Continuity and differentiability of expected value functions in dynamic discrete choice models

Andriy Norets

Department of Economics, Princeton University

This paper explores the properties of expected value functions in dynamic discrete choice models. The continuity with respect to state variables and parameters, and the differentiability with respect to state variables are established under fairly general conditions. The differentiability with respect to parameters is proved when some state variables do not affect the state transition probabilities and, thus, the expected value functions. It is shown that such variables are needed so as to apply the implicit function theorem used in the proof. The results are of particular relevance to estimable dynamic discrete choice models.

Keywords. Dynamic discrete choice models, continuity, differentiability.

JEL classification. C35, C61.

1. Introduction

A dynamic discrete choice model (DDCM) is a dynamic program with discrete controls. These models are widely employed in many different areas of economics, especially in empirical work. Early applications and methodological contributions in this area include Wolpin (1984), Miller (1984), Pakes (1986), Rust (1987), Hotz and Miller (1993), Keane and Wolpin (1994), Aguirregabiria and Mira (2002), Bajari, Benkard, and Levin (2007), Su and Judd (2008), Ackerberg (2009), Imai, Jain, and Ching (2009), and Norets (2009). For a much more extensive list of references, see the literature surveys by Eckstein and Wolpin (1989), Rust (1994), and Aguirregabiria and Mira (2007).

Continuity and differentiability of the expected value functions in DDCMs are sometimes assumed and exploited without justification. Popular solution methods for DDCMs involve the approximation of the expected value functions by polynomials or splines in state variables (Keane and Wolpin (1994)). In this case, at least continuity in state variables is required. In more recent solution and estimation methods for DDCMs (Imai, Jain, and Ching (2009), Norets (2009, forthcoming)), continuity and, perhaps, differentiability of the expected value functions with respect to model parameters can be fruitfully exploited. The likelihood function of DDCMs can involve the expected value functions; see Rust (1994). Although a standard proof of the asymptotic normality of the maximum likelihood estimator assumes the third-order differentiability of the likelihood, first-order differentiability combined with some regularity conditions can suffice;

Andriy Norets: `anorets@princeton.edu`

see, for example, Theorems 5.39 and 7.2 in van der Vaart (1998). Similarly, differentiability can be helpful in deriving the asymptotic properties of the generalized method of moments estimator (Hansen (1982)). Differentiability is also important in implementation of estimation procedures, as it permits the application of efficient gradient-based numerical optimization algorithms.

Thus, there seems to be a need in the literature on DDCMs for a set of general and easy to verify sufficient conditions for continuity and differentiability of the expected value functions. There are well known results for continuity and differentiability of the value function as a function of state variables for dynamic models with continuous controls; see Benveniste and Scheinkman (1979) or Stokey and Lucas (1989) for a textbook treatment. To the best of my knowledge, Rust (1988) is the only researcher who deals with these issues for DDCMs. He established the differentiability of the expected value function as a function of parameters for a special class of models with conditionally independent and additive utility shocks.

In this paper, I show that the value function is jointly continuous in state variables and parameters under fairly general conditions. The differentiability of the expected value functions with respect to the state variables is trivial, and amounts to interchanging the order of differentiation and integration operations. To establish sufficient conditions for the differentiability of the expected value function with respect to parameters, I use an implicit function theorem for Banach spaces. Sufficient conditions for differentiability include smoothness of per-period utility functions and state transition densities. The most restrictive part of the sufficient conditions for differentiability is that some state variables are assumed not to affect the state transition probabilities and thus the expected value functions. The obtained sufficient conditions are more general than those in Rust (1988): the utility shocks do not have to be additive, they do not have to be conditionally independent under all possible decisions, and there might be fewer shocks than there are alternatives.

Somewhat surprisingly, the state variables that do not affect the state transition probabilities are needed for applicability of the implicit function theorem; in particular, for the differentiability of the Emax operator. This seems to be the paper's most interesting result. It provides an additional justification for using state variables that do not affect the state transition probabilities in DDCMs. Including these variables is also useful for making the model consistent with any possible realization of the data (Rust (1994, p. 3102)), for constructing feasible Bayesian estimation algorithms (Norets (2009)), and for making conditional choice probabilities well defined.

The obtained sufficient conditions for continuity and differentiability are easy to check in applications. This is illustrated by examples.

In the following section, I set up a dynamic discrete choice model. Sections 3 and 4 present continuity and differentiability results. Section 5 concludes with a brief discussion. Proofs are gathered in the Appendix.

2. Setup of DDCMs

In a dynamic discrete choice model, the agent chooses an alternative d_t from a finite set of available alternatives D in each period t. The per-period utility $u(s_t, d_t; \theta)$ depends on

the chosen alternative, current state variables $s_t \in S$, and a vector of parameters $\theta \in \Theta$. The state variables are assumed to evolve according to a controlled first-order Markov process with a transition probability denoted by $F(ds_{t+1}|s_t, d_t; \theta)$. A density of the transition probability with respect to a generic measure ν is denoted by $f(s_{t+1}|s_t, d_t; \theta)$. Time is discounted with a factor β. The agent maximizes a lifetime utility given by the expected discounted sum of the per-period utilities. In the recursive formulation, the lifetime utility of the agent or the value function is given by the Bellman equation

$$V(s_t; \theta) = \max_{d_t \in D}\{u(s_t, d_t; \theta) + \beta \, \mathrm{EV}(s_t, d_t; \theta)\}, \tag{1}$$

where the expected value function, $\mathrm{EV}(s, d; \theta)$, is defined by

$$\mathrm{EV}(s, d; \theta) = \int V(s'; \theta) f(s'|s, d; \theta) \nu(ds'). \tag{2}$$

This formulation embraces a finite horizon case if time t is included in the vector of the state variables.

In estimable DDCMs, the vector of state variables includes variables unobserved by econometricians. These variables play the role of econometric errors, which are necessary in estimable models. They are often denoted by ε, while the states observed by the econometrician are denoted by x. In a popular framework of Rust (1987), these variables enter utility functions additively ($u(s_t, d; \theta) = u(x_t, d; \theta) + \varepsilon_{t,d}$) and are assumed to be conditionally independent ($f(s_{t+1}|s_t, d_t; \theta) = f(\varepsilon_{t+1}|x_{t+1}; \theta) f(x_{t+1}|x_t, d_t; \theta)$). Under these assumptions and other regularity conditions, Rust (1988) established the differentiability of the expected value functions. In this paper, I do not start with these assumptions. Moreover, the distinction between the observed and unobserved states is not crucial for the results obtained, herein. However, it might be useful to keep this distinction in mind in interpreting some examples and assumptions below.

Bhattacharya and Majumdar (1989) showed that the optimal lifetime utility satisfies the Bellman equation (1) and that there exists an optimal Markovian policy that solves the Bellman equation under the following assumption.

Assumption 1.

(i) S is a nonempty, complete, and separable metric space.

(ii) $u(s, d; \theta)$ is continuous in s.

