SEMIPARAMETRIC IDENTIFICATION OF DYNAMIC MULTINOMIAL CHOICE MODELS

Andriy Norets

This paper considers dynamic discrete choice models with conditionally independent and additively separable unobserved state variables as in Rust (1987) and Hotz and Miller (1993). Previous literature commonly assumed specific parametric distributions for the unobserved states, for example, extreme value distribution. Norets and Tang (2010) proposed an approach to identification and inference in dynamic binary choice models that does not impose distributional assumptions on the unobserved state variables. This paper generalizes this approach to dynamic multinomial choice models.

Keywords: Dynamic discrete choice models, Markov decision processes, semiparametric inference, identification, Bayesian estimation, MCMC.

1. INTRODUCTION

This paper considers dynamic discrete choice models with conditionally independent and additively separable unobserved states as in Rust (1987) and Hotz and Miller (1993). The observed states are assumed to be discrete. The per-period payoffs can be specified parametrically or non-parametrically with optional shape restrictions such as monotonicity or concavity. Norets and Tang (2010) introduced an approach for semiparametric inference in such models when there are only two choice alternatives. Previous literature commonly assumed specific parametric distributions for the state variables unobserved by the econometrician, for example, extreme value distribution. Norets and Tang (2010) do not impose distributional assumptions on the unobserved state variables. This paper generalizes this approach to dynamic multinomial choice models. The reader is referred to Norets and Tang (2010) for a more detailed discussion

[1]Current version: May 4, 2011.

[2]I am grateful to Satoru Takahashi and Xun Tang for helpful discussions.

of how this work is related to previous literature. See also Eckstein and Wolpin (1989), Rust (1994), Pakes (1994), Aguirregabiria and Mira (2007a) and Keane et al. (2010) for surveys of the literature.

The organization of the paper is as follows. Section 2 describes the model and assumptions. Section 3 provides a characterization of the conditional choice probabilities when the distribution of the unobserved state variables is known. Section 4 extends this characterization to the case when the distribution of the unobserved states is unknown. These results are then used for characterizing the identified sets for parameters of per-period utilities and counterfactual choice probabilities. An algorithm for computing the identified sets based on Markov chain Monte Carlo (MCMC) methods is briefly discussed. See Norets and Tang (2010) for applications of the methodology to several dynamic binary choice examples. Dynamic multinomial choice applications are a subject of future work.

2. MODEL SETUP

In an infinite horizon dynamic discrete choice model, the agent maximizes the expected discounted sum of the per-period utilities

$$V(x_t, \epsilon_t) = \max_{d_t, \cdots, d_T} E_t\left(\sum_{j=0}^{T} \beta^j u(x_{t+j}, \epsilon_{t+j}, d_{t+j})\right),$$

where $T = \infty$, $d_t \in D = \{0, 1, \ldots, J\}$ is the control variable, $x_t \in X$ are state variables observed by the econometrician, $\epsilon_t = (\epsilon_{0t}, \ldots, \epsilon_{Jt}) \in \mathbb{R}^{J+1}$ are state variables unobserved by the econometrician, β is the time discount factor, and $u(x_t, \epsilon_t, d_t)$ is the time-invariant per-period utility function. The state variables evolve according to a controlled first-order Markov process. Under mild regularity conditions (see Bhattacharya and Majumdar (1989)), the optimal lifetime utility of the agent has a recursive representation:

$$V(x_t, \epsilon_t) = \max_{d_t \in D}[u(x_t, \epsilon_t, d_t) + \beta E\{V(x_{t+1}, \epsilon_{t+1}) | x_t, \epsilon_t, d_t\}] \tag{1}$$

The following assumptions are standard in the literature.

ASSUMPTION 1 *The state space for the observed states is finite and denoted by* $X = \{1, \ldots, K\}$.

