Econometrica, Vol. 77, No. 1 (January, 2009), 135–175

NONPARAMETRIC IDENTIFICATION OF FINITE MIXTURE MODELS OF DYNAMIC DISCRETE CHOICES

BY HIROYUKI KASAHARA AND KATSUMI SHIMOTSU[1]

In dynamic discrete choice analysis, controlling for unobserved heterogeneity is an important issue, and finite mixture models provide flexible ways to account for it. This paper studies nonparametric identifiability of type probabilities and type-specific component distributions in finite mixture models of dynamic discrete choices. We derive sufficient conditions for nonparametric identification for various finite mixture models of dynamic discrete choices used in applied work under different assumptions on the Markov property, stationarity, and type-invariance in the transition process. Three elements emerge as the important determinants of identification: the time-dimension of panel data, the number of values the covariates can take, and the heterogeneity of the response of different types to changes in the covariates. For example, in a simple case where the transition function is type-invariant, a time-dimension of $T = 3$ is sufficient for identification, provided that the number of values the covariates can take is no smaller than the number of types and that the changes in the covariates induce sufficiently heterogeneous variations in the choice probabilities across types. Identification is achieved even when state dependence is present if a model is stationary first-order Markovian and the panel has a moderate time-dimension ($T \geq 6$).

KEYWORDS: Dynamic discrete choice models, finite mixture, nonparametric identification, panel data, unobserved heterogeneity.

1. INTRODUCTION

IN DYNAMIC DISCRETE CHOICE ANALYSIS, controlling for unobserved heterogeneity is an important issue. Finite mixture models, which are commonly used in empirical analyses, provide flexible ways to account for it. To date, however, the conditions under which finite mixture dynamic discrete choice models are nonparametrically identified are not well understood. This paper studies nonparametric identifiability of finite mixture models of dynamic discrete choices when a researcher has access to panel data.

Finite mixtures have been used in numerous applications, especially in estimating dynamic models. In empirical industrial organization, Crawford and Shum (2005) used finite mixtures to control for patient-level unobserved heterogeneity in estimating a dynamic matching model of pharmaceutical demand. Gowrisankaran, Mitchell, and Moro (2005) estimated a dynamic model of voter behavior with finite mixtures. In labor economics, finite mixtures are a popular choice for controlling for unobserved person-specific effects when

[1]The authors are grateful to the co-editor and three anonymous referees whose comments greatly improved the paper. The authors thank Victor Aguirregabiria, David Byrne, Seung Hyun Hong, Hidehiko Ichimura, Thierry Magnac, and the seminar participants at Hitotsubashi University, University of Tokyo, University of Toronto, New York Camp Econometrics II, and 2006 JEA Spring Meeting for helpful comments. The financial support from SSHRC is gratefully acknowledged.

 DOI: 10.3982/ECTA6763

dynamic discrete choice models are estimated (e.g., Keane and Wolpin (1997), Cameron and Heckman (1998)). Heckman and Singer (1984) used finite mixtures to approximate more general mixture models in the context of duration models with unobserved heterogeneity.

In most applications of finite mixture models, the components of the mixture distribution are assumed to belong to a parametric family. The nonparametric maximum likelihood estimator (NPMLE) of Heckman and Singer (1984) treats the distribution of unobservables nonparametrically but assumes parametric component distributions. Most existing theoretical work on identification of finite mixture models either treats component distributions parametrically or uses training data that are from known component distributions (e.g., Titterington, Smith, and Makov (1985), Rao (1992)). As Hall and Zhou (2003) stated, "very little is known of the potential for consistent nonparametric inference in mixtures without training data."

This paper studies nonparametric identifiability of type probabilities and type-specific component distributions in finite mixture dynamic discrete choice models. Specifically, we assess the identifiability of type probabilities and type-specific component distributions when no parametric assumption is imposed on them. Our point of departure is the work of Hall and Zhou (2003), who proved nonparametric identifiability of two-type mixture models with independent marginals:

$$F(y) = \pi \prod_{t=1}^{T} F_t^1(y_t) + (1-\pi) \prod_{t=1}^{T} F_t^2(y_t), \tag{1}$$

where $F(y)$ is the distribution function of a T-dimensional variable Y, and $F_t^j(y_t)$ is the distribution function of the tth element of Y conditional on type j. Hall and Zhou showed that the type probability π and the type-specific components F_t^j are nonparametrically identifiable from $F(y)$ and its marginals when $T \geq 3$, while they are not when $T = 2$. The intuition behind their result is as follows. Integrating out different elements of y from (1) gives lower-dimensional submodels,

$$F(y_{i_1}, y_{i_2}, \dots, y_{i_l}) = \pi \prod_{s=1}^{l} F_{i_s}^1(y_{i_s}) + (1-\pi) \prod_{s=1}^{l} F_{i_s}^2(y_{i_s}), \tag{2}$$

where $1 \leq l \leq T$, $1 \leq i_1 < \cdots < i_l \leq T$, and $F(y_{i_1}, y_{i_2}, \dots, y_{i_l})$ is the l-variate marginal distribution of $F(y)$. Each lower-dimensional submodel implies a different restriction on the unknown elements, that is, π and the F_t^j's. F and its marginals imply $2^T - 1$ restrictions, while there are $2T + 1$ unknown elements. When $T = 3$, the number of restrictions is the same as the number of unknowns, and one can solve these restrictions to uniquely determine π and the F_t^j's.

While Hall and Zhou's analysis provides the insight that lower-dimensional submodels (2) provide important restrictions for identification, it has limited applicability to the finite mixture models of dynamic discrete choices in economic applications. First, it is difficult to generalize their analysis to three or more types.[2] Second, their model (1) does not have any covariates, while most empirical models in economics involve covariates. Third, the assumption that elements of y are independent in (1) is not realistic in dynamic discrete choice models.

This paper provides sufficient conditions for nonparametric identification for various finite mixture models of dynamic discrete choices used in applied work. Three elements emerge as the important determinants of identification: the time-dimension of panel data, the number of the values the covariates can take, and the heterogeneity of the response of different types to changes in the covariates. For example, in a simple case where the transition function is type-invariant, a time-dimension of $T = 3$ is sufficient for identification, provided that the number of values the covariates can take is no smaller than the number of types and that the changes in the covariates induce sufficiently heterogeneous variations in the choice probabilities across types.

The key insight is that, in models with covariates, different *sequences* of covariates imply different identifying restrictions in the lower-dimensional submodels; in fact, if d is the number of support points of the covariates and T is the time-dimension, then the number of restrictions becomes on the order of d^T. As a result, the presence of covariates provides a powerful source of identification in panel data even with a moderate time-dimension T.

We study a variety of finite mixture dynamic discrete choice models under different assumptions on the Markov property, stationarity, and type-invariance in the transition process. Under a type-invariant transition function and conditional independence, we analyze the nonstationary case that conditional choice probabilities change over time because time-specific aggregate shocks are present or agents are finitely lived. We also examine the case where state dependence is present (for instance, when the lagged choice affects the current choice and/or the transition function of state variables is different across types), and show that identification is possible when a model is stationary first-order Markovian and the panel has a moderate time-dimension $T \geq 6$. This result is important since distinguishing unobserved heterogeneity and state dependence often motivates the use of finite mixture models in empirical studies. On the other hand, our approach has a limitation in that it does not simultaneously allow for both state dependence and nonstationarity.

[2]When the number of types, M, is more than three, Hall, Neeman, Pakyari, and Elmore (2005) showed that for any number of types, M, there exists T_M such that type probabilities and type-specific component distributions are nonparametrically identifiable when $T \geq T_M$, and that T_M is no larger than $(1 + o(1))6M \ln(M)$ as M increases. However, such a T_M is too large for typical panel data sets.

We also study nonparametric identifiability of the number of types, M. Under the assumptions on the Markov property, stationarity, and type-invariance used in this paper, we show that the lower bound of M is identifiable and, furthermore, M itself is identified if the changes in covariates provide sufficient variation in the choice probabilities across types.

Nonparametric identification and estimation of finite mixture dynamic discrete choice models are relevant and useful in practical applications for, at least, the following reasons. First, choosing a parametric family for the component distributions is often difficult because of a lack of guidance from economic theory; nonparametric estimation provides a flexible way to reveal the structure hidden in the data. Furthermore, even when theory offers guidance, comparing parametric and nonparametric estimates allows us to examine the validity of the restrictions imposed by the underlying theoretical model.

Second, analyzing nonparametric identification helps us understand the identification of parametric or semiparametric finite mixture models of dynamic discrete choices. Understanding identification is not a simple task for finite mixture models even with *parametric* component distributions, and formal identification analysis is rarely provided in empirical applications. Once type probabilities and component distributions are nonparametrically identified, the identification analysis of parametric finite mixture models often becomes transparent as it is reduced to the analysis of models without unobserved heterogeneity. As we demonstrate through examples, our nonparametric identification results can be applied to check the identifiability of some parametric finite mixture models.

Third, the identification results of this paper will open the door to applying semiparametric estimators for structural dynamic models to models with unobserved heterogeneity. Recently, by building on the seminal work by Hotz and Miller (1993), computationally attractive semiparametric estimators for structural dynamic models have been developed (Aguirregabiria and Mira (2002), Kasahara and Shimotsu (2008a)), and a number of papers in empirical industrial organization have proposed two-/multistep estimators for dynamic games (e.g., Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007), Pesendorfer and Schmidt-Dengler (2008), Bajari and Hong (2006), and Aguirregabiria and Mira (2007)). To date, however, few of these semiparametric estimators have been extended to accommodate unobserved heterogeneity. This is because these estimators often require an initial nonparametric consistent estimate of type-specific component distributions, but it has not been known whether one can obtain a consistent nonparametric estimate in finite mixture models.[3] The identification results of this paper provide an apparatus

[3]It is believed that it is not possible to obtain a consistent estimate of choice probabilities. For instance, Aguirregabiria and Mira (2007) proposed a pseudo maximum likelihood estimation algorithm for models with unobserved heterogeneity, but stated that (p. 15) "for [models with unobservable market characteristics] it is not possible to obtain consistent nonparametric

that enables researchers to apply these semiparametric estimators to the models with unobserved heterogeneity. This is important since it is often crucial to control for unobserved heterogeneity in dynamic models (see Aguirregabiria and Mira (2007)).

In a closely related paper, Kitamura (2004) examined nonparametric identifiability of finite mixture models with covariates. Our paper shares his insight that the variation in covariates may provide a source of identification; however, the setting as well as the issues we consider are different from Kitamura's. We study discrete choice models in a dynamic setting with panel data, while Kitamura considered regression models with continuous dependent variables with cross-sectional data. We address various issues specific to dynamic discrete choice models, including identification in the presence of state dependence and type-dependent transition probabilities for endogenous explanatory variables.

Our work provides yet another angle for analysis that relates current and previous work on dynamic discrete choice models. Honoré and Tamer (2006) studied identification of dynamic discrete choice models, including the initial conditions problem, and suggested methods to calculate the identified sets. Rust (1994), Magnac and Thesmar (2002), and Aguirregabiria (2006) studied the identification of structural dynamic discrete choice models. Our analysis is also related to an extensive literature on identification of duration models (e.g., Elbers and Ridder (1982), Heckman and Singer (1984), Ridder (1990), and Van den Berg (2001)).

The rest of the paper is organized as follows. Section 2 discusses our approach to identification and provides the identification results using a simple "baseline" model. Section 3 extends the identification analysis of Section 2, and studies a variety of finite mixture dynamic discrete choice models. Section 4 concludes. The proofs are collected in the Appendix.

2. NONPARAMETRIC IDENTIFICATION OF FINITE MIXTURE MODELS OF DYNAMIC DISCRETE CHOICES

Every period, each individual makes a choice a_t from the discrete and finite set A, conditioning on $(x_t, x_{t-1}, a_{t-1}) \in X \times X \times A$, where x_t is observable individual characteristics that may change over time and the lagged choice a_{t-1} is included as one of the conditioning variables. Each individual belongs to one of M types, and his/her type attribute is unknown. The probability of belonging to type m is π^m, where the π^m's are positive and sum to 1.

Throughout this paper, we impose a first-order Markov property on the conditional choice probability of a_t and denote type m's conditional choice probability by $P^m(a_t|x_t, x_{t-1}, a_{t-1})$. The initial distribution of (x_1, a_1) and the transition probability function of x_t are also different across types. For each type

estimates of [choice probabilities]." Furthermore, Geweke and Keane (2001, p. 3490) wrote that "the [Hotz and Miller] methods cannot accommodate unobserved state variables."

m, we denote them by $p^{*m}(x_1, a_1)$ and $f_t^m(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1})$, respectively. With a slight abuse of notation, we let $p^{*m}(x_1, a_1)$ and $f_t^m(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1})$ denote the density of the continuously distributed elements of x_t and the probability mass function of the discretely distributed elements of x_t.

Suppose we have a panel data set with time-dimension equal to T. Each individual observation, $w_i = \{a_{it}, x_{it}\}_{t=1}^T$, is drawn randomly from an M-term mixture distribution,

$$
\begin{aligned}
(3)\qquad & P(\{a_t, x_t\}_{t=1}^T) \\
&= \sum_{m=1}^M \pi^m p^{*m}(x_1, a_1) \prod_{t=2}^T f_t^m(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1}) P_t^m(a_t|x_t, \{x_\tau, a_\tau\}_{\tau=1}^{t-1}) \\
&= \sum_{m=1}^M \pi^m p^{*m}(x_1, a_1) \prod_{t=2}^T f_t^m(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1}) P_t^m(a_t|x_t, x_{t-1}, a_{t-1}),
\end{aligned}
$$

where the first equality presents a general mixture model, while the second equality imposes the Markovian assumption on the conditional choice probabilities, $P_t^m(a_t|x_t, \{x_\tau, a_\tau\}_{\tau=1}^{t-1}) = P_t^m(a_t|x_t, x_{t-1}, a_{t-1})$. This is the key identifying assumption of this paper. The left-hand side of (3) is the distribution function of the observable data, while the right-hand side of the second equality contains the objects we would like the data to inform us about.

REMARK 1: In models where a_t and x_t follow a stationary first-order Markov process, it is sometimes assumed that the choice of the distribution of the initial observation, $p^{*m}(x_1, a_1)$, is the stationary distribution that satisfies the fixed point constraint

$$
(4)\qquad p^{*m}(x_1, a_1) = \sum_{x' \in X} \sum_{a' \in A} P^m(a_1|x_1, x', a') f^m(x_1|x', a') p^{*m}(x', a'),
$$

when all the components of x have finite support. When x is continuously distributed, we replace the summation over x' with integration. Our identification result does not rely on the stationarity assumption of the initial conditions.

The model (3) includes the following examples as special cases.

EXAMPLE 1—Dynamic Discrete Choice Model With Heterogeneous Coefficients: Denote a parameter vector specific to type m's individual by $\theta^m = (\beta^{m\prime}, \rho^m)'$. Consider a dynamic binary choice model for individual i who belongs to type m:

$$
\begin{aligned}
(5)\qquad P^m(a_{it} = 1|x_{it}, \{x_{i\tau}, a_{i\tau}\}_{\tau=1}^{t-1}) &= P^m(a_{it} = 1|x_{it}, a_{i,t-1}) \\
&= \Phi(x_{it}'\beta^m + \rho^m a_{i,t-1}),
\end{aligned}
$$

where the first equality imposes the Markovian assumption and the second follows from the parametric restriction with $\Phi(\cdot)$ denoting the standard normal cumulative distribution function (c.d.f.). The distribution of x_{it} conditional on $(x_{i,t-1}, a_{i,t-1})$ is specific to the value of θ^m. Since the evolution of (x_{it}, a_{it}) in the presample period is not independent of random coefficient θ^m, the initial distribution of (x_{i1}, a_{i1}) depends on the value of θ^m (cf. Heckman (1981)).

Browning and Carro (2007) estimated a continuous mixture version of (5) for the purchase of milk using a Danish consumer "long" panel ($T \geq 100$), and provided evidence for heterogeneity in coefficients. Their study illustrates that allowing for such heterogeneity can make a significant difference for outcomes of interest such as the marginal dynamic effect. In practice however, researchers quite often only have access to a short panel. The results of this paper are therefore useful to understand the extent to which unobserved heterogeneity in coefficients is identified in such a situation.

Our identification results are not applicable, however, to a parametric dynamic discrete choice model with serially correlated idiosyncratic shocks; for example, $a_{it} = 1(x_{it}'\beta^m + \rho^m a_{i,t-1} + \varepsilon_{it})$, where ε_{it} is serially correlated.

EXAMPLE 2—Structural Dynamic Discrete Choice Models: Type m's agent maximizes the expected discounted sum of utilities, $E[\sum_{j=0}^{\infty} \beta^j \{u(x_{t+j}, a_{t+j}; \theta^m) + \varepsilon_{t+j}(a_{t+j})\} | a_t, x_t; \theta^m]$, where x_t is an observable state variable and $\varepsilon_t(a_t)$ is a state variable that are known to the agent but not to the researcher. The Bellman equation for this dynamic optimization problem is

$$V(x) = \int \max_{a \in A} \left\{ u(x, a; \theta^m) + \varepsilon(a) + \beta \sum_{x' \in X} V(x') f(x' | x, a; \theta^m) \right\} \times g(d\varepsilon | x), \tag{6}$$

where $g(\varepsilon|x)$ is the joint distribution of $\varepsilon = \{\varepsilon(j) : j \in A\}$ and $f(x'|x, a; \theta^m)$ is a type-specific transition function. The conditional choice probability is

$$P_{\theta^m}(a|x) = \int 1\left\{ a = \arg\max_{j \in A} \left[u(x, j; \theta^m) + \varepsilon(j) + \beta \sum_{x' \in X} V_{\theta^m}(x') f(x'|x, j; \theta^m) \right] \right\} \times g(d\varepsilon | x), \tag{7}$$

where V_{θ^m} is the fixed point of (6). Let $P_t^m(a_t | x_t, x_{t-1}, a_{t-1}) = P_{\theta^m}(a_t | x_t)$ and $f_t^m(x_t | \{x_\tau, a_\tau\}_{\tau=1}^{t-1}) = f(x_t | x_{t-1}, a_{t-1}; \theta^m)$ in (3). The initial distribution of (x_1, a_1) is given by the stationary distribution (4). Then the likelihood function for $\{a_t, x_t\}_{t=1}^{T}$ is given by (3) with (4).

