THE IDENTIFICATION AND ECONOMIC CONTENT OF ORDERED CHOICE
MODELS WITH STOCHASTIC THRESHOLDS

Flavio Cunha
James J. Heckman
Salvador Navarro

The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds
Flavio Cunha, James J. Heckman, and Salvador Navarro
NBER Technical Working Paper No. 340
July 2007
JEL No. C31

ABSTRACT

This paper extends the widely used ordered choice model by introducing stochastic thresholds and interval-specific outcomes. The model can be interpreted as a generalization of the GAFT (MPH) framework for discrete duration data that jointly models durations and outcomes associated with different stopping times. We establish conditions for nonparametric identification. We interpret the ordered choice model as a special case of a general discrete choice model and as a special case of a dynamic discrete choice model.

Flavio Cunha
University of Pennsylvania
Department of Economics
160 McNeil Building
3718 Locust Walk
Philadelphia PA 19104
flaviocunha@gmail.com

James J. Heckman
Department of Economics
The University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and NBER
jjh@uchicago.edu

Salvador Navarro
University of Wisconsin, Madison
Department of Economics
1180 Observatory Drive
Madison WI 53706
snavarro@ssc.wisc.edu

1   Introduction

Throughout his career, Daniel McFadden has stressed the importance of economic theory in formulating and interpreting econometric models. He has also stressed the value of stating the exact conditions under which an econometric model is identified. The best known example of his approach is his analysis of discrete choice (1974; 1981), but there are other examples (e.g. Fuss and McFadden, 1978). In some of his earliest work (1963), he exposited the implicit economic assumptions used by Theil in the Rotterdam model of consumer demand.[2]

This paper continues the McFadden tradition by examining the economic foundations of the widely used ordered discrete choice model. We extend this model to allow for thresholds that depend on observables and unobservables to jointly analyze discrete choices and associated choice outcomes and to accommodate uncertainty at the agent level.

Ordered choice models arise in many areas of economics. Goods can sometimes be defined in terms of their quality as measured along a one-dimensional spectrum. In this case, consumer choice of a good can be modeled as the choice of an interval of the quality spectrum (Bresnahan, 1987; Prescott and Visscher, 1977; Shaked and Sutton, 1982). Schooling choices are often modeled using an ordered choice model (see, e.g. Machin and Vignoles, 2005). Cameron and Heckman (1998) present an economic analysis that justifies the application of the ordered choice model to schooling choices and a proof of the semiparametric identification of their model.[3]

In the analysis of taxation and labor supply with kinked convex constraints, choices of intervals of hours of work and segments of the consumer's budget set are often modeled using ordered choice models (Heckman and MaCurdy, 1981). Ordered choice models encompass a widely used class of duration models. Ridder (1990) established the equivalence of the conventional ordered choice model and GAFT (Generalized Accelerated Failure Time) models

---

[2]Theil's work is summarized in his collected papers on consumer demand (1975,1976).

[3]Our model generalizes Cameron and Heckman (1998) by allowing the cutoffs or thresholds that define the ordered choice model to depend on regressors and unobservables. We also establish that the ordered choice model can represent forward looking economic models, contrary to claims made by those authors.

for discrete time duration data which include the MPH (Mixed Proportional Hazard) model as a special case.

The conventional ordered choice model is assumed to be additively separable in observables (Z) and in unobservables (U_I) and is generated by an index

$$I = \varphi(Z) + U_I \tag{1}$$

where the observed-unobserved distinction is made from the point of view of the econometrician. U_I is a mean zero scalar random variable that is assumed to be independent of Z. It is traditionally assumed that I is separable in Z and U_I.[4] Individuals select a state $s \in \{1, \ldots, \bar{S}\}$ if the index lies between certain threshold or cutoff values c_s, which are assumed to be constants. We let $D(s) = 1$ if the agent chooses $S = s$. Cutoffs c_s are ordered so that $c_s \leq c_{s+1}$, $s = 1, \ldots, \bar{S} - 1$. In this notation, the ordered choice model can be written as

$$D(1) = \mathbf{1}(I \leq c_1), \ldots, D(s) = \mathbf{1}(c_{s-1} < I \leq c_s), \ldots, D(\bar{S}) = \mathbf{1}(c_{\bar{S}-1} < I), \tag{2}$$

where $c_{\bar{S}} = \infty$. The defining feature of the classical ordered choice model is that choices are generated by ordered sections of the support of a scalar latent continuous random variable I (e.g. durations or hours of work).[5]

In a number of contexts, it is plausible that the cutoff values differ among persons depending on variables that cannot be observed by the econometrician. In an analysis of taxation and labor supply, the locations of the kink points of the budget set, c_s, $s = 1, \ldots, \bar{S} - 1$, depends on assets and exemptions to which the agent is entitled. These may not be fully observable, especially if wages or assets are imputed.[6] In an analysis of schooling, there may

---

[4]If $I^* = g(Z, U_I)$, there may exist one or more monotonic transformations h, such that $h(I^*) = I = \varphi(Z) + U_I$. The conventional approach works with this representation.

[5]Separability of the index in Z and U_I as in (1) is a secondary requirement, but is a part of the specification of the classical ordered choice model.

[6]This problem is discussed by Heckman and MaCurdy (1981) and Heckman (1983).

be grade-specific subsidies and genuine grade-specific uncertainty at the agent level arising from learning about abilities and labor market shocks. Uncertainty is an essential feature of job search models.

To capture these possibilities, Carneiro, Hansen and Heckman (2003) generalize the ordered choice model by allowing the cutoffs c_s to depend on (a) state-s-specific regressors (Q_s) and (b) variables unobserved by the econometrician (η_s).[7] The thresholds are written as $c_s(Q_s, \eta_s)$. To preserve the separability of the classical ordered choice model, we assume that $c_s(Q_s, \eta_s) = c_s(Q_s) + \eta_s$, $s = 1, \ldots, \bar{S}$. We array the Q_s into a vector $Q = (Q_1, \ldots, Q_{\bar{S}})$. Carneiro, Hansen and Heckman (2003) adjoin systems of both discrete and continuous outcomes associated with the choice of each state, $s = 1, \ldots, \bar{S}$.

This paper builds on their analysis. We develop conditions for nonparametric identification of ordered choice models with stochastic thresholds and associated outcomes that are applicable to a variety of economic problems. We consider classes of economic models that can be represented by the ordered choice model. We also develop the restrictions on information processing and the arrival of new information that are required to produce a separable-in-observables-and-unobservables ordered choice duration model with stochastic thresholds that can be used to analyze dynamic discrete choices and associated outcomes.

We generate the ordered choice model from an index of marginal returns. The marginal returns must be monotone across the ordered states to preserve the structure of the ordered choice model. The unobservables must satisfy a stochastic monotonicity property. More formally, we define the generalized ordered choice model by extending (2) to:

$$(3) \qquad D(s) = \mathbf{1}\left(c_{s-1}(Q_{s-1}) + \eta_{s-1} \leq \varphi(Z) + U_I \leq c_s(Q_s) + \eta_s\right), \quad s = 1, \ldots, \bar{S},$$

where $c_0(Q_0) = -\infty$, $c_{\bar{S}}(Q_{\bar{S}}) = \infty$, and $\eta_0 = \eta_{\bar{S}} = 0$. When the η_s, $s = 1, \ldots, \bar{S}$, are non-degenerate, they can absorb U_I (i.e., setting $U_I = 0$ is innocuous). The model is separable

---

[7]See Carneiro, Hansen and Heckman (2003, footnote 23). This model is also discussed in Heckman, LaLonde and Smith (1999). For a recent analysis of this model and its relationship to the treatment effect literature, see Vytlacil (2006).

in terms of observables (functions of Q_s, $s = 1, \ldots, \bar{S} - 1$) and the unobservables (U_I, η_s, $s = 1, \ldots, \bar{S} - 1$). For this representation to be probabilistically meaningful, it is required that the upper and lower limits on $\varphi(Z) + U_I$ be ordered across all choices. We call this property *stochastic monotonicity* and we define it in assumption (A-1), where we condition on $Q = q$ and $Z = z$:

(A-1) $\Pr\left(c_s(q_{s-1}) + \eta_{s-1} - \varphi(z) - U_I \leq c_s(q_s) + \eta_s - \varphi(z) - U_I \mid Q = q, Z = z\right) = 1$,
for all $s = 1, \ldots, \bar{S}$.[8]

Assumption (A-1) defines ordered stochastic intervals that replace the non-stochastic intervals assumed in (2). (A-1) ensures that probabilities associated with the events characterized by (3) sum to one and are non-negative. (A-1) is a coherency condition for ordered choice models. This paper analyzes an array of well defined economic models that can be characterized by (3) and (A-1).

The plan of this paper is as follows. Section 2 presents four ordered choice models to demonstrate the range of economic phenomena that the ordered choice model can capture. The first is a model under perfect certainty for the choice of goods when qualities are heterogeneous. A version of this model can be used to analyze labor supply in the presence of discontinuous tax schedules. The second is a prototypical model of discrete choice under perfect certainty. The third is a model of agent decision making under uncertainty with sequential revision of information. Section 3 establishes conditions for nonparametric identification of the generalized ordered choice model. Section 4 discusses identification of ordered choice models with adjoined state-specific outcomes. Section 5 concludes.

---

[8]An alternative and equivalent formulation is that

(A-1)′ $\Pr\left(c_s(q_{s-1}) + \eta_{s-1} \leq c_s(q_s) + \eta_s \mid Q = q, Z = z\right) = 1$ *for all* $s = 1, \ldots, \bar{S}$,

where the conditioning on Z is redundant for this condition, but we maintain it to unify the notation in this paper.

2 Ordered Choice Models

Let "s" denote a state generated by some latent variable falling in an interval. The latent variable can be an index associated with different lengths of durations as it falls into different segments of an underlying continuum as in the GAFT model of Ridder (1990). "s" can be a stage in a process or a quality interval that defines a good as in Prescott and Visscher (1977), Shaked and Sutton (1982) and Bresnahan (1987). It can also represent intervals of hours of work as in Heckman (1974) and Heckman and MaCurdy (1981). Schooling with $\bar{S}$ stages is another example where the latent index is a marginal return function. The framework is general and can be used to model the choice of the time at which a drug is taken or the date (stage) at which a machine is installed. We present four examples. One is a model for the choice of differentiated goods. The second is a version of the deterministic Wicksell (1934) capital model applied to the cutting of a tree. The third is a stochastic tree cutting problem. The fourth is an optimal schooling model that captures the essential features of the model of Keane and Wolpin (1997).

2.1 Choice of Differentiated Goods

Following the analysis of Prescott and Visscher (1977), let τ_i be consumer i's marginal valuation of quality X. Goods come in discrete packages with quality X_g and price P_g, $g = 1, \ldots, G$. A quality-price bundle (X_g, P_g) defines a good. Consumers can buy at most one unit of the good. Bundles are ordered so that $X_{g+1} > X_g$ and $P_{g+1} > P_g$. Assume that all of the goods are purchased in equilibrium. Consumer preferences are over X and the rest of consumption M:

$$U(X_g, M) = \tau_i X_g + M.$$

For income Y, if a person buys good g at price P_g, $M = Y - P_g$.

Consumer i is indifferent between two goods $g + 1$ and g if

$$X_{g+1} \tau_i - P_{g+1} = X_g \tau_i - P_g.$$

Thus, persons are indifferent between goods $g + 1$ and g if their value of $\tau_i = c_g$, where

$$c_g = \frac{P_{g+1} - P_g}{X_{g+1} - X_g}.$$

The "cutoff value" c_g has the interpretation of the marginal price per unit quality. If $c_g < \tau_i \leq c_{g+1}$, the consumer buys good $g + 1$. As an equilibrium condition, the marginal price of quality must be nondecreasing in the level of quality. If there are some agents at each margin of indifference, an ordered choice model is generated with these threshold values. In the notation of the ordered choice model, $I_i = \tau_i$, $c_s = \frac{P_{s+1} - P_s}{X_{s+1} - X_s}$, $s = 1, \ldots, \bar{S} - 1$, and the goods are ordered by their price per unit quality. The demand function in terms of τ is generated as the envelope of $\tau X_g - P_g$, $g = 1, \ldots, G$.[9] The cutoffs may depend on both observed and unobserved variables. Prices may depend on the characteristics of the buyer. Quality may be measured with error so the thresholds may be stochastic.

In the analysis of taxes and labor supply (Heckman, 1974; Heckman and MaCurdy, 1981), the ordered choice model arises as the natural econometric framework for analyzing labor supply in the presence of progressive taxation associated with different tax brackets at different levels of earnings. Cutoffs correspond to points of discontinuity of the tax schedule that are determined by exemptions and asset levels, and that may be only partially observed by the econometrician. We next develop a stopping time example which is a vehicle for introducing uncertainty into the framework of the ordered choice model and thereby extending it. We begin by developing the case of perfect certainty. This is a version of a tree cutting problem, originally analyzed by Wicksell (1934) and applied to the analysis of human capital by Rosen (1977).

---

[9]One good might have zero price at zero quality.