ASSUMPTION 2 *The per-period utility is* $u(x_t, \epsilon_t, d_t) = u_{ji} + \epsilon_j$ *when* $(x_t, d_t) = (i, j)$, *with* $E(\epsilon_j|x) = 0$ *for any* $x \in X$ *and* $j \in D$.

ASSUMPTION 3 $Pr(x_{t+1} = i|x_t = k, \epsilon_t, d_t = j) = G^j_{ki}$ *is independent of* ϵ_t. *The distribution of* ϵ_{t+1} *given* $(x_{t+1}, x_t, \epsilon_t, d_t)$ *depends only on* x_{t+1} *and is denoted by* $F_{\epsilon|x}$.

ASSUMPTION 4 *The distribution of* $(\epsilon_0, \ldots, \epsilon_J)$ *given any* x *has a positive density on* $\mathbb{R}^{J+1}$ *with respect to (w.r.t.) the Lebesgue measure.*

Assumption 3 of conditional independence is, perhaps, the strongest one. However, it seems hard to avoid. Without Assumption 3 it is not clear whether the model can explain any possible choice pattern in the data (Rust (1994)) and whether the expected value functions are differentiable with respect to parameters (Norets (2010)). The assumption is also very convenient for computationally feasible classical (Rust (1994), Hotz and Miller (1993)) and Bayesian (Norets (2009)) estimation of parametrically specified models.

3. CHARACTERIZATION OF THE CCPS WHEN THE UNOBSERVED STATE DISTRIBUTION IS KNOWN

This section characterizes the CCPs assuming that the unobserved state distribution is known to econometricians. Under Assumptions 1-3, the Bellman equation (1) can be rewritten in vector notation as follows,

$$v_j = u_j + \beta G^j \int \max\{v_0 + \epsilon_0, \ldots, v_J + \epsilon_J\} dF_{\epsilon|x}(\epsilon|X) \tag{2}$$

where $u_j = (u_{j1}, \ldots, u_{jK})'$ is a vector of stacked deterministic parts of the per-period utilities with $d_t = j$; $v_j = (v_{j1}, \ldots, v_{jK})'$ is a vector of stacked deterministic parts of the alternative specific life time utilities $v_{ji} = u_{ji} + \beta E\{V(x_{t+1}, \epsilon_{t+1})|x_t = i, d_t = j\}$; and $G^j = [G^j_{ki}]$ is the Markov transition matrix for the observed states conditional on $d_t = j$. I also adopt a Matlab convention to simplify notation: for scalar ϵ_j and vector v_j, $v_j + \epsilon_j = (v_{j1} + \epsilon_j, \ldots, v_{jK} + \epsilon_j)'$, for a scalar function/expression $f(x)$ with a scalar argument x, $f(x_1, \ldots, x_K) = (f(x_1), \ldots, f(x_K))'$, and $\int \max\{v_0 + \epsilon_0, v_1 + \epsilon_1\} dF_{\epsilon|x}(\epsilon|X)$ stands for a $K \times 1$ vector of integrals correspondingly.

Let $\Delta\epsilon_j = \epsilon_0 - \epsilon_j$, $\delta_j = v_j - v_0$, and

$$R(\delta) = \int \max\{0, \delta_1 - \Delta\epsilon_1, \ldots, \delta_J - \Delta\epsilon_J\} dF_{\epsilon|x}(\epsilon|X).$$

Then (2) can be rewritten as follows

$$v_0 = (I - \beta G^0)^{-1}[u_0 + \beta G^0 R(\delta)]$$
$$\delta_j = u_j - u_0 + \beta(G^j - G^0)(v_0 + R(\delta))$$

Note that $I + (I - \beta G^0)^{-1}\beta G^0 = (I - \beta G^0)^{-1}$. Then, for $j = 1, \ldots, J$,

$$\delta_j = u_j - u_0 + \beta(G^j - G^0)(I - \beta G^0)^{-1}[u_0 + R(\delta)]. \tag{3}$$

