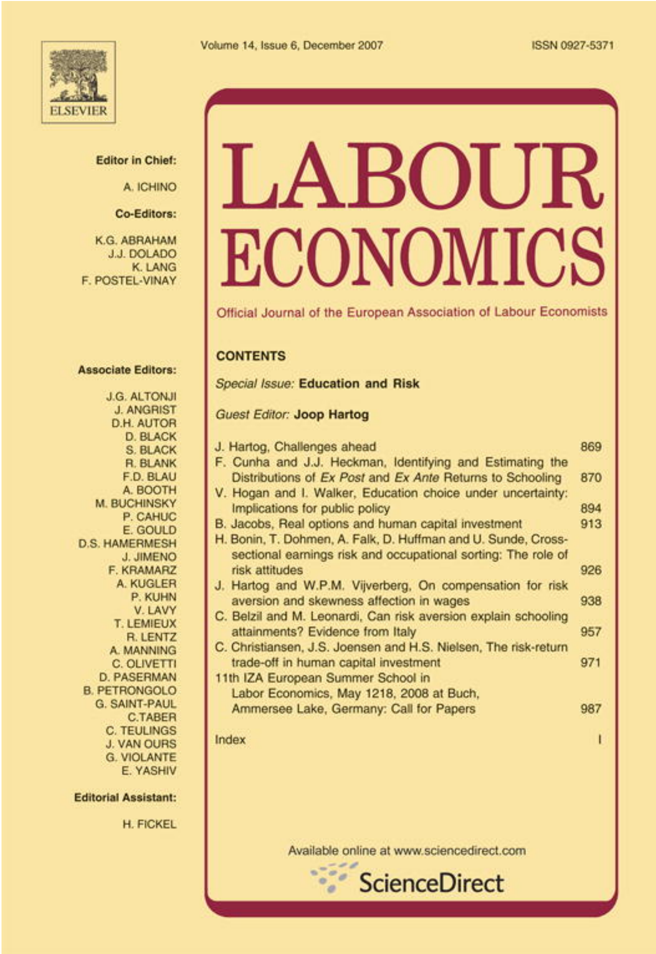

Volume 14, Issue 6, December 2007 ISSN 0927-5371

ELSEVIER

LABOUR ECONOMICS

Official Journal of the European Association of Labour Economists

CONTENTS

Special Issue: **Education and Risk**

Guest Editor: **Joop Hartog**

Available online at www.sciencedirect.com

ScienceDirect

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Labour Economics 14 (2007) 870–893
LABOUR ECONOMICS
www.elsevier.com/locate/econbase

Identifying and Estimating the Distributions of *Ex Post* and *Ex Ante* Returns to Schooling☆

Flavio Cunha [a], James J. Heckman [b,c,*]

[a] University of Pennsylvania, United States
[b] University of Chicago, United States
[c] University College Dublin, Ireland

Received 21 February 2006; received in revised form 18 May 2007; accepted 5 June 2007
Available online 19 June 2007

Abstract

This paper surveys a recent body of research by Carneiro, Hansen, and Heckman [Carneiro, P., K. Hansen, and J.J. Heckman, 2001, Fall. Removing the veil of ignorance in assessing the distributional impacts of social policies. Swedish Economic Policy Review 8 (2), 273–301., Carneiro, P., K. Hansen, and J.J. Heckman, 2003, May. 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], Cunha and Heckman [Cunha, F. and J.J. Heckman, 2006. The evolution of earnings risk in the US economy. Presented at the 9th World Congress of the Econometric Society, London], Cunha, Heckman, and Navarro [Cunha, F., J.J. Heckman, and S. Navarro, 2004, March. Separating heterogeneity from uncertainty in an aiyagari–laitner economy. Presented at the Goldwater Conference on Labor Markets, Arizona., Cunha, F., J.J. Heckman, and S. Navarro, 2005, April. Separating uncertainty from heterogeneity in life cycle earnings, The 2004 Hicks Lecture. Oxford Economic Papers 57 (2), 191–261., Cunha, F., J.J. Heckman, and S. Navarro, 2006. Counterfactual analysis of inequality and social mobility. In S.L. Morgan, D.B. Grusky, and G.S. Fields (Eds.), Mobility and Inequality: Frontiers of Research in Sociology and Economics, Chapter 4, pp. 290–348. Stanford, CA: Stanford University Press], Heckman and Navarro [Heckman, J.J. and S. Navarro, 2007, February. Dynamic discrete choice and dynamic treatment effects. Journal of Econometrics 136 (2), 341–396] and Navarro [Navarro, S., 2005. Understanding Schooling: Using Observed Choices to Infer Agent's Information in a Dynamic Model of Schooling Choice When Consumption Allocation is Subject to

☆ This research was supported by NIH R01-HD043411, NSF SES-024158 and the Geary Institute, University College Dublin, Ireland. The views expressed in this paper are those of the authors and not necessarily those of the funders listed here. We wish to thank the editor and two anonymous referees, as well as Lars Hansen, Lance Lochner, Salvador Navarro, Robert Townsend, Sergio Urzua, and Petra Todd for helpful comments.

* Corresponding author. University of Chicago, Department of Economics, 1126 E. 59th Street, Chicago IL 60637 United States. Tel.: +1 773 702 0634.

E-mail address: jheckman@uchicago.edu (J.J. Heckman).

0927-5371/$ - see front matter
doi:10.1016/j.labeco.2007.06.002

Borrowing Constraints. Ph.D. Dissertation, University of Chicago, Chicago, IL] that identifies and estimates the *ex post* distribution of returns to schooling and determines *ex ante* distributions of returns on which agents base their schooling choices. We discuss methods and evidence, and state a fundamental identification problem concerning the separation of preferences, market structures and agent information sets. For a variety of market structures and preference specifications, we estimate that over 50% of the *ex post* variance in returns to college are forecastable at the time agents make their schooling choices.

JEL classification: C31
Keywords: *Ex Ante* Returns to Schooling; *Ex Post* Returns to Schooling; Agent Information Sets

1. Introduction

The literature on the returns to schooling attempts to estimate the *ex post* rate of return. *Ex post* returns are interesting historical facts that describe how economies reward schooling. *Ex ante* returns are, however, what agents act on. To explain choices and evaluate their optimality, it is necessary to know what is in the agent's information set in order to determine the *ex ante* rate of return.

This paper describes new methods developed to estimate *ex ante* returns to schooling. We describe methods that characterize what is in the agent's information set at the time schooling decisions are made. The literature surveyed in this paper exploits the key idea that if agents know something and use that information in making their schooling decisions, it will affect their schooling choices. With panel data on earnings we can measure realized outcomes and assess what components of those outcomes are known at the time schooling choices are made.

The literature on panel data earnings dynamics (e.g. Lillard and Willis, 1978; MaCurdy, 1982) is not designed to estimate what is in agent information sets. It estimates earnings equations of the following type:

$$Y_{i,t} = X_{i,t}\beta + S_i\tau + U_{i,t}, \tag{1}$$

where $Y_{i,t}$, $X_{i,t}$, S_i, and $U_{i,t}$ denote (for person i at time t) the realized earnings, observable characteristics, educational attainment, and unobservable characteristics, respectively, from the point of view of the observing economist. The variables generating outcomes realized at time t may or may not have been known to the agents at the time they made their schooling decisions. Many economists mistakenly equate their ignorance about the $U_{i,t}$ with what the agents they study know about it.

The error term $U_{i,t}$ is often decomposed into two or more components. For example, it is common to specify that

$$U_{i,t} = \phi_i + \delta_{i,t}. \tag{2}$$

The term ϕ_i is a person-specific effect. The error term $\delta_{i,t}$ is often assumed to follow an ARMA (p, q) process (see Hause, 1980; MaCurdy, 1982) such as $\delta_{i,t}=\rho\delta_{i,t-1}+m_{i,t}$, where $m_{i,t}$ is a mean zero innovation independent of $X_{i,t}$ and the other error components. The components $X_{i,t}$, ϕ_i, and $\delta_{i,t}$ all contribute to measured *ex post* variability across persons. However, the literature is silent about the difference between heterogeneity or variability among persons from the point of view of the observer economist and uncertainty, the unforecastable part of earnings as of a given age. The literature on income mobility and on inequality measures all variability *ex post* as in Chiswick (1974), Mincer (1974) and Chiswick and Mincer (1972).

An alternative specification of the error process postulates a factor structure for earnings,

$$U_{i,t} = \theta_i \alpha_t + \varepsilon_{i,t}, \tag{3}$$

where θ_i is a vector of skills (e.g., ability, initial human capital, motivation, and the like), α_t is a vector of skill prices, and the $\varepsilon_{i,t}$ are mutually independent mean zero shocks independent of θ_i. Hause (1980) and Heckman and Scheinkman (1987) analyze such earnings models. Any process in the form of Eq. (2) can be written in terms of (3). The latter specification is more directly interpretable as a pricing equation than is (2).

The predictable components of $U_{i,t}$ will have different effects on choices and economic welfare than the unpredictable components, if people are risk averse and cannot fully insure against uncertainty. Statistical decompositions based on (1), (2), and (3) or versions of them describe *ex post* variability but tell us nothing about which components of (1), (2), or (3) are forecastable by agents *ex ante*. Is ϕ_i unknown to the agent? $\delta_{i,t}$? Or $\phi_i+\delta_{i,t}$? Or $m_{i,t}$? In representation (3), the entire vector θ_i, components of the θ_i, the $\varepsilon_{i,t}$, or all of these may or may not be known to the agent at the time schooling choices are made.

The methodology developed in Carneiro, Hansen, and Heckman (2003), Cunha, Heckman, and Navarro (2004, 2005) and Cunha and Heckman (2006) provides a framework within which it is possible to identify components of life cycle outcomes that are forecastable and acted on at the time decisions are taken from ones that are not. In order to choose between high school and college, agents forecast future earnings (and other returns and costs) for each schooling level. Using information about an educational choice at the time the choice is made, together with the *ex post* realization of earnings and costs that are observed at later ages, it is possible to estimate and test which components of future earnings and costs are forecast by the agent. This can be done provided we know, or can estimate, the earnings of agents under both schooling choices and provided we specify the market environment under which they operate as well as their preferences over outcomes.

For market environments where separation theorems are valid, so that consumption decisions are made independently of wealth maximizing decisions, it is not necessary to know agent preferences to decompose realized earnings outcomes in this fashion. Carneiro, Hansen, and Heckman (2003), Cunha, Heckman, and Navarro (2004, 2005) and Cunha and Heckman (2006) use choice information to extract *ex ante*, or forecast, components of earnings and to distinguish them from realized earnings for different market environments. The difference between forecast and realized earnings allows them to identify the distributions of the components of uncertainty facing agents at the time they make their schooling decisions.

2. A Generalized Roy Model

To state the problem addressed in the recent literature more precisely, consider a version of the generalized Roy (1951) economy with two sectors.[1] Let S_i denote different schooling levels. $S_i=0$ denotes choice of the high school sector for person i, and $S_i=1$ denotes choice of the college sector. Each person chooses to be in one or the other sector but cannot be in both. Let the two potential outcomes be represented by the pair $(Y_{0,i}, Y_{1,i})$, only one of which is observed by the analyst for any agent. Denote by C_i the direct cost of choosing sector 1, which is associated with

[1] See Heckman (1990) and Heckman and Smith (1998) for discussions of the generalized Roy model. In this paper we assume only two schooling levels for expositional simplicity, although our methods apply more generally.

choosing the college sector (e.g., tuition and non-pecuniary costs of attending college expressed in monetary values).

