CCP Estimation of Dynamic Discrete/Continuous Choice Models with Generalized Finite Dependence and Correlated Unobserved Heterogeneity

Wayne-Roy Gayle*

November 14, 2013

Abstract

This paper investigates conditional choice probability estimation of dynamic structural discrete and continuous choice models. We extend the concept of finite dependence in a way that accommodates non-stationary, irreducible transition probabilities. We show that, under this new definition of finite dependence, one-period dependence is obtainable in any dynamic model. This finite dependence property also provides a convenient and computationally cheap representation of the optimality conditions for the continuous choice variables. We allow for general form of discrete-valued unobserved heterogeneity in utilities, transition probabilities, and production functions. The unobserved heterogeneity may be correlated with the observable state variables. We show the estimator is root-n–asymptotically normal. We develop a new and computationally cheap algorithm to compute the estimator.

KEYWORDS: Conditional Choice Probabilities Estimator, Discrete/Continuous Choice, Finite Dependence, Correlated Random Effects.
JEL: C14, C31, C33, C35.

*Department of Economics, University of Virginia, Monroe Hall, McCormick Rd, Room 208A, Charlottesville, VA 22903, E-mail: wg4b@virginia.edu. The author is grateful to Lala Zun Ma insightful comments. I also thank the seminar participants at the University of Virginia. All errors are my own.

1 Introduction

In this paper, we investigate conditional choice probability (CCP) estimation of dynamic structural discrete/continuous choice models with unobserved individual heterogeneity. We show that a modest extension to the definition finite dependence in Hotz and Miller (1993), Altug and Miller (1998), and Arcidiacono and Miller (2010) accommodates general non-stationary and irreducible transition probabilities, as well as a general form of correlated unobserved heterogeneity in the utility functions, production functions, and the transition probabilities. We propose a generalized method of moments (GMM) estimator for the structural parameters of the model and derive their asymptotic distributions. We also propose a simple algorithm to implement the estimator.

Since its introduction by Hotz and Miller (1993), CCP estimation of dynamic structural models has flourished in empirical labor economics and industrial organization, largely because of its potentially immense reduction in computational cost compared to the more traditional backward recursive- and contraction mapping-based full maximum likelihood estimation pioneered by Rust (1987), referred to as the nested fixed point algorithm (NFXP). The CCP estimator circumvents having to solve the dynamic programming problem for each trial value of the structural parameters by making use of a one-to-one mapping between the normalized value functions and the CCPs established in Hotz and Miller (1993). Therefore, nonparametric estimates of the CCPs can be inverted to obtain estimates of the normalized value functions, which can then be used in estimating the structural parameters.

Empirical application the early formulation of CCP estimation had important limitations relative to the NFXP method. The emerging literature has focused on separate, but related drawbacks. The first is nonparametric estimation of the CCPs results in less efficient estimates of the structural parameters, as well as, relatively poor finite sample performance. The second is the difficulty of accounting for unobserved individual heterogeneity, mainly due to having to estimate the CCPs by nonparametric methods. A limitation of both approaches to estimation is the difficulty of both the CCP and NFXP estimators is they are largely restricted to discrete choice, discrete states models.

Aguirregabiria and Mira (2002) proposed a solution to the issue of efficiency and finite sample performance of the CCP estimator relative to NFXP estimator. They show that, for a given value of the preference parameters, the fixed point problem in the value function space can be transformed into a fixed point problem in the probability space. Aguirregabiria and Mira (2002) propose swapping the nesting of the NFXP, and show the resulting estimator is asymptotically equivalent to the NFXP estimator. Furthermore, Aguirregabiria and Mira (2002) show in simulation studies that their method produce estimates 5 to 15 times faster than NFXP. The method proposed by Aguirregabiria and Mira (2002) is restricted to discrete choice models in stationary environments, and is not designed to account for unobserved individual heterogeneity.

Recent developments in accounting for unobserved heterogeneity in CCP estimators include Aguirregabiria and Mira (2007), Arcidiacono and Miller (2010). Aguirregabiria and Mira (2007) allow for permanent unobserved heterogeneity in stationary, dynamic discrete games. Their method requires multiple inversion of potentially large dimensional matrices. Arcidiacono and Miller (2010) propose a more general method for incorporating time-specific or time-invariant unobserved heterogeneity into CCP estimators. Their method modifies the Expectations-Maximization algorithm proposed in Acidiacono (2002). However, Arcidiacono and Miller's method is only applicable to discrete dynamic models.

Altug and Miller (1998) proposed a method for allowing for continuous choices in the CCP framework. By assuming complete markets, estimates of individual effects and aggregate shocks are obtained, which are then used in the second stage to form (now) observationally equivalent individuals. These observationally equivalent individuals are used to compute counterfactual continuous choices. Bajari et al. (2007) modify the methods of Hotz and Miller (1993) and Hotz et al. (1994), to estimating dynamic games. Their method of modeling unobserved heterogeneity in continuous choices is inconsistent with the dynamic selection.

The finite dependence property; when two different policies associated with different initial choices lead to the same distribution of states after a few periods, is critical for the computational feasibility and finite sample performance of CCP estimators. Finite dependence combined with the invertibility result of Hotz and Miller (1993) results in significant reduction in computational cost of estimating dynamic structural models. Essentially, the smaller the order of dependence, the faster and more precise the estimator, because fewer future choice probabilities that either have to be estimated or updated, depending of the method of estimation. The concept of finite dependence was first introduced by Hotz and Miller (1993), extended by Altug and Miller (1998), and further by Arcidiacono and Miller (2010). Despite these generalizations, the concept of finite dependence is largely restricted to discrete choice models with either stationary transitions or the renewal property.

This paper make three separate, but closely related contributions to the literature on CCP estimation of dynamic structural models. We extend the concept of finite dependence to allow for general non-stationary and irreducible transition probabilities. While its definition is precise and well understood, the strategy to construct finite dependence in dynamic structural models have been largely ad hoc and imprecise, often relying on assumptions that are either theoretically unjustified, or significantly restricting the data. Altug and Miller (1998), Gayle and Miller (2003), and Gayle (2006) rely on complete market and degenerate transition probability assumptions to form counterfactual strategies that obtain finite dependence. A key insight of Arcidiacono and Miller (2010) is: "the expected value of future utilities from optimal decision making can always be expressed as functions of the flow payoffs and conditional choice probabilities for *any* sequence of future choices, optimal or not." This insight is the basis of our extension of the finite dependence property. We show the expected value of future utilities from optimal decision making can be expressed as *any linear combination* of flow payoffs and conditional CCPs, as long as the weights sum to one. This insight converts the difficult problem of finding one pair of sequences of choices that obtains finite dependence to a continuum of finite dependencies from which to choose.

Given we are now able to choose from a continuum of finite dependence representations, the question becomes whether there is a choice that obtains one-period finite dependence. Indeed, one-period finite dependency is achievable regardless of the form of the transition probabilities. The resulting form of the conditional value function has the advantage of being elegant and intuitive, as well as providing a simple method to accommodate continuous choices. Our approach to accounting for continuous choices do not rely on first stage estimation as in Altug and Miller (1998), and Bajari et al. (2007), nor does it require forward simulation as in Hotz et al. (1994), and Bajari et al. (2007). The proposed method for estimating discrete/continuous dynamic structural models parallels the method for estimating discrete/continuous static structural models of Dubin and McFadden (1984), and Hanemann (1984), operationalized by the inversion result of Hotz and Miller (1993), Arcidiacono and Miller (2010), and the generalized finite dependence result of this paper.

