cemmap
centre for microdata methods and practice

Identification of structural dynamic discrete choice models

Le-Yu Chen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP08/09

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Identification of Structural Dynamic Discrete Choice Models

Le-Yu Chen[1 2]
Institute of Economics, Academia Sinica

Revised : April 2009

[**Abstract**]

This paper presents new identification results for the class of structural dynamic discrete choice models that are built upon the framework of the structural discrete Markov decision processes proposed by Rust (1994). We demonstrate how to semiparametrically identify the deep structural parameters of interest in the case where utility function of one choice in the model is parametric but the distribution of unobserved heterogeneities is nonparametric. The proposed identification method does not rely on the availability of terminal period data and hence can be applied to infinite horizon structural dynamic models. For identification we assume availability of a continuous observed state variable that satisfies certain exclusion restrictions. If such excluded variable is accessible, we show that the structural dynamic discrete choice model is semiparametrically identified using the control function approach.

KEYWORDS : structural dynamic discrete choice models, semiparametric identification, control function

[1] I am grateful to my supervisor, Dr. Sokbae Lee for his helpful advice on this research work. I also thank Prof. Oliver Linton and Prof. Pedro Mira for their various comments on this work.

[2] E-mail : lychen@econ.sinica.edu.tw

1 Introduction

Over the decades, structural estimation of dynamic discrete choice models has been more and more employed in empirical economics and for the assessment of various policies of interest. Well-known empirical applications include Wolpin (1984) and Holz and Miller (1993) in fertility; Pakes (1986) in patent renewal; Rust (1987), Das (1992), Cooper, Haltiwanger, and Power (1999), and Adda and Cooper (2000) in capital retirement and replacement; Wolpin (1987) in job search; Berkovec and Stern (1991), Daula and Moffitt (1995), Rust and Phelan (1997), and Karlstrom, Palm and Svensson in (2004) in retirement from labor force; Erdem and Keane (1996) in brand choice; Keane and Wolpin (1997) in career choice; Eckstein and Wolpin (1999) in education choice. See Aguirregabiria and Mira (2007) for a survey of recent work on structural dynamic discrete choice models[3] and Wolpin (1996) for the use of such models in public policy evaluation.

Structural estimation of dynamic discrete choice models is attractive to researchers because it tightly links economic theories of rational decision making under uncertainty in a dynamic setting to the interpretation and prediction of the stochastic process that generates the economic data of interest. The parameters in the model and the estimation methods are structural in the sense that they are derived from solutions of an explicitly postulated economic behavioral model. Hence structural estimation avoids Lucas critique about the use of reduced form analysis and it can allow researchers to simulate the consequences of vaious policy experiments after the underlying structural parameters are estimated.

However, a common feature in the structural estimation literature is the parametric specification of the underlying structural objects such as utility functions, transition probabilities of state variables and distributions of the unobservables. Estimated results from parametric models may be sensitive to changes of specification and hence suffer from problems of misspecification. Therefore, later work in the literature tries to study the nonparametric identification of the dynamic discrete choice model. An influential work by Rust (1994) shows that the dynamic discrete choice model is nonparametrically unidentified. Magnac and Thesmar (2002) extend Rust's framework and show that if the history of observed variables is discrete, the dynamic discrete choice model is nonidentified. They then further characterize the degree of nonidentification and show that the model is identified subject to ad hoc assumptions on distributions of unobservables and functional forms of agents' preferences. In particular, their identification result indicates that parametric specifications on the distributions of unobservables are indispensable for identifying the deep structural parameters of interest, which consequently motivates the maximum likelihood estimation

[3] See Eckstein and Wolpin (1989) for a survey of earlier empirical work on structural dynamic discrete choice models.

approach to most of the empirical work on structural estimation of the dynamic discrete choice model.

This paper builds on the framework of Rust (1994) and Magnac and Thesmar (2002) and develops new identification results for the case where the distribution of unobservables is nonparametric. As in static discrete choice models, the structural dynamic discrete choice model is not nonparametrically identified when all regressors (observed state variables) are discrete and the distribution of unobservables is unknown and continuous. To allow for nonparametric distribution of the unobservables, we assume the availability of continuous observables so that we can still gain identification by exploiting continuous variation from observed continuous state variables. However, as noted in Magnac and Thesmar (2002), the dynamic discrete choice model is still unidentified even if the distribution of unobservables is given a priori. To secure identification, the researcher needs to further assume the functional form of the per-period utility for one of the choices. Therefore, throughout this paper the identification analysis is semiparametric in the sense that one of the per-period utility functions of the choices is given a priori. Under this assumption, we show that the structural dynamic discrete choice model is semiparametrically identified if there is an exclusion restriction that provides identification power when information about distribution of unobservables is not available. Therefore, in this respect, our analysis provides semiparametric identification for the dynamic counterpart of the already well developed semiparametric static discrete choice models in which preference shock distributions are nonparametric but systematic utility functions are restricted to certain function space[4].

In the identification analysis of structural dynamic discrete choice models with nonparametric unobservables, our work is related to the analysis of Heckman and Navarro (2007). In their paper, they consider semiparametric identification of a variety of reduced form optimal stopping time models and a finite horizon structural optimal stopping model[5]. Their structural model allows for richer time series dependence between nonparametric unobservables than is assumed in Rust's (1994) framework. However, they assume future values of some observed continuous state variables are in the agent's current information set[6] so that they can achieve identification by varying these variables and using identification-at-infinity strategy in a fashion similar to the one entertained by Taber (2000). Their identification-at-infinity strategy requires large support

[4]See Manski (1975, 1985), Cosslett (1983), Stoker (1986), Klein and Spady (1988), and Ichimura (1993) for studies of semiparametric identification and estimation of static discrete choice models in which the random shock distribution is nonparametric but systematic utility is parametric. Matzkin (1992, 1993) and Lewbel and Linton (2007) consider nonparametric discrete choice models by further relaxing parametric assumptions on the systematic utility to a function space defined under certain shape restrictions such as linear homogeneity that are usually motivated from economic theories.

[5]An optimal stopping time model is a dynamic discrete choice model in which one of the choices is irreversible.

[6]In their schooling decision example, the agent knows all current and future schooling cost shifters and these variables are in the agent's current information set.

of the utility functions that may also be demanding for practical applications[7]. Furthermore, such large support assumption is not applicable to infinite horizon dynamic models where the solution of value functions is due to the application of Blackwell sufficient conditions that require (sup-norm) bounded per-period utility functions. In this paper, we consider identification of the general structural dynamic discrete choice model. We establish identification of structural models in which all choices can be recurrent. While the model admits an irreversible choice, our identification assumption can be further relaxed by exploiting such extra behavioral restriction.

Regarding the dynamics of the model, we assume that future values of state variables may be uncertain to the agent so that the agent's current information set contains only current and past realized state variables. We follow Rust (1994)'s conditional independence framework to model the agent's belief about evolution of the state variables. We develop identification strategy that is applicable to both finite and infinite horizon structural models. Note that there is no terminal decision period for the class of infinite horizon models. Admitting the terminal period yields nonstationarity in the model so that one can gain extra identification power by discriminating the model structure that may vary across periods. For instance, Taber (2000), Heckman and Navarro (2007), and Aguirregabiria (2008) exploit the modeling assumption that the agent's choice problem is static in the terminal period and therefore one can identify using data from terminal period the distribution of unobserved state variables and other structural objects by applying standard arguments of identification of static discrete choice models and then solve the previous period problems using backward induction. Lacking the terminal period data poses a challenge in identification of structural dynamic models. In this paper, we discuss the use of exclusion restrictions to circumvent identification problems in these cases within Rust (1994)'s structural Markov decision process framework. There is little literature that addresses identification of Rust (1994)'s structural models with nonparametric unobservables[8]. This paper is the first work that does not rely on the use of terminal period data but still provides positive result for semiparametrically identifying all deep structural parameters of dynamic discrete choice models within Rust (1994)'s framework with nonparametric unobserved state variables.

The rest of the paper is organized as follows. Section 2 presents the framework and assumptions of the structural dynamic discrete choice model. Section 3 studies the semiparametric identification of the basic model. Section 4 discusses identification of the optimal stopping time model which is a variant of the basic model that allows for the presence of an irreversible choice. Section 5 discusses an illustrating example to which the identification analysis of this

[7]In the empirical analysis of their paper (Heckman and Navarro 2007), they report that they do not have the required limit sets in their data.

[8]Aguirregabiria (2005, 2008) also studies identification of Rust (1994)'s dynamic discrete choice model with nonparametric unobserved state variables. However, he is only interested in identifying counterfactual choice probability when the counterfactual policy experiments take some specific formats.

paper can be applied. Section 6 proposes a estimation strategy based on the identification analysis. Section 7 concludes the paper.

2 The framework and assumptions

Time is discrete and indexed by $t \in \{1, 2, ...\}$. Consider an economic agent with intertemporal utility that is additive separable over time. At each period, the agent makes a decision over two mutually exclusive choices[9], denoted as Choice 0 and Choice 1. Denote as s_t the state variables at time t that the agent considers in this structural dynamic discrete choice model. From the econometrician's point of view, some components of s_t are observables that are denoted by x_t. Others are unobserved random shocks $\varepsilon_t \equiv (\varepsilon_{0,t}, \varepsilon_{1,t})$ for each choice. Therefore, $s_t = (x_t, \varepsilon_t)$. At the beginning of each period t, s_t is revealed to the agent who then chooses $d_t \in \{0, 1\}$ and receives the instantaneous return $u_t(d_t, s_t)$. However, next period state variables s_{t+1} are still uncertain to the agent. Following Rust (1994)'s dynamic Markov discrete decision process framework, we assume the transition of state variables follows controlled first-order Markov property. In other words, the next period state variables s_{t+1} are drawn from the Markov transition probability density $f_s(\ s_{t+1}|s_t, d_t)$, which represents the law of motion of the state variables in the model. The agent has belief, $\mu(\ s_{t+1}|s_t, d_t)$ about the evolution of the state variables. We assume that the agent's belief is rational in the sense that it coincides with the true transition probability $f_s(\ s_{t+1}|s_t, d_t)$[10]. To proceed, as in Rust (1994) and Magnac and Thesmar (2002, pp. 802-804), we make the following assumptions on the agent's preference and laws of motion of the state variables.