We study the nonparametric identifiability of the type probabilities, the initial distribution, the type-specific conditional choice probabilities, and the type-specific transition function in equation (3), which we denote by $\theta = \{\pi^m, p^{*m}(\cdot), \{P_t^m(\cdot|\cdot), f_t^m(\cdot|\cdot)\}_{t=2}^T\}_{m=1}^M$. Following the standard definition of nonparametric identifiability, θ is said to be nonparametrically identified (or identifiable) if it is uniquely determined by the distribution function $P(\{a_t, x_t\}_{t=1}^T)$, without making any parametric assumption about the elements of θ. Because the order of the component distributions can be changed, θ is identified only up to a permutation of the components. If no two of the π's are identical, we may uniquely determine the components by assuming $\pi^1 < \pi^2 < \cdots < \pi^M$.

2.1. *Our Approach and Identification of the Baseline Model*

The finite mixture models studied by Hall and Zhou (2003) have no covariates as discussed in the Introduction. In this subsection, we show that the presence of covariates in our model creates a powerful source of identification.

First, we impose the following simplifying assumptions on the general model (3) and analyze the nonparametric identifiability of the resulting "baseline model." Analyzing the baseline model helps elucidate the basic idea of our approach and clarifies the logic behind our main results. In the subsequent sections, we relax Assumption 1 in various ways and study how it affects the identifiability of the resulting models.

ASSUMPTION 1: (a) *The choice probability of* a_t *does not depend on time*. (b) *The choice probability of* a_t *is independent of the lagged variable* (x_{t-1}, a_{t-1}) *conditional on* x_t. (c) $f_t^m(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1}) > 0$ *for all* $(x_t, \{x_\tau, a_\tau\}_{\tau=1}^{t-1}) \in X^t \times A^{t-1}$ *and for all* m. (d) *The transition function is common across types*; $f_t^m(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1}) = f_t(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1})$ *for all* m. (e) *The transition function is stationary*; $f_t(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1}) = f(x_t|x_{t-1}, a_{t-1})$ *for all* m.

Under Assumptions 1(a) and (b), the choice probabilities are written as $P_t^m(a_t|x_t, x_{t-1}, a_{t-1}) = P^m(a_t|x_t)$, where a_{t-1} is not one of the elements of x_t. Under Assumption 1(b), the lagged variable (x_{t-1}, a_{t-1}) affects the current choice a_t only through its effect on x_t via $f_t^m(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1})$. Assumption 1(c) implies that, starting from any combinations of the past state and action, any state $x' \in X$ is reached in the next period with positive probability.

With Assumption 1 imposed, the baseline model is

$$(8) \qquad P(\{a_t, x_t\}_{t=1}^T) = \sum_{m=1}^M \pi^m p^{*m}(x_1, a_1) \prod_{t=2}^T f(x_t|x_{t-1}, a_{t-1}) P^m(a_t|x_t).$$

Since $f(x_t|x_{t-1}, a_{t-1})$ is nonparametrically identified directly from the observed data (cf. Rust (1987)), we may assume $f(x_t|x_{t-1}, a_{t-1})$ is known without affecting the other parts of the argument. Divide $P(\{a_t, x_t\}_{t=1}^T)$ by the transition functions and define

$$\tilde{P}(\{a_t, x_t\}_{t=1}^T) = \frac{P(\{a_t, x_t\}_{t=1}^T)}{\prod_{t=2}^T f(x_t|x_{t-1}, a_{t-1})} = \sum_{m=1}^M \pi^m p^{*m}(x_1, a_1) \prod_{t=2}^T P^m(a_t|x_t), \tag{9}$$

which can be computed from the observed data. Assumption 1 guarantees that $\tilde{P}(\{a_t, x_t\}_{t=1}^T)$ is well defined for any possible sequence of $\{a_t, x_t\}_{t=1}^T \in (A \times X)^T$.

Let $\mathcal{I} = \{i_1, \dots, i_l\}$ be a subset of the time indices, so that $\mathcal{I} \subseteq \{1, \dots, T\}$, where $1 \le l \le T$ and $1 \le i_1 < \cdots < i_l \le T$. Integrating out different elements from (9) gives the l-variate marginal version of $\tilde{P}(\{a_t, x_t\}_{t=1}^T)$, which we call *lower-dimensional submodels*

$$\tilde{P}\big(\{a_{i_s}, x_{i_s}\}_{i_s \in \mathcal{I}}\big) = \sum_{m=1}^M \pi^m p^{*m}\big(x_1, a_1\big) \prod_{s=2}^l P^m(a_{i_s}|x_{i_s}), \quad \text{when} \quad \{1\} \in \mathcal{I}, \tag{10}$$

and

$$\tilde{P}\big(\{a_{i_s}, x_{i_s}\}_{i_s \in \mathcal{I}}\big) = \sum_{m=1}^M \pi^m \prod_{s=1}^l P^m\big(a_{i_s}|x_{i_s}\big), \quad \text{when} \quad \{1\} \notin \mathcal{I}. \tag{11}$$

In model (9), a powerful source of identification is provided by the difference in each type's response patterns to the variation of the covariate $(x_1, \dots, x_T)$. The key insight is that for each different value of $(x_1, \dots, x_T)$, (10) and (11) imply different restrictions on the type probabilities and conditional choice probabilities. Let $|X|$ denote the number of elements in X. The variation of $(x_1, \dots, x_T)$ generates different versions of (10) and (11), providing restrictions whose number is on the order of $|X|^T$, while the number of the parameters $\{\pi^m, p^{*m}(a, x), P^m(a|x) : (a, x) \in A \times X\}_{m=1}^M$ is on the order of $|X|$. This identification approach is much more effective than one without covariates, in particular, when T is small.[4]

In what follows, we assume that the support of the state variables is discrete and known. This is assumed for the sake of clarity: our identification results are easier to understand in the context of a discrete state space, although they hold more generally. We also focus on the case where $A = \{0, 1\}$ to simplify notation. It is straightforward to extend our analysis to the case with a multinomial

[4]For example, when $T = 3$ and $A = \{0, 1\}$, (10) and (11) imply at least $\binom{|X|+2}{3}$ different restrictions while there are $3M|X| - 1$ parameters.

choice of a, but with heavier notation. Note also that Chandra (1977) shows that a multivariate finite mixture model is identified if all the marginal models are identified.

It is convenient to collect notation first. Define, for $\xi \in X$,

$$\lambda_{\xi}^{*m} = p^{*m}((a_1, x_1) = (1, \xi)) \quad \text{and} \quad \lambda_{\xi}^{m} = P^m(a = 1|x = \xi). \tag{12}$$

Let ξ_j, $j = 1, \dots, M-1$, be elements of X. Let k be an element of X. Define a matrix of type-specific distribution functions and type probabilities as

$$\underset{(M\times M)}{L} = \begin{bmatrix} 1 & \lambda_{\xi_1}^{1} & \cdots & \lambda_{\xi_{M-1}}^{1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_{\xi_1}^{M} & \cdots & \lambda_{\xi_{M-1}}^{M} \end{bmatrix}, \tag{13}$$
$$D_k = \mathrm{diag}(\lambda_k^{*1}, \dots, \lambda_k^{*M}), \quad V = \mathrm{diag}(\pi^1, \dots, \pi^M).$$

The elements of L, D_k, and V are parameters of the underlying mixture models to be identified.

Now we collect notation for matrices of observables. Fix $a_t = 1$ for all t in $\tilde{P}(\{a_t, x_t\}_{t=1}^3)$, and define the resulting function as

$$F_{x_1,x_2,x_3}^{*} = \tilde{P}(\{1, x_t\}_{t=1}^3) = \sum_{m=1}^{M} \pi^m \lambda_{x_1}^{*m} \lambda_{x_2}^{m} \lambda_{x_3}^{m}, \tag{14}$$

where λ_x^{*m} and λ_x^m are defined in (12). Next, integrate out (a_1, x_1) from $\tilde{P}(\{a_t, x_t\}_{t=1}^3)$, fix $a_2 = a_3 = 1$, and define the resulting function as

$$F_{x_2,x_3} = \tilde{P}(\{1, x_t\}_{t=2}^3) = \sum_{m=1}^{M} \pi^m \lambda_{x_2}^{m} \lambda_{x_3}^{m}. \tag{15}$$

Similarly, define the following “marginals” by integrating out other elements from $\tilde{P}(\{a_t, x_t\}_{t=1}^3)$ and setting $a_t = 1$:

$$F_{x_1,x_2}^{*} = \tilde{P}(\{1, x_t\}_{t=1}^2) = \sum_{m=1}^{M} \pi^m \lambda_{x_1}^{*m} \lambda_{x_2}^{m}, \tag{16}$$
$$F_{x_1,x_3}^{*} = \tilde{P}(\{1, x_1, 1, x_3\}) = \sum_{m=1}^{M} \pi^m \lambda_{x_1}^{*m} \lambda_{x_3}^{m},$$
$$F_{x_1}^{*} = \tilde{P}(\{1, x_1\}) = \sum_{m=1}^{M} \pi^m \lambda_{x_1}^{*m},$$

$$F_{x_2} = \tilde{P}(\{1, x_2\}) = \sum_{m=1}^{M} \pi^m \lambda_{x_2}^m,$$

$$F_{x_3} = \tilde{P}(\{1, x_3\}) = \sum_{m=1}^{M} \pi^m \lambda_{x_3}^m.$$

Note that $F_{\cdot}^*$ involves (a_1, x_1) while $F_{\cdot}$ does not contain (a_1, x_1). In fact, $F_{x_1,x_2}^* = F_{x_1,x_3}^*$ if $x_2 = x_3$ because $P^m(a|x)$ does not depend on t, but we keep separate notation for the two because later we analyze the case where the choice probability depends on t. Evaluate F_{x_1,x_2,x_3}^*, F_{x_2,x_3}, and their marginals at $x_1 = k$, $x_2 = \xi_1, \dots, \xi_{M-1}$, and $x_3 = \xi_1, \dots, \xi_{M-1}$, and arrange them into two $M \times M$ matrices:

$$P = \begin{bmatrix} 1 & F_{\xi_1} & \cdots & F_{\xi_{M-1}} \\ F_{\xi_1} & F_{\xi_1,\xi_1} & \cdots & F_{\xi_1,\xi_{M-1}} \\ \vdots & \vdots & \ddots & \vdots \\ F_{\xi_{M-1}} & F_{\xi_{M-1},\xi_1} & \cdots & F_{\xi_{M-1},\xi_{M-1}} \end{bmatrix}, \tag{17}$$

$$P_k = \begin{bmatrix} F_k^* & F_{k,\xi_1}^* & \cdots & F_{k,\xi_{M-1}}^* \\ F_{k,\xi_1}^* & F_{k,\xi_1,\xi_1}^* & \cdots & F_{k,\xi_1,\xi_{M-1}}^* \\ \vdots & \vdots & \ddots & \vdots \\ F_{k,\xi_{M-1}}^* & F_{k,\xi_{M-1},\xi_1}^* & \cdots & F_{k,\xi_{M-1},\xi_{M-1}}^* \end{bmatrix}.$$

The following proposition and corollary provide simple and intuitive sufficient conditions for identification under Assumption 1. Proposition 1 extends the idea of the proof of nonparametric identifiability of finite mixture models from Anderson (1954) and Gibson (1955) to models with covariates.[5] Proposition 1 gives a sufficient condition for identification in terms of the rank of the matrix L and the type-specific choice probabilities evaluated at k. In practice, however, it may be difficult to check this rank condition because the elements of L are functions of the component distributions. Corollary 1 provides a sufficient condition in terms of the observable quantities P and P_k. The proofs are constructive.

PROPOSITION 1: *Suppose that Assumption 1 holds and assume $T \geq 3$. Suppose further that there exist some $\{\xi_1, \dots, \xi_{M-1}\}$ such that L is nonsingular and that there exists $k \in X$ such that $\lambda_k^{*m} > 0$ for all m and $\lambda_k^{*m} \neq \lambda_k^{*n}$ for any $m \neq n$. Then*

[5]Anderson (1954) and Gibson (1955) analyzed nonparametric identification of finite mixture models similar to (9) but *without covariates* and derived a sufficient condition for nonparametric identifiability under the assumption $T \geq 2M - 1$. Madansky (1960) extended their analysis to obtain a sufficient condition under the assumption $2^{(T-1)/2} \geq M$. When T is small, the number of identifiable types by their method is quite limited.

$\{\pi^m, \{\lambda_\xi^{*m}, \lambda_\xi^m\}_{\xi\in X}\}_{m=1}^M$ *is uniquely determined from* $\{\tilde{P}(\{a_t, x_t\}_{t=1}^3) : \{a_t, x_t\}_{t=1}^3 \in (A \times X)^3\}$.

COROLLARY 1: *Suppose that Assumption 1 holds, and assume $T \geq 3$. Suppose further that there exist some $\{\xi_1, \dots, \xi_{M-1}\}$ and $k \in X$ such that P is of full rank and that all the eigenvalues of $P^{-1}P_k$ take distinct values. Then* $\{\pi^m, \{\lambda_\xi^{*m}, \lambda_\xi^m\}_{\xi\in X}\}_{m=1}^M$ *is uniquely determined from* $\{\tilde{P}(\{a_t, x_t\}_{t=1}^3) : \{a_t, x_t\}_{t=1}^3 \in (A \times X)^3\}$.

REMARK 2:

(i) The condition of Proposition 1 implies that all columns in L must be linearly independent. Since each column of L represents the conditional choice probability of different types for a given value of x, the changes in x must induce sufficiently heterogeneous variations in the conditional choice probabilities across types. In other words, the covariate must be relevant, and different types must respond to its changes differently.

(ii) When $\lambda_k^{*m} = 0$ for some m, its identification fails, because we never observe (x_1, a_1) for such type. The condition that $\lambda_k^{*m} \neq \lambda_k^{*n}$ for some $k \in X$ is satisfied if the initial distributions are different across different types. If either of these conditions is violated, then the initial distribution cannot be used as a source of identification and, as a result, the requirement on T becomes $T \geq 4$ instead of $T \geq 3$.

(iii) One needs to find only one set of $M - 1$ points to construct a nonsingular L. The identification of choice probabilities at all other points in X follows without any further requirement.

(iv) When X has $|X| < \infty$ support points, the number of identifiable types is at most $|X| + 1$. When x is continuously distributed, we may potentially identify as many types as we wish.

(v) By partitioning X into $M - 1$ disjoint subsets $(\Xi_1, \Xi_2, \dots, \Xi_{M-1})$, we may characterize a sufficient condition in terms of the conditional choice probabilities given a subset Ξ_j of X rather than an element ξ_j of X.

(vi) We may check the conditions of Corollary 1 empirically by computing the sample counterpart of P and P_k for various $\{\xi_1, \dots, \xi_{M-1}\}$'s and/or for various partitions Ξ_j's. The latter procedure is especially useful when x is continuously distributed.

The foundation for our identification method lies in the following relationship between the observables, P and P_k, and the parameters L, D_k, and V, which we call the *factorization equations*:

$$P = L'VL, \quad P_k = L'D_kVL. \tag{18}$$

Note that the $(1, 1)$th element of $P = L'VL$ is $1 = \sum_{m=1}^M \pi^m$. These two equations determine L, D_k, and V uniquely. The first equation of (18) alone does not give a unique decomposition of P in terms of L and V, because this equation provides $M(M+1)/2$ restrictions due to the symmetry of P, while there are $M^2 - M + M = M^2$ unknowns in L and V. Indeed, when $M = 2$, there are three restrictions and four unknowns, and L and V are just not identified.

To shed further light on our identification method, we provide a sketch of how we constructively identify L, D_k, and V from P and P_k. Suppose P is invertible or, equivalently, L is invertible. As is apparent from equation (18), P_k is similar to P except that P_k contains an extra diagonal matrix D_k. Since $P^{-1}P_k = L^{-1}D_kL$, the eigenvalues of $P^{-1}P_k$ identify the elements of D_k. Furthermore, multiplying both sides of $P^{-1}P_k = L^{-1}D_kL$ by L^{-1}, we have $(P^{-1}P_k)L^{-1} = L^{-1}D_k$, suggesting that the columns of L^{-1} are identified with the eigenvectors of $P^{-1}P_k$. Finally, once L is identified, V is identified since $V = (L')^{-1}PL^{-1}$.

By applying the above algorithm to a sample analogue of P and P_k, we may construct an estimator for $\{\pi^m, \{\lambda_\xi^{*m}, \lambda_\xi^m\}_{\xi\in X}\}_{m=1}^M$, which will have the same rate of convergence as the estimates of P and P_k. Alternatively, once identification is established, we may use various nonparametric estimation procedures, such as a series-based mixture likelihood estimator.

Magnac and Thesmar (2002, Proposition 6) studied a finite mixture dynamic discrete choice model similar to our baseline model and showed that their model is not nonparametrically identified. They assumed that the transition probability is common across types and that the initial distribution is independent of the types. Hence, we may express their model in terms of our notation as

$$P(\{a_t, x_t\}_{t=1}^T) = \sum_{m=1}^M \pi^m f(x_1)P^m(a_1|x_1)\prod_{t=2}^T f(x_t|x_{t-1}, a_{t-1})P^m(a_t|x_t). \tag{19}$$

Setting $p^{*m}(x_1, a_1) = f(x_1)P^m(a_1|x_1)$ gives our baseline model (3).

Our results differ from those of Magnac and Thesmar in two ways: the length of the periods considered and the variation of x_t. First, Magnac and Thesmar considered a two-period model, whereas our identifiability result requires at least three periods.[6] Second, Magnac and Thesmar restricted the variation of x_t by assuming that there are only two states, one of which is absorbing. In terms of our notation, this restriction is the same as assuming $x_t \in \{0, 1\}$, and $x_t = 1$ with probability 1 if $x_{t-1} = 1$. This reduces the possible variation of the sequences of x_t substantially, making identification difficult because only $T+1$ different sequences of x_t are observable. For example, when $T = 3$, the only possible sequences of (x_1, x_2, x_3) are $(1, 1, 1)$, $(0, 1, 1)$, $(0, 0, 1)$, and $(0, 0, 0)$.