2.2 An Optimal Stopping Model Under Perfect Certainty

Let $S = s$ denote the individual's choice of stopping time, where $s \in \{1, \ldots, \bar{S}\}$. Let $R(s, X)$ denote the discounted net lifetime reward associated with stopping at stage s, where the discounting is done at the end of period s. An example would be a model of the choice of schooling s where each schooling level is assumed to take one year and the opportunity cost of schooling is the foregone earnings.[10]

In an environment of perfect certainty, the agent solves the problem $\max_{s \in \{1, \ldots, \bar{S}\}} \left\{ \frac{R(s, X)}{(1+r)^s} \right\}$, where r is the interest rate and $R(s, X)$ is the reward from stopping at stage s. The value function at stage $S - 1$ is $V(S - 1, X) = \max \left\{ R(S - 1, X), \frac{V(S, X)}{1+r} \right\}$, where the agent's value function at $\bar{S} - 1$, $V(\bar{S} - 1, X)$, is $V(\bar{S} - 1, X) = \max \left\{ R(\bar{S} - 1, X), \frac{R(\bar{S}, X)}{1+r} \right\}$. In the general case, an individual will stop at stage s if $R(s, X) \geq \frac{V(s+1, X)}{1+r}$. For the agent to reach stage s, it is required that $R(s - 1, X) < \frac{V(s, X)}{1+r}$. This rule produces the global optimum.

An ordered choice model representation of the general choice problem can be written when pairwise comparisons of returns $R(s, X)$ across adjacent states characterize the optimum and additional separability assumptions are invoked. This model is based on the *marginal return function*

$$f(s, X) = \frac{R(s, X)}{1 + r} - R(s - 1, X).$$

Assuming concavity, i.e.

(A-2) *the marginal return function $f(s, X)$ is nonincreasing in s for all X,*

the optimum for the general problem is characterized by $s = s^*$ if and only if $f(s^* + 1, X) \leq 0 \leq f(s^*, X)$. The optimum is unique if the weak inequality on the left of zero is replaced by a strict inequality.

---

[10]Agents pay a fixed cost $C(s, X)$ after completing each grade of school and $R(s, X)$ is the reward to schooling net of these costs. When $C(s, X) = 0$, the agent's only costs are foregone earnings. This model includes both Card's (1999) and Rosen's (1977) versions of Becker's Woytinsky Lecture (1967). We work with present values of earnings associated with schooling states and Card works with annualized returns.

The general rule for locating the optimum s^* is

$$\frac{V(s^*+1,X)}{1+r} - R(s^*,X) \leq 0 \leq \frac{V(s^*,X)}{1+r} - R(s^*-1,X).$$

Because of the concavity assumed in (A-2), one can replace $V(s^*,X)$ and $V(s^*-1,X)$ in these expressions with $R(s^*,X)$ and $R(s^*-1,X)$, respectively. We now introduce unobservables into the model. With additional separability assumptions about the unobservables, (A-1) and (A-2) produce the conventional ordered choice model.

2.2.1 Introducing unobservables into the model

Unobservables are introduced into the model in two distinct ways. Both preserve additive-separability-in-unobservables that is a defining feature of the conventional ordered choice model. First, we introduce a scalar random variable U_I representing an invariant individual-specific shifter of the net gain function that is observed and acted on by the individual but is not observed by the econometrician.[11] Second, there may be transition-specific regressors that determine the net return (e.g. tuition in a schooling model), some of which are unobserved.

For the optimal stopping model defined in the preceding section to be represented by separable ordered choice model (3), we need to invoke separability in the marginal return function $f(s,X)$ in addition to monotonicity in s.

(A-3) *Assume that the marginal return depends on individual characteristics where $f(s+1,X) = \frac{R(s+1,X)}{1+r} - R(s,X) = -(c_s(Q_s) + \eta_s) + \varphi(Z) + U_I$, $s \in \{1,\ldots,\bar{S}\}$, where $X = (Q_1,\ldots,Q_{\bar{S}-1},Z)$, and $E(U_I) = 0$. The Z variables are common across all states, $s = 1,\ldots,\bar{S}$. The Q_s are the state-specific arguments of $R(s+1,X)$ and $R(s,X)$, and components of X are observed. The η_s are unobservables from the point of view of the econometrician.*

The η_s and the $c_s(Q_s)$ can be interpreted as cost shocks. Under (A-3), the choice of schooling

---

[11]Such invariant random variables are sometimes called components of "heterogeneity" in the literature.

level s is characterized by (3). The cutoffs must satisfy the stochastic monotonicity assumption. This restriction imposes constraints on the model that are not present in standard discrete choice models. The traditional ordered choice model treats the $c_s(Q_s)$ as constants and sets $\eta_j \equiv 0$, $j = 1, \ldots, \bar{S}$.

It is fruitful to compare this model with a general discrete choice model with net rewards for choice s written as

$$R(s, X) = \mu_R(s, X) - \varepsilon(s), \ \ s = 1, \ldots, \bar{S},$$

where preference shocks satisfy $\varepsilon(s) \perp\!\!\!\perp X$ for all s, where "$\perp\!\!\!\perp$" denotes independence conditional on the arguments to the right of the symbol. The optimal s, defined as s^*, is

$$s^* = \operatorname*{argmax}_j \{R(j)\}_{j=1}^{\bar{S}}.$$

In the general model, the states are unordered. In the ordered choice specialization of this model,

$$f(s, X) = R(s, X) - R(s-1, X) = \mu_R(s, X) - \mu_R(s-1, X) - (\varepsilon(s) - \varepsilon(s-1)), s = 1, \ldots, \bar{S}-1,$$

where $\varphi(Z)$ in equation (3) consists of components of $\mu_R(s, X) - \mu_R(s - 1, X)$ that are functionally independent of s, and $-c_s(Q_s)$ are the components of $\mu_R(s, X) - \mu_R(s - 1, X)$ that are $(s - 1, s)$-specific and $\eta_s - U_I = \varepsilon(s) - \varepsilon(s - 1)$. These shocks can be interpreted as either negative marginal return shocks or as marginal cost shocks added to gross returns. Condition (A-1) restricts the admissible shocks in a general discrete choice model to satisfy the *ordered discrete choice condition* for $X = x$:

(**OD**) $\qquad \Pr\left(c_{j+1}(q_{j+1}) - c_j(q_j) \geq 2\varepsilon(j) - \varepsilon(j-1) - \varepsilon(j+1) \mid X = x\right) = 1.$[12]

9

This condition is testable because Matzkin (1994) and Heckman and Navarro (2007) show, under different assumptions, that the general multinomial choice model is nonparametrically identified, so that it is possible to identify the joint distribution of the $\varepsilon(s)$, $s = 1, \ldots, \bar{S}$ up to pairwise contrasts, and the $c_j(q_j)$, $j = 1, \ldots, \bar{S}$ up to pairwise contrasts.[13]

In this specification, agents are assumed to be making choices in an atemporal setting. They draw shocks $\varepsilon(s)$ across all states, $s = 1, \ldots, \bar{S}$, subject to condition (**OD**) and maximize their utility. In section 3, we establish conditions under which this model is nonparametrically identified.

The ordered choice model is a version of the mixed proportional hazards for discrete durations (Ridder, 1990), which is widely used in applied work on unemployment and other dynamic outcomes. It is thus of interest to examine whether the ordered choice model can be modified to capture the sequential arrival of information under uncertainty. We generalize the analysis of this section to account for uncertainty and agent information updating. Additional assumptions are required to justify the ordered choice framework of this section as a well-defined economic model for the analysis of uncertain environments in which agents update their information about their future choices.

2.3   Adding Sequential Revelation of Information

This section extends the ordered choice model under certainty to a model with period-specific shocks that are not known by the agents in advance. We consider a prototypical tree-cutting problem which motivates economically richer models. We give conditions under which the ordered choice framework accurately captures the economic model. In section 2.4 we generalize our analysis to a model that is a version of the general framework of Keane and Wolpin (1997).

Let L_t denote the length of a tree t periods after it is planted. $L_{t+1} = (1-\rho)\bar{L} + \rho L_t + \varepsilon_{t+1}$,

---

[12]Recall that X contains both Q and Z components.

[13]Crawford, Pollak and Vella (1998) analyze restrictions on the derivatives of ordered logit and general logit choice models for an ordered logit model with nonstochastic thresholds that do not depend on the regressors.

where $E\left(\varepsilon_{t+1} \mid L_t\right) = 0$, $-1 < \rho < 1$, and $\bar{L}$ is the steady state mean.[14] Assume that the agent has to decide when to cut the tree. The agent has a finite number of periods to make this decision. Thus, if the agent does not cut the tree by the end of period T, he loses the right to cut the tree and his payoff is zero. If the individual cuts the tree in period $t (\leq T)$, he collects L_t, where we assume that the price of one foot of lumber is unity in each period. The individual may decide not to cut the tree, and keep the option of selling it tomorrow when it may be a little longer. As in Wicksell's (1934) model, the opportunity cost of not cutting the tree is generated by the foregone interest, which accumulates at a deterministic rate r. We may write the value function of the individual at stage t as

(4)
$$V\left(L_t\right) = \max\left\{L_t, \frac{1}{1+r}E\left[V\left(L_{t+1}\right)\middle| L_t\right]\right\} \text{ if } t = 1, ..., T-1$$

$$V\left(L_T\right) = L_T.$$

The agent will always cut the tree. If by the last period he has not yet cut the tree, it is always preferable to cut the tree rather than to forego the right to do so. Assume that the growth process of the tree satisfies the condition that

(5)
$$\frac{1}{1+r}E\left[L_{t+1}\middle| L_t\right] - L_t \leq \frac{1}{1+r}E\left[L_t\middle| L_{t-1}\right] - L_{t-1}.$$

Condition (5) guarantees that the stochastic monotonicity condition (A-1) is satisfied. Thus, this problem can be econometrically formulated as an ordered choice model with stochastic thresholds.

Consider an agent who at the beginning of period $\tau = 1, \ldots, T-1$ is contemplating whether to cut the tree or not. If $\frac{1}{1+r}E\left[L_{\tau+1}\middle| L_\tau\right] - L_\tau > 0$ he should not cut the tree. To see why, note that $L_{t+1} \leq V\left(L_{t+1}\right) = \max\left\{L_{t+1}, \frac{E\left[L_{t+2}\middle| L_{t+1}\right]}{1+r}\right\}$. Consequently, $E\left[L_{t+1}\middle| L_t\right] \leq E\left[V\left(L_{t+1}\right)\middle| L_t\right]$ so that $L_t < \frac{1}{1+r}E\left[L_{t+1}\middle| L_t\right] \leq \frac{1}{1+r}E\left[V\left(L_{t+1}\right)\middle| L_t\right]$ and

---

[14]Our version of the tree cutting problem is closely related to the case discussed by Brock et al. (1989), who analyze a model with a random walk error term. Our model is stationary in tree lengths and their model is nonstationary in tree lengths, but stationary in present values.

thus $V(L_t) = \max\left\{L_t, \frac{1}{1+r}E\left[V(L_{t+1})|L_t\right]\right\} = \frac{1}{1+r}E\left[V(L_{t+1})|L_t\right]$. If, on the other hand, $\frac{1}{1+r}E\left[L_{\tau+1}|L_\tau\right] - L_\tau \leq 0$, the agent should cut the tree. To see why, consider the value function in the next to last period. By definition,

$$V(L_{T-1}) = \max\left\{L_{T-1}, \frac{1}{1+r}E\left[L_T|L_{T-1}\right]\right\} = L_{T-1}.$$

The first equality arises because in the last period, $V(L_T) = L_T$. The second equality arises because if $t < T-1$ and $\frac{1}{1+r}E\left[L_{t+1}|L_t\right] - L_t \leq 0$, then $\frac{1}{1+r}E\left[L_T|L_{T-1}\right] - L_T \leq 0$ is implied by (5). Proceeding by backward induction, at period τ we have

$$V(L_\tau) = \max\left\{L_\tau, \frac{1}{1+r}E\left[V(L_{\tau+1})|L_\tau\right]\right\} = \max\left\{L_\tau, \frac{1}{1+r}E\left[L_{\tau+1}|L_\tau\right]\right\} = L_\tau.$$

The second equality shows that we can write the problem of the agent using a one stage look ahead rule substituting the conditional expectation of the value function $E\left[V(L_{t+1})|L_t\right]$ by the conditional expectation of the length of the tree, $E\left[L_{t+1}|L_t\right]$.[15]

The agent cuts the tree at the period $\tau = 1, \ldots, T-1$ that satisfies

$$\frac{1}{1+r}E\left[L_{\tau+1}|L_\tau\right] - L_\tau \leq 0 \qquad \text{and} \qquad \frac{1}{1+r}E\left[L_\tau|L_{\tau-1}\right] - L_{\tau-1} \geq 0.$$

If we define

$$\frac{1}{1+r}E\left[L_{\tau+1}|L_\tau\right] - L_\tau = \varphi(Z) + U_I - c_\tau(Q_\tau) - \eta_\tau,$$

it follows from (5) that

$$\Pr\left(c_\tau(Q_t) + \eta_\tau - (\varphi(Z) + U_I) \geq c_{\tau-1}(Q_{\tau-1}) + \eta_{\tau-1} - (\varphi(Z) + U_I) \mid Q = q, Z = z\right) = 1,$$

which is the *stochastic monotonicity* condition (A-1). Consequently, the tree-cutting problem can be formulated as a generalized ordered choice model, and the tree is cut in period τ if

---

[15]Ferguson (2003) shows that the one step ahead rule is only optimal in monotone optimal stopping problems.

$c_{\tau-1}\left(Q_t\right) + \eta_{\tau-1} \leq \varphi(Z) + U_I \leq c_\tau\left(Q_\tau\right) + \eta_\tau$. We next consider a more general model of dynamic discrete choice under uncertainty.