[M1] (Additive Separability) :
for $k \in \{0, 1\}$ *and for all* t*,*

$$u_t(k, s_t) \equiv u_{k,t}(s_t) = u^*_{k,t}(x_t) + \varepsilon_{k,t}.$$

[M2] (Conditional Independence) :
for all t*,*

$$f_s(x_{t+1}, \varepsilon_{t+1}|x_t, \varepsilon_t, d_t) = f_\varepsilon(\varepsilon_{t+1}) f_x(x_{t+1}|x_t, d_t). \tag{1}$$

The additive separability assumption allows us to decompose the utility into a systematic part that depends only on observable state variables and a preference shock that is unobserved to the econometrician. This is a standard

[9] To present the main idea, we focus on binary choice models. The analysis in this chapter can be generalized to the multinomial choice context, which is left for further research work.

[10] In general, the agent's belief $\mu(s_{t+1}|s_t, d_t)$ is subject to his degree of rationality and hence may not necessarily represent the true law of motion of the states. However, identification of a general belief requires more structural assumptions about the agent's belief formation process.

assumption in static discrete choice models[11]. [M2] assumes that the unobservables ε_t are i.i.d. exogenous random shocks to the model and future values of the observed states x_{t+1} can depend on the current control variable d_t and values of current observed states x_t but they do not direct depend on current values of the exogenous shocks ε_t. Since [M2] restricts the time series dependence of unobserved state variables, it precludes unobserved persistent heterogeneity and dynamic selection bias discussed in Cameron and Heckman (1998), Taber (2000), and Heckman and Navarro (2007)[12]. However, general time series dependence between observables are allowed. Besides, preference shocks can also be dependent across alternatives. [M2] allows rational expectation belief to be derived from a learning process based on observations in the sense that the agent can infer joint distribution of next period states based on observable information from other agents and on the assumption that he knows distribution of his private signal ε_t (Manski 1993, Magnac and Thesmar 2002).

Let $\beta \in [0, 1)$ be the discount factor. Assuming at each period choices are made to maximize the agent's expected life time utility, under [M1], [M2] and rational expectation belief, by Bellman principle of optimality, π_t,the policy function (the optimal decision rule) can be characterized as the follows.

$$\pi_t(s_t) = \underset{k\in\{0,1\}}{\arg\max}\{v_{k,t}(s_t)\}.^{13}$$

and the value function at each period t can be obtained via the following recursive expressions.

$$v_t(s_t) = \pi_t(s_t)v_{1,t}(s_t) + (1 - \pi_t(s_t))v_{0,t}(s_t),$$

where $v_{k,t}(s_t)$ are choice-specific value functions represented by the following Bellman equations

$$v_{k,t}(s_t) \equiv u_{k,t}(s_t) + \beta E(v_{t+1}(s_{t+1}, k)|s_t, d_t = k) \text{ for } k \in \{0, 1\}.$$

Under rationality the observed time series of individual choice is the sequence of optimal decisions $\{\delta_t = \pi_t(s_t)\}$ that satisfy

$$\delta_t = 1\{v_{1,t}(s_t) \geq v_{0,t}(s_t)\}. \tag{2}$$

It is clear that under [M1] and [M2], for $k \in \{0, 1\}$ and for all t, there is $v^*_{k,t}(x_t)$, which is a function of x_t only, such that the choice specific value functions, $v_{k,t}(s_t)$ can be written as :

$$v_{k,t}(s_t) = v^*_{k,t}(x_t) + \varepsilon_{k,t},$$

[11] See Vytlacil (2002, 2005) for discussions about the use of additively separable utility functions in static discrete choice models.

[12] See Rust (1994) for a detailed and graphical explanation of controlled Markov process under such conditional independence assumption.

[13] When distribution of ε_t is continous (see assumption [M4]), ties between the two choice specific value functions do not occur almost surely so that there is a (almost surely) unique solution to this optimization problem.

where $v^*_{k,t}(x_t)$ are choice-specific systematic value functions that follow Bellman equations

$$v^*_{k,t}(x_t) = u^*_{k,t}(x_t) + \beta E(v_{t+1}(s_{t+1}, k)|x_t, d_t = k).$$

In this paper, we study stationary infinite horizon models. In other words, the time horizon of the decision problem is infinite and the per-period utilities and Markov transition probabilities are time-invariant. The stationary Markovian structure of the dynamic model implies that the decision problem faced by the agent is the same whether the agent is in state s_t at period t or in state s_{t+j} at period $t+j$ provided that $s_t = s_{t+j}$ (Rust 1994). Therefore, we can suppress the time index and for the rest of this paper we denote variables y and y' as the current and next period objects, respectively. Under this notation, we can formulate the model as follows. For $k \in \{0, 1\}$,

$$\begin{aligned}
v(s) &= \pi(s)v_1(s) + (1 - \pi(s))v_0(s) && (3)\\
u_k(s) &= u^*_k(x) + \varepsilon_k && (4)\\
v_k(s) &= v^*_k(x) + \varepsilon_k && (5)\\
v^*_k(x) &= u^*_k(x) + \beta E(v(s')|x, d = k) && (6)\\
\pi(s) &= 1\{v_1(s) \geq v_0(s)\} && (7)
\end{aligned}$$

Let $\Delta v^*(x) \equiv v^*_1(x) - v^*_0(x)$ and $\Delta\varepsilon \equiv \varepsilon_1 - \varepsilon_0$. Using the following lemma, we can further rewrite (6) to get a more convenient representation of the choice specific value functions.

Lemma 1

Under [M1] and [M2], the systematic Bellman equations (6) can be written as the follows. For $k \in \{0, 1\}$,

$$v^*_k(x) = u^*_k(x) + \beta E(v^*_0(x') + \varepsilon'_0|x, d = k) + \beta[E(\Delta v^*(x')\delta'|x, d = k) + E(\Delta\varepsilon'\delta'|x, d = k)] \tag{8}$$

Proof. Lemma 1 follows immediately by putting equations (3) and (7) into (6) and noting that $\delta' = \pi(s')$. ■

Equation (8) expresses the choicewise systematic value function as the sum of three terms. The first term on the right hand side of (8) is the instantaneous utility the agent receives when he makes choice k. The second term is the discounted expected future value when at next period the agent makes choice 0 given that his current action is choice k. The third term is the discounted expected gain when it is optimal for the agent to deviate from choice 0 at next period given that his current action is choice k. This discounted expected gain consists of two components: $\beta E(\Delta v^*(x')\delta'|x, d = k)$ represents the discounted expected gain from the systematic component and $\beta E(\Delta\varepsilon'\delta'|x, d = k)$ represents that from the random unobserved preference shock.

The reduced form equation of this model is the conditional choice probability,

$$P(\delta = 1|x) = P(\Delta v^*(x) + \Delta\varepsilon \geq 0|x). \tag{9}$$

It is clear that the distribution of $\Delta\varepsilon$ is more critical than the joint distribution of $\varepsilon = (\varepsilon_i)_{i\in\{0,1\}}$ in characterizing the reduced form equation. Furthermore, if $E(\varepsilon_0)$ is normalized to be zero, then the systematic Bellman equations (8), as we will show later, will be completely characterized by the distribution of $\Delta\varepsilon$.

[M3] : $E(\varepsilon_0) = 0$.

[M3] provides location normalization for the random shock of one of the choices[14][15]. By [M2], we have $E(\varepsilon_0'|x, d = 0) = E(\varepsilon_0')$ so that under [M3] we can further simplify the systematic Bellman equations as follows[16]. For $k \in \{0, 1\}$,

$$v_k^*(x) = u_k^*(x) + \beta E(v_0^*(x')|x, d = k) + \beta[E(\Delta v^*(x')\delta'|x, d = k) + E(\Delta\varepsilon'\delta'|x, d = k)]. \tag{11}$$

To proceed, we need further assumptions on the distribution of $\Delta\varepsilon$.

[M4] : $-\Delta\varepsilon$ has strictly increasing and absolutely continuous (with respect to Lebesgue measure) distribution function G that induces a density function g, whose support is Γ_G.

Restricting G to be strictly increasing ensures that G is invertible, which is a key condition for identification of the model (Hotz and Miller 1993). The absolute continuity property of G ensures that the dynamic programming model has a unique optimal solution almost surely.

2.1 Parameters of interest

Let B be the set of all measurable, real-valued and bounded functions under sup norm. Note that B is a Banach space. We assume that for $k \in \{0, 1\}$, $u_k^*(x) \in B$. Then using Blackwell sufficient conditions (see Theorem 3.3 in Stokey and Lucas 1989), it can be shown that the value functions $v_k^*(x)$, $k \in \{0, 1\}$ are also in B and are the unique fixed point of the system of Bellman equations (11).

[14] As in the static discrete choice models, the locations of the distributions for ε_1 and ε_0 are not separately identified since only the difference $\Delta\varepsilon$ matters in the conditional choice probability.

[15] Magnac and Thesmar (2002, pp803) assume both $E(\varepsilon_1)$ and $E(\varepsilon_0)$ are zero. However, their location normalization assumption is more than required.

[16] If the assumption [M3] is not maintained, then (11) becomes

$$v_k^*(x) = u_k^*(x) + \mu_0 + \beta E(v_0^*(x')|x, d = k) + \beta[E(\Delta v^*(x')\delta'|x, d = k) + E(\Delta\varepsilon'\delta'|x, d = k)], \tag{10}$$

where $\mu_0 \equiv \beta E(\varepsilon_0')$ is the location parameter. In counterfactual policy analysis, μ_0 is needed to compute the counterfactual value function via Bellman equation (10). Note that the location parameter μ_0 is not identified since the choice behavior is driven by difference of value functions in which μ_0 is always differenced out.

The parameters of interest in this model are $(u_k^*(x), v_k^*(x), \beta, G, f_x(x'|x,k))$ for $k \in \{0,1\}$. We refer to $v_k^*(x)$ as the *derived* structural parameters as they are derived from the *primitive* structural parameters $(u_k^*(x), \beta, G, f_x(x'|x,k))$ via the Bellman equations (11). Obtaining these structural parameters allows researchers to answer a variety of policy questions[17]. For example, in doing welfare analysis, we may need $\partial E(v(x,\varepsilon)|x)/\partial x_k$, a measure to assess the impact of changing a particular state variable x_k on the average social surplus function. Note that

$$\frac{\partial E(v(x,\varepsilon)|x)}{\partial x_k} = \frac{\partial v_1^*(x)}{\partial x_k} P(\delta = 1|x) + \frac{\partial v_0^*(x)}{\partial x_k} P(\delta = 0|x)^{18}. \tag{12}$$

Therefore, the structural parameters $v_k^*(x)$ allow us to do such comparative statics analysis. Furthermore, in some contexts, the researcher may want to study a counterfactual policy experiment τ such that under policy τ, the agent's behavior is generated from the new structure characterized by $(u_k^{*\tau}(x), v_k^{*\tau}(x), \beta^\tau, G^\tau, f_x^\tau(x'|x,k))$[19]. Assuming the counterfactual structural parameters have a known mapping to the set of $(u_k^*(x), \beta, G, f_x(x'|x,k))$, then we can simulate the agent's behavior under the counterfactual policy if the set of structural parameters can be identified.

