Identification of Structural Discrete Dynamic Programming Models *

Dr. Lakshmi K. Raut
Visiting Fellow, University of Chicago
Center for the Economics of Human Development
1126 E. 59th St., Chicago, IL 60637,USA
mailto:lakshmiraut@gmail.com

December 9, 2020

Abstract

Structural Dynamic Programming Models are not in general econometrically identified. The paper finds general conditions for identification for a commonly used family of such models.

Keywords: Structural Dynamic Programming
JEL Classification Nos.: J24, J62, O15, I21

First Draft: December 2009
This Draft: April 20, 2015.

* Will be added

Identification of Structural Discrete Dynamic Programming Models

1 Introduction

To incorporate Hotz and Miller, 1993

The rest of the paper is organized as follows. Section 2 describes the basic framework of structural dynamic programming models and notations of the paper. Section 4 deals with the issues related to identification of the structural parameters of the dynamic programming model. Section **??** concludes the paper.

Example 1 (An Example)**.**

We give a preschool investment decision problem of an altruistic parent from Heckman and Raut, 2016.

2 The Basic Framework

The choices of an individual with observable characteristics x are restricted to the set $A(x,\varepsilon) \equiv \{a \in A | c(w(x,\varepsilon), a) > 0\}$. The choice a yields direct utility from life-time annualized consumption and indirect utility through its effect on child outcomes and welfare, as represented in the following Bellman equation corresponding to the parent's preschool investment decision problem

$$V(x,\varepsilon) = \max_{a \in A(x,\varepsilon)} u(x,\varepsilon,a) + \beta \int V(x',\varepsilon') \, p(dx', d\varepsilon' | x, \varepsilon, a) \tag{1}$$

where $V(.)$ is the intergenerational welfare function, known in the dynamic programming literature as the value function, $u(.)$ is the felicity index of yearly permanent consumption over the whole lifetime of the parent, and the parameter β measures the degree of parental altruism toward the child.

Under general regularity conditions on $u(.)$, $p(dx', d\varepsilon' | x, \varepsilon, a)$ and β, the value function $V(x,\varepsilon)$, and a measurable optimal decision rule $a^*(x,\varepsilon)$ exist (see, for instance, Bhattacharya and Majumdar, 1989, Theorem 3.2). Given $u(.)$, $p(dx', d\varepsilon' | x, \varepsilon, a)$ and β, satisfying the regularity conditions, we carry out a Lucas-Critique free policy evaluation by examining a policy's effect on the individual optimal decision a, on the intergenerational welfare level V, and we also examine the intergenerational long-run aggregate effect of the policy on the economy by aggregating individual choices with respect to the long-run population distribution, also known as the invariant population distribution, of the equilibrium transition probability distribution $p\left(dx', d\varepsilon' | x, \varepsilon, a^*\left(x, \varepsilon\right)\right)$.

To be able to do this, we need to estimate the structural parameters. Our data consists of a sample of parent's observable state x, child's observable state x', and parents decision a. Suppose a vector of parameters ξ_p specifies the transition probability distribution $p\left(dx', d\varepsilon' | x, \varepsilon, a\right)$. Our problem is then to statistically estimate the structural parameters $\zeta = \left\{u(.), \xi_p, \beta\right\}$ given observable information on a random sample of parent-child pairs $y = \left\{(x_i, x_i'), a_i\right\}_{i=1}^n$ such that the predicted behaviors of the sample from the model are close to observed behavior. We denote the log-likelihood function of the sample by $\mathcal{L}_y(\zeta)$. Estimation of the model involves two steps: For a given ζ, calculate the probability distribution of the endogenous variables $a_i | x_i$ and $x_i' | x_i, a_i$ using the model to form the log-likelihood of the sample $\mathcal{L}_y(\zeta)$ and then use an appropriate estimation procedure to choose a ζ.

Two questions need be addressed to that end. First, is the computation of the likelihood $\mathcal{L}_y(\zeta)$, which involves solving the dynamic programming problem in Eq. (1) repeatedly for each (x, ε), feasible with the currently available computing technology, especially when ε is a continuous multivariate random variable? Second, are the structural parameters of the model identified (the definition of identification is stated later)?