[6]The possibility of identifying many types using the variation of $|X|$ under $T = 2$ is currently under investigation. A related study on nonidentifiability of multivariate mixtures by Kasahara and Shimotsu (2008b) suggests, however, that $T \geq 3$ is necessary for identification even when $|X| \geq 2$.

If we assume $T \geq 3$ and that there are more than two nonabsorbing states, then we can apply Proposition 1 and Corollary 1 to (19). Alternately, identifying the single nonabsorbing state model is still possible if $T \geq 2M - 1$ by applying Remark 3 below with $\bar{x} = 0$, although Remark 3 uses the stationarity of $P^m(a|x)$.

For the sake of brevity, in the subsequent analysis we provide sufficient conditions only in terms of the rank of the matrix of the type-specific component distributions (e.g., L). In each of the following propositions, sufficient conditions in terms of the distribution function of the observed data can easily be deduced from the conditions in terms of the type-specific component distributions.

The identification method of Proposition 1 uses a set of restrictions implied by the joint distribution of only $(a_1, x_1, a_2, x_2, a_3, x_3)$. When the variation of $(x_1, x_2, \dots, x_T)$ for $T \geq 5$ is available, we may adopt the approach of Madansky (1960) to use the information contained in all x_t's. Define $u = (T-1)/2$, and write the functions corresponding to (14) and (15) as

$$F^*_{x_1\cdots x_T} = \tilde{P}(\{1, x_t\}_{t=1}^T) = \sum_{m=1}^M \pi^m \lambda^{*m}_{x_1} \lambda^m_{x_2} \cdots \lambda^m_{x_T}$$
$$= \sum_{m=1}^M \pi^m \lambda^{*m}_{x_1} (\lambda^m_{x_2} \cdots \lambda^m_{x_{u+1}})(\lambda^m_{x_{u+2}} \cdots \lambda^m_{x_T}) \tag{20}$$

and

$$F_{x_2\cdots x_T} = \tilde{P}(\{1, x_t\}_{t=2}^T) = \sum_{m=1}^M \pi^m (\lambda^m_{x_2} \cdots \lambda^m_{x_{u+1}})(\lambda^m_{x_{u+2}} \cdots \lambda^m_{x_T}). \tag{21}$$

Equations (20) and (21) have the same form as (14) and (15) if we view $\lambda^m_{x_2} \cdots \lambda^m_{x_{u+1}}$ and $\lambda^m_{x_{u+2}} \cdots \lambda^m_{x_T}$ as marginal distributions with $|X|^u$ support points. Consequently, we can construct factorization equations similar to (18), in which the elements of a matrix corresponding to the matrix L are based on $\lambda^m_{x_2} \cdots \lambda^m_{x_{u+1}}$ and $\lambda^m_{x_{u+2}} \cdots \lambda^m_{x_T}$ and their subsets. This extends the maximum number of identifiable types from on the order of $|X|$ to on the order of $|X|^{(T-1)/2}$. Despite being more complex than Proposition 1, the following proposition is useful when T is large, making it possible to identify a large number of types even if $|X|$ is small. For notational simplicity, we assume $|X|$ is finite and $X = \{1, 2, \dots, |X|\}$.

PROPOSITION 2: *Suppose that Assumption 1 holds. Assume $T \geq 5$ is odd and define $u = (T-1)/2$. Suppose $X = \{1, 2, \dots, |X|\}$ and define*

$$\Lambda_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \Lambda_1 = \begin{bmatrix} \lambda^1_1 & \cdots & \lambda^1_{|X|} \\ \vdots & & \vdots \\ \lambda^M_1 & \cdots & \lambda^M_{|X|} \end{bmatrix}.$$

For $l = 2, \ldots, u$, define Λ_l to be a matrix, each column of which is formed by choosing l columns (unordered, with replacement) from the columns of Λ_1 and taking their Hadamard product. There are $\binom{|X|+l-1}{l}$ ways to choose such columns; thus the dimension of Λ_l is $M \times \binom{|X|+l-1}{l}$. For example, Λ_2 and Λ_3 take the form

$$\Lambda_2 = \begin{bmatrix} \lambda_1^1\lambda_1^1 & \cdots & \lambda_1^1\lambda_{|X|}^1 & \lambda_2^1\lambda_2^1 & \cdots & \lambda_2^1\lambda_{|X|}^1 & \cdots & \lambda_{|X|}^1\lambda_{|X|}^1 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ \lambda_1^M\lambda_1^M & \cdots & \lambda_1^M\lambda_{|X|}^M & \lambda_2^M\lambda_2^M & \cdots & \lambda_2^M\lambda_{|X|}^M & \cdots & \lambda_{|X|}^M\lambda_{|X|}^M \end{bmatrix},$$

$$\Lambda_3 = \left[\begin{matrix} \lambda_1^1\lambda_1^1\lambda_1^1 & \cdots & \lambda_1^1\lambda_1^1\lambda_{|X|}^1 \\ \vdots & & \vdots \\ \lambda_1^M\lambda_1^M\lambda_1^M & \cdots & \lambda_1^M\lambda_1^M\lambda_{|X|}^M \end{matrix} \right.$$
$$\left. \begin{matrix} \lambda_2^1\lambda_1^1\lambda_2^1 & \cdots & \lambda_2^1\lambda_1^1\lambda_{|X|}^1 & \cdots & \lambda_{|X|}^1\lambda_{|X|}^1\lambda_{|X|}^1 \\ \vdots & & \vdots & & \vdots \\ \lambda_2^M\lambda_1^M\lambda_2^M & \cdots & \lambda_2^M\lambda_1^M\lambda_{|X|}^M & \cdots & \lambda_{|X|}^M\lambda_{|X|}^M\lambda_{|X|}^M \end{matrix} \right].$$

Define an $M \times (\sum_{l=0}^{u} \binom{|X|+l-1}{l})$ matrix Λ as

$$\Lambda = [\Lambda_0, \Lambda_1, \Lambda_2, \ldots, \Lambda_u].$$

Suppose (a) $\sum_{l=0}^{u} \binom{|X|+l-1}{l} \geq M$, (b) *we can construct a nonsingular $M \times M$ matrix $L^\diamond$ by setting its first column as Λ_0 and choosing other $M-1$ columns from the columns of Λ other than Λ_0, and* (c) *there exists $k \in X$ such that $\lambda_k^{*m} > 0$ for all m and $\lambda_k^{*m} \neq \lambda_k^{*n}$ for any $m \neq n$. Then $\{\pi^m, \{\lambda_j^{*m}, \lambda_j^m\}_{j=1}^{|X|}\}_{m=1}^{M}$ is uniquely determined from $\{\tilde{P}(\{a_t, x_t\}_{t=1}^T) : \{a_t, x_t\}_{t=1}^T \in (A \times X)^T\}$.*

REMARK 3: In a special case where there are no covariates and $|X| = 1$, the matrix Λ becomes

$$\Lambda = \begin{bmatrix} 1 & \lambda_1^1 & (\lambda_1^1)^2 & \cdots & (\lambda_1^1)^u \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_1^M & (\lambda_1^M)^2 & \cdots & (\lambda_1^M)^u \end{bmatrix},$$

and the sufficient condition of Proposition 2 reduces to (a) $T \geq 2M - 1$, (b) $\lambda_1^m \neq \lambda_1^n$ for any $m \neq n$, and (c) $\lambda_1^{*m} > 0$ and $\lambda_1^{*m} \neq \lambda_1^{*n}$ for any $m \neq n$. Not surprisingly, the condition $T \geq 2M - 1$ coincides with the sufficient condition of nonparametric identification of finite mixtures of binomial distributions (Blischke (1964)). This set of sufficient condition also applies to the case where the covariates have no time variation ($x_1 = \cdots = x_T$), such as race and/or sex. In this case, because of the stationarity, the time-series variation of a_t substitutes for the variation of x_t.

Houde and Imai (2006) studied nonparametric identification of finite mixture dynamic discrete choice models by fixing the value of the covariate x (to $\bar{x}$, for instance) and derived a sufficient condition for T. They also considered a model with terminating state.

If the conditional choice probabilities of different types are heterogeneous and the column vectors $(\lambda_x^1, \dots, \lambda_x^M)'$ for $x = 1, \dots, |X|$ are linearly independent, the rank condition of this proposition is likely to be satisfied, since the Hadamard products of these column vectors are unlikely to be linearly dependent, unless by chance.

Since the construction of the matrices in Proposition 2 is rather complex, we provide a simple example with $T = 5$ to illustrate its connection to the L matrix in Proposition 1.

EXAMPLE 3—An Example for Proposition 2: Suppose that $T = 5$ and $X = \{1, 2\}$. In this case, we can identify $M = \sum_{l=0}^{2} \binom{1+l}{l} = 6$ types. Consider a matrix

$$L^\diamond = \Lambda = \begin{bmatrix} 1 & \lambda_1^1 & \lambda_2^1 & \lambda_1^1\lambda_1^1 & \lambda_1^1\lambda_2^1 & \lambda_2^1\lambda_2^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_1^M & \lambda_2^M & \lambda_1^M\lambda_1^M & \lambda_1^M\lambda_2^M & \lambda_2^M\lambda_2^M \end{bmatrix}.$$

Then the factorization equations that correspond to (18) are given by

$$P^\diamond = (L^\diamond)' V L^\diamond, \qquad P_k^\diamond = (L^\diamond)' D_k V L^\diamond, \tag{22}$$

where $V = \operatorname{diag}(\pi^1, \dots, \pi^M)$ and $D_k = \operatorname{diag}(\lambda_k^{*1}, \dots, \lambda_k^{*M})$, as defined in (13). We can verify that the elements of $P^\diamond$ and $P_k^\diamond$ can be constructed from the distribution function of the observed data. For instance,

$$P^\diamond = \begin{bmatrix} 1 & F_1 & F_2 & F_{11} & F_{12} & F_{22} \\ F_1 & F_{11} & F_{12} & F_{111} & F_{112} & F_{122} \\ F_2 & F_{21} & F_{22} & F_{211} & F_{212} & F_{222} \\ F_{11} & F_{111} & F_{112} & F_{1111} & F_{1112} & F_{1122} \\ F_{12} & F_{121} & F_{122} & F_{1211} & F_{1212} & F_{1222} \\ F_{22} & F_{221} & F_{222} & F_{2211} & F_{2212} & F_{2222} \end{bmatrix},$$

where $F_i = \sum_{m=1}^M \pi^m \lambda_i^m$, $F_{ij} = \sum_{m=1}^M \pi^m \lambda_i^m \lambda_j^m$, $F_{ijk} = \sum_{m=1}^M \pi^m \lambda_i^m \lambda_j^m \lambda_k^m$, and $F_{ijkl} = \sum_{m=1}^M \pi^m \lambda_i^m \lambda_j^m \lambda_k^m \lambda_l^m$ for $i, j, k, l \in \{1, 2\}$ are identifiable from the population. Once the factorization equations (22) are constructed, we may apply the argument following Corollary 1 to determine $L^\diamond$, V, and D_k uniquely from $P^\diamond$ and $P_k^\diamond$.

2.2. *Identification of the Number of Types*

So far, we have assumed that the number of mixture components M is known. How to choose M is an important practical issue because economic theory usually does not provide much guidance. We now show that it is possible to nonparametrically identify the number of types from panel data with two periods.[7]

Assume $T \geq 2$ and $X = \{1, \dots, |X|\}$. Define a $(|X|+1) \times (|X|+1)$ matrix which is analogous to P in (17) but uses the first two periods and all the support points of X:

$$P^* = \begin{bmatrix} 1 & F_1 & \cdots & F_{|X|} \\ F_1^* & F_{11}^* & \cdots & F_{1,|X|}^* \\ \vdots & \vdots & \ddots & \vdots \\ F_{|X|}^* & F_{|X|,1}^* & \cdots & F_{|X|,|X|}^* \end{bmatrix},$$

where, as defined in (16), $F_i^* = \tilde{P}(\{(a_1, x_1) = (1, i)\}) = \sum_{m=1}^{M} \pi^m \lambda_i^{*m}$, $F_i = \tilde{P}(\{(a_2, x_2) = (1, i)\}) = \sum_{m=1}^{M} \pi^m \lambda_i^m$, and $F_{i,j}^* = \tilde{P}(\{(a_1, x_1, a_2, x_2) = (1, i, 1, j)\}) = \sum_{m=1}^{M} \pi^m \lambda_i^{*m} \lambda_j^m$. The matrix P^* contains information on how different types react differently to the changes in covariates for all possible x's. The following proposition shows that we may nonparametrically identify the number of types from P^* under Assumption 1.

PROPOSITION 3: *Suppose that Assumption* 1 *holds*. *Assume $T \geq 2$ and $X = \{1, \dots, |X|\}$. Then $M \geq \operatorname{rank}(P^*)$. Furthermore*, *in addition to Assumption* 1, *suppose that the two matrices L_1^* and L_2^* defined below both have rank M*:

$$\underset{(M\times(|X|+1))}{L_1^*} = \begin{bmatrix} 1 & \lambda_1^{*1} & \cdots & \lambda_{|X|}^{*1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_1^{*M} & \cdots & \lambda_{|X|}^{*M} \end{bmatrix},$$

$$\underset{(M\times(|X|+1))}{L_2^*} = \begin{bmatrix} 1 & \lambda_1^{1} & \cdots & \lambda_{|X|}^{1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_1^{M} & \cdots & \lambda_{|X|}^{M} \end{bmatrix}.$$

Then $M = \operatorname{rank}(P^*)$.

REMARK 4:

(i) The rank condition of L_1^* and L_2^* is not empirically testable from the observed data. The rank of P^*, which is observable, gives the lower bound of the number of types.

[7]We thank the co-editor and a referee for suggesting that we investigate this problem.

(ii) Surprisingly, two periods of panel data, rather than three periods, may suffice for identifying the number of types.

(iii) The rank condition on L_1^* implies that no row of L_1^* can be expressed as a linear combination of the other rows of L_1^*. The same applies to the rank condition on L_2^*. Since the mth row of L_1^* or L_2^* completely summarizes type m's conditional choice probability within each period, this condition requires that the changes in x provide sufficient variation in the choice probabilities across types and that no type is "redundant" in one-dimensional submodels.

(iv) The rank condition on L_2^* is equivalent to the rank condition on L in Proposition 1 when $X = \{1, \dots, |X|\}$. In other words, $\text{rank}(L_2^*) = M$ if and only if $\text{rank}(L) = M$.

(v) We may partition X into disjoint subsets and compute P^* with respect to subsets of X rather than elements of X.

(vi) When $T \geq 4$, we may use an approach similar to Proposition 2 to construct a $(\sum_{l=0}^{u} \binom{|X|+l-1}{l} \times \sum_{l=0}^{u} \binom{|X|+l-1}{l})$ matrix P^* (similar to $P^\diamond$ in Example 3 but using (x_1, x_2) and (x_3, x_4)) and increase the number of identifiable M to the order of $|X|^{(T-1)/2}$.

3. EXTENSIONS OF THE BASELINE MODEL

In this section, we relax Assumption 1 of the baseline model in various ways to accommodate real-world applications. In the following subsections, we relax Assumption 1(a) and (e) (stationarity), Assumption 1(b) and (d) (type-invariant transition), and Assumption 1(c) (unrestricted transition) in turn and analyze nonparametric identifiability of resulting models. In all cases, identification is achieved by constructing a version of the factorization equation similar to (18), specific to the model under consideration, and then applying an argument that follows the one presented in (18). The differences arise solely from the ways in which the factorization equations are constructed across the various models.

3.1. *Time-Dependent Conditional Choice Probabilities*

The baseline model (8) assumes that conditional choice probabilities and the transition function do not change over periods. However, the agent's decision rules may change over periods in some models, such as a model with time-specific aggregate shocks or a model of finitely lived individuals. In this subsection, we keep the assumption of the common transition function, but relax Assumption 1(a) and (e) to extend our analysis to models with time-dependent choice probabilities.

When Assumption 1(a) and (e) (stationarity) are relaxed but Assumption 1(b) and (d) (conditional independence and type-invariant transition) are maintained, the choice probabilities and the transition function are written as

$P_t^m(a_t|x_t, x_{t-1}, a_{t-1}) = P_t^m(a_t|x_t)$ and $f_t^m(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1}) = f_t(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1})$, respectively, where a_{t-1} is not an element of x_t. Equation (9) then becomes

$$\tilde{P}(\{a_t, x_t\}_{t=1}^T) = \frac{P(\{a_t, x_t\}_{t=1}^T)}{\prod_{t=2}^T f_t(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1})} = \sum_{m=1}^M \pi^m p^{*m}(x_1, a_1) \prod_{t=2}^T P_t^m(a_t|x_t). \tag{23}$$

The next proposition states a sufficient condition for nonparametric identification of the mixture model (23). In the baseline model (8), the sufficient condition is summarized to the invertibility of a matrix consisting of the conditional choice probabilities. In the time-dependent case, this matrix of conditional choice probabilities becomes time-dependent, and hence its invertibility needs to hold for each period. We consider the case of $A = \{0, 1\}$. Define, for $\xi \in X$,

$$\lambda_\xi^{*m} = p^{*m}((a_1, x_1) = (1, \xi)) \quad \text{and}$$
$$\lambda_{t,\xi}^m = P_t^m(a_t = 1|x_t = \xi), \quad t = 2, \dots, T.$$

PROPOSITION 4: *Suppose that Assumptions 1(b)–(d) hold and assume $T \geq 3$. For $t = 2, \dots, T-1$, let ξ_j^t, $j = 1, \dots, M-1$, be elements of X and define*

$$\underset{(M\times M)}{L_t} = \begin{bmatrix} 1 & \lambda_{t,\xi_1^t}^1 & \cdots & \lambda_{t,\xi_{M-1}^t}^1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_{t,\xi_1^t}^M & \cdots & \lambda_{t,\xi_{M-1}^t}^M \end{bmatrix}.$$

Suppose there exist $\{\xi_1^t, \dots, \xi_{M-1}^t\}$ such that L_t is nonsingular for $t = 2, \dots, T$ and there exists $k \in X$ such that $\lambda_k^{*m} > 0$ for all m and $\lambda_k^{*m} \neq \lambda_k^{*n}$ for any $m \neq n$. Then $\{\pi^m, \{\lambda_\xi^{*m}, \{\lambda_{t,\xi}^m\}_{t=2}^T\}_{\xi\in X}\}_{m=1}^M$ is uniquely determined from $\{\tilde{P}(\{a_t, x_t\}_{t=1}^T) : \{a_t, x_t\}_{t=1}^T \in (A \times X)^T\}$.