To avoid stochastic degeneracy, the econometric specification of discrete/continuous structural models requires at least as many choice-specific unobserved shocks as there are choices. The estimator developed in this paper conveniently accommodates these traditional i.i.d. shocks, as well as, other correlated unobserved heterogeneity. The advantages of the algorithm proposed in this paper are, it does not require specifying initial conditions, it does not require discrete approximation of the continuous choice variables and the value functions and, its convergence is well understood.

2 Model

2.1 General framework

The general setup of a dynamic structural discrete/continuous choice model that we consider is as follows. In each period, t, an individual chooses among J discrete, mutually exclusive, and exhaustive alternatives. Let d_{tj} be one if the discrete action $j \in \{1, \cdots, J\}$ is taken in period t, and zero otherwise, and define $d_t = (d_{t1}, \cdots, d_{tJ})$. Associated with each discrete alternative, j, the individual chooses L_j continuous alternatives. Let $c_{tl_j} \in \Re_+$, $l_j \in 1, \cdots, L_j$, be the continuous actions associated with alternative j, with $c_{tl_j} > 0$ if $d_{tj} = 1$. Define $c_{tj} = (c_{t1}, \cdots, c_{tL_j}) \in \Re_+^{L_j}$, and $c_t = (c_{t1}, \cdots, c_{tJ}) \in \Re_+^L$, where $L = \sum_{j=1}^J L_j$. Also, let (j, c_{tj}) be the vector of discrete and continuous actions associated with alternative j. The current period payoff associated with action (j, c_{tj}) depends on the observed state $x_t \in \Re^{D_x}$, where D_x is the dimension of x_t, the unobserved state $s_t \in \Re^{D_s}$, where D_s is the dimension of s_t, the unidimensional discrete-choice–specific shock $\varepsilon_{jt} \in \Re$, and the L_j-dimensional vector of continuous-choice–specific shocks $r_{tj} = (r_{t1}, \cdots, r_{tL_j}) \in \Re^{L_j}$. Let $z_t = (x_t, s_t)$, $e_{tj} = (\varepsilon_{tj}, r_{tj})$, and $e_t = (e_{t1}, \cdots, e_{tJ})$. The probability density function of (z_{t+1}, e_{t+1}) given (z_t, e_t) and action (j, c_{tj}) is taken in period t is denoted by $f_{jt}(z_{t+1}, e_{t+1} | z_t, e_{tj}, c_{tj})$. The shocks associated with alternative (j, c_{tj}) in period t, e_{tj}, are observed to the individual at the beginning of period t. The individual's conditional direct current period payoff from choosing alternative (j, c_{tj}) in period t is denoted by $u_{tj}(z_t, c_{tj}, r_{tj}) + \varepsilon_{tj}$.

Define $y_{tj} = (d_{tj}, c_{tj})$. The individual chooses the vector $y_t = (y_{t1}, \cdots, y_{tJ})$ to sequentially maximize the expected discounted sum of payoffs:

$$E\left\{\sum_{t=1}^{T}\sum_{j=1}^{J}\beta^{t-1}d_{tj}[u_{tj}(z_t,c_{tj},r_{tj})+\varepsilon_{tj}]\right\}, \tag{2.1}$$

where $\beta\in(0,1)$ is the discount factor. In each period, t, the expectation is taken over $z_{t+1},\cdots,z_T$ and $e_{t+1},\cdots,e_T$. The solution to maximizing expression (2.1) is a Markov decision rule for optimal choice conditional on the time-specific state vectors and i.i.d. shocks. Let the optimal decision rule at period t be given by $(d^0_{tj}(z_t,e_t),c^0_{tj}(z_t,e_t))$. Let the ex-ante value function in period t, $V_t(z_t,r_t)$, be the discounted sum of expected future payoffs, before ε_t is revealed, given the optimal decision rule:

$$V_t(z_t,r_t)=E\left\{\sum_{\tau=t}^{T}\sum_{j=1}^{J}\beta^{\tau-t}d^0_{\tau j}(z_\tau,e_\tau)[u_{\tau j}(z_\tau,c^0_{\tau j}(z_\tau,e_\tau),r_{\tau j})+\varepsilon_{\tau j}]\right\}.$$

As is standard in dicrete/continuous models, the additive separability of the utility function implies the discrete-choice–specific continuous choice is a function of their associated shocks and not of ε_t. Assume that $f_{jt}(z_{t+1},e_{t+1}|z_t,e_{tj},c_{tj})=f_{jt}(z_{t+1}|z_t,c_{tj})g_r(r_{t+1})g_\varepsilon(\varepsilon_{t+1})$, where g_r is the density function of r_t and g_ε be the density function of ε. The expected value function in period $t+1$, given z_t, r_{tj}, the discrete choice, j, and corresponding optimal continuous choice, $c^0_{tj}(z_t,r_{tj})$, is

$$\bar{V}_{t+1,j}(z_t,r_t)=\beta\int V_{t+1}(z_{t+1},r_{t+1})f_{jt}(z_{t+1}|z_t,c^0_{tj}(z_t,r_{tj}))g_r(r_{t+1})dr_{t+1}dz_{t+1}.$$

If behavior is governed by a Markov decision rule, then $V_t(z_t)$ can be written recursively:

$$\begin{aligned}V_t(z_t,r_t)&=E\left\{\sum_{j=1}^{J}d^0_{tj}(z_t,e_t)\left[u_{tj}(z_t,c^0_{tj}(z_t,r_{tj}),r_{tj})+\varepsilon_{tj}+\beta\bar{V}_{t+1,j}(z_t,r_t)\right]\right\}\\&=\int\sum_{j=1}^{J}d^0_{tj}(z_t,e_t)\left[u_{tj}(z_t,c^0_{tj}(z_t,r_{tj}),r_{tj})+\varepsilon_{tj}+\beta\bar{V}_{t+1,j}(z_t,r_t)\right]g_\varepsilon(\varepsilon_t)d\varepsilon_t,\\&=\int\sum_{j=1}^{J}d^0_{tj}(z_t,e_t)\left[v_{tj}(z_t,c^0_{tj}(z_t,r_{tj}),r_{tj})+\varepsilon_{tj}\right]g_\varepsilon(\varepsilon_t)d\varepsilon_t\end{aligned}$$

where

$$v_{tj}(z_t, c^0_{tj}(z_t, r_{tj}), r_{tj}) = u_{tj}(z_t, c^0_{tj}(z_t, r_{tj}), r_{tj}) + \bar{V}_{t+1,j}(z_t, r_t), \tag{2.2}$$

the choice-specific conditional value function without ε_{tj}. The optimal conditional continuous choices, given the discrete alternative j being chosen in period t, satisfy

$$\frac{\partial}{\partial c_{tl_j}} v_{tj}(z_t, c^0_{tj}(z_t, r_{tj}), r_{tj}) = 0, \tag{2.3}$$

for $l_j = 1, \cdots, L_j$. Given the optimal conditional continuous choice, $c^0_{tj}(z_t, r_{tj})$, the individual's discrete choice of alternative j is optimal if

$$d^0_{tj}(z_t, e_t) = \begin{cases} 1 \quad \text{if} \quad v_{tj}(z_t, c^0_{tj}(z_t, r_{tj}), r_{tj}) + \varepsilon_{tj} > v_{tk}(z_t, c^0_{tk}(z_t, r_{tk}), r_{tk}) + \varepsilon_{tk} \quad \forall k \neq j \\ 0 \qquad \text{otherwise} \end{cases} \tag{2.4}$$

