The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds

By Flavio Cunha, James J. Heckman and Salvador Navarro*
University of Chicago;
University of Chicago, University College Dublin
and the American Bar Foundation;
University of Wisconsin–Madison

This Draft, March 13, 2007

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 of the model. 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.

JEL code: C31

*This paper was previously circulated as part of "Dynamic Treatment Effects" (Heckman and Navarro, 2004) and "Dynamic Discrete Choice and Dynamic Treatment Effects" (Heckman and Navarro, 2005, published 2007). This research was supported by NIH R01-HD043411 and NSF SES-024158. Cunha also acknowledges support from the Claudio Haddad Dissertation Fund at the University of Chicago. The views expressed in this paper are those of the authors and not necessarily those of the funders listed here. Versions of this paper were presented at the UCLA Conference on Panel Data in April 2004, at the Econometrics Study Group, UCL, in London in June 2004, at econometrics seminars at the University of Toulouse in November 2004 and at Northwestern University in April 2005, and at the Festschrift in honor of Daniel McFadden at the University of California at Berkeley in May 2005. Jeremy Fox, Han Hong, Rosa Matzkin and Aureo de Paula provided useful comments on this version. We also thank two anonymous referees and the special editor, Charles Manski, all of whom made very helpful comments on the penultimate draft. Address correspondence to: James Heckman, University of Chicago, Department of Economics, 1126 East 59th Street, Chicago, IL 60637, USA; Senior Fellow, American Bar Foundation, 750 North Lake Shore Drive, Chicago, IL 60611, USA. Tel.: +1-773-702-0634, Fax: +1-773-702-8490, E-mail: *jjh@uchicago.edu*. Flavio Cunha is at the University of Chicago, Department of Economics, E-mail: *flavio@uchicago.edu*. 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*.

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 many 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.[1]

This paper continues the McFadden tradition by examining the economic foundations of widely used ordered discrete choice models. 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 (Prescott and Visscher, 1977; Shaked and Sutton, 1982; Bresnahan, 1987). 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.[2]

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 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 separable in observables (Z) and in

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

[2] Our model generalizes Cameron and Heckman (1998) by introducing regressors and unobservables in the cutoffs or thresholds that generate the ordered choice model.

unobservables (U_I) and is generated by an index

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

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. Individuals select a state $s \in \{1, \ldots, \bar{S}\}$ if the index lies between certain threshold or cutoff values c_s. 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

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

The essential feature of the ordered choice model is that choices are generated by ordered sections of the support of a scalar latent continuous random variable (e.g. durations or hours of work).

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.[3] In an analysis of schooling, there may be grade-specific subsidies and genuine transition-specific uncertainty at the agent level arising from learning about abilities and labor market shocks. Uncertainty is at the core of conventional 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 and (b) variables unobserved by the econometrician.[4] In addition, they adjoin systems of both discrete and continuous outcomes associated with the choice of each state.

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

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

[4] See Carneiro, Hansen, and Heckman (2003, footnote 17). This model is also discussed in Heckman, LaLonde, and Smith (1999). For a recent analysis of this model and its relation to the treatment effect literature, see Vytlacil (2006).

to a variety of economic problems. 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.

An ordered choice model is generated by an index of marginal returns. Marginal returns must be monotone across the ordered states to preserve the ordered choice model structure. In an uncertain environment, the ordered choice structure adequately captures discrete choices only if a "stochastic monotonicity" property developed in this paper holds.[5]

The plan of this paper is as follows. Section 2 presents three ordered choice models to demonstrate the range of economic phenomena that our analysis addresses. The first is a model 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 brackets. The second is a model of agent decision making 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.

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 that generates 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

[5]Heckman and Navarro (2007) develop a more general unordered model with a richer information revelation structure. They establish nonparametric identification of the model.

machine is installed. We present three examples.

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:

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

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

Consumer i is indifferent between two traded 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 emerges for those 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 $\tau X_g - P_g$, $g = 1, \ldots, G$.[6] 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.

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

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 which are determined by exemptions and asset levels, and which may be only partially observed by the econometrician. We next develop a stopping time example which is a vehicle for introducing uncertainty into the model. We begin by developing the case of perfect certainty.

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 the period. 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 earnings foregone.[7]

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 return 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 $\bar{S}-1$, $V\left(\bar{S}-1, X\right)$, is $V\left(\bar{S}-1, X\right) = \max \left\{ R\left(\bar{S}-1, X\right), \frac{R\left(\bar{S},X\right)}{1+r} \right\}$. 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, $R(s-1, X) < \frac{V(s,X)}{1+r}$. This rule produces the global optimum if $R(s)$ is concave in s.

The ordered choice model considers a special case of the general choice problem where pairwise comparisons of returns $R(s, X)$ across adjacent states characterize the optimum. Specifically, it works with the *marginal return function*

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

[7] Agents pay a fixed cost $C(s, X)$ after completing each grade of school and $R(s, X)$ is the return to schooling net of these costs. When $C(s, X) = 0$, the agent 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.

Assuming

(A-1) *The marginal return function $f(s, X)$ is nonincreasing in s for all X,*

the optimum 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 to the left of zero is replaced by a strict inequality.