The answer to both questions is in general no. Following the literature, we make simplifying assumptions to transform the above structural dynamic programming problem into a random utility model of discrete choices. We will show that these assumptions greatly simplify the computation and the identification of the structural parameters of the model. Given those assumptions, we will see two facts: First, the set of structural parameters ξ_p determines the transition distribution $p\left(x' | x, a\right)$ of the observable state variables, which is the mixture distribution of the original transition probability distribution, more specifically $p\left(x'|x, a\right) = \int p\left(x', \varepsilon' | x, \varepsilon, a\right)\ d\varepsilon | x\ d\varepsilon' | x'$. Second, the set of optimal choice probabilities $P\left(a|x\right), a \in A\left(x\right), x \in X$ over the observed discrete choices depends on ξ_p only through $p\left(x'|x, a\right)$.

Notice that the optimal choice a is treated as an exogenous variable in the estimation of $p\left(x'|x, a\right)$, the maximization of joint likelihood of two components is more efficient. To make estimation task computationally manageable, however, again following the trend in the literature, in place of ξ_p, we take an estimate of $p\left(x'|x, a\right)$ as our fixed parameters in the vector of parameters ζ, and in place of β, we calibrate β from other information, and then form the likelihood of the sample of observed discrete choices $a_i|x_i$ for identification and estimation of the remaining parameters.

2.1 Notation

In the rest of the paper, our parameter vector is $\zeta = \{u(x,a), p(x'|x,a), \beta\}, a \in A(x), x \in X$ where $p(x'|x,a)$ and β are fixed. Denote by Ξ the set of all such parameter values. We denote by $\mathcal{L}_y(\zeta)$ the log-likelihood of the sample of observed choices $y = \{a_i|x_i, i = 1...n\}$. Given a set of conditional choice probabilities $\{P(a|x), a \in A(x), x \in X\}$ which depends on ζ, the log-likelihood function $\mathcal{L}_y(\zeta)$ of the sample is defined.

Let J_x denote the number of elements in the feasible choice set $A(x)$. Denote by $J = \sum_{x \in X} J_x$. Assume that X is a finite ordered set of M elements.

Denote by $F(a) = [f(x'|x,a)]_{x',x \in X}$ the $J_x \times J_{x'}$ conditional transition probability matrix given a choice $a \in A(x)$ where the element $f(x'|x,a)$ corresponding to the row x and the column x' is the probability of the child moving to state x' given that his parent is from the state x and he had made a choice $a \in A(x)$. We denote by $F(x,a)$ the row vector of $F(a)$ corresponding to the parent's state x.

The vector of conditional choice probabilities denoted by $\mathcal{P} = \{P(a|x), a \in A(x), x \in X\}$ is ordered by the primary index of ordering in X and the secondary index of the ordering in A. For each x, the component vector of conditional choice probabilities $\{P(a|x), a \in A(x)\}$ belongs to a $J_x - 1$ dimensional simplex. The set of all vectors $\mathcal{P}$ of conditional choice probabilities $\triangle$ is a subset of $\Re^{MJ}_{++}$ which is restricted to the interior of the M-fold cross product of the $J_x - 1$ dimensional simplices.

For any function $v(x,a)$, its vector representation is a $J \times 1$ vector v (i.e., with the same symbol v) in which the function values $v(x,a)$'s are ordered in the same way as in $\mathcal{P}$. For any scalar or a vector function $w(x)$, we denote by w (again using the same symbol w to denote it) the values of w stacked in rows in the same order as in the ordered set X.