When the choice probabilities are time-dependent, the factorization equations (that correspond to (18)) are also time-dependent:

$$P_t = L_t' V L_{t+1} \quad \text{and} \quad P_{t,k} = L_t' D_k V L_{t+1} \quad \text{for} \quad t = 2, \dots, T-1,$$

where V and D_k are defined as before. The elements of P_t and $P_{t,k}$ are the "marginals" of $\tilde{P}(\{a_t, x_t\}_{t=1}^T)$ in equation (23) with $a_t = 1$ for all t. Specifically,

P_t and $P_{t,k}$ are defined as

$$P_t = \begin{bmatrix} 1 & F^{t+1}_{\xi_1^{t+1}} & \cdots & F^{t+1}_{\xi_{M-1}^{t+1}} \\ F^t_{\xi_1^t} & F^{t,t+1}_{\xi_1^t,\xi_1^{t+1}} & \cdots & F^{t,t+1}_{\xi_1^t,\xi_{M-1}^{t+1}} \\ \vdots & \vdots & \ddots & \vdots \\ F^t_{\xi_{M-1}^t} & F^{t,t+1}_{\xi_{M-1}^t,\xi_1^{t+1}} & \cdots & F^{t,t+1}_{\xi_{M-1}^t,\xi_{M-1}^{t+1}} \end{bmatrix},$$

$$P_{t,k} = \begin{bmatrix} F^*_k & F^{*t+1}_{k,\xi_1^{t+1}} & \cdots & F^{*t+1}_{k,\xi_{M-1}^{t+1}} \\ F^{*t}_{k,\xi_1^t} & F^{*t,t+1}_{k,\xi_1^t,\xi_1^{t+1}} & \cdots & F^{*t,t+1}_{k,\xi_1^t,\xi_{M-1}^{t+1}} \\ \vdots & \vdots & \ddots & \vdots \\ F^{*t}_{k,\xi_{M-1}^t} & F^{*t,t+1}_{k,\xi_{M-1}^t,\xi_1^{t+1}} & \cdots & F^{*t,t+1}_{k,\xi_{M-1}^t,\xi_{M-1}^{t+1}} \end{bmatrix},$$

where, similar to (15), $F^t_{\xi^t} = \tilde{P}(\{(a_t, x_t) = (1, \xi^t)\}) = \sum_{m=1}^M \pi^m \lambda^m_{t,\xi^t}$, $F^{t,t+1}_{\xi^t,\xi^{t+1}} = \tilde{P}(\{(a_t, x_t, a_{t+1}, x_{t+1}) = (1, \xi^t, 1, \xi^{t+1})\}) = \sum_{m=1}^M \pi^m \lambda^m_{t,\xi^t}\lambda^m_{t+1,\xi^{t+1}}$, $F^*_k = \tilde{P}(\{(a_1, x_1) = (1, k)\}) = \sum_{m=1}^M \pi^m \lambda^{*m}_k$, $F^{*t}_{k,\xi^t} = \tilde{P}(\{(a_1, x_1, a_t, x_t) = (1, k, 1, \xi^t)\}) = \sum_{m=1}^M \pi^m \lambda^{*m}_k \lambda^m_{t,\xi^t}$, and $F^{*t,t+1}_{k,\xi^t,\xi^{t+1}} = \tilde{P}(\{(a_1, x_1, a_t, x_t, a_{t+1}, x_{t+1}) = (1, k, 1, \xi^t, 1, \xi^{t+1})\}) = \sum_{m=1}^M \pi^m \lambda^{*m}_k \lambda^m_{t,\xi^t}\lambda^m_{t+1,\xi^{t+1}}$. Since P_t and $P_{t,k}$ are identifiable from the data, we may construct V, D_k, L_t, and L_{t+1} from P_t and $P_{t,k}$ for $t = 2, \ldots, T-1$ by applying an argument that follows the one presented in (18) to each period.

The following proposition corresponds to Proposition 2 and relaxes the identification condition of Proposition 4 when $T \geq 5$ by utilizing all the marginals of $\tilde{P}(\{a_t, x_t\}_{t=1}^T)$. The proof is omitted because it is similar to that of Proposition 2. The difference from Proposition 2 is (i) the conditions are stated in terms of $\lambda^m_{t,\xi}$ instead of λ^m_ξ because of time-dependence and (ii) the number of restrictions implied by the submodels, analogously defined to (10) and (11) but with time subscripts, is larger because the order of the choices becomes relevant. As a result, the condition on $|X|$ of Proposition 5 is weaker than that of Proposition 2.

PROPOSITION 5: *Suppose Assumption* 1(b)–(d) *hold. Assume $T \geq 5$ is odd and define $u = (T-1)/2$. Suppose $X = \{1, \ldots, |X|\}$ and further define*

$$\bar{\Lambda}_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \bar{\Lambda}_{1,1} = \begin{bmatrix} \lambda^1_{2,1} & \cdots & \lambda^1_{2,|X|} \\ \vdots & & \vdots \\ \lambda^M_{2,1} & \cdots & \lambda^M_{2,|X|} \end{bmatrix}.$$

For $l = 2, \ldots, u$, define $\bar{\Lambda}_{1,l}$ to be a matrix whose elements consist of the l-variate product of the form $\lambda^m_{2,j_2}\lambda^m_{3,j_3}\cdots\lambda^m_{l,j_{l+1}}$, covering all possible l ordered combinations (*with replacement*) *of* $(j_2, j_3, \ldots, j_{l+1})$ *from* $(1, \ldots, |X|)$. *For example*,

$$\bar{\Lambda}_{1,2} = \begin{bmatrix} \lambda^1_{2,1}\lambda^1_{3,1} & \cdots & \lambda^1_{2,1}\lambda^1_{3,|X|} & \lambda^1_{2,2}\lambda^1_{3,1} & \cdots & \lambda^1_{2,2}\lambda^1_{3,|X|} & \lambda^1_{2,|X|}\lambda^1_{3,1} & \cdots & \lambda^1_{2,|X|}\lambda^1_{3,|X|} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \lambda^M_{2,1}\lambda^M_{3,1} & \cdots & \lambda^M_{2,1}\lambda^M_{3,|X|} & \lambda^M_{2,2}\lambda^M_{3,1} & \cdots & \lambda^M_{2,2}\lambda^M_{3,|X|} & \lambda^M_{2,|X|}\lambda^M_{3,1} & \cdots & \lambda^M_{2,|X|}\lambda^M_{3,|X|} \end{bmatrix}.$$

Similarly, for $l = 1, \ldots, u$, *define* $\bar{\Lambda}_{2,l}$ *to be a matrix whose elements consist of the l-variate product of the form* $\lambda^m_{u+1,j_2}\lambda^m_{u+2,j_3}\cdots\lambda^m_{u+l,j_{l+1}}$, *covering all possible l ordered combinations* (*with replacement*) *of* $(j_2, j_3, \ldots, j_{l+1})$ *from* $(1, \ldots, |X|)$. *Let*

$$\bar{\Lambda}_1 = [\bar{\Lambda}_0, \bar{\Lambda}_{1,1}, \bar{\Lambda}_{1,2}, \ldots, \bar{\Lambda}_{1,u}] \quad \text{and} \quad \bar{\Lambda}_2 = [\bar{\Lambda}_0, \bar{\Lambda}_{2,1}, \bar{\Lambda}_{2,2}, \ldots, \bar{\Lambda}_{2,u}].$$

Define $\bar{L}^\diamond_1$ *to be an* $M \times M$ *matrix whose first column is* $\bar{\Lambda}_0$ *and whose other* $M-1$ *columns are from the columns of* $\bar{\Lambda}_1$ *other than* $\bar{\Lambda}_0$. *Define* $\bar{L}^\diamond_2$ *to be an* $M \times M$ *matrix whose first column is* $\bar{\Lambda}_0$ *and whose other columns are from* $\bar{\Lambda}_2$.

Suppose (a) $\sum_{l=0}^{u} |X|^l \geq M$, (b) $\bar{L}^\diamond_1$ *and* $\bar{L}^\diamond_2$ *are nonsingular, and* (c) *there exists* $k \in X$ *such that* $\lambda^{*m}_k > 0$ *for all* m *and* $\lambda^{*m}_k \neq \lambda^{*n}_k$ *for any* $m \neq n$. *Then* $\{\pi^m, \{\lambda^{*m}_j, \{\lambda^m_{t,j}\}_{t=2}^T\}_{j=1}^{|X|}\}_{m=1}^M$ *is uniquely determined from* $\{\tilde{P}(\{a_t, x_t\}_{t=1}^T) : \{a_t, x_t\}_{t=1}^T \in (A \times X)^T\}$.

3.2. *Lagged Dependent Variable and Type-Specific Transition Functions*

In empirical applications, including the lagged choice in explanatory variables for the current choice is a popular way to specify dynamic discrete choice models. Furthermore, we may encounter a case where the transition pattern of state variables is heterogeneous across individuals, even after controlling for other observables. In such cases, the transition function of both a_t and x_t becomes type-dependent.

In this subsection, we relax Assumption 1(b) and (d) of the baseline model (8) to accommodate type-specific transition functions as well as the dependence of current choice on lagged variables. In place of Assumption 1(b) and (d), we impose stationarity and a first-order Markov property on the transition process of x_t. Assumption 2(a) and (c) are identical to Assumption 1(a) and (c).

ASSUMPTION 2: (a) *The choice probability of* a_t *does not depend on time*. (b) x_t *follows a stationary first-order Markov process*; $f^m_t(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1}) = f^m(x_t|x_{t-1}, a_{t-1})$ *for all* t *and* m. (c) $f^m(x'|x, a) > 0$ *for all* $(x', x, a) \in X \times X \times A$ *and for all* m.

Under Assumption 2, the model is

$$P(\{a_t, x_t\}_{t=1}^{T}) = \sum_{m=1}^{M} \pi^m p^{*m}(x_1, a_1) \prod_{t=2}^{T} f^m(x_t|x_{t-1}, a_{t-1}) P^m(a_t|x_t, x_{t-1}, a_{t-1}) \tag{24}$$

and the transition process of (a_t, x_t) becomes a stationary first-order Markov process. Define $s_t = (a_t, x_t)$, $q^{*m}(s_1) = p^{*m}(x_1, a_1)$, and $Q^m(s_t|s_{t-1}) = f^m(x_t| x_{t-1}, a_{t-1}) P^m(a_t|x_t, x_{t-1}, a_{t-1})$, and rewrite the model (24) as

$$P(\{s_t\}_{t=1}^{T}) = \sum_{m=1}^{M} \pi^m q^{*m}(s_1) \prod_{t=2}^{T} Q^m(s_t|s_{t-1}). \tag{25}$$

Unlike the transformed baseline model (9), s_t appears both in $Q^m(s_t|s_{t-1})$ and $Q^m(s_{t+1}|s_t)$, and creates the dependence between these terms. Consequently, the variation of s_t affects $P(\{s_t\}_{t=1}^{T})$ via both $Q^m(s_t|s_{t-1})$ and $Q^m(s_{t+1}|s_t)$. This dependence makes it difficult to construct factorization equations that correspond to (18), which is the key to obtaining identification.

We solve this dependence problem by using the Markov property of s_t. The idea is that if s_t follows a first-order Markov process, looking at *every other period* breaks the dependence of s_t across periods. Specifically, consider the sequence (s_{t-1}, s_t, s_{t+1}) for various values of s_t, while fixing the values of s_{t-1} and s_{t+1}. Once s_{t-1} and s_{t+1} are fixed, the variation of s_t does not affect the state variables in other periods because of the Markovian structure of $Q^m(s_t|s_{t-1})$. As a result, we can use this variation to distinguish different types.

Let $\bar{s} \in S = A \times X$ be a fixed value of s and define

$$\pi_{\bar{s}}^m = \pi^m q^{*m}(\bar{s}), \quad \lambda_{\bar{s}}^m(s) = Q^m(\bar{s}|s) Q^m(s|\bar{s}), \quad \lambda_{\bar{s}}^{*m}(s_T) = Q^m(s_T|\bar{s}). \tag{26}$$

Assume T is even and consider $P(\{s_t\}_{t=1}^{T})$ with $s_t = \bar{s}$ for odd t:

$$P(\{s_t\}_{t=1}^{T}|s_t = \bar{s} \text{ for } t \text{ odd}) = \sum_{m=1}^{M} \pi_{\bar{s}}^m \left(\prod_{t=2,4,\ldots}^{T-2} \lambda_{\bar{s}}^m(s_t) \right) \lambda_{\bar{s}}^{*m}(s_T). \tag{27}$$

This conditional mixture model shares the property of independent marginals with (9). Consequently, we can construct factorization equations similar to (18) and, hence, can identify the components of the mixture model (27) for each $\bar{s} \in S$.

Assume $T = 6$. Let ξ_j, $j = 1, \ldots, M-1$, be elements of S and let $k \in S$. Define

$$\underset{(M\times M)}{L_{\bar{s}}} = \begin{bmatrix} 1 & \lambda_{\bar{s}}^1(\xi_1) & \cdots & \lambda_{\bar{s}}^1(\xi_{M-1}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_{\bar{s}}^M(\xi_1) & \cdots & \lambda_{\bar{s}}^M(\xi_{M-1}) \end{bmatrix},$$

$$V_{\bar{s}} = \text{diag}(\pi_{\bar{s}}^1, \dots, \pi_{\bar{s}}^M), \quad D_{k|\bar{s}} = \text{diag}(\lambda_{\bar{s}}^{*1}(k), \dots, \lambda_{\bar{s}}^{*M}(k)).$$

Then, from (27), the factorization equations that correspond to (18) are

$$P_{\bar{s}} = L_{\bar{s}}' V_{\bar{s}} L_{\bar{s}}, \quad P_{\bar{s},k} = L_{\bar{s}}' D_{k|\bar{s}} V_{\bar{s}} L_{\bar{s}}, \tag{28}$$

where the elements of $P_{\bar{s}}$ and $P_{\bar{s},k}$ are various marginals of the left-hand side of (27) and are identifiable from the data. Then we can construct $V_{\bar{s}}$, $D_{k|\bar{s}}$, and $L_{\bar{s}}$ uniquely from $P_{\bar{s}}$ and $P_{\bar{s},k}$ by applying the argument following Corollary 1.

The following proposition establishes a sufficient condition for nonparametric identification of model (27). Because of the temporal dependence in s_t, the requirement on T becomes $T \geq 6$ instead of $T \geq 3$.

PROPOSITION 6: *Suppose Assumption 2 holds and assume $T \geq 6$. Suppose that $q^{*m}(\bar{s}) > 0$ for all m, there exist some $\{\xi_1, \dots, \xi_{M-1}\}$ such that $L_{\bar{s}}$ is nonsingular, and there exists $k \in S$ such that $\lambda_{\bar{s}}^{*m}(k) > 0$ for all m and $\lambda_{\bar{s}}^{*m}(k) \neq \lambda_{\bar{s}}^{*n}(k)$ for any $m \neq n$. Then $\{\pi_{\bar{s}}^m, \{\lambda_{\bar{s}}^m(s), \lambda_{\bar{s}}^{*m}(s)\}_{s\in S}\}_{m=1}^M$ is uniquely determined from $\{P(\{s_t\}_{t=1}^T) : \{s_t\}_{t=1}^T \in S^T\}$.*

REMARK 5:

(i) The assumption of stationarity and a first-order Markov property is crucial. When s_t follows a second-order Markov process (e.g., $P^m(a_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1}) = P^m(a_t|x_{t-1}, a_{t-1}, x_{t-2}, a_{t-2})$), the requirement on T becomes $T \geq 9$ instead of $T \geq 6$ because we need to look at *every two other periods* so as to obtain the "independent" variation of s_t across periods.

(ii) The transition functions and conditional choice probabilities are identified from $\lambda_{\bar{s}}^{*m}(s)$ as follows. Recall $Q^m(s|\bar{s}) = \lambda_{\bar{s}}^{*m}(s)$ by definition (see (26)) and $Q^m(s|\bar{s}) = f^m(x|\bar{x}, \bar{a}) P^m(a|x, \bar{x}, \bar{a})$ with $(\bar{a}, \bar{x}) = \bar{s}$. Summing $Q^m(s|\bar{s})$ over $a \in A$ gives $f^m(x|\bar{x}, \bar{a})$, and we then identify the conditional choice probabilities by $P^m(a|x, \bar{x}, \bar{a}) = Q^m(s|\bar{s})/f^m(x|\bar{x}, \bar{a})$.

(iii) If $|S| \gg M$ and the transition pattern of s is sufficiently heterogeneous across different types, the sufficient conditions in Proposition 6 are likely to hold for all $\bar{s} \in S$, and we may therefore identify the primitive parameters π^m, $p^{*m}(a, x)$, $f^m(x'|x, a)$, and $P^m(a'|x', x, a)$. Specifically, repeating Proposition 6 for all $\bar{s} \in S$, we obtain $\pi^m q^{*m}(s) = \pi^m p^{*m}(a, x)$ for all $(a, x) \in A \times X$. Then π^m is determined by $\pi^m = \sum_{(a,x)\in A\times X} \pi^m p^{*m}(a, x)$ and we identify $p^{*m}(a, x) = (\pi^m p^{*m}(a, x))/\pi^m$.