$Y_{1,i}$ is the *ex post* present value of earnings in the college sector, discounted over horizon T for person i, assumed for convenience to be zero,

$$Y_{1,i} = \sum_{t=0}^{T} \frac{Y_{1,i,t}}{(1+r)^t},$$

and $Y_{0,i}$ is the *ex post* present value of earnings in the high school sector at age zero,

$$Y_{0,i} = \sum_{t=0}^{T} \frac{Y_{0,i,t}}{(1+r)^t},$$

where r is the one-period risk-free interest rate. $Y_{1,i}$ and $Y_{0,i}$ can be constructed from time series of *ex post* potential earnings streams in the two states: $(Y_{0,i,0}, \ldots, Y_{0,i,T})$ for high school and $(Y_{1,i,0}, \ldots, Y_{1,i,T})$ for college. A practical problem with constructing both $Y_{0,i}$ and $Y_{1,i}$ is that we observe at most one or the other of these streams. This partial observability creates a fundamental identification problem that can be solved using the methods described in Heckman, Lochner, and Todd (2006), Abbring and Heckman (in press), and the references they cite.

The variables $Y_{1,i}$, $Y_{0,i}$, and C_i are *ex post* realizations of returns and costs, respectively. At the time agents make their schooling choices, these may be only partially known to the agent. Let $\mathcal{I}_{i,0}$ denote the information set of agent i at the time the schooling choice is made, which is time period $t=0$ in our notation. If agents act as if they are risk neutral, the decision rule governing sectoral choices at decision time "0" is

$$S_i = \begin{cases} 1, & \textit{if } E(Y_{1,i} - Y_{0,i} - C_i | \mathcal{I}_{i,0}) \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Under perfect foresight, the postulated information set $\mathcal{I}_{i,0}$ includes $Y_{1,i}$, $Y_{0,i}$, and C_i. Agents can use decision rule (4) if there is no uncertainty (and they are free to lend and borrow), if there is full insurance (complete contingent claims markets exist) or if the agents are risk neutral (have linear utility functions) and can lend and borrow freely.[2]

The decision rule is more complicated in the absence of full risk diversifiability and depends on the curvature of utility functions, the availability of markets to spread risk, and possibilities for storage. (See Cunha and Heckman, 2006, and Navarro, 2005.) In these more realistic economic settings, the components of earnings and costs required to forecast the gain to schooling depend on higher moments than the mean. In this paper we use a simple choice model to motivate the identification analysis of other environments analyzed elsewhere (Carneiro et al., 2003; Cunha et al., 2004).

Suppose that we seek to determine $\mathcal{I}_{i,0}$. This is a difficult task. Typically we can only partially identify $\mathcal{I}_{i,0}$ and generate a list of candidate variables that belong in the information set. Usually, we can only estimate the distributions of the unobservables in $\mathcal{I}_{i,0}$ (from the standpoint of the econometrician) across individuals and not person-specific information sets (the random variables

[2] If there are aggregate sources of risk, full insurance would require a linear utility function.

agents actually know). Before describing the analysis of Cunha, Heckman, and Navarro, we consider how this question might be addressed in the linear-in-the-parameters Card (2001) model.

We use the Card model as a familiar and convenient starting point within which to make some basic points. No criticism of the Card model is intended. Card does not seek to distinguish *ex ante* from *ex post* mean returns. Depending on the choice of instruments, one can estimate either or neither.

3. Identifying Information Sets in Card's Model of Schooling

Consider decomposing the "returns" coefficient on schooling in an earnings equation into components that are known at the time schooling choices are made and components that are not known. Write the log of annualized discounted lifetime earnings of person i as

$$\ln y_i = \alpha + \rho_i S_i + U_i, \tag{5}$$

where ρ_i is the person-specific *ex post* return, S_i is years of schooling, and U_i is a mean zero unobservable. We seek to decompose ρ_i into two components $\rho_i = \eta_i + \nu_i$, where η_i is a component known to the agent when he/she makes schooling decisions and ν_i is revealed after the choice is made. Schooling choices are assumed to depend on what is known to the agent at the time decisions are made, $S_i = \lambda\ (\eta_i, Z_i, \tau_i)$, where the Z_i are other observed determinants of schooling known to the agent and τ_i represents additional factors unobserved by the analyst but known to the agent. Both of these variables are assumed to be in the agent's information set at the time schooling choices are made. We seek to determine what components of *ex post* lifetime earnings Y_i enter the schooling choice equation.

If η_i is known to the agent and acted on, it enters the schooling choice equation. Even if it is known, it may not be acted on. If it is not known *or* acted on, it does not enter the schooling equation. Thus, if agents do not act on the information they know, any method that uses choices to infer what is in agent information sets overstates the amount of uncertainty. Component ν_i and any measurement errors in $Y_{1,i}$ or $Y_{0,i}$ should not be determinants of schooling choices. Neither should future skill prices that are unknown at the time agents make their decisions. Determining the correlation between realized Y_i and schooling choices based on *ex ante* forecasts enables economists to identify components known to agents and acted on in making their schooling decisions. Even if we cannot identify ρ_i, η_i, or ν_i for each person, under conditions specified in this paper, we might be able to identify their distributions.

If we correctly specify the variables that enter the outcome equation (X) and the variables in the choice equation (Z) that are known to the agent at the time schooling choices are made, local instrumental variable estimates of the Marginal Treatment Effect identify *ex ante* gross returns.[3] Any dependence between the variables in the schooling equation and returns arises from information known to the agent at the time schooling choices are made. If the econometrician's conditioning set is misspecified by using information on X and Z that accumulates after schooling choices are made and that predicts realized earnings (but not *ex ante* earnings), the estimated return is an *ex post* return relative to that information set. Thus, it is important to specify the conditioning set correctly to obtain the appropriate *ex ante* return. The question is how to pick the information set. We consider this problem in the context of the Card model, which, as previously noted, was designed only to estimate *ex post* returns.

[3] See Heckman, Urzua, and Vytlacil (2006) for a discussion of local instrumental variables.

3.1. The Card Model

Card presents a version of the Mincer (1974) model, which writes log earnings for person i with schooling level S_i as

$$\ln y_i = \alpha_i + \rho_i S_i, \tag{6}$$

where the "rate of return" ρ_i varies among persons as does the intercept, α_i. For the purposes of this discussion think of y_i as an annualized flow of lifetime earnings.[4] Let $\alpha_i = \bar{\alpha} + \varepsilon_{\alpha_i}$ and $\rho_i = \bar{\rho} + \varepsilon_{\rho_i}$ where $\bar{\alpha}$ and $\bar{\rho}$ are the means of α_i and ρ_i. Thus the means of ε_{α_i} and ε_{ρ_i} are zero. Earnings Eq. (6) can be written as

$$\ln y_i = \bar{\alpha} + \bar{\rho}\, S_i + \{\varepsilon_{\alpha_i} + \varepsilon_{\rho_i} S_i\}. \tag{7}$$

Allowing for ρ_i to be correlated with S_i (so S_i is correlated with ε_{ρ_i}) raises substantial problems that have just begun to be addressed in a systematic fashion in the recent literature. Card's (2001) random coefficient model of the growth rate of earnings with schooling is derived from economic theory and is based on the analysis of Rosen (1977).[5] We consider conditions under which it is possible to estimate mean *ex ante* returns in his model.

Card's model generalizes Rosen's (1977) model to allow for psychic costs of schooling. Assuming person-specific interest rate r_i, he obtains optimal schooling as

$$S_i = \frac{(\rho_i - r_i)}{k}, \tag{8}$$

where k is related to the curvature of psychic costs in schooling.[6]

In the Card model, ρ_i would be a rate of return if there were no direct costs of schooling and everyone faces a constant borrowing rate. In this model, ρ_i is the person-specific growth rate of earnings and overstates the true rate of return if there are direct and psychic costs of schooling.[7] Since schooling depends on ρ_i and r_i, any covariance between $\rho_i - r_i$ (in the schooling equation) and ρ_i (in the earnings function) produces a random coefficient model. Least squares will not estimate the mean growth rate of earnings with schooling $E(\rho_i)$ unless $\mathrm{Cov}(\rho_i, \rho_i - r_i) = 0$.

Define for this model $\rho_i = \eta_i + v_i$. The cost is r_i. Suppose r_i depends on observables (Z_i) and unobservables (ε_i) in the following fashion:

$$r_i = \gamma_0 + \gamma_1 Z_i + \varepsilon_i,$$

where ε_i has mean zero and is assumed to be independent of Z_i. If we are uncertain about which components of ρ_i enter the schooling equation, we may rewrite (8) as

$$S_i = \lambda_0 + \lambda_1 \eta_i + \lambda_2 v_i + \lambda_3 Z_i + \tau_i, \tag{9}$$

[4] Unless the only costs of schooling are earnings foregone, and markets are perfect, ρ_i is a percentage growth rate in earnings with schooling and not a rate of return to schooling. See Heckman, Lochner, and Todd (2006) for conditions under which a Mincer coefficient is a rate of return.

[5] Random coefficient models with coefficients correlated with the regressors are systematically analyzed in Heckman and Robb (1985, 1986). See also Heckman and Vytlacil (1998). They originate in labor economics with the work of Lewis (1963). Heckman and Robb analyze training programs but their analysis applies to estimating the returns to schooling.

[6] It is necessary to assume the absence of diminishing market returns to schooling to obtain Mincer specification (6).

[7] See Heckman, Lochner, and Todd (2006).

where $\lambda_0 = -\frac{\gamma_0}{k}, \lambda_1 = \frac{1}{k}, \lambda_2 = \frac{1}{k}$ if v_i is in the information set at the time schooling choices are taken and $\lambda_2=0$ otherwise. The remaining coefficients are $\lambda_3 = -\frac{\gamma_1}{k}$ and $\tau_i = -\frac{\varepsilon_i}{k}$.

Suppose that we observe the cost of funds, r_i, and assume that r_i is independent of (ρ_i, α_i). This assumes that the costs of schooling are independent of the "return" ρ_i and the payment to raw ability, α_i. Suppose that agents do not know ρ_i at the time they make their schooling decisions but instead know $E(\rho_i)=\bar{\rho}$.[8] If agents act on the expected return to schooling, decisions are given by

$$S_i = \frac{\bar{\rho} - r_i}{k}$$

and *ex post* earnings observed after schooling are

$$\ln y_i = \bar{\alpha} + \bar{\rho}\, S_i + \{(\alpha_i - \bar{\alpha}) + (\rho_i - \bar{\rho})S_i\}.$$

In the notation introduced in the Card model, $\eta_i=\bar{\rho}$ and $v_i=\rho_i-\bar{\rho}$. $\lambda_2=0$ in Eq. (9) and $\lambda_1 = \frac{1}{k}$.

In this case,

$$Cov(\ln y, S) = \bar{\rho}\, Var(S)$$

because $(\rho_i-\bar{\rho})$ is independent of S_i. Note that, under this information assumption, $(\bar{\alpha}, \bar{\rho})$ can be identified by least squares because $S_i \perp\!\!\!\perp [(\alpha_i-\bar{\alpha}), (\rho_i-\bar{\rho})\, S_i]$ where "$\perp\!\!\!\perp$" denotes independence.

If, on the other hand, agents know ρ_i at the time they make their schooling decisions, OLS breaks down for identifying $\bar{\rho}$ because ρ_i is correlated with S_i. We can identify $\bar{\rho}$ and the distribution of ρ_i using the method of instrumental variables. Under our assumptions, r_i is a valid instrument for S_i.