2.4   A Dynamic Schooling Choice Model

We next consider a richer dynamic model where individuals decide between two choices which are stochastically updated. This is a two sector version of the general model estimated by Keane and Wolpin (1997).[16]

Consider a model in which an individual with ability a at each period t decides whether to enroll in school or not, where $t = 1, ..., T$. Instead of schooling, we can analyze other types of discrete states in which agents decide to remain in the state 0 and then drop out (e.g. a spell of training or a physical therapy program). We denote the current schooling level of the individual by s_t, $s_t = 1, ..., \bar{S}$. Let $d_t = 0$ if the agent decides not to enroll in school in period t and $d_t = 1$ otherwise. If an agent with schooling level s does not enroll in school, he works full time and increases his schooling-sector-s specific experience by one unit. If the same agent decides to enroll in school, he works part time, but does not accumulate experience. In each period t and at each schooling level s, let $x_{s,t} \in \Re_+$ denote the accumulated experience of the individual in schooling sector s. Let $x_t = \left(x_{1,t}, ..., x_{\bar{S},t}\right) \in \Re_+^{\bar{S}}$ denote the vector of accumulated experience in all schooling sectors $s = 1, \ldots, \bar{S}$. It simplifies notation to define the vector $e_{s,t} \in \Re_+^{\bar{S}}$ and vector $\bar{x}_{s,t} \in \Re_+^{\bar{S}}$ as $e_{s,t} = (0, ..., 1, ..., 0)$ and $\bar{x}_{s,t} = (0, ..., x_{s,t}, ..., 0) = x_{s,t}e_{s,t}$, respectively.

If the individual decides not to enroll in school in period t, he works full time and has earnings $r_0\left(t, s, x_{s,t}, a\right)$ which depends on schooling level s, accumulated experience by period t in sector s, $x_{s,t}$, and ability a. If the individual decides to enroll in school in period t his earnings are $r_1\left(t, s, x_{s,t}, a\right)$. We assume that the earnings functions $r_k\left(t, s, x_{s,t}, a\right)$, $k = 0, 1$, satisfy the following conditions:

---

[16]Keane and Wolpin do not consider nonparametric identification of their model.

(A-4) *(A)* $r_0(t, s, x_{s,t}, a) = \alpha(s, a) + \gamma x_{s,t} + \Delta r_0(t, s, x_{s,t})$, *where* $\Delta r_0(t, s, x_{s,t})$ *is unknown from the point of view of the agent at all periods* $\tau < t$. *We assume that* $\gamma \geq 0$.

(B) $r_1(t, s, x_{s,t}, a) = \delta(s, a) + \Delta r_1(t, s, x_{s,t})$, *where* $\Delta r_1(t, s, x_{s,t})$ *is unknown from the point of view of the agent at all periods* $\tau < t$.

(C) For every s, a: $\alpha(s+1, a) \geq \alpha(s, a)$.

(D) For every s, a: $\alpha(s+1, a) - \alpha(s, a) \leq \alpha(s, a) - \alpha(s-1, a)$;

(E) For every s, a: $\alpha(s, a) \geq 0$. *If* $a' \geq a$ *then* $\alpha(s, a') \geq \alpha(s, a)$;

(F) For every a, $\alpha(\bar{S}+1, a) - \alpha(\bar{S}, a) \leq 0$;

(G) For every s, a: $\alpha(s, a) - \delta(s, a) \geq 0$;

(H) For every t, s, x_t, a: $\Pr[\Delta r_0(t, s, x_{s,t}) - \Delta r_1(t, s, x_{s,t}) \leq \delta(s, a) - \alpha(s, a)] = 0$.

(I) Let $\Delta r(t) = (\Delta r_0(t, 1, x_{1,t}), ..., \Delta r_0(t, \bar{S}, x_{\bar{S},t}), \Delta r_1(t, 1, x_{s,t}), ..., \Delta r_1(t, \bar{S}, x_{\bar{S},t}))$.
We assume that $\Delta r(t)$ *is independent from* $\Delta r(t')$ *for* $t \neq t'$;

Conditions *(A)* and *(B)* state that the earnings functions $r_k(t, s, x_{s,t}, a)$ are linear and separable in schooling and experience. This specification is commonly invoked in labor economics (see, e.g., Keane and Wolpin, 1997, or Heckman, Lochner and Todd, 2006a). The stochastic components are also assumed to be separable from the deterministic components. However, we do not require s and a to be separable. To maintain mathematical tractability and to simplify the argument, we assume the "Mincer" model that assumes that the experience profiles are parallel—that is, that $\gamma_s = \gamma$ for all s. For simplicity, we also assume that the earnings of the agents enrolled in school do not depend on their work experience. The "Mincer" assumptions can be relaxed at the cost of greater notational complexity.

Condition *(C)* states that $\alpha(s, a)$ is increasing in schooling s for all ability levels a. From Condition *(D)*, the marginal returns to schooling, $\alpha(s+1, a) - \alpha(s, a)$ are decreasing in schooling, for all ability levels a. Condition *(E)* says that $\alpha(s, a)$ is increasing in ability a for all schooling levels s. Condition *(F)* states that the returns to school above maximum schooling level $\bar{S}$ are nonpositive for all levels of ability a. Conditions *(G)* and *(H)* impose the requirement that the current opportunity costs of attending school are non-negative.

14

Assuming that the agent starts with schooling level s_0 and accumulated experience x_0, the individual is assumed to maximize the following criterion:

$$V\left(1, s_0, x_0, a\right) = \max_{\{d_\tau\}_{\tau=1}^{T}} E\left\{\sum_{\tau=1}^{T} \beta^{\tau-1}\left[d_\tau r_1\left(\tau, s_\tau, x_{s_\tau,\tau}, a\right) + \left(1 - d_\tau\right) r_0\left(\tau, s_\tau, x_{s_\tau,\tau}, a\right)\right] \middle| s_0, x_0, a\right\}$$

subject to

$$s_0 = 0, x_0 = 0;$$

$$s_{\tau+1} = s_\tau + d_\tau;$$

$$x_{s_{\tau+1},\tau+1} = x_{s_{\tau+1},\tau+1} + 1 \text{ if } d_\tau = 0;$$

$$x_{s_\tau,\tau+1} = x_{s_\tau,\tau} \text{ if } d_\tau = 1.$$

Consider an agent in period t who has accumulated a total of s years of schooling and experience vector x_t. Let $z_t = \left(x_{1,t}, ..., x_{s,t} + 1, ..., x_{\bar{S},t}\right)$. We can write the problem of the agent recursively as

(6)

$$V\left(t, s, x_t, a\right) = \max\left\{\begin{array}{c} \alpha\left(s, a\right) + \gamma x_{s,t} + \Delta r_0\left(t, s, x_{s,t}\right) + \beta E\left[V\left(t+1, s, z_t, a\right) \middle| s, x_t, a\right], \\ \delta\left(s, a\right) + \Delta r_1\left(t, s\right) + \beta E\left[V\left(t+1, s+1, x_t, a\right) \middle| s, x_t, a\right] \end{array}\right\}.$$

The agent decides to enroll in school, $d_t = 1$, if and only if

$$\left\{\alpha\left(s, a\right) + \gamma x_{s,t} - \delta\left(s, a\right) + \Delta r_0\left(t, s, x_{s,t}\right) - \Delta r_1\left(t, s\right)\right\}$$

$$+ \left\{\beta E\left[V\left(t+1, s, z_t, a\right) - V\left(t+1, s, x_t, a\right) \middle| s, x_t, a\right]\right\}$$

$$\leq \beta E\left[V\left(t+1, s+1, x_t, a\right) - V\left(t+1, s, x_t, a\right) \middle| s, x_t, a\right].$$

The left-hand side of the inequality has two components in braces that represent the costs of enrolling in school. The first is the foregone earnings of being enrolled in school, which is nonnegative according to Conditions (G) and (H) of (A-4). The second term in braces

15

arises from returns to investment. If the agent decides to enroll in school today he will not collect the higher pay associated with work experience he accumulates if he decides not to enroll in school. The opportunity cost of enrolling in school is the current earnings foregone plus the future return to current work experience. The future benefits, which appear on the right-hand side, arise from higher levels of education. The agent decides to enroll in school if the net benefit is positive.

In general, the solution to the dynamic schooling choice model may involve dropping out of school for some periods to take advantage of favorable labor market conditions. Note that if there are T periods, there are a total of 2^T possible paths. The solution of the model is obtained by generating a set of decision rules $\{d_t^* (s_t, \mathcal{I}_t)\}_{\tau=1}^T$ by backward induction.

2.4.1 Sufficient Conditions for the Ordered Choice Model of Schooling to Represent the Stochastic Dynamic Schooling Choice Model.

As shown in our analysis of the tree-cutting problem, the ordered choice model can sometimes represent an optimal stopping problem. This is not always true in a general stochastic dynamic schooling choice model, where agents may drop out of school and return at a later date. The first step required to justify an ordered choice model in this more general setup is to obtain conditions that guarantee that if the agent finds it optimal not to enroll in school at date t, *i.e* $d_t = 0$, then he will also find it optimal not to return, so that $d_\tau = 0$ for any $\tau > t$. We establish a series of propositions that justify application of the ordered choice model.

Proposition 1. *Under conditions (D) and (F) of (A-4), for each ability level a there exists a schooling level $s^* (a)$ such that:*

$$\alpha (s^* (a) + 1, a) - \alpha (s^* (a), a) \leq \gamma.$$

Proof. This is a consequence of the assumption that $\alpha (s, a)$ is concave in s and that for all

16

$a, \alpha\left(\bar{S}+1, a\right) - \alpha\left(\bar{S}, a\right) \le 0.$ □

Proposition 2. *Under (A-4), for any ability level a, schooling level $s \ge s^*(a)$, experience vector $\bar{x}_{s,t}$, and period $t \ge s^*(a)$:*

$$(7) \quad V\left(t, s, \bar{x}_{s,t}, a\right) = E\left[\sum_{\tau=t}^{T} \beta^{\tau-t} r_0\left(\tau, s, x_{s,t} + \tau - t, a\right) \middle| s, \bar{x}_{s,t}, a\right] + \Delta r_0\left(t, s, x_{s,t}\right)$$
$$= \sum_{\tau=t}^{T} \beta^{\tau-t} \left[\alpha\left(s, a\right) + \gamma\left(x_{s,t} + \tau - t\right)\right] + \Delta r_0\left(t, s, x_{s,t}\right)$$

Proof. See Appendix A. □

From this proposition it follows that the value of the program after $s^*(a)$ is just the present value of earnings in the no schooling state. This proposition implies the following useful corollary.

Corollary 1. *(To Proposition 2). Under (A-4), if the agent reaches school level $s^*(a)$ at some period $t < T$, then at period t, he drops out of school with $s^*(a)$ years of education and he never returns to school at any period $\tau = t+1, \ldots, T$.* □

We next analyze the behavior of an agent with schooling level $s < s^*(a)$. To do so, we impose the following additional assumption that guarantees that the returns to experience grow faster than the returns to education:

(A-5) *(Entrapment): For any school level $s = 1, \ldots, \bar{S}$ and ability level a,*

$$\alpha(s+k, a) - \alpha\left(s, a\right) \le (k+1)\gamma,$$

where k is a nonnegative integer.

This condition, joined with (A-4), allows us to simplify the value function, as we record in the next proposition.

17

Proposition 3. *Under (A-4) and (A-5), at every period t, schooling level s such that $s^*(a) = s + k(a)$, $k(a) \geq 0$, for persons with experience vector $x_t = 0$ (no work experience up to t), and ability level a:*

(8)

$$V(t, s, 0, a) =$$

$$= \max \left\{ \begin{array}{c} \Delta r_0(t, s, 0) + \sum_{\tau=t}^{T} \beta^{\tau-t} \left[\alpha(s, a) + \gamma(\tau - t) \right], \\ \delta(s, a) + \Delta r_1(t, s) + \beta \sum_{\tau=t+1}^{T} \beta^{\tau-t-1} \left[\alpha(s+1, a) + \gamma(\tau - t - 1) \right], \\ \vdots \\ \sum_{l=1}^{k(a)-1} \beta^{l-1} \delta(s+l, a) + \Delta r_1(t, s) + \beta^{k(a)} \sum_{\tau=t+k(a)}^{T} \beta^{\tau-t-k(a)} \left[\alpha(s^*(a), a) + \gamma(\tau - t - k(a)) \right] \end{array} \right\}.$$

Proof. See Appendix A. □

From Proposition 3, we obtain the following corollary.

Corollary 2. *(To Proposition 3). Under (A-4) and (A-5), at every period t, schooling level s such that $s^*(a) = s + k(a)$, $k(a) \geq 0$, experience vector $x_t = 0$, and ability level a, if the agent decides to drop out of school at period t with schooling level s, he/she never returns to school.*

Define the expected reward to permanently dropping out at schooling level s at time t as

$$R(t, s, 0, a) = \alpha(s, a) + \Delta r_0(t, s, 0) + \sum_{\tau=t+1}^{T} \beta^{\tau-t} \left[\alpha(s, a) + \gamma(\tau - t) \right].$$

Under (A-4) and (A-5), we can rewrite the value function in this notation as

$$V(t, s, 0, a) = \max \left\{ R(t, s, 0, a), \delta(s, a) + \Delta r_1(t, s, 0) + \beta E_t \left[V(t+1, s+1, 0, a) \right] \right\}.$$

We have just shown that under our assumptions, for any period t, schooling level s, experience x_t, and ability level a, once the agent leaves school he never returns. The optimal schooling

18

problem is an optimal stopping problem. Note that if $R(t, s, 0, a) \leq \delta(s, a) + \Delta r_1(t, s, 0) +$

$\beta E_t [R(t+1, s+1, 0, a)]$ then the agent would not leave school. This is a straightforward

consequence of Proposition (2), which establishes that $E[R(t+1, s+1, 0, a)] \leq E_t[V(t+1, s+1, 0, a)]$.