EXAMPLE 4—An Example for Proposition 6: Consider a case in which $T = 6$, $A = \{0, 1\}$, and $X = \{0, 1\}$. Then $M = |S| + 1 = 5$ types can be identified. Fix $s_1 = s_3 = s_5 = \bar{s} \in A \times X$. Then $L_{\bar{s}}$ is given by

$$\underset{(M\times M)}{L_{\bar{s}}} = \begin{bmatrix} 1 & \lambda_{\bar{s}}^1(0,0) & \lambda_{\bar{s}}^1(0,1) & \lambda_{\bar{s}}^1(1,0) & \lambda_{\bar{s}}^1(1,1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_{\bar{s}}^M(0,0) & \lambda_{\bar{s}}^M(0,1) & \lambda_{\bar{s}}^M(1,0) & \lambda_{\bar{s}}^M(1,1) \end{bmatrix}.$$

For example, the $(5,5)$th elements of $P_{\bar{s}}$ and $P_{\bar{s},k}$ in (28) are given by $P(\{s_1 = s_3 = s_5 = \bar{s}, s_2 = s_4 = (1,1)\}) = \sum_{m=1}^{5} \pi_{\bar{s}}^m (\lambda_{\bar{s}}^m(1,1))^2$ and $P(\{s_1 = s_3 = s_5 = \bar{s}, s_2 = s_4 = (1,1), s_6 = k\}) = \sum_{m=1}^{5} \pi_{\bar{s}}^m (\lambda_{\bar{s}}^m(1,1))^2 \lambda_{\bar{s}}^{*m}(k)$, respectively.

When $T \geq 8$, we can relax the condition $|S| \geq M-1$ of Proposition 6 by applying the approach of Proposition 2. Define $\lambda_{\bar{s}}^{*m}(s) = Q^m(s|\bar{s})$, $\lambda_{\bar{s},1}^m(s_1) = Q^m(\bar{s}|s_1)Q^m(s_1|\bar{s})$, and $\lambda_{\bar{s},2}^m(s_1,s_2) = Q^m(\bar{s}|s_1)Q^m(s_1|s_2)Q^m(s_2|\bar{s})$, and similarly define $\lambda_{\bar{s},l}^m(s_1,\dots,s_l)$ for $l \geq 3$ as an $(l+1)$-variate product of $Q^m(s'|s)$'s of the form $Q^m(\bar{s}|s_1)\cdots Q^m(s_{l-1}|s_l)Q^m(s_l|\bar{s})$ for $\{s_t\}_{t=1}^l \in S^l$.

PROPOSITION 7: *Suppose Assumption 2 holds. Assume $T \geq 8$ and is even, and define $u = (T-4)/2$. Suppose that $S = \{1, 2, \dots, |S|\}$ and further define*

$$\tilde{\Lambda}_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \tilde{\Lambda}_1 = \begin{bmatrix} \lambda_{\bar{s},1}^1(1) & \cdots & \lambda_{\bar{s},1}^1(|S|) \\ \vdots & \ddots & \vdots \\ \lambda_{\bar{s},1}^M(1) & \cdots & \lambda_{\bar{s},1}^M(|S|) \end{bmatrix}.$$

For $l = 2, \dots, u$, define $\tilde{\Lambda}_l$ to be a matrix whose elements consists of $\lambda_{\bar{s},l}^m(s_1,\dots,s_l)$, covering all possible unordered combinations (with replacement) of $(s_1,\dots,s_l)$ from S^l. For example,

$$\underset{(M\times\binom{|S|+1}{2})}{\tilde{\Lambda}_2} = \left[\begin{matrix} \lambda_{\bar{s},2}^1(1,1) & \cdots & \lambda_{\bar{s},2}^1(1,|S|) \\ \vdots & & \vdots \\ \lambda_{\bar{s},2}^M(1,1) & \cdots & \lambda_{\bar{s},2}^M(1,|S|) \end{matrix}\right. \\ \left.\begin{matrix} \lambda_{\bar{s},2}^1(2,2) & \cdots & \lambda_{\bar{s},2}^1(2,|S|) & \cdots & \lambda_{\bar{s},2}^1(|S|,|S|) \\ \vdots & & \vdots & & \vdots \\ \lambda_{\bar{s},2}^M(2,2) & \cdots & \lambda_{\bar{s},2}^M(2,|S|) & \cdots & \lambda_{\bar{s},2}^M(|S|,|S|) \end{matrix}\right].$$

Define an $M \times \sum_{l=0}^{u}\binom{|S|+l-1}{l}$ matrix $\tilde{\Lambda}$ as $\tilde{\Lambda} = [\tilde{\Lambda}_0, \tilde{\Lambda}_1, \tilde{\Lambda}_2, \dots, \tilde{\Lambda}_u]$ and define $L_{\bar{s}}^{\diamond}$ to be a $M \times M$ matrix consisting of M columns from $\tilde{\Lambda}$ but with the first column unchanged.

Suppose (a) $\sum_{l=0}^{u}\binom{|S|+l-1}{l} \geq M$, (b) $q^{*m}(\bar{s}) > 0$ *for all m*, (c) $L_{\bar{s}}^{\diamond}$ *is nonsingular, and* (d) *there exists $k \in S$ such that $\lambda_{\bar{s}}^{*m}(k) > 0$ for all m and $\lambda_{\bar{s}}^{*m}(k) \neq \lambda_{\bar{s}}^{*n}(k)$ for any $m \neq n$. Then $\{\pi_{\bar{s}}^m, \{\lambda_{\bar{s},l}^m(s_1,\dots,s_l) : (s_1,\dots,s_l) \in S^l\}_{l=1}^u, \{\lambda_{\bar{s}}^{*m}(s)\}_{s=1}^{|S|}\}_{m=1}^M$ is uniquely determined from $\{P(\{s_t\}_{t=1}^T) : \{s_t\}_{t=1}^T \in (A \times X)^T\}$.*

The identification of the primitive parameters π^m, $p^{*m}(a,x)$, $f^m(x'|x,a)$, $P^m(a|x)$ follows from Remark 5(ii) and (iii).

EXAMPLE 5—An Example for Proposition 7: Browning and Carro (2007, Section 4) considered a stationary first-order Markov chain model of $a_{it} \in$

$\{0, 1\}$ without covariates and showed that their model is not nonparametrically identified when $T = 3$ and $M = 9$. In our notation, Browning and Carro's model is written as

$$P(a_1, \dots, a_T) = \sum_{m=1}^{M} \pi^m p^{*m}(a_1) \prod_{t=2}^{T} P^m(a_t|a_{t-1}).$$

Note that $s = a$ because there are no covariates. If $T = 8$, we can identify $M = \sum_{l=0}^{u} \binom{2+l-1}{l} = 6$ types, provided that, for $\bar{s} = \{0, 1\}$, $p^{*m}(\bar{s}) > 0$ for all m, $L_{\bar{s}}^{\diamond}$ is nonsingular, and $P^m(1|\bar{s}) \neq P^n(1|\bar{s})$ for any $m \neq n$. Here, $L_{\bar{s}}^{\diamond}$ is given by

$$L_{\bar{s}}^{\diamond} = \begin{bmatrix} 1 & \lambda_{\bar{s},1}^{1}(0) & \lambda_{\bar{s},1}^{1}(1) & \lambda_{\bar{s},2}^{1}(0,0) & \lambda_{\bar{s},2}^{1}(0,1) & \lambda_{\bar{s},2}^{1}(1,1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_{\bar{s},1}^{M}(0) & \lambda_{\bar{s},1}^{M}(1) & \lambda_{\bar{s},2}^{M}(0,0) & \lambda_{\bar{s},2}^{M}(0,1) & \lambda_{\bar{s},2}^{M}(1,1) \end{bmatrix}$$

and the elements of $L_0^{\diamond}$ are given by, for example,

$$\lambda_{0,1}^{m}(0) = P^m(0|0)P^m(0|0), \quad \lambda_{0,2}^{m}(1,1) = P^m(0|1)P^m(1|1)P^m(1|0).$$

The factorization equations that correspond to (18) are $P_{\bar{s}}^{\diamond} = (L_{\bar{s}}^{\diamond})' V_{\bar{s}} L_{\bar{s}}^{\diamond}$ and $P_{\bar{s},k}^{\diamond} = (L_{\bar{s}}^{\diamond})' D_{k|\bar{s}} V_{\bar{s}} L_{\bar{s}}^{\diamond}$, where $\tilde{V}_{\bar{s}}$ and $D_{k|\bar{s}}$ are defined as before. $P_{\bar{s}}^{\diamond}$ and $P_{\bar{s},k}^{\diamond}$ are identifiable from the data, and we can construct $\tilde{V}_{\bar{s}}$, $D_{k|\bar{s}}$, and $L_{\bar{s}}^{\diamond}$ from these factorization equations. Similarly, if $T = 10, 12, 14$, then the maximum number of identifiable types by Proposition 7 is $10, 15, 21$, respectively.

The following example demonstrates that nonparametric identification of component distributions may help us understand the identification of parametric finite mixture models of dynamic discrete choices.

EXAMPLE 6—Identification of Models With Heterogeneous Coefficients: Consider the model of Example 1. For an individual who belongs to type m, $P^m(a_t = 1|x_t, a_{t-1}) = \Phi(x_t'\beta^m + \rho^m a_{t-1})$ and the initial observation, (a_1, x_1), is randomly drawn from $p^{*m}(a_1, x_1)$ while the transition function of x_t is given by $f^m(x_t|x_{t-1}, a_{t-1})$.

If the conditions in Proposition 6 including $T \geq 6$, $|S| \geq M - 1$, and the rank of $L_{\bar{s}}$ are satisfied, then $p^{*m}(a_1, x_1)$, $f^m(x_t|x_{t-1}, a_{t-1})$, and $P^m(a_t = 1|x_t, a_{t-1})$ are identified for all m. Once $P^m(a_t = 1|x_t, a_{t-1})$ is identified, taking an inverse mapping gives $x_t'\beta^m + \rho^m a_{t-1} = \Phi^{-1}(P^m(a_t = 1|x_t, a_{t-1}))$. Evaluating this at all the points in $A \times X$ gives a system of $|A||X|$ linear equations with $\dim(\beta^m) + 1$ unknown parameters (β^m, ρ^m), and solving this system for (β^m, ρ^m) identifies (β^m, ρ^m).

For instance, consider a model $P^m(a_t = 1|x_t, a_{t-1}) = \Phi(\beta_0^m + \beta_1^m x_t + \rho^m a_{t-1})$ with $A = X = \{0, 1\}$. If $P^m(a_t = 1|x, a)$ is identified for all $(a, x) \in A \times X$, then the type-specific coefficient $(\beta_0^m, \beta_1^m, \rho^m)$ is identified as the unique solution to the linear system

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0^m \\ \beta_1^m \\ \rho^m \end{bmatrix} = \begin{bmatrix} \Phi^{-1}(P^m(a_t = 1|0, 0)) \\ \Phi^{-1}(P^m(a_t = 1|0, 1)) \\ \Phi^{-1}(P^m(a_t = 1|1, 0)) \\ \Phi^{-1}(P^m(a_t = 1|1, 1)) \end{bmatrix}.$$

The next example shows that the degree of underidentification in structural dynamic models *with* unobserved heterogeneity can be reduced to that in models *without* unobserved heterogeneity. Furthermore, a researcher can now apply various two-step estimators for structural models developed by Hotz and Miller (1993) (and others listed in the Introduction) to models with unobserved heterogeneity since, with our identification results, one can obtain an initial nonparametric consistent estimate of type-specific component distributions. Kasahara and Shimotsu (2008a) provided an example of such an application.[8]

EXAMPLE 7—Dynamic Discrete Games (Aguirregabiria and Mira (2007, Section 3.5)): Consider the model of dynamic discrete games with unobserved market characteristics studied by Aguirregabiria and Mira (2007, Section 3.5). There are N ex ante identical "global" firms competing in H local markets. There are M market types and each market's type is common knowledge to all firms, but unknown to a researcher. In market $h \in \{1, 2, \dots, H\}$, firm i maximizes the expected discounted sum of profits $E[\sum_{s=t}^{\infty} \beta^{s-t}\{\Pi_i(x_{hs}, a_{hs}, a_{h,s-1}; \theta_h) + \varepsilon_{his}(a_{his})\}|x_{ht}, a_{ht}, a_{h,t-1}; \theta_h]$, where x_{ht} is a state variable that is common knowledge for all firms, $\theta_h \in \{1, 2, \dots, M\}$ is the type attribute of market h, $a_{hs} = (a_{h1s}, \dots, a_{hNs})$ is the vector of firms' decisions, and $\varepsilon_{hit}(a_{hit})$ is a state variable that is private information to firm i. The profit function may depend on, for example, the past entry/exit decision of firms. The researcher observes x_{ht} and a_{ht}, but neither θ_h nor ε_{hit}. There is no interaction across different markets.

Let a^{-1} denote the vector of firms' decision in the preceding period. Assume that the ε_i's are independent from x and independent and identically distributed (i.i.d.) across firms. Let $\sigma^*(\theta_h) = \{\sigma_i^*(x, a^{-1}, \varepsilon_i; \theta_h) : i = 1, \dots, N\}$ denote a set of strategy functions in a stationary Markov perfect equilibrium (MPE). Then, the equilibrium conditional choice probabilities are given by $P_i^{\sigma^*}(a_i|x, a^{-1}; \theta_h) = \int 1\{a_i = \sigma_i^*(x, a^{-1}, \varepsilon_i; \theta_h)\}g(\varepsilon_i)\, d\varepsilon_i$, where $g(\varepsilon_i)$ is the density function for $\varepsilon = \{\varepsilon(a) : a \in A\}$. A MPE induces a transition function of x, which we denote by $f^{\sigma^*}(x_t|x_{t-1}, a^{-1}; \theta_h)$.

[8]Kasahara and Shimotsu (2008a) showed that in structural discrete Markov decision models with unobserved heterogeneity, it is possible to obtain an estimator that is higher-order equivalent to the maximum likelihood estimator (MLE) by iterating the nested pseudo-likelihood (NPL) algorithm of Aguirregabiria and Mira (2002) sufficiently many, but finite times.

Suppose that panel data $\{\{a_{ht}, x_{ht}\}_{t=1}^{T}\}_{h=1}^{H}$ are available. As in Aguirregabiria and Mira (2007), consider the case where $H \to \infty$ with N and T fixed. The initial distribution of (a, x) differs across market types and is given by $p^{*m}(a, x)$. Let $P^m(a_{ht}|x_{ht}, a_{h,t-1}) = \prod_{i=1}^{N} P_i^{\sigma^*}(a_{hit}|x_{ht}, a_{h,t-1}; m)$ and $f^m(x_{ht}|x_{h,t-1}, a_{h,t-1}) = f^{\sigma^*}(x_{ht}|x_{h,t-1}, a_{h,t-1}; m)$. Then the likelihood function for market h becomes a mixture across different unobserved market types,

$$P(\{a_{ht}, x_{ht}\}_{t=1}^{T}) = \sum_{m=1}^{M} \pi^m p^{*m}(a_{h1}, x_{h1}) \prod_{t=2}^{T} P^m(a_{ht}|x_{ht}, a_{h,t-1}) f^m(x_{ht}|x_{h,t-1}, a_{h,t-1}),$$

for which Propositions 6 and 7 are applicable.

EXAMPLE 8—An Empirical Example of Aguirregabiria and Mira (2007, Section 5): Aguirregabiria and Mira (2007, Section 5) considered an empirical model of entry and exit in local retail markets based on the model of Example 7. Each market is indexed by h. The firms' profits depend on the logarithm of the market size, x_{ht}, and their current and past entry/exit decisions, a_{ht} and $a_{h,t-1}$. The profit of a nonactive firm is zero. When active, firm i's profit is $\Pi_i^o(x_{ht}, a_{ht}, a_{h,t-1}) + \omega_h + \varepsilon_{hit}(a_{hit})$, where the function Π_i^o is common across all the markets and ε_{hit} is i.i.d. across markets. The parameter ω_h captures the unobserved market characteristics and has a discrete distribution with 21 points of support. The logarithm of market size follows an exogenous first-order Markov process.

Their panel covers 6 years at annual frequency (i.e., $T = 6$). This satisfies the requirement for T in Proposition 6. Given the nonlinear nature of the dynamic games models, the rank condition on the $L_{\bar{s}}$ matrix in Proposition 6 is likely to be satisfied. In their specification however, the transition function for x_{ht}, $f_h(x_{ht}|x_{h,t-1})$, is market-specific, so that the number of types is equal to the number of markets. Consequently, Proposition 6 does not apply to this case, and the type-specific conditional choice probabilities may not be nonparametrically identified.[9]

If we limit the number of types for transition functions, then we may apply Proposition 6 to identify the type-specific conditional choice probabilities and the type-specific transition probabilities. For the mth type market, the joint conditional choice probabilities across all firms are $P^m(a_{ht}|x_{ht}, a_{h,t-1}) =$

[9]Aguirregabiria and Mira estimated the market-specific transition function using 14 years of data on market size from other data sources. Given the relatively long time length, the identification of market-specific transition functions may come from the time variation of each market. When the transition function is market-specific, however, the conditional choice probabilities also become market-specific, leading to the incidental parameter problem. *Even if the market-specific transition function is known*, nonparametrically identifying the market-specific conditional choice probabilities is not possible given a short panel.

$\prod_{i=1}^{N} P_i(a_{hit}|x_{ht}, a_{h,t-1}; \theta^m)$, where $a_{ht} = (a_{h1t}, \ldots, a_{hNt})'$. The market size is discretized with 10 support points (i.e., $|X| = 10$) and $A = \{0, 1\}^N$. Consequently, the size of the state space of $s_{ht} = (a_{ht}, x_{ht})$ for this model is $|A||X| = 2^N \times 10$, and we may identify up to $M = 2^N \times 10 + 1$ types. For instance, their Table VI reports that 63.5 percent of markets have no less than five potential firms (i.e., $N \geq 5$) in the restaurant industry. Hence, $M = 2^5 \times 10 + 1 = 321$ market types can be identified for these markets.[10]

The following proposition extends Proposition 3 for identification of M under Assumption 2. Because of the state dependence, the required panel length becomes $T = 5$. We omit the proof because it is essentially the same as that of Proposition 3.

