54	0.04
Age 26-33	-0.68	0.04	-0.68	0.04	-0.68	0.04	-0.68	0.04	-0.68	0.04	-0.68	0.04
Age 34-41	-0.64	0.04	-0.64	0.04	-0.64	0.04	-0.64	0.04	-0.64	0.04	-0.64	0.04
Age 42-49	-0.52	0.04	-0.52	0.04	-0.52	0.04	-0.52	0.04	-0.52	0.04	-0.52	0.04
Age 50-57	-0.46	0.05	-0.46	0.05	-0.46	0.05	-0.46	0.05	-0.46	0.05	-0.46	0.05
Age 58-65	-0.63	0.07	-0.63	0.07	-0.63	0.07	-0.63	0.07	-0.63	0.07	-0.63	0.07
Married	-0.10	0.02	-0.10	0.02	-0.10	0.02	-0.10	0.02	-0.10	0.02	-0.10	0.02
Number of Children	-0.18	0.01	-0.18	0.01	-0.18	0.01	-0.18	0.01	-0.18	0.01	-0.18	0.01
Age of Youngest Child	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
Mixture Parameters:												
Mean 1	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-
Mean 2	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-
Variance 1	0.88	1.17	0.04	0.00	1.77	0.27	0.04	0.00	0.05	0.01	0.02	0.02
Variance 2	0.31	0.11	0.86	0.13	0.04	0.00	0.79	0.45	4.31	8.10	0.46	0.19
Mixture Weight 1	0.16	0.43	0.86	0.02	0.06	0.02	0.94	0.02	0.94	0.05	0.41	0.16
Mixture Weight 2	0.84	-	0.14	-	0.94	-	0.06	-	0.06	-	0.59	-

Table A-6

Estimated Parameters for Factor Distributions

	Factor 1		Factor 2		Factor 3	
Mixture Parameters:	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*	*Estimate*	*Std. Err.*
Mean 1	-0.65	0.00	0.09	0.00	0.12	0.00
Mean 2	0.44	-	-1.60	-	-1.55	-
Variance 1	0.48	0.00	0.49	0.00	0.07	0.00
Variance 2	0.14	0.00	0.55	0.00	1.12	0.00
Mixture Weight 1	0.41	0.00	0.95	0.00	0.93	0.00
Mixture Weight 2	0.59	-	0.05	-	0.07	-

Table A-7

Estimated Parameters for Cost Equation and Utility Functions

Cost		
Variable	*Estimate*	*Std. Err.*
Constant	3.09	0.55
Lived in South at age 14	0.55	0.11
Lived in Urban at age 14	-1.13	0.12
Mother's Education	-1.32	0.08
Year of Birth	0.01	0.01
PSID	0.61	0.15
Number of Siblings	0.29	0.03
Local HS Unemp at age 17	6.90	1.84
Local HS Wage at age 17	0.62	0.06
Local Col Unemp at age 17	2.85	2.60
Local Col Wage at age 17	-0.30	0.04
θ_1	-2.44	0.42
θ_2	-0.91	0.35
Variance	1.00	-
Utility		
Variable	*Estimate*	*Std. Err.*
Risk Aversion (ψ)	0.82	0.00
Risk Aversion Terminal *(χ)*	3.35	0.00
Beta Terminal *(B)*	4.05	0.00
Discount Factor *($1/(1+\rho)$)*	0.58	0.00

References

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109(3), 659–684.

Ashenfelter, O., and A. B. Krueger (1994): "Estimates of the Economic Returns to Schooling from a New Sample of Twins," *American Economic Review*, 84(5), 1157–1173.

Ashenfelter, O., and C. Rouse (1998): "Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins," *Quarterly Journal of Economics*, 113(1), 253–284.

Attanasio, O., and H. Low (2004): "Estimating Euler Equations," *Review of Economic Dynamics*, 7, 405–435.

Belzil, C., and J. Hansen (2002): "Unobserved Ability and the Return to Schooling," *Econometrica*, 70(5), 2075–2091.

Blundell, R., L. Pistaferri, and I. Preston (2004): "Consumption inequality and partial insurance," Discussion Paper WP04/28, Institute for Fiscal Studies.

Blundell, R., and I. Preston (1998): "Consumption Inequality and Income Uncertainty," *Quarterly Journal of Economics*, 113(2), 603–640.

Brown, M., J. K. Scholz, and A. Seshadri (2007): "A New test of Borrowing Constraints for Education," Unpublished Manuscript, University of Wisconsin-Madison.

Browning, M., L. P. Hansen, and J. J. Heckman (1999): "Micro Data and General Equilibrium Models," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1A, chap. 8, pp. 543–633. Elsevier.

Browning, M., and S. Leth-Petersen (2003): "Imputing consumption from income and wealth information," *Economic Journal*, 113, 282–301.

Browning, M., and A. Lusardi (1996): "Household Saving: Micro Theories and Micro Facts," *Journal of Economic Literature*, 34(4), 1797–1855.

Cameron, S. V., and J. J. Heckman (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106(2), 262–333.

——— (2001): "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109(3), 455–99.

Cameron, S. V., and C. Taber (2004): "Estimation of Educational Borrowing Constraints Using Returns to Schooling," *Journal of Political Economy*, 112(1), 132–182.

Card, D. (1995): "Earnings, Schooling, and Ability Revisited," in *Research in Labor Economics*, ed. by S. Polachek, pp. 23–48. JAI Press, Greenwich, CT.

——— (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.