Let us define a collection of conditional choice probabilities (CCP), $p = \{p_{jk},\ j = 1, \ldots, J,\ k = 1, \ldots, K\} = \mathcal{F}(\delta)$, as follows

$$p_{jk} = \Pr(E_{jk}|x = k) = \int_{E_{jk}} dF_{\epsilon|x}(\epsilon|x = k),\ j = 0, 1, \ldots, J,\ k = 1, \ldots, K,$$

where

$$E_{jk} = \{(\Delta\epsilon_{1k}, \ldots, \Delta\epsilon_{Jk}) : \delta_{jk} + \epsilon_{jk} \geq \delta_{ik} + \epsilon_{ik},\ i = 0, 1, \ldots, J\},$$

is a set of unobserved states on which alternative j is optimal. It follows from this definition that under known $F_{\epsilon|x}$, CCP p is completely specified by $\delta = \{\delta_{jk},\ j = 1, \ldots, J,\ k = 1, \ldots, K\} \in R^{JK}$. Thus, we can define a mapping $\mathcal{F} : R^{JK} \to [0, 1]^{JK}$ such that $p = \mathcal{F}(\delta)$ ($\delta_{0k} = 0$ and $p_{0k} = 1 - \sum_{j=1}^J p_{jk}$ are not included in δ and

p). Hotz and Miller (1993) (Proposition 1) show that under Assumptions 1 - 4 there exists an inverse of $\mathcal{F}$, $\mathcal{F}^{-1}$. Thus, the following lemma, which gives the necessary and sufficient conditions for some p to be the CCP given structural parameters, follows immediately.

LEMMA 1 *A $J \times K$ matrix p is the collection of CCPs for structural parameters $u = (u_0, \dots, u_J)$, $G = (G^0, \dots, G^J)$, β, and $F_{\Delta\epsilon|x}$ if and only if p is in the image of $\mathcal{F}$, $\mathcal{F}(R^{JK})$, and $\delta = \mathcal{F}^{-1}(p)$ satisfies (3).*

Analogs of Lemma 1 for dynamic discrete choice games were obtained by Aguirregabiria and Mira (2007b) and Pesendorfer and Schmidt-Dengler (2008) .

4. CHARACTERIZATION OF THE CCPS WHEN THE UNOBSERVED STATE DISTRIBUTION IS UNKNOWN

It is convenient to rewrite $R(\delta)$ in (3) as follows.

$$R(\delta) = \sum_{j=1}^{J} (p_j \circ \delta_j - e_j)$$

where "$\circ$" is element-by-element vector multiplication and k^{th} component of e_j is defined as

$$e_{jk} = \int_{E_{jk}} \Delta\epsilon_{jk} dF_{\epsilon|x}(\epsilon|k)$$

Let $e = \{e_{jk},\ j = 1, \dots, J,\ k = 1, \dots, K\} \in R^{JK}$. Note that Norets and Tang (2010) used symbol e for the negative of the definition here.

From Lemma 1, a characterization of CCP depends on $F_{\epsilon|x}$ only through a finite number of variables (δ, e). Below I establish conditions on δ and e so that p, δ, and e correspond to some distribution $F_{\epsilon|x}$. First, I will establish such conditions on $(p_{jk}, \delta_{jk}, e_{jk},\ j = 0, 1, \dots, J,\ k = 1, \dots, L)$ when $F_{\epsilon|x}(\cdot|k)$ does not depend on k for $k = 1, \dots, L$ $(F_{\epsilon|x}(\cdot|k) = F(\cdot))$. These conditions will be useful for a characterization of counterfactual CCPs under the assumption of independence ($L = 2K$) and possible dependence ($L = 2$) between x and ϵ.

The starting point of the argument is presented in the following lemma, which suggests how $p_{jk} = \int_{E_{jk}} dF(s)$ and $e_{jk} = \int_{E_{jk}} s_j dF(s_1, \ldots, s_J)$ have to be related to each other.