PROPOSITION 8: *Suppose that Assumption 2 holds. Assume $T \geq 5$ and $S = \{1, \ldots, |S|\}$. Fix $s_1 = s_3 = s_5 = \bar{s} \in S$. Define, for $s, s' \in S$,*

$$P_{\bar{s}}(s) = P(s_2 = s, s_1 = s_3 = \bar{s}),$$
$$P_{\bar{s}}(s, s') = P\big((s_2, s_4) = (s, s'), s_1 = s_3 = s_5 = \bar{s}\big),$$

and define a $(|S| + 1) \times (|S| + 1)$ matrix

$$P^*_{\bar{s}} = \begin{bmatrix} 1 & P_{\bar{s}}(1) & \cdots & P_{\bar{s}}(|S|) \\ P_{\bar{s}}(1) & P_{\bar{s}}(1, 1) & \cdots & P_{\bar{s}}(1, |S|) \\ \vdots & \vdots & \ddots & \vdots \\ P_{\bar{s}}(|S|) & \tilde{P}_{\bar{s}}(|S|, 1) & \cdots & P_{\bar{s}}(|S|, |S|) \end{bmatrix}.$$

Suppose $q^{*m}(\bar{s}) > 0$ for all m. Then $M \geq \operatorname{rank}(P^*_{\bar{s}})$. Furthermore, if the matrix $L^*_{\bar{s}}$ defined below has rank M, then $M = \operatorname{rank}(P^*_{\bar{s}})$:

$$\underset{(M\times(|S|+1))}{L^*_{\bar{s}}} = \begin{bmatrix} 1 & \lambda^1_{\bar{s}}(1) & \cdots & \lambda^1_{\bar{s}}(|S|) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda^M_{\bar{s}}(1) & \cdots & \lambda^M_{\bar{s}}(|S|) \end{bmatrix}.$$

In some applications, the model has two types of covariates, z_t and x_t, where the transition function of x_t depends on types, while the transition function of z_t is common across types. In such a case, we may use the variation of z_t as a main source of identification and relax the requirement on T in Proposition 6.

We impose an assumption analogous to Assumption 2, as well as the conditional independence assumption on the transition function of (x', z'):

[10]Even if the type-specific conditional choice probabilities are nonparametrically identified, it is generally not possible to nonparametrically identify the primitive objects such as the discount factor β, profit functions Π_i, and the distribution of shocks, ε_{iht}, in structural dynamic models (Rust (1994), Magnac and Thesmar (2002)).

ASSUMPTION 3: (a) *The choice probability of* a_t *does not depend on time and is independent of* z_{t-1}. (b) *The transition function of* (x_t, z_t) *conditional on* $\{x_\tau, z_\tau, a_\tau\}_{\tau=1}^{t-1}$ *takes the form* $g(z_t|x_{t-1}, z_{t-1}, a_{t-1})f^m(x_t|x_{t-1}, a_{t-1})$ *for all* t. (c) $f^m(x'|x, a) > 0$ *for all* $(x', x, a) \in X \times X \times A$ *and* $g(z'|x, z, a) > 0$ *for all* $(z', x, z, a) \in Z \times X \times Z \times A$ *and for* $m = 1, \dots, M$.

Under Assumption 3, consider a model

$$
\begin{aligned}
&P(\{a_t, x_t, z_t\}_{t=1}^T) \\
&\quad = \sum_{m=1}^M \pi^m p^{*m}(x_1, z_1, a_1) \prod_{t=2}^T g(z_t|x_{t-1}, z_{t-1}, a_{t-1}) f^m(x_t|x_{t-1}, a_{t-1}) \\
&\qquad\qquad \times P^m(a_t|x_t, x_{t-1}, a_{t-1}, z_t).
\end{aligned}
$$

Assuming $g(z_t|x_{t-1}, z_{t-1}, a_{t-1})$ is known and defining $s_t = (a_t, x_t)$, $\tilde{q}^{*m}(s_1, z_1) = p^{*m}(x_1, z_1, a_1)$, and $\tilde{Q}^m(s_t|s_{t-1}, z_t) = f^m(x_t|x_{t-1}, a_{t-1})P^m(a_t|x_t, x_{t-1}, a_{t-1}, z_t)$, we can write this equation as

$$
\begin{aligned}
\text{(29)} \qquad \tilde{P}(\{s_t, z_t\}_{t=1}^T) &= \frac{P(\{a_t, x_t, z_t\}_{t=1}^T)}{\prod_{t=2}^T g(z_t|x_{t-1}, z_{t-1}, a_{t-1})} \\
&= \sum_{m=1}^M \pi^m \tilde{q}^{*m}(s_1, z_1) \prod_{t=2}^T \tilde{Q}^m(s_t|s_{t-1}, z_t).
\end{aligned}
$$

We fix the value of $\{s_t\}_{t=1}^T$ and use the "independent" variation of z_t to identify unobserved types. The next proposition provides a sufficient condition for nonparametric identification of the model (29). Define, for $\bar{s} \in S$ and $h, \xi \in Z$,

$$
\tilde{\pi}_{\bar{s},h}^m = \pi^m \tilde{q}^{*m}(\bar{s}, h), \quad \tilde{\lambda}_{\bar{s}}^m(\xi) = \tilde{Q}^m(\bar{s}|\bar{s}, \xi).
$$

PROPOSITION 9: *Suppose that Assumption* 3 *holds and assume* $T \geq 4$. *Define*

$$
\underset{(M\times M)}{\bar{L}_{\bar{s}}} = \begin{bmatrix} 1 & \tilde{\lambda}_{\bar{s}}^1(\xi_1) & \cdots & \tilde{\lambda}_{\bar{s}}^1(\xi_{M-1}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \tilde{\lambda}_{\bar{s}}^M(\xi_1) & \cdots & \tilde{\lambda}_{\bar{s}}^M(\xi_{M-1}) \end{bmatrix}.
$$

Suppose that $\tilde{q}^{*m}(\bar{s}, h) > 0$ *for all* m, *there exist some* $\{\xi_1, \dots, \xi_{M-1}\}$ *such that* $\bar{L}_{\bar{s}}$ *is nonsingular, and there exist* $(r, k) \in S \times Z$ *such that* $\tilde{Q}^m(r|\bar{s}, k) > 0$ *for all* m *and* $\tilde{Q}^m(r|\bar{s}, k) \neq \tilde{Q}^n(r|\bar{s}, k)$ *for any* $m \neq n$. *Then* $\{\tilde{\pi}_{\bar{s},h}^m, \{\lambda_{\bar{s}}^m(\xi)\}_{\xi\in Z}, \{\tilde{Q}^m(s|\bar{s}, \xi)\}_{(s,\xi)\in S\times Z}\}_{m=1}^M$ *is uniquely determined from* $\{\tilde{P}(\{s_t, z_t\}_{t=1}^T) : \{s_t, z_t\}_{t=1}^T \in (S \times Z)^T\}$.

We may identify the primitive parameters π^m, $p^{*m}(a,x)$, $f^m(x'|x,a)$, $P^m(a|x)$ using an argument analogous to those of Remark 5(ii) and (iii). The requirement of $T=4$ in Proposition 9 is weaker than that of $T=6$ in Proposition 6 because the variation of z_t, rather than (x_t,a_t), is used as a main source of identification. When $T>4$, we may apply the argument of Proposition 2 to relax the sufficient condition for identification in Proposition 9, but we do not pursue it here; Proposition 2 provides a similar result.

3.3. *Limited Transition Pattern*

This section analyzes the identification condition of the baseline model when Assumption 1(c) is relaxed. In some applications, the transition pattern of x is limited, as not all $x'\in X$ are reachable with a positive probability. In such instances, the set of sequences $\{a_t,x_t\}_{t=1}^T$ that can be realized with a positive probability also becomes limited and the number of restrictions from a set of the submodels falls, making identification more difficult.

EXAMPLE 9—Bus Engine Replacement Model (Rust (1987)): Suppose $a\in\{0,1\}$ is the replacement decision for a bus engine, where $a=1$ corresponds to replacing a bus engine. Let x denote the mileage of a bus engine with $X=\{1,2,\dots\}$. The transition function of x_t is

$$f(x_{t+1}|x_t,a_t;\theta)=\begin{cases}\theta_{f,1}, & \text{for } x_{t+1}=(1-a_t)x_t+a_t,\\ \theta_{f,2}, & \text{for } x_{t+1}=(1-a_t)x_t+a_t+1,\\ 1-\theta_{f,1}-\theta_{f,2}, & \text{for } x_{t+1}=(1-a_t)x_t+a_t+2,\\ 0, & \text{otherwise,}\end{cases}$$

and not all $x'\in X$ can be realized from (x,a).

Henceforth, we assume the transition function of x is stationary and takes the form $f(x'|x,a)$ to simplify the exposition. If $f(x'|x,a)=0$ for some (x',x,a) and not all $x'\in X$ can be reached from (a,x), then some values of $\{a_t,x_t\}$ are never realized. For such values, $\tilde{P}(\{a_t,x_t\}_{t=1}^T)$ in (9) and its lower-dimensional submodels in (10) and (11) are not well defined. Hence, their restrictions cannot be used for identification. Thus, we fix the values of (a_1,x_1) and (a_τ,x_τ), and focus on the values of (a_t,x_t) that are realizable between (a_1,x_1) and (a_τ,x_τ). The difference in response patterns between (a_1,x_1) and (a_τ,x_τ) provides a source of identification.

To fix the idea, assume $T=4$, and fix $a_t=0$ for all t, $x_1=h$, and $x_\tau=k$. Of course, it is possible to choose different sequences of $\{a_t\}_{t=1}^T$. Let B^h and C^h be subsets of X, of which elements are realizable between (a_1,x_1) and (a_τ,x_τ). We use the variations of x within B^h and C^h as a source of identification. Define, for $h,\xi\in X$,

$$\tilde{\pi}_h^m=\pi^m p^{*m}(a_1=0,x_1=h)\quad\text{and}\quad \tilde{\lambda}_\xi^m=P^m(a=0|x=\xi), \tag{30}$$

and $\tilde{V}_h = \text{diag}(\tilde{\pi}_h^1, \dots, \tilde{\pi}_h^M)$ and $\tilde{D}_k = \text{diag}(\tilde{\lambda}_k^1, \dots, \tilde{\lambda}_k^M)$. We identify $\tilde{V}_h$, $\tilde{D}_k$, and $\tilde{\lambda}_\xi^m$'s from the factorization equations corresponding to (18):

$$P^h = \tilde{L}_b' \tilde{V}_h \tilde{L}_c \quad \text{and} \quad P_k^h = \tilde{L}_b' \tilde{D}_k \tilde{V}_h \tilde{L}_c, \tag{31}$$

where $\tilde{L}_b$ and $\tilde{L}_c$ are defined analogously to L in (13), but using $\tilde{\lambda}_\xi^m$, and with $\xi \in B^h$ and $\xi \in C^h$, respectively. As we discuss below, we choose B^h, C^h, and k so that P^h and P_k^h are identifiable from the data.

Each equation of these factorization equations (31) represents a submodel in (9) and (10) for a sequence of a_t's and x_t's that belongs to one of the sets

$$\begin{aligned} A_1 &= \{x_1 = h, (x_2, x_3) \in B^h \times C^h, x_4 = k; a_t = 0 \text{ for all } t\}, \\ A_2 &= \{x_1 = h, x_2 \in B^h, x_3 = k; a_t = 0 \text{ for all } t\}, \\ A_3 &= \{x_1 = h, x_2 \in C^h, x_3 = k; a_t = 0 \text{ for all } t\}, \\ A_4 &= \{x_1 = h, x_2 = k; a_t = 0 \text{ for all } t\}. \end{aligned} \tag{32}$$

For instance, a submodel for a sequence $q_1 \in A_1$ in (32) is

$$\begin{aligned} \tilde{P}(q_1) &= \frac{P(q_1)}{f(k|x_3, 0) f(x_3|x_2, 0) f(x_2|h, 0)} \\ &= \sum_{m=1}^{M} \pi^m p^{*m}(h, 0) P^m(0|x_2) P^m(0|x_3) P^m(0|k) \\ &\qquad \text{for } (x_2, x_3) \in B^h \times C^h, \end{aligned}$$

which represents one of the equations of $P_k^h = \tilde{L}_b' \tilde{D}_k \tilde{V}_h \tilde{L}_c$ in (31).

For all the submodels implied by (31) to provide identifying restrictions, all the sequences of x_t's in A_1–A_4 in (32) must have positive probability; otherwise, some elements of P^h and P_k^h in (31) cannot be constructed from the data, and our identification strategy fails. This requires that all the points in B^h must be reachable from h, while all the points in C^h must be reachable from h and all the points in B^h. Finally, k must be reachable from h and all the points in B^h and C^h.

EXAMPLE 9—Continued: In Example 9, assume the initial distribution $p^{*m}(x, a)$ is defined as the type-specific stationary distribution. Set $a_t = 0$ for $t = 1, \dots, 4$ and $x_1 = h$. Choose $B^h = \{h, h+1\}$ and $C^h = \{h+1, h+2\}$, and $k = h+2$. For this choice of B^h, C^h, and k, the corresponding transition probabilities are nonzero, and we may construct all the elements of P^h and P_k^h in (31) from the observables. For each $h \in X$, these submodels provide $4 + 3 + 1 = 8$ restrictions for identification.

We now state the restrictions on B^h and C^h formally. First we develop useful notation. For a singleton $\{x\} \subset X$, let $\Gamma(a, \{x\}) = \{x' \in X : f(x'|x, a) > 0\}$ denote a set of $x' \in X$ that can be reached from (a, x) in the next period with a positive probability. For a subset $W \subseteq X$, define $\Gamma(a, W)$ as the intersection of $\Gamma(a, \{x\})$'s across all x's in $W : \Gamma(a, W) = \bigcap_{x \in W} \Gamma(a, \{x\})$.

We summarize the assumptions of this subsection including the restrictions on B^h and C^h:

ASSUMPTION 4: (a) *The choice probability of a_t does not depend on time.* (b) *The choice probability of a_t is independent of the lagged variable (x_{t-1}, a_{t-1}) conditional on x_t.* (c) $P^m(a|x) > 0$ *for all* $(a, x) \in A \times X$ *and* $m = 1, \dots, M$. (d) $f_t^m(x_t|\{x_\tau, a_\tau\}_{\tau=1}^{t-1}) = f(x_t|x_{t-1}, a_{t-1})$ *for all* m. (e) $h, k \in X$, B^h *and* C^h *satisfy* $p^{*m}(a_1 = 0, x_1 = h) > 0$ *for all* m, *and*

$$B^h \subseteq \Gamma(0, \{h\}), \quad C^h \subseteq \Gamma(0, B^h) \cap \Gamma(0, \{h\}),$$
$$\{k\} \subseteq \Gamma(0, C^h) \cap \Gamma(0, B^h) \cap \Gamma(0, \{h\}).$$

Assumption 4(a) and (b) are identical to Assumption 1(a) and (b). Assumption 4(c) is necessary for the submodels to be well defined. Assumption 4(d) strengthens Assumption 1(d) by imposing stationarity and a first-order Markov property. It may be relaxed, but doing so would add substantial notational complexity. Assumption 4(e) guarantees that all the sequences we consider in the subsets in (32) have nonzero probability. Note that the choice of C^h is affected by how B^h is chosen. If Assumption 1(c) holds, it is possible to set $B^h = C^h = X$.

The next proposition provides a sufficient condition for identification under Assumption 4.

PROPOSITION 10: *Suppose that Assumption 4 holds $T = 4$, and $|B^h|$, $|C^h| \geq M - 1$. Let $\{\xi_1^b, \dots, \xi_{M-1}^b\}$ and $\{\xi_1^c, \dots, \xi_{M-1}^c\}$ be elements of B^h and C^h, respectively. Define $\tilde{\pi}_h^m$ and $\tilde{\lambda}_\xi^m$ as in (30), and define*

$$\underset{(M\times M)}{\tilde{L}_b} = \begin{bmatrix} 1 & \tilde{\lambda}^1_{\xi_1^b} & \tilde{\lambda}^1_{\xi_2^b} & \cdots & \tilde{\lambda}^1_{\xi_{M-1}^b} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \tilde{\lambda}^M_{\xi_1^b} & \tilde{\lambda}^M_{\xi_2^b} & \cdots & \tilde{\lambda}^M_{\xi_{M-1}^b} \end{bmatrix},$$
$$\underset{(M\times M)}{\tilde{L}_c} = \begin{bmatrix} 1 & \tilde{\lambda}^1_{\xi_1^c} & \tilde{\lambda}^1_{\xi_2^c} & \cdots & \tilde{\lambda}^1_{\xi_{M-1}^c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \tilde{\lambda}^M_{\xi_1^c} & \tilde{\lambda}^M_{\xi_2^c} & \cdots & \tilde{\lambda}^M_{\xi_{M-1}^c} \end{bmatrix}.$$

Suppose that $\tilde{L}_b$ and $\tilde{L}_c$ are nonsingular for some $\{\xi_1^b, \dots, \xi_{M-1}^b\}$ and $\{\xi_1^c, \dots, \xi_{M-1}^c\}$, and that $\tilde{\lambda}_k^m > 0$ for all m and $\tilde{\lambda}_k^m \neq \tilde{\lambda}_k^n$ for any $m \neq n$. Then $\{\tilde{\pi}_h^m, \tilde{\lambda}_\xi^m : \xi \in B^h \cup C^h\}_{m=1}^M$ is uniquely determined from $\{\tilde{P}(\{a_t, x_t\}_{t=1}^T) : \{a_t, x_t\}_{t=1}^T \in (A \times X)^T\}$.

Assuming that all the values of x can be realized in the initial period, we may repeat the above argument for all possible values of x_1 to identify $\tilde{\lambda}^m_\xi$ for any $\xi \in \bigcup_{h\in X} B^h$. Furthermore, we can repeat the argument for different sequences of $\{a_t\}_{t=1}^4$ to increase the identifiable elements of $P^m(a|x)$. For instance, by choosing $B^h = \Gamma(a, \{h\})$, $\tilde{\lambda}^m_l$ is identified for all $l \in X$ if the union of $\Gamma(a, \{h\})$ across different $(a, h) \in A \times X$ includes all the elements of X so that $X = \bigcup_{(a,h)\in A\times X} \Gamma(a, \{h\})$. This is a weak condition and is satisfied if X is an ergodic set. However, setting $B^h = \Gamma(a, \{h\})$ may lead to a small number of identifiable types.