Finally, the optimal unconditional continuous choice, $c_{tj}(z_t, r_{tj})$, is given by

$$c_{tj}(z_t, e_{tj}) = d^0_{tj}(z_t, e_t) c^0_{tj}(z_t, r_{tj}). \tag{2.5}$$

2.2 Alternative representation

The probability of choosing alternative j at time t, conditional on z_t, r_t, and the vector of choice-specific optimal conditional continuous choices, $c^0_t = (c^0_{t1}, \cdots, c^0_{tJ})$ is given by

$$p_{tj}(z_t, r_t) = E[d^0_{tj}(z_t, e_t) | z_t, r_t] = \int d^0_{tj}(z_t, r_t, \varepsilon_t) g_\varepsilon(\varepsilon_t) d\varepsilon_t, \tag{2.6}$$

so that, for all (z_t, r_t), $\sum_{j=1}^{J} p_{tj}(z_t, r_t) = 1$, and $p_{tj}(z_t, r_t) > 0$ for all j. Let $p_t(z_t, r_t) = (p_{t1}(z_t, r_t), \cdots, p_{tJ}(z_t, r_t))$ be the vector of conditional choice probabilities. Lemma 1 of Arcidiacono and Miller (2010) show a function $\psi : [0,1]^J \mapsto \Re$ exists such that, for $k = 1, \cdots, J$

$$\psi_k(p_t(z_t, r_t)) \equiv V_t(z_t, r_t) - v_{tk}(z_t, c^0_{tk}(z_t, r_{tk}), r_{tk}). \tag{2.7}$$

Equation (2.7) is simply equation (3.5) of Arcidiacono and Miller (2010), modified so the choice probabilities and value functions are also conditional on the i.i.d. shocks associated with the conditional continuous choices. Our key insight is, given equation (2.7) holds for

$k = 1, \cdots, J$, then for any J-dimensional vector of real numbers $a_t = (a_{t1}, \cdots, a_{tJ})$ such that $\sum_{k=1}^{J} a_{tk} = 1$, we have

$$V_t(z_t, r_t) = \sum_{k=1}^{J} a_{tk}[v_{tk}(z_t, c_{tk}^0(z_t, r_{tk}), r_{tk}) + \psi_k(p_t(z_t, r_t))]. \tag{2.8}$$

Let $a_{t+1,j} = (a_{t+1,1j}, \cdots, a_{t+1,J,j})$, possibly depending on $(z_t \cdots, z_T)$ be the weights associated with the initial discrete choice, j, in period t. Substituting equation (2.8) into equation (2.2) gives:

$$\begin{aligned} & v_{tj}(z_t, c_{tj}^0(z_t, r_{tj}), r_{tj}) = u_{tj}(z_t, c_{tj}^0(z_t, r_{tj}), r_{tj}) \\ & + \beta \sum_{k=1}^{J} \int [v_{t+1,k}(z_{t+1}, c_{tk}^0(z_t, r_{tk}), r_{t+1,k}) \\ & + \psi_k(p_{t+1}(z_{t+1}, r_{t+1}))] a_{t+1,kj} g_r(r_{t+1}) dr_{t+1} f_{jt}(z_{t+1}|z_t, c_{tj}^0(z_t, r_{tj})) dz_{t+1}, \end{aligned} \tag{2.9}$$

Equation (2.9) shows the value function conditional on (z_t, r_t) can be written as the flow payoff of the choice plus *any* weighted sum of a function of the one period ahead CCPs plus the one period ahead conditional value functions, where the weights sum to one. This modest extension of the results of Arcidiacono and Miller (2010) provides a powerful tool for obtaining finite dependence in any model which can be formulated as the one developed in the previous section.

Clarifying example

In order to make clear the alternative representation, we provide a "stripped down" example of the model formation. In this example we abstract away from the conditional continuous choice and consider the case where $J = 2$. We also assume the individual-time–specific discrete-choice shock, ε_{itj}, is distributed i.i.d., type 1 logit. Under these assumptions, the choice-specific conditional value function in equation (2.2) becomes

$$v_{tj}(z_t) = u_{tj}(z_t) + \bar{V}_{t+1,j}(z_t), \tag{2.10}$$

where

$$\bar{V}_{t+1,j}(z_t) = \beta \int \ln \sum_{k=1}^{2} e^{v_{t+1,k}(z_{t+1})} f_{jt}(z_{t+1}|z_t) dz_{t+1} + \beta\gamma, \tag{2.11}$$

where γ is the Euler constant. Equation (2.2) becomes

$$v_{tj}(z_t) = u_{tj}(z_t) + \beta \int \ln \sum_{k=1}^{2} e^{v_{t+1,k}(z_{t+1})} f_{jt}(z_{t+1}|z_t) dz_{t+1} + \beta\gamma. \tag{2.12}$$

Also, the conditional choice probability of alternative $j = 1,2$ is given by

$$p_{tj}(z_t) = \frac{e^{v_{tj}(z_t)}}{\sum_{k=1}^{2} e^{v_{tk}(z_t)}}. \tag{2.13}$$

From equation (2.13), we have the following equality for $j = 1,2$,

$$\ln \sum_{k=1}^{2} e^{v_{tk}(z_t)} = v_{tj}(z_t) - \ln p_{tj}(z_t). \tag{2.14}$$

Notice equation (2.14) is simply equation (2.7) under the assumptions of this example. Also, note the LHS of equation (2.14) evaluated at period $t+1$ is the term inside the integral on the RHS of equation (2.11). For alternative $j = 1,2$, let $a_{t+1,kj}$ be weights associated with alternative j in period t and alternative k in period $t+1$, with $a_{t+1,1j} + a_{t+1,2j} = 1$, $j = 1,2$. Then from equation (2.13) we have

$$\ln \sum_{k=1}^{2} e^{v_{t+1,k}(z_{t+1})} = \sum_{k=1}^{2} a_{t+1,kj}[v_{t+1,k}(z_{t+1}) - \ln p_{t+1,k}(z_{t+1})]. \tag{2.15}$$

Substituting equation (2.15) into equation (2.11) obtains

$$\begin{aligned} \bar{V}_{t+1,j}(z_t) = \beta \int \sum_{k=1}^{2} [v_{t+1,k}(z_{t+1}) - \ln p_{t+1,k}(z_{t+1})] a_{t+1,kj} f_{jt}(z_{t+1}|z_t) dz_{t+1} \\ + \beta\gamma. \end{aligned} \tag{2.16}$$

Now, substituting $\bar{V}_{t+1,j}$ from equation (2.16) into equation (2.12), obtains

$$\begin{aligned} v_{tj}(z_t) = u_{tj}(z_t) + \beta \int \sum_{k=1}^{2} [v_{t+1,k}(z_{t+1}) - \ln p_{t+1,k}(z_{t+1})] a_{t+1,kj} f_{jt}(z_{t+1}|z_t) dz_{t+1} \\ + \beta\gamma. \end{aligned} \tag{2.17}$$

2.3 Generalized finite dependence

The purpose of this section is to show how the weights, $\{a_{\tau,k,j}\ ,\tau \geq t+1,\ k,j=1,\cdots,J\}$, may be used to obtain finite dependence. We begin by showing this result holds for the clarifying example.