LEMMA 2 *Let E be a non-empty closed convex subset of $\mathcal{R}^J$ and F be a probability measure on E ($F(E) = 1$). Suppose $\int_E s dF = \left(\int_E s_1 dF, \ldots, \int_E s_J dF\right)'$ exists. Then, $\int_E s dF \in E$. If F has a positive on E density, then $\int_E s dF$ is an interior point of E. Conversely, if s^* is an interior point of a closed convex set E, then there exists a probability measure F on E that has positive density on E and $s^* = \int_E s dF$.*

PROOF: By the supporting hyperplane theorem (Proposition 2.4.4 in Bertsekas (2003)), $E = \cap_{(c,\alpha)\in\Gamma}\{s \in \mathcal{R}^J : \alpha' s \leq c\}$ for some set $\Gamma \subset \mathcal{R}^{J+1}$. For any $(c, \alpha) \in \Gamma$,

$$\alpha' \int_E s dF = \int_E \alpha' s dF \leq c,$$

which means $\int_E s dF \in E$. When F has a density, $F(s : \alpha' s = c) = 0$. Thus, the inequality in the above expression is strict for all $(c, \alpha) \in \Gamma$ and $\int_E s dF$ is an interior point.

To prove the converse, define an arbitrary distribution on E, F_1, that has a positive density. Let $s_1 = \int_E s dF_1$. Since s^* is interior there exists s_2 and $\gamma \in (0, 1)$ such that $s^* = \gamma s_1 + (1 - \gamma)s_2$ and a ball with a positive radius and center at s_2, $B_2 \subset E$. Let F_2 be a distribution with a density that puts probability 1 on B_2 and $\int s dF_2 = s_2$ (for example, a multivariate normal with a diagonal covariance matrix and center s_2 truncated to B_2). Then, the mixture $F = \gamma F_1 + (1 - \gamma)F_2$ satisfies the conditions of the converse.

Q.E.D.

Note that E_{jk} are closed and convex. When $L = 1$ they form a partition of $\mathcal{R}^J$ (up to boundaries) and Lemma 2 can be directly applied to E_{jk} for characterizing the relationship between p, δ, and e. For general L, E_{jk}'s can overlap and they need to be divided into smaller non-overlapping pieces first. Thus, consider the following collection of subsets of $\mathcal{R}^J$

$$\mathcal{A}(\delta) = \Big\{A: \ A = \bigcap_{k=1}^{L} E_{j_k k}, \ j_k \in \{0,\dots,J\}, \ \lambda(A) > 0\Big\} = \{A_1,\dots,A_{N_\delta}\},$$

where λ is the Lebesgue measure and N_δ is the number of sets in $\mathcal{A}(\delta)$. Since E_{jk} are determined by δ, $\mathcal{A}(\delta)$ is also determined by δ. One can interpret $\mathcal{A}(\delta)$ as a refinement of L partitions $(E_{0k}, E_{1k},\dots,E_{Jk})$ $k = 1,\dots,L$. Note that since A_n is an intersection of convex closed sets it is also closed and convex. Figure 1 provides an illustration for $J = 2$, $L = 2$, and $N_\delta = 6$.