Note that if we normalize the variables in a way that the first period is "1" and the first

schooling level is also "1", then an agent who never drops out of school and advances one

grade per period will reach schooling level "s" at period "s."

Assumptions (A-4) and (A-5) guarantee that the dynamic discrete choice model is an

optimal stopping model. They are not enough to deliver an ordered choice model represen-

tation. Additional assumptions are required. We invoke the following additional assumption

(A-6):

(A-6) *For any schooling level s and ability a :*

$$\alpha(s+k, a) - \delta(s+k, a) \geq \frac{1 - \beta^{T-s-k}}{1 - \beta^{T-s}} (\alpha(s, a) - \delta(s, a))$$

and

$$\Pr\left(\Delta r_0(s+k, s+k, 0) - \Delta r_1(s+k, s+k, 0) \geq \frac{1 - \beta^{T-s-k}}{1 - \beta^{T-s}} \Delta r_0(s, s, 0) - \Delta r_1(s, s, 0)\right) = 1.$$

These conditions ensure that the schooling problem is a monotone optimal stopping problem

as defined in Ferguson (2003). In a monotone optimal stopping problem, local comparisons

of returns generate globally optimal choices. We establish the following claim:

Proposition 4. *Under (A-4) – (A-6), dynamic discrete schooling choice model is a monotone*

optimal stopping problem.

Proof. See Appendix A. □

Under (A-4) – (A-6) we can solve the global optimization problem by making local or one

stage ahead comparisons:

Proposition 5. *Under (A-4)-(A-6), we can write the value function of the dynamic discrete schooling choice model as*

$$V(t, s, 0, a) = \max \{R(t, s, 0, a), \delta(s, a) + \Delta r_1(t, s, 0) + \beta E_t[R(t+1, s+1, 0, a)]\}.$$

Proof. This is true because for monotone optimal stopping problems, the one stage look ahead rule is optimal. For a proof, see Ferguson (2003). □

We now collect the results established in this section on conditions on a general dynamic discrete choice model that justify the ordered choice model.

Theorem 1. *Under (A-4) – (A-6), the dynamic discrete schooling choice model can be represented by an ordered choice model with stochastic thresholds that satisfies stochastic monotonicity condition (A-1).*

Proof. Define

$$\varphi(Z, U_I) = -\gamma$$

$$C_s(Q_s) = \beta \left(\frac{1 - \beta^{T-s}}{1 - \beta} \right) [\alpha(s, a) - \delta(s, a)] - [\alpha(s+1, a) - \alpha(s, a)]$$

$$\eta_s = \beta \left(\frac{1 - \beta^{T-s}}{1 - \beta} \right) [\Delta r_0(s, s, 0) - \Delta r_1(s, s, 0)].$$

To show that it satisfies stochastic monotonicity, by the definition of $C_s(Q_s)$ and η_s, together with (A-6), we have for any $s = 1, ..., \bar{S}$:

$$C_s(Q_s) + \eta_s \leq C_{s+1}(Q_{s+1}) + \eta_{s+1}$$

and hence (A-1) is satisfied.[17] □

---

[17]In addition to the shocks associated with uncertainty at the agent level, we can add components of heterogeneity known to the agent but not observed by the econometrician. U_I can be interpreted as such a component. Cameron and Heckman (1998) and Carneiro, Hansen and Heckman (2003) analyze mixture versions of the ordered choice model. See the analysis in section 3 below.

We next consider some simple examples to illustrate the various economic models analyzed in this paper.

2.5   A Three Period Example

A three period (schooling level) example helps to fix ideas developed for different models. Suppose that the reward function associated with each schooling level can be written as

$$R(1) = \mu_1(X) + \varepsilon(1), \qquad R(2) = \mu_2(X) + \varepsilon(2), \qquad R(3) = \mu_3(X) + \varepsilon(3).$$

Assume no discounting. A standard discrete choice model postulates that the agent draws $\varepsilon = (\varepsilon(1), \varepsilon(2), \varepsilon(3))$ and $s = \operatorname{argmax}_j \{R(j)\}_{j=1}^3$. There is no restriction on the $\varepsilon(j)$, $j = 1, 2, 3$. The ordered choice model applied to this setting imposes the restriction (**OD**) given in section 2.2.1.

Next, we take the same reward functions but use them in a simple version of the sequential dynamic discrete choice model with information updating developed in section 2.4. Agents are assumed to choose among states, and states and periods are the same. Assume that agents know X. The only uncertainty at the agent level is about the $\varepsilon(j)$. Let $\mathcal{I}_j$ denote the agent's information set. In period i, the agent knows $\varepsilon(i)$ but not $\varepsilon(i')$, $i' > i$. The agent stops at stage 1 if

(9a)
$$\mu_1(X) - \varepsilon(1) > E\left[(\mu_2(X) - \varepsilon(2))\,\mathbf{1}[\mu_2(X) - \varepsilon(2) > \mu_3(X) - E(\varepsilon(3) \mid \mathcal{I}_2)]\right.$$
$$\left. + (\mu_3(X) - E(\varepsilon(3) \mid \mathcal{I}_2))\,\mathbf{1}[\mu_2(X) - \varepsilon(2) \le \mu_3(X) - E(\varepsilon(3) \mid \mathcal{I}_2)] \mid \mathcal{I}_1\right].$$

The agent stops at stage 2 if the inequality is reversed in the previous expression and

(9b)
$$\mu_2(X) - \varepsilon(2) > \mu_3(X) - E(\varepsilon(3) \mid \mathcal{I}_2).$$

The agent stops at stage 3 if the inequality is reversed in both previous expressions.

Observe that if the $\varepsilon(i)$ are independently distributed, as is assumed in Keane and Wolpin (1997), the expression on the right-hand side of inequality (9a) can be written as

$$(\mu_2(X) - E[\varepsilon(2) \mid \mu_2(X) - \varepsilon(2) > \mu_3(X)]) \Pr(\mu_2(X) - \varepsilon(2) > \mu_3(X))$$
$$+ \mu_3(X) \Pr(\mu_2(X) - \varepsilon(2) \leq \mu_3(X)).$$

The right hand side of (9a) in this case does not depend on $\varepsilon(1)$ and is clearly separable in X and $\varepsilon(1)$. Thus it is possible to represent this version of the dynamic discrete choice model under uncertainty using the ordered choice model. However, the independence of the shocks is crucial to this example. Suppose instead that $\varepsilon(2)$ and $\varepsilon(1)$ are dependent. Learning about $\varepsilon(1)$ changes the expression on the right-hand side of equation (9a). In general, $\varepsilon(1)$ is in the conditioning set on the right-hand side interacted with X (via $\mu_2(X)$ and $\mu_3(X)$) and is clearly on the left-hand side of the expression. This modification generates a fundamental nonseparability in the unobservables of the model. A key requirement of the classical ordered choice model is violated. If $\varepsilon(j)$ is a random walk

$$\varepsilon(2) = \omega(2) + \varepsilon(1), \qquad\qquad \omega(2) \perp\!\!\!\perp \varepsilon(1)$$
$$\varepsilon(3) = \omega(3) + \varepsilon(2), \qquad\qquad \omega(3) \perp\!\!\!\perp \varepsilon(2),$$

the expression on the right-hand side of (9a) simplifies, since $E(\varepsilon(3) \mid \mathcal{I}_2) = \varepsilon(2)$, so it can be written as

$$(\mu_2(X) - \varepsilon(1)) \, \mathbf{1}\, (\mu_2(X) > \mu_3(X)) + (\mu_3(X) - \varepsilon(1)) \, \mathbf{1}\, (\mu_2(X) > \mu_3(X)),$$

which is clearly nonseparable in $\varepsilon(1), X$. However, the optimal decision does not depend on $\varepsilon(2)$ and $\varepsilon(3)$.

These examples illustrate the sensitivity of the stochastic structure of the choice model

to the specification of agent information sets and learning rules. As developed in section 2.4, the requirements built into decision rules (9a) and (9b) that a person who drops out cannot return to a state, is sometimes artificial. For example, if at the end of period 3, suppose that the agent who has dropped out in a previous period gets a very favorable draw of $\varepsilon(3)$. Suppose that there is recall. Since $\varepsilon(1)$, $\varepsilon(2)$, $\varepsilon(3)$ are known at the end of the period, the agent will be back in a static decision world. He could optimally return to school. This possibility is ruled out in the ordered choice model, but it arises in a variety of dynamic discrete choice models in economics.[18]

However, there are many events where reentry is not possible, e.g., a person can die only once, a company can only be founded once. For these and other examples, the ordered choice model is a useful framework. Even for events that are not irreversible, after a stage, irreversibility may characterize the generating process and the ordered choice model may adequately characterize it.

2.6 Summary of the results of this section

The ordered choice model (3) with stochastic thresholds that satisfy the stochastic monotonicity condition (A-1), can be used to represent a variety of interesting economic choice models. An essential feature of these models is that decisions about the choice of a state can be made by comparing (expected) returns in the adjacent ordered states. In a model with sequential information updating, this requires that one stage look ahead rules be optimal. We next present conditions for the nonparametric identification of our generalization of the ordered choice model.

---

[18]See the survey in Abbring and Heckman (2007).

3 Nonparametric Identification of the Generalized Ordered Choice Model

Assumption (A-1) is essential for the definition of a coherent discrete choice model. In general, (A-1) imposes restrictions on the dependence between the η_j and the Q_j for $j > 1$. One cannot freely specify the c_j, Q_j and η_j without violating the assumption. The dependence induced by (A-1) must be addressed in any proof of identification of the ordered choice model.

It is easy to satisfy (A-1) in a variety of leading cases. Thus in the conventional ordered choice model with $\eta_j \equiv 0$, $j = 1, \ldots, \bar{S} - 1$ and with $c_j(Q_j) = \bar{c}_j$, the same constant for all Q_j, condition (A-1) is satisfied if the $\bar{c}_j$ are properly ordered. Even if $c_j(Q_j)$ is a nontrivial function of Q_j, (A-1) is satisfied, and the model is coherent if the restriction is imposed in estimation. When $c_j(Q_j) = \bar{c}_j$, a constant, and η_j is general, (A-1) requires that $\eta_{j+1} + \bar{c}_{j+1} \geq \eta_j + \bar{c}_j$ for all $j = 1, \ldots, \bar{S} - 1$. When $c_j(Q_j)$ is a nondegenerate function of Q_j and the η_j are nondegenerate, establishing nonparametric identifiability becomes more difficult, but is still possible. One case where (A-1) is satisfied and η_j, η_{j+1} are independent of Q_j, Q_{j+1} occurs when $\eta_{j+1} \geq \eta_j$, $j = 1, \ldots, \bar{S}-1$, almost everywhere and $c_{j+1}(Q_{j+1}) > c_j(Q_j)$ almost everywhere. This case is a strong form of the "no news is good news" assumption.

We first prove nonparametric identification under assumptions that cover all of these cases. We denote the support of a variable as "Supp". We collect all of these cases into assumption (A-7):

(A-7) *For all $j = 1, \ldots, \bar{S} - 1$, one of the following holds*

   i. $\eta_j \equiv 0$, $c_j(Q_j) = \bar{c}_j$, $\bar{c}_{j+1} \geq \bar{c}_j$; or

   ii. $\eta_j \equiv 0$, $c_{j+1}(Q_{j+1}) \geq c_j(Q_j)$; or

   iii. $\Pr(\eta_{j+1} \geq \eta_j) = 1$, $c_j(Q_j) = \bar{c}_j$, $\bar{c}_{j+1} \geq \bar{c}_j$; or

   iv. $\Pr(\eta_{j+1} + \bar{c}_{j+1} \geq \eta_j + \bar{c}_j) = 1$, $\bar{c}_{j+1} \geq \bar{c}_j$; or

$v.$ $\Pr(\eta_{j+1} \geq \eta_j) = 1$, $c_{j+1}(Q_{j+1}) \geq c_j(Q_j)$.

Clearly the first three cases are special versions of *(iv)* and *(v)*. We distinguish them because these simpler cases are likely to be used in applied work. For any of these cases, we prove the following theorem.