The general rule for locating an optimum s^* is

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

Under its assumptions, the ordered choice model allows us to replace $V(s^*, X)$ and $V(s^*-1, X)$ in these expressions with $R(s^*, X)$ and $R(s^*-1, X)$, respectively. We analyze these conditions below. Before doing this, we first introduce unobservables into the model.

2.2.1 Introducing unobservables into the model

Unobservables are introduced into the model in two distinct ways. Both preserve the additive-separability-in-unobservables — a scalar random variable structure (1) — that is an essential feature of the 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.[8] 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 an optimal stopping model to be represented by a conventional separable ordered choice model (1), we need to invoke separability in the marginal return function $f(s, X)$ in addition to monotonicity in s.

(A-2) *Assume that the marginal return depends on individual characteristics where $f(s+1, X) = \frac{R(s+1,X)}{1+r} - R(s, X) = -c_s(Q_s) + \varphi(Z) + U_I + \eta_s$, $s \in \{1, \ldots, \bar{S}\}$, where $X = (Q_1, \ldots, Q_{\bar{S}-1}, Z)$, and $E(U_I) = 0$. The Z variables are common across all stopping states. 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.*

[8] Such invariant random variables are called "heterogeneity" in the literature.

Under (A-2), the choice of schooling level s is characterized by

$$(2) \qquad D(s) = \mathbf{1}\left(c_{s-1}(Q_{s-1}) + \eta_{s-1} < \varphi(Z) + U_I \leq c_s(Q_s) + \eta_s\right), \qquad s = 1, \ldots, \bar{S} - 1.$$

The cutoffs must satisfy the restriction $c_{s-1}(Q_{s-1}) + \eta_{s-1} \leq c_s(Q_s) + \eta_s$. We show that 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 set $\eta_j \equiv 0$, $j = 1, \ldots, \bar{S} - 1$. The ordered choice model with stochastic cutoffs was introduced in Carneiro, Hansen, and Heckman (2003) and Heckman, LaLonde, and Smith (1999). It extends the traditional ordered choice model in two fundamental ways: (a) threshold values depend on observed regressors and (b) thresholds are stochastic. The full model is completed by specifying the joint distribution of U_I and the η_s, $s \in \{1, \ldots, \bar{S}\}$.

It is fruitful to compare this model with a general discrete choice model with net returns 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. The optimal s, s^*, is $s^* = \text{argmax}\{R(j)\}_{j=1}^{\bar{S}}$. In the general model, the states are unordered. In the ordered choice specialization of this model,

$$f(s, X) = R(s, X) - R(s-1, X) = \mu_R(s, X) - \mu_R(s-1, X) - (\varepsilon(s) - \varepsilon(s-1))$$

where $\varphi(Z)$ 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 $R(s, X) - R(s-1, X)$ that are $(s-1, s)$-specific and $\eta_s = \varepsilon(s) - \varepsilon(s-1)$. The condition $c_j(Q_j) + \eta_j \leq c_{j+1}(Q_{j+1}) + \eta_{j+1}$ thus restricts the admissible shocks in a general discrete choice model to satisfy

$$(\mathbf{RD}) \qquad c_{j+1}(Q_{j+1}) - c_j(Q_j) \geq 2\varepsilon(j) - \varepsilon(j-1) - \varepsilon(j+1).$$

The agent is envisioned as making choices in an atemporal setting. They draw shocks $\varepsilon(s)$

across all states, $s = 1, \ldots, \bar{S}$, subject to condition (**RD**) and choose according to equation (2). In section 3, we establish conditions under which such a model is nonparametrically identified.

Since the ordered choice model is a version of the GAFT or mixed proportional hazards (MPH) model for discrete durations (Ridder, 1990), which is widely used in applied work on unemployment and other dynamic outcomes, it is of interest to examine how well it captures the sequential arrival of information under uncertainty. We next generalize our analysis to account for uncertainty and agent information updating. Additional assumptions are required to justify the ordered choice model 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

We now extend the separable ordered choice model to accommodate information updating. We first derive the optimal stopping rule in an environment of uncertainty. We assume that s (schooling) is an irreversible investment and that once an agent stops (drops out), he never again continues (drops out). We establish in Appendix A that this is an essential feature of the ordered choice structure. Its inability to handle dropout and return is a limitation of the ordered choice model.

Denote the agent's information set at stage s as $\mathcal{I}_s$. For stage $s = 1, 2, \ldots, \bar{S} - 1$, the value function is

$$V(s, X, \mathcal{I}_s) = \max\left\{R(s, X), \frac{1}{1+r} E\left[V(s+1, X, \mathcal{I}_{s+1}) \mid \mathcal{I}_s\right]\right\}.$$

$R(s, x)$ is assumed to be in the agent's information set $\mathcal{I}_s$. An agent choosing between $\bar{S}$ and $\bar{S} - 1$ chooses to stop at stage $\bar{S}$ if

$$\text{(3a)} \qquad \frac{1}{1+r} E\left[V\left(\bar{S}, X, \mathcal{I}_{\bar{S}}\right) | \mathcal{I}_{\bar{S}-1}\right] - R(\bar{S} - 1, X) \geq 0.$$

Because $\bar{S}$ is a terminal state,

$$\text{(3b)} \qquad E\left[V\left(\bar{S}, X, \mathcal{I}_{\bar{S}}\right) | \mathcal{I}_{\bar{S}-1}\right] = E\left(R\left(\bar{S}, X\right) | \mathcal{I}_{\bar{S}-1}\right).$$

An agent at stage $s-1$ chooses to go on to stage s before stopping if

$$\frac{1}{1+r}E\left[V\left(s,X,\mathcal{I}_s\right)\mid\mathcal{I}_{s-1}\right]-R(s-1,X)\geq 0, \tag{4a}$$

and stops at s if

$$\frac{1}{1+r}E\left[V\left(s+1,X,\mathcal{I}_{s+1}\right)\mid\mathcal{I}_s\right]-R(s,X)<0. \tag{4b}$$

This rule produces the global optimum. The ordered choice model replaces the "V" with "R" in these expressions, i.e., it works with pairwise comparisons of the return functions. In addition, it is additively separable in observables and unobservables. We now present conditions that justify this formulation.