(iii) $R(s, \theta) = \sum_{k=0}^{\infty} \beta^k u_k(s, \theta) < \infty$, for all s and θ, where

$$u_k(s, \theta) = \max_d \int u_{k-1}(s', \theta) F(ds'|s, d, \theta) \quad \text{and} \quad u_0(s, \theta) = \max_d |u(s, d, \theta)|.$$

(iv) $\int \phi(s') F(ds'|s, d, \theta)$ is continuous in s for $|\phi(s)| \leq R(s, \theta)$.

Parts (i) and (ii) are assumed hereafter. All the results in this paper will assume conditions guaranteeing that parts (iii) and (iv) are satisfied.

3. Continuity

PROPOSITION 1. *If Θ and S are compact, $\nu(S) < \infty$, $u(s, d; \theta)$ is continuous in (s, θ), and $f(s'|s, d; \theta)$ is continuous in (s', s, θ), then $V(s; \theta)$ and $E\{V(s'; \theta)|s, d; \theta\}$ are continuous in (s, θ).*

The details of the proof are given in the Appendix. The argument closely follows the standard proof of the continuity of value functions with respect to the state variables. Under compactness and continuity, Assumption 1 is clearly satisfied. The Bellman operator Γ is defined on the Banach space of bounded functions $V : S \times \Theta \rightarrow R$ (the standard argument for continuity in s defines Γ on functions from S):

$$\Gamma(V)(s; \theta) = \max_d \left\{ u(s, d; \theta) + \beta \int V(s'; \theta) f(s'|s, d; \theta) \nu(ds') \right\}. \tag{3}$$

It is a contraction and it takes the set of continuous bounded functions into itself. Since the set of continuous bounded functions is closed, the unique fixed point of the Bellman operator is continuous in (s, θ).

Dynamic discrete choice models used in empirical work often have unbounded state space and unbounded per-period utility functions. In this case, the Bellman operator might not have the contraction property and the approach described in the previous paragraph does not directly apply. Lippman (1975) and Bhattacharya and Majumdar (1989) proposed sufficient conditions under which a power of the Bellman operator Γ^J is a contraction on a Banach space with a weighted sup norm, where J is a finite positive integer. I extend this approach to establish the continuity of the value functions in the state variables and parameters for DDCMs with unbounded state space. The idea of the extension is the same as for the bounded case. The following restrictions on the primitives of DDCMs are assumed.

ASSUMPTION 2. *There exist a continuous function $w : S \times \Theta \rightarrow R$, $1 \leq w(s, \theta) < \infty$ $\forall (s, \theta) \in S \times \Theta$ and a positive integer m such that $\| \max_d |u(s, d; \theta)| \|_{w^m} \leq M < \infty$.*

The weighted sup norm is defined by $\|V(s, \theta)\|_{w^m} = \sup_{s, \theta} |V(s, \theta)| w(s, \theta)^{-m}$. The following facts about the weighted sup norm can be established along the same lines as similar facts about the usual sup norm: (i) the space of functions bounded in $\| \cdot \|_{w^m}$, B_{w^m}, is a Banach space; (ii) continuous functions in B_{w^m} are a closed set. The weighted sup norm is introduced to make the value functions bounded (in problems with unbounded per-period utilities, the value functions are unbounded in the usual norm). The boundedness is useful for establishing the contraction property. Intuitively, the boundedness of the value function in the weighted sup norm is achieved when per-period utilities do not grow too fast in states relative to the probability of realization for these states. Precise mathematical conditions are given in the assumptions below. Example 1, which follows, demonstrates that it is not difficult to verify these assumptions in models with explicit parametric forms of transition probabilities and per-period utilities. In particular, the example constructs a weight function for a model with normally distributed and serially correlated utility shocks.

ASSUMPTION 3. *There is $b > 0$ such that for $(s, \theta) \in S \times \Theta$,*

$$\max_d \int w(s', \theta)^n f(s'|s, d; \theta)\nu(ds') \leq (w(s, \theta) + b)^n, \quad n = 1, 2, \ldots, m.$$

ASSUMPTION 4. *The transition density $f(s'|s, d; \theta)$ is continuous in (s, θ) for any $d \in D$. For any s, d, θ there exist $\varepsilon > 0, \overline{f}_{s,d,\theta}(s')$, and $\overline{w}_{s,d,\theta}(s')$ such that $\int \overline{w}_{s,d,\theta}(s')^m \cdot \overline{f}_{s,d,\theta}(s') \times \nu(ds') < \infty$ and if $\|(\tilde{s}, \tilde{\theta}) - (s, \theta)\| < \varepsilon$, then*

$$\overline{w}_{s,d,\theta}(s')\overline{f}_{s,d,\theta}(s') \geq w(\tilde{\theta}, \tilde{s})f(s'|\tilde{s}, d; \tilde{\theta}). \tag{4}$$

PROPOSITION 2. *Under Assumptions 2–4, (i) the lifetime utility of the agent satisfies the Bellman equation for a fixed parameter θ; (ii) there exists a positive integer J such that the operator Γ^J is a contraction on B_{w^m}; thus the Bellman equation has a unique solution; (iii) $V(s; \theta)$ and $E\{V(s'; \theta)|s, d; \theta\}$ are continuous in (s, θ).*

The following example illustrates that the imposed assumptions are sufficiently general and easy to verify for models used in empirical applications of DDCMs.

EXAMPLE 1. Let the per-period utility be given by $u(s_t, d; \theta) = u(\theta, d) + s_t^d$, where $u(\theta, d)$ is continuous in θ. The vector of state variables includes card(D) components: $s = \{s_d, d \in D\}$. The components of s evolve according to independent AR(1) processes: $s'_d \sim N(\rho_d s_d, \sigma_d^2)$. In this model, the state variables can be interpreted as serially correlated preference shocks unobserved by econometricians. The correlation coefficients $\rho_d \in [-1, 1]$ and the variances $\sigma_d^2 \leq \overline{\sigma}^2 < \infty$ are included in the vector of parameters θ. Let $m = 1$ and $w(s, \theta) = \max_d |u(\theta, d)| + 1 + \sum_d s_d^2$. Obviously, $\|\max_d |u(s, d; \theta)|\|_{w^m} \leq 1$ and Assumption 2 is satisfied. Assumption 3 holds as well since

$$\int w(s', \theta)f(s'|s, d; \theta)\, ds' = \max_d u(\theta, d) + 1 + \sum_d [(\rho_d s_d)^2 + \sigma_d^2] \leq w(s, \theta) + b,$$

where $b = \text{card}(D)\overline{\sigma}^2$. Let $\overline{w}_{s,d,\theta}(s') = w(s, \theta) + \delta, \delta > 0$ and

$$\overline{f}_{s,d,\theta}(s') = \prod_d c_d \exp\{-0.5(\sigma_d + 1)^{-2}(s'_d - \rho_d s_d)^2\},$$

where c_d is sufficiently large for (4) to hold. Such c_d exists since $[-0.5(\sigma_d \pm \varepsilon)^{-2}(s'_d - (\rho_d \pm \varepsilon)(s_d \pm \varepsilon))^2 + 0.5(\sigma_d + 1)^{-2}(s'_d - \rho_d s_d)^2]$ is bounded above in s'_d for sufficiently small ε. Thus, Assumption 4 is satisfied.