Theorem 2. *Assume that one of the conditions in (A-7) holds, and in addition,*

    i. The $\{\eta_s\}_{s=1}^{\bar{S}-1}$ are absolutely continuous with respect to Lebesgue measure and have finite means, $E(\eta_1) = 0$ (alternatively, the median or mode is zero), $\eta_{\bar{S}} \equiv 0$; $\mathrm{Supp}(\eta_s) \subseteq [\underline{\eta}_s, \bar{\eta}_s]$ and $\mathrm{Supp}(\eta) = \mathrm{Supp}(\eta_1, \ldots, \eta_{\bar{S}-1}) = \mathrm{Supp}(\eta_1) \times \cdots \times \mathrm{Supp}(\eta_{\bar{S}-1})$;

    ii. $\eta_j \perp\!\!\!\perp (Z, Q)$;

    iii. $\mathrm{Supp}(\eta_s) \subseteq \mathrm{Supp}(\varphi(Z) - c_s(Q_s))$ for each $Q = q$, $s = 1, \ldots, \bar{S} - 1$;

    iv. (A-1), where $c_0(Q_0) = -\infty$; and $c_{\bar{S}}(Q_{\bar{S}}) = \infty$ for all Q_0 and $Q_{\bar{S}}$; and $\eta_{\bar{S}} = 0$;

    v. $\mathrm{Supp}(Q, Z) = \mathrm{Supp}(Q_1) \times \cdots \times \mathrm{Supp}(Q_{\bar{S}-1}) \times \mathrm{Supp}(Z)$;

    vi. $c_s(Q_s) = 0$ at known $Q_s = \bar{q}_s$, $s = 1, \ldots, \bar{S} - 1$; $\bar{q}_s$ is in the support of $c_s(Q_s)$;

    vii. $\varphi(Z)$, $c_s(Q_s)$, $s = 1, \ldots, \bar{S} - 1$, are members of the Matzkin class of functions (1992) defined in Appendix B (i.e., they satisfy one of the conditions 1–4 in that Appendix);

    viii. $U_I \equiv 0$ (Normalization).

    Then the $\varphi(Z)$, $c_s(Q_s)$, $s = 1, \ldots, \bar{S} - 1$, are identified over their supports and the distributions of the η_j, F_{η_j}, $j = 1, \ldots, \bar{S} - 1$ are identified up to an unknown mean.

Proof. See Appendix A. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

Matzkin's assumptions set the scale of the functions. One can weaken her assumptions and obtain identification up to scale. If we relax *(v)*, we can still identify components of $\varphi(Z)$ and the $c_j(Q_j)$, $j = 1, \ldots, \bar{S} - 1$, or the combined functions $\varphi(Z) - c_j(Q_j)$, without identifying

the individual components. Assumption *(vi)* and the normalization of the mean of η_1 set the location parameters. The classical ordered choice model $c_j(Q_j) \equiv 0$, $j = 1, \ldots, \bar{S} - 1$, $\eta_1 = U_I$, follows as a trivial case of Theorem 2. The case of deterministic thresholds ($\eta_1 = U_I$ but $c_j(Q_j)$ nontrivial functions of the Q_j) follows as a separate case of the theorem. So does a model with $c_j(Q_j) = 0$, $j = 1, \ldots, \bar{S} - 1$, and stochastic thresholds. (The η_j are nondegenerate random variables with $\eta_{j+1} \geq \eta_j$, $j = 1, \ldots, \bar{S} - 1$.) The theorem also applies when $\eta_{j+1} \geq \eta_j$ and $c_{j+1}(Q_{j+1}) \geq c_j(Q_j)$, $j = 1, \ldots, \bar{S} - 1$, independently of each other. Under the alternative set of assumptions embodied in (A-7), there is no contradiction between condition *(ii)* and condition (A-1).

The model can be nonparametrically identified for more general cases that satisfy assumption (A-1). We now produce a model where η_j and Q_j are dependent and hence fail assumption *(ii)* in Theorem 2, but the ordered choice model is nonparametrically identified. It constructs the η_j from a hyperpopulation of latent random variables that in general do not satisfy (A-1), but are sampled by a known rule to generate a population that satisfies (A-1). The population so generated represents an economic environment where cost shocks increase at progressive stages. This could be associated with deteriorating skills or marketability or rising direct and psychic costs of schooling with age. Since the sampling rule is known, it is possible to account for it and establish identification.

Assume a hyperpopulation of latent random variables (η_j^*, Q_j^*), $j = 1, \ldots, \bar{S} - 1$, where the population of observed (η_j, Q_j) is generated by a recursive sampling rule from the hyperpopulation that generates random variables that satisfy condition (A-1). We call this model **(S)**.

(S) $\quad \begin{cases} (\eta_1, Q_1) = (\eta_1^*, Q_1^*) \\ (\eta_j, Q_j) = (\eta_j^*, Q_j^*) \quad \text{if} \quad \eta_{j-1}^* + c_{j-1}(Q_{j-1}^*) \leq \eta_j^* + c_j(Q_j^*) \quad j = 2, \ldots, \bar{S} - 1. \end{cases}$

No restrictions are imposed on (η_1^*, Q_1^*) by the sampling rule.

We assume that $\eta^* = (\eta_1^*, \ldots, \eta_{\bar{S}-1}^*)$ has mutually independent components and is inde-

26

pendent of $Q^* = (Q_1^*, \dots, Q_{\bar{S}-1}^*)$ and Z.[19] Letting "$\perp\!\!\!\perp$" denote independence, we assume that in the hyperpopulation,

(A-8) $\eta^* \perp\!\!\!\perp (Q^*, Z)$.

As a consequence of (**S**) and (A-8), the density of η_2 given $Q_2 = q_2$ and $Q_1 = q_1$ is

$$(10) \qquad g(\eta_2 \mid Q_2 = q_2, Q_1 = q_1) = \frac{f_{\eta_2^*}(\eta_2) \int_{-\infty}^{\eta_2 + c_2(q_2) - c_1(q_1)} f_{\eta_1^*}(\tau) \, d\tau}{K(q_2, q_1)},$$

where

$$K(q_2, q_1) = \int_{-\infty}^{\infty} f_{\eta_2^*}(\eta_2) \int_{-\infty}^{\eta_2 + c_2(q_2) - c_1(q_1)} f_{\eta_1^*}(\tau) \, d\tau \, d\eta_2.$$

The dependence among the η_j and the Q arises from the sampling process (**S**).

The Q_s, $s = 1, \dots, \bar{S} - 1$, are assumed to be observed by the econometrician. As done before, we can absorb U_I into the η_j; alternatively, we set $U_I \equiv 0$. We now establish nonparametric identification of this model. As in the proof of Theorem 2, we use many standard assumptions from the discrete choice literature. We prove the following theorem under assumption (**S**).

Theorem 3. *Assume that*

    i. The $\{\eta_s^*\}_{s=1}^{\bar{S}-1}$ are mutually independent absolutely continuous random variables and have finite means. Assume $E(\eta_1^*) = 0$. (Alternatively, the median or mode of η_1 is known.) $\eta_{\bar{S}}^* \equiv 0$; $\eta_s^* \in [\underline{\eta}_s^*, \bar{\eta}_s^*]$ for $s = 1, \dots, \bar{S} - 1$;

    ii. (A-8);

    iii. $\mathrm{Supp}(\eta_s^*) \subseteq \mathrm{Supp}(\varphi(Z) - c_s(Q_s^*))$ for $s = 1, \dots, \bar{S} - 1$ for each $Q_s^* = q_s$ and for each $Z = z$;

    *iv. Selection rule (**S**) holds;*

---

[19]It is possible to relax the independence assumption, but it simplifies the analysis to maintain it. Sampled η are dependent.

v. $\varphi(Z)$, $c_s(Q_s^*)$, $s = 1, \ldots, \bar{S} - 1$, *are members of the Matzkin class of functions (1992) defined in Appendix B (i.e., they satisfy one of the conditions 1–4 in that Appendix);*

vi. $\mathrm{Supp}(Q^*, Z) = \mathrm{Supp}(Q_1^*) \times \cdots \times \mathrm{Supp}(Q_{\bar{S}-1}^*) \times \mathrm{Supp}(Z)$, $s = 1, \ldots, \bar{S} - 1$;

vii. $c_s(Q_s) = 0$ *at known* $Q_s = \bar{q}_s$, $s = 1, \ldots, \bar{S} - 1$; $\bar{q}_s$ *is in the support of* $c_s(Q_s)$;

viii. $U_I \equiv 0$ *(normalization).*

Then the $\varphi(Z)$, $c_s(Q_s)$, $s = 1, \ldots, \bar{S} - 1$ *are identified over their supports and the distributions of the* η_j, F_{η_j}, $j = 1, \ldots, \bar{S} - 1$ *are identified as are the distributions* $F_{\eta_j^*}$, $j = 1, \ldots, \bar{S} - 1$.

Proof. See Appendix A. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Other assumptions about the arrival of new information rationalize the ordered choice model and produce a model that can be nonparametrically identified. These assumptions allow for some news to be good news, but not too good. One can generate the η_j from the process

$$(11) \qquad \eta_j = -c_j(Q_j) + \eta_{j-1} + \omega_j, \ \eta_0 = 0, \ c_0(Q_0) = 0 \quad \text{and} \quad j > 1$$

where $\omega_j \geq 0$, $j = 1, \ldots, \bar{S} - 1$ is a nonnegative random variable assumed to be independent of Q and η_{j-1} and $c_j(Q_j) \geq c_{j-1}(Q_{j-1})$, $j = 2, \ldots, \bar{S} - 1$. Array the ω_j into a vector ω. Assume for this process that

$(A\text{-}9) \quad (a) \ \omega \perp\!\!\!\perp (Q, Z) \ and \ (b) \ \omega_j \perp\!\!\!\perp \omega_{j'} \ \forall \ j \neq j', \ j, j' = 1, \ldots, \bar{S} - 1.$

It is straightforward to establish identification of the model using the argument in Theorem 2. Effectively, this model replaces η_j with $\sum_{\ell=1}^{j} \omega_\ell$ and eliminates the $c_j(Q_j)$ so that it is a version of case *(iii)* of assumption (A-7). Generating the η_j in this fashion essentially removes transition-specific regressors from the model and hence we lose identifiability of $c_j(Q_j)$. We can identify the marginal distributions of the ω_j, $j = 1, \ldots, \bar{S} - 1$ by applying deconvolution to specification (11) applied to the successive marginal distributions of the η_j.

4 Adjoining s-Specific Outcomes

Associated with each choice s is an associated outcome vector $Y(s, W)$. The outcomes can be binary (e.g. employment indicators), continuous variables (present values), durations or any combination of such variables.[20] This includes the case where the $Y(s, W)$ are, for example, the net present values associated with each completed schooling level, $Y(s, W) = R(s, W)$ in the notation of Section 2. Write

$$Y(s, W) = \mu(s, W) + U(s), \; s = 1, \ldots, \bar{S},$$

where $U(s) \in [\underline{U}(s), \bar{U}(s)]$. In addition to choice-specific outcomes we may have access to a vector of measurements $M(W)$ that do not depend on s. We write

$$M(W) = \mu_M(W) + U_M,$$

where $U_M \in [\underline{U}_M, \bar{U}_M]$. We assume that $E(U(s)) = 0$, $s = 1, \ldots, \bar{S}$, $E(U_M) = 0$ and

(A-10) $(Z, Q, W) \perp\!\!\!\perp (U(s), U_M)$, $s = 1, \ldots, \bar{S}$.

In this section, we allow for the possibility that W contains variables distinct from (Z, Q).

The analysis of Section 3 presents conditions for identifying the marginal distribution of each η_s, $s = 1, \ldots, \bar{S} - 1$, up to scale. We can identify the marginal distribution of the $U(s)$ using the limit set arguments developed in Carneiro, Hansen and Heckman (2003, Theorem 3). Thus we can identify $\mu(s, W)$, $s = 1, \ldots, \bar{S}$, the marginal distributions of $U(s)$, $\mu_M(W)$, the marginal distribution of U_M, and the joint distribution of $\left(U(s), U_M, \{\eta_j\}_{j=1}^s\right)$ using the analysis in their Theorem 3. They assume that it is possible to vary $\mu_M(W)$, $\mu(s, W)$, $\varphi(Z)$ and $c_j(Q_j)$ freely and attain a limit set that produces $\Pr(S = s \mid Z, Q) = 1$.

To sketch their proof structure, note that from information on $D(s) = 1, W, Y(s, W), Z, Q,$

---

[20]We can develop the analysis for discrete components of outcomes using the analysis of Carneiro, Hansen and Heckman (2003). They use latent variables crossing thresholds to generate the discrete variables and identify the latent variables and their distribution up to an unknown scale.

we can construct $\Pr(D(s) = 1 \mid W, Z, Q)$ and $\Pr(Y(s, W) \leq y(s, W), M(W) \leq m(x) \mid D(s) = 1, W = w, Z = z)$. In this notation, the joint distribution of $Y(s, W)$, $M(W)$, $D(s) = 1$, $s = 1, \ldots, \bar{S} - 1$, given $W = w$, $Z = z$, and $Q = q$ multiplied by the probability that $D(s) = 1$ can be written as

$$\Pr \left(\begin{array}{c|c} Y(s, W) \leq y(s, W) & D(s) = 1, W = w, \\ M(W) \leq m(x) & Q = q, Z = z \end{array} \right) \cdot \Pr(D(s) = 1 \mid W = w, Q = q, Z = z)$$

$$= \int_{\underline{U}(s)}^{y(s,w) - \mu(s,w)} \int_{\underline{U}_m}^{m(w) - \mu_m(w)} \int_{(\eta_1, \ldots, \eta_s) \in \Gamma} f_{U(s), U_M, \eta}(u(s), u_m, \eta_1, \ldots, \eta_s) \, d\eta_s \cdots d\eta_1 \, du_m \, du_s,$$

where $\Gamma = \{(\eta_1, \ldots, \eta_s) \mid \eta_1 + c_1(q_1) - \varphi(z) < \eta_2 + c_2(q_2) - \varphi(z) < \cdots < \eta_s + c_s(q_s) - \varphi(z)\}$. We assume either (A-7) characterizes the model or condition (**S**), in which case we interpret the η_j as η_j^* in this section.