Clarifying example contd.

Evaluating equation (2.17) at period $t+1$, and substituting into equation (2.10) obtains

$$\begin{aligned} v_{tj}(z_t) &= u_{tj}(z_t) + \beta \int \sum_{k=1}^{2} [u_{t+1,k}(z_{t+1}) - \ln p_{t+1,k}(z_{t+1})] a_{t+1,kj} f_{jt}(z_{t+1}|z_t) dz_{t+1} \\ &\quad + \beta^2 \int V_{t+2}(z_{t+2}) \left[\int \sum_{k=1}^{2} a_{t+1,kj} f_{k,t+1}(z_{t+2}|z_{t+1}) f_{jt}(z_{t+1}|z_t) dz_{t+1} \right] dz_{t+2} \\ &\quad + \beta\gamma. \end{aligned} \tag{2.18}$$

Equation (2.18) can be used to write the difference in the choice-specific conditional value function as follows,

$$\begin{aligned} & v_{t2}(z_t) - v_{t1}(z_t) = u_{t2}(z_t) - u_{t1}(z_t) \\ & + \beta \int \sum_{k=1}^{2} [u_{t+1,k}(z_{t+1}) - \ln p_{t+1,k}(z_{t+1})] \\ & \times [a_{t+1,k2} f_{2t}(z_{t+1}|z_t) - a_{t+1,k1} f_{1t}(z_{t+1}|z_t)] dz_{t+1} \\ & + \beta^2 \int V_{t+2}(z_{t+2}) \\ & \times \left[\int \sum_{k=1}^{2} f_{k,t+1}(z_{t+2}|z_{t+1}) [a_{t+1,k2} f_{2t}(z_{t+1}|z_t) - a_{t+1,k1} f_{1t}(z_{t+1}|z_t)] dz_{t+1} \right] dz_{t+2}. \end{aligned} \tag{2.19}$$

Finite dependence is obtained if $\{a_{t+1,kj},\ \ k,j=1,2\}$ satisfies

$$\int \sum_{k=1}^{2} f_{k,t+1}(z_{t+2}|z_{t+1}) [a_{t+1,k2} f_{2t}(z_{t+1}|z_t) - a_{t+1,k1} f_{1t}(z_{t+1}|z_t)] dz_{t+1} = 0, \tag{2.20}$$

$$\sum_{k=1}^{2} a_{t+1,kj} = 1, \text{ and,} \tag{2.21}$$

$$a_{t+1,k^*2} f_{2t}(z_{t+1}|z_t) \neq a_{t+1,k^*1} f_{1t}(z_{t+1}|z_t) \quad \text{for at least one } k^* \in \{1,2\}. \tag{2.22}$$

The following presents an example of how to find $\{a_{t+1,kj}, k, j = 1, 2\}$ that satisfies equations (2.20)-(2.22). First, for any $c \in \Re$, setting $a_{t+1,11} = c$, substituting this and equation (2.21) into equation (2.20), and solving for $a_{t+1,12}$ gives

$$a_{t+1,12} = \frac{f_{2,t+1}(z_{t+2}|z_{t+1})\, f_{2t}(z_{t+1}|z_t) + c\, f_{1,t+1}(z_{t+2}|z_{t+1})\, f_{1t}(z_{t+1}|z_t) + (c-1)\, f_{2,t+1}(z_{t+2}|z_{t+1})\, f_{1t}(z_{t+1}|z_t)}{f_{2t}(z_{t+1}|z_t)[f_{1,t+1}(z_{t+2}|z_{t+1}) + f_{2,t+1}(z_{t+2}|z_{t+1})]}$$

$$a_{t+1,22} = 1 - \frac{f_{2,t+1}(z_{t+2}|z_{t+1})\, f_{2t}(z_{t+1}|z_t) + c\, f_{1,t+1}(z_{t+2}|z_{t+1})\, f_{1t}(z_{t+1}|z_t) + (c-1)\, f_{2,t+1}(z_{t+2}|z_{t+1})\, f_{1t}(z_{t+1}|z_t)}{f_{2t}(z_{t+1}|z_t)[f_{1,t+1}(z_{t+2}|z_{t+1}) + f_{2,t+1}(z_{t+2}|z_{t+1})]}.$$

Second, check if equation (2.22) is satisfied. Indeed, equation (2.22) is satisfied so long as $f_{1t}(z_{t+1}|z_t) \neq f_{2t}(z_{t+1}|z_t)$.

Interestingly, for the weights calculated in the previous paragraph,

$$\begin{aligned} & v_{t2}(z_t) - v_{t1}(z_t) = u_{t2}(z_t) - u_{t1}(z_t) \\ & + \beta \int \Bigg([u_{t+1,1}(z_{t+1}) - \ln p_{t+1,1}(z_{t+1})] \frac{f_{2,t+1}(z_{t+2}|z_{t+1})}{f_{1,t+1}(z_{t+2}|z_{t+1}) + f_{2,t+1}(z_{t+2}|z_{t+1})} \\ & + [u_{t+1,2}(z_{t+1}) - \ln p_{t+1,2}(z_{t+1})] \frac{f_{1,t+1}(z_{t+2}|z_{t+1})}{f_{1,t+1}(z_{t+2}|z_{t+1}) + f_{2,t+1}(z_{t+2}|z_{t+1})} \Bigg) \\ & \times [f_{2t}(z_{t+1}|z_t) - f_{1t}(z_{t+1}|z_t)] dz_{t+1}. \end{aligned} \tag{2.23}$$

Note that equation (2.23) holds for any z_{t+2}, so for any density function, $h(z_{t+2})$, defined on the support of z_{t+2},

$$\begin{aligned} & v_{t2}(z_t) - v_{t1}(z_t) = u_{t2}(z_t) - u_{t1}(z_t) \\ & + \beta \int \int \Bigg([u_{t+1,1}(z_{t+1}) - \ln p_{t+1,1}(z_{t+1})] \frac{f_{2,t+1}(z_{t+2}|z_{t+1})}{f_{1,t+1}(z_{t+2}|z_{t+1}) + f_{2,t+1}(z_{t+2}|z_{t+1})} \\ & + [u_{t+1,2}(z_{t+1}) - \ln p_{t+1,2}(z_{t+1})] \frac{f_{1,t+1}(z_{t+2}|z_{t+1})}{f_{1,t+1}(z_{t+2}|z_{t+1}) + f_{2,t+1}(z_{t+2}|z_{t+1})} \Bigg) h(z_{t+2}) dz_{t+2} \\ & \times [f_{2t}(z_{t+1}|z_t) - f_{1t}(z_{t+1}|z_t)] dz_{t+1}. \end{aligned} \tag{2.24}$$

Also, notice that, for at least in this example, while the weights depend on the choice of c, the difference in the conditional value functions does not. It is an open question whether this invariance holds for the general case, which we now present.