4. DIFFERENTIABILITY

4.1 *Differentiability with respect to state variables*

This section provides sufficient conditions for the differentiability of the expected value function $EV(s, d; \theta)$ defined in (2) with respect to state variables. A partial derivative $\partial EV(s, d; \theta)/\partial s_j$ with respect to a component of s, s_j, exists if differentiation and integration operations can be interchanged in (2). Sufficient conditions for that are given in the following proposition.

PROPOSITION 3. *Under Assumptions 2–4,* $EV(s, d; \theta)$ *is continuously differentiable in s if for all j,* $\partial f(s'|s, d; \theta)/\partial s_j$ *exists and satisfies all the conditions on* $f(s'|s, d; \theta)$ *imposed in Assumption 4.*

4.2 *Differentiability with respect to parameters*

The following simple example illustrates that differentiability with respect to (w.r.t.) parameters can fail even if per-period utility functions and transition densities are smooth.

EXAMPLE 2. Let the per-period utility be given by $u(s, d_1; \theta) = \theta_1 s$ and $u(s, d_2; \theta) = \theta_2$, and let s be independent and identically distributed (i.i.d.). For $\theta_1 = 0$, the future expected value function (which does not depend on the current state and decision) $EV(\theta) = \max\{0, \theta_2\}/(1 - \beta)$ is not differentiable at $(\theta_1, \theta_2) = (0, 0)$.

Thus, there is a need to provide a precise characterization of the conditions that guarantee differentiability. For this purpose, I apply the implicit function theorem for Banach spaces to the Bellman equation. Rust (1988) also used this theorem to prove the differentiability of the expected value function in parameters for a special class of DDCMs with conditionally independent and additive utility shocks. Below, I introduce the notation and describe the implicit function theorem. In Proposition 4, I show that the sufficient conditions for differentiability in the implicit function theorem require the policy function to be almost surely unique, even when small perturbations are introduced in the expected value functions. This in turn implies that there must be some state variables that do not affect state transition probabilities and thus the expected value functions (Proposition 5). After I establish that the state variables that do not affect the expected value function are needed for the applicability of the implicit function theorem, I assume the presence of such variables in developing a set of sufficient conditions for differentiability (Proposition 6). In Proposition 7, I provide a set of sufficient conditions that are easier to verify in applications. I illustrate this in Example 3, which verifies sufficient conditions for the differentiability in several versions of a stylized model of work hours choice. The example also demonstrates that the sufficient conditions for differentiability developed here considerably generalize Rust's results for additive conditionally independent utility shocks. In particular, the example shows that the utility shocks do not have to be additive, they do not have to be serially independent under all possible decisions, and there might be fewer shocks than there are alternatives.

Let C_{w^m} denote the set of continuous functions from S to R bounded in $\|\cdot\|_{w^m}$. Let $v(\cdot) = \{v(\cdot, d) \in C_{w^m}, d \in D\}$ and let $C_{w^m}^{\text{card}(D)}$ denote a Cartesian product of card(D) copies of C_{w^m}. Assume that the weight function w^m does not depend on parameters θ, an assumption that is not restrictive if the parameter space is compact. Define an analog of the Bellman operator for the expected value function (also called Emax operator) $\Lambda : C_{w^m}^{\text{card}(D)} \times \Theta \to C_{w^m}^{\text{card}(D)}$ as

$$\Lambda_d(v(\cdot), \theta)(s) = \int \max_{d'}\{u(s', d'; \theta) + \beta v(s', d')\} f(s'|s, d; \theta) v(ds') \tag{5}$$

and $\Lambda(v(\cdot), \theta) = \{\Lambda_d(v(\cdot), \theta), d \in D\}$. For a given value of the parameter vector θ, the expected value functions $\mathrm{EV}(s, \theta) = \{\mathrm{EV}(s, d, \theta), d \in D\}$ solve the functional equation in $v(\cdot)$,

$$\Phi(v(\cdot), \theta) = O, \tag{6}$$

where $\Phi = (I - \Lambda)$, and O and I are 0 and 1 in the space of operators on $C_{w^m}^{\mathrm{card}(D)}$. Equation (6) is an alternative representation of the Bellman equation.

For a fixed parameter vector θ_0, let $\mathrm{EV}_0(s) = \mathrm{EV}(s, \theta_0)$ and let U be a neighborhood of $(\mathrm{EV}_0, \theta_0)$. According to the implicit function theorem in Kolmogorov and Fomin (1989), $\mathrm{EV}(s, \theta)$ is differentiable in θ at θ_0 if the following conditions hold:

IFT1. $\Phi(v, \theta)$ is continuous at $(\mathrm{EV}_0, \theta_0)$.

IFT2. $\Phi(\mathrm{EV}_0, \theta_0) = O$.

IFT3. The Fréchet derivative $\partial \Phi(v, \theta)/\partial v$ exists in U.

IFT4. $\partial \Phi(v, \theta)/\partial v$ is continuous at $(\mathrm{EV}_0, \theta_0)$.

IFT5. The inverse of the operator $\partial \Phi(\mathrm{EV}_0, \theta_0)/\partial v$, denoted by $[\partial \Phi(\mathrm{EV}_0, \theta_0)/\partial v]^{-1}$, exists.

IFT6. $\partial \Phi(v, \theta)/\partial \theta$ exists in U.

IFT7. $\partial \Phi(v, \theta)/\partial \theta$ is continuous at $(\mathrm{EV}_0, \theta_0)$.

The following proposition describes an important implication of the differentiability of $\Phi(v, \theta)$ in v.

PROPOSITION 4. *Condition* IFT3 *implies Assumption* 5.

ASSUMPTION 5. *Let* $\hat{S}(v, \theta) = \{s' : \arg\max_{d' \in D}[u(s', d'; \theta) + \beta v(s', d')]$ *is single valued*}. *For any* $(v, \theta) \in U$ *and any* (s, d), $F(\hat{S}(v, \theta)|s, d; \theta) = 1$.

Assumption 5 is very strong. It is stronger than the requirement of the almost sure (a.s.) uniqueness of the policy function.[1] The a.s. uniqueness of the policy function has to hold even if some small perturbations are introduced into the expected value function in the Bellman equation. Proposition 4 is proved in the Appendix. The proof can be described as follows. When the arguments of max in the Emax operator Λ_d are equal on a set of states of positive probability (Assumption 5 fails), then even very small changes in v in different directions can lead to different maximizers inside Λ_d and, correspondingly, to quite different changes in values of Λ_d. The formal proof makes these ideas more precise.

---

[1] Almost sure uniqueness of the policy function by itself is an important property for estimable DDCMs since it is needed for choice probabilities to be well defined. Also, from the proof of Proposition 4, it appears to be necessary for the differentiability of the expected value functions. Whether its stronger version in Assumption 5 is necessary for differentiability and whether it can be established without states that do not affect transition probabilities are interesting questions for future research.