For any random vector or a random variable $w(x,a)$, we denote its expectation with respect to a by $\bar{w}(x)$, i.e., $\bar{w}(x) \equiv \sum_{a \in A(x)} w(x,a) P(a|x)$, (with the convention that when w is a random vector, the product inside this summation is element-by-element). Define the $M \times J$ matrix Π derived from a vector of conditional choice probabilities $\mathcal{P}$ by

$$\underset{M\times J}{\Pi} = \begin{pmatrix} P(a=1|x_1) & \dots & P(a=J|x_1) & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & P(a=1|x_M) & \dots & P(a=J|x_M) \end{pmatrix}$$

and the transition matrices in matrix notation as a $J \times M$ matrix F by,

$$\underset{J\times M}{F} = \begin{pmatrix} f(x_1'|x_1, a=1) & \dots & f(x_M'|x_1, a=1) \\ & \dots & \\ f(x_1'|x_1, a=J_{x_1}) & \dots & f(x_M'|x_1, a=J_{x_1}) \\ & \dots & \\ f(x_1'|x_M, a=1) & \dots & f(x_M'|x_M, a=1) \\ & \dots & \\ f(x_1'|x_M, a=J_{x_M}) & \dots & f(x_M'|x_M, a=J_{x_M}) \end{pmatrix}$$

3 Class of Structural Models

The structural estimation of the original problem is computationally intractable. Similar to Rust, 1994, we make the following simplifying assumptions to transform the original model in Eq. (1) to a random utility model. In the next two sections, we utilize these simplifications to find conditions for identification and estimation of structural parameters.

We assume that $w(x,\varepsilon)$ and hence $A(x,\varepsilon)$ does not depend on ε, i.e., $w()$ does not contain any unobservable idiosyncratic shocks. However, we assume that ε represents a taste shifter for individual preferences and constitutes our only source of unobserved heterogeneity, the specific nature of which is stated formally in the following assumption.

Assumption 1. $u(x,\varepsilon,a) = u(x,a) + \varepsilon(a)$, and support of $\varepsilon(a)$ is the real line for all $a \in A(x)$.

We also make the following additional assumptions.

Assumption 2. The transition probability $p(x',\varepsilon'|x,\varepsilon,a) = g(\varepsilon'|x') f(x'|x,a)$ for some twice continuously differentiable density function g with finite first moment.

Assumption 3. The set of observable individual characteristics $X = \{x^1, ..x^M\}$ is a finite ordered set.

Under Assumption 1 - Assumption 3 , we have

$$V(x,\varepsilon) = \max_{a\in A(x)} u(x,a) + \varepsilon(a) + \beta \sum_{x'\in X} \int V(x',\varepsilon')\, g(d\varepsilon'|x')\, f(x'|x,a) \quad (2)$$

Denote the value function, after integrating out the unobservable component of the state variable, by $v(x) \equiv \int V(x,\varepsilon)\, g(d\varepsilon|x)$. Integrating both sides of Equation (2) with respect to the conditional density $g(d\varepsilon|x)$, and utilizing this notation for $v(x)$, we have

$$v(x) = \int \max_{a\in A} [\tilde{v}(x,a) + \varepsilon(a)]\, g(d\varepsilon|x) \quad (3)$$

where

$$\begin{aligned} \tilde{v}(x,a) &\equiv u(x,a) + \beta \sum_{x'\in X} v(x')\, f(x'|x,a) \\ &= u(x,a) + \beta F(x,a).v \end{aligned} \quad (4)$$

Eq. (3) above is a random utility model in which the function $\tilde{v}(x,a)$ measures the common utility that an individual of observable characteristics x derives from a choice $a \in A(x)$. Denote by

$$\Omega(x,a) = \{\varepsilon | \tilde{v}(x,a) + \varepsilon(a) \geq \tilde{v}(x,a') + \varepsilon(a'), \text{for all } a' \in A(x)\} \quad (5)$$

the set of individuals with observed characteristics x who made a as their optimal choice. The conditional choice probabilities are then given by

$$P(a|x) = \int_{\Omega(x,a)} g(d\varepsilon|x). \quad (6)$$

By partitioning the domain of integral in Eq. (3) into disjoint regions $\Omega(x,a), a \in A(x), x \in X$ and then integrating we have the following,

$$\begin{aligned} v(x) &= \sum_{a\in A(x)} P(a|x) \left[u(x,a) + \frac{\int_{\Omega(x,a)} \varepsilon(a)\, g(d\varepsilon|x)}{P(a|x)} + \beta \sum_{x'\in X} v(x')\, f(x'|x,a) \right] \\ &= \sum_{a\in A(x)} P(a|x) [u(x,a) + e(x,a) + \beta F(x,a) \cdot v] \text{ ...(*)} \\ &= \bar{u}(x) + \bar{e}(x) + \beta \bar{F}(x) \cdot v \end{aligned} \quad (7)$$