3 Semiparametric identification of the structural model

We have a sample of individuals indexed by i who follow the constituted structural dynamic discrete choice model. Data consist of the observed state variables and optimal choices for all individuals and for two consecutive periods of the decision horizon[20]. Assume random sampling, we can suppress the individual index i.

Under conditional independence assumption [M2], Magnac and Thesmar (2002, pp. 803-804) observe that the agent's belief $f_x(x'|x, d = k)$ for $k \in \{0,1\}$ can be identified as $f_x(x'|x, \delta = k)$ using the data (δ, x, x'). Thus we can replace the control variable d with the observed choice δ in Bellman equations (11). In other words, We can then rewrite these Bellman equations as

[17] Of course, we assume that the policy does not change the nature of the decision problem such that the agent still faces the same dynamic discrete choice problem with the same choice set.

[18] Since the model implies that $E(v(x,\varepsilon)|x) = \int \max(v_1^*(x) + \varepsilon_1, v_0^*(x) + \varepsilon_0) dF(\varepsilon)$, (12) follows by interchanging the integral and differentiation, which can be justified under Lebesgue Dominated Convergence Theorem when both $\frac{\partial v_1^*(x)}{\partial x_k}$ and $\frac{\partial v_0^*(x)}{\partial x_k}$ are bounded.

[19] All variables superscripted with τ denote the same variables under the counterfactual policy τ.

[20] Since the model is assumed stationary, short panel data that consists of only two consecutive periods are sufficient for identification (subject to other identification conditions to be discussed). The structural parameters will thus be over-identified if panel data of more than two periods are available. In this case, more efficient estimators of the structural parameters can be constructed by combining those estimators obtained using data from different consecutive periods.

$$v_k^*(x) = u_k^*(x)+\beta E(v_0^*(x')|x,\delta=k)+\beta[E(\Delta v^*(x')\delta'|x,\delta=k)+E(\Delta\varepsilon'\delta'|x,\delta=k)] \tag{13}$$

Therefore, for the rest of the paper, we will direct work on the Bellman equations (13).

3.1 Identification when the distribution G is known

We first characterize the degree of identification for the dynamic structural discrete choice model. The following lemma is an extension to stationary infinite horizon models based on Magnac and Thesmar (2002)'s identification result (Proposition 2 (i) pp. 806) that shows identification of finite horizon dynamic discrete choice models when the distribution G is known.

Lemma 2

Under [M2], the transition probability $f_x(x'|x,k)$ is identified and given β, G and $u_0^*(x)$, the remaining structural parameters $(u_1^*(x), v_1^*(x), v_0^*(x))$ are identified.

Proof. Given G, under [M2] and [M4], using (9) we can identify $\Delta v^*(x')$ as

$$\Delta v^*(x') = G^{-1}(P(\delta'=1|x')).$$

Furthermore, $E(\Delta\varepsilon'\delta'|x,\delta=k)$ in (13) can be written as

$$E(\Delta\varepsilon'\delta'|x,\delta=k) = E(\Delta\varepsilon'|\delta'=1,x,\delta=k)P(\delta'=1|x,\delta=k), \tag{14}$$

where the term $E(\Delta\varepsilon'|\delta'=1,x,\delta=k)$ in (14) satisfies

$$\begin{aligned} E(\Delta\varepsilon'|\delta' &= 1,x,\delta=k) = E(E(\Delta\varepsilon'|\delta'=1,x',x,\delta=k)|\delta'=1,x,\delta=k) \\ &= E(E(\Delta\varepsilon'|\Delta v^*(x')+\Delta\varepsilon'\geq 0,x',x,\delta=k)|\delta'=1,x,\delta=k). \end{aligned}$$

Under [M2], for any point τ in the support of x',

$$\begin{aligned} E(\Delta\varepsilon'|\Delta v^*(x')+\Delta\varepsilon' &\geq 0,x'=\tau,x,\delta=k) = E(\Delta\varepsilon'|\Delta v^*(\tau)+\Delta\varepsilon'\geq 0) \\ &= E(\Delta\varepsilon'|G^{-1}(P(\delta'=1|x'=\tau))+\Delta\varepsilon'\geq 0). \end{aligned}$$

is identified since G is given a priori. Thus for $k\in\{0,1\}$, the term $E(\Delta\varepsilon'\delta'|x,\delta=k)$ is identified. Therefore, we can identify the expected gain

$$eg_k(x) \equiv E(\Delta v^*(x')\delta'|x,\delta=k)+E(\Delta\varepsilon'\delta'|x,\delta=k)$$

for $k\in\{0,1\}$ in Bellman equation (13). We can then write (13) as

$$v_k^*(x) = u_k^*(x)+eg_k(x)+\beta E(v_0^*(x')|x,\delta=k). \tag{15}$$

Given $\beta, u_0^*(x)$ and the identification of $eg_0(x)$, $v_0^*(x)$ can then be identified as the unique fixed point to the (reduced form) Bellman equation of Choice 0 using (15). Thus the identification of $v_1^*(x)$ follows by noting that

$$v_1^*(x) = G^{-1}(P(\delta = 1|x)) + v_0^*(x). \tag{16}$$

Using β and the identified v_1^*, v_0^* and eg_1, the remaining structural parameter, $u_1^*(x)$ is then identified as

$$u_1^*(x) = v_1^*(x) - eg_1(x) - \beta E(v_0^*(x')|x, \delta = 1).$$

■

We comment on the sufficient conditions for identification given in Lemma 2. First, the discount factor β is generally assumed rather than identified since one cannot distinguish between myopic ($\beta = 0$) and forward-looking ($\beta > 0$) agents if information on agents' utility primitives is not available. To see this, consider a myopic agent with $\beta = 0$. This agent chooses Choice 1 if $\widetilde{u}_1^*(x)+\varepsilon_1 > \widetilde{u}_0^*(x)+\varepsilon_0$, where $\widetilde{u}_k^*, k \in \{0, 1\}$ are his per-period systematic utility functions. If $\widetilde{u}_k^*$ (as functions) coincide with v_k^*, then this agent's choice behavior is exactly the same as that of a forward-looking agent with non-zero discount factor and per-period utilities u_k^*[21].

Second, as in static discrete choice model, utility primitives are not separately identified since it is the difference rather than the level that drives the choice behavior. To separately identify the remaining structural parameters, utility primitive of one reference alternative (Choice 0, for example) has to be given a priori[22].

We observe that both β and u_0^* are inevitable for nonparametric identification of the structural model. However, the assumption of the distribution G in Lemma 2 seems sufficient rather than necessary for identification. Note that the original analysis of Magnac and Thesmar (2002) is conducted based on the discrete-support assumption in which all observed state variables x take only finite values. In that context, a continuous distribution G cannot be identified with only discrete-valued regressors. There is scope of further relaxing the assumption on G if the structural model admits continuous observed state variables. Note that most empirical literature makes ad-hoc parametric assumptions on G. Though such practice facilitates the identification task in view of the results in Lemma 2, these assumptions of G are rarely justified a priori and hence suffers from misspecification problems. Thus it is important to develop

[21] This argument assumes that the researcher does not know the primitive utilities u_k^* and G. If the researcher is willing to assume parametric functional forms on (u_k^*, G), then β may be identified under certain parametric restrictions. However, in this paper we are also interested in nonparametric identification of (u_k^*, G) and thus we will assume β in line with the literature (Magnac and Thesmar 2002, Aguirregabiria 2008) for nonparametric identification of the remaining structural parameters.

[22] Note that the assumption on utility of a reference alternative is not needed if one is only interested in identifying counterfactual choice probability when counterfactual policy experiments affect the utility primitives in an additive way (Aguirregabiria 2005, 2008).

positive identification that allows for nonparametric G such that the resulting structural inference can be robust against mis-specification of distribution of unobservables. In the subsequent sections, we will discuss how to identify the complete structure of the model using β, $u_0^*(x)$ and observed continuous state variables. Since we relax the distribution assumption of G, extra information in terms of modeling assumptions should be supplied in lieu of G so that identification is still ensured. To do this, a natural approach parallel to semiparametric methods of the static discrete choice models is to further restrict the systematic utilities that can be motivated from economic theory while keeping nonparametric the distribution of unobservables about which economic theory is usually silent. This paper aims to develop such semiparametric identification results. In particular, we discuss the degree of identification in the case where only the utility $u_0^*(x)$ is assumed and then provide sufficient conditions that guarantee nonparametric identification of G, from which we can proceed to identify more structural parameters using the results in Lemma 2.

3.2 Identification when the distribution G is unknown

Assuming continuous state variables x are available, can we identify G and other structural parameters by exploiting the continuous variation of x ? In Lemma 2, given β, G and $u_0^*(x)$, we can first identify v_0^* and then solve v_1^* using (16). When information of G is not available, since the term $E(\Delta\varepsilon'\delta'|x, \delta = k)$ appears in (13) and depends on the distribution G, v_0^* is generally not identified. It is interesting to ask whether obtaining extra information of v_0^* helps to identify G and the remaining structural parameters when continuous state variables x are available. The answer is yet negative. In contrast with the identification result in Lemma 2, the following lemma shows that assuming one value function alone provides no identification power.

Lemma 3

Assume [M4], given only β, u_0^* and v_0^*, then G, u_1^* and v_1^* are not identified.

Proof. Let G, u_1^* and v_1^* be a set of admissible structure parameters that satisfy Bellman equation (13) and

$$P(\delta = 1|x) = G(v_1^*(x) - v_0^*(x)). \tag{17}$$

Consider another distribution $\widetilde{G} \neq G$. For example, one can take $\widetilde{G}(y) = [G(y)]^\alpha$ for some $\alpha > 1$. Then the value function $\widetilde{v_1^*}(x) \equiv v_0^*(x) + \widetilde{G}^{-1}(G(v_1^*(x) - v_0^*(x)))$ and $\widetilde{G}$ generate the same conditional choice probability as that generated by v_1^* and G. Using Bellman equation (13), $\widetilde{v_1^*}(x)$ and $\widetilde{G}$ then implicitly define a utility function $\widetilde{u_1^*}$. Therefore, the parameters (u_1^*, v_1^*, G) and $(\widetilde{u_1^*}, \widetilde{v_1^*}, \widetilde{G})$ are observationally equivalent and they are not identified. ■

It is clear that the result of Lemma 3 applies regardless of the continuity property of x. This result arises from the fact that the index function in (17) is the difference, $v_1^* - v_0^*$ which is still an unknown object even if v_0^* is given a priori. When G is nonparametric, non-identification then follows from lack of variation to distinguish between the unknown link function G and the unknown index function $\Delta v^* = v_1^* - v_0^*$.