Proposition 5 below shows that under very reasonable conditions, Assumption 5 requires that some state variables not affect state transition probabilities and thus the expected value functions. The intuition behind this is that when all state variables do affect the expected value functions, it is possible to construct slightly perturbed versions of the expected value functions that make arg max in Assumption 5 multivalued on a set of positive probability, which violates Assumption 5.

PROPOSITION 5. *Suppose that $F(\cdot|s, d, \theta_0)$ and the Lebesgue measure on S are mutually absolutely continuous, and Assumption 5 holds. If the optimal decision at θ_0 is not the same for all states, then there must be at least one state variable that does not affect the transition probability and, as a result, the expected value function.*

Proposition 6 lists sufficient conditions for the differentiability of the expected value function w.r.t. parameters.

PROPOSITION 6. *If Assumption 5 holds, Θ and S are compact (w_m is a constant), $\nu(S) < \infty$, $u(s, d; \theta)$ and $\partial u(s, d; \theta)/\partial \theta_j$ are continuous in (s, θ), $f(s'|s, d; \theta)$ and $\partial f(s'|s, d; \theta)/\partial \theta_j$ are continuous in (s', s, θ), then* IFT1–IFT7 *hold and $E\{V(s'; \theta)|s, d; \theta\}$ is differentiable in θ on Θ.*

It might be possible to generalize this result to the unbounded state space case with methods similar to those used to establish the continuity results for the unbounded state space in Section 3. It is certainly possible to do so for some special cases (this was done in Rust (1988) for additive conditionally independent utility shocks).

Assumption 5 is the only nontrivial condition for the differentiability of the expected value functions. Let us try to find useful sufficient conditions for Assumption 5 to hold. From Proposition 5, we know that such conditions have to include the existence of state variables that do not affect the transition probabilities and the future expected value functions. For a given $d \in D$, let us denote such state variables by ε_d and denote the rest of the state variables by x_d, so that $s = (x_d, \varepsilon_d)$ and

$$f(s'|s, d; \theta) = f(s'|x_d, d; \theta) \tag{7}$$

for any d. This is more general than Rust's conditional independence assumption: $\varepsilon_d = \varepsilon$, $x_d = x$, and $f(s'|s, d; \theta) = f(\varepsilon'|x'; \theta)f(x'|x, d; \theta)$ for any d. Let ε_{d_1,d_2} denote the set of state variables that are contained in both ε_{d_1} and ε_{d_2}, and let x_{d_1,d_2} denote the rest; thus, $s = (\varepsilon_{d_1,d_2}, x_{d_1,d_2})$.

PROPOSITION 7. (i) *Assumption 5 holds if for any $d, d_1, d_2 \in D$, s, x'_{d_1,d_2}, $\theta \in \Theta$, and $c \in R$, the set of variables ε'_{d_1,d_2} is nonempty and*

$$\Pr[\varepsilon'_{d_1,d_2} : u(s', d_1; \theta) - u(s', d_2; \theta) = c|x'_{d_1,d_2}, s, d; \theta] = 0. \tag{8}$$

(ii) *Equation (8) is satisfied when the distribution of ε'_{d_1,d_2} conditional on $(x'_{d_1,d_2}, s, d, \theta)$ is absolutely continuous and equation $u(s', d_1; \theta) - u(s', d_2; \theta) = c$ defines one of the components of ε'_{d_1,d_2} as a continuous function of the other components.*

The following example illustrates that condition (8) is easy to verify in applications using Proposition 7(ii).

EXAMPLE 3. Consider a stylized model of work hours choice. Let us assume that there are three possible alternatives: full-time work, part-time work, and leisure. Define the per-period payoff of choosing alternative d by

$$u(s, d) = \exp\{\alpha_d \cdot s_0 + \gamma_{d,1} \cdot s_1 + \gamma_{d,2} \cdot s_2 + \gamma_{d,3} \cdot s_3\},$$

where vector s_0 can include age, experience, and other personal characteristics, and scalars s_d, $d \in \{1, 2, 3\}$, are stochastic shocks. In an estimable DDCM, s_0 can represent state variables observed by econometricians and the other shocks can represent unobserved state variables. The transition law for s_0 is not important for the results below and, therefore, is not specified here. Let $f(\cdot|\cdot)$ and $f_0(\cdot)$ be densities with respect to the Lebesgue measure that are positive on their compact support S.

Case 1 ($\gamma_{i,j} = 1_{\{i=j\}}$): In this case, s_d is an alternative specific utility shock. If s_d is i.i.d. $f_0(\cdot)$, then $\varepsilon_{d_1, d_2} = (s_1, s_2, s_3)$ and Proposition 7(ii) applies immediately.

Case 2 ($\gamma_{1,1} = \gamma_{3,3} = 1$, $\gamma_{2,1}, \gamma_{2,3} \in (0, 1)$, and $\gamma_{1,2} = \gamma_{1,3} = \gamma_{2,2} = \gamma_{3,1} = \gamma_{3,2} = 0$): Thus, s_2 is not present in the model and s_1 can be interpreted as a shock to the utility of work and s_3 can be interpreted as a shock to the utility of leisure. Assume that s_3' is i.i.d. $f_0(\cdot)$ and s_1 follows a first-order Markov process $s_1' \sim f(\cdot|s_1)$ for all d. In this case, $\varepsilon_{d_1, d_2} = s_3$ for any (d_1, d_2) and Proposition 7 applies, since for any (d_1, d_2), $u(s, d_1) - u(s, d_2) = c$ either never holds or holds for a unique value of s_3.

Case 3 ($\gamma_{1,3} = \gamma_{2,1} = \gamma_{3,2} = 0$ and $\gamma_{1,1} = \gamma_{1,2} = \gamma_{2,2} = \gamma_{2,3} = \gamma_{3,3} = \gamma_{3,1} = 1$): Assume that when alternative d is chosen, s_d follows a first-order Markov process $s_d' \sim f(\cdot|s_d)$, and when $d_1 \neq d$ is chosen, in the next period s_d is an i.i.d. draw $s_d' \sim f_0(\cdot)$. In this setup, $\varepsilon_{1,2} = \{s_3\}$, $\varepsilon_{1,3} = \{s_2\}$, and $\varepsilon_{2,3} = \{s_1\}$. Note that

$$u(s, 1) - u(s, 2) = \exp\{\alpha_1 \cdot s_0 + s_1 + s_2\} - \exp\{\alpha_2 \cdot s_0 + s_2 + s_3\} = c$$

for some c either cannot hold or implies

$$s_3 = \log[\exp\{\alpha_1 \cdot s_0 + s_1 + s_2\} - c] - (\alpha_2 \cdot s_0 + s_2),$$

which holds with probability 0. The other two cases, $u(s, 1) - u(s, 3) = c$ and $u(s, 2) - u(s, 3) = c$, are symmetric to this one. Thus, Proposition 7 applies.