where

$$e(x,a) \equiv \int_{\Omega(x,a)} \varepsilon(a)\, g(d\varepsilon|x) / P(a|x) \quad (8)$$

in line (*) is the conditional expectation of the component $\varepsilon(a)$ of the random vector ε given x and a. Writing the above in matrix notation, we have

$$v = \bar{u} + \bar{e} + \beta \bar{F} \cdot v \equiv \Phi(v, \zeta) \tag{9}$$

Let $v(\zeta)$ be a fixed point of the map $\Phi(v, \zeta)$ for given $\zeta \in \Xi$, and denote by $\mathcal{P}(v)$ the conditional choice probabilities in Eq. (6) for a given value function v. Then the computation of the likelihood of the sample is simplified to the computation of the fixed point of the above map $\Phi(v, \zeta)$. The computation of $P(a|x)$, and $e(x, a)$ involve multi-dimensional numerical integration, which may make computations extremely slow. Both computational tasks are, however, substantially simplified under the following assumption:

Assumption 4. The components of ε are independently and identically distributed as extreme value distribution with location parameter 0 and scale parameter 1.

McFadden, 1981 has shown that under Assumption 4, $e(x, a) = (\lambda - \ln P(a|x))$, where λ is the Euler-Mascheroni constant, with a numerical value of $\lambda = 0.57721566$, and the conditional choice probability $P(a|x)$ has the following Logit representation,

$$P(a|x) = \frac{e^{\tilde{v}(x,a)}}{\sum_{a \in D} e^{\tilde{v}(x,a')}} \tag{10}$$

The above strategy of computational simplification was pioneered by Rust, 1987. The computational burdens could be, however, further simplified as follows: From Eq. (9) it follows that $v = [I_M - \beta \bar{F}]^{-1} [\bar{u} + \bar{e}]$. Substituting this in Eq. (4), we have

$$\tilde{v}(x, a) = u(x, a) + \beta F(x, a) [I_M - \beta \bar{F}]^{-1} [\bar{u} + \bar{e}] \tag{11}$$

It is easy to see that given $\mathcal{P}_0 \in \triangle$, the right hand side of the above, and hence, a new vector of conditional choice probabilities say $\mathcal{P}_1 \in \triangle$ can easily be computed by substituting it in Eq. (10). We represent this relationship for each structural parameter $\zeta \in \Xi$ by $\mathcal{P}_1 = \Psi(\mathcal{P}_0, \zeta)$. Following the line of argument in Aguirregabiria and Mira, 2002, it is easy to show that for each $\zeta \in \Xi$, there exists a unique fixed point $\mathcal{P}(\zeta)$ to the mapping $\Psi(\mathcal{P}, \zeta)$, and starting from any initial $\mathcal{P}_0 \in \triangle$, the iterative process $\mathcal{P}_{n+1} = \Psi(\mathcal{P}_n, \zeta)$, $n \geq 0$ converges to the fixed point $\mathcal{P}(\zeta) \in \triangle$. Thus, for each structural parameter $\zeta \in \Xi$, there exists a unique likelihood of the sample $\mathcal{L}_y(\zeta)$, the computation of which is brought down to computation of the fixed point of the mapping Ψ on the finite dimensional space $\triangle$.

4 Identification of Structural Parameters

In the previous section we saw that given $\zeta \in \Xi$, there exists a unique likelihood function $\mathcal{L}_y(\zeta)$. To be able to estimate $\zeta \in \Xi$, the model should be identified in the sense that

$$\mathcal{L}_y(\zeta) = \mathcal{L}_y(\zeta') \text{ } a.e. \text{ if and only if } \zeta = \zeta', \tag{12}$$

the a.e. is with respect to the dominant probability measure defining the likelihood of the sample. Following Rao, 1992, we say that our model is *globally identified* if the relationship in Eq. (12) holds for any two $\zeta, \zeta' \in \Xi$, and is *locally identified* around a particular parameter $\zeta \in \Xi$, if the relationship in Eq. (12) holds for all $\zeta' \in \Xi$ in a neighborhood of ζ.