In this case

$$Cov(\ln y, S) = \bar{\rho}\, Var(S) + Cov(S(\rho - \bar{\rho})S).$$

Since we observe S and r_i, we can identify $\bar{\rho}$ and can construct the value of $(\rho-\bar{\rho})$ associated with each S, we can form both terms on the right hand side. Under the assumption that agents do not know ρ but forecast it by $\bar{\rho}$, ρ is independent of S so we can test for independence directly. In this case the second term on the right hand side is zero and does not contribute to the explanation of $Cov(\ln y, S)$. Note further that a Durbin (1954)-Wu (1973)-Hausman (1978) test can be used to compare the OLS and IV estimates, which should be the same under the model that assumes that ρ_i is not known at the time schooling decisions are made and that agents base their choice of schooling on $E(\rho_i)=\bar{\rho}$. If the economist does not observe r_i, but instead observes determinants of r_i that are exogenous, it is still possible to conduct a Durbin-Wu-Hausman test to discriminate between the two hypotheses, but one cannot form $Cov(\rho, S)$ directly. This shows that, provided one has a good instrument, it is possible to test for the information in the agent's information set. However, the method is somewhat fragile.

If we add selection bias to the Card model (so $E(\alpha \mid S)$ depends on S, something ruled out up to this point), we can identify $\bar{\rho}$ by IV (Heckman and Vytlacil, 1998), but OLS is no longer consistent for $\bar{\rho}$ even if, in making their schooling decisions, agents forecast ρ_i using $\bar{\rho}$. Selection bias could occur, for example, if fellowship aid is given on the basis of raw ability, which

[8] This is a rational expectations assumption. Under rational expectations, the mean *ex ante* return is the same as the mean *ex post* return, but the distributions of these returns may be very different.

presumably affects the level (α_i) of the earnings equation.[9] In this case, the Durbin-Wu-Hausman test is not helpful in assessing what is in the agent's information set.

Even ignoring selection bias, in the case where r_i is not observed, so that $E(\alpha \mid S)$ does not depend on S, the proposed testing approach based on the Durbin-Wu-Hausman test breaks down if we misspecify the information set. Thus if we include the predictors of r_i that predict *ex post* gains ($\rho_i - \bar{\rho}$) and are correlated with S_i, we do not identify $\bar{\rho}$. The Durbin-Wu-Hausman test is not informative on the stated question. For a familiar example, if local labor market variables proxy the opportunity cost of school (the r_i), and also predict the realization of *ex post* earnings ($\rho_i - \bar{\rho}$), they are invalid instruments. The question of determining the appropriate information set is front and center and unfortunately cannot, in general, be inferred using IV methods and standard model specification tests.

The method developed by Cunha, Heckman, and Navarro (2004, 2005) and Cunha and Heckman (2006) exploits the covariance between S and the realized $\ln(y)$ to determine which components of $\ln(y)$ are known at the time schooling decisions are made. Covariance restrictions and not IV exclusions secure identification. Their approach explicitly models selection bias and allows for measurement error in earnings. It does not rely on linearity of the schooling relationship in terms of $\rho - r$. Their method recognizes the discrete nature of the schooling decision. We reiterate that our analysis is not intended as a criticism of the Card model, which does not address the question raised in this paper, but clarifies the limitations of an IV approach for agent information sets.

4. The Method of Cunha, Heckman and, Navarro

Cunha, Heckman, and Navarro (2004, 2005, henceforth CHN) and Cunha and Heckman (2006), exploit covariances between schooling and realized earnings that arise under different agent information structures to test which information structure characterizes the data. They build on the analysis of Carneiro, Hansen, and Heckman (2003). To see how the method works, simplify the model back to two schooling levels: $S_i = 1$ (college); $S_i = 0$ (high school). Heckman and Navarro (2007) extend this analysis to multiple schooling levels.

Suppose, contrary to what is possible, that the analyst observes $Y_{0,i}$, $Y_{1,i}$, and C_i. In the rest of the paper, we work with the present value of earnings in high school ($Y_{0,i}$) and college ($Y_{1,i}$). Such information would come from an ideal data set in which we could observe two different lifetime earnings streams for the same person in high school and in college as well as the costs they pay for attending college. From such information, we could construct $Y_{1,i} - Y_{0,i} - C_i$. If we knew the information set $\mathcal{I}_{i,0}$ of the agent that governs schooling choices, we could also construct $E(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0})$. Under the correct model of expectations, we could form the residual

$$V_{\mathcal{I}_{i,0}} = (Y_{1,i} - Y_{0,i} - C_i) - E(Y_{1,i} - Y_{0,i} - C_i | \mathcal{I}_{i,0}),$$

and from the *ex ante* college choice decision, we could determine whether S_i depends on $V_{\mathcal{I}_{i,0}}$. It should not if we have specified $\mathcal{I}_{i,0}$ correctly. Analogous to the model of Eqs. (5) and (9), if there are no direct costs of schooling, $E(Y_{1,i} - Y_{0,i} \mid \mathcal{I}_{i,0})$ corresponds to η_i, and $V_{\mathcal{I}_{i,0}}$ corresponds to v_i.

A test for correct specification of candidate information set $\widetilde{\mathcal{I}}_{i,0}$ is a test of whether S_i depends on $V_{\widetilde{\mathcal{I}}_{i,0}}$, where $V_{\widetilde{\mathcal{I}}_{i,0}} = (Y_{1,i} - Y_{0,i} - C_i) - E(Y_{1,i} - Y_{0,i} - C_i | \widetilde{\mathcal{I}}_{i,0})$. More precisely, the information set is valid if $S_i \perp\!\!\!\perp V_{\widetilde{\mathcal{I}}_{i,0}} \mid \widetilde{\mathcal{I}}_{i,0}$. ($A \perp\!\!\!\perp B \mid C$ means A is independent of B given C). In terms of the log linear schooling model of Eqs. (5) and (9), the analogous condition says that v_i should not enter the

[9] In addition, ability might affect $\bar{\rho}$ if agents know their own ability.

schooling choice equation ($\lambda_2=0$). A test of misspecification of $\widetilde{\mathcal{I}}_{i,0}$ is a test of whether the coefficient of $V_{\widetilde{\mathcal{I}}_{i,0}}$ is statistically significantly different from zero in the schooling choice equation.

More generally, $\widetilde{\mathcal{I}}_{i,0}$ is the correct information set if $V_{\widetilde{\mathcal{I}}_{i,0}}$ does not predict schooling. One can search among candidate information sets $\widetilde{\mathcal{I}}_{i,0}$ to determine which ones satisfy the requirement that the generated $V_{\widetilde{\mathcal{I}}_{i,0}}$ does not predict S_i and what components of $Y_{1,i}-Y_{0,i}-C_i$ (and $Y_{1,i}-Y_{0,i}$) are predictable at the age schooling decisions are made for the specified information set.[10] There may be several information sets that satisfy this property.[11] For a properly specified $\widetilde{\mathcal{I}}_{i,0}$, $V_{\widetilde{\mathcal{I}}_{i,0}}$ should not predict schooling choices. The components of $V_{\widetilde{\mathcal{I}}_{i,0}}$ that are unpredictable are called intrinsic components of uncertainty, as defined in this paper.

It is difficult to determine the exact content of $\mathcal{I}_{i,0}$ known to each agent. If we could, we would perfectly predict S_i given our decision rule. More realistically, we might find variables that proxy $\mathcal{I}_{i,0}$ or the distribution of variables in the agent's information set. Thus, in the model of Eqs. (5) and (9) we would seek to determine the distribution of v_i and the allocation of the variance of ρ_i to η_i and v_i rather than trying to estimate ρ_i, η_i, or v_i for each person. We now review the strategy developed in Cunha, Heckman, and Navarro (2005, 2006) for a two-choice model of schooling that is generalized by Cunha and Heckman (2006) and Heckman and Navarro (2007).

4.1. An Approach Based on Factor Structures

We develop a simple linear-in-parameters model for T periods. Write the earnings of agent i in each counterfactual state as

$$\begin{aligned} Y_{0,i,t} &= X_{i,t}\beta_{0,t} + U_{0,i,t}, \\ Y_{1,i,t} &= X_{i,t}\beta_{1,t} + U_{1,i,t}, \quad t = 0, \ldots, T. \end{aligned}$$

Let costs of attending college be represented by

$$C_i = Z_i\gamma + U_{i,C}.$$

Assume that the life cycle of the agent ends after period T. Linearity of outcomes in terms of parameters is convenient but not essential to the method of CHN.

Suppose that there exists a vector of factors $\theta_i=(\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,L})$ such that $\theta_{i,k}$ and $\theta_{i,j}$ are mutually independent random variables for $k, j=1, \ldots, L$, $k\neq j$. They represent the error term in earnings at age t for agent i in the following manner:

$$\begin{aligned} U_{0,i,t} &= \theta_i\alpha_{0,t} + \varepsilon_{0,i,t}, \\ U_{1,i,t} &= \theta_i\alpha_{1,t} + \varepsilon_{1,i,t}, \end{aligned}$$

where $\alpha_{0,t}$ and $\alpha_{1,t}$ are vectors and θ_i is a vector distributed independently across persons. The $\varepsilon_{0,i,t}$ and $\varepsilon_{1,i,t}$ are mutually independent of each other and independent of the θ_i. We can also decompose the cost function C_i in a similar fashion:

$$C_i = Z_i\gamma + \theta_i\alpha_C + \varepsilon_{i,C}.$$

[10] This procedure is a Sims (1972) version of a Wiener–Granger causality test.

[11] Thus different combinations of variables may contain the same information. The issue of the existence of a smallest information set is a technical one concerning the existence of a minimum σ-algebra that satisfies the condition on $\mathcal{I}_{i,0}$.

The essential statistical feature of the factor vector θ is that it captures all of the dependence across *all* unobservables in the model.

All of the statistical dependence across potential outcomes and costs is assumed to be generated by θ, X and Z. Thus, if we could condition on θ, X, and Z, all outcome variables would be independent. We could use the method of matching to infer the distribution of counterfactuals and capture all of the dependence across the counterfactual states through the θ_i. However, in general, CHN allow for the possibility that not all of the required elements of θ_i are observed.

The parameters α_C and $\alpha_{s,t}$ for $s=0, 1$, and $t=0, \ldots, T$ are called factor loadings. $\varepsilon_{i,C}$ is independent of the θ_i and the other ε components. In this notation, the choice equation can be written as:

$$
\begin{aligned}
S_i^* &= E\left(\sum_{t=0}^{T} \frac{Y_{1,i,t}-Y_{0,i,t}}{(1+r)^t} - C_i \middle| \mathcal{I}_{i,0}\right) \\
&= E\left(\sum_{t=0}^{T} \frac{(X_{i,t}\beta_{1,t}+\theta_i\alpha_{1,t}+\varepsilon_{1,i,t})-(X_{i,t}\beta_{0,t}+\theta_i\alpha_{0,t}+\varepsilon_{0,i,t})}{(1+r)^t} - (Z_i\gamma+\theta_i\alpha_C+\varepsilon_{iC}) \middle| \mathcal{I}_{i,0}\right); \\
S_i &= 1 \text{ if } S^*\geq 0; S_i = 0 \text{ otherwise.}
\end{aligned} \tag{10}
$$

The term inside the parentheses is the discounted earnings of agent i in college minus the discounted earnings of the agent in high school. We observe $Y_{1,i,t}$, $t=0, \ldots, T$ if $S_i=1$ and $Y_{0,i,t}$, $t=0, \ldots, T$ if $S_i=0$. In the case of perfect certainty, this is the model of Willis and Rosen (1979), which is an application of the Roy model (1951) as developed by Gronau (1974) and Heckman (1974, 1976).