EXAMPLE 9—Continued: Setting $a_t = 0$ for $t = 1, \dots, 4$, we have $\Gamma(0, \{h\}) = \{h, h+1, h+2\}$ for any $h \in X$. To satisfy Assumption 4(e), choose $B^h = \{h, h+1\}$, $C^h = \{h+1, h+2\}$, and $k = h+2$. If the other assumptions of Proposition 10 are satisfied, we can identify $M = 3$ types. From the factorization equations (31), we can uniquely determine $\tilde{V}_h$, $\tilde{D}_k$, $\tilde{L}_b$, and $\tilde{L}_c$, and identify $\{\pi^m p^{*m}(0, x), P^m(0|x) : x = h, h+1, h+2\}_{m=1,2,3}$. Repeating for all $h \in X$, we identify $P^m(a|x)$ for all $(a, x) \in A \times X$. We then identify $p^{*m}(x, a)$ using $P^m(a|x)$, $f(x'|x, a)$, and the fixed point constraint, while π^m is determined as $\pi^m p^{*m}(0, x)/p^{*m}(0, x)$.

The sufficient condition of Proposition 10 does not allow one to identify many types when the size of B^h or C^h is small. It is possible to identify more types when we can find a subset D of X that is reachable from itself, namely $D \subseteq \Gamma(0, D)$. For example, if the transition pattern is such that $\Gamma(0, \{x\}) = \{x-2, x-1, x, x+1, x+2\}$ for some $x \in X$, then the set $\{x-1, x, x+1\}$ serves as D. In such cases, we can apply the logic of Proposition 2 to identify many types if $T \geq 5$.

ASSUMPTION 5: (a) *Assumptions* 4(a)–(d) *hold*. (b) *A subset D of X satisfies $D \subseteq \Gamma(0, D)$.*

Set $D = \{d_1, \dots, d_{|D|}\}$, and define $\lambda^{*m}_d = p^{*m}((a, x) = (1, d))$ and $\lambda^m_d = P^m(a = 1|x = d)$ for $d \in D$. Under Assumption 5, replacing X with D and simply repeating the proof of Proposition 2 gives the following proposition:

PROPOSITION 11: *Suppose Assumption 5 holds. Assume $T \geq 5$ is odd and define $u = (T-1)/2$. Define Λ_r, $r = 0, \dots, u$, analogously to Proposition 2 except $(X, \lambda^{*m}_{\xi_j}, \lambda^m_{\xi_j})$ is replaced with $(D, \lambda^{*m}_{d_j}, \lambda^m_{d_j})$. Define an $M \times (\sum_{l=0}^{u} \binom{|D|+l-1}{l})$ matrix Λ as $\Lambda = [\Lambda_0, \Lambda_1, \Lambda_2, \dots, \Lambda_u]$.*

Suppose (a) $\sum_{l=0}^{u} \binom{|D|+l-1}{l} \geq M$, (b) *we can construct a nonsingular $M \times M$ matrix $L^\diamond$ by setting its first column as Λ_0 and choosing the other $M-1$ columns from the columns of Λ but Λ_0, and* (c) *there exists $d_k \in D$ such that $\lambda^{*m}_{d_k} > 0$ for all m and $\lambda^{*m}_{d_k} \neq \lambda^{*n}_{d_k}$ for any $m \neq n$. Then $\{\pi^m, \{\lambda^{*m}_{d_j}, \lambda^m_{d_j}\}_{j=1}^{|D|}\}_{m=1}^M$ is uniquely determined from $\{\tilde{P}(\{a_t, x_t\}_{t=1}^T) : \{a_t, x_t\}_{t=1}^T \in (A \times X)^T\}$.*

For example, if $|D| = 3$ and $T = 5$, the number of identifiable types becomes $\binom{2}{0} + \binom{3}{1} + \binom{4}{2} = 10$. Identifying more types is also possible when the model has an additional covariate z_t whose transition pattern is not limited and there is a state $\bar{x}$ such that $P(x_1 = \cdots = x_T = \bar{x}) > 0$ for some sequence of a_t. Then, for $x = \bar{x}$, we can use the variation of z_t and apply Proposition 9. This increases the number of identifiable types to $|Z| + 1$.

4. CONCLUDING REMARK

This paper studies dynamic discrete choice models with unobserved heterogeneity that is represented in the form of finite mixtures. It provides sufficient conditions under which such models are identified without parametric distributional assumptions.

While we emphasize that the variation in the covariate and in time provides important identifying information, our identification approach does require assumptions on the Markov property, stationarity, and type-invariance in transition processes. To clarify our identification results, consider a general nonstationary finite mixture model of dynamic discrete choices:

$$(33)\qquad P(\{a_t, x_t\}_{t=1}^T) = \sum_{m=1}^M \pi^m p^{*m}(x_1, a_1) \prod_{t=2}^T f_t^m(x_t | \{x_\tau, a_\tau\}_{\tau=1}^{t-1}) P_t^m(a_t | x_t, \{x_\tau, a_\tau\}_{\tau=1}^{t-1}).$$

Such a general mixture model (33) cannot be nonparametrically identified without imposing further restrictions.[11] One possible nonparametric restriction is a first-order Markovian assumption on (x_t, a_t), that yields a less general nonstationary model:

$$(34)\qquad P(\{a_t, x_t\}_{t=1}^T) = \sum_{m=1}^M \pi^m p^{*m}(x_1, a_1) \prod_{t=2}^T f_t^m(x_t | x_{t-1}, a_{t-1}) P_t^m(a_t | x_t, x_{t-1}, a_{t-1}).$$

We do not know whether this model is nonparametrically identified without additional assumptions. Section 3.1 shows that the identification of the nonsta-

[11]The model (33) is equivalent to a mixture model $P(\{a_t, x_t\}_{t=1}^T) = \sum_{m=1}^M \pi^m P^m(\{a_t, x_t\}_{t=1}^T)$, because it is always to possible to decompose

$$P^m(\{a_t, x_t\}_{t=1}^T) = p^{*m}(x_1, a_1) \prod_{t=2}^T f_t^m(x_t | \{x_\tau, a_\tau\}_{\tau=1}^{t-1}) P_t^m(a_t | x_t, \{x_\tau, a_\tau\}_{\tau=1}^{t-1}).$$

The number of restrictions implied by $P(\{a_t, x_t\}_{t=1}^T)$ is $(|A||X|)^T - 1$, while the number of unknowns in $\sum_{m=1}^M \pi^m P^m(\{a_t, x_t\}_{t=1}^T)$ is $M - 1 + M((|A||X|)^T - 1)$.

tionary model (34) is possible under the assumptions of type-invariant transition processes and conditional independence of discrete choices. In Section 3.2, we provide identification results under the stationarity assumption on the transition function and choice probabilities in (34). Relaxing these identifying assumptions as well as investigating identifiability, or perhaps nonidentifiability, of finite mixture model (34) is an important future research area.

Estimation and inference on the number of components (types), M, is an important topic because of the lack of guidance from economic theory. It is known that the likelihood ratio statistic has a nonstandard limiting distribution when applied to testing the number of components of a mixture model (see, for example, Liu and Shao (2003)). Leroux (1992) considered a maximum-penalized-likelihood estimator for the number of components, which includes the Akaike information criterion and Bayesian information criterion as a special case. McLachlan and Peel (2000, Chapter 6) surveyed the methods for determining the number of components in parametric mixture models. To our best knowledge, all of these existing methods assume that the component distributions belong to a parametric family. Developing a method for testing and selecting the number of components without imposing any parametric assumption warrants further research.

A statistical test of the number of components may be possible by testing the rank of matrix P^* in Proposition 3. When the covariate has a large number of support points, we may test the number of components by testing a version of matrix P in (17) across different partitions of X. In Kasahara and Shimotsu (2008b), we pursued this idea, and proposed a selection procedure for the number of components by sequentially testing the rank of matrices.

APPENDIX: PROOFS

PROOF OF PROPOSITION 1 AND COROLLARY 1: Define $V = \mathrm{diag}(\pi^1, \dots, \pi^M)$ and $D_k = \mathrm{diag}(\lambda_k^{*1}, \dots, \lambda_k^{*M})$ as in (13). Define P and P_k as in (17). Then P and P_k are expressed as (see (14)–(16))

$$P = L'VL, \quad P_k = L'VD_kL.$$

We now uniquely determine L, V, and D_k from P and P_k constructively. Since L is nonsingular, we can construct a matrix $A_k = P^{-1}P_k = L^{-1}D_kL$. Because $A_kL^{-1} = L^{-1}D_k$, the eigenvalues of A_k determine the diagonal elements of D_k while the right eigenvectors of A_k determine the columns of L^{-1} up to multiplicative constants; denote the right eigenvectors of A_k by $L^{-1}K$, where K is some diagonal matrix. Now we can determine VK from the first row of $PL^{-1}K$ because $PL^{-1}K = L'VK$ and the first row of L' is a vector of ones. Then L' is determined uniquely by $L' = (PL^{-1}K)(VK)^{-1} = (L'VK)(VK)^{-1}$.

Having obtained L', we may determine V from the first column of $(L')^{-1}P$ because $(L')^{-1}P = VL$ and the first column of L is a vector of ones. Therefore, we identify $\{\pi^m, \{\lambda^m_{\xi_j}\}_{j=1}^{M-1}\}_{m=1}^M$ as the elements of V and L.

Once V and L are determined, we can uniquely determine $D_\zeta = \mathrm{diag}(\lambda^{*1}_\zeta, \dots, \lambda^{*M}_\zeta)$ for any $\zeta \in X$ by constructing P_ζ in the same way as P_k and using the relationship $D_\zeta = (L'V)^{-1}P_\zeta L^{-1}$. Furthermore, for arbitrary $\zeta, \xi_j \in X$, evaluate F_{x_2,x_3}, F_{x_2}, and F_{x_3} defined in (15) and (16) at $(x_2, x_3) = (\zeta, \xi_j)$, and define

$$\underset{(M\times 2)}{L^\zeta} = \begin{bmatrix} 1 & \lambda^1_\zeta \\ \vdots & \vdots \\ 1 & \lambda^M_\zeta \end{bmatrix}, \quad \underset{(2\times M)}{P^\zeta} = \begin{bmatrix} 1 & F_{\xi_1} & \dots & F_{\xi_{M-1}} \\ F_\zeta & F_{\zeta,\xi_1} & \dots & F_{\zeta,\xi_{M-1}} \end{bmatrix}. \tag{35}$$

Since $P^\zeta = (L^\zeta)'VL$, we can uniquely determine $(L^\zeta)' = P^\zeta(VL)^{-1}$. Therefore, $\{\lambda^{*m}_\zeta\}_{m=1}^M$ and $\{\lambda^m_\zeta\}_{m=1}^M$ are identified for any $\zeta \in X$. This completes the proof of Proposition 1, and Corollary 1 follows immediately. *Q.E.D.*

PROOF OF PROPOSITION 2: The proof is similar to the proof of Proposition 1. Let $\mathcal{T} = (\tau_2, \dots, \tau_p)$, $2 \le p \le T$, be a subset of $\{2, \dots, T\}$. Let $\mathcal{X}(\mathcal{T})$ be a subset of $\{x_t\}_{t=2}^T$ with $t \in \mathcal{T}$. For example, if $\mathcal{T} = \{2, 4, 6\}$, then $\mathcal{X}(\mathcal{T}) = \{x_2, x_4, x_6\}$. Starting from $\tilde{P}(\{a_t, x_t\}_{t=1}^T)$, integrating out (a_t, x_t) if $t \notin \mathcal{T}$, and evaluating it at $(a_1, x_1) = (1, k)$ and $a_t = 1$ for $t \in \mathcal{T}$ gives a "marginal" $F^*_{k,\mathcal{X}(\mathcal{T})} = \tilde{P}(\{a_1, x_1\} = \{1, k\}, \{1, x_t\}_{\tau\in\mathcal{T}}) = \sum_{m=1}^M \pi^m \lambda^{*m}_k \prod_{t\in\mathcal{T}} \lambda^m_{x_t}$. For example, if $\mathcal{T} = \{2, 4, 6\}$, then $F^*_{k,\mathcal{X}(\mathcal{T})} = \sum_{m=1}^M \pi^m \lambda^{*m}_k \lambda^m_{x_2} \lambda^m_{x_4} \lambda^m_{x_6}$. Integrating out (a_1, x_1) additionally and proceeding in a similar way gives $F_{\mathcal{X}(\mathcal{T})} = \tilde{P}(\{1, x_t\}_{\tau\in\mathcal{T}}) = \sum_{m=1}^M \pi^m \prod_{t\in\mathcal{T}} \lambda^m_{x_t}$.

Define $V = \mathrm{diag}(\pi^1, \dots, \pi^M)$ and $D_k = \mathrm{diag}(\lambda^{*1}_k, \dots, \lambda^{*M}_k)$ as in (13). Define $P^\diamond = (L^\diamond)'VL^\diamond$ and $P^\diamond_k = (L^\diamond)'VD_kL^\diamond$. Then the elements of $P^\diamond$ take the form $\sum_{m=1}^M \pi^m \prod_{t\in\mathcal{T}} \lambda^m_{x_t}$ and can be expressed as $F_{\mathcal{X}(\mathcal{T})}$ for some $\mathcal{T}$ and $\{x_t\}_{t\in\mathcal{T}} \in X^{|\mathcal{T}|}$. Similarly, the elements of $P^\diamond_k$ can be expressed as $F^*_{k,\mathcal{X}(\mathcal{T})}$. For instance, if $u = 3$, $T = 7$, and both Λ and $L^\diamond$ are $M \times M$, then $P^\diamond$ is given by

$$\begin{bmatrix}
1 & F_1 & \cdots & F_{|X|} & F_{11} & \cdots & F_{|X||X|} & F_{111} & \cdots & F_{|X||X||X|} \\
F_1 & & & & & & & & & \\
\vdots & & & & & & & & & \\
F_{|X|} & & & & F_{|X|11} & & & & & F_{|X||X||X||X|} \\
F_{11} & & & & & & & & & \\
\vdots & & & & & \ddots & & & & \vdots \\
F_{|X||X|} & & & & & & & & & \\
F_{111} & & & & & & & & & F_{111|X||X||X|} \\
\vdots & & & & & & & & & \\
F_{|X||X||X|} & & & & F_{|X||X||X|11} & \cdots & & & & F_{|X||X||X||X||X||X|}
\end{bmatrix},$$

where the (i, j)th element of $P^\diamond$ is F_σ, where σ consists of the combined subscripts of the $(i, 1)$th and $(1, j)$th element of $P^\diamond$. For example, the $(|X|+1, 2)$th element of $P^\diamond$ is $F_{|X|1}$ $(= F_{1|X|})$. $P_k^\diamond$ is given by replacing F_σ in $P^\diamond$ with $F^*_{k,\sigma}$ and setting the $(1, 1)$th element to F^*_k.

Consequently, $P^\diamond$ and $P_k^\diamond$ can be computed from the distribution function of the observed data. By repeating the argument of the proof of Proposition 1, we determine $L^\diamond$, V, and D_k uniquely from $P^\diamond$ and $P_k^\diamond$ first, and then $D_\zeta = \operatorname{diag}(\lambda_\zeta^{*1}, \dots, \lambda_\zeta^{*M})$ and L^ζ for any $\zeta \in X$ from $P^\diamond$, $P_\zeta^\diamond$, $L^\diamond$, and P^ζ, where L^ζ and P^ζ are defined in (35). *Q.E.D.*

PROOF OF PROPOSITION 3: Let $V = \operatorname{diag}(\pi^1, \dots, \pi^M)$. Then $P^* = (L_1^*)' \times VL_2^*$. It follows that $\operatorname{rank}(P^*) \le \min\{\operatorname{rank}(L_1^*), \operatorname{rank}(L_2^*), \operatorname{rank}(V)\}$. Since $\operatorname{rank}(V) = M$, it follows that $M \ge \operatorname{rank}(P^*)$, where the inequality becomes strict when $\operatorname{rank}(L_1^*)$ or $\operatorname{rank}(L_2^*)$ is smaller than M.