To find reasonable conditions for identification, from Eq. (5) note that the optimal choices are invariant if we add a location m_x and divide both sides by a scale factor $\sigma_x > 0$, for each $x \in X$. Thus it follows that we can recover the utility function only up to a scale and location. Given this fact, we restrict the one period utility function $(u(x,a), a \in A(x))$ to lie in a $J_x - 1$ dimensional open submanifold of $\Re^{J_x}$ for each $x \in X$. We take each possible utility vector $(u(x,a), a \in A(x), x \in X)$ to lie in the cross product (or equivalently in the direct sum, if we view $\Re^{J_x}$ to be embedded in $\Re^J$) of these $J_x - 1$ dimensional submanifolds over all $x \in X$. There are many such manifolds, and up to diffeomorphisms they are all equivalent. We define one such manifold $\mathcal{U}$ using the map $\varphi : \triangle \ni \mathcal{P} \mapsto u \in \Re^J$ (which reads as, φ takes a member $\mathcal{P}$ in $\triangle$ to a member u in $\Re^J$) by

$$u = \left[I_J + \beta F \left(I_M - \beta \bar{F}\right)^{-1} \Pi\right]^{-1} [\tilde{v} - \tilde{e}] \equiv \varphi(\mathcal{P}) \tag{13}$$

where $\tilde{v}(x,a) = \ln P(a|x)$ and $\tilde{e} = \beta F \left(I_M - \beta \bar{F}\right)^{-1} \Pi e$. Take $\mathcal{U} = \varphi^{-1}(\triangle)$. It can be shown that the set $\mathcal{U}$ is a $J - M$ dimensional smooth manifold. Given parameters β, and F fixed, we restrict our parameter space Ξ to be such that the u-component of a parameter vector $\zeta \in \Xi$ is restricted to lie in $\mathcal{U}$. The most general non-parametric family that we can restrict our parameters u to lie in is $\mathcal{U}$. Our nonparametric identification issue boils down to the question, under what conditions can we identify our structural model in this non-parametric family of $\mathcal{U}$? Theorem 1 addresses this, using the following assumption

Assumption 5. Given the vector of transition probabilities F, the degree of altruism parameter β is such that (1) $0 \leq \beta < 1$ and (2) $I_J + \beta F \left(I_M - \beta \bar{F}\right)^{-1} \Pi$ is of full rank.

Note that there always exist such $\beta' s$ at least near $\beta = 0$. Also note that $\beta = 1$ will violate condition (2) since in that case $I_M - \beta\bar{F}$ is not invertible, as each row will add-up to zero.

Theorem 1 (Nonparametric Identification). Suppose the components β and F of the parameter vectors are fixed. Let $\mathcal{P} \in \triangle$ be a vector of conditional choice probabilities that satisfy Assumption 5. Then there exists a unique utility function $(u(x,a), a \in A(x), x \in X)$ $\in \mathcal{U}$ that generates $\mathcal{P}$ as the optimal solution to the choice problem in Eq. (1). Furthermore, the model in Eq. (1) is globally or locally non-parametrically identified depending on whether Assumption 5 holds globally or locally.

Proof. Let $\mathcal{P} \in \triangle$ be a vector of conditional choice probabilities that satisfy Assumption 5. Note that writing Eq. (11) in matrix notation, we have $\tilde{v} = \left[I_J + \beta F \left(I_M - \beta\bar{F}\right)^{-1}\Pi\right] u + \beta F \left(I_M - \beta\bar{F}\right)^{-1}\Pi e$, where $\bar{F}$ is the expectation of $F(a)$ with respect to $\mathcal{P}$. Taking $\tilde{v}(x,a) \equiv \ln P(a|x)$, and denoting by $\tilde{e} = \beta F \left(I_M - \beta\bar{F}\right)^{-1}\Pi e$, we have

$$u = \left[I_J + \beta F \left(I_M - \beta\bar{F}\right)^{-1}\Pi\right]^{-1}\left[\tilde{v} - \tilde{e}\right] \tag{14}$$

Thus by Assumption 5, for each $\mathcal{P}$, there exists a unique $u \in \mathcal{U}$.