Choice Eq. (10) is based on a counterfactual comparison. Even if earnings in one schooling level are observed over the lifetime using panel data, the earnings in the counterfactual state are not. After the schooling choice is made, some components of the $X_{i,t}$, the θ_i, and the $\varepsilon_{i,t}$ may be revealed (e.g., unemployment rates, macro shocks) to both the observing economist and the agent, although different components may be revealed to each and at different times. For this reason, application of IV even in the linear schooling model of the previous section is problematic. If the wrong information set is used, the IV method will not identify the true *ex ante* mean returns.

Examining alternative information sets, one can determine which ones produce models for outcomes that fit the data best in terms of producing a model that predicts date $t=0$ schooling choices and at the same time passes the CHN test for misspecification of predicted earnings and costs. Some components of the error terms may be known or not known at the date schooling choices are made. The unforecastable components are intrinsic uncertainty as CHN define it. The forecastable information is called heterogeneity.[12]

To formally characterize the CHN empirical procedure, it is useful to use some notation from linear algebra. Let $\odot$ denote the Hadamard product ($a \odot b=(a_1b_1, \ldots, a_Lb_L)$) for vectors a and b of length L. This product is a componentwise multiplication of vectors of the same length that produces a vector of the same length. Let Δ_{X_t}, $t=0, \ldots, T$, Δ_Z, Δ_θ, Δ_{ε_t}, Δ_{ε_C}, denote coefficient vectors associated with the X_t, $t=0, \ldots, T$, the Z, the θ, the $\varepsilon_{1,t}-\varepsilon_{0,t}$, $t=0, \ldots, T$, and the ε_C, respectively. These coefficients will be estimated to be nonzero in a schooling choice equation if a proposed information set is not the actual information set used by agents.

[12] The term 'heterogeneity' is somewhat unfortunate. Under this term, CHN include trends common across all people (e.g., macrotrends). The real distinction they are making is between components of realized earnings forecastable by agents at the time they make their schooling choices vs. components that are not forecastable.

For a proposed information set $\widetilde{\mathcal{I}}_{i,0}$ which may or may not be the true information set on which agents act, the proposed choice index $\widetilde{S}_i^*$ is broken down into forecastable and unforecastable components with the unforecastable components associated with $\odot$.

$$\begin{aligned}
\widetilde{S}_i^* = {} & \underbrace{\left[\sum_{t=0}^{T}\frac{E(X_{i,t}|\widetilde{\mathcal{I}}_{i,0})}{(1+r)^t}(\beta_{1,t}-\beta_{0,t})\right]}_{\textit{forecastable under } \widetilde{\mathcal{I}}_{i,0}} + \underbrace{\sum_{t=0}^{T}\frac{[X_{i,t}-E(X_{i,t}|\widetilde{\mathcal{I}}_{i,0})]}{(1+r)^t}(\beta_{1,t}-\beta_{0,t})\odot\Delta_{X_t}}_{\textit{unforecastable under } \widetilde{\mathcal{I}}_{i,0}} \\
& \underbrace{+E\left(\theta_i|\widetilde{\mathcal{I}}_{i,0}\right)\left[\sum_{t=0}^{T}\frac{(\alpha_{1,t}-\alpha_{0,t})}{(1+r)^t}-\alpha_C\right]}_{\textit{forecastable under } \widetilde{\mathcal{I}}_{i,0}} \\
& \underbrace{+\left[\theta_i-E(\theta_i|\widetilde{\mathcal{I}}_{i,0})\right]\left\{\left[\sum_{t=0}^{T}\frac{\alpha_{1,t}-\alpha_{0,t}}{(1+r)^t}-\alpha_C\right]\odot\Delta_\theta\right\}}_{\textit{unforecastable under } \widetilde{\mathcal{I}}_{i,0}} \\
& +\underbrace{\frac{\sum_{t=0}^{T}E(\varepsilon_{1,i,t}-\varepsilon_{0,i,t}|\widetilde{\mathcal{I}}_{i,0})}{(1+r)^t}}_{\textit{forecastable under } \widetilde{\mathcal{I}}_{i,0}} + \underbrace{\sum_{t=0}^{T}\frac{[(\varepsilon_{1,i,t}-\varepsilon_{0,i,t})-E(\varepsilon_{1,i,t}-\varepsilon_{0,i,t}|\widetilde{\mathcal{I}}_{i,0})]}{(1+r)^t}\Delta_{\varepsilon_t}}_{\textit{unforecastable under } \widetilde{\mathcal{I}}_{i,0}} \\
& -\underbrace{E(Z_i|\widetilde{\mathcal{I}}_{i,0})\gamma}_{\textit{forecastable under } \widetilde{\mathcal{I}}_{i,0}} - \underbrace{[Z_i-E(Z_i|\widetilde{\mathcal{I}}_{i,0})]\gamma\odot\Delta_z}_{\textit{unforecastable under } \widetilde{\mathcal{I}}_{i,0}} \\
& -\underbrace{E(\varepsilon_{iC}|\widetilde{\mathcal{I}}_{i,0})}_{\textit{forecastable under } \widetilde{\mathcal{I}}_{i,0}} - \underbrace{[\varepsilon_{iC}-E(\varepsilon_{iC}|\widetilde{\mathcal{I}}_{i,0})]\Delta_{\varepsilon_C}}_{\textit{unforecastable under } \widetilde{\mathcal{I}}_{i,0}}.
\end{aligned} \tag{11}$$

In this expression, if candidate information set $\widetilde{\mathcal{I}}_{i,0}$ is correctly specified ($\mathcal{I}_{i,0}=\widetilde{\mathcal{I}}_{i,0}$), then each term with a Δ_j would have a zero effect in a schooling equation — i.e., components with Δ_j would not affect schooling. If the variable being predicted is in the agent's information set, the deviation between the actual and the predicted values of each variable would be zero and clearly cannot affect schooling. Thus, if $X_{i,t}$ is in the information set, $X_{i,t}-E\left(X_{i,t}|\widetilde{\mathcal{I}}_{i,0}\right)=0$. If some components of $X_{i,t}$ are not in the agent's information set, so that $X_{i,t}-E\left(X_{i,t}|\widetilde{\mathcal{I}}_{i,0}\right)\neq 0$, the coefficient $\Delta_{X_t}=0$ in a schooling equation. A similar analysis can be applied term-by-term to each argument of the schooling choice equation. Our analysis can be extended to apply to the coefficients in (10) and (11). In the Gorman (1980)-Lancaster (1966) theory of attributes, the coefficients can be interpreted as prices and the agents can be envisioned as predicting future prices. To simplify the analysis we do not make this extension in this paper.

To conduct their test, CHN fit a schooling choice model based on the proposed model (11). They estimate the parameters of the model including the Δ_j parameters associated with the candidate unforecastable components. A test of no misspecification of information set $\widetilde{\mathcal{I}}_{i,0}$ is a joint test of the hypothesis that all of the Δ_j are zero. That is, when $\widetilde{\mathcal{I}}_{i,0}=\mathcal{I}_{i,0}$ the proposed choice index $\widetilde{S}_i^*=S_i^*$. In a correctly specified model, the components associated with zero Δ_j are the unforecastable elements or the elements that, even if known to the agent, are not acted on in making schooling choices. To operationalize the test, we must take a position on agent expectations, i.e., what is in the agent's information set and how agents forecast future variables.

4.2. Operationalizing the Test

To illustrate how to operationalize the method of CHN, assume for simplicity that the $X_{i,t}$, the Z_i, the $\varepsilon_{i,C}$, the $\beta_{1,t}$, $\beta_{0,t}$, the $\alpha_{1,t}$, $\alpha_{0,t}$, and α_C are known to the agent, and the $\varepsilon_{j,i,t}$ are unknown and are set at their mean zero values ($E(\varepsilon_{j,i,t} \mid \mathcal{I}_{i,0})=0$). We show how to relax these assumptions below. Agents may know some or all of the components of θ_i. We can infer which components of the θ_i are known and acted on in making schooling decisions if we postulate that some components of θ_i are known perfectly at date $t=0$ while others are not known at all, and their forecast values have mean zero given $\mathcal{I}_{i,0}$.

If there is an element of the vector θ_i, say $\theta_{i,2}$ (factor 2), that has nonzero "loadings" (the α coefficients) in the schooling choice equation and a nonzero loading on one or more potential future earnings, then one can say that at the time the schooling choice is made, the agent knows the unobservable captured by factor 2 that affects future earnings. If $\theta_{i,2}$ does not enter the schooling choice equation but explains future earnings, then $\theta_{i,2}$ is unknown (not predictable by the agent) at the age schooling decisions are made. An alternative interpretation is that the second component of $\left[\sum_{t=0}^{T} \frac{(\alpha_{1,t}-\alpha_{0,t})}{(1+r)^t} - \alpha_C\right]$ is zero, i.e., that even if the component is known, it is not acted on. CHN can only test for what the agent knows and acts on. This means that if agents do not act on information, the method of CHN will understate it.

Contrary to what we assumed, if there are components of the $\varepsilon_{j,i,t}$ that are predictable at age $t=0$, they will induce additional dependence between S_i and future earnings not captured by the initially specified factors. CHN allow for this dependence by introducing new factors that capture the dependence and that appear in the agent information set. Thus, if some of the components of $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ are known to the agent at the date schooling decisions are made and enter (11), then additional dependence between S_i and future $Y_{1,i}-Y_{0,i}$ due to the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ would be estimated. This is captured by adding additional factors to the model. These components of $\varepsilon_{0,i,t}$ and $\varepsilon_{1,i,t}$ become factors. The CHN procedure can be generalized to test whether all of the components of (11) are in the agent's information set. With it, the analyst can test the predictive power of each subset of the overall possible information set at the date the schooling decision is being made.[13,14] The CHN approach allows the analyst to determine which components of θ_i (and $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$) are known and acted on at the time schooling decisions are made. It allows us to test among candidate information sets.

[13] This test has been extended to a nonlinear setting, allowing for credit constraints, preferences for risk, and the like. See Cunha, Heckman, and Navarro (2004) and Navarro (2005).

[14] A similar but distinct idea motivates the Flavin (1981) test of the permanent income hypothesis and her measurement of unforecastable income innovations. She picks a particular information set $\widetilde{\mathcal{I}}_{i,0}$ (permanent income constructed from an assumed ARMA (p, q) time series process for income, where she estimates the coefficients given a specified order of the AR and MA components) and tests if $V_{\widetilde{\mathcal{I}}_{i,0}}$ (our notation) predicts consumption. Her test of 'excess sensitivity' can be interpreted as a test of the correct specification of the ARMA process that she assumes generates $\widetilde{\mathcal{I}}_{i,0}$ which is unobserved (by the economist), although she does not state it that way. Blundell and Preston (1998) and Blundell, Pistaferri, and Preston (2004) extend her analysis but, like her, maintain an *a priori* specification of the stochastic process generating $\mathcal{I}_{i,0}$. Blundell, Pistaferri, and Preston (2004) claim to test for 'partial insurance.' In fact their procedure can be viewed as a test of their specification of the stochastic process generating the agent's information set. More closely related to the analysis of CHN is the analysis of Pistaferri (2001), who uses the distinction between expected starting wages (to measure expected returns) and realized wages (to measure innovations) in a consumption analysis. Hansen (1987) shows that Flavin's model is observationally equivalent to a complete markets model.