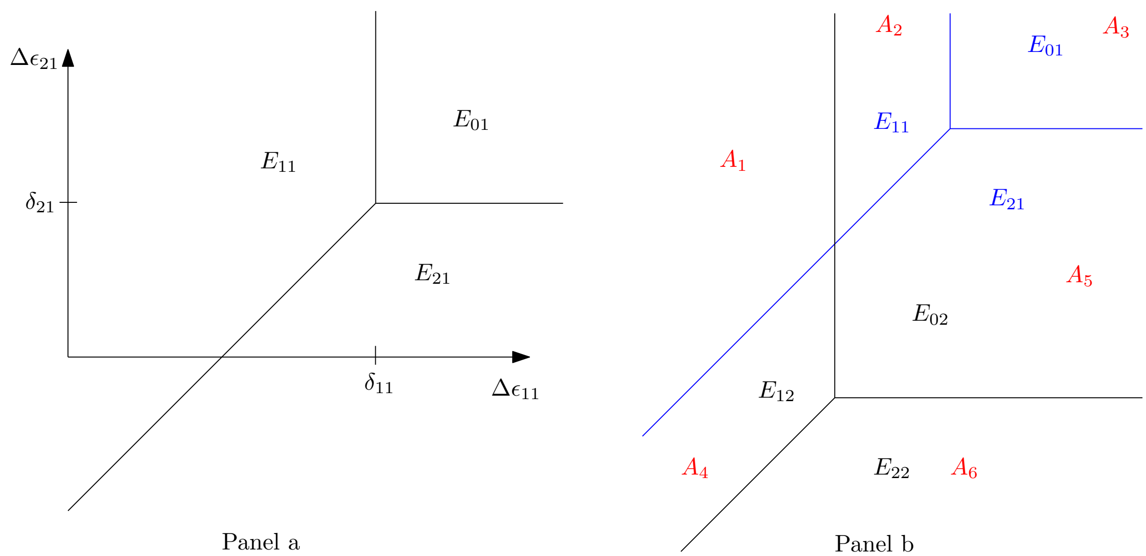

FIGURE 1.— Panel a shows sets E_{jk} for fixed $k = 1$ and $j = 0, 1, \dots, J$ ($J = 2$). Panel b shows E_{jk}, $k = 1, 2$, $J = 2$ and corresponding $\mathcal{A}(\delta) = \{A_1,\dots,A_6\}$.

For every $n = 1,\dots,N_\delta$ let

$$q_n = \int_{A_n} dF \quad \text{and} \quad r_n = \int_{A_n} y dF(y) = \left(\int_{A_n} y_1 dF(y),\dots,\int_{A_n} y_J dF(y)\right)'. \tag{4}$$

Lemma 2 implies that (4) can hold for given $(q_n, r_n,\ n = 1, \dots, N_\delta)$ and some probability distribution F with positive density if and only if $r_n/q_n \in \text{int}(A_n)$ ($\text{int}(A_n)$ stands for the interior of A_n). From this, the following lemma follows immediately.

Lemma 3 *Given $(p_{jk}, \delta_{jk}, e_{jk},\ j = 0, 1, \dots, J,\ k = 1, \dots, L)$, where $\sum_{j=0}^{J} p_{jk} = 1$ and $\delta_{0k} = 0$ for all k, there exists F such that*

(5) F *has a density* $f > 0$ *on* $\mathcal{R}^J$ *w.r.t. the Lebesgue measure,*

$$\int s dF = 0 \tag{6}$$

$$p_{jk} = \int_{E_{jk}} dF,\ \forall (j,k) \tag{7}$$

$$e_{jk} = \int_{E_{jk}} s_j dF(s),\ \forall (j,k) \tag{8}$$

if and only if for $\mathcal{A}(\delta) = \{A_1, \dots, A_{N_\delta}\}$ there exist $(r_1, q_1, \dots, r_{N_\delta}, q_{N_\delta})$ such that

$$q_n \in (0,1),\ \forall n \tag{9}$$

$$r_n/q_n \in int(A_n),\ \forall n \tag{10}$$

$$\sum_{n=1}^{N_\delta} r_{nj} = 0,\ j = 1, \dots, J \tag{11}$$

$$p_{jk} = \sum_{n:\ \lambda(A_n \cap E_{jk}) \neq 0} q_n,\ \forall (j,k) \tag{12}$$

$$e_{jk} = \sum_{n:\ \lambda(A_n \cap E_{jk}) \neq 0} r_{nj},\ \forall (j,k) \tag{13}$$

First, consider the case when the distribution of ϵ does not depend on x. Let us denote this distribution by F. The following proposition characterizes the CCPs that are consistent with the model and some F.