As can be seen from the above example, Propositions 6 and 7 considerably generalize Rust's results for additive conditionally independent utility shocks. In particular, the utility shocks can be multiplicative, they can be serially correlated, and there might be fewer shocks than there are alternatives.

5. DISCUSSION

The continuity with respect to state variables and parameters, and the differentiability with respect to state variables of the expected value functions in DDCMs are established in this paper under fairly general conditions. The differentiability with respect

to parameters is established when some state variables do not affect the state transition probabilities and thus the expected value functions. The sufficient conditions of the implicit function theorem used in the proof are actually not satisfied unless such variables are present. It seems unlikely that the expected value functions are differentiable and the implicit function theorem does not apply (in general, it is of course possible). Perhaps if the expected value functions can be shown to belong to some special classes with more structure than the class of bounded continuous functions (for example, monotone expected value functions in dynamic binary choice models), then differentiability can be established without the implicit function theorem for the Banach space of bounded continuous functions. This conjecture is left for future research.

<div align="center">Appendix</div>

Proof Proposition 1. The proof closely follows the standard proof of the continuity of value functions with respect to the state variables (see, for example, Chapters 3 and 4 of Stokey and Lucas (1989)). Let us consider the Bellman operator Γ on the Banach space of bounded functions B with the sup norm: $V : S \times \Theta \to R$ defined in (3). Blackwell's sufficient conditions for contraction are satisfied for this operator, so Γ is a contraction mapping on B. The set of continuous functions C is a closed subset in B. Thus, it suffices to show that $\Gamma(C) \subset C$ (this trivially implies that the fixed point of Γ is a continuous function).

Let $V(s; \theta)$ be a continuous function in B ($V \in C$). Let us show that $\Gamma(V)$ is also continuous.

$$
\begin{aligned}
&|\Gamma(V)(s_1; \theta_1) - \Gamma(V)(s_2; \theta_2)| \\
&\quad \leq \max_d \Big| u(s_1, d; \theta_1) - u(s_2, d; \theta_2) \\
&\qquad + \beta \int V(s'; \theta_1) f(s'|s_1, d; \theta_1) \nu(ds') - \beta \int V(s'; \theta_2) f(s'|s_2, d; \theta_2) \nu(ds') \Big| \\
&\quad \leq \max_d |u(s_1, d; \theta_1) - u(s_2, d; \theta_2)| \\
&\qquad + \beta \max_d \Big| \int [V(s'; \theta_1) f(s'|s_1, d; \theta_1) - V(s'; \theta_2) f(s'|s_2, d; \theta_2)] \nu(ds') \Big|.
\end{aligned}
\tag{9}
$$

Given $\varepsilon > 0$, there exists $\delta_1 > 0$ such that $\|(s_1; \theta_1) - (s_2; \theta_2)\| < \delta_1$ implies $\max_d |u(s_1, d; \theta_1) - u(s_2, d; \theta_2)| < \varepsilon/2$:

$$
\begin{aligned}
&\Big| \int [V(s'; \theta_1) f(s'|s_1, d; \theta_1) - V(s'; \theta_2) f(s'|s_2, d; \theta_2)] \nu(ds') \Big| \\
&\quad \leq \max_d \sup_{s'} |V(s'; \theta_1) f(s'|s_1, d; \theta_1) - V(s'; \theta_2) f(s'|s_2, d; \theta_2)| \cdot \nu(S).
\end{aligned}
\tag{10}
$$

Since $V(s'; \theta)f(s'|s, d; \theta)$ is continuous on compact $S \times S \times \Theta$, for $\varepsilon > 0$ there exists $\delta_2^d > 0$ such that $\|(s_1, s'; \theta_1) - (s_2, s'; \theta_2)\| = \|(s_1; \theta_1) - (s_2; \theta_2)\| < \delta_2^d$ implies

$$\sup_{s'} |V(s'; \theta_1)f(s'|s_1, d; \theta_1) - V(s'; \theta_2)f(s'|s_2, d; \theta_2)| < \frac{\varepsilon}{2\nu(S)}.$$

Thus, for $\delta = \min\{\delta_1, \min_d \delta_2^d\}$, $\|(s_1; \theta_1) - (s_2; \theta_2)\| < \delta$ implies $|\Gamma(V)(s_1; \theta_1) - \Gamma(V)(s_2; \theta_2)| < \varepsilon$. So $\Gamma(V)$ is a continuous function. The continuity of $E\{V(s'; \theta)|s, d; \theta\}$ follows from the continuity of $V(s'; \theta)$ by an analogous argument. □

PROOF OF PROPOSITION 2. (i) Bhattacharya and Majumdar (1989, p. 376) showed that

$$R(s, \theta) \leq M2^{m-1} \left[(1 - \beta)^{-1} w(s, \theta)^m + b^m \sum_{k=0}^{\infty} k^m \beta^k \right] < \infty. \tag{11}$$

Therefore, part (iii) of Assumption 1 is satisfied. Consider $s^n \to s$ and

$$\int \phi(s')f(s'|s^n, d, \theta) \, d\nu(s') \tag{12}$$

for $|\phi(s)| \leq R(s, \theta)$. From (11), $|\phi(s)|$ is bounded by a linear function of $w(s, \theta)^m$. Therefore, Assumption 4 implies that the dominated convergence theorem is applicable to (12) and part (iv) of Assumption 1 follows.

(ii) For fixed θ, the result is established in Lippman (1975). See also Bhattacharya and Majumdar (1989, pp. 376–377). Since Lippman (1975) used different notation, the outline of the argument is presented below to rule out any confusion. For $V_1, V_2 \in B_{w^m}$,

$$|\Gamma(V_1) - \Gamma(V_2)(s; \theta)| \leq \beta \max_d \left| \int [V_1(s'; \theta) - V_2(s'; \theta)]f(s'|s, d; \theta)\nu(ds') \right| \tag{13}$$

$$\leq \beta \|V_1 - V_2\|_{w^m} \max_d \left| \int w(s'; \theta)^m f(s'|s, d; \theta)\nu(ds') \right| \tag{14}$$

$$\leq \beta \|V_1 - V_2\|_{w^m} (w(s, \theta) + b)^m. \tag{15}$$

The induction argument of Lippman (1975, pp. 1228–1229) applies from here on without any changes. It implies

$$\|\Gamma^J V_1 - \Gamma^J V_2\|_{w^m} \leq \beta^J [1 + Jb]^m \|V_1 - V_2\|_{w^m}.$$

We can choose J sufficiently large so that $\beta^J [1 + Jb]^m < 1$. The claim is proved.