We now prove the second part regarding the nonparametric identification. Note that the data on distribution of choices given a fixed number of individuals $n(x)$ (a positive integer) for each observed value of individual characteristics $x \in X$ can be summarized as an ordered vector y defined similar to $\mathcal{P}$ by $y = (n(a|x), a \in A(x), x \in X)$ where $n(a|x)$ is the number of individuals who chose $a \in A(x)$ given their characteristics $x \in X$. The likelihood of the sample can be written as follows

$$\begin{aligned}
L_y(\mathcal{P}) &= \prod_{x\in X}\frac{n(x)!}{\prod_{a\in A(x)} n_a(x)!}\exp\left(\sum_{x\in X} n(x)\ln\left(1-\sum_{a=1}^{J_x-1}P(a|x)\right)\right)\times \\
&\quad \exp\left(\sum_{x\in X}\sum_{a=1}^{J_x-1} n(a|x)\ln\left(\frac{P(a|x)}{1-\sum_{a=1}^{J_x-1}P(a|x)}\right)\right) \\
&= h(y)\,g(\eta)\exp(y'\eta)\,,\ \text{where } \eta = (\eta(a|x), a\in A(x), x\in X)\,,\text{with} \\
\eta(a|x) &= \ln\left(\frac{P(a|x)}{1-\sum_{a=1}^{J_x-1}P(a|x)}\right),\text{and } g(\eta) = -\sum n(x)\ln\left(1+\sum_{a=1}^{J_x-1}\exp\eta(a|x)\right),
\end{aligned}$$

and $h(y)$ is the multiplicative component in the first expression. It follows from the above that $L_y(\mathcal{P})$ is an exponential distribution. The determinant $\det(\mathfrak{I}(\mathcal{P}))$ of the Fisher information matrix $\mathfrak{I}(\mathcal{P})$ of $L_y(\mathcal{P})$ at any parameter vector $\mathcal{P} \in \triangle$ can be shown to be

$\det(\Im(\mathcal{P})) = \left[\prod_{x\in X}\prod_{a=1}^{J_x-1} P(a|x)\right]^{-1}$, which is always > 0 since each $P(a|x) > 0$. Since $\det(\Im(\mathcal{P}))$ is a continuous function of $\mathcal{P}$, there exists a neighborhood of $\mathcal{P}$ in $\triangle$ such that the Fisher information matrix is of full rank for all $\mathcal{P}$ in that neighborhood. Moreover, note that the function $g(\eta)$ is continuously differentiable in η. Hence by Prakash Rao (1992, Theorem 6.3.2), for any $\mathcal{P}'$ in a neighborhood of $\mathcal{P}$, we have $L_y(\mathcal{P}) = L_y(\mathcal{P}')$ *a.e.* $\Leftrightarrow \mathcal{P} = \mathcal{P}'$. But $u = \varphi(\mathcal{P})$ in Eq. (13) is a 1-1 function from $\triangle$ to $\mathcal{U}$ around $\mathcal{P} \in \triangle$ that satisfies Assumption 5. Hence for any $\zeta \in \Xi$ such that the corresponding $\mathcal{P}(\zeta)$ satisfies Assumption 5, there exists a neighborhood of ζ in Ξ such that for any ζ' in that neighborhood, $L_y(\mathcal{P}(\zeta)) = L_y(\mathcal{P}(\zeta'))$ *a.e.* $\Leftrightarrow \zeta = \zeta'$. Hence the model in Eq. (1) is locally nonparametrically identified around a ζ whose associated $\mathcal{P}(\zeta)$ satisfies Assumption 5. It is also clear that if Assumption 5 is true for all $\mathcal{P} \in \triangle$, the model in Eq. (1) is also globally identified. **Q.E.D.**

The conditional choice probabilities $\mathcal{P} = \{P(a|x), a \in A, x \in X\}$ are nothing but the aggregate demand functions of discrete choices $a \in A$ as a function of individual characteristics $x \in X$. The characteristics $x \in X$ is acting like a price of the Marshallian demand function. Nonparametric identification problem in our set-up can be viewed as the well-known aggregation problem of the consumer theory: *Given a system of demand functions $P \in \triangle$, when does there exist a utility function $u(x,a)$ that generates P as the optimal solution of problem in Eq. (1)*? The above theorem provides conditions for an analogous aggregation problem in the present context of structural dynamic programming problem.