Define $f_{jt}(z_{t+1}|z_t) = \int f_{jt}(z_{t+1}|z_t, r_{tj}) g_r(r_{tj}) dr_{tj}$. For any initial choice (j, c_{tj}), for periods $\tau = \{t+1, \cdots, t+\rho\}$, and any corresponding sequence $a_\tau = \{a_{\tau kj}, k, j = 1, \cdots, J\}$ with $\sum_{k=1}^{J} a_{\tau kj} = 1$, define

$$\kappa_{\tau j}(z_{\tau+1}, |z_t, r_{tj}) = \begin{cases} f_{jt}(z_{t+1}|z_t, r_{tj}) & \text{for } \tau = t \\ \int \sum_{k=1}^{J} a_{\tau+1,kj} f_{k\tau}(z_{\tau+1}|z_\tau) \kappa_{\tau-1,j}(z_\tau|z_t, r_{tj}) dz_\tau & \text{for } \tau = t+1, \cdots, t+\rho \end{cases}. \tag{2.25}$$

Because $\sum_{k=1}^{J} a_{\tau kj} = 1$, $\int \kappa_{\tau j}(z_{\tau+1}, |z_t) dz_{\tau+1} = 1$. This restriction does not require $a_j \geq 0$. By forward substitution, equations (2.9) and (2.25) obtain

$$\begin{aligned} & v_{tj}(z_t, c_{tj}^0(z_t, r_{tj}), r_{tj}) = u_{tj}(z_t, c_{tj}^0(z_t, r_{tj}), r_{tj}) \\ & + \sum_{\tau=t+1}^{t+\rho} \sum_{k=1}^{J} \int \beta^{\tau-t} [u_{\tau k}(z_\tau, c_{\tau k}^0(z_\tau, r_{\tau k}), r_{\tau k}) + \psi_k[p_\tau(z_\tau, r_\tau)]] a_{\tau kj} g_r(r_\tau) \kappa_{\tau-1,j}(z_\tau|z_t, r_{tj}) dr_\tau dz_\tau \\ & + \beta^{t+\rho+1} \int V_{t+\rho+1}(z_{t+\rho+1}, r_{t+\rho+1}) g_r(r_{t+\rho+1}) \kappa_{t+\rho+1,j}(z_{t+\rho+1}|z_t, r_{tj}) dr_{t+\rho+1} dz_{t+\rho+1}. \end{aligned} \tag{2.26}$$

Equation (2.26) shows the alternative-specific conditional value functions can be represented as depending on the ρ future sequence of (possibly non-optimal) choice probabilities $\mathbf{p}_{t+1} = (p_{t+1}(z_{t+1}, r_{t+1}), \cdots, p_{t+\rho}(z_{t+\rho}, r_{t+\rho}))$, the $\rho+1$ optimal alternative-specific continuous choices $\mathbf{c}_t^0 = (c_t^0(z_t, r_t), \cdots, c_{t+\rho}^0(z_{t+\rho}, r_{t+\rho}))$, and the $t+\rho+1$ continuation value so that

$$v_{tj}(z_t, c_{tj}^0(z_t, r_{tj}), r_{tj}) = v_{tj}(z_t, \mathbf{c}_t^0, \mathbf{p}_{t+1}, r_{tj}). \tag{2.27}$$

In what follows, we suppress this dependence on $(\mathbf{c}_t^0, \mathbf{p}_{t+1})$ and reintroduce them when clarity is required. Using equation (2.26), we can therefore express the difference in the conditional value functions associated with two alternative initial choices, j and j' as

$$
\begin{aligned}
& v_{tj}(z_t, r_{tj}) - v_{tj'}(z_t, r_{tj'}) = u_{tj}(z_t, r_{tj}) - u_{tj'}(z_t, r_{tj'}) \\
& + \sum_{\tau=t+1}^{t+\rho} \sum_{k=1}^{J} \int \beta^{\tau-t} [u_{\tau k}(z_\tau, r_{\tau k}) + \psi_k[p_\tau(z_\tau, r_\tau)]] g_r(r_\tau) dr_\tau \\
& \times [a_{\tau kj} \kappa_{\tau-1,j}(z_\tau | z_t, r_{tj}) - a_{\tau kj'} \kappa_{\tau-1,j'}(z_\tau | z_t, r_{tj'})] dz_\tau \\
& + \beta^{t+\rho+1} \int V_{t+\rho+1}(z_{t+\rho+1}, r_{t+\rho+1}) g_r(r_{t+\rho+1}) dr_{t+\rho+1} \\
& \times [\kappa_{t+\rho,j}(z_{t+\rho+1} | z_t, r_{tj}) - \kappa_{t+\rho,j'}(z_{t+\rho+1} | z_t, r_{tj'})] dz_{t+\rho+1}.
\end{aligned} \tag{2.28}
$$

Therefore, we say a pair of initial choices, (j, c_{tj}) and $(j', c_{tj'})$ exhibit **generalized ρ-period dependence** if corresponding sequences $(a_{t+1,j}, \cdots, a_{t+\rho,j})$, and $(a_{t+1,j'}, \cdots, a_{t+\rho,j'})$ exist such that

$$
\kappa_{t+\rho,j}(z_{t+\rho+1} | z_t, r_{tj}) = \kappa_{t+\rho,j'}(z_{t+\rho+1} | z_t, r_{tj'}),
$$

almost everywhere, and for at least one $k^* \in \{1, \cdots, J\}$ and $\tau \in \{t+1, \cdots, t+\rho\}$,

$$
a_{\tau k^* j} \kappa_{\tau-1,j}(z_\tau | z_t, r_{tj}) \neq a_{\tau k^* j'} \kappa_{\tau-1,j'}(z_\tau | z_t, r_{tj'}).
$$

We now show that this generalization of the finite dependence property can be used to obtain one-period dependence for any model that satisfies the setup given in the previous section. For initial choice (j, c_{tj}),

$$
\kappa_{t+1,j}(z_{t+2} | z_t) = \int \sum_{k=1}^{J} a_{t+1,kj} f_{kt+1}(z_{t+2} | z_{t+1}) f_{jt}(z_{t+1} | z_t, r_{tj}) dz_{t+1},
$$

so for any pair of initial choices, (j, c_{tj}) and $(j', c_{tj'})$,

$$
\begin{aligned}
& \kappa_{t+1,j}(z_{t+2} | z_t, r_{tj}) - \kappa_{t+1,j'}(z_{t+2} | z_t, r_{tj'}) \\
& = \int \sum_{k=1}^{J} [a_{t+1,kj} f_{kt+1}(z_{t+2} | z_{t+1}) f_{jt}(z_{t+1} | z_t, r_{tj}) - a_{t+1,kj'} f_{kt+1}(z_{t+2} | z_{t+1}) f_{j't}(z_{t+1} | z_t, r_{tj'})] dz_{t+1} \\
& = \int \sum_{k=1}^{J} f_{kt+1}(z_{t+2} | z_{t+1}) [a_{t+1,kj} f_{jt}(z_{t+1} | z_t, r_{tj}) - a_{t+1,kj'} f_{j't}(z_{t+1} | z_t, r_{tj'})] dz_{t+1}.
\end{aligned} \tag{2.29}
$$