Assume that we can freely vary the arguments of this expression in the following sense:

(A-11) $\text{Supp}(\mu(s, W), \mu_m(W), \varphi(Z) - c_1(Q_1), \ldots, \varphi(Z) - c_s(Q_{\bar{S}-1})) =$

$\text{Supp}(\mu(s, W)) \times \text{Supp}(\mu_m(W)) \times \text{Supp}(\varphi(Z) - c_1(Q_1)) \times \cdots \times \text{Supp}(\varphi(Z) - c_{\bar{S}-1}(Q_{\bar{S}-1}))$

and that the supports of the latent random variables in the underlying hyperpopulation are not restricted:

(A-12) $\text{Supp}(U(1), \ldots, U(\bar{S}), U_M, \eta_1, \ldots, \eta_{\bar{S}-1})$
$= \text{Supp}(U(1)) \times \cdots \times \text{Supp}(U(\bar{S})) \times \text{Supp}(U_M) \times \text{Supp}(\eta_1) \times \cdots \times \text{Supp}(\eta_{\bar{S}-1})$, *where this condition applies to all components.*

Assumptions (A-10), (A-11) and (A-12), coupled with the assumptions used in either Theorem 2 or Theorem 3, along with the requirement that there are no restrictions on the support of the components of $M(W)$ and $Y(s, W)$, produce identification of the means, the joint distributions of the $\left(U(s), U_M, \{\eta_j\}_{j=1}^s\right)$, $s = 1, \ldots, \bar{S} - 1$ and to identify the joint distribution of $\left(U(\bar{S}), U_M, \{\eta_j\}_{j=1}^{\bar{S}-1}\right)$. The proof is a straightforward extension of proofs in the

published literature.[21] For the sake of brevity, it is deleted.

From the limit sets that drive $\Pr(D(s) = 1|Z = z, Q = q)$ to 1, $s = 1, \ldots, \bar{S}$, one can identify the average treatment effects across different outcome states $E(Y(s) - Y(s') \mid W)$. The marginal treatment effects for transitions $(s, s+1)$, $s = 1, \ldots, \bar{S} - 1$ can be identified by applying the local instrumental variable method following Heckman, Urzua and Vytlacil (2006b), or Heckman and Vytlacil (2007) or directly by using the argument of Carneiro, Hansen and Heckman (2003). The parameters Treatment on the Treated or Treatment on the Untreated require information on the joint distributions of random variables like $(U(\ell), \eta_1, \ldots, \eta_{\bar{S}-1})$.[22]

If we use the model based on independent latent censored variables as described in condition (**S**), we can identify the joint density of the $\eta = (\eta_1, \ldots, \eta_{\bar{S}-1})$ under the conditions of Theorem 3. We can identify the scales and all of the marginal densities, given the normalizations for the η, using the limit set argument. We can identify the joint distributions of $(U(s), \eta_s)$ for each outcome state, $s = 1, \ldots, \bar{S} - 1$, by setting $c_{s-1}(Q_{s-1}) \rightarrow \underline{\eta}_{s-1}$ and $c_{s+1}(Q_{s+1}) \rightarrow \bar{\eta}_s$. These joint distributions do not contain the information required to form the full joint distribution of $(U(s), \eta_1, \ldots, \eta_s)$.[23]

5 Summary and Conclusions

This paper examines the economic foundations of ordered discrete choice models. The classical ordered discrete choice model is generalized to accommodate stochastic thresholds and associated outcome variables. We develop conditions for nonparametric identification. We

---

[21] See Carneiro, Hansen and Heckman (2003) and Heckman and Navarro (2007).

[22] See the discussion in Heckman and Navarro (2007).

[23] Under a factor structure assumption and under conditions specified in Carneiro, Hansen and Heckman (2003) and Heckman and Navarro (2007), we can identify the factor loadings as well as distribution of the factors and uniquenesses from data on $Y(s, W)$, for each $s = 1, \ldots, \bar{S}$, and any associated measurements $M(W)$. If we assume a factor model for the choice process and a corresponding structure for measurements and outcomes, then we can identify the covariances between $U(s)$ and $(\eta_1, \ldots, \eta_s)$. This requires a restriction on the dimension of the admissible factors. Under the factor structure assumption and with suitable restrictions on the dimension of the model, we can identify the joint distribution of $(U(1), \ldots, U(\bar{S}), U_M, \eta_1, \ldots \eta_{\bar{S}-1})$. From this information, and the parameters previously identified, we can form all of the desired counterfactuals, applying the analysis of Carneiro, Hansen and Heckman (2003) and Heckman and Navarro (2007).

discuss classes of interesting economic models that can be represented by the generalized ordered choice model. We also develop restrictions on information processing and the arrival of new information that are required to justify the application of the generalized ordered choice model to adequately represent dynamic discrete choice models.

There are two key requirements for the model. The first is that local comparisons between the rewards of adjacent states locate the global optimum. In a deterministic setting, this is justified by global concavity, where the unobservables respect a stochastic monotonicity condition. In an environment of uncertainty, assumptions that produce a monotone optimal stopping condition justify a naive one-step-ahead forecasting rule as a way of characterizing optimal policies. The local comparisons used in the ordered choice model contrast with more general choice frameworks, which rely on global comparisons. A second requirement is separability between observables and unobservables. We conjecture that it is possible to relax separability, but we leave the analysis for another occasion.[24] Separability is a hallmark feature of the classical ordered choice model and we maintain it for the sake of familiarity and ease of analysis.

Address correspondence to: James Heckman, University of Chicago, Department of Economics, 1126 East 59th Street, Chicago, IL 60637, USA; Tel.: +1-773-702-0634, Fax: +1-773-702-8490, E-mail: jjh@uchicago.edu. Also affiliated with the American Bar Foundation and University College, Dublin. Flavio Cunha is at the University of Pennsylvania, Department of Economics, 160 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104, USA, E-mail: flaviocunha@gmail.com. Salvador Navarro is at the University of Wisconsin–Madison, Department of Economics, 1180 Observatory Drive, Madison, WI 53706, USA, E-mail: snavarro@ssc.wisc.edu.

---

[24]Cameron and Heckman (1998) relax separability in one version of the ordered choice model. See also Carneiro, Hansen and Heckman (2003).

Appendix

A   Proofs of Theorems and Propositions

Proof of Proposition (2)

Proof. We show that this is true by backward induction. Consider an individual with ability level a. We want to calculate the value function when the schooling level is $s > s^*(a)$. We first derive the value functions for the last period $t = T$. Note that

$$
\begin{aligned}
\text{(AP.1)} \quad V(T, s, \bar{x}_{s,T}, a) &= \max\left\{\alpha(s, a) + \gamma x_{s,T} + \Delta r_0(T, s, x_{s,T}), \delta(s, a) + \Delta r_1(T, s)\right\} \\
&= \alpha(s, a) + \gamma x_{s,T} + \Delta r_0(T, s, x_{s,T}).
\end{aligned}
$$

So the proposition is true for $t = T$. Note that the value function at period T of a person who is working, with schooling s, until period $T - 1$ and decides to return to school at period $T - 1$ is

$$
\begin{aligned}
\text{(AP.2)} \quad V(T, s + 1, \bar{x}_{s,T}, a) &= \max\left\{\alpha(s + 1, a) + \Delta r_0(T, s, x_{s,T}), \delta(s, a) + \Delta r_1(T, s)\right\} \\
&= \alpha(s + 1, a) + \Delta r_0(T, s, x_{s,T}).
\end{aligned}
$$

Next we show that the proposition is also true for period $T - 1$. Note that $x_{s,T} = x_{s,T-1} + 1$ if the agent decides to work in period $T - 1$. In period $T - 1$, the problem of the agent is

(AP.3)
$$
V(T - 1, s, \bar{x}_{s,T-1}, a) = \max \left\{ \begin{aligned} &\alpha(s, a) + \gamma x_{s,T-1} + \Delta r_0(T - 1, s, x_{s,T-1}) + \beta E_{T-1}\left[V(T, s, \bar{x}_{s,T}, a)\right], \\ &\delta(s, a) + \Delta r_1(T - 1, s) + \beta E_{T-1}\left[V(T, s + 1, \bar{x}_{s,T-1}, a)\right] \end{aligned} \right\}.
$$

If we substitute (AP.1) and (AP.2) into (AP.3), we conclude that the agent decides not to enroll in school in period $T-1$ if, and only if

$$\alpha\left(s,a\right)+\gamma\left(1+\beta\right)x_{s,T-1}+\Delta r_{0}\left(T-1,s,x_{s,T-1}\right)+\beta\alpha\left(s,a\right)+\beta\gamma\geq$$
$$\delta\left(s,a\right)+\Delta r_{1}\left(T-1,s\right)+\beta\alpha\left(s+1,a\right),$$

which is guaranteed by Proposition 1 for any schooling level $s>s^{*}\left(a\right)$. Consequently,

$$V\left(T-1,s,\bar{x}_{s,T-1},a\right)$$
$$=\sum_{\tau=T-1}^{T}\beta^{\tau-T+1}E\left[r_{0}\left(\tau,s,x_{s,T-1}+\tau-T+1,a\right)\mid s,x_{\tau},a\right]+\Delta r_{0}(T-1,s,x_{s,T-1}).$$

We now show that the proposition also holds for the value function $V\left(T-1,s+1,\bar{x}_{s,T-2},a\right)$. This is the value function of an agent who until period $T-2$ was working at school level s, but at the beginning of period $T-2$ decides to go back to school. Define z_{T} as the vector of length $\bar{S}$ that describes the experience of this agent, who works at period $T-1$ with schooling $s+1$:

$$z_{T}=\left(0,...,x_{s,T-2},1,...,0\right).$$

The value function $V\left(T-1,s+1,\bar{x}_{s,T-2},a\right)$ satisfies

$$V\left(T-1,s+1,\bar{x}_{s,T-2},a\right)=\max\left\{\begin{array}{l}\alpha\left(s+1,a\right)+\Delta r_{0}\left(T-1,s+1,0\right)+\beta E_{T-1}\left[V\left(T,s+1,z_{T},a\right)\right],\\ \delta\left(s+1,a\right)+\Delta r_{1}\left(T-1,s\right)+\beta E_{T-1}\left[V\left(T,s+2,\bar{x}_{s,T-2},a\right)\right]\end{array}\right\}.$$

Now, it can be shown that

$$E_{T-1}\left[V\left(T,s+1,z_{T},a\right)\right]=\alpha\left(s+1,a\right)+\gamma$$

and

$$E_{T-1}\left[V\left(T,s+2,\bar{x}_{s,T-2},a\right)\right]=\alpha\left(s+2,a\right).$$

34

Consequently, the agent does not enroll in school if and only if

$$\alpha\left(s+1,a\right)+\Delta r_0\left(T-1,s+1,0\right)+\beta\alpha\left(s+1,a\right)+\beta\gamma\geq$$

$$\delta\left(s+1,a\right)+\Delta r_1\left(T-1,s\right)+\beta\alpha\left(s+2,a\right).$$

Note that by Proposition 1, this inequality is always true for any schooling level $s\geq s^*\left(a\right)$. Therefore (7) is true for $V\left(T-1,s+1,\bar{x}_{s,T-2},a\right)$.

We seek to prove that the proposition is true for a generic period t, schooling level s, and experience x_t. Before we proceed, we define the jth component of the experience vector x_{t+1} as

$$x_{j,t+1}=\begin{cases}x_{j,t}\text{ if }j<s\\x_{j,t}+1\text{ if }j=s\\0\text{ if }j>s\end{cases}.$$

Note that if the agent decides to work, his next period experience will be denoted by x_{t+1}. If he decides not to work, then his experience vector is summarized by x_t.