2.3.1 Conditions that justify the ordered choice model as a representation of a dynamic discrete choice model

To preserve the separability intrinsic to the ordered choice model, we augment assumption (A-2) to allow for separable components of uncertainty in a sequential information updating setting:

(A-3) *Assume that the marginal return to stopping at choice $s+1$ depends on both observed and unobserved individual characteristics* $\frac{R(s+1,X)}{1+r}-R(s,X)=-c_s(Q_s)+\varphi(Z)+U_I+\frac{\varepsilon(s+1)}{1+r}-\varepsilon(s)$, $j\in\{1,\ldots,\bar{S}\}$, *where* $X=(Q_1,\ldots,Q_{\bar{S}-1},Z)$, *and* $E(U_I)=0$. *The Z variables are common across all stopping times. The Q_s may depend on the period-specific arguments of $R\left(s+1,X\right)$ and $R\left(s,X\right)$.*

Uncertainty is introduced into the separable ordered choice model by assuming that $\varepsilon(s+1)\notin\mathcal{I}_s$. We assume that $Z\in\mathcal{I}_s$ and $Q_s\in\mathcal{I}_s$. Define marginal cost shocks (negative marginal return shocks) as $\eta_s=-\left[\frac{E(\varepsilon(s+1)|\mathcal{I}_s)}{1+r}-\varepsilon(s)\right]$. Without restrictions on the $\varepsilon(s)$, it is easy to violate the ordered choice structure. In general, the innovations in $\varepsilon(s)$ have no necessary order. To produce an ordered choice model, the stochastic thresholds must be ordered over all stages. This requires a stochastic version of the monotonicity condition (A-1) that was assumed in the deterministic case:

(A-4) *(Stochastic Monotonicity for $j \geq 1$)*

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

where $Q = (Q_1, \ldots, Q_{\bar{S}-1})$. *Dropping common constants on both sides of the inequality, this can be written as* $\Pr\left(c_{j+1}(Q_{j+1}) + \eta_{j+1} \geq c_j(Q_j) + \eta_j | Q = q, Z = z\right) = 1$, $j \geq 1$, *where* $Q = (Q_1, \ldots, Q_{\bar{S}-1})$.

This assumption imposes a restriction on the sample path of the information arriving to agents. It states that the distribution of $\eta_{j+1} - \eta_j$ has to be truncated from below by $c_j(Q_j) - c_{j+1}(Q_{j+1})$. This requires that marginal "cost" shocks have to get worse (larger) over time (if $c_j(Q_j) - c_{j+1}(Q_{j+1}) \geq 0$), or that they cannot improve (if $c_j(Q_j) - c_{j+1}(Q_{j+1}) < 0$). Thus, the news about cost shocks cannot be very good. Under stochastic monotonicity, a sequence of local comparisons of reward functions characterizes the optimum choice as is required by the ordered choice model. Assumption (A-4) does not rule out option values.

The option value at stage s for an agent deciding whether to go forward to $s + 1$ is the expected return of the decision to go forward in excess of the return that would occur from stopping at $s + 1$. In general, agents at stage $s < s^*$, where s^* is the optimum, have positive option values.[9]

Condition (A-4) guarantees that the analyst can write down a coherent likelihood for an ordered model with nonnegative probabilities attached to each state with the property that $\Pr\left(\sum_{s=1}^{\bar{S}} D(s) = 1\right) = 1$. The sequence of marginal returns across stopping times crosses zero at most once.

Conditions (A-3) and (A-4) are implicitly maintained in applying widely used ordered choice and GAFT (or MPH) duration models to represent dynamic discrete choice models. Without allowing for stochastic thresholds, the conventional ordered choice model does not allow for information updating.[10] Thus all versions of the ordered choice model make strong assumptions about the arrival of information and how agents process it.

[9] Strictly speaking, the option values are nonnegative.

[10] If the information being updated by the agent is observable to the econometrician, he can model the agent's expectations in a non-stochastic manner.