Suppose instead of most general non-parametric utility specifications for the parameter vector ζ, we parametrize u (and also possibly β, but F is still assumed to be fixed) to have a parametric form $\zeta : \Theta \to \Xi$, where $\Theta \subset \Re^k$, $k < J - M + 1$ is an open set. When can we identify such parametric models? To state our sufficient condition for this, we recall a definition from the Differential Geometry. A map $f : \Theta \to \triangle$ is an *immersion* at $\theta \in \Theta$, an open subset of $\Re^k$, if the differential map $df_\theta : \Re^k \to T_{f(\theta)}(\triangle)$ is injective, i.e., one-to-one, where $T_{f(\theta)}(\triangle)$ is the tangent space of the manifold $\triangle$ at $f(\theta)$.

Theorem 2 (**Parametric Identification**)**.** Let $\Theta \subset \Re^k$ be an open set. Let $\zeta : \Theta \to \Xi$ denotes a family of parametric models. A parametric model is locally identified at $\theta \in \Theta$ if and only if the map $\mathcal{P}(\zeta(\theta)) : \Theta \to \triangle$ is an *immersion* at θ. The parametric model is globally identified if and only if the map $\mathcal{P}(\zeta(\theta))$ is an injective map.

Proof. Since$\mathcal{P}(\zeta(\theta))$ is an immersion at θ, there exists a neighborhood around θ in Θ

such that $\mathcal{P}(\zeta(\theta))$ is one-one in this neighborhood. For any θ' in this neighborhood of θ, $\mathcal{L}_y(\mathcal{P}(\zeta(\theta))) = \mathcal{L}_y(\mathcal{P}(\zeta(\theta')))$ a.e. implies $\mathcal{P}(\zeta(\theta)) = \mathcal{P}(\zeta(\theta'))$ since $\mathcal{L}_y(\mathcal{P})$ is globally identified in the parameter space $\triangle$ by theorem 1. Hence $\theta = \theta'$ since $\mathcal{P}(\zeta(\theta))$ is 1-1 in this neighborhood. The second part follows immediately. **Q.E.D.**

References

Aguirregabiria, V. and P. Mira (July 2002). Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models, *Econometrica*, **70**, no. 4, 1519–1543. DOI: 10.1111/1468-0262.00340 (cit. on p. 7).

Bhattacharya, R. N. and M. Majumdar (1989). Controlled semi-markov models: The discounted case, *Journal of Statistical Planning and Inference*, **21**, no. 3, 365–381. DOI: 10.1016/0378-3758(89)90053-0 (cit. on p. 2).

Heckman, J. J. and L. K. Raut (2016). Intergenerational long-term effects of preschool-structural estimates from a discrete dynamic programming model, *Journal of Econometrics*, **191**, no. 1, 164–175. DOI: 10.1016/j.jeconom.2015.10.001 (cit. on p. 2).

Hotz, V. J. and R. A. Miller (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models, *The Review of Economic Studies*, **60**, no. 3, 497–529. DOI: 10.2307/2298122 (cit. on p. 2).

McFadden, D. (1981). Econometric models of probabilistic choice, *Structural analysis of discrete data with econometric applications*, **198272** (cit. on p. 7).

Rao, B. P. (1992). Identifiability in stochastic models: characterization of probability distributions. Academic Press (cit. on p. 8).

Rust, J. (Sept. 1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher, *Econometrica*, **55**, no. 5, 999. DOI: 10.2307/1911259 (cit. on p. 7).

Rust, J. (1994). "Chapter 51 Structural estimation of markov decision processes", *Handbook of Econometrics*. Elsevier, 3081–3143. DOI: 10.1016/s1573-4412(05)80020-0 (cit. on p. 5).