Then a sufficient condition for one-period dependence is $\{(a_{t+1,kj}, a_{t+1,kj'},\ k = 1,\cdots,J\}$ satisfies

$$\int \sum_{k=1}^{J} f_{kt+1}(z_{t+2}|z_{t+1})[a_{t+1,kj} f_{jt}(z_{t+1}|z_t, r_{tj}) - a_{t+1,kj'} f_{j't}(z_{t+1}|z_t, r_{tj'})]dz_{t+1} = 0, \quad j,k = 1,2,$$

$$\sum_{k=1}^{J} a_{t+1,kj} = 1, \quad j = 1\cdots,J, \text{ and,}$$

$$a_{t+1,k^*j} f_{jt}(z_{t+1}|z_t) \neq a_{t+1,k^*j'} f_{j't}(z_{t+1}|z_t) \quad \text{for at least one } k^* \in \{1,\cdots,J\}.$$

Given the volume of alternative choices of weights that obtains one-period finite dependence, we proceed by assuming $\rho = 1$, so

$$\begin{aligned} v_{tj}(z_t, r_{tj}) - v_{tj'}(z_t, r_{tj'}) &= u_{tj}(z_t, r_{tj}) - u_{tj'}(z_t, r_{tj'}) \\ &+ \beta \int \left(\int \sum_{k=1}^{J} [u_{t+1,k}(z_{t+1}, r_{t+1,k}) + \psi_k[p_{t+1}(z_{t+1}, r_{t+1})]] g_r(r_{t+1}) dr_{t+1} \right) \\ &\times [a_{t+1kj} f_{tj}(z_{t+1}|z_t, r_{tj}) - a_{t+1kj'} f_{tj'}(z_{t+1}|z_t, r_{tj'})]dz_{t+1}. \end{aligned} \tag{2.30}$$

2.4 Optimal continuous choice

The alternative representation of the difference in conditional value functions provide a simple and convenient representation of the condition for optimal conditional continuous choice, $c^0_{tj}(z_t, r_{tj})$ given alternative j is chosen. The key is to note $\partial v_{tj'}(z_t, c^0_{tj'}(z_t, r_{tj'}), r_{tj'})/\partial c_{tl_j} = 0$

for $j' \neq j$ and $l_j = 1, \cdots, L_j$. This equality, and equation (2.30) imply $c^0_{tj}(z_t, r_{tj})$ solves

$$
\begin{aligned}
0 &= \frac{\partial}{\partial c_{tl_j}} u_{tj}(z_t, c^0_{tj}(z_t, r_{tj}), r_{tj}) \\
&+ \beta \int \left(\int \sum_{k=1}^{J} \frac{\partial}{\partial c_{tl_j}} [u_{t+1,k}(z_{t+1}, c^0_{t+1,k}(z_{t+1}, r_{t+1,k}), r_{t+1,k}) + \psi_k[p_{t+1}(z_{t+1}, r_{t+1})]] g_r(r_{t+1}) dr_{t+1} \right) \\
&\times [a_{t+1kj} f_{tj}(z_{t+1}|z_t, r_{tj}) - a_{t+1kj'} f_{tl}(z_{t+1}|z_t, r_{tj'})] dz_{t+1} \\
&+ \beta \int \left(\int \sum_{k=1}^{J} [u_{t+1,k}(z_{t+1}, c^0_{t+1,k}(z_{t+1}, r_{t+1,k}), r_{t+1,k}) + \psi_k[p_{t+1}(z_{t+1}, r_{t+1})]] g_r(r_{t+1}) dr_{t+1} \right) \\
&\times \frac{\partial}{\partial c_{tl_j}} [a_{t+1kj} f_{tj}(z_{t+1}|z_t, r_{tj}) - a_{t+1kj'} f_{tl}(z_{t+1}|z_t, r_{tj'})] dz_{t+1} \qquad (2.31)
\end{aligned}
$$

The key to this solution is to note $a_{t+1kj'}$ will typically be a function of $c^0_{tj'}$.

Clarifying example contd.

To continue the clarifying example, suppose a unidimensional continuous choice is associated with alternative 2, and period specific utilities are not functions of lagged continuous choice. Then equation (2.24) implies the optimality condition for $c^0_{t2}(z_t, r_{t2})$ is

$$
\begin{aligned}
0 = & \frac{\partial}{\partial c_{t2}} u_{t2}(z_t, c^0_{t2}(z_t, r_{t2}), r_{t2}) \\
& + \beta \int \int [u_{t+1,2}(z_{t+1}, r_{t+1,2}) - \ln p_{t+1,2}(z_{t+1}, r_{t+1})] \\
& \times \left(\int \frac{f_{1,t+1}(z_{t+2}|z_{t+1}, r_{t+1,1})}{f_{1,t+1}(z_{t+2}|z_{t+1}, r_{t+1,1}) + f_{2,t+1}(z_{t+2}|z_{t+1}, r_{t+1,2})} h(z_{t+2}) dz_{t+2} \right) g_r(r_{t+1}) dr_{t+1} \\
& \times \frac{\partial}{\partial c_{t2}} f_{t2}(z_{t+1}|z_t, r_{tj}) dz_{t+1}. \qquad (2.32)
\end{aligned}
$$

2.5 Correlated unobserved heterogeneity

Recall that $z_t = (x_t, s_t)$, where x_t is a vector of observable state variables and s_t is a vector of unobserved state variables. Let w_t by a subset of x_t, and assume that for $j = 1, \cdots, J$

$$f_{jt}(z_{t+1}|z_t, c^0_{tj}(z_t, r_{tj}) = f_{jt}(x_{t+1}|x_t, s_t, c^0_{tj}(z_t, r_{tj}))\pi(s_{t+1}|s_t, w_t).$$

Assume that s_t is discretely distributed with Q support points, $s_t \in \{s_1, \cdots, s_Q\}$. Define $\pi_{q'q}(w_t)$ be the probability of being in state q in period t and q' in period $t+1$, conditional on w_t. Define $\pi_{q'|q}(w_t)$ to be the probability of being in state q' given being is state q in period t and w_t. Then equation (2.30) becomes

$$\begin{aligned} & v_{tj}(x_t, s_q, r_{tj}) - v_{tj'}(x_t, s_q, r_{tj'}) = u_{tj}(x_t, s_q, r_{tj}) - u_{tj'}(x_t, s_q, r_{tj'}) \\ & + \beta \sum_{q'=1}^{Q} \int \left(\int \sum_{k=1}^{J} [u_{t+1,k}(x_{t+1}, s_{q'}, r_{t+1,k}) + \psi_k[p_{t+1}(x_{t+1}, s_{q'}, r_{t+1})]] g_r(r_{t+1}) dr_{t+1} \right) \\ & \times [a_{t+1kj} f_{tj}(x_{t+1}|x_t, s_q, r_{tj}) - a_{t+1kj'} f_{tj'}(x_{t+1}|x_t, s_q, r_{tj'})] dz_{t+1} \pi_{q'|q}(w_t). \end{aligned} \tag{2.33}$$