(iii) The set of continuous functions $C \subset B_{w^m}$ is closed. Thus, it suffices to show that $\Gamma(C) \subset C$. Let $(s^n, \theta^n) \to (s, \theta)$. The claim follows if for a continuous $V \in C$,

$$\lim_{n \to \infty} \int V(s', \theta^n)f(s'|s^n, d, \theta^n) \, d\nu(s') = \int V(s', \theta)f(s'|s, d, \theta) \, d\nu(s').$$

First, note that by the continuity, $V(s', \theta^n)f(s'|s^n, d, \theta^n) \to V(s', \theta)f(s'|s, d, \theta)$. Second, from Assumption 4, for a sufficiently large N and all $n \geq N$,

$$|V(s', \theta^n)f(s'|s^n, d, \theta^n)| \leq \|V\|_{w^m} w(s', \theta^n)f(s'|s^n, d, \theta^n) \leq \overline{w}_{s,d,\theta}(s')\overline{f}_{s,d,\theta}(s').$$

Thus, the dominated convergence theorem applies. The continuity of $E\{V(s';\theta)|s,d;\theta\}$ follows from the continuity of $V(s';\theta)$ by an analogous argument.   □

PROOF OF PROPOSITION 3. Consider $s_j^n \to s_j$:

$$\frac{EV(s_{-j}, s_j^n, d; \theta) - EV(s, d; \theta)}{s_j^n - s_j}$$

$$= \int V(s';\theta) \frac{f(s'|s_{-j}, s_j^n, d; \theta) - f(s'|s, d; \theta)}{s_j^n - s_j} \nu(ds') \tag{16}$$

$$= \int V(s';\theta) \frac{\partial f(s'|s_{-j}, \tilde{s}_j^n(s'), d; \theta)}{\partial s_j} \nu(ds'),$$

where the last equality follows from the intermediate value theorem. For each s',

$$\frac{\partial f(s'|s_{-j}, \tilde{s}_j^n(s'), d; \theta)}{\partial s_j} \to \frac{\partial f(s'|s_{-j}, s_j, d; \theta)}{\partial s_j}.$$

Note that $|V(s';\theta)| \le R(s',\theta)$. As we discussed in the proof of Proposition 2, $R(s',\theta)$ is bounded by a linear function of $w(s',\theta)^m$. This fact and Assumption 4 for $\partial f(s'|s_{-j}, s_j, d; \theta)/\partial s_j$ imply that the absolute value of the integrand in (16) is bounded by an integrable function for all sufficiently large n. Thus, the existence of partial derivatives follows by the dominated convergence theorem. Essentially the same argument establishes continuity in (s,θ) of the partial derivatives, since $V(s;\theta)$ is continuous under Assumptions 2–4.   □

PROOF OF PROPOSITION 4. The proposition is proved by contradiction. For a bounded nonnegative function h, consider the right limit

$$\lim_{t \to +0} [\Lambda_d(v + th, \theta)(s) - \Lambda_d(v, \theta)(s)]/t$$

$$= \lim_{t \to +0} \beta \int \Big[\max_{d'} [u(s', d'; \theta) + \beta(v + th)(s', d')]$$

$$- \max_{d'} [u(s', d'; \theta) + \beta v(s', d')] \Big]/t F(ds'|s, d; \theta) \tag{17}$$

$$= \beta \int \max_{d' \in \arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta) + \beta v(s', \tilde{d})]} h(s', d') F(ds'|s, d; \theta),$$

where the last equality follows by the dominated convergence theorem. Similarly, the left limit is

$$\beta \int \min_{d' \in \arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta) + \beta v(s', \tilde{d})]} h(s', d') F(ds'|s, d; \theta).$$

If Assumption 5 does not hold (in other words, $\arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta) + \beta v(s', \tilde{d})]$ is multi-valued on a set of positive $F(\cdot|s, d; \theta)$ measure), then the right and the left limits will be

different for h satisfying the condition $h(s', d_1) > h(s', d_2)$ for any s' and $d_1 > d_2$. Therefore, the Gâteaux derivative $\partial\Lambda_d/\partial v$ does not exist. Since the existence of the Fréchet derivative implies the existence of the Gâteaux derivative, the Fréchet derivative of Λ_d w.r.t. v also does not exist. □

PROOF OF PROPOSITION 5. The proposition is proved by contradiction. We assume that for all $d \in D$, all of the state variables affect the transition probability and then show that under the proposition conditions, there exists an alternative that gives the decision-maker strictly higher lifetime utility than any other alternative at all state variables. For $\varepsilon > 0$ define

$$S_\varepsilon = \Big\{ s' : d_1 \in \arg\max_{\tilde{d}}[u(s', \tilde{d}, \theta_0) + \beta\,\mathrm{EV}_0(s', \tilde{d})], d_2 \neq d_1,$$

$$0 \le u(s', d_1, \theta_0) - u(s', d_2, \theta_0) + \beta(\mathrm{EV}_0(s', d_1) - \mathrm{EV}_0(s', d_2)) < \varepsilon \Big\}. \tag{18}$$

Suppose S_ε is nonempty for all $\varepsilon > 0$. Choose $\varepsilon > 0$ so that if $\|v - \mathrm{EV}_0\| < \varepsilon$, then $(v, \theta_0) \in U$. By continuity of u and EV_0, S_ε contains a ball with positive radius. This ball contains a point s^* with the unique optimal decision $d_1^* = \arg\max_{\tilde{d}}[u(s^*, \tilde{d}, \theta_0) + \beta\,\mathrm{EV}_0(s^*, \tilde{d})]$ because $F(\hat{S}(\mathrm{EV}_0, \theta_0)|s, d, \theta_0) = 1$ by Assumption 5. By continuity of u and EV_0, d_1^* is the unique optimal decision for any s' in a ball $B_\delta(s^*)$ with center s^* and sufficiently small radius $\delta > 0$. The radius δ can be chosen so that $B_\delta(s^*) \subset \hat{S}(\mathrm{EV}_0, \theta_0) \cap S_\varepsilon$ and for all $s' \in B_\delta(s^*)$, the condition in (18) holds for d_1^* and for some fixed d_2^*.

Define function v^* as $v^*(s', d) = \mathrm{EV}_0(s', d)$ if $d \neq d_2^*$; $v^*(s', d_2^*) = \mathrm{EV}_0(s', d_2^*)$ for $s' \notin B_\delta(s^*)$ and $v^*(s', d_2^*) = \beta^{-1}(u(s', d_1^*, \theta_0) - u(s', d_2^*, \theta_0)) + \beta\,\mathrm{EV}_0(s', d_1^*)$ for $s' \in B_{\delta/2}(s^*)$. For $s' \in B_\delta(s^*) \setminus B_{\delta/2}(s^*)$, choose $v^*(s', d_2^*)$ so that $v^*(s', d_2^*)$ is continuous and $\|\mathrm{EV}_0 - v^*\| < \varepsilon$, for example, by linear interpolation between points on the boundaries of $B_\delta(s^*)$ and $B_{\delta/2}(s^*)$ along the line going through s' and s^* as shown in Figure 1. This definition of v^* requires that the expected value function depend on all the state variables and that is where the corresponding assumption is used. This continuous v^* violates Assumption 5 since $\|\mathrm{EV}_0 - v^*\| < \varepsilon$ and $(v^*, \theta_0) \in U$, and $\arg\max_d[u(s', d, \theta_0) + \beta(v^*(s', d)]$ is not single valued on $B_{\delta/2}(s^*)$ and $F(B_{\delta/2}(s^*)|s, d, \theta_0) > 0$ as the Lebesgue measure is absolutely
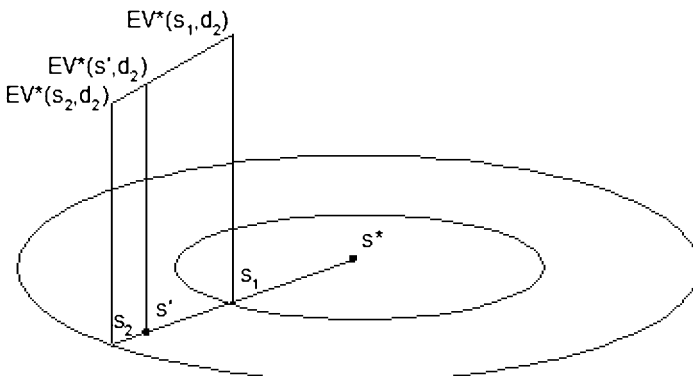