PROPOSITION 1 *For given* (β, u, G)*,* $(p_{jk},\ j = 0, 1, \ldots, J,\ k = 1, \ldots, L)$*,* $\sum_j p_{jk} = 1$ *for all* k*, are CCPs for some* F *satisfying* (5)-(6) *if and only if there exist* $(e_{jk}, \delta_{jk}, j = 0, \ldots, J,\ k = 1, \ldots, K)$ *such that*
(i) for every $j = 1, \ldots, J$*,* (p, δ, e) *satisfy*

$$\delta_j = u_j - u_0 + \beta(G^j - G^0)(I - \beta G^0)^{-1}[u_0 + \sum_{j=1}^{J}(p_j \circ \delta_j - e_j)]$$

(ii) conditions (9)-(13) *hold for* $(p, e, \delta, \mathcal{A}(\delta))$*.*

The proposition is implied by Lemmas 1 and 3. Proposition 1 can be used for characterizing the identified set for the parameters of the per-period utility functions. Suppose the utility functions are parameterized: $u(\theta)$. Then, for given (p, G, β) the identified set for θ includes values for which p are the CCPs for $(\beta, u(\theta), G)$ and some F (the proposition describes a way to verify this).
Structural models are particularly useful in analysis of counterfactual changes in structural parameters. Suppose we are interested in the CCPs, $\tilde{p}$, when the per-period payoffs and the observed state transition probabilities are changed to $\tilde{u}$ and $\tilde{G}$ while F remains the same. The counterfactual transition probabilities and per-period payoffs may either be assigned known numerical values or be known functions of the primitives G and u in the actual (data-generating) environment. Let p denote the actual CCPs corresponding to (u, G).

PROPOSITION 2 *For given* $(\beta, u, G, \tilde{u}, \tilde{G})$*,* p *and* $\tilde{p}$ *are actual and counterfactual CCPs for some* F *satisfying* (5)-(6) *if and only if there exist* $(e, \delta, \tilde{e}, \tilde{\delta})$ *such that*
(i) (p, δ, e) *satisfy condition (i) in Proposition 1,*

(ii) $(\tilde{p}, \tilde{\delta}, \tilde{e})$ *satisfy condition (i) in Proposition 1,*

(iii) conditions (9)-(13) *of Lemma 3 hold for stacked* $((p, \tilde{p}), (e, \tilde{e}), (\delta, \tilde{\delta}), \mathcal{A}(\delta, \tilde{\delta}))$ *(L=2K).*

Proposition 2 provides a characterization of the identified set for the counterfactual CCPs when the distribution of the unobserved states does not depend on the observed states. The following proposition provides a similar characterization of the identified set when the unobserved state distribution can depend on the observed state.

Proposition 3 *For given* $(\beta, u, G, \tilde{u}, \tilde{G})$, p *and* $\tilde{p}$ *are actual and counterfactual CCPs for some* $F(\cdot|k)$, $k = 1, \ldots, K$, *satisfying* (5)-(6) *if and only if there exist* $(e, \delta, \tilde{e}, \tilde{\delta})$ *such that*

(i) (p, δ, e) *satisfy condition (i) in Proposition 1,*

(ii) $(\tilde{p}, \tilde{\delta}, \tilde{e})$ *satisfy condition (i) in Proposition 1,*

(iii) for each $k \in \{1, \ldots, K\}$ *conditions* (9)-(13) *of Lemma 3 hold for actual and counterfactual variables corresponding to k:* $(p_{jk}, e_{jk}, \delta_{jk}, \tilde{p}_{jk}, \tilde{e}_{jk}, \tilde{\delta}_{jk},\ j = 0, \ldots, J)$ *and* $\mathcal{A}(\delta_{jk}, \tilde{\delta}_{jk},\ j = 0, \ldots, J)$, *(*$L = 2$*).*

The only difference between Propositions 2 and 3 is that in the latter the distributions of the unobserved states can be different for each value of the observed state k and thus the conditions of Lemma 3 need to be verified separately for every k.