Statistical decompositions of earnings equations do not tell us which components of (3) are known at the time agents make their schooling decisions. A model of expectations and schooling is needed. Alternative models can be tested.

The contrast between the sources generating realized earnings outcomes and the sources generating dependence between S_i and realized earnings is the essential idea in the analysis of CHN. The method can be generalized to deal with nonlinear preferences and imperfect market environments.[15] A central issue, discussed next, is how far one can go in identifying income information processes without specifying preferences, insurance, and market environments.

5. More general preferences and market settings

To focus on the main ideas in the literature, we have used the simple market structures of complete contingent claims markets. What can be identified in more general environments? In the absence of perfect certainty or perfect risk sharing, preferences and market environments also determine schooling choices. The separation theorem allowing consumption and schooling decisions to be analyzed in isolation that has been used thus far breaks down.

If we postulate information processes *a priori*, and assume that preferences are known up to some unknown parameters as in Flavin (1981), Blundell and Preston (1998) and Blundell, Pistaferri, and Preston (2004), we can identify departures from specified market structures. Cunha, Heckman, and Navarro (2004) postulate an Aiyagari (1994)-Laitner (1992) economy with one asset and parametric preferences to identify the information processes in the agent's information set. They take a parametric position on preferences and a nonparametric position on the economic environment and the information set.

An open question, not yet fully resolved in the literature, is how far one can go in non-parametrically jointly identifying preferences, market structures and information sets. Cunha, Heckman, and Navarro (2004) add consumption data to the schooling choice and earnings data to secure identification of risk preference parameters (within a parametric family) and information sets, and to test among alternative models for market environments. Alternative assumptions about what analysts know produce different interpretations of the same evidence. The lack of full

[15] In a model with complete autarky with preferences Ψ, ignoring costs,

$$I_i = \sum_{t=0}^{T} E\left[\frac{\Psi(X_{i,t}\beta_{1,t} + \theta_i\alpha_{1,t} + \varepsilon_{1,i,t}) - \Psi(X_{i,t}\beta_{0,t} + \theta_i\alpha_{0,t} + \varepsilon_{0,i,t})}{(1+\rho)^t}\middle| \widetilde{\mathcal{I}}_{i,0}\right],$$

where ρ is the time rate of discount, we can make a similar decomposition but it is more complicated given the nonlinearity in Ψ. For this model we could perform a Sims noncausality test where

$$\begin{aligned} V_{\widetilde{\mathcal{I}}_{i,0}} &= \sum_{t=0}^{T} \frac{\Psi(X_{i,t}\beta_{1,t} + \theta_i\alpha_{1,t} + \varepsilon_{1,i,t}) - \Psi(X_{i,t}\beta_{0,t} + \theta_i\alpha_{0,t} + \varepsilon_{0,i,t})}{(1+\rho)^t} \\ &\quad - \sum_{t=0}^{T} E\left[\frac{\Psi(X_{i,t}\beta_{1,t} + \theta_i\alpha_{1,t} + \varepsilon_{1,i,t}) - \Psi(X_{i,t}\beta_{0,t} + \theta_i\alpha_{0,t} + \varepsilon_{0,i,t})}{(1+\rho)^t}\middle| \widetilde{\mathcal{I}}_{i,0}\right]. \end{aligned}$$

This requires some specification of Ψ. See Carneiro, Hansen, and Heckman (2003), who assume $\Psi(Y) = \ln Y$ and that the equation for $\ln Y$ is linear in parameters. Cunha, Heckman, and Navarro (2004) and Navarro (2005) generalize that framework to a model with imperfect capital markets where some lending and borrowing is possible.

insurance interpretation given to the empirical results by Flavin (1981) and Blundell, Pistaferri, and Preston (2004) may be a consequence of their misspecification of the generating process for the agent's information set. We now present some evidence on *ex ante* vs. *ex post* returns based on the analysis of Cunha and Heckman (2006) that uses the framework of Section 4.

6. Evidence on Uncertainty and Heterogeneity of Returns

Few data sets contain the full life cycle of earnings along with the test scores and schooling choices needed to directly estimate the CHN model and extract components of uncertainty. It is necessary to pool data sets. See CHN who discuss how to combine NLSY and PSID data sets. We summarize the analysis of Cunha and Heckman (2006) in this subsection. See their paper for their exclusions and identification conditions.

Following the preceding theoretical analysis, they consider only two schooling choices: high school and college graduation.[16] For simplicity and familiarity, we focus on their results based on complete contingent claims markets. Because they assume that all shocks are idiosyncratic and the operation of complete markets, schooling choices are made on the basis of expected present value income maximization. Carneiro, Hansen, and Heckman (2003) assume the absence of any credit markets or insurance. Navarro (2005) checks whether empirical findings about components of income inequality are robust to different assumptions about the availability of credit markets and insurance markets. He estimates an Aiyagari–Laitner economy with a single asset and borrowing constraints and discuss risk aversion and the relative importance of uncertainty. We summarize the evidence from alternative assumptions about market structures below.

6.1. Identifying Joint Distributions of Counterfactuals and the Role of Costs and Ability as Determinants of Schooling

Suppose that the error term for $Y_{s,t}$ is generated by a two factor model, so that we may write the outcome equation as

$$Y_{s,t} = X\beta_{s,t} + \theta_1\alpha_{s,t,1} + \theta_2\alpha_{s,t,2} + \varepsilon_{s,t}. \tag{12}$$

We omit the "i" subscripts to eliminate notational burden. Cunha and Heckman (2006) report that two factors are all that is required to fit the pooled PSID and NLSY data that they use.

They use a test score system of K ability tests:

$$A_k = X_A\omega_k + \theta_1\alpha_k + \varepsilon_k, \qquad k = 1, \ldots, K. \tag{13}$$

Abbring and Heckman (in press) and Cunha and Heckman (2007) show that test scores are not required to implement the method. Earnings data of sufficient length will suffice. Other proxies for θ_1 can be used. An advantage of using test scores is that it helps set the scale of θ_1 in an interpretable metric. In system (13), factor 1 is identified as an ability component. The cost function C is specified by:

$$C = Z\gamma + \theta_1\alpha_{C,1} + \theta_2\alpha_{C,2} + \varepsilon_C. \tag{14}$$

[16] Heckman and Navarro (2007) present a model with multiple schooling levels.

Table 1
Ex Ante Conditional Distributions for the NLSY/1979 (College Earnings Conditional on High School Earnings) $\Pr(d_i<Y_c<d_i+1|d_j<Y_h<d_j+1)$ where d_i is the ith decile of the College Lifetime *Ex Ante* Earnings Distribution and d_j is the jth decile of the High School *Ex Ante* Lifetime Earnings Distribution

High School	College									
	1	2	3	4	5	6	7	8	9	10
1	0.1833	0.1631	0.1330	0.1066	0.0928	0.0758	0.0675	0.0630	0.0615	0.0535
2	0.1217	0.1525	0.1262	0.1139	0.1044	0.0979	0.0857	0.0796	0.0683	0.0498
3	0.1102	0.1263	0.1224	0.1198	0.1124	0.0970	0.0931	0.0907	0.0775	0.0506
4	0.0796	0.1083	0.1142	0.1168	0.1045	0.1034	0.1121	0.1006	0.0953	0.0652
5	0.0701	0.0993	0.1003	0.1027	0.1104	0.1165	0.1086	0.1112	0.1043	0.0768
6	0.0573	0.0932	0.1079	0.1023	0.1110	0.1166	0.1130	0.1102	0.1059	0.0825
7	0.0495	0.0810	0.0950	0.1021	0.1101	0.1162	0.1202	0.1174	0.1134	0.0950
8	0.0511	0.0754	0.0770	0.1006	0.1006	0.1053	0.1244	0.1212	0.1297	0.1147
9	0.0411	0.0651	0.0841	0.0914	0.1039	0.1117	0.1162	0.1216	0.1442	0.1206
10	0.0590	0.0599	0.0622	0.0645	0.0697	0.0782	0.0770	0.1028	0.1181	0.3087

Individual fixes known θ at their means, so Information Set$=\{\theta_1=0,\ \theta_2=0\}$.
Corrrelation$(Y_C,Y_H)=0.4083$.

Cunha and Heckman (2006) assume that agents know the coefficients of the model and X, Z, ε_C and some, but not necessarily all, components of θ. Let the components known to the agent be $\bar{\theta}$. The decision rule for attending college is based on

$$S^* = E\left(\sum_{t=0}^{T}\frac{Y_{1,t}-Y_{0,t}}{(1+r)^t}|X,\bar{\theta}\right) - E\left(C|Z,X,\bar{\theta},\varepsilon_C\right), \tag{15}$$

where $S=1$ $(S^*\geq 0)$. Cunha and Heckman (2006) report evidence that the estimated factors are highly nonnormal.[17] They present the coefficient estimates and discuss how to construct the counterfactual distributions reported below. Abbring and Heckman (in press), Cunha and Heckman (2006), and Heckman, Lochner, and Todd (2006) present exact conditions for identification and estimation of this model. Here, we just report estimates based on their approach.

Table 1 presents the conditional distribution of *ex post* potential college earnings given *ex post* potential high school earnings, decile by decile, as reported by Cunha and Heckman (2006). The table displays a positive dependence between the relative positions of individuals in the two distributions.[18] The dependence is far from perfect. For example, about 10% of those who are at the first decile of the high school distribution would be in the fourth decile of the college distribution. Note that this comparison is not made in terms of positions in the overall distribution of earnings. CHN, unlike Willis and Rosen (1979), identify joint distributions of potential outcomes and can determine where individuals are located in the distribution of population potential high school earnings and the distribution of potential college earnings, although in the

[17] They assume that each factor $k \in \{1, 2\}$ is generated by a mixture of J_k normal distributions,

$$\theta_k \sim \sum_{j=1}^{J_k} p_{kj}\phi(\theta_k|\mu_{kj},\tau_{kj}),$$

where the $p_{k,j}$ are the weights on the normal components.

[18] Willis and Rosen (1979) do not identify this joint distribution.