Then the probability of choosing alternative j at time t, conditional on x_t, s_q, r_t, and the vector of choice-specific optimal conditional continuous choices, $c_t^0 = (c_{t1}^0, \cdots, c_{tJ}^0)$ is given by

$$p_{tj}^0(x_t, s_q, r_t) = E[d_{tj}^0(z_t, e_t)|x_t, s_q, r_t] = \int d_{tj}^0(z_t, r_t, \varepsilon_t) g_\varepsilon(\varepsilon_t) d\varepsilon_t. \tag{2.34}$$

The probability of choosing alternative j at time t, conditional on s_q and x_t is

$$p_{tj}^0(x_t, s_q) = \int p_{tj}^0(x_t, s_q, r_t) g_r(r_t) dr_t. \tag{2.35}$$

The probability of choosing alternative j at time t, conditional on x_t is

$$p_{tj}^0(x_t) = \sum_{q=1}^{Q} p_{tj}^0(x_t, s_q) \pi_q(w_t), \tag{2.36}$$

where $\pi_q(w_t) = \sum_{q'=1}^{Q} \pi_{q'q}(w_t)$, the (marginal) probability of being in state q in period t given w_t. Finally, define

$$p_{tt+1,jk}^0(x_{t+1}, x_t, s_{q'}, s_q) = p_{t+1,j}^0(x_{t+1}, s_{q'}) p_{t,k}^0(x_t, s_q), \tag{2.37}$$

and

$$p_{tt+1,jk}^0(x_{t+1}, x_t) = \sum_{q'=1}^{Q} \sum_{q=1}^{Q} p_{t+1,j}^0(x_{t+1}, s_{q'}) p_{t,k}^0(x_t, s_q) \pi_{q'q}(w_t). \tag{2.38}$$

3 Identification

In this section, we discuss sufficient conditions for identification the parameters of the model. Define $\Pi(w_t) = \{(s_q, \pi_{q'q}(w_t)),\ q', q = 1, \cdots, Q\}$. Identification is semiparametric in the sense that we impose parametric restrictions on u_{tj}, g_ε, g_r, and $f_{jt}(x_{t+1}|x_t, s_t, r_{tj})$, but we only impose exclusion restrictions on $\Pi(w_t)$.

For each individual unit, the random variables (d_t, c_t, x_t), $t = 1, \cdots, T$ are observable. Hence in the population, the joint distribution $F(d_t, c_t, x_t)$ is observed. For $t = 1, \cdots, T$, $j = 1, \cdots, J, k \neq j$, define $u_{tjk}(z_t, c_t, r_t) = u_{tj}(z_t, c_{tj}, r_{tj}) - u_{tk}(z_t, c_{tk}, r_{tk})$, and $v_{tjk}(z_t, c_t, r_t) = v_{tj}(z_t, c_{tj}, r_{tj}) - v_{tk}(z_t, c_{tk}, r_{tk})$. In what follows, we will use the shorthand notations u_{tjk} and v_{tjk} for $u_{tjk}(z_t, c_t, r_t)$ and $v_{tjk}(z_t, c_t, r_t)$. Let $x_t \in \mathcal{X}_t \subseteq \Re^{D_{x_t}}$, where D_{x_t} is the dimension of x_t. Define $\mathcal{X} = \prod_{t=1}^{T} \mathcal{X}_t$.

Assumption 3.1. *For $t = 1, \cdots, T$, $j = 1, \cdots, J, k \neq j$,*
1. $\beta \in [0, 1)$ is known.
2. Rank $E[x_t' x_t] = D_{x_t}$, Rank $E[w_t' w_t] = D_{w_t}$, and the conditional density function of x_t given w_t, $f_{x_t|w_t} > 0$.
3. ε_j and ε_k are independent with known common density function, g_ε, which is twice continuously differentiable and log-concave with support $\Re$.
4. $g_r(r_t) = g_r(r_t; \theta_2)$, is continuous and known up to θ_2.
5. $f_{tj}(x_{t+1}|x_t, s_t, c_{tj}) = f_{tj}(x_{t+1}|x_t, s_t, c_{tj}; \theta_3)$ is known up to θ_3 and is twice continuously differentiable in $c_{tl_j}, l_j = 1, \cdots, L_j$.
6. $u_{tj}(z_t, c_{tj}, r_{tj}) = u_{tj}(z_t, c_{tj}, r_{tj}; \theta_1)$ is known up to $\theta_1 \in \Re^{D_{\theta_1}}$, increasing and strictly concave in c_{tl_j}, with $\lim_{c_{tl_j} \to 0} u_{tj}(z_t, c_{tj}, r_{tj}; \theta_1) = -\infty$, $l_j = 1, \cdots, L_j$.
7. Let $\theta = (\theta_1, \theta_2, \theta_3)$. For any $w_t \in \mathcal{X}_t$, and for some $j^* \in \{1, \cdots, J\}$, there is a non-empty set of x_t for which (i) for any θ, either $v_{tj^*k}(x_t, c_{tj}, r_{tj}; \theta, s_{q'}) > v_{tj^*k}(x_t, c_{tj}, r_{tj}; \theta, s_q)$, $s_{q'} > s_q$, or $v_{tj^*k}(x_t, c_{tj}, r_{tj}; \theta, s_{q'}) < v_{tj^*k}(x_t, c_{tj}, r_{tj}; \theta, s_q)$, $s_{q'} > s_q$, and (ii) for $\tilde{\theta} \neq \theta$, and any pair $(\tilde{s}_q, s_q)$, $(\tilde{x}, \overline{x})$ exists in this set for which $v_{tj^*k}(\tilde{x}_t, c_{tj}, r_{tj}; \tilde{\theta}, \tilde{s}_q) < v_{tj^*k}(x_t, c_{tj}, r_{tj}; \theta, s_q)$, and $v_{tj^*k}(\overline{x_t}, c_{tj}, r_{tj}; \tilde{\theta}, \tilde{s}_q) > v_{tj^*k}(x_t, c_{tj}, r_{tj}; \theta, s_q)$.

Define $\mathbf{P}(x; \theta, \Pi) = (p_{t,j}(x_t; \theta, \Pi), t = 1, \cdots, T, j = 1, \cdots, J)$. Let (θ_0, Π_0) be the true parameter vector, that is, the probabilities generated from the model at (θ_0, Π_0) coincides with the true probabilities: $\mathbf{P}(x; \theta_0, \Pi_0) = \mathbf{P}^0(x)$

Theorem 3.2. *Suppose assumption 3.1 holds. Then* (θ_0, Π_0) *is identified in the sense that any* $(\tilde{\theta}, \tilde{\Pi})$ *satisfying* $\mathbf{P}(x; \tilde{\theta}, \tilde{\Pi}) = \mathbf{P}^0(x)$ *implies* $(\tilde{\theta}, \tilde{\Pi}) = (\theta_0, \Pi_0)$.

The proof of theorem 3.2 is found in Gayle (2013).

4 Estimator

In this section, we propose a GMM estimator for the parameters of the model, θ and Π. We choose to propose a GMM estimator instead of the ML estimator for two reasons. First, definition of the GMM estimator does not require specifying the distribution of measurement errors. Second, the GMM estimator is implementable without the need to specify the initial conditions for the process of s_t. We begin by imposing a restrictions on w.