FIGURE 1. Linear interpolation for $v^*(s', d_2^*)$, $s' \in B_\delta(s^*) \setminus B_{\delta/2}(s^*)$.

continuous w.r.t. $F(\cdot|s, d, \theta_0)$. Therefore, under Assumption 5, there must exist $\varepsilon > 0$ such that S_ε is empty. Then the optimal decision must be unique for all $s' \in S$. If the optimal decision is unique for all states, then the continuity of u and EV_0 implies that the optimal decision is the same for all states.   □

PROOF OF PROPOSITION 6.

IFT1. It suffices to show that if $(v^n, \theta^n) \to (EV_0, \theta_0)$, then $\Lambda_d(v^n, \theta^n) \to \Lambda_d(EV_0, \theta_0)$ for any $d \in D$ and

$$
|\Lambda_d(v^n, \theta^n)(s) - \Lambda_d(EV_0, \theta_0)(s)|
$$

$$
= \left| \int \left[\max_{d'}\{u(s', d'; \theta^n) + \beta v^n(s', d')\} f(s'|s, d; \theta^n) \right. \right.
$$

$$
\left. \left. - \max_{d'}\{u(s', d'; \theta_0) + \beta EV_0(s', d')\} f(s'|s, d; \theta_0) \right] v(ds') \right|
$$

$$
\leq \int \left| \max_{d'}\{u(s', d'; \theta^n) + \beta v^n(s', d')\} - \max_{d'}\{u(s', d'; \theta_0) + \beta EV_0(s', d')\} \right|
$$

$$
\times f(s'|s, d; \theta^n) v(ds')
$$

$$
+ \int \left| \max_{d'}\{u(s', d'; \theta_0) + \beta EV_0(s', d')\} \right| |f(s'|s, d; \theta^n) - f(s'|s, d; \theta_0)| v(ds').
$$

Note that $|\max_{d'}\{u(s', d'; \theta^n) + \beta v^n(s', d')\} - \max_{d'}\{u(s', d'; \theta_0) + \beta EV_0(s', d')\}| \leq \max_{d'}|u(s', d'; \theta^n) - u(s', d'; \theta_0)| + \beta \max_{d'}|v^n(s', d') - EV_0(s', d')|$. Since $u(\cdot)$ and $f(\cdot|\cdot)$ are uniformly continuous, $\|v^n - EV_0\| \to 0$, and $\|u(\cdot) + \beta EV_0(\cdot)\| < \infty$, then $|\Lambda_d(v^n, \theta^n)(s) - \Lambda_d(EV_0, \theta_0)(s)|$ converges to zero uniformly in s.

IFT2. The condition holds by the formulation of the problem.

IFT3. If the Gâteaux derivative exists and is linear, bounded, and continuous, then the Fréchet derivative exists (and is equal to the Gâteaux derivative). First, let us show that following expression is equal to the Gâteaux derivative of Λ_d with respect to v at $(v, \theta) \in U$ evaluated at s:

$$
\frac{\partial \Lambda_d(v, \theta)(h)(s)}{\partial v} = \beta \int \left[\sum_{d'} 1\left\{ d' = \arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta) + \beta v(s', \tilde{d})] \right\} h(s', d'; \theta) \right]
$$

$$
\times f(s'|s, d; \theta) v(ds'),
$$

where $1\{d' = \arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta) + \beta v(s', \tilde{d})]\} = 0$ whenever the arg max is not single valued. It is the case if the following expression converges to zero uniformly in s:

$$
\frac{\Lambda_d(v + th, \theta)(s) - \Lambda_d(v, \theta)(s)}{t}
$$

$$
- \beta \int \left[\sum_{d'} 1\left\{ d' = \arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta) + \beta v(s', \tilde{d})] \right\} h(s', d'; \theta) \right]
$$

$$
\times f(s'|s, d; \theta) v(ds')
$$

$$= \int 1_{\hat{S}}(s') \Big(t^{-1} \Big[\max_{d'}[u(s', d'; \theta) + \beta(v + th)(s', d')] $$

$$- \max_{d'}[u(s', d'; \theta) + \beta v(s', d')] \Big]$$

$$- \beta \Big[\sum_{d'} 1 \Big\{ d' = \arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta) + \beta v(s', \tilde{d})] \Big\} h(s', d'; \theta) \Big] \Big)$$

$$\times f(s'|s, d; \theta)\nu(ds')$$

$$= \int B_t(s') f(s'|s, d; \theta)\nu(ds') \leq \int |B_t(s')|\overline{f}\nu(ds'),$$

where a finite upper bound $\overline{f}$ on $f(s'|s, d; \theta)$ exists since $f(\cdot|\cdot)$ is continuous on a compact. By Assumption 5, for all $s' \in \hat{S}$, $\arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta) + \beta v(s', \tilde{d})]$ is single valued and $B_t(s')$ becomes equal to zero for all sufficiently small t. The integral $\int |B_t(s')|\overline{f}\nu(ds')$ does not depend on s and converges to zero by the dominated convergence theorem ($|B_t(s')| \leq 2\|h\|$ and ν is assumed to be finite in the bounded case). The obtained Gâteaux derivative of Λ is linear in h. Its norm is equal to β. Linear bounded operators are continuous. Thus, the Fréchet derivative of Λ exists.