Under (A-3) and (A-4), the agent decides to stop at period s if

$$c_{s-1}(Q_{s-1}) - \left[\frac{E\left(\varepsilon(s) \mid \mathcal{I}_{s-1}\right)}{1+r} - \varepsilon(s-1)\right] < \varphi(Z) + U_I \leq c_s(Q_s) - \left[\frac{E\left(\varepsilon(s+1) \mid \mathcal{I}_s\right)}{1+r} - \varepsilon(s)\right],$$

or, using the definition of marginal cost shocks

$$c_{s-1}(Q_{s-1}) + \eta_{s-1} < \varphi(Z) + U_I \leq c_s(Q_s) + \eta_s. \tag{5}$$

Inequality (5) can be rewritten as

$$-\varphi(Z) + c_{s-1}(Q_{s-1}) + \eta_{s-1} - U_I \leq -\varphi(Z) + c_s(Q_s) + \eta_s - U_I, \tag{6}$$

emphasizing that the sample paths of the expected marginal rate of return functions are ordered at the optimum. If $\eta_j \equiv 0$, $j = 1, \ldots, \bar{S}$, so there is no uncertainty, this ordering is implied by concavity assumption (A-1) joined with separability assumption (A-2). Condition (5) is both necessary and sufficient for an ordered choice model to characterize choices, i.e., for local comparisons of expected reward functions to characterize the optimum.

2.3.2 A Three Period Example

A three period (schooling level) example helps to fix ideas. Suppose that the reward function 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 $s = \text{argmax}_j \{R(j)\}_{j=1}^3$. There is no restriction on the $\varepsilon(j)$, $j = 1, 2, 3$. The ordered choice model applied to this setting imposes the restriction (**RD**) given in section 2.2.1.

Next, take the same reward functions but use them in a sequential dynamic discrete choice model with information updating. 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

(7a)

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

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

(7b)
$$\mu_2(X) - \varepsilon(2) > \mu_3(X) - E(\varepsilon(3) \mid \mathcal{I}_2).$$

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 (7a) can be written as

$$\mu_2(X) - E\left(\varepsilon(2) \mid \mu_2(X) - \varepsilon(2) > \mu_3(X)\right)\Pr\left(\mu_2(X) - \varepsilon(2) > \mu_3(X)\right)$$
$$+ \mu_3(X)\Pr\left(\mu_2(X) - \varepsilon(2) > \mu_3(X)\right).$$

The logic of the general choice model does not impose any order on $\varepsilon(1)$, $\varepsilon(2)$, and $\varepsilon(3)$. Now suppose that $\varepsilon(2)$ and $\varepsilon(1)$ are dependent. Learning about $\varepsilon(1)$ changes the expression on the right-hand side of equation (7a). 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. This generates a fundamental nonseparability in the unobservables of the model. A key requirement of the ordered choice model is violated. If $\varepsilon(j)$ is a random walk

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

the expression on the right-hand side of (7a) simplifies, since $E(\varepsilon(3) \mid \mathcal{I}_2) = \varepsilon(2)$, so it can be

written as

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

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 delicacy of the stochastic structure of the choice model to the specification of agent information sets and learning rules. As developed in Appendix A, the requirements built into decision rules (7a) and (7b) that a person who drops out cannot return to the state, is artificial. If at the end of period 3, the agent gets a very favorable draw of $\varepsilon(3)$, since $\varepsilon(1)$, $\varepsilon(2)$, $\varepsilon(3)$ are known, the agent will be in the static decision world of 2.2.1 and could optimally return to school. This possibility is ruled out by the ordered choice model. We next present conditions for the nonparametric identification of the ordered choice model.

3 Nonparametric Identification of the Generalized Ordered Choice Model

Assumption (A-4) is essential for the definition of a coherent discrete choice model. In general, (A-4) 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-4) must be addressed in any proof of identification.

It is easy to satisfy (A-4) 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-4) is satisfied if the $\bar{c}_j$ are properly ordered. Even if $c_j(Q_j)$ is a nontrivial function of Q_j, (A-4) 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-4) 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-4) 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-5):

(A-5) *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. $\Pr(\eta_{j+1} \geq \eta_j) = 1$, $c_{j+1}(Q_{j+1}) \geq c_j(Q_j)$.

For any of these cases, we prove the following theorem.

Theorem 1. *Assume that one of the conditions in (A-5) 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-4), 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 $Q_s = \bar{q}_s$, $s = 1, \ldots, \bar{S} - 1$;

vii. $\varphi(Z)$, $c_s(Q_s)$, $s = 1, \ldots, \bar{S} - 1$, are members of the Matzkin class of functions (1992; 1993; 1994), 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. 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, 1993, 1994), 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 *(vi)* 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 $\varphi(Z)$ provided that $\bar{q}_1$ is in $\text{Supp}(Q_1)$, 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 $\eta_2 + c_2(q_2)$. Assume that $\bar{q}_2$ is in $\text{Supp}(Q_2)$. 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. □

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 1. 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 pure 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-5), there is no contradiction between condition *(ii)* and condition (A-4).

The model can be nonparametrically identified for more general cases that satisfy assumption (A-4). We now produce a model where η_j and Q_j are dependent and hence fail assumption *(ii)* in Theorem 1, but the ordered choice model remains nonparametrically identified. It constructs the η_j from a hyperpopulation of latent random variables that in general do not satisfy (A-4), but are sampled to satisfy (A-4). Since the sampling rule is known, it is possible to account for it and achieve 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-4). We call this model (**S**).

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

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 independent of $Q^* = (Q_1^*, \ldots, Q_{\bar{S}-1}^*)$ and Z.[11] Letting "$\perp\!\!\!\perp$" denote independence, we assume that in the hyperpopulation,

(A-6) $\eta^* \perp\!\!\!\perp (Q^*, Z)$.