By (13),

$$\begin{aligned} \Delta v^*(x) &= u_1^*(x) + m(x) - u_0^*(x) \\ &= \lambda(x) - u_0^*(x), \end{aligned}$$

where $\lambda(x) = u_1^*(x) + m(x)$ and $m(x) = m_1(x) + m_2(x) + m_3(x)$ defined as

$$m_1(x) = \beta[E(v_0^*(x')|x, \delta = 1) - E(v_0^*(x')|x, \delta = 0)], \tag{18}$$
$$m_2(x) = \beta[E(\Delta v^*(x')\delta'|x, \delta = 1) - E(\Delta v^*(x')\delta'|x, \delta = 0)], \tag{19}$$
$$m_3(x) = \beta[E(\Delta\varepsilon'\delta'|x, \delta = 1) - E(\Delta\varepsilon'\delta'|x, \delta = 0)]. \tag{20}$$

Note that $m(x)$ represents the difference between discounted expected future values when current action is to choose Choice 1 and when Choice 0 is chosen at current period. This difference is zero if future state variables that determine the continuation values are (conditionally) independent of the current action and in this case the model becomes static since the choice behavior is driven only by difference of instantaneous utilities. Nevertheless, in general, $m(x)$ is a non-zero function of x and is an unknown object even if both β and u_0^* are given. To overcome the identification problem, since u_0^* is always assumed throughout the analysis, effects from Δv^* and G on the conditional choice probability can be disentangled if we can fix $\lambda(x)$ but at the same time freely move u_0^*. In this case, we can trace out the distribution G and thus G can be identified. As shown in Lemma 3, there is no such variation-free condition for identification when $\lambda(x)$ and $u_0^*(x)$ share completely the same set of regressors. Therefore, to achieve semiparametric identification when G is nonparametric, we propose to impose an exclusion restriction to provide such source of variation. In other words, if there is one continuous variable that is in the arguments of u_0^* but is excluded from those of $\lambda(x)$, then we can separately identify $\lambda(x)$ and G up to a location normalization. Theorem 4 demonstrates this semiparametric identification strategy.

Denote the support of x as $\Gamma_X = \Gamma_W \times \Gamma_Z$, where Γ_W and Γ_Z are the supports of w and z, respectively. Let $u_0^*(\Gamma_X)$ and $\Delta v^*(\Gamma_X)$ be the supports of u_0^* and Δv^*, respectively. We assume that $\Delta v^*(\Gamma_X) \subset u_0^*(\Gamma_X)$.

Theorem 4

Let $x = (w, z)$ in which the subvectors w and z have no common component and both of them are non-empty. Let $u_0^*(x) = u_0^*(w, z) \in B$ be a known function. Assume also the following : (i) (excluded regressors): $u_1^*(x) = u_1^*(w)$ (ii) (conditional independence between observables): for $k \in \{0, 1\}$, $(w', z') \perp z|w, \delta = k$

(iii) (continuous regressors): $\exists$ at least one continuous variable z_s in z such that the distribution of z_s conditional on w is non-degenerate and $u_0(x)$ is differentiable with respect to z_s with $\frac{\partial u_0^*(x)}{\partial z_s} \neq 0$ almost surely over Γ_X. (iv) (location normalization): $\lambda(c) = 0$ for some point c in the interior of Γ_X and the support of u_0^* conditional on $w = c$ is the same as $u_0^*(\Gamma_X)$. Then under [M4], $\lambda(x) = \lambda(w)$ is identified on Γ_W and the distribution G is identified on $u_0^*(\Gamma_X)$.

Proof. Under [M2] and assumption (ii), $m(x) = m(w)$ does not depend on z. This together with assumption (i) implies that $\lambda(x) = \lambda(w)$ is a function of w only. Hence the reduced form equation of this model is

$$P(\delta = 1|w, z) = G(\lambda(w) - u_0^*(w, z)) = P(\delta = 1|w, u_0^*(w, z)).$$

Using assumption (iv), we have

$$G(t) = P(\delta = 1|w = c, u_0^*(w, z) = -t).$$

Assumptions (iii) and (iv) imply that u_0^* is a continuous random variate and the distribution of u_0^* conditional on $w = c$ is non-degenerate with support $u_0^*(\Gamma_X)$. Therefore, by varying u_0^*, the distribution G is then identified on $u_0^*(\Gamma_X)$. Regarding identification of $\lambda(w)$, note that

$$\lambda(w) = G^{-1}(P(\delta = 1|w, z)) + u_0^*(w, z). \tag{21}$$

Since G is identified on $u_0^*(\Gamma_X)$ and $\Delta v^*(\Gamma_X) \subset u_0^*(\Gamma_X)$, (21) implies that $\lambda(w)$ is identified on Γ_W. ■

The assumptions of excluded regressors (i) and conditional independence (ii) in Theorem 4 can produce the required exclusion restriction between the two functions $\lambda(x)$ and $u_0^*(x)$ so that we can distinguish the source of variation between the unknown Δv^* and G by moving only the excluded variables. Assumption (i) may be justified if there are choice specific attributes. Note that assumptions (i) does not preclude common attributes since the vector of attributes w is allowed to appear in u_0^* and hence v_0^*. Only the attributes z are excluded. However, assumption (i) alone is not sufficient to generate the required exclusion restriction in the conditional choice probability equation since the choice specific attributes may enter both value functions via the information set that the agent uses to form their expected future value. Assumption (ii) is sufficient to remove such effect by regulating the predicability of these attributes through conditional independence assumption. A sufficient condition to validate assumption (ii) is the case in which the transition probability densities $f(w', z'|w, z, \delta = k)$ for $k \in \{0, 1\}$ can be factored out as the product of $f(z'|w', \delta = k)f(w'|w, \delta = k)$. Hence assumption (ii) essentially requires that once conditional on the choice, w serves as a sufficient statistic for predicting all observed state variables of next period. Note that assumption (i) in Theorem 4 is not testable and should be regarded as the researcher's belief in modeling a particular application. However, assumption (ii) is in principle testable because this restriction is imposed on observed state variables. For instance, as simple diagnosis, one can plot the joint conditional densities $f(w', z'|w, z, \delta = k)$ and $f(w', z'|w, \delta = k)$ to see whether these two density functions look similar. Note that semiparametric identification in Theorem 4 can be achieved by requiring only one excluded variable z and over-identification may arise if more than one exclusion restrictions are available. Therefore, the identification restrictions stated in Theorem 4 may not be very demanding for practical applications.

Theorem 4 requires u_0^* is specified a priori. However, complete specification of the instantaneous utility function u_0^* may not be necessary in the sense that one can specify it up to finite dimensional unknown parameters. Of course, if u_0^* depends on some unknown parameters, one also needs to guarantee these parameters are identified. Let $u_0^*(x) = u_0^*(w, z; \theta)$ be known up to a finite J-dimensional vector of parameters, $\theta \in \Theta$, where Θ is a compact subset of R^J. Instead of providing identification of θ in the general case, we give the identification result for some popular specifications of utility functions.

Theorem 5

Assume all conditions in Theorem 4 still hold except that $u_0^*(x) = u_0^*(w, z; \theta) = \theta' h(w, z)$ where $h(w, z)$ is a J-dimensional vector of known functions $h_j(w, z) \in B$ and each component θ_j in the vector θ is not zero and each $h_j(w, z)$ function has non-zero partial derivative with respect to z. Then θ is identified up to a scale normalization.

Proof. Given $u_0^*(x) = u_0^*(w, z; \theta) = \theta' h(w, z)$, the conditional choice probability is

$$P(\delta = 1|w, z) = G(\lambda(w) - \theta' h(w, z)) = P(\delta = 1|w, h(w, z)).$$

Taking partial derivative with respect to $h_j(w, z)$, we have

$$\frac{\partial P(\delta = 1|w, h(w, z))}{\partial h_j} = -g(\lambda(w) - \theta' h(w, z))\theta_j.$$

Using the average derivative arguments, we have

$$E(\frac{\partial P(\delta = 1|w, h(w, z))}{\partial h_j}) = -E(g(\lambda(w) - \theta' h(w, z)))\theta_j$$

Therefore, θ_j is identified up to a scale normalization. ■

Theorem 5 provides the identification result when the instantaneous utility u_0^* is specified as linear in parameters. Note that in order to provide the required exclusion restriction, each $h_j(w, z)$ in Theorem 5 needs to be a non-trivial function in z. Therefore, it is now clear to note that the variable z essentially serves as the "instrumental variable" in semiparametric identification of this structural model in the sense that it provides variation for the observed but endogenous object $h(w,z)$ (the "rank condition") but cannot affect the "unobserved" object $\lambda(w)$ (the "exclusion" condition). Theorem 5 adopts the usual average derivative approach to identify the unknown parameters θ. As in static discrete choice models, the average derivative method can at best identify the unknown parameters up to scale normalization. Once the scale of θ is determined, using Theorem 4 we can proceed to identify G. To identify the remaining structural parameters, we assume Γ_G, the support of G is a subset of $u_0^*(\Gamma_X)$ so that G is completely identified on its support[23]. Then by Lemma 2, given β, u_0^* and the identified G, the remaining structural parameters $(u_1^*(x), v_1^*(x), v_0^*(x))$ are also identified.

4 Identification of the dynamic optimal stopping model

In this section, we study identification of a popular variant of the basic structural dynamic discrete choice model that allows for one of the choices to be irreversible (absorbing). Such model is called an optimal stopping time model because the agent has to decide when it is optimal to stop (at the absorbing choice).