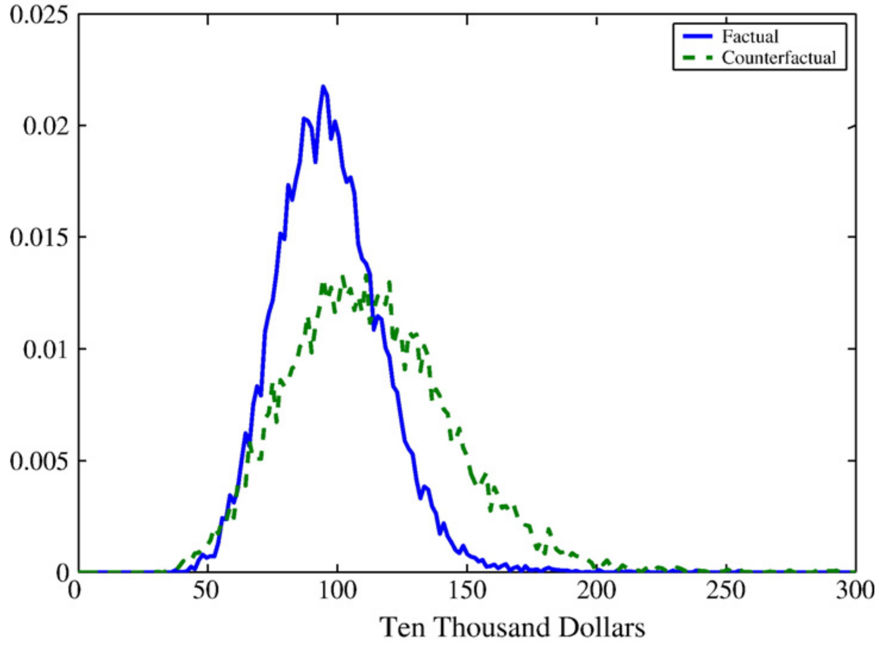

Fig. 1. Densities of present value of lifetime earnings for High School Graduates. Factual and Counterfactual NLSY/1979 Sample. Present Value of Lifetime Earnings from age 18 to 65 for high school graduates using a discount rate of 3%. Let Y_0 denote present value of earnings in high school sector. Let Y_1 denote present value of earnings in college sector. In this graph we plot the factual density function $f(y_0 \mid S=0)$ (the solid line), against the counterfactual density function $f(y_1 \mid S=0)$. We use kernel density estimation to smooth these functions.

data we only observe them in either one or the other state. Their evidence shows that the assumption of perfect dependence across components of counterfactual distributions that is maintained in much of the recent literature (e.g. Juhn et al., 1993) is far too strong.

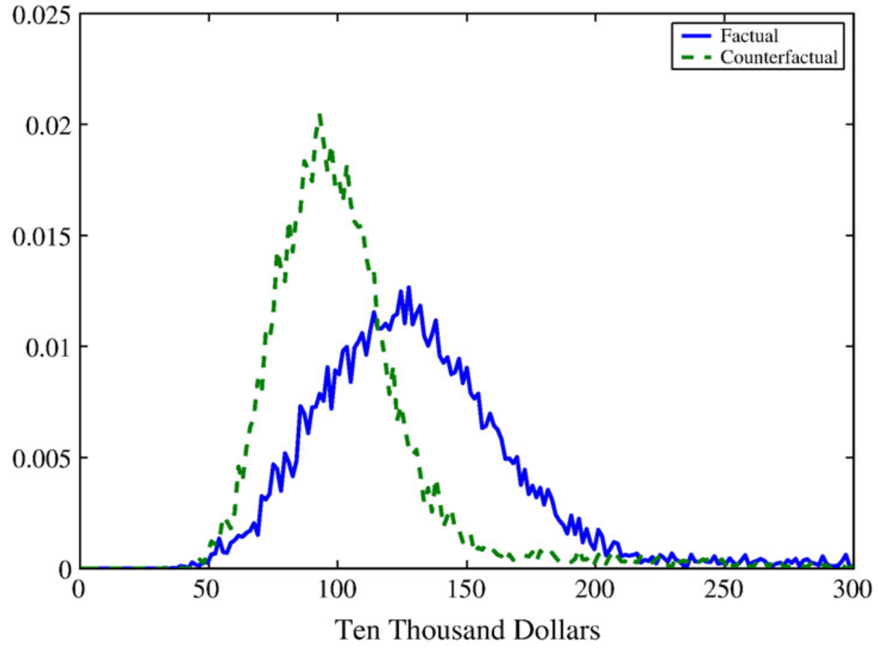

Fig. 2. Densities of present value of earnings for College Graduates, Factual and Counterfactual. NLSY/1979 Sample. Present Value of Lifetime Earnings from age 18 to 65 for college graduates using a discount rate of 3%. Let Y_0 denote present value of earnings in high school sector. Let Y_1 denote present value of earnings in college sector. In this graph we plot the counterfactual density function $f(y_0 \mid S=1)$ (the dashed line), against the factual density function $f(y_1 \mid S=1)$. We use kernel density estimation to smooth these functions.

Table 2
Average present value of *ex post* earnings[1] for high school graduates. Fitted and Counterfactual.[2] White males from NLSY79

	High School (Fitted)	College (Counterfactual)	Returns[3]
Average	968.5100	1125.7870	0.2055
Std. Err.	*7.9137*	*9.4583*	*0.0113*

[1]Thousands of dollars. Discounted using a 3% interest rate.
[2]The counterfactual is constructed using the estimated college outcome equation applied to the population of persons selecting high school.
[3]As a fraction of the base state, i.e., (PVearnings(Col)-PVearnings(HS))/PVearnings(HS).

Fig. 1 presents the marginal densities of predicted (actual) earnings for high school students and counterfactual college earnings for actual high school students. When we compare the densities of present value of earnings in the college sector for persons who choose college against the counterfactual densities of college earnings for high school graduates we can see that many high school graduates would earn more as college graduates. In Fig. 2 we repeat the exercise, this time for college graduates.

Table 2 from Cunha and Heckman (2006) reports the fitted and counterfactual present value of earnings for agents who choose high school. The typical high school student would earn \$968.51 thousand dollars over the life cycle. She would earn \$1,125.78 thousand if she had chosen to be a college graduate.[19] This implies a return of 20% to a college education over the whole life cycle (i.e., a monetary gain of \$157.28 thousand dollars). In Table 3, from Cunha and Heckman (2006), the typical college graduate earns \$1,390.32 thousand dollars (above the counterfactual earnings of what a typical high school student would earn in college), and would make only \$1,033.72 thousand dollars over her lifetime if she chose to be a high school graduate instead. The returns to college education for the typical college graduate (which in the literature on program evaluation is referred to as the effect of Treatment on the Treated) is around 38% above that of the return for a high school graduate.

Fig. 3 plots the density of *ex post* gross returns to education excluding direct costs and psychic costs for agents who are high school graduates (the solid curve), and the density of returns to education for agents who are college graduates (the dashed curve). In reporting our estimated returns, CH follow conventions in the literature on the returns to schooling and present growth rates in terms of present values, and not true rates of return. Thus they ignore option values. These figures report the growth rates in present values $\left(\frac{PV(1)-PV(0)}{PV(0)}\right)$ where “1” and “0” refer to college and high school and all present values are discounted to a common benchmark level. Tuition and psychic costs are ignored. College graduates have returns distributed somewhat to the right of high school graduates, so the difference is not only a difference for the mean individual but is actually present over the entire distribution. An economic interpretation of Fig. 3 is that agents who choose a college education are the ones who tend to gain more from it.

With their methodology, CH can also determine returns to the marginal student. Table 4 reveals that the average individual who is just indifferent between a college education and a high school diploma earns \$976.04 thousand dollars as a high school graduate or \$1,208.26 thousand dollars as a college graduate. This implies a return of 28%. The returns to people at the margin are above those of the typical high school graduate, but below those for the typical college graduate. Since persons at the margin are more likely to be affected by a policy that encourages college

[19] These numbers may appear to be large but are a consequence of using a 3% discount rate.

Table 3
Average present value of *ex post* earnings[1] for college graduates. Fitted and Counterfactual.[2] White males from NLSY79

	High School (Counterfactual)	College (fitted)	Returns[3]
Average	1033.721	1390.321	0.374
Std. Err.	*14.665*	*30.218*	*0.280*

[1]Thousands of dollars. Discounted using a 3% interest rate.
[2]The counterfactual is constructed using the estimated high school outcome equation applied to the population of persons selecting college.
[3]As a fraction of the base state, i.e., (PVearnings(Col)-PVearnings(HS))/PVearnings(HS).

attendance, their returns are the ones that should be used in order to compute the marginal benefit of policies that induce people into schooling.

6.2. Ex ante and Ex post Returns: Heterogeneity versus Uncertainty

Figs. 4–6, from Cunha and Heckman (2006) separate the effect of heterogeneity from uncertainty in earnings. The figures plot the distribution of *ex ante* and *ex post* outcomes. The information set of the agent is $\mathcal{I}=\{X, Z, X_T, \varepsilon_C, \Theta\}$, Θ contains some or all of the factors. In their papers, the various information sets consist of different components of θ. We first consider Fig. 4. It presents results for a variety of information sets. First assume that agents do not know their factors; consequently, $\Theta=\emptyset$. This is the case in which all of the unobservables are treated as unknown by the agent, and, as a result, the density has a large variance. If we assume that the agents know factor 1 but not factor 2,[20] so that $\Theta=\{\theta_1\}$, there is a reduction in the forecast variance, but it is small. Factor 1, which is associated with cognitive ability, is important for forecasting educational choices, but does not do a very good job in forecasting earnings. The third case is the one in which the agent knows both factors, which is the case we test and cannot reject. The agent is able to substantially reduce the forecast variance of earnings in high school. Note that the variance in this case is much smaller than in the other two cases. Fig. 6 reveals much the same story about the college earnings distribution.

Table 5 presents the variance of potential earnings in each state, and returns under different information sets available to the agent. We conduct this exercise for lifetime earnings, and report baseline variances and covariances without conditioning and state the remaining uncertainty as a fraction of the baseline no-information state variance when different components of θ are known to the agents. CH — who estimate a three factor model — show that both θ_1 and θ_2 are known to the agents at the time they make their schooling decisions, but not θ_3.

This discussion sheds light on the issue of distinguishing predictable heterogeneity from uncertainty. CH demonstrate that there is a large dispersion in the distribution of the present value of earnings. This dispersion is largely due to heterogeneity, which is forecastable by the agents at the time they are making their schooling choices. CH provide tests that determine that agents know θ_1 and θ_2. The remaining dispersion is due to luck, or uncertainty or unforecastable factors as of age 17. Its contribution is smaller.

It is interesting to note that knowledge of the factors enables agents to make better forecasts. Fig. 6 presents an exercise for returns to college (Y_1-Y_0) similar to that presented in Figs. 4 and 5 regarding information sets available to the agent. Knowledge of factor 2 also greatly improves the forecastability of returns. 56% of the variability in returns is forecastable at age 18. The levels also

[20] As opposed to the econometrician who never gets to observe either θ_1 or θ_2.

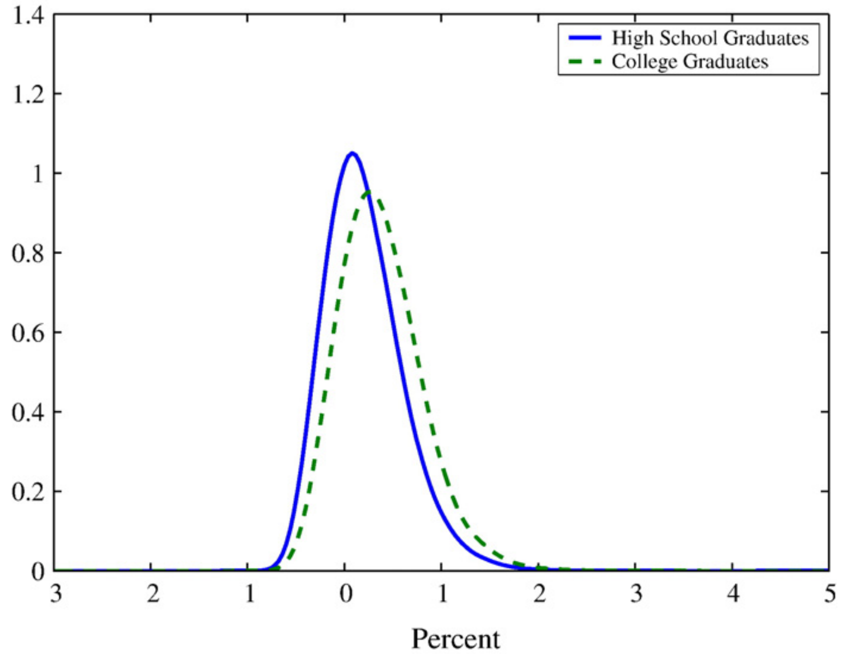

Fig. 3. Densities of Returns to College NLSY/1979 Sample. Let Y_0 denote present value of earnings in high school sector. Let Y_1 denote present value of earnings in college sector. Let $R=(Y_1-Y_0)/Y_0$ denote the gross rate of return to college. In this graph we plot the density function of the returns to college conditional on being a high school graduate, $f(r \mid S=0)$ (the solid line), against the density function of returns to college conditional on being a college graduate, $f(r \mid S=1)$. We use kernel density estimation to smooth these functions.

show high predictability (65% for high school; 56% for college). Most variability across people is due to heterogeneity and not uncertainty. However there is still a lot of variability in agent earnings after accounting for what is in the agent's information set. This is intrinsic uncertainty at the time agents make their schooling choices.