As a consequence of (**S**) and (A-6), the density of η_2 given $Q_2 = q_2$ and $Q_1 = q_1$ is

$$(8) \qquad g(\eta_2 \mid Q_2 = q_2, Q_1 = q_1) = \frac{f_{\eta_2^*}(\eta_2) \int_{-\infty}^{\eta_2 + c_2(q_2) - c_1(q_1)} f_{\eta_1^*}(\tau)\, d\tau}{K(q_2, q_1)},$$

where

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

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

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

The Q_s, $s = 1, \ldots, \bar{S} - 1$, are assumed to be observed by the econometrician. 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 1, we use many standard assumptions from the discrete choice literature. We prove the following theorem under assumption (**S**).

Theorem 2. *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* $s = 1, \ldots, \bar{S} - 1$;

ii. (A-6);

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

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

v. $\varphi(Z)$, $c_s(Q_s^*)$, $s = 1, \ldots, \bar{S} - 1$, *are members of the Matzkin class of functions (1992; 1993; 1994), 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* $q_s = \bar{q}_s$, $s = 1, \ldots, \bar{S} - 1$;

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

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

Condition (A-4) and sampling rule (**S**) impose no restriction on (Q_1^*, η_1^*). Using Matzkin's extension of Manski (1988) and the Matzkin class of functions, we invoke *(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 (8), we obtain for each value of η_2,

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

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

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

This limit set drives $c_1(q_1)$ small enough that sampling rule (**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 the Theorem. □

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

$$\eta_j = -c_j\left(Q_j\right) + \eta_{j-1} + \omega_j, \eta_0 = 0,\ c_0\left(Q_0\right) = 0 \quad \text{and} \quad j > 1 \tag{9}$$

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-7) *(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 1. Effectively, this model replaces η_j with $\sum_{\ell=1}^{j} \omega_\ell$ and eliminates the $c(Q_j)$ so that it is a version of case *(iii)* of assumption (A-5). 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 (9) applied to the successive marginal distributions of the η_j.

We can link stochastic process (9) to the model of information updating discussed in section 2.3 by defining η_j as

$$\eta_j = -\left[\frac{E\left(\varepsilon\left(j+1\right) \mid \mathcal{I}_j\right)}{1+r} - \varepsilon\left(j\right)\right],\ j = 1, \ldots, \bar{S} - 1.$$

In the particular case in which $\varepsilon\left(j\right)$ follows a random walk where $\mathcal{I}_j = \{\varepsilon\left(j\right), Q, Z\}$, then $\eta_j = \left(\frac{r}{1+r}\right)\varepsilon\left(j\right)$. To satisfy requirement (A-4) in this case, the random walk on net returns operates under a negative drift.[12]

4 Adjoining s-Specific Outcomes

Associated with each choice s is an associated outcome vector $Y\left(s, X\right)$. The outcomes can be binary (e.g. employment indicators), continuous variables (present values), durations or any

[12] This interpretation requires that there be no unobserved components in Q_j. Otherwise, the η_j contain both the shocks and these components.

combination of such variables.[13] This includes the case where the $Y(s,X)$ are, for example, the net present values associated with each completed schooling level, $Y(s,X) = R(s,X)$ in the notation of Section 2. Write

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

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 $M(X)$ that do not depend on s. We write

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

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-8) $(Z,Q,X) \perp\!\!\!\perp (U(s), U_M)$, $s = 1,\ldots,\bar{S}$.

In this section, we allow for the possibility that X 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 $\mu(s,X)$, $s = 1,\ldots,\bar{S}$, the marginal distributions of $U(s)$, $\mu_M(X)$, 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 $\mu_M(X)$, $\mu(s,X)$, $\varphi(Z)$ 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 $D(s) = 1, X, Y(s,X), Z, Q$, we can construct $\Pr(D(s) = 1 \mid X,Z,Q)$ and $\Pr(Y(s,X) \leq y(s,X), M(X) \leq m(x) \mid D(s) = 1, X = x, Z = z)$. In this notation, the joint distribution of $Y(s,X)$, $M(X), D(s) = 1$, $s = 1,\ldots,\bar{S}-1$, given $X = x$, $Z = z$, and $Q = q$ multiplied by the probability that $D(s) = 1$

[13] 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.

can be written as

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

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

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-5) 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-9) $\text{Supp}\left(\mu(s,X), \mu_m(X), \varphi(Z) - c_1(Q_1), \ldots, \varphi(Z) - c_s(Q_{\bar{S}-1})\right) =$
$\text{Supp}\left(\mu(s,X)\right) \times \text{Supp}\left(\mu_m(X)\right) \times \text{Supp}\left(\varphi(Z) - c_1(Q_1)\right) \times \cdots \times \text{Supp}\left(\varphi(Z) - c_{\bar{S}-1}(Q_{\bar{S}-1})\right)$

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

(A-10) $\text{Supp}\left(U(1), \ldots, U\left(\bar{S}\right), U_M, \eta_1, \ldots, \eta_{\bar{S}-1}\right) = \text{Supp}\left(U(1)\right) \times \cdots \times \text{Supp}\left(U\left(\bar{S}\right)\right) \times \text{Supp}\left(U_M\right) \times \text{Supp}\left(\eta_1\right) \times \cdots \times \text{Supp}\left(\eta_{\bar{S}-1}\right)$, *where this condition applies to all components.*