4.1 Semiparametric identification via exclusion restriction

The structural dynamic optimal stopping time model has appeared as an important modeling framework in many empirical applications in which the agent faces an irreversible choice. See for example Pakes (1986) in patent renewal, Das (1992) in capital disposal, Daula and Moffitt (1995) in military reenlistment, Rothwell and Rust (1997) in nuclear plant operation, Karlstrom, Palm and Svensson in (2004) and Heyma (2004) in retirement from labor force models. The distribution of unobservables in these models is often parametrically specified and estimated. In this section, we investigate the identification for such models when the researcher does not hold a priori information on the distribution of unobservables. Let Choice 0 be the irreversible (absorbing) choice. Lemma 3 shows that the dynamic optimal stopping time model is not identified. To remedy this problem, we can resort to the use of exclusion restriction as proposed in Section 3.2. To see this, observe that the time series of observed choice in this optimal stopping time model follows

$$\delta' = 1\{\Delta v^*(x') + \Delta\varepsilon' \geq 0\}\delta \tag{22}$$

and the associated Bellman equations are

$$v_1^*(x) = u_1^*(x)+\beta E(v_0^*(x')|x,\delta=1)+\beta[E(\Delta v^*(x')\delta'|x,\delta=1)+E(\Delta\varepsilon'\delta'|x,\delta=1)] \tag{23}$$

[23] Since utilities are assumed sup-norm bounded, this assumption essentially precludes those distributions G with unbounded support. See Section 4.2 for identification analysis when G has large support and hence is not identified over its entire support.

and

$$v_0^*(x) = u_0^*(x) + \beta E(v_0^*(x')|x, \delta = 0). \tag{24}$$

Note that the Bellman equation (24) can be interpreted in a similar fashion as (23) except that the term of the expected gain vanishes due to the behavioral restriction that choice 0 is irreversible and hence deviation is not allowed once choice 0 is made at previous period. Clearly, in this model once β and u_0^* are given, v_0^* can be identified as the fixed point solution of the functional equation (24) that does not depend on the distribution G. The reduced form equation of this optimal stopping time model is the conditional choice probability

$$P(\delta' = 1|x', \delta = 1) = G(v_1^*(x') - v_0^*(x')). \tag{25}$$

Following the discussions in Section 3.2, we decompose the state variables x into two mutually exclusive components, w and z. Since v_0^* is identified when β and u_0^* are given, identification with exclusion restriction can be achieved based on similar arguments as given in Theorem 4 if (25) can be rewritten as

$$P(\delta' = 1|w', z', \delta = 1) = G(v_1^*(w') - v_0^*(w', z')). \tag{26}$$

The next theorem presents sufficient conditions to validate (26) and thus establish the identification of G and v_1^* in this optimal stopping time model. Denote the support of x as $\Gamma_X = \Gamma_W \times \Gamma_Z$, where Γ_W and Γ_Z are the supports of w and z, respectively. Let $v_0^*(\Gamma_X)$ and $\Delta v^*(\Gamma_X)$ be the supports of v_0^* and Δv^*, respectively. We assume that $\Delta v^*(\Gamma_X) \subset v_0^*(\Gamma_X)$.

Theorem 6

Let $x = (w, z)$ in which the subvectors w and z have no common component and both of them are non-empty. Let $u_0^*(x) = u_0^*(w, z) \in B$ be a known function. Assume also the following : (i) (excluded regressors): $u_1^*(x) = u_1^*(w)$ (ii) (conditional independence between observables): $(w', z') \perp z|w, \delta = 1$ (iii) (continuous regressors): $\exists$ at least one continuous variable z_s in z such that the distribution of z_s conditional on w is non-degenerate and $v_0(x)$ is differentiable with respect to z_s with $\frac{\partial v_0^*(x)}{\partial z_s} \neq 0$ almost surely over Γ_X. (iv) (location normalization): $v_1^*(c) = 0$ for some point c in the interior of Γ_X and the support of v_0^* conditional on $w = c$ is the same as $v_0^*(\Gamma_X)$. Then under [M4], given β, $v_0^*(x) = v_0^*(w, z)$ is identified on Γ_X and $v_1^*(x) = v_1^*(w)$ is identified on Γ_W and the distribution G is identified on $v_0^*(\Gamma_X)$.

Proof. Given β and $u_0^*(x) = u_0^*(w, z) \in B$, we can identify $v_0^*(x) = v_0^*(w, z)$ by solving the fixed point of Bellman equation (24). Under [M2], assumptions (i) and (ii), the Bellman equation (23) implies that $v_1^*(x) = v_1^*(w)$. Hence the reduced form equation of this optimal stopping time model is

$$P(\delta' = 1|w', z', \delta = 1) = G(v_1^*(w') - v_0^*(w', z')).$$

Using assumption (iv), we have

$$G(t) = P(\delta' = 1 | w' = c, v_0^*(w', z') = -t, \delta = 1).$$

Assumptions (iii) and (iv) imply that v_0^* is a continuous random variate and the distribution of v_0^* conditional on $w = c$ is non-degenerate with support $v_0^*(\Gamma_X)$. Therefore, by varying v_0^*, the distribution G is then identified on $v_0^*(\Gamma_X)$. Regarding identification of $v_1^*(w')$, note that

$$v_1^*(w') = G^{-1}(P(\delta' = 1 | w', z', \delta = 1)) + v_0^*(w', z'). \tag{27}$$

Since G is identified on $v_0^*(\Gamma_X)$ and $\Delta v^*(\Gamma_X) \subset v_0^*(\Gamma_X)$, (27) implies that $v_1^*(w')$ is identified on Γ_W. $\blacksquare$

To produce the required exclusion restriction, we still need the assumption (i) of excluded regressors (i) and (ii) of conditional independence between observables. However, unlike Theorem 4 that aims to identify the structural model in which all choices are recurrent, the availability of an irreversible choice provides an extra behavioral restriction such that Bellman equation (24) for Choice 0 does not depend on the distribution of unobservables, thus assumption (ii) in Theorem 6 requires conditional independence between the next period observed state variables $x' = (w', z')$ and current period excluded variable z to hold only for the subpopulation $\delta = 1$ who chooses the non-absorbing choice. Such weaker version of conditional independence is important since in most dynamic optimal stopping time models evolution of z is not serially conditionally independent for those who chooses the absorbing choice ($\delta = 0$). For instance, consider a structural dynamic retirement model[24] in which the agent has to decide between to continue working ($\delta = 1$) or to retire ($\delta = 0$). Suppose retirement decision is irreversible. Note that this is an optimal stopping time model in which the agent has to decide when it is optimal to stop working. A potential excluded variable z in this example can be the individual pension allowance. Since the agent does not receive any pension before he retires, assumption (i) in Theorem 6 holds. Note that assumption (ii) of Theorem 4 requires that the time series of pension is serially conditionally independent for both working and retired subpopulations. Such assumption is clearly violated for most pension schemes in which the individual receives the same (yearly) pension allowance after retirement. In contrast, assumption (ii) in Theorem 6 completely allows for any serial dependence of the post-retirement pension series as long as the law of (potential) pension before retirement satisfies the required conditional independence assumption (ii). A sufficient condition to validate this assumption is the case in which the transition probability density $f(w', z' | w, z, \delta = 1)$ can be factored out as the product of $f(z' | w', \delta = 1) f(w' | w, \delta = 1)$. In other words, assumption (ii) of Theorem 6 essentially requires that once conditional on $\delta = 1$, w serves as a sufficient statistic for predicting all observed state variables of next period.

[24]See Karlstrom, Palm and Svensson in (2004) for a more detailed description of such retirement model in the parametric framework.

Note that we can put any relevant observed explanatory variables in w so that assumption (ii) can be fulfilled. In the retirement example, assumption (ii) is likely to hold if the evolution of pension z before retirement is designed to depend on other observed individual or job-specific characteristics such as marital status, education, or rank of the job. Note that as in Theorem 4, assumption (i) is not testable and should be regarded as the researcher's belief in modeling a particular application. But assumption (ii) is in principle testable because this restriction is imposed on observed state variables. As simple diagnosis, one can plot the joint conditional densities $f(w', z'|w, z, \delta = 1)$ and $f(w', z'|w, \delta = 1)$ to see whether these two density functions look similar. Note that semiparametric identification in Theorem 6 can be achieved by requiring only one excluded variable z and over-identification may arise if more than one exclusion restrictions are available. Therefore, the identification restrictions stated in Theorem 6 may not be very demanding for practical applications.

Theorem 6 identifies v_1^* and the distribution G up to a location normalization (assumption iv). Since identification of v_0^* in an optimal stopping time model does not require the information of G, unlike Theorem 4, we can identify of the partial derivatives of the value functions without location normalizaton. These marginal values may already be the structural objects of interest and are free from arbitrariness of the location normalization assumption. For example, in doing welfare analysis, we may need $\partial E(v(x', \varepsilon')|x', \delta = 1)/\partial x_k$, a measure to assess the impact of changing a particular state variable x_k on the average social surplus function for the $\delta = 1$ subpopulation. Note that

$$\frac{\partial E(v(x', \varepsilon')|x', \delta = 1)}{\partial x_k} = \frac{\partial v_1^*(x')}{\partial x_k} P(\delta' = 1|x', \delta = 1) + \frac{\partial v_0^*(x')}{\partial x_k} P(\delta' = 0|x', \delta = 1)$$

is identified since the conditional choice probability is identified from the data and all the partial derivatives are identified without the need of location normalization. To see this, using the control function approach, we can obtain direct identification of these derivatives. Assume $v_1^*(w)$ and $v_0^*(w, z)$ are differentiable with respect to some continuous component w_r. The conditional choice probability equation is

$$P(\delta' = 1|w', z', \delta = 1) = G(v_1^*(w') - v_0^*(w', z')) = P(\delta' = 1|w', v_0^*(w', z'), \delta = 1)$$

Take partial derivative with respect to w_r and get

$$\frac{\partial P(\delta' = 1|w', z', \delta = 1)}{\partial w_r} = g(v_1^*(w') - v_0^*(w', z')) \left[\frac{\partial v_1^*(w')}{\partial w_r} - \frac{\partial v_0^*(w', z')}{\partial w_r} \right] [25] \tag{28}$$

On the other hand, we can identify $g(v_1^*(w') - v_0^*(w', z'))$ by taking derivative with respect to v_0^* as follows.

$$\frac{\partial P(\delta' = 1|w', v_0^*(w', z'), \delta = 1)}{\partial v_0^*} = -g(v_1^*(w') - v_0^*(w', z')) \tag{29}$$

[25] Recall that g is the density of $-\Delta\varepsilon$

So putting (28) and (29) together, we have

$$\frac{\partial v_1^*(w')}{\partial w_r} = -\left(\frac{\partial P(\delta'=1|w',z',\delta=1)}{\partial w_r}\right) / \left(\frac{\partial P(\delta'=1|w',v_0^*(w',z'),\delta=1)}{\partial v_0^*}\right) + \frac{\partial v_0^*(w',z')}{\partial w_r} \tag{30}$$

Therefore, identification of $\frac{\partial v_1^*(w')}{\partial w_r}$ immediately follows since given β and u_0^* the right hand side objects of (30) are identified.

As given in Theorem 5, for identification of this dynamic optimal stopping time model, complete specification of the instantaneous utility function u_0^* is not necessary and one can specify it up to finite dimensional unknown parameters. Let $u_0^*(x) = u_0^*(w,z;\theta)$ be known up to a finite J-dimensional vector of parameters, $\theta \in \Theta$, where Θ is a compact subset of R^J. We also give the identification result for the dynamic optimal stopping time model when u_0^* is specified as linear in parameters.