To continue with the proof by backward induction, we assume that the claim is true for the value functions $V\left(t+1,s,x_{t+1},a\right)$ and $V\left(t+1,s+1,x_t,a\right)$. From the definition of the Bellman equation,

$$V\left(t,s,x_t,a\right)=\max\begin{cases}\alpha\left(s,a\right)+\Delta r_0\left(t,s,x_{s,t}\right)+\beta E\left[V\left(t+1,s,x_{t+1},a\right)|\,s,x_t,a\right],\\\delta\left(s,a\right)+\Delta r_1\left(t,s\right)+\beta E\left[V\left(t+1,s+1,x_t,a\right)|\,s,x_t,a\right]\end{cases}.$$

But recall that

$$V\left(t,s,x_t,a\right)=\max\begin{cases}\alpha\left(s,a\right)+\Delta r_0\left(t,s,x_{s,t}\right)+\beta\sum_{\tau=t+1}^T\beta^{\tau-t-1}\left[\alpha\left(s,a\right)+\gamma\left(\tau-\left(t+1\right)+x_{s,t}+1\right)\right],\\\delta\left(s,a\right)+\Delta r_1\left(t,s\right)+\beta\sum_{\tau=t+1}^T\beta^{\tau-t-1}\left[\alpha\left(s+1,a\right)+\gamma\left(\tau-t-1\right)\right]\end{cases}.$$

Note that the agent decides not to enroll if and only if

$$\alpha\left(s,a\right)+\Delta r_0\left(t,s,x_{s,t}\right)+\beta\sum_{\tau=t+1}^T\beta^{\tau-t-1}\left[\alpha\left(s,a\right)+\gamma\left(\tau-\left(t+1\right)+x_{s,t}+1\right)\right]$$

$$\geq \delta\left(s, a\right) + \Delta r_1\left(t, s\right) + \beta \sum_{\tau=t+1}^{T} \beta^{\tau-t-1}\left[\alpha\left(s+1, a\right) + \gamma\left(\tau - t - 1\right)\right].$$

This inequality reduces to

$$\alpha\left(s, a\right) + \Delta r_0\left(t, s, x_{s,t}\right) - \delta\left(s, a\right) - \Delta r_1\left(t, s\right) + \frac{1 - \beta^{T-t}}{1 - \beta} x_{s,t} + \frac{1 - \beta^{T-t}}{1 - \beta} \gamma \geq \frac{1 - \beta^{T-t}}{1 - \beta}\left[\alpha\left(s+1, a\right) - \alpha\left(s, a\right)\right],$$

which is true for any $s > s^*\left(a\right)$. Consequently,

$$V\left(t, s, x_t, a\right) = \sum_{\tau=t}^{T} \beta^{\tau-t}\left[\alpha\left(s, a\right) + \gamma\left(x_{s,t} + \tau - t\right)\right] + \Delta r_0\left(t, s, 0\right).$$

□

Proof of Proposition (3)

Proof. Again, by backward induction. Note that from Proposition (2), we conclude that Corollary (1) of Proposition (2), it is also true for $V\left(t, s^*\left(a\right), 0, a\right)$. Next, we assume that (8) is true in period $t' + 1$ with schooling level $s + 1$, such that $s^*\left(a\right) = s + k\left(a\right) - 1$, where $k(a)$ is a nonnegative integer, experience vector $x_{t'+1} = 0$, and ability level is a and show that it is also valid for period t', schooling level s, experience $x_t = 0$, and ability a. By definition,

$$V\left(t', s, 0, a\right) = \max \left\{ \begin{array}{l} \alpha\left(s, a\right) + \Delta r_0\left(t', s, 0\right) + \beta E_{t'}\left[V\left(t' + 1, s, x_{t'+1}, a\right)\right], \\ \delta\left(s, a\right) + \Delta r_1\left(t', s\right) + \beta E_{t'}\left[V\left(t' + 1, s + 1, 0, a\right)\right], \end{array} \right\}$$

where

$$x_{j,t'+1} = \left\{ \begin{array}{ll} 0 & \text{if } j \neq s \\ 1 & \text{if } j = s \end{array} \right. .$$

Again, suppose that the agent does not enroll in school at period t' and starts working with schooling s. We next investigate whether he ever returns to school again, perhaps at period

$t' + p$, for $p = 1, ..., T - t'$. We show that, under our assumptions, this cannot happen. Let

$$
x_{j,t'+p} = \begin{cases} 0 & \text{if } j \neq s \\ p & \text{if } j = s, \text{ for } p = 1, \ldots, T - t' \end{cases}.
$$

Note that at period $t' + p$, the problem of the agent is

$$
V(t' + p, s, x_{t'+p}, a)
$$
$$
= \max \left\{ \begin{array}{c} \alpha(s, a) + \gamma p + \Delta r_0(t' + p, s, x_{s,t'+p}) + \beta E_{t'+p}[V(t' + p + 1, s, x_{t'+p+1}, a)], \\ \delta(s, a) + \Delta r_1(t', s) + \beta E_{t'+p}[V(t' + p + 1, s + 1, x_{t'+p}, a)] \end{array} \right\}.
$$

Note that the agent has accumulated experience only at schooling level s. Because s-specific experience is not useful in sector $s' \neq s$,

$$
V(t' + p + 1, s + 1, x_{t'+p}, a) = V(t' + p + 1, s + 1, 0, a).
$$

The agent does not enroll in school again at period $t' + p$ if, and only if

$$
\text{(AP.4)} \quad \begin{aligned} & \alpha(s, a) + \gamma p - \delta(s, a) + \Delta r_0(t' + p, s, x_{s,t'+p}) - \Delta r_1(t' + p, s) \\ & \geq \beta E_{t'+p}[V(t' + p + 1, s + 1, 0, a) - V(t' + p + 1, s, x_{t'+p+1}, a)]. \end{aligned}
$$

Note that the right-hand side of this equation is the expectation of the difference of two value functions. We assume that the proposition is true for the first value function, $V(t' + p + 1, s + 1, 0, a)$, propose a lower bound for the second value function, $V(t' + p + 1, s, x_{t'+p+1}, a)$, and show that it is not optimal to return to school at period $t' + p$.

Define n^* as the optimal date of dropping out of school:

$$
n^* = \operatorname*{argmax}_{n} \left\{ \begin{array}{c} \sum_{l=1}^{n-1} \beta^{l-1} \delta(s + l, a) + \Delta r_1(t' + p, s) + \beta^n + \\ \sum_{\tau=t'+p+n}^{T} \beta^{\tau-t'-p-n} [\alpha(s + n, a) + \gamma(\tau - t' - p - n)] \end{array} \right\}.
$$

37

Assuming that the proposition is true for period $t' + p + 1$ and schooling level $s + 1$, we can write

$$V(t' + p + 1, s + 1, 0, a)$$

$$= \max \left\{ \begin{array}{l} \Delta r_0 (t' + p + 1, s + 1, 0) + \sum_{\tau=t'+p+1}^{T} \beta^{\tau-t'-p-1} \left[\alpha (s + 1, a) + \gamma (\tau - t' - p - 1) \right], \\ \sum_{l=1}^{n^*-1} \beta^{l-1} \delta (s + l, a) + \Delta r_1 (t' + p, s) \\ \qquad + \beta^{n^*} \sum_{\tau=t'+p+n^*}^{T} \beta^{\tau-t'-p-n^*} \left[\alpha (s + n^*, a) + \gamma (\tau - t' - p - n^*) \right] \end{array} \right\}.$$

Suppose that

(AP.5) $V(t' + p + 1, s + 1, 0, a)$

$$= \Delta r_0 (t' + p + 1, s + 1, 0) + \sum_{\tau=t+p+1}^{T} \beta^{\tau-t'-p-1} \left[\alpha (s + 1, a) + \gamma (\tau - t' - p - 1) \right].$$

Then the agent does not enroll in school in period $t' + p$. To see why, note that we have the following bound:

(AP.6) $E_{t'+p} \left[V(t' + p + 1, s, x_{t'+p+1}, a) \right] \geq \sum_{\tau=t'+p+1}^{T} \beta^{\tau-t'-p-1} \left[\alpha (s, a) + \gamma (\tau - t) \right].$

Now, if (AP.5) is true, we can bound the difference:

$$E_{t'+p} \left[V(t' + p + 1, s + 1, 0, a) - V(t' + p + 1, s, x_{t'+1}, a) \right]$$

$$\leq \sum_{\tau=t+p+1}^{T} \beta^{\tau-t'-p-1} \left[\alpha (s + 1, a) - \alpha (s, a) + \gamma (\tau - t) \right] \leq 0,$$

where the last inequality is guaranteed by (A-5). Because of Condition *(G)* of (A-4), we can also bound the difference:

$$\alpha (s, a) + \gamma p - \delta (s, a) + \Delta r_0 (t' + p, s, 0) - \Delta r_1 (t' + p, s) \geq 0.$$

Thus, under (AP.5), inequality (AP.4) holds and it is not optimal to return to school. Now,

suppose that

$$V\left(t'+p+1, s+1, 0, a\right)$$

$$= \sum_{l=1}^{n^*-1} \beta^{l-1}\delta\left(s+l, a\right) + \Delta r_1\left(t'+p, s\right) + \beta^{n^*} \sum_{\tau=t'+p+n^*}^{T} \beta^{\tau-t'-p-n^*}\left[\alpha\left(s+n^*, a\right) + \gamma\left(\tau - t' - p - n^*\right)\right].$$

Again, we can use (AP.6) to claim that

$$E_{t'+p}\left[V\left(t'+p+1, s, x_{t'+p+1}, a\right)\right]$$

$$\geq \sum_{\tau=t+p+1}^{t+p+n^*-1} \beta^{\tau-t'-p-1}\left[\alpha\left(s, a\right) + \gamma\left(\tau - t\right)\right] + \beta^{n^*} \sum_{\tau=t+p+n^*}^{T} \beta^{\tau-t'-p-n^*}\left[\alpha\left(s, a\right) + \gamma\left(\tau - t\right)\right].$$

Consequently, because of (A-5) the following difference can be bounded above by zero

$$\left\{\sum_{\tau=t'+p+n^*}^{T} \beta^{\tau-t'-p-n^*}\left[\alpha\left(s+n^*, a\right) + \gamma\left(\tau - t' - p - n^*\right)\right] - \sum_{\tau=t+p+n^*}^{T} \beta^{\tau-t'-p-n^*}\left[\alpha\left(s, a\right) + \gamma\left(\tau - t\right)\right]\right\}$$

$$= \sum_{\tau=t'+p+n^*}^{T} \beta^{\tau-t'-p-n^*}\left[\alpha\left(s+n^*, a\right) - \alpha\left(s, a\right) - \gamma\left(p + n^*\right)\right]$$

$$= \frac{1 - \beta^{T-t'-p-n^*+1}}{1 - \beta}\left[\alpha\left(s+n^*, a\right) - \alpha\left(s, a\right) - \gamma\left(p + n^*\right)\right] \leq 0.$$

So the gross returns to going back to school are negative. On the other hand, note that the

gross costs are positive:

$$\alpha\left(s, a\right) - \delta\left(s, a\right) + \gamma p + \Delta r_0\left(t'+p, s, 0\right) - \Delta r_1\left(t'+p, s\right) + \beta^{n^*} \sum_{l=1}^{n^*-1} \beta^{l-1}\left[\alpha\left(s, a\right) + \gamma\left(\tau - t\right) - \delta\left(s+l, a\right)\right]$$

$$= \left\{\alpha\left(s, a\right) - \delta\left(s, a\right) + \gamma p + \Delta r_0\left(t'+p, s, 0\right) - \Delta r_1\left(t', s\right)\right\} + \left\{\beta^{n^*} \sum_{l=1}^{n^*-1} \beta^{l-1}\left[\alpha\left(s+l, a\right) - \delta\left(s+l, a\right)\right]\right\}$$

$$+ \left\{\beta^{n^*} \sum_{l=1}^{n^*-1} \beta^{l-1}\left[\alpha\left(s, a\right) + \gamma\left(\tau - t\right) - \alpha\left(s+l, a\right)\right]\right\} \geq 0.$$

Note that because of Condition *(G)* of (A-4), the first term (in braces) in the final expression

before the inequality is nonnegative. The second term is nonnegative because of condition (F) of (A-4). The third part is nonnegative because of (A-5). Consequently, if the agent drops out of school at period t' with school level s, he will never return. This implies that for any period $t' + p$, schooling level s, and experience vector $x_{t'+p} \neq 0$ that satisfies

$$x_{j,t'+p} = \begin{cases} 0 \text{ if } j \neq s \\ p \text{ if } j = s \end{cases} \quad \text{for } p = 1, ..., T - t'.$$

The value function $V(t' + p, s, x_{t'+p}, a)$ satisfies

$$V(t' + p, s, x_{t'+p}, a) = \alpha(s, a) + \gamma p + \Delta r_0(t' + p, s, x_{s,t'+p}) + \beta \sum_{\tau=t'+p+1}^{T} \beta^{\tau-t'-p-1} [\alpha(s, a) + \gamma(\tau - t')].$$

In particular, note that

(AP.7)

$$V(t' + 1, s, x_{t'+1}, a) = \alpha(s, a) + \gamma + \Delta r_0(t' + 1, s, x_{s,t'+1}) + \beta \sum_{\tau=t'+1}^{T} \beta^{\tau-t'-1} [\alpha(s, a) + \gamma(\tau - t' + 1)].$$

This is important because by definition of the Bellman equation,

$$V(t' + 1, s + 1, 0, a) = \max \left\{ \begin{array}{c} \alpha(s + 1, a) + \Delta r_0(t' + 1, s + 1, 0) + \beta E_{t'}[V(t' + 2, s, x_{t'+2}, a)], \\ \delta(s + 1, a) + \Delta r_1(t' + 1, s + 1) + \beta E_{t'}[V(t' + 2, s + 2, 0, a)] \end{array} \right\}.$$

Thus, we can write

$$V(t' + 1, s + 1, 0, a)$$
$$= \max \left\{ \begin{array}{c} \alpha(s, a) + \Delta r_0(t' + 1, s, x_{s,t'+1}) + \beta \sum_{\tau=t'+2}^{T} \beta^{\tau-t'-2} [\alpha(s, a) + \gamma(\tau - t' - 1)], \\ \delta(s + 1, a) + \Delta r_1(t' + 1, s + 1) + \beta E_{t'}[V(t' + 2, s + 2, 0, a)] \end{array} \right\},$$

and by substituting the equality (AP.7) sequentially, we conclude that (8) is true for all $s \leq s^*(a)$ as we sought to prove. $\square$

Proof of Proposition (5)

Proof. To prove the claim we must show that an agent who decides to drop out at period t with schooling level s would also choose to remain out of school at school level $s + k$. To see why, consider the one stage look ahead rule

$$\max \left\{ R\left(t, s, 0, a\right), \delta\left(s, a\right) + \Delta r_1\left(t, s, 0\right) + \beta E_t\left[R\left(t + 1, s + 1, 0, a\right)\right] \right\}.$$