Assumptions (A-8), (A-9) and (A-10), coupled with the assumptions used in either Theorem 1 or Theorem 2, along with the requirement that there are no restrictions on the support of the components of $M(X)$ and $Y(s,X)$, 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 $\left(U\left(\bar{S}\right), U_M, \{\eta_j\}_{j=1}^{\bar{S}-1}\right)$. The proof is a straightforward extension of proofs in the published literature.[14] 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 $E(Y(s) - Y(s') \mid X)$. 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 (2006), or Heckman and

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

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 $(U(\ell), \eta_1, \ldots, \eta_{\bar{S}-1})$.[15]

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 2. 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 $c_{s-1}(Q_{s-1}) \to \underline{\eta}_{s-1}$ and $c_{s+1}(Q_{s+1}) \to \bar{\eta}_s$. These joint distributions do not contain the information required to form the full joint distribution of $(U(s), \eta_1, \ldots, \eta_s)$.[16]

5 Summary and Conclusions

This paper presents a nonparametric identification analysis for ordered choice models with stochastic thresholds. Such models arise in a variety of areas of applied economics. We develop the assumptions on the economic model that justify application of ordered choice models to characterize choices made by economic agents. We analyze models with atemporal resolution of uncertainty and stopping time models with temporal resolution of uncertainty. These models are prototypes for a general choice model among possible states.

Separability between observables and unobservables is an essential aspect of the ordered choice structure as is the scalar nature of the random variable. So is the requirement that once an agent drops out in a state, he/she does not return. The ordered choice model is based on the assumption that the first time a latent index crosses zero the agent stops the process (e.g. chooses a schooling level), and that the index is monotone across the ordered choices so there is only one

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

[16] 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 $Y(s, X)$, for each $s = 1, \ldots, \bar{S}$, and any associated measurements $M(X)$. 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 $\left(U(1), \ldots, U\left(\bar{S}\right), U_M, \eta_1, \ldots \eta_{\bar{S}-1}\right)$. 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).

zero for the process. In an environment of perfect certainty, this property requires concavity of the criterion function, or equivalently, that the marginal return function decline across choices. In an environment of uncertainty, application of the ordered choice model requires an extension of concavity which we term "stochastic concavity."

Drawing on the analysis of Carneiro, Hansen, and Heckman (2003) and Heckman and Navarro (2007), we adjoin a system of outcome equations associated with each stopping time. This allows identification of average treatment effects and stage-specific marginal treatment effects.

Appendix

A A Stochastic Dynamic Schooling Choice Model

Consider a model in which the individual at each period t decides whether to enroll in school or not.[17] Instead of schooling, we can work with other types of discrete states for which agents decide to remain in the state 0 and then drop out (e.g. a spell of training or a physical therapy program). Let $\mathcal{I}_t$ denote the agent's information set (the state space) at period t. We denote the current schooling level of the individual by s_t. Let $\delta_t = 0$ if the agent decides not to enroll in school in period t and $\delta_t = 1$ otherwise. For each schooling level s_t and enrollment choice δ_t there is an instantaneous reward function $r_t(s_t, \delta_t)$ which summarizes the current net benefits. This can be zero during a year of schooling. The reward function at t is in the information set $\mathcal{I}_t$, $r_t(s_t, k) \in \mathcal{I}_t$. However, the instantaneous reward function can be unknown from the viewpoint of an agent with information set $\mathcal{I}_\tau$ for $\tau < t$. For simplicity, we ignore discounting.

The lifetime problem of the individual is to solve:

$$V_1(s_1, \mathcal{I}_1) = \max_{\{\delta_\tau\}_{\tau=1}^{T}} E\left\{ \sum_{\tau=1}^{T} \left[\delta_\tau r_\tau(s_\tau, 1) + (1-\delta_\tau) r_\tau(s_\tau, 0)\right] \middle| s_1, \mathcal{I}_1 \right\}$$

subject to:

$$s_{\tau+1} = s_\tau + \delta_\tau. \tag{AP.1}$$

[17] See, for example, Heckman and Navarro (2007) and Keane and Wolpin (1997).

The optimal schooling decision generated by this model may require that the agent drop out of school at some year to take advantage of unusually high instantaneous rewards for those not enrolled at school, $r_t(s_t, 0)$, and then return to school at a later date.

Writing the problem of the agent recursively, we obtain

(AP.2)

$$V_t(s_t, \mathcal{I}_t) = \max\{r_t(s_t, 0) + E[V_{t+1}(s_t, \mathcal{I}_{t+1})|\mathcal{I}_t], r_t(s_t, 1) + E[V_{t+1}(s_t + 1, \mathcal{I}_{t+1})|\mathcal{I}_t]\}.$$

Consequently, $\delta_t = 1$ if and only if:

$$r_t(s_t, 1) + E[V_{t+1}(s_t + 1, \mathcal{I}_{t+1})|\mathcal{I}_t, s_t] \geq r_t(s_t, 0) + E[V_{t+1}(s_t, \mathcal{I}_{t+1})|\mathcal{I}_t, s_t].$$

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 $\{\delta_\tau^*(s_t, \mathcal{I}_t)\}_{\tau=1}^T$ by backward induction. In general, conditions (AP.1) and (AP.2) are not sufficient to guarantee that the decision rules generated by the ordered choice model with stochastic thresholds represent the decision rules generated by the general dynamic discrete choice model. The required conditions impose restrictions on the return function $R(s)$, which reports the lifetime rewards associated with choice s.