Theorem 7

Assume all conditions in Theorem 6 still hold except that $u_0^*(x) = u_0^*(w,z;\theta) = \theta' h(w,z)$ where $h(w,z)$ is a J-dimensional vector of known functions $h_j(w,z) \in B$ and each component θ_j in the vector θ is not zero and each $h_j(w,z)$ function has non-zero partial derivative with respect to z. Then $v_0^*(x) = \theta' r(w,z)$, where $r(w,z)$ is a J-dimensional vector of functions $r_j(w,z)$ with each $r_j(w,z) \in B$ satisfying the Bellman equation $r_j(w,z) = h_j(w,z) + \beta E(r_j(w',z')|w,z,\delta=0)$. Hence, given β, θ is identified up to a scale normalization.

Proof. Given $u_0^*(x) = u_0^*(w,z;\theta) = \theta' h(w,z)$, we shall first verify the conjecture that $v_0^*(x) = \theta' r(w,z)$ does satisfy Bellman equation (23). Plugging $v_0^*(x) = \theta' r(w,z)$ into equation (23), we have

$$\theta' r(x) = \theta' h(x) + \beta E(\theta' r(x')|x,\delta=0) = \theta'(h(x) + \beta E(r(x')|x,\delta=0))$$

Therefore,

$$r(x) = h(x) + \beta E(r(x')|x,\delta=0) \tag{31}$$

Since $h(x) \in B$, $r(x)$ is the unique fixed point of Bellman equation (31). So $v_0^*(x) = \theta' r(x)$ is the unique fixed point of Bellman equation (23). Since $h(x)$ is known, given β, $r(x)$ is then identified using equation (31) and hence $v_0^*(x)$ is identified up to θ. The conditional choice probability in this case is

$$P(\delta'=1|w',z',\delta=1) = G(v_1^*(w') - v_0^*(w',z')) = G(v_1^*(w') - \theta' r(w',z')).$$

Hence, we have

$$P(\delta'=1|w',r(w',z'),\delta=1) = G(v_1^*(w') - \theta' r(w',z')).$$

Taking partial derivative with respect to $r_j(w',z')$, we have

$$\frac{\partial P(\delta'=1|w',r(w',z'),\delta=1)}{\partial r_j} = -g(v_1^*(w') - \theta' r(w',z'))\theta_j$$

Using the average derivative arguments, we have

$$E(\frac{\partial P(\delta' = 1|w', r(w', z'), \delta = 1)}{\partial r_j}) = -E(g(v_1^*(w') - \theta' r(w', z')))\theta_j$$

Therefore, θ_j is identified up to a scale normalization. ■

4.2 Partial identification when the distribution G has large support

To identify the remaining structural parameters $(u_1^*(x), v_1^*(x))$, if Γ_G, the support of G is a subset of $v_0^*(\Gamma_X)$ so that G is completely identified on its support, by Lemma 2, given β, u_0^* and the identified G, $(u_1^*(x), v_1^*(x))$ are also identified. However, such support assumption essentially precludes those popular specifications G of unbounded support. In identifying the basic model as discussed in Section 3.2, complete identification of G is essential to compute the discounted expected gain $E(\Delta\varepsilon'\delta'|x, \delta = 0)$ in Bellman equation (13) for Choice 0 so that v_0^* can be solved and hence identified. Obtaining v_0^* is a key step to the identification of v_1^* because v_1^* can be solved from (**??**) using information of G and v_0^*. When Γ_G is unbounded, G is then only partially identified over the relevant subset of its support. In this case, there is lack of identification of the basic model since $E(\Delta\varepsilon'\delta'|x, \delta = 0)$ is not identified and v_0^* in this case cannot be uniquely determined. Therefore, we cannot yet separately identify v_1^* and v_0^* though the difference Δv^* is identifiable over the relevant support where G is identified.

However, in contrast with the basic model, the extra behavioral restriction from the presence of an irreversible choice allows us to characterize the degree of under-identification when G has large support. To see this, note that given β and u_0^*, we can always identify v_0^* as the fixed point solution of Bellman equation (24). Theorem 6 identifies G on the support of $v_0^*(\Gamma_X)$. This information allows us to identify the difference Δv^* and thus v_1^* can be separately identified since v_0^* is already known. With information of v_1^*, we can proceed to discuss the identification of $u_1^*(x)$. From Bellman equation (23), we have

$$u_1^*(x) = v_1^*(x) - \beta E(v_0^*(x')|x, \delta = 1) - \beta[E(\Delta v^*(x')\delta'|x, \delta = 1) + E(\Delta\varepsilon'\delta'|x, \delta = 1)] \tag{32}$$

Given β, after applying Theorem 6, all righthand side objects except the last term of equation (32) can be identified. The term $E(\Delta\varepsilon'\delta'|x, \delta = 1)$ may not be identified because the support of the distribution G may not be completely contained in the support of v_0^*. By further investigating the unidentified term $E(\Delta\varepsilon'\delta'|x, \delta = 1)$, although not point identified, we can show that the upper bound of $u_1^*(x)$ can be identified. To see this, let the support of v_0^* be $[L_v, U_v]$

and assume the support of Δv^* is contained in $[L_v, U_v]$. Note that

$$\begin{aligned}
E(\Delta\varepsilon'\delta'|x,\delta=1) &= E(\Delta\varepsilon'|\delta'=1,x,\delta=1)P(\delta'=1|x,\delta=1)\\
&= E(E(\Delta\varepsilon'|\delta'=1,x',x,\delta=1)|\delta'=1,x,\delta=1)P(\delta'=1|x,\delta=1)\\
&= E(E(\Delta\varepsilon'|\Delta v^*(x')+\Delta\varepsilon'\geq 0,x',x,\delta=1)|\delta'=1,x,\delta=1)P(\delta'=1|x,\delta=1)\\
&= -\left[\int_{\Gamma_x}\frac{\int_{-\infty}^{\Delta v^*(\tau)}(-\Delta\varepsilon')g(-\Delta\varepsilon')d(-\Delta\varepsilon')}{P(\delta'=1|x'=\tau,\delta=1)}f(\tau|\delta'=1,x,\delta=1)d\tau\right]P(\delta'=1|x,\delta=1)
\end{aligned} \tag{33}$$

We can further analyze the term $\int_{-\infty}^{\Delta v^*(\tau)}(-\Delta\varepsilon')g(-\Delta\varepsilon')d(-\Delta\varepsilon')$ in (33) as follows. Let $\eta = -\Delta\varepsilon'$ and $W \equiv \int_{-\infty}^{L_v} G(\eta)d\eta$. Then

$$\begin{aligned}
\int_{-\infty}^{\Delta v^*(\tau)}(-\Delta\varepsilon')g(-\Delta\varepsilon')d(-\Delta\varepsilon') = \int_{-\infty}^{\Delta v^*(\tau)}\eta g(\eta)d\eta &= \int_{-\infty}^{L_v}\eta g(\eta)d\eta + \int_{L_v}^{\Delta v^*(\tau)}\eta g(\eta)d\eta\\
&= L_v G(L_v) + \int_{L_v}^{\Delta v^*(\tau)}\eta g(\eta)d\eta - W
\end{aligned} \tag{34}$$

Note that the value of W is not identified since the distribution G is not identified for $\eta \notin [L_v, U_v]$. However, the "bias" term W is non-negative and shrinks to zero when the lower bound of the support of v_0^* approach the lower bound of the support of G. Plugging (34) into (33), we have

$$E(\Delta\varepsilon'\delta'|x,\delta=1) = (W - L_v G(L_v))E(\frac{\delta'}{P(\delta'=1|x',\delta=1)}|x,\delta=1)) - AP(\delta'=1|x,\delta=1), \tag{35}$$

where A is defined as

$$A = \int_{\Gamma_x}\frac{\int_{L_v}^{\Delta v^*(\tau)}\eta g(\eta)d\eta}{P(\delta'=1|x'=\tau,\delta=1)}f(\tau|\delta'=1,x,\delta=1)d\tau$$

Since W is non-negative, setting $W = 0$ will give a lower bound of $E(\Delta\varepsilon'\delta'|x,\delta = 1)$ and hence the upper bound of $u_1^*(x)$ is identified. Therefore, treating the tail area of G as zero (setting $W = 0$) gives a nonparametric approximation of $u_1^*(x)$ and this approximation gets more precise when G is less heavy-tailed. In practice, one would expect that the bias W does not matter if the support of v_0^* is wide enough.

It is clear that given the discount factor β, W can be identified if one is willing to parameterize $u_1^*(x)$. For example, if $u_1^*(x) = u_1^*(x,\alpha)$ for some finite dimensional vector of parameters α, then W and α can be (over-) identified by plugging (35) into (32) and then solving for W and α[26].

[26] Of course, the usual rank condition should be satisfied to guarantee a unique solution of W and α.

5 An illustrating example : optimal replacement of production capital

We present in this section a simple motivating example in which the analysis of this paper may be useful.

Consider Rust (1987)'s classic model of optimal replacement of bus engines. A bus operation company at every period has to decide for each bus whether the engine of the bus should be replaced with a new one. Let Choice 1 be the decision of continuing bus operation with its current engine and Choice 0 be that of replacing with a completely new engine. Following Rust (1987), we assume that the bus maintenance manager behaves as a cost minimizer with instantaneous choicewise utilities

$$\begin{aligned} u_1(x,\varepsilon_1) &= -c(w)+\varepsilon_1, \\ u_0(x,\varepsilon_1) &= z-p+\varepsilon_0, \end{aligned}$$

where w is the cumulative mileage of the bus, z is the manager's booked (estimated) remaining value of the bus engine that may be observed in the company's financial statements, p is the price of the new engine, ε_1 can be interpreted as unobserved cost of operating the bus with an old engine and ε_0 can account for any discrepancy between z and actual scrapping value of the old engine. Note that in this model $u_1^*(x)=-c(w)$ and $u_0^*(x)=z-p$. The parameters of interest are the nonparametric cost function $c(w)$ and the value functions v_0^* and v_1^*. Clearly, the price process of p is generally serially correlated[27] and hence it does not satisfy the conditional independence assumption (ii) in Theorem 4 though it does fulfil the excluded regressor assumption (i). Therefore, in this model only z is the potential candidate for a valid excluded variable that satisfies both assumptions (i) and (ii). Assumption (i) is trivially satisfied for z. The conditional independence assumption may be satisfied when the company estimates the remaining value of the engine based on its used mileage w, which does serve as a key factor of its reselling value, plus some independent random noise that reflects the company's assessment of the current status of the engine[28].