Suppose that the agent decides to drop out of school. Then, it must be true that

$$\alpha\left(s, a\right) + \Delta r_0\left(s, s, 0\right) - \delta\left(s, a\right) - \Delta r_1\left(s, s, 0\right) \geq \beta \frac{1 - \beta^{T-s}}{1 - \beta} \left[\alpha\left(s + 1, a\right) - \alpha\left(s, a\right) - \gamma\right].$$

Now, suppose that the agent faces the same choice at some school level $s + k$, $k \geq 1$. Then, the agent would still drop out of school if and only if

$$\alpha\left(s + k, a\right) + \Delta r_0\left(s + k, s, 0\right) - \delta\left(s + k, a\right) - \Delta r_1\left(s, s, 0\right) \geq \beta \frac{1 - \beta^{T-s-k}}{1 - \beta} \left[\alpha\left(s + 1, a\right) - \alpha\left(s, a\right) - \gamma\right].$$

Note that it is always true that, for all nonnegative integers k,

$$\left[\alpha\left(s + k, a\right) - \alpha\left(s, a\right) - \gamma\right] \geq \left[\alpha\left(s + k, a\right) - \alpha\left(s, a\right) - \gamma\right].$$

So, if we know that

$$\alpha\left(s + k, a\right) - \delta\left(s + k, a\right) \geq \frac{1 - \beta^{T-s-k}}{1 - \beta^{T-s}} \left(\alpha\left(s, a\right) - \delta\left(s, a\right)\right)$$

and

$$\Delta r_0\left(s + k, s + k, 0\right) - \Delta r_1\left(s + k, s + k, 0\right) \geq \frac{1 - \beta^{T-s-k}}{1 - \beta^{T-s}} \Delta r_0\left(s, s, 0\right) - \Delta r_1\left(s, s, 0\right),$$

41

then we know that the agent would still decide to drop out of school. But these inequalities are guaranteed by Condition (8). $\square$

Proof of Theorem 2

Proof. Normalize $U_I = 0$ (alternatively absorb it into the η_j). From the assumptions,

$$\Pr\left(D(1) = 1 \mid Z = z, Q = q\right) = \Pr\left(\varphi(z) - c_1(q_1) \leq \eta_1\right).$$

Using Matzkin's (1992) extension of Manski (1988), for the class of functions for $\varphi(Z)$ and $c_1(Q_1)$ defined by Matzkin (1992), we invoke assumptions *(i)*, *(ii)*, *(iii)* to identify F_{η_1} up to an unknown mean and $\varphi(Z) - c_1(Q_1)$ over its support. From *(vii)* we can separately identify $\varphi(Z)$ and $c_1(Q_1)$ up to constants. From *(i)* and *(vi)* we can pin down the constants in $\varphi(Z)$ given that $\bar{q}_1$ is in $\text{Supp}(Q_1)$ by assumption *(vi)* since we fix the location of η_1 by *(i)*. Next consider the event $D(1) + D(2) = 1$ given $Q = q, Z = z$. This can be written as $\Pr(D(1) + D(2) = 1 \mid Z = z, Q = q) = \Pr(\varphi(z) - c_2(q_2) \leq \eta_2)$. We can repeat the argument made for $\Pr(D(1) = 1 \mid Z, Q)$ for this probability. Alternatively, we can vary $\varphi(Z)$ and identify the distribution of $\eta_2 + c_2(q_2)$. $\bar{q}_2$ is in $\text{Supp}(Q_2)$ from assumption *(vi)*. We can identify the distribution of η_2 up to an unknown location parameter. We can identify the location parameter since we know the constant in $\varphi(Z)$. Proceeding in this fashion for $\Pr(D(1) + D(2) + D(3) = 1 \mid Z = z, Q = q)$ and successive probabilities of this type, we establish identifiability of the model. $\square$

Proof of Theorem 3

Proof. Instead of normalizing $U_I = 0$, we can absorb it into the definition of η_j. From the assumptions,

$$\Pr\left(D\left(1\right) = 1 \mid Z = z, Q_1 = q_1\right) = \Pr\left(\varphi\left(z\right) - c_1(q_1) \leq \eta_1\right).$$

Condition (A-1) and sampling rule ($\mathbf{S}$) impose no restriction on (Q_1^*, η_1^*). Using Matzkin's extension of Manski (1988) and the Matzkin class of functions, we invoke conditions *(i)*, *(ii)*, *(iii)*, *(v)* and *(vi)*, we identify F_{η_1} up to its mean ($= 0$), the $\varphi(Z)$ and the $c_1(Q_1)$. The constants in $\varphi(Z)$ and $c_1(Q_1)$ cannot be separated without the information provided by assumption *(vii)*.

Proceeding sequentially, consider the event $D(1) + D(2) = 1$, given $Z = z$, $Q_2 = q_2$, $Q_1 = q_1$. Its probability can be written as $\Pr(D(1) + D(2) = 1 \mid Z = z, Q_2 = q_2, Q_1 = q_1) = \Pr(\varphi(z) - c_2(q_2) \leq \eta_2 \mid Q_2 = q_2, Q_1 = q_1)$. Absorb $c_2(q_2)$ into η_2: $\tilde{\eta}_2 = \eta_2 + c_2(q_2)$. Since we know $\varphi(Z)$ from the first step of the proof, under *(vi)*, we can identify $F_{\tilde{\eta}_2}$. At the point of evaluation $q_2 = \bar{q}_2$, $c_2(q_2) = 0$. We thus obtain the distribution of η_2 and its density as a consequence of *(i)*.

Using $c_2(\bar{q}_2) = 0$ for $Q_2 = \bar{q}_2$, and (10), we obtain for each value of η_2,

$$\frac{f_{\eta_2^*}(\eta_2)}{K(\bar{q}_2, q_1)} = \frac{g(\eta_2 \mid Q_2 = \bar{q}_2, Q_1 = q_1)}{\int\limits_{-\infty}^{\eta_2 - c_1(q_1)} f_{\eta_1^*}(\tau) \, d\tau},$$

where the right-hand side is known for each value of q_1 and η_2. Since $\int_{-\infty}^{\infty} f_{\eta_2^*}(\eta_2) \, d\eta_2 = 1$, we can identify $K(\bar{q}_2, q_1)$ for each q_1 and hence we can identify $f_{\eta_2^*}(\eta_2)$ over the full support of η_2.

To recover $c_2(q_2)$, invoke *(iii)*. Then there exists a limit set $\mathcal{S}(\lim Q_1)$ such that

$$\Pr(D(1) + D(2) = 1 \mid Z = z, Q_1 = q_1, Q_2 = q_2, Q_1 \in \mathcal{S}(\lim Q_1))$$
$$= \Pr(D(2) = 1 \mid Z = z, Q_1 = q_1, Q_2 = q_2, Q_1 \in \mathcal{S}(\lim Q_1))$$
$$= \Pr(\varphi(Z) - c_2(q_2) \leq \eta_2).$$

This limit set drives $c_1(q_1)$ small enough that sampling rule ($\mathbf{S}$) for $j = 2$ is satisfied almost

everywhere and

$$\lim_{c_1(q_1) \to \mathcal{S}(\lim Q_1)} \Pr(D(1) = 1 \mid Z = z, Q_1 = q_1, Q_2 = q_2, Q_1 \in \mathcal{S}(\lim Q_1)) = 0.$$

Proceeding sequentially, we establish the claim in Theorem 3. □

B   The Matzkin Class of Functions

Consider a binary choice model, $D = \mathbf{1}(\varphi(Z) > V)$, where Z is observed and V is unobserved. Let φ^* denote the true φ and let F_V^* denote the true cdf of V. Let $z \in \mathcal{Z}$. Let Γ denote the set of monotone increasing functions from $\Re$ into $[0, 1]$. Matzkin (1992) establishes conditions for identifiability of $\varphi(Z)$.

She shows that the following alternative representations of functional forms satisfying the conditions for exact identification for $\varphi(Z)$. We refer to these as the Matzkin class of functions in the text.

1. $\varphi(Z) = Z\gamma$, $\|\gamma\| = 1$ or $\gamma_1 = 1$, or

2. $\varphi(Z)$ is homogeneous of degree one and attains a given value α, at $Z = z^*$ (e.g. cost functions where $\alpha = 0$ when $Z = 0$), or

3. the $\varphi(Z)$ is a member of a class of least-concave function that attains common values at two points in their domain, or

4. additively separable function, for $\varphi(Z)$:

   (a) functions additively separable into a continuous and monotone increasing function and a continuous monotone increasing, concave and homogeneous of degree one function;

   (b) functions additively separable into the value of one variable and a continuous, monotone increasing function of the remaining variables;

(c) a set of additively separable functions (see Matzkin, 1992, example 5, p.255).

References

Abbring, J. H. and J. J. Heckman, "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, 2007), forthcoming.

Becker, G. S., "Human Capital and the Personal Distribution of Income: An Analytical Approach," Woytinsky Lecture no. 1. Ann Arbor: University of Michigan, Institute of Public Administration, 1967.

Bresnahan, T. F., "Competition and Collusion in the American Automobile Industry: The 1955 Price War," *Journal of Industrial Economics* 35 (1987), 457–482.

Brock, W. A., M. Rothschild and J. E. Stiglitz, "Stochastic Capital Theory," in G. R. Feiwel, ed., *Joan Robinson and Modern Economic Theory* (New York: New York University Press, 1989), 591–622.

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

Card, D., "The Causal Effect of Education on Earnings," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*volume 5 (New York: North-Holland, 1999), 1801–1863.

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

Crawford, D., R. Pollak and F. Vella, "Simple Inference in Multinomial and Ordered Logit," *Econometric Reviews* 17 (1998), 289–299.

Ferguson, T. S., "Optimal Stopping and Applications," Electronic text, found at: http://www.math.ucla.edu/ tom/Stopping/Contents.html, 2003.

Fuss, M. A. and D. McFadden, *Production Economics: A Dual Approach to Theory and Applications* (New York: North-Holland Publishing Company, 1978).

Heckman, J. J., "Effects of Child-Care Programs on Women's Work Effort," *Journal of Political Economy* 82 (1974), S136–S163.

Heckman, J. J., "Comment," in M. Feldstein, ed., *Behavioral Simulation Methods in Tax Policy Analysis* (Chicago: University of Chicago Press, 1983), 70–82.

Heckman, J. J., R. J. LaLonde and J. A. Smith, "The Economics and Econometrics of Active Labor Market Programs," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics* (New York: North-Holland, 1999), 1865–2097.

Heckman, J. J., L. J. Lochner and P. E. Todd, "Earnings Equations and Rates of Return: The Mincer Equation and Beyond," in E. A. Hanushek and F. Welch, eds., *Handbook of the Economics of Education* (Amsterdam: North-Holland, 2006a), 307–458.

Heckman, J. J. and T. E. MaCurdy, "New Methods for Estimating Labor Supply Functions," in R. Ehrenberg, ed., *Research in Labor Economics* (Greenwich, Connecticut: JAI Press, 1981), 65–102.

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

Heckman, J. J., S. Urzua and E. J. Vytlacil, "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics* 88 (2006b), 389–432.

Heckman, J. J. and E. J. Vytlacil, "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in J. Heckman

and E. Leamer, eds., *Handbook of Econometrics, Volume 6* (Amsterdam: Elsevier, 2007), forthcoming.

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

Machin, S. and A. Vignoles, *What's the Good of Education? The Economics of Education in the UK* (Princeton, N.J.: Princeton University Press, 2005).

Manski, C. F., "Identification of Binary Response Models," *Journal of the American Statistical Association* 83 (1988), 729–738.

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

Matzkin, R. L., "Restrictions of Economic Theory in Nonparametric Methods," in R. Engle and D. McFadden, eds., *Handbook of Econometrics* (New York: North-Holland, 1994), 2523–2558.

McFadden, D., "Existence Conditions for Theil-Type Preferences," Unpublished manuscript, Center for Mathematical Economics, University of Chicago, June 1963.

McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics* (New York: Academic Press, 1974).

McFadden, D., "Econometric Models of Probabilistic Choice," in C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications* (Cambridge, MA: MIT Press, 1981).

Prescott, E. C. and M. Visscher, "Sequential Location among Firms with Foresight," *Bell Journal of Economics* 8 (1977), 378–893.

Ridder, G., "The Non-parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies* 57 (1990), 167–181.

Rosen, S., "Human Capital: A Survey of Empirical Research," in R. Ehrenberg, ed., *Research in Labor Economics* (Greenwich, CT: JAI Press, 1977), 3–40.

Shaked, A. and J. Sutton, "Relaxing Price Competition through Product Differentiation," *Review of Economic Studies* 49 (1982), 3–13.

Theil, H., *Theory and Measurement of Consumer Demand* (New York: American Elsevier Publishing Company, 1975).

Theil, H., *Theory and Measurement of Consumer Demand* (New York: American Elsevier Publishing Company, 1976).

Vytlacil, E. J., "Ordered Discrete-Choice Selection Models and Local Average Treatment Effect Assumptions: Equivalence, Nonequivalence, and Representation Results," *Review of Economics and Statistics* 88 (2006), 578–581.

Wicksell, K., *Lectures on Political Economy* (New York: Macmillan Company, 1934).