IFT4. Let $(\theta^n, v^n) \to (\theta_0, \text{EV}_0)$. We need to show that

$$\sup_{\|h\| \leq 1} \sup_s \left| \frac{\partial \Lambda_d(\theta^n, v^n)(h)(s)}{\partial v} - \frac{\partial \Lambda_d(\theta_0, \text{EV}_0)(h)(s)}{\partial v} \right| \to 0,$$

$$\left| \frac{\partial \Lambda_d(\theta^n, v^n)(h)(s)}{\partial v} - \frac{\partial \Lambda_d(\theta_0, \text{EV}_0)(h)(s)}{\partial v} \right|$$

$$\leq \int \Big[\sum_{d'} 1 \Big\{ d' = \arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta^n) + \beta v^n(s', \tilde{d})] \Big\} h(s', d') \Big]$$

$$\times (f(s'|s, d; \theta^n) - f(s'|s, d; \theta_0))\nu(ds')$$

$$+ \int 1_{\hat{S}}(s') \Big[\sum_{d'} 1 \Big\{ d' = \arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta^n) + \beta v^n(s', \tilde{d})] \Big\} h(s', d') \tag{19}$$

$$- \sum_{d'} 1 \Big\{ d' = \arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta_0) + \beta \text{EV}_0(s', \tilde{d})] \Big\} h(s', d') \Big]$$

$$\times f(s'|s, d; \theta_0)\nu(ds')$$

$$\leq \|h\| \int |f(s'|s, d; \theta^n) - f(s'|s, d; \theta_0)|\nu(ds') + \int B_n^1(s') f(s'|s, d; \theta)\nu(ds'),$$

where B_n^1 is the integrand in (19). By the same argument as in IFT3 for B_t and by the uniform continuity of $f(\cdot|\cdot)$, the last two integrals converge to zero uniformly in s and $\|h\| \leq 1$.

IFT5. Since

$$\left\| \frac{\partial \Lambda(\text{EV}_0, \theta_0)}{\partial v} \right\| = \beta < 1$$

and

$$\frac{\partial \Phi(\mathrm{EV}_0, \theta_0)}{\partial v} = I - \frac{\partial \Lambda(\mathrm{EV}_0, \theta_0)}{\partial v}$$

the inverse exists and is given by

$$\left[\frac{\partial \Phi(\mathrm{EV}_0, \theta_0)}{\partial v}\right]^{-1} = \sum_{t=0}^{\infty} \left[\frac{\partial \Lambda(\mathrm{EV}_0, \theta_0)}{\partial v}\right]^t.$$

IFT6. Let g be an element of the Banach space that includes Θ. By an argument similar to the one in IFT3, the following expression is the Fréchet derivative of Λ_d with respect to the parameter vector θ:

$$\frac{\partial \Lambda_d(v, \theta)(g)(s)}{\partial \theta}$$

$$= \int \max_{d'}[u(s', d'; \theta) + \beta v(s', d')] \cdot \left[\sum_j \frac{\partial f(s'|s, d; \theta)}{\partial \theta_j} g_j\right] v(ds')$$

$$+ \int \left(\sum_{d'} 1\{d' = \arg\max_{\tilde{d}}[u(s', \tilde{d}; \theta) + \beta v(s', \tilde{d})]\} \left[\sum_j \frac{\partial u(s', d'; \theta)}{\partial \theta_j} g_j\right]\right)$$

$$\times f(s'|s, d; \theta) v(ds').$$

In this argument, interchanging the differentiation and integration operations is also justified by the dominated convergence theorem, since the partial derivatives of $u(s', d'; \theta)$ and $f(s'|s, d; \theta)$ are assumed to be continuous on a compact and thus are bounded.

IFT7. The continuity of $\partial \Lambda_d(v, \theta)(g)(s)/\partial \theta$ in (v, θ) is established by the same argument as the one in IFT4.   □

PROOF OF PROPOSITION 7.   (i) Define operator Λ on the space of vector functions whose d_i component does not depend on ε'_{d_i}. Since x'_{d_1, d_2} includes x'_{d_1} and x'_{d_2}, we can set $c = \beta[v(x'_{d_2}, d_2) - v(x'_{d_1}, d_1)]$. Then (8) immediately implies Assumption 5.

Part (ii) of the proposition holds because the Lebesgue measure of a graph of a continuous function is zero.   □

REFERENCES

Ackerberg, D. (2009), "A new use of importance sampling to reduce computational burden in simulation estimation." *Quantitative Marketing and Economics*, 7, 343–376. [305]

Aguirregabiria, V. and P. Mira (2002), "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models." *Econometrica*, 70, 1519–1543. [305]

Aguirregabiria, V. and P. Mira (2007), "Dynamic discrete choice structural models: A survey." Working Paper 297, University of Toronto. [305]

Bajari, P., L. Benkard, and J. Levin (2007), "Estimating dynamic models of imperfect competition." *Econometrica*, 75, 1331–1370. [305]

Benveniste, L. M. and J. A. Scheinkman (1979), "On the differentiability of the value function in dynamic models of economics." *Econometrica*, 47, 727–732. [306]

Bhattacharya, R. and M. Majumdar (1989), "Controlled semi-Markov models—The discounted case." *Journal of Statistical Planning and Inference*, 21, 365–381. [307, 308, 315]

Eckstein, Z. and K. Wolpin (1989), "The specification and estimation of dynamic stochastic discrete choice models: A survey." *Journal of Human Resources*, 24, 562–598. [305]

Hansen, L. P. (1982), "Large sample properties of generalized method of moments estimators." *Econometrica*, 50, 1029–1054. [306]

Hotz, J. and R. Miller (1993), "Conditional choice probabilities and the estimation of dynamic models." *Review of Economic Studies*, 60, 497–530. [305]

Imai, S., N. Jain, and A. Ching (2009), "Bayesian estimation of dynamic discrete choice models." *Econometrica*, 77, 1865–1899. [305]

Keane, M. and K. Wolpin (1994), "The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence." *Review of Economics and Statistics*, 76, 648–672. [305]

Kolmogorov, A. and C. Fomin (1989), *Elements of the Theory of Functions and Functional Analysis*. Nauka, Moskva. [311]

Lippman, S. (1975), "On dynamic programming with unbounded rewards." *Management Science*, 21 (11), 1225–1233. [308, 315]

Miller, R. (1984), "Job matching and occupational choice." *Journal of Political Economy*, 92 (6), 1086–1120. [305]

Norets, A. (2009), "Inference in dynamic discrete choice models with serially correlated unobserved state variables." *Econometrica*, 77, 1665–1682. [305, 306]

Norets, A. (forthcoming), "Estimation of dynamic discrete choice models using artificial neural networks approximations." *Econometric Reviews*. [305]

Pakes, A. (1986), "Patents as options: Some estimates of the value of holding European patent stocks." *Econometrica*, 54, 755–784. [305]

Rust, J. (1987), "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher." *Econometrica*, 55, 999–1033. [305, 307]

Rust, J. (1988), "Maximum likelihood estimation of discrete control processes." *SIAM Journal on Control and Optimization*, 26, 1006–1023. [306, 307, 310, 312]

Rust, J. (1994), "Structural estimation of Markov decision processes." In *Handbook of Econometrics* (R. Engle and D. McFadden, eds.), North-Holland, Amsterdam. [305, 306]

Stokey, N. and R. Lucas, with E. Prescott (1989), *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge, Massachusetts. [306, 314]

Su, C.-L. and K. L. Judd (2008), "Constrained optimization approaches to estimation of structural models." Working paper, SSRN. [305]

van der Vaart, A. (1998), *Asymptotic Statistics*. Cambridge University Press, Cambridge. [306]

Wolpin, K. I. (1984), "An estimable dynamic stochastic model of fertility and child mortality." *Journal of Political Economy*, 92, 852–874. [305]