It is worth to investigate using this example the support condition and location normalization assumed in Theorem 4. The observed state variable $x=(w,z,p)$. Let Γ_X be the support of x. Note that we require that both $u_1^*(x)$ and $u_0^*(x)$ are sup-norm bounded to ensure unique fixed point to the Bellman equations (13). This condition generally requires that the support of x be bounded[29]. Let $u_0^*(\Gamma_X)$ and $\Delta v^*(\Gamma_X)$ be the supports of u_0^* and Δv^*, respectively. Since G is identified only on $u_0^*(\Gamma_X)$ and identification of Δv^*

[27] Note the (equilibrium) price of the engine is a macro-level variable to be determined in the market. If the market of bus engines is efficient, the price p is generally a random walk process.

[28] As discussed before, this conditional independence restriction is indeed testable because it is imposed on observed state variables.

[29] Rust (1987) discretizes w into finite number of grids and thus the support of w in his model is essentially bounded.

follows from identification of G, for complete identification we requires that $\Delta v^*(\Gamma_X) \subset u_0^*(\Gamma_X)$ and Γ_G,the support of G be a subset of $u_0^*(\Gamma_X)$. Since $u_0^*(\Gamma_X)$ is bounded, these support conditions precludes those unobservables whose distributions have unbounded support. If the support conditions do not hold, then complete identification is not available and Theorem 4 should be regarded as a result of partial identification in which the distribution G is only identified on $u_0^*(\Gamma_X)$[30]. The location normalization assumption (iv) of Theorem 4 essentially requires that there be a known point c in the interior of Γ_X and a known constant b such that[31]

$$\lambda(c) = u_1^*(c) + m(c) = b,$$

where $m(c) = m_1(c) + m_2(c) + m_3(c)$ and m_k, $k \in \{1, 2, 3\}$ is given in (18), (19) and (20). Since each m_k depends on dynamic value functions which are not primitive structural objects, such location normalization is not innocuous and should be interpreted with caution. Consider a researcher who estimates a parametric specification of the bus engine replace model. The structural model can then be computed using the consistent estimates and predicted conditional choice probabilities can be plotted. The researcher may find that the plot of conditional choice probabilities at a point c in the interior of Γ_X matches the data very well. The researcher can then compute $\lambda(c)$ using the estimated model parameters. Note that $\lambda(c)$ depends on the postulated parameterization of the model and we denote such dependence by $\lambda(c) = \lambda_{M,c}$, where M is the vector of parameters that characterizes the model. Therefore, the researcher may be willing to keep the fit of the originally parametric model at the point c but allow for more flexible specification of structural functions evaluated at other points. In this case, the researcher may choose to achieve the location normalization required in Theorem 4 by setting $b = \lambda_{M,c}$. In other words, the location normalization assumption can be interpreted as an assumption that the postulated model in terms of M is correctly specified at the point c and thus nonparametric identification using Theorem 4 can remedy potential misspecification of the structural model M at other points in Γ_X.

When all conditions in Theorem 4 are fulfilled, we can nonparametrically identify the cost function and thus our inference of this model of optimal replacement policy of capital can be more robust against misspecification of distribution of the unobserved modeling components.

6 A sketched estimation strategy

In this section we sketch a preliminary estimation strategy based on the identification analysis that has been presented so far.

[30] If Choice 0 is absorbing, then the upper bound of $u_1^*(x)$ is also identified (See Section 4.2). We leave for further research the characterization of the partially identified set of $u_1^*(x)$ of a general recurrent model.

[31] In Theorem 4, the constant b is assumed to be zero for convenience. However, we can easily modify the theorem to accommodate location normalization to other non-zero value of b.

6.1 Estimating the basic model

The estimation procedure of the basic model proceeds as follows. Assume data consists a random sample of n individual observations $(\delta_i, w_i, w_i', z_i, z_i')$. For simplicity, we assume both w and z are scalar continuous random variates. Let $K(.)$ be the usual symmetric univariate kernel function and h be the bandwidth parameter. Theorem 4 implies that

$$G(t) = P(\delta = 1|w = c, u_0^*(w, z) = -t).$$

Thus, we can estimate the distribution G using sample analog estimator as

$$\widehat{G}(t) \equiv \frac{\sum_{i=1}^{n} \delta_i K(\frac{w_i - c}{h}) K(\frac{u_0^*(w_i, z_i) + t}{h})}{\sum_{i=1}^{n} K(\frac{w_i - c}{h}) K(\frac{u_0^*(w_i, z_i) + t}{h})}.$$

Using $\widehat{G}$, we can estimate Δv^* as follows.

$$\widehat{\Delta v^*}(w, z) \equiv \widehat{G}^{-1}(\widehat{P}(\delta = 1|w, z)),$$

where $\widehat{P}(\delta = 1|w, z)$ is the kernel estimator of the conditional choice probability

$$\widehat{P}(\delta = 1|w, z) \equiv \frac{\sum_{i=1}^{n} \delta_i K(\frac{w_i - w}{h}) K(\frac{z_i - z}{h})}{\sum_{i=1}^{n} K(\frac{w_i - w}{h}) K(\frac{z_i - z}{h})}.$$

Estimation of v_0^* is more involved. We shall estimate the terms of discounted expected gain, $\psi_0(w) \equiv E(\Delta v^*(w', z')\delta'|w, \delta = 0)$ and $e_0(w) \equiv E(\Delta\varepsilon'\delta'|w, \delta = 0)$[32]. The first term can be estimated as

$$\widehat{\psi}_0(w) \equiv \frac{\sum_{i=1}^{n} \widehat{\Delta v^*}(w_i', z_i')\delta_i' K(\frac{w_i - w}{h})(1 - \delta_i)}{\sum_{i=1}^{n} K(\frac{w_i - w}{h})(1 - \delta_i)}. \tag{36}$$

The estimation of $e_0(w)$ can be based on a simulation procedure. Note that

$$E(\Delta\varepsilon'\delta'|w, \delta = 0) = E(\Delta\varepsilon' 1\{\Delta v^*(w', z') + \Delta\varepsilon' \geq 0\}|w, \delta = 0).$$

We can first simulate $\widehat{\Delta\varepsilon_i'}$ by drawing from $\widehat{G}$. For each individual i, we can compute the simulated choice $\widehat{\delta_i'} \equiv 1\{\widehat{\Delta v^*}(w_i', z_i') + \widehat{\Delta\varepsilon_i'} \geq 0\}$ and then form the

[32]Note that the assumptions in Theorem 4 imply that both ψ_0 and e_0 are functions of w only.

estimator

$$\widehat{e}_0(w) \equiv \frac{\sum_{i=1}^{n} \widehat{\Delta\varepsilon'_i}\widehat{\delta}'_i K(\frac{w_i-w}{h})(1-\delta_i)}{\sum_{i=1}^{n} K(\frac{w_i-w}{h})(1-\delta_i)}. \tag{37}$$

Therefore, we can construct the following empirical Bellman equation of Choice 0[33].

$$\widehat{v_0^*}(w,z) = u_0^*(w,z) + \beta(\widehat{\psi}_0(w) + \widehat{e}_0(w)) + \beta\widehat{E}_{w',z'|w,0}(\widehat{v_0^*}(w',z')), \tag{38}$$

where for any function $\lambda(.)$, $\widehat{E}_{w',z'|w,d}(\lambda(w',z'))$ is an empirical conditional expectation operator of λ that is constructed as follows

$$\widehat{E}_{w',z'|w,d}(\lambda(w',z')) \equiv \frac{\sum_{i=1}^{n} \lambda(w'_i,z'_i)K(\frac{w_i-w}{h})1\{\delta_i = d\}}{\sum_{i=1}^{n} K(\frac{w_i-w}{h})1\{\delta_i = d\}}.$$

Given β and u_0^*, the estimator $\widehat{v_0^*}$ of v_0^* is then solved as the unique fixed point of the empirical Bellman equation (38). The following proposition establishes the existence of the fixed point estimator, $\widehat{v_0^*}$.

Proposition 8

Given $\beta \in [0,1)$ and $u_0^* \in B$, the space of all sup-normed bounded functions, then for any sample size n, we have $\widehat{v_0^*} \in B$ as the unique fixed point solution of Bellman equation (38).

Proof. Define the operator $\widehat{T}$ as

$$\widehat{T}(\widehat{v_0^*})(w,z) = u_0^*(w,z) + \beta(\widehat{\psi}_0(w) + \widehat{e}_0(w)) + \beta\widehat{E}_{w',z'|w,0}(\widehat{v_0^*}(w',z'))$$

Because $u_0^* \in B$ and in any finite sample both $\widehat{\psi}_0$ and $\widehat{e}_0$ are bounded, if $\widehat{v_0^*} \in B$, then $\widehat{T}(\widehat{v_0^*}) \in B$ and thus the operator $\widehat{T}$ is a mapping from B to B. To show that this operator induces a contraction mapping, we can check Blackwell sufficient conditions (see Theorem 3.3 in Stockey and Lucas 1989) of monotonicity and discounting. For monotonicity, due to the monotone property of the operator $\widehat{E}_{w',z'|w,0}$, it is clear that $\widehat{T}(v_2) \geq \widehat{T}(v_1)$ if $v_2 \geq v_1$ and both v_1 and v_2 are in B. The discounting condition is also satisfied since for any $a \in R$,

$$\begin{aligned} \widehat{T}(\widehat{v_0^*} + a)(w,z) &= u_0^*(w,z) + \beta(\widehat{\psi}_0(w) + \widehat{e}_0(w)) + \beta\widehat{E}_{w',z'|w,0}(\widehat{v_0^*}(w',z') + a) \\ &= \widehat{T}(\widehat{v_0^*})(w,z) + \beta a \end{aligned}$$

[33] The assumptions in Theorem 4 imply that $E(v_0^*(w',z')|w,z,\delta=0)$ is also a function of w only.

due to the linear property of the operator $\widehat{E}_{w',z'|w,0}$. Therefore, Proposition 8 follows by noting that Blackwell sufficient conditions hold for the operator $\widehat{T}$ and hence $\widehat{T}$ is a contraction mapping. ■