6.3. Ex Ante versus Ex Post

Once the distinction between heterogeneity and uncertainty is made, it is possible to be precise about the distinction between *ex ante* and *ex post* decision making. From their analysis, CH conclude that, at the time agents pick their schooling, the ε's in their earnings equations are unknown to them. These are the components that correspond to "luck." It is clear that decision making would be different, at least for some individuals, if the agent knew these chance components when choosing schooling levels, since the decision rule would be based on (4) where all components of $Y_{1,i}$, $Y_{0,i}$, and C_i are known, and no expectation need be taken.

If individuals could pick their schooling level using their *ex post* information (i.e., after learning their luck components in earnings) 13.81% of high school graduates would rather be

Table 4
Average present value of *ex post* earnings[1] for individuals at margin. Fitted and Counterfactual.[2] White males from NLSY79

	High School	College	Returns[3]
Average	976.04	1208.26	0.2828
Std. Err.	*21.503*	*33.613*	*0.0457*

[1]Thousands of dollars. Discounted using a 3% interest rate.
[2]It defines the result of taking a person at random from the population regardless of his schooling choice.
[3]As a fraction of the base state, i.e., (PVearnings(Col)-PVearnings(HS))/PVearnings(HS).

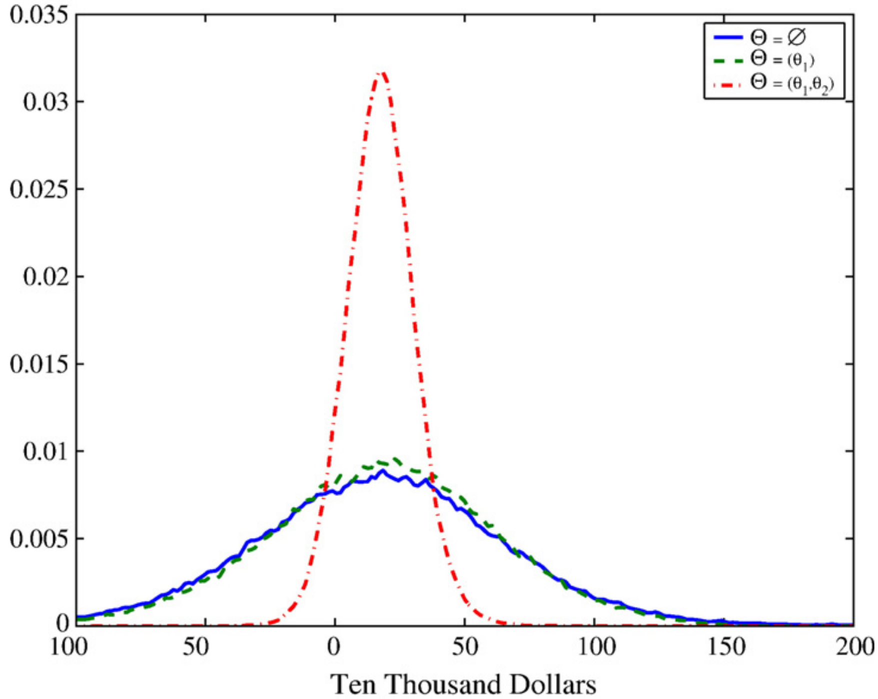

Fig. 4. Densities of present value of returns - NLSY/1979 under different information sets for the agent calculated for the entire population regardless of schooling choice. Let Θ denote the information set of the agent. Let Y_0 denote the present value of returns (discounted at a 3% interest rate). Let $f(y_0|\Theta)$ denote the density of Y_0 conditional on information set Θ. The solid line plots the density of Y_0 when $\Theta=\phi$. The dashed line plots the density of Y_0 when $\Theta=\{\theta_1\}$.The dotted and dashed line plots the density of of Y_0 when $\Theta=\{\theta_1,\theta_2\}$. The X variables are in the information set of the agent. The factors θ, when known, are evaluated at their mean, which is zero.

college graduates and 17.15% of college graduates would have stopped their schooling at the high school level. Using the estimated counterfactual distributions, it is possible to consider the effects of a variety of policy experiments on distributions of outcomes locating persons in pre- and post-policy distributions. They analyze, for example, how tuition subsidies move people from one quantile of a Y_0 distribution to another quantile of a Y_1 distribution. See Carneiro, Hansen, and Heckman (2001, 2003), Cunha, Heckman, and Navarro (2005, 2006) and Cunha and Heckman (2006) for examples of this work.

6.4. Comparison of Our Analysis with the Analysis of Willis and Rosen

The approach developed in the literature just surveyed extends the application of the Roy model to schooling by Willis and Rosen (1979) in two ways.[21] (1) Using the factor model, the recent literature identifies the joint distribution of counterfactual outcomes associated with different schooling states. It is possible to estimate the dependence in outcomes across counterfactual states, and hence to test for comparative advantage in the labor market and construct counterfactual distributions for new policies.[22] Willis and Rosen do not identify or estimate joint counterfactual distributions and cannot identify comparative advantage parameters, even though they discuss comparative advantage. (2) Willis and Rosen assume an environment of

[21] Earlier applications of the Roy model to labor economics are Heckman (1974, 1976) and Gronau (1974).
[22] See Cunha and Heckman (in press).

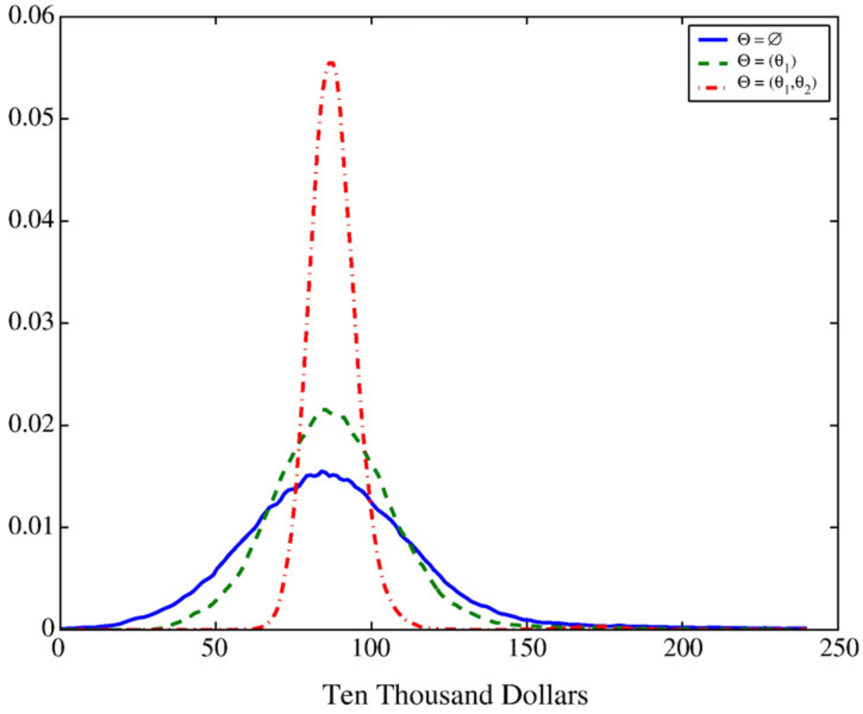

Fig. 5. Densities of present value of high school earnings - NLSY/1979 under different information sets for the agent calculated for the entire population regardless of schooling choice. Let Θ denote the information set of the agent. Let Y_0 denote the present value of earnings in the high school sector (discounted at a 3% interest rate). Let $f(y_0|\Theta)$ denote the density of Y_0 conditional on information set Θ. The solid line plots the density of Y_0 when $\Theta=\phi$. The dashed line plots the density of Y_0 when $\Theta=\{\theta_1\}$. The dotted and dashed line plots the density of of Y_0 when $\Theta=\{\theta_1,\theta_2\}$. The X variables are in the information set of the agent. The factors θ, when known, are evaluated at their mean, which is zero.

perfect certainty. The literature surveyed in this paper separates components of earnings realized after schooling is completed from components that are known to the agent at the time schooling choices are made. In this way, it is possible to separate *ex ante* earnings forecasts from *ex post* realizations, a distinction not made in the Willis–Rosen analysis.

7. Extensions and Alternative Specifications

Carneiro, Hansen, and Heckman (2003) estimate a version of the model just discussed for an environment of complete autarky. Individuals have to live within their means each period. Cunha, Heckman, and Navarro (2004) estimate a version of this model with restrictions on intertemporal trade as in the Aiyagari–Laitner economy. Different assumptions about credit markets and preferences produce a range of estimates of the proportion of the total variability of returns to schooling that are unforecastable, ranging from 37% (Carneiro et al., 2003) for a model with complete autarky and log preferences, to 53% (CHN) for complete markets, to 44% (Cunha and Heckman, 2006) for another complete market economy.

This line of work has just begun. It shows what is possible with rich panel data. The empirical evidence on the importance of uncertainty is not settled, but the range of estimates from alternative specifications is not large. Most of the papers developed within this research program suggest a substantial role for uncertainty in producing returns. Accounting for uncertainty and psychic costs may help to explain the sluggish response of schooling enrollment rates to rising returns to schooling that is documented in Ellwood, 2001 and Card and Lemieux (2001) because of the wedge between utility and money returns.