To exploit Propositions 1, 2, or 3 for computing identified sets one needs to be able to verify the feasibility of systems of equalities and inequalities (9) - (13). Note that except (10), (9) - (13) are linear in (e, r, q, δ) and (10) can be made linear in r but would contain cross products $\delta_{jk} q_n$ (sets E_{jk} and thus A_n are defined by inequalities linear in δ_{jk}). In binary choice case ($J = 1$), one can express q_n as a function of p only as shown in Norets and Tang (2010). In that case all the conditions in Lemma 3 are linear and linear programming methods can be used for verifying them. For general J, q cannot be expressed as functions of p and one has to deal with cross-products $\delta_{jk} q_n$. Checking feasibility of non-convex quadratic system is a very hard problem

(NP-hard). The task of solving the quadratic problem can be avoided if we invoke MCMC.

For example, consider the problem of recovering the identified sets for the counterfactual CCPs. First, we specify a very flexible "prior" distribution for $\tilde{p}, \delta, \tilde{\delta}$. Experiments in Norets and Tang (2010) demonstrate that a mixture of beta distributions works well in dynamic binary choice models as a hierarchical prior for $\tilde{p}$. In multinomial models, one could use a mixture of Dirichlet distributions in a similar fashion. The prior for $\tilde{\delta}$ conditional on $\tilde{p}$ can be a multivariate normal centered at values implied by CCPs $\tilde{p}$ under extreme value distributed unobserved states. Additional flexibility can be obtained by specifying a prior on the variance-covariance matrix of the normal distribution. A prior for δ can specified similarly. Such a prior for $(\tilde{p}, \delta, \tilde{\delta})$ should be truncated to satisfy the conditions from Proposition 2. An MCMC algorithm producing draws from this truncated prior can be used for exploring the identified set. To produce the draws one needs to verify the truncation restrictions. Conditional on $(\delta, \tilde{\delta})$ the restrictions are equivalent to verifying feasibility of a system of linear equalities and inequalities (11) - (13). This can be done by standard linear programming algorithms as described in Appendix A of Norets and Tang (2010). Analogous algorithm is shown to work for recovering the identified set for the counterfactual CCPs in Norets and Tang (2010) for binary case. Applications of the methodology to multinomial case are a subject of future work.

REFERENCES

AGUIRREGABIRIA, V. AND P. MIRA (2007a): "Dynamic Discrete Choice Structural Models: A Survey," University of Toronto, Working Paper 297.

——— (2007b): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1–53.

BERTSEKAS, D. (2003): *Convex Analysis and Optimization*, Athena Scientific, Belmont, Mass., USA.

BHATTACHARYA, R. AND M. MAJUMDAR (1989): "Controlled Semi-Markov Models - The Discounted Case," *Journal of Statistical Planning and Inference*, 365–381.

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.

Hotz, J. and R. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–530.

Keane, M., P. Todd, and K. Wolpin (2010): "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications," Unpublished manuscript, University of Pennsylvania.

Norets, A. (2009): "Inference in Dynamic Discrete Choice Models With Serially Correlated Unobserved State Variables," *Econometrica*, 77, 1665–1682.

——— (2010): "Continuity and Differentiability of Expected Value Functions in Dynamic Discrete Choice Models," *Quantitative economics*, 1.

Norets, A. and X. Tang (2010): "Semiparametric Inference in Dynamic Binary Choice Models," Unpublished manuscript, Princeton and University of Pennsylvania.

Pakes, A. (1994): "The Estimation of Dynamic Structural Models: Problems and Prospects," in *Advances in Econometrics: Proceedings of the 6th World Congress of the Econometric Society*, ed. by J. Lafont and C. Sims.

Pesendorfer, M. and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75, 901–928.

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

——— (1994): "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics*, ed. by R. Engle and D. McFadden, North Holland.