By using $\widehat{v_0^*}$ and $\widehat{\Delta v^*}$, the estimator $\widehat{v_1^*}$ for v_1^* can be immediately obtained as[34]

$$\widehat{v_1^*}(w) = \widehat{\Delta v^*}(w,z) + \widehat{v_0^*}(w,z).$$

The remaining parameter to be estimated is $u_1^*(w)$, which can be obtained via

$$\widehat{u_1^*}(w) \equiv \widehat{v_1^*}(w) - \beta \widehat{E}_{w',z'|w,1}(\widehat{v_0^*}(w',z')) - \beta[\widehat{\psi}_1(w) + \widehat{e}_1(w)], \tag{39}$$

where $\widehat{\psi}_1(w)$ and $\widehat{e}_1(w)$ are estimators of $\psi_1(w) \equiv E(\Delta v^*(w',z')\delta'|w,\delta=1)$ and $e_1(w) \equiv E(\Delta\varepsilon'\delta'|w,\delta=1)$, respectively and they can be obtained as

$$\widehat{\psi}_1(w) \equiv \frac{\sum_{i=1}^{n} \widehat{\Delta v^*}(w_i',z_i')\delta_i'\delta_i K(\frac{w_i-w}{h})}{\sum_{i=1}^{n} \delta_i K(\frac{w_i-w}{h})}, \tag{40}$$

$$\widehat{e}_0(w) \equiv \frac{\sum_{i=1}^{n} \widehat{\Delta\varepsilon_i'}\widehat{\delta_i'}\delta_i K(\frac{w_i-w}{h})}{\sum_{i=1}^{n} \delta_i K(\frac{w_i-w}{h})}. \tag{41}$$

6.2 Estimating the dynamic optimal stopping time model

The estimation strategy of the dynamic optimal stopping time model can be done in a similar way as discussed in previous section. However, for the optimal stopping time model the identification power comes from the control function $v_0^*(w,z)$ and hence the first step is to get the estimator $\widehat{v_0^*}$, which can be solved as the fixed point of its empirical Bellman equation

$$\widehat{v_0^*}(w,z) = u_0^*(w,z) + \beta \widehat{E}_{w',z'|w,0}(\widehat{v_0^*}(w',z')). \tag{42}$$

Note that existence of $\widehat{v_0^*}$ follows from Proposition 8 by setting both $\widehat{\psi}_0(w)$ and $\widehat{e}_0(w)$ to be zero. The next step is to estimate G using $\widehat{v_0^*}$ and the implication of Theorem 6 as

$$\widehat{G}(t) \equiv \frac{\sum_{i=1}^{n} \delta_i'\delta_i K(\frac{w_i'-c}{h})K(\frac{v_0^*(w_i',z_i')+t}{h})}{\sum_{i=1}^{n} \delta_i K(\frac{w_i'-c}{h})K(\frac{u_0^*(w_i',z_i')+t}{h})}.$$

[34]Note that the assumptions in Theorem 4 imply that v_1^* is also a function of w only.

Using $\widehat{G}$, we can estimate Δv^* as follows.

$$\widehat{\Delta v^*}(w', z') \equiv \widehat{G}^{-1}(\widehat{P}(\delta' = 1|w', z', \delta = 1)),$$

where $\widehat{P}(\delta' = 1|w, z, \delta = 1)$ is the kernel estimator of the conditional choice probability

$$\widehat{P}(\delta' = 1|w', z', \delta = 1) \equiv \frac{\sum_{i=1}^{n} \delta_i' \delta_i K(\frac{w_i' - w'}{h}) K(\frac{z_i' - z'}{h})}{\sum_{i=1}^{n} \delta_i K(\frac{w_i' - w'}{h}) K(\frac{z_i' - z'}{h})}.$$

By using $\widehat{v_0^*}$ and $\widehat{\Delta v^*}$, the estimator $\widehat{v_1^*}$ for v_1^* can be immediately obtained as

$$\widehat{v_1^*}(w) = \widehat{\Delta v^*}(w, z) + \widehat{v_0^*}(w, z).$$

We assume that G is identified over its entire support. Then the estimator $\widehat{u_1^*}(w)$ of $u_1^*(w)$ is the same as the expression given in (39).

7 Conclusions

This paper develops semiparametric identification results for structural dynamic discrete choice models. The main parametric assumption for this semiparametric identification method is the (parametric) specification of the per-period return function of one of the choices. The distribution of unobserved state variables and per-period return function for the other choice are both nonparametric. Thus the semiparametric approach adopted in this paper can serve as a diagnostic device to check for parametric assumptions of distribution of unobservables. Our identification strategy does not rely on the availability of terminal period data and hence can be applied to infinite horizon structural dynamic models. The identification crucially depends on the availability of at least one excluded continuous variable that is excluded from the unknown function but enters the known object that serves as control function to identify the model. Primitive conditions to validate such exclusion restriction can be relaxed when there is extra behavioral restriction from the presence of an irreversible choice. We illustrate the use of identification strategy for the classic model of optimal replacement of production capital. We also discuss a potential estimation strategy. The proposed implementation method results in mutli-stage estimation procedure which requires the task of solving a contraction mapping for the empirical counterpart of Bellman equations (13). The asymptotic properties of the estimation procedure proposed in Section 6 are a topic for further research beyond the present paper. Note that $\widehat{v_0^*}$ as a solution of (38) or (42) generally has no closed form representation and can only be obtained iteratively as a numerical fixed point of the corresponding empirical Bellman equation. Hence delta method or asymptotic expansions of $\widehat{v_0^*}$ cannot be direct applicable. Such non-standard problem poses a challenge and an investigation of its potential solutions is the next work on our research agenda.

References

[1] Adda, J. and Cooper R. (2000), "Balladurette and Juppette: A Discrete Analysis of Scrapping Subsidies", *Journal of Political Economy*, 108, 778-806.

[2] Aguirregabiria, V. (2005), "Nonparametric Identification of Behavioral Responses to Counterfactual Policy Interventions in Dynamic Discrete Decision Processes", *Economics Letters*, 87, 393-398.

[3] Aguirregabiria, V. (2008), "Another Look at the Identification of Dynamic Discrete Decision Processes, with an Application to Retirement Behavior", *Journal of Business and Economic Statistics*, forthcoming.

[4] Aguirregabiria, V. and Mira, P. (2008), " Dynamic Discrete Choice Structural Models: A Survey", *Journal of Econometrics*, forthcoming.

[5] Berkovec, J. and Stern, S. (1991), "Job Exit Behavior of Old Man", *Econometrica*, 59, 189-210.

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

[7] Cooper, R., Haltiwanger, J. and Power, L. (1999), "Machine Replacement and the Business Cycle: Lumps and Bumps", *American Economic Review*, 89, 921-946.

[8] Daula, T. and Moffitt, R. (1995), "Estimating Dynamic Models of Quit Behavior: The Case of Military Reenlistment", *Journal of Labor Economics,* 13, 499-523.

[9] Das, S. (1992), "A Micro-Econometric Model of Capital Utilization and Retirement: the Case of the U.S. Cement Industry", *Review of Economic Studies*, 59, 277-297.

[10] Eckstein, Z. and Wolpin, K. (1989), "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey", *Journal of Human Resources*, 24, 562-598.

[11] Eckstein, Z. and Wolpin, K. (1999), "Why Youth Drop out of High School: The Impact of Preferences, Opportunities and Abilities", *Econometrica*, 67, 1295-1339.

[12] Erdem, T. and Keane, M. (1996), "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets", *Marketing Science*, 15, 1-20.

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

[14] Heyma, A. (2004), "A Structural Dynamic Analysis of Retirement Behaviour in the Netherlands", *Journal of Applied Econometrics*, 19, 739-759.

[15] Hotz, V. J. and Miller, R. (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models", *Review of Economic Studies*, 60, 497-529.

[16] Ichimura, H. (1993), "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models", *Journal of Econometrics*, 58, 71-120.

[17] Karlstrom, A., Palm, M. and Svensson, I. (2004), "A Dynamic Programming Approach to Model the Retirement Behavior of Blue-Collar Workers in Sweden", *Journal of Applied Econometrics*, 19, 795-807.

[18] Keane, M. and Wolpin, K. (1997), "The Career Decisions of Young Men", *Journal of Political Economy*, 105, 473-521.

[19] Klein, R. and Spady, R. (1993), "An Efficient Semiparametric Estimator for Binary Response Models", *Econometrica*, 61, 387-421.

[20] Lewbel, A. and Linton, O. (2007), "Nonparametric Matching and Efficient Estimators of Homothetically Separable Functions", *Econometrica*, 75, 1209-1227.

[21] Magnac, T. and Thesmar, D. (2002), "Identifying Dynamic Discrete Decision Processes", *Econometrica*, 70, 801-816.

[22] Manski, C. F. (1975), "Maximum Score Estimation of the Stochastic Utility Model of Choice", *Journal of Econometrics*, 3, 205-228.

[23] Manski, C. F. (1985), "Semiparametric Analysis of Discrete Responses: Asymptotic Properties of the Maximum Score Estimator", *Journal of Econometrics*, 27,313-334.

[24] Manski, C. F. (1993), "Dynamic Choices in Social Settings", *Journal of Econometrics*, 58, 121-136.

[25] Matzkin, R. (1991), "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models", *Econometrica*, 59, 1315-1327.

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

[27] Miller, R. (1984), "Job Matching and Occupational Choice", *Journal of Political Economy*, 92, 1086-1120..

[28] Pakes, A. (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks", *Econometrica*, 54, 755-784.

[29] Rothwell, G. and Rust, J. (1997), "On the Optimal Lifetime of Nuclear Power Plants", *Journal of Business and Economic Statistics*, 15, 195-208.

[30] Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica*, 57, 1027-1058.

[31] Rust (1988), "Maximum Likelihood Estimation of Discrete Control Processes", *SIAM Journal of Control and Optimization*, 26, 1006-1024.

[32] Rust, J. (1994), "Structural Estimation of Markov Decision Processes", in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. McFadden, Amsterdam: North Holland, 3081-3143.

[33] Rust, J. and Phelan, C. (1997), "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets", *Econometrica*, 65, 781-832.

[34] Stoker, T. M. (1986), "Consistent Estimation of Scaled Coefficients", *Econometrica*, 54, 1461-1481.

[35] Stokey, N. and Lucas, R. (1989), "Recursive Methods in Economics Dynamics", Harvard University Press, Cambridge, Massachusetts.

[36] Taber, C. (2000), "Semiparametric Identification and Heterogeneity in Discrete Choice Dynamic Programming Models", *Journal of Econometrics*, 96, 201-229.

[37] Vytlacil, E. (2002), "Independence, Monotonicity, and Latent Index Models: An Equivalence Result", *Econometrica,* 70, 331-341.

[38] Vytlacil, E. (2005), "A Note on Additive Separability and Latent Index Models of Binary Choice: Representation Results", unpublished manuscript, Standford University.

[39] Wolpin, K. (1984), "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality", *Journal of Political Economy*, 92, 852-874.

[40] Wolpin, K. (1987), "Estimating A Structural Search Model: The Transition From School To Work", *Econometrica*, 55, 801-817.

[41] Wolpin, K. (1996), "Public Policy Uses of Discrete Choice Dynamic Programming Models", *American Economic Review*, 86, 427-432.