A.1 Necessary and Sufficient Conditions for the Ordered Choice Model of Schooling to represent the Stochastic Dynamic Schooling Choice Model.

The ordered choice model assumes that once agents decide to drop out of school, it is never optimal for them to return. This is not necessarily the case in a general stochastic dynamic schooling choice model, where agents may drop out of school and return at a later date. As a result, for the ordered choice model to represent the general model of schooling, it is necessary to impose the condition that if the agent finds it optimal not to enroll in school at date t^*, so that $\delta_{t^*} = 0$, then he must also find it optimal not to return, so that $\delta_\tau = 0$ for any $\tau > t^*$. We call this condition the "no-return" condition.

Condition 1. *Let t^* be the first period that the agent decides not to enroll in school, i.e., $\delta_t = 1$ for $t < t^*$ and $\delta_{t^*} = 0$. The agent does not return if $\delta_\tau = 0$ for all $\tau = t^* + 1, \ldots, T$.*

The no-return condition states that the schooling level t^* is an absorbing state.

Proposition 2. *If the no-return condition holds and period s is such that $\delta_\tau = 1$ for $\tau = 1, 2, \ldots, s-1$ and $\delta_s = 0$, it follows that:*

$$V_s(s, \mathcal{I}_s) = r_s(s, 0) + E\left[\sum_{\tau=s+1}^{T} r_\tau(s, 0) \middle| \mathcal{I}_s\right].$$

Proof. By backward induction. Assume that period s is such that $\delta_\tau = 1$ for $\tau < s$ and $\delta_s = 0$. Under the no-return condition, $\delta_{s+k} = 0$ for $k = 1, \ldots, T-s$. In particular, $\delta_T = 0$. Then,

$$V_T(s, \mathcal{I}_T) = \max\{r_T(s, 0), r_T(s, 1)\} = r_T(s, 0). \tag{AP.3}$$

Note that:

$$V_{T-1}(s, \mathcal{I}_{T-1}) = \max\{r_{T-1}(s, 0) + E[V_T(s, \mathcal{I}_T)|\mathcal{I}_{T-1}], r_T(s, 1) + E[V_T(s+1, \mathcal{I}_T)|\mathcal{I}_{T-1}]\},$$

but because of the no-return condition, $\delta_{T-1} = 0$, and we know that:

$$V_{T-1}(s, \mathcal{I}_{T-1}) = r_{T-1}(s, 0) + E[V_T(s, \mathcal{I}_T)|\mathcal{I}_{T-1}]. \tag{AP.4}$$

Substitute (AP.3) into (AP.4) to obtain

$$V_{T-1}(s, \mathcal{I}_{T-1}) = r_{T-1}(s, 0) + E[r_T(s, 0)|\mathcal{I}_{T-1}].$$

Assume now that

$$V_{s+1}(s, \mathcal{I}_{s+1}) = r_{s+1}(s, 0) + E\left[\sum_{\tau=s+2}^{T} r_\tau(s, 0) \middle| \mathcal{I}_{s+1}\right].$$

Then,

$$V_s(s, \mathcal{I}_s) = \max\{r_s(s, 0) + E[V_{s+1}(s, \mathcal{I}_{s+1})|\mathcal{I}_s], r_s(s, 1) + E[V_{s+1}(s+1, \mathcal{I}_{s+1})|\mathcal{I}_s]\}.$$

But because $\delta_s = 0$,

$$V_s\left(s, \mathcal{I}_{s+1}\right) = r_s\left(s, 0\right) + E\left[V_{s+1}\left(s, \mathcal{I}_{s+1}\right) \middle| \mathcal{I}_s\right]$$

and consequently:

$$V_s\left(s, \mathcal{I}_s\right) = r_s\left(s, 0\right) + E\left[\sum_{\tau=s+1}^{T} r_\tau\left(s, 0\right) \middle| \mathcal{I}_s\right].$$

□

Proposition 3. *If the no-return condition holds and period s is such that $\delta_\tau = 1$ for $\tau = 1, 2, \ldots, s-1$ and $\delta_s = 0$, it follows that:*

$$V_1\left(1, \mathcal{I}_1\right) = E\left[\sum_{\tau=1}^{s-1} r_\tau\left(\tau, 1\right) + \sum_{\tau=s}^{T} r_\tau\left(s, 0\right) \middle| \mathcal{I}_1\right]$$

Proof. Assume that period s is such that $\delta_\tau = 1$ for $\tau < s$ and $\delta_s = 0$. Then repeat the steps in the previous proof. □

Corollary 4. *If the no-return condition holds and period s is such that $\delta_\tau = 1$ for $\tau = 1, 2, \ldots, s-1$ and $\delta_s = 0$ then s is the optimal schooling level implied by the stochastic dynamic schooling choice model.*

Definition 5. *The ordered choice model of schooling decisions represents the general stochastic dynamic schooling choice model if the decision rules generated by the former coincide with the ones generated by the latter.*

Proposition 6. *If the ordered choice model of schooling decisions represents the stochastic dynamic schooling choice model, then the no-return condition holds.*