Carneiro, P., K. Hansen, and J. J. Heckman (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44(2), 361–422, 2001 Lawrence R. Klein Lecture.

Carneiro, P., and J. J. Heckman (2002): "The Evidence on Credit Constraints in Post-Secondary Schooling," *Economic Journal*, 112(482), 705–734.

Carranza, J. E. (2007): "Product innovation and adoption in market equilibrium: The case of digital cameras," Unpublished manuscript, University of Wisconsin-Madison, Department of Economics.

Cawley, J., K. Conneely, J. J. Heckman, and E. J. Vytlacil (1997): "Cognitive Ability, Wages and Meritocracy," in *Intelligence, Genes, and Success : Scientists Respond to The Bell Curve*, ed. by B. Devlin, S. Fienberg, D. Resnick, and K. Roeder, pp. 179–192. Springer, New York.

Cawley, J., J. J. Heckman, L. J. Lochner, and E. J. Vytlacil (2000): "Understanding the Role of Cognitive Ability in Accounting for the Recent Rise in the Return to Education," in *Meritocracy and Economic Inequality*, ed. by K. Arrow, and S. Bowles. Princeton University Press, Princeton, NJ.

Cawley, J., J. J. Heckman, and E. J. Vytlacil (2001): "Three Observations on Wages and Measured Cognitive Ability," *Labour Economics*, 8(4), 419–442.

Chioda, L. (2004): "Estimating Euler Equations with Measurement Error: A Nonparametric Approach," Unpublished Manuscript, University of California Berkeley, Department of Economics.

Cunha, F., J. J. Heckman, and S. Navarro (2005): "Separating Uncertainty from Heterogeneity in Life Cycle Earnings, The 2004 Hicks Lecture," *Oxford Economic Papers*, 57(2), 191–261.

——— (2006): "Counterfactual Analysis of Inequality and Social Mobility," in *Mobility and Inequality: Frontiers of Research in Sociology and Economics*, ed. by S. L. Morgan, D. B. Grusky, and G. S. Fields, chap. 4, pp. 290–348. Stanford University Press, Stanford, CA.

Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998): "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," *Journal of Human Resources*, 33(2), 251–299.

Flavin, M. A. (1981): "The Adjustment of Consumption to Changing Expectations About Future Income," *Journal of Political Economy*, 89(5), 974–1009.

Gourinchas, P.-O., and J. Parker (2002): "Consumption over the LifeCycle," *Econometrica*, 70(1), 47–89.

Guvenen, F. (2007): "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?," *The American Economic Review*, 97(3), 687–712.

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.

Hansen, L. P., and K. J. Singleton (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journalof Political Economy*, 91(2), 249–265.

Heckman, J. J. (2005): "The Scientific Model of Causality," *Sociological Methodology*, 35(1), 1–97.

Heckman, J. J., L. J. Lochner, and C. Taber (1998a): "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," *Review of Economic Dynamics*, 1(1), 1–58.

——— (1998b): "General-Equilibrium Treatment Effects: A Study of Tuition Policy," *American Economic Review*, 88(2), 381–386.

——— (1999): "General-Equilibrium Cost-Benefit Analysis of Education and Tax Policies," in *Trade, Growth and Development: Essays in Honor of T.N. Srinivasan*, ed. by G. Ranis, and L. Raut, chap. 14, pp. 291–349. Elsevier Science B.V.

Heckman, J. J., R. L. Matzkin, S. Navarro, and S. Urzua (2004): "Nonseparable Factor Analysis," Unpublished manuscript, Department of Economics, University of Chicago.

Heckman, J. J., and S. Navarro (2007): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136(2), 341–396.

Heckman, J. J., and E. J. Vytlacil (1998): "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return Is Correlated with Schooling," *Journal of Human Resources*, 33(4), 974–987.

Hotz, V. J., and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60(3), 497–529.

Kane, T. J. (1996): "College Cost, Borrowing Constraints and the Timing of College Entry," *Eastern Economic Journal*, 22(2), 181–194.

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

——— (2001): "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment," *International Economic Review*, 42(4), 1051–1103.

Kotlarski, I. I. (1967): "On Characterizing the Gamma and Normal Distribution," *Pacific Journal of Mathematics*, 20, 69–76.

Laitner, J. (1992): "Random Earnings Differences, Lifetime Liquidity Constraints, and Altruistic Intergenerational Transfers," *Journal of Economic Theory*, 58(2), 135–170.

MaCurdy, T. E., T. Mroz, and R. M. Gritz (1998): "An Evaluation of the National Longitudinal Survey on Youth," *Journal of Human Resources*, 33(2), 345–436.

Magnac, T., and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70(2), 801–816.

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

——— (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339–1375.

Park, S. (2004): "Quantitative Analysis of Network Externalities in Competing Technologies: The VCR Case," *Review of Economics and Statistics*, 86(4), 937–945.

Pistaferri, L. (2001): "Superior Information, Income Shocks, and the Permanent Income Hypothesis," *Review of Economics and Statistics*, 83(3), 465–476.

Roehrig, C. S. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56(2), 433–447.

Sims, C. A. (1972): "Money, Income, and Causality," *American Economic Review*, 62(4), 540–552.

Taber, C. R. (2000): "Semiparametric Identification and Heterogeneity in Discrete Choice Dynamic Programming Models," *Journal of Econometrics*, 96(2), 201–229.

——— (2001): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?," *Review of Economic Studies*, 68(3), 665–691.

Ventura, E. (1994): "A Note on Measurement Error and Euler Equations," *Economics Letters*, 45, 305 – 308.