When $\operatorname{rank}(L_1^*) = \operatorname{rank}(L_2^*) = M$, multiplying both sides of $P^* = (L_1^*)'VL_2^*$ from the right by $(L_2^*)'(L_2^*(L_2^*)')^{-1}$ gives $P^*(L_2^*)'(L_2^*(L_2^*)')^{-1} = (L_1^*)'V$. There are M linearly independent columns in $(L_1^*)'V$, because $(L_1^*)'$ has M linearly independent columns while V is a diagonal matrix with strictly positive elements. Therefore, $\operatorname{rank}(P^*(L_2^*)'(L_2^*(L_2^*)')^{-1}) = M$. It follows that $\operatorname{rank}(P^*) = M$ because $M \le \min\{\operatorname{rank}(P^*), \operatorname{rank}(L_2^*), \operatorname{rank}(L_2^*(L_2^*)')^{-1}\}$, and $\operatorname{rank}(L_2^*) = M$ imply $\operatorname{rank}(P^*) \ge M$. *Q.E.D.*

PROOF OF PROPOSITION 4: The proof is similar to the proof of Proposition 1. Define P_t and $P_{t,k}$ analogously to P and P_k but with λ_{x_2} and λ_{x_3} replaced with λ_{t,x_t} and $\lambda_{t+1,x_{t+1}}$ in the definition of $F.$ and $F^*_.$. Define V and D_k as before. Then P_t and $P_{t,k}$ are expressed as $P_t = L_t'VL_{t+1}$ and $P_{t,k} = L_t'VD_kL_{t+1}$. Since L_t and L_{t+1} are nonsingular, we have $A_k = P_t^{-1}P_{t,k} = L_{t+1}^{-1}D_kL_{t+1}$. Because $A_kL_{t+1}^{-1} = L_{t+1}^{-1}D_k$, the eigenvalues of A_k determine the diagonal elements of D_k while the right eigenvectors of A_k determine the columns of L_{t+1}^{-1} up to multiplicative constants; denote the right eigenvectors of A_k by $L_{t+1}^{-1}K$, where K is some diagonal matrix. Now we can determine VK from the first row of $P_tL_{t+1}^{-1}K$ because $P_tL_{t+1}^{-1}K = L_t'VK$ and the first row of L_t' is a vector of ones. Then L_t' is determined uniquely by $L_t' = (L_t'VK)(VK)^{-1}$. Having obtained L_t', we may determine V and L_{t+1} from $VL_{t+1} = (L_t')^{-1}P$ because the first column of VL_{t+1} equals the diagonal of V and $L_{t+1} = V^{-1}(VL_{t+1})$. Therefore, we determine $\{\pi^m, \{\lambda^m_{t,\xi_j^t}, \lambda^m_{t+1,\xi_j^{t+1}}\}_{j=1}^{M-1}\}_{m=1}^M$ as elements of V, L_t, and L_{t+1}. Once V, L_t and L_{t+1} are determined, we can uniquely determine $D_\zeta = \operatorname{diag}(\lambda_\zeta^{*1}, \dots, \lambda_\zeta^{*M})$ for any $\zeta \in X$ by constructing $P_{t,\zeta}$ in the same way as $P_{t,k}$ and using the relationship $D_\zeta = (L_t'V)^{-1}P_{t,\zeta}(L_{t+1})^{-1}$. Furthermore, for arbitrary $\zeta \in X$, define

$$\underset{(M\times 2)}{L_t^\zeta} = \begin{bmatrix} 1 & \lambda^1_{t,\zeta} \\ \vdots & \vdots \\ 1 & \lambda^M_{t,\zeta} \end{bmatrix}.$$

Then $P_t^{\zeta} = (L_t^{\zeta})' V L_{t+1}$ is a function of the distribution function of the observable data, and we can uniquely determine $(L_t^{\zeta})'$ for $2 \le t \le T-1$ as $P_t^{\zeta}(VL_{t+1})^{-1}$. For $t = T$, we can use the fact that $(L_{T-1})' V L_T^{\zeta}$ is also a function of the distribution function of the observable data and proceed in the same manner. Therefore, we can determine $\{\lambda_{\zeta}^{*m}, \lambda_{t,\zeta}^{m}\}_{j=1}^{M-1}$ for any $\zeta \in X$ and $2 \le t \le T$. *Q.E.D.*

PROOF OF PROPOSITION 6: Without loss of generality, set $T = 6$. Integrating out s_t's backward from $P(\{s_t\}_{t=1}^6)$ and fixing $s_1 = s_3 = s_5 = \bar{s}$ gives the "marginals"

$$\tilde{F}^*_{s_2,s_4,s_6} = \sum_{m=1}^M \pi_{\bar{s}}^m \lambda_{\bar{s}}^m(s_2)\lambda_{\bar{s}}^m(s_4)\lambda_{\bar{s}}^{*m}(s_6), \quad \tilde{F}^*_{s_2,s_6} = \sum_{m=1}^M \pi_{\bar{s}}^m \lambda_{\bar{s}}^m(s_2)\lambda_{\bar{s}}^{*m}(s_6),$$

$$\tilde{F}^*_{s_6} = \sum_{m=1}^M \pi_{\bar{s}}^m \lambda_{\bar{s}}^{*m}(s_6), \quad \tilde{F}_{s_2,s_4} = \sum_{m=1}^M \pi_{\bar{s}}^m \lambda_{\bar{s}}^m(s_2)\lambda_{\bar{s}}^m(s_4),$$

$$\tilde{F}_{s_2} = \sum_{m=1}^M \pi_{\bar{s}}^m \lambda_{\bar{s}}^m(s_2), \quad \tilde{F} = \sum_{m=1}^M \pi_{\bar{s}}^m.$$

As in the proof of Proposition 1, evaluate these $\tilde{F}$.'s at $s_2 = \xi_1, \dots, \xi_{M-1}$, $s_4 = \xi_1, \dots, \xi_{M-1}$, and $s_6 = r$, and arrange them into two $M \times M$ matrices:

$$P_{\bar{s}} = \begin{bmatrix} \tilde{F} & \tilde{F}_{\xi_1} & \cdots & \tilde{F}_{\xi_{M-1}} \\ \tilde{F}_{\xi_1} & \tilde{F}_{\xi_1,\xi_1} & \cdots & \tilde{F}_{\xi_1,\xi_{M-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{F}_{\xi_{M-1}} & \tilde{F}_{\xi_{M-1},\xi_1} & \cdots & \tilde{F}_{\xi_{M-1},\xi_{M-1}} \end{bmatrix},$$

$$P_{\bar{s},k} = \begin{bmatrix} \tilde{F}^*_k & \tilde{F}^*_{\xi_1,k} & \cdots & \tilde{F}^*_{\xi_{M-1},k} \\ \tilde{F}^*_{\xi_1,k} & \tilde{F}^*_{\xi_1,\xi_1,k} & \cdots & \tilde{F}^*_{\xi_1,\xi_{M-1},k} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{F}^*_{\xi_{M-1},k} & \tilde{F}^*_{\xi_{M-1},\xi_1,k} & \cdots & \tilde{F}^*_{\xi_{M-1},\xi_{M-1},k} \end{bmatrix}.$$

Define $V_{\bar{s}} = \mathrm{diag}(\pi_{\bar{s}}^1, \dots, \pi_{\bar{s}}^M)$ and $D_{k|\bar{s}} = \mathrm{diag}(\lambda_{\bar{s}}^{*1}(k), \dots, \lambda_{\bar{s}}^{*M}(k))$. Then $P_{\bar{s}}$ and $P_{\bar{s},k}$ are expressed as $P_{\bar{s}} = L_{\bar{s}}' V_{\bar{s}} L_{\bar{s}}$ and $P_{\bar{s},k} = L_{\bar{s}}' V_{\bar{s}} D_{k|\bar{s}} L_{\bar{s}}$. Repeating the argument of the proof of Proposition 1 shows that $L_{\bar{s}}$, $L_{\bar{s}}$, $V_{\bar{s}}$, and $D_{k|\bar{s}}$ are uniquely determined from $P_{\bar{s}}$ and $P_{\bar{s},k}$, and that $D_{s|\bar{s}}$ and $\lambda_{\bar{s}}^m(s)$ can be uniquely determined for any $s \in S$ and $m = 1, \dots, M$. *Q.E.D.*

PROOF OF PROPOSITION 7: Define $V_{\bar{s}} = \mathrm{diag}(\pi_{\bar{s}}^1, \dots, \pi_{\bar{s}}^M)$ and $D_{k|\bar{s}} = \mathrm{diag}(\lambda_{\bar{s}}^{*1}(k), \dots, \lambda_{\bar{s}}^{*M}(k))$. Applying the argument of the proof of Proposition 6 with $L_{\bar{s}}$ replaced by $L^{\diamond}_{\bar{s}}$, we can identify $L^{\diamond}_{\bar{s}}$, $V_{\bar{s}}$, and $D_{k|\bar{s}}$, and then $D_{s|\bar{s}}$ and $\lambda^m_{\bar{s}}(s)$ for any $s \in S$ and $m = 1, \dots, M$. The stated result immediately follows. *Q.E.D.*

PROOF OF PROPOSITION 9: The proof uses the logic of the proof of Proposition 6. Consider a sequence $\{s_t, z_t\}_{t=1}^4$ with $(s_1, s_2, s_3, s_4) = (\bar{s}, \bar{s}, \bar{s}, r)$ and $(z_1, z_4) = (h, k)$. Summarize the value of s_4 and z_4 into $\zeta = (r, k)$. For $(z_2, z_3) \in Z^2$, define $\tilde{F}^{h*}_{z_2,z_3,\zeta} = \sum_{m=1}^M \tilde{\pi}^m_{\bar{s},h}\tilde{\lambda}^m_{\bar{s}}(z_2)\tilde{\lambda}^m_{\bar{s}}(z_3)\tilde{Q}^m(r|\bar{s}, k)$ and $\tilde{F}^h_{z_2,z_3} = \sum_{m=1}^M \tilde{\pi}^m_{\bar{s},h}\tilde{\lambda}^m_{\bar{s}}(z_2)\tilde{\lambda}^m_{\bar{s}}(z_3)$. Define $\tilde{F}^{h*}_{z_2,\zeta} = \sum_{m=1}^M \tilde{\pi}^m_{\bar{s},h}\tilde{\lambda}^m_{\bar{s}}(z_2)\tilde{Q}^m(r|\bar{s}, k)$, and define $\tilde{F}^{h*}_{\zeta}$, $\tilde{F}^h_{z_2}$, and $\tilde{F}^h$ analogously to the proof of Proposition 6.

As in the proof of Proposition 6, arrange these marginals into two matrices $\bar{P}^h$ and $\bar{P}^h_{\zeta}$. $\bar{P}^h$ and $\bar{P}^h_{\zeta}$ are the same as $P_{\bar{s}}$ and $P_{\bar{s},k}$, but $\tilde{F}_{\cdot}$ and $\tilde{F}^*_{\cdot,k}$ replaced with $\tilde{F}^h_{\cdot}$ and $\tilde{F}^{h*}_{\cdot,\zeta}$ and subscripts are elements of Z instead of S. Define $\tilde{V}^h_{\bar{s}} = \mathrm{diag}(\tilde{\pi}^1_{\bar{s},h}, \dots, \tilde{\pi}^M_{\bar{s},h})$ and $\tilde{D}_{\zeta|\bar{s}} = \mathrm{diag}(\tilde{Q}^1(r|\bar{s}, k), \dots, \tilde{Q}^M(r|\bar{s}, k))$. It then follows that $\bar{P}^h = \bar{L}'_{\bar{s}}\tilde{V}^h_{\bar{s}}\bar{L}_{\bar{s}}$ and $\bar{P}^h_{\zeta} = \bar{L}'_{\bar{s}}\tilde{V}^h_{\bar{s}}\tilde{D}_{\zeta|\bar{s}}\bar{L}_{\bar{s}}$. By repeating the argument of the proof of Proposition 1, we can uniquely determine $\bar{L}_{\bar{s}}$, $\tilde{V}^h_{\bar{s}}$, and $\tilde{D}_{\zeta|\bar{s}}$ from $\bar{P}^h$ and $\bar{P}^h_{\zeta}$, and, having determined $\bar{L}_{\bar{s}}$, determine $\tilde{D}_{(s,z)|\bar{s}}$ for any $(s, z) \in S \times Z$. *Q.E.D.*

PROOF OF PROPOSITION 10: For $(x_2, x_3) \in B^h \times C^h$ and $x_c \in B^h \cup C^h$, define $F^{h*}_{x_2,x_3,k} = \sum_{m=1}^M \tilde{\pi}^m_h\tilde{\lambda}^m_{x_2}\tilde{\lambda}^m_{x_3}\tilde{\lambda}^m_k$, $F^{h*}_{x_c,k} = \sum_{m=1}^M \tilde{\pi}^m_h\tilde{\lambda}^m_{x_c}\tilde{\lambda}^m_k$, $F^{h*}_k = \sum_{m=1}^M \tilde{\pi}^m_h\tilde{\lambda}^m_k$, $F^h_{x_2,x_3} = \sum_{m=1}^M \tilde{\pi}^m_h\tilde{\lambda}^m_{x_2}\tilde{\lambda}^m_{x_3}$, $F^h_{x_c} = \sum_{m=1}^M \tilde{\pi}^m_h\tilde{\lambda}^m_{x_c}$, and $F^h = \sum_{m=1}^M \tilde{\pi}^m_h$. They can be constructed from sequentially integrating out $P(\{a_t, x_t\}_{t=1}^4)$ backward and then dividing them by a product of $f(x_t|x_{t-1}, 0)$. Note that Assumption 4(b) guarantees $f(x_t|x_{t-1}, 0) > 0$ for all x_t and x_{t-1} in the subsets of X considered.

As in the proof of Proposition 1, arrange these marginals into two matrices P^h and P^h_k. P^h and P^h_k are the same as P and P_k but $F_{\cdot}$ and $F^*_{k,\cdot}$ are replaced with $F^h_{\cdot}$ and $F^{h*}_{\cdot,k}$. Define $\tilde{V}_h = \mathrm{diag}(\tilde{\pi}^1_h, \dots, \tilde{\pi}^M_h)$ and $\tilde{D}_k = \mathrm{diag}(\tilde{\lambda}^1_k, \dots, \tilde{\lambda}^M_k)$. By applying the argument in the proof of Proposition 4, we may show that $\tilde{L}_b$, $\tilde{L}_c$, $\tilde{V}_h$, and $\tilde{D}_k$ are uniquely determined from $\tilde{P}(\{a_t, x_t\}_{t=1}^4)$ and its marginals, and then show that $\{\tilde{\lambda}^m_{\xi}\}_{m=1}^M$ is determined for $\xi \in B^h \cup C^h$. *Q.E.D.*

REFERENCES

AGUIRREGABIRIA, V. (2006): "Another Look at the Identification of Dynamic Discrete Decision Processes," Unpublished Manuscript, University of Toronto. [139]

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. [138,160]

——— (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1–54. [138,139,160,161]

ANDERSON, T. W. (1954): "On Estimation of Parameters in Latent Structure Analysis," *Psychometrika*, 19, 1–10. [145]

BAJARI, P., AND H. HONG (2006): "Semiparametric Estimation of a Dynamic Game of Incomplete Information," Technical Working Paper 320, NBER. [138]

BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75, 1331–1370. [138]

BLISCHKE, W. R. (1964): "Estimating the Parameters of Mixtures of Binomial Distributions," *Journal of the American Statistical Association*, 59, 510–528. [149]

BROWNING, M., AND J. CARRO (2007): "Heterogeneity and Microeconometrics Modelling," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Vol. 3, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge, U.K.: Cambridge University Press, 47–74. [141,158]

CAMERON, S. V., AND J. J. HECKMAN (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106, 262–333. [136]

CHANDRA, S. (1977): "On the Mixtures of Probability Distributions," *Scandinavian Journal of Statistics*, 4, 105–112. [144]

CRAWFORD, G. S., AND M. SHUM (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73, 1137–1173. [135]

ELBERS, C., AND G. RIDDER (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 49, 403–409. [139]

GEWEKE, J., AND M. KEANE (2001): "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics*, Vol. 5, ed. by J. Heckman and E. Leamer. Amsterdam: North-Holland. [139]

GIBSON, W. A. (1955): "An Extension of Anderson's Solution for the Latent Structure Equations," *Psychometrika*, 20, 69–73. [145]

GOWRISANKARAN, G., M. F. MITCHELL, AND A. MORO (2005): "Why Do Incumbent Senators Win? Evidence From a Dynamic Selection Model," Unpublished Manuscript, Washington University in St. Louis. [135]

HALL, P., AND X.-H. ZHOU (2003): "Nonparametric Estimation of Component Distributions in a Multivariate Mixture," *Annals of Statistics*, 31, 201–224. [136,142]

HALL, P., A. NEEMAN, R. PAKYARI, AND R. ELMORE (2005): "Nonparametric Inference in Multivariate Mixtures," *Biometrika*, 92, 667–678. [137]

HECKMAN, J. J. (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Panel Data With Econometric Applications*, ed. by C. F. Manski and D. McFadden. Cambridge, MA: MIT Press, 179–195. [141]

HECKMAN, J. J., AND B. SINGER (1984): "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320. [136,139]

HONORÉ, B. E., AND E. TAMER (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 73, 611–629. [139]

HOTZ, J., AND R. A. MILLER (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529. [138,160]

HOUDE, J.-F., AND S. IMAI (2006): "Identification and 2-Step Estimation of DDC Models With Unobserved Heterogeneity," Unpublished Manuscript, Queen's University. [149,150]

KASAHARA, H., AND K. SHIMOTSU (2008a): "Pseudo-Likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models," *Journal of Econometrics*, 146, 92–106. [138,160]

——— (2008b): "Nonparametric Identification and Estimation of Multivariate Mixtures," Unpublished Manuscript, Queen's University. [147,169]

KEANE, M. P., AND K. I. WOLPIN (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522. [136]

KITAMURA, Y. (2004): "Nonparametric Identifiability of Finite Mixtures," Unpublished Manuscript, Yale University. [139]

LEROUX, B. G. (1992): "Consistent Estimation of a Mixing Distribution," *Annals of Statistics*, 20, 1350–1360. [169]

LIU, X., AND Y. SHAO (2003): "Asymptotics for Likelihood Ratio Tests Under Loss of Identifiability," *Annals of Statistics*, 31, 807–832. [169]

MADANSKY, A. (1960): "Determinantal Methods in Latent Class Analysis," *Psychometrika*, 25, 183–198. [145,148]

MAGNAC, T., AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816. [139,147,162]

MCLACHLAN, G. J., AND D. PEEL (2000): *Finite Mixture Models*. New York: Wiley. [169]

PAKES, A., M. OSTROVSKY, AND S. BERRY (2007): "Simple Estimators for the Parameters of Discrete Dynamic Games (With Entry/Exit Examples)," *RAND Journal of Economics*, 38, 373–399. [138]

PESENDORFER, M., AND P. SCHMIDT-DENGLER (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75, 901–928. [138]

RAO, P. (1992): *Identifiability in Stochastic Models*. San Diego: Academic Press. [136]

RIDDER, G. (1990): "The Non-Parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies*, 57, 167–181. [139]

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033. [142,164]

——— (1994): "Estimation of Dynamic Structural Models, Problems and Prospects: Discrete Decision Processes," in *Advances in Econometrics: Sixth World Congress of the Econometric Society*, ed. by C. Sims. Cambridge, U.K.: Cambridge University Press, 119–170. [139,162]

TITTERINGTON, D. M., A. F. M. SMITH, AND U. E. MAKOV (1985): *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley. [136]

VAN DEN BERG, G. J. (2001): "Duration Models: Specification, Identification and Multiple Durations," in *Handbook of Econometrics*, Vol. 5, ed. by J. Heckman and E. Leamer. Amsterdam: North-Holland. [139]

Dept. of Economics, University of Western Ontario, London, Ontario, N6A 5C2 Canada; hkasahar@uwo.ca

and

Dept. of Economics, Queen's University, 94 University Avenue, Kingston, Ontario, K7L 3N6 Canada; shimotsu@econ.queensu.ca.

Manuscript received October, 2006; final revision received April, 2008.