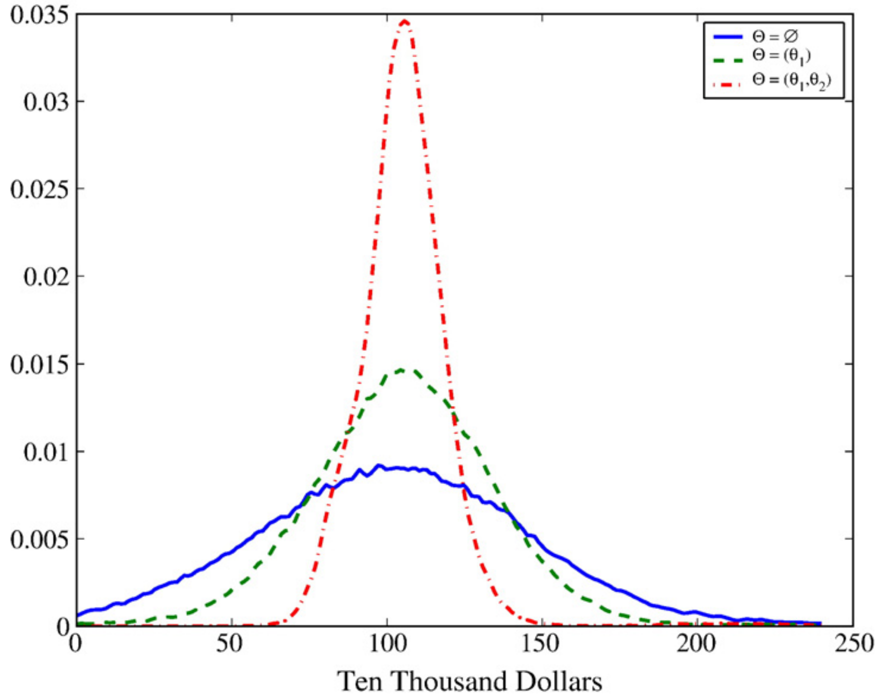

Fig. 6. Densities of present value of college earnings - NLSY/1979 under different information sets for the agent calculated for the entire population regardless of schooling choice. Let Θ denote the information set of the agent. Let Y_0 denote the present value of earnings in the college sector (discounted at a 3% interest rate). Let $f(y_0|\Theta)$ denote the density of Y_0 conditional on information set Θ. The solid line plots the density of Y_0 when $\Theta=\phi$. The dashed line plots the density of Y_0 when $\Theta=\{\theta_1\}$. The dotted and dashed line plots the density of of Y_0 when $\Theta=\{\theta_1,\theta_2\}$. The X variables are in the information set of the agent. The factors θ, when known, are evaluated at their mean, which is zero.

The analysis discussed in this paper is for one-shot models of schooling choice. In truth, schooling is a sequential decision process made with increasingly richer information sets at later stages of the choice process. Heckman and Navarro (2007) discuss identification of a semiparametric sequential schooling models based on the factor structures exposited in this paper. In their framework, it is possible in principle to test among alternative information structures about the arrival of information on the components of vector θ at different stages of the life cycle, but no empirical results are available from their approach at the time of this writing.[23]

8. Summary and Conclusions

This paper surveys the main models and methods developed in Carneiro, Hansen, and Heckman (2003), Cunha, Heckman, and Navarro (2004, 2005, 2006) and Cunha and Heckman (2006) for estimating models of heterogeneity and uncertainty in the returns to schooling. The goal of this work is to separate variability from uncertainty and to estimate the distributions of *ex ante* and *ex post* returns to schooling. The key idea in the recent literature is to exploit the relationship between realized earnings and schooling choice equations to determine which components of realized earnings are in agent information sets at the time they make their

[23] Keane and Wolpin (1997) develop a dynamic discrete choice model of schooling. In their setup they assume that all new information is serially independent, and hence they do not allow for updating of persistent components of agent information sets (i.e., in their setup agents learn about temporally independent shocks, but the shocks do not predict future earnings). Thus they cannot estimate serially correlated information updating.

Table 5
Agent's Forecast Variance of Present Value of Earnings[1] Under Different Information Sets - NLSY/1979

	Var(Y_c)	Var (Y_h)	Var($Y_c - Y_h$)
For lifetime: Total Residual Variance	290.84	103.13	334.02
Share of Total Variance due to Forecastable Components	65.13%	55.94%	56.04%
Share of Total Variance due to Unforecastable Components	34.87%	44.06%	43.94%

[1]We use a discount rate ρ of 3% to calculate the present value of earnings.

schooling decisions. For a variety of market environments and assumptions about preferences, a robust empirical regularity is that over 50% of the *ex post* variance in the returns to schooling are forecastable at the time students make their college choices.

References

Abbring, J.H., Heckman, J.J., in press. Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. Heckman and E. Leamer (Eds.), Handbook of Econometrics, Volume 6. Amsterdam: Elsevier.

Aiyagari, S.R., 1994. Uninsured idiosyncratic risk and aggregate saving. Quarterly Journal of Economics 109 (3), 659–684 (August).

Blundell, R., Preston, I., 1998. Consumption inequality and income uncertainty. Quarterly Journal of Economics 113 (2), 603–640 (May).

Blundell, R., Pistaferri, L., Preston, I., 2004. Consumption inequality and partial insurance. Technical Report WP04/28. Institute for Fiscal Studies. October.

Card, D., 2001. Estimating the return to schooling: Progress on some persistent econometric problems. Econometrica 69 (5), 1127–1160 (September).

Card, D., Lemieux, T., 2001. Can falling supply explain the rising return to college for younger men? A cohort-based analysis. Quarterly Journal of Economics 116 (2), 705–746 (May).

Carneiro, P., Hansen, K., Heckman, J.J., 2001. Removing the veil of ignorance in assessing the distributional impacts of social policies. Swedish Economic Policy Review 8 (2), 273–301 (Fall).

Carneiro, P., Hansen, K., Heckman, J.J., 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 (May).

Chiswick, B.R., 1974. Income Inequality: Regional Analyses Within a Human Capital Framework. National Bureau of Economic Research, New York.

Chiswick, B.R., Mincer, J., 1972. Time-series changes in personal income inequality in the United States from 1939, with projections to 1985. Journal of Political Economy 80 (3 (Part II)), S34–S66 (May/June).

Cunha, F., Heckman, J.J., 2006. The evolution of earnings risk in the US economy. Presented at the 9th World Congress of the Econometric Society, London.

Cunha, F., Heckman, J.J., 2007. The technology of skill formation. American Economic Review 97 (2), 31–47 (May).

Cunha, F., Heckman, J.J., in press. A framework for the analysis of inequality. Macroeconomic Dynamics.

Cunha, F., Heckman, J.J., Navarro, S., 2004. Separating heterogeneity from uncertainty in an Aiyagari–Laitner economy. Presented at the Goldwater Conference on Labor Markets, Arizona. March.

Cunha, F., Heckman, J.J., Navarro, S., 2005. Separating uncertainty from heterogeneity in life cycle earnings, The 2004 Hicks Lecture. Oxford Economic Papers 57 (2), 191–261 (April).

Cunha, F., Heckman, J.J., Navarro, S., 2006. Counterfactual analysis of inequality and social mobility. In: Morgan, S.L., Grusky, D.B., Fields, G.S. (Eds.), Mobility and Inequality: Frontiers of Research in Sociology and Economics, Chapter 4. Stanford University Press, Stanford, CA, pp. 290–348.

Durbin, J., 1954. Errors in variables. Review of the International Statistical Institute 22, 23–32.

Ellwood, D.T., 2001. The sputtering U.S. Labor Market. Kennedy School of Government, Harvard (Unpublished paper).

Flavin, M.A., 1981. The adjustment of consumption to changing expectations about future income. Journal of Political Economy 89 (5), 974–1009 (October).

Gorman, W.M., 1980. A possible procedure for analysing quality differentials in the egg market. Review of Economic Studies 47 (5), 843–856 (October).

Gronau, R., 1974. Wage comparisons — a selectivity bias. Journal of Political Economy 82 (6), 1119–1143 (November–December).

Hansen, L.P., 1987. Calculating Asset Prices. In: Bewley, T. (Ed.), Advances in Econometrics, Fifth World Congress, vol. 1, pp. 207–243.

Hause, J.C., 1980. The fine structure of earnings and the on-the-job training hypothesis. Econometrica 48 (4), 1013–1029 (May).

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46 (6), 1251–1272 (November).

Heckman, J.J., 1974. Shadow prices, market wages, and labor supply. Econometrica 42 (4), 679–694 (July).

Heckman, J.J. (1976, August). A life-cycle model of earnings, learning, and consumption. Journal of Political Economy 84 (4, Part 2), S11–S44. Journal Special Issue: Essays in Labor Economics in Honor of H. Gregg Lewis.

Heckman, J.J., 1990. Varieties of selection bias. American Economic Review 80 (2), 313–318 (May).

Heckman, J.J., Navarro, S., 2007. Dynamic discrete choice and dynamic treatment effects. Journal of Econometrics 136 (2), 341–396 (February).

Heckman, J.J., Robb, R., 1985. Alternative methods for evaluating the impact of interventions. In: Heckman, J., Singer, B. (Eds.), Longitudinal Analysis of Labor Market Data, vol. 10. Cambridge University Press, New York, pp. 156–245.

Heckman, J.J., Robb, R., 1986. Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In: Wainer, H. (Ed.), Drawing Inferences from Self-Selected Samples. Springer-Verlag, New York, pp. 63–107. Reprinted in 2000, Mahwah, NJ: Lawrence Erlbaum Associates.

Heckman, J.J., Scheinkman, J., 1987. The importance of bundling in a Gorman-Lancaster model of earnings. Review of Economic Studies 54 (2), 243–355 (April).

Heckman, J.J., Smith, J.A., 1998. Evaluating the welfare state. In: Strom, S. (Ed.), Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium. Cambridge University Press, New York, pp. 241–318.

Heckman, J.J., Vytlacil, E.J., 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 (Fall).

Heckman, J.J., Lochner, L.J., Todd, P.E., 2006. Earnings equations and rates of return: The Mincer equation and beyond. In: Hanushek, E.A., Welch, F. (Eds.), Handbook of the Economics of Education, Chapter 7. North-Holland, Amsterdam, pp. 307–458.

Heckman, J.J., Urzua, S., Vytlacil, E.J., 2006. Understanding instrumental variables in models with essential heterogeneity. Review of Economics and Statistics 88 (3), 389–432.

Juhn, C., Murphy, K.M., Pierce, B., 1993. Wage inequality and the rise in returns to skill. Journal of Political Economy 101 (3), 410–442 (June).

Keane, M.P., Wolpin, K.I., 1997. The career decisions of young men. Journal of Political Economy 105 (3), 473–522 (June).

Laitner, J., 1992. Random earnings differences, lifetime liquidity constraints, and altruistic intergenerational transfers. Journal of Economic Theory 58 (2), 135–170 (December).

Lancaster, K.J., 1966. A new approach to consumer theory. Journal of Political Economy 74 (2), 132–157 (April).

Lewis, H.G., 1963. Unionism and Relative Wages in the United States: An Empirical Inquiry. University of Chicago Press, Chicago.

Lillard, L.A., Willis, R.J., 1978. Dynamic aspects of earning mobility. Econometrica 46 (5), 985–1012 (September).

MaCurdy, T.E., 1982. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. Journal of Econometrics 18 (1), 83–114 (January).

Mincer, J., 1974. Schooling, Experience and Earnings. Columbia University Press for National Bureau of Economic Research, New York.

Navarro, S., 2005. Understanding Schooling: Using Observed Choices to Infer Agent's Information in a Dynamic Model of Schooling Choice When Consumption Allocation is Subject to Borrowing Constraints. Ph.D. Dissertation, University of Chicago, Chicago, IL.

Pistaferri, L., 2001. Superior information, income shocks, and the permanent income hypothesis. Review of Economics and Statistics 83 (3), 465–476 (August).

Rosen, S., 1977. Human capital: A survey of empirical research. In: Ehrenberg, R. (Ed.), Research in Labor Economics, vol. 1. JAI Press, Greenwich, CT, pp. 3–40.

Roy, A., 1951. Some thoughts on the distribution of earnings. Oxford Economic Papers 3 (2), 135–146 (June).

Sims, C.A., 1972. Money, income, and causality. American Economic Review 62 (4), 540–552 (September).

Willis, R.J., Rosen, S., 1979. Education and self-selection. Journal of Political Economy 87 (5, Part 2), S7–S36 (October).

Wu, D., 1973. Alternative tests of independence between stochastic regressors and disturbances. Econometrica 41 (4), 733–750 (July).