Proof. Suppose that the no-return condition does not hold. Let s denote the optimal schooling level generated by the ordered choice model. Now, take an agent whose behavior is governed by the dynamic schooling choice model for which $\delta_t = 1$ for $t = 1, \ldots, s-1$ and $\delta_s = 0$. If the no-return condition does not hold, there exists a period t^*, $t^* > s$, such that for some information set $\mathcal{I}_{t^*}$ we have:

$$r_{t^*}\left(s, 1\right) + E\left[V_{t^*+1}\left(s+1, \mathcal{I}_{t^*+1}\right) \middle| \mathcal{I}_t\right] \geq r_{t^*}\left(s, 0\right) + E\left[V_{t^*+1}\left(s, \mathcal{I}_{t^*+1}\right) \middle| \mathcal{I}_{t^*}\right],$$

and consequently s is not the optimal schooling level predicted by the stochastic dynamic schooling choice model. □

The proof exploits the fact that the ordered choice model assumes an irreversible decision, while that is not so in the stochastic dynamic schooling choice model in which the no-return condition does not hold.

The no-return condition alone is not sufficient to guarantee that the ordered choice model of schooling decisions represents the stochastic dynamic schooling choice model. The decision rule of the ordered choice model is based on the return function $R(s)$. The decision rule in the stochastic dynamic schooling choice model depends on the value functions $\{V_t(s_t, \mathcal{I}_t)\}_{t=1}^{T}$ and the instantaneous reward function $\{r_t(s_t, \delta_t)\}_{t=1}^{T}$.

Condition 7. *Suppose that the no-return condition holds. Assume that $R(s)$, defined in the text, satisfies:*

$$R(s) = E\left[\sum_{\tau=1}^{s-1} r_\tau(\tau, 1) + \sum_{\tau=s}^{T} r_\tau(s, 0) \middle| \mathcal{I}_1\right] = V_1(1, \mathcal{I}_1) \tag{AP.5}$$

This condition states that if s is the optimal schooling implied by the stochastic dynamic schooling choice model, then $R(s)$, the return function for the ordered choice model, is also the lifetime return implied by the stochastic dynamic schooling choice model.

Proposition 8. *If the no-return condition is satisfied and $R(s)$ is as defined in (AP.5), then the ordered choice model decision structure represents the stochastic dynamic schooling choice model.*

Proof. If the no-return condition holds, the stochastic dynamic schooling choice model has the decision structure such that once the agent decides to drop out of school, he does not return. Consequently, for any optimal s, $R(s) > R(s')$ for any $s' = 1, \ldots, \bar{S}$, $s' \neq s$. In particular, note that $R(s) \geq R(s-1)$ and $R(s) \geq R(s+1)$ from which it follows that $f(s+1) \leq 0 \leq f(s)$, which is the decision rule implied by the ordered choice model. □

As noted in the text, the ordered choice model also requires separability between the observables and the unobservables.

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 $\mathbb{R}$ into $[0, 1]$. Matzkin (1992, 1993, 1994) 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) additively separable functions, e.g. $\varphi(Z) = Z_1 + \tau(Z_2, \ldots, Z_K)$.

References

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

Bresnahan, T. F. (1987): "Competition and Collusion in the American Automobile Industry: The 1955 Price War," *Journal of Industrial Economics*, 35(4), 457–482.

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

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

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

Fuss, M. A., and D. McFadden (1978): *Production Economics: A Dual Approach to Theory and Applications*. North-Holland Publishing Company, New York, v. 1. The theory of production. v. 2. Applications of the theory of production.

Heckman, J. J. (1974): "Effects of Child-Care Programs on Women's Work Effort," *Journal of Political Economy*, 82(2), S136–S163, Reprinted in T.W.Schultz (ed.) *Economics of the Family: Marriage, Children and Human Capital*, University of Chicago Press, 1974.

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

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The Economics and Economet-

rics of Active Labor Market Programs," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3A, chap. 31, pp. 1865–2097. North-Holland, New York.

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

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

Heckman, J. J., S. Urzua, and E. J. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88(3), 389–432.

Heckman, J. J., and E. J. Vytlacil (2007): "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics, Volume 6*, ed. by J. Heckman, and E. Leamer. Elsevier, Amsterdam, Forthcoming.

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

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

Manski, C. F. (1988): "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83(403), 729–738.

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

——— (1993): "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58(1-2), 137–168.

——— (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, vol. 4, pp. 2523–58. North-Holland, New York.

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

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

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

Prescott, E. C., and M. Visscher (1977): "Sequential Location among Firms with Foresight," *Bell Journal of Economics*, 8(2), 378–893.

Ridder, G. (1990): "The Non-parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies*, 57(2), 167–181.

Rosen, S. (1977): "Human Capital: A Survey of Empirical Research," in *Research in Labor Economics*, ed. by R. Ehrenberg, vol. 1, pp. 3–40. JAI Press, Greenwich, CT.

Shaked, A., and J. Sutton (1982): "Relaxing Price Competition through Product Differentiation," *Review of Economic Studies*, 49(1), 3–13.

Theil, H. (1975): *Theory and Measurement of Consumer Demand*, vol. 1. American Elsevier Publishing Company, New York.

——— (1976): *Theory and Measurement of Consumer Demand*, vol. 2. American Elsevier Publishing Company, New York.

Vytlacil, E. J. (2006): "Ordered Discrete-Choice Selection Models and Local Average Treatment Effect Assumptions: Equivalence, Nonequivalence, and Representation Results," *Review of Economics and Statistics*, 88(3), 578–581.