Assumption 4.1. *The random vector* w *is discrete valued with* R *distinct values.*

We impose assumption 4.1 for two reasons. First this structure of the distribution of types affords relatively simple implementation in estimation. Second, this assumption reduces Π to belong in a parametric class, which implies standard parametric asymptotics can be implemented to derive the asymptotic distribution, which in tern leads to the standard asymptotic covariance matrix estimator. Under assumption 4.1, $\Pi(w_t)$ consists of $Q((Q-1)R+1)$ parameters to be estimated. Suppose n observations of the random vector $\{y_{it} = (d_{it}, c_{it}, x_{it}), t = 1, \cdots, T\}$ are independently drawn from $F(d_t, c_t, x_t)$. Let $y_i = (y'_{i3}, \cdots, y'_{iT})'$, $p_i^0 = (p_{i3}^0, \cdots, p_{iT}^0)'$, and $c_i^0 = (c_{i3}^0, \cdots, c_{iT}^0)'$. For each i, and for $t = 3, \cdots, T$, define the residuals

$$\rho_{1itj}(y_i; \theta, \Pi) = d_{itj} - p_{tj}^0(x_{it}; \theta, \Pi), j = 2, \cdots, J,$$
$$\rho_{2itj}(y_i; \theta, \Pi) = c_{itj} - c_{tj}^0(x_{it}; \theta, \Pi), j = 1, \cdots, J,$$
$$\rho_{3tjk}(y_i; \theta, \Pi) = d_{itj} d_{it+1,k} - p_{tt+1,jk}^0(x_{it}; \theta, \Pi), \quad j = 1, \cdots, J-1, k = j+1, \cdots, J.$$

Define

$$\rho_{1it}(y_i;\theta,\Pi) = (\rho_{1itj}(y_i;\theta,\Pi), j = 2,\cdots,J),$$
$$\rho_{2it}(y_i;\theta,\Pi) = (\rho_{2itj}(y_i;\theta,\Pi), j = 1,\cdots,J), \quad \text{and}$$
$$\rho_{3it}(y_i;\theta,\Pi) = (\rho_{3tjk}(y_i;\theta,\Pi), j = 1,\cdots,J-1, k = j+1\cdots,J).$$

Define the $L+(J+2)(J-1)/2$-dimensional residual vector $\rho_{it}(y_i;\theta,\Pi) = (\rho_{1it}, \rho_{2it}, \rho_{3it})'$. Let X_{it} be a $(L+(J+2)(J-1)/2)\times N_t^X$ matrix of instruments, and define the N_t^X-dimensional vector.

$$m_{it}(\theta,\Pi) = X_{it}'\rho_{it}(y_i;\theta,\Pi). \tag{4.1}$$

Define the N^X-dimensional vector $m_i(\theta,\Pi) = (m_{i2}(\theta,\Pi)',\cdots,m_{iT-1}(\theta,\Pi)')'$, where $N^X = \sum_{t=2}^{T-1} N_t^X$, and the N^X-dimensional vector of moments $m(\theta,\Pi) = E[m_i(\theta,\Pi)]$. Let Ω be a $N^X \times N^X$-dimensional symmetric, positive definite weighting matrix. Then from the results of section 3, θ_0 minimizes

$$S(\theta,\Pi) = m(\theta,\Pi)'\Omega m(\theta,\Pi). \tag{4.2}$$

To implement the estimator, we assume the number of types, Q, is known to the investigator. Let

$$\hat{m}(\theta,\Pi) = \frac{1}{n}\sum_{i=1}^{n} m_i(\theta,\Pi). \tag{4.3}$$

Then the GMM estimator for (θ,Π_0), $(\hat{\theta},\hat{\Pi})$ minimizes

$$\hat{S}(\theta,\Pi) = \hat{m}(\theta,\Pi)'\hat{\Omega}\hat{m}(\theta,\Pi), \tag{4.4}$$

where $\hat{\Omega}$ is a consistent estimator for Ω.

5 Computing The Estimator

In this section, we present a method for computing the estimator proposed in the previous section. We describe the algorithm starting with the $o+1$ iteration with $\lambda^{[o]} := (\pi^{[o]}, \mathbf{c}^{0,[o]}, \mathbf{p}^{0,[o]}, s^{[o]}, \theta^{[o]})$ in hand.

Define

$$L^d_{i,tt+1}(d_{it+1}, d_{it}|x_{it+1}, x_{it}, s_{q'}, s_q; \lambda^{[o]}) = \prod_{j=1}^{J}\prod_{k=1}^{J} p^0_{i,tt+1,jk}(x_{it+1}, x_{it}, s_{q'}, s_q; \lambda^{[o]})^{d_{itj}d_{it+1,k}}. \quad (5.1)$$

Then, by Bayes' rule

$$\begin{aligned} &L^s_{i,tt+1}(s_{q'}, s_q|d_{it+1}, d_{it}, x_{it+1}, x_{it}; \lambda^{[o]}) \\ &= \frac{L^d_{i,tt+1}(d_{it+1}, d_{it}|x_{it+1}, x_{it}, s_{q'}, s_q; \lambda^{[o]})\pi^{[o]}_t(s_{q'}, s_q|w_{it})}{\sum_{q'=1}^{Q}\sum_{q=1}^{Q} L^d_{i,tt+1}(d_{it+1}, d_{it}|x_{it+1}, x_{it}, s_{q'}, s_q; \lambda^{[o]})\pi^{[o]}_t(s_{q'}, s_q|w_{it})}. \end{aligned} \quad (5.2)$$

The conditional type probabilities are therefore updated as follows.

$$\pi^{[o+1]}_t(s_{q'}, s_q|w) = \frac{\sum_{i=1}^{n} L^s_{i,tt+1}(s_{q'}, s_q|d_{it+1}, d_{it}, x_{it+1}, x_{it}; \lambda^{[o]}) I_{it}(w)}{\sum_{i=1}^{n} I_{it}(w)}, \quad (5.3)$$

where $I_{it}(w)$ is equal to one if $w_{it} = w$, and zero otherwise.

The alternative-specific continuous choice variables are updated by iterating (over o')

$$\begin{aligned} &c^{0,[o'+1]}_{itj}(x_{it}, r_{itj}; \pi^{[o+1]}, \mathbf{c}^{0,[o]}, \mathbf{p}^{0,[o]}, s^{[o]}, \theta^{[o]}) = c^{0,[o']}_{itj}(x_{it}, r_{itj}; \pi^{[o+1]}, \mathbf{c}^{0,[o]}, \mathbf{p}^{0,[o]}, s^{[o]}, \theta^{[o]}) \\ &- \left[\frac{\partial^2}{\partial c^2_{itj}} v_{itj}(x_{it}, c^{[o']}_{itj}(x_{it}, r_{itj}; \pi^{[o+1]}, \mathbf{c}^{0,[o]}, \mathbf{p}^{0,[o]}, s^{[o]}, \theta^{[o]})\right]^{-1} \\ &\times \left[\frac{\partial}{\partial c_{itj}} v_{itj}(x_{it}, c^{[o']}_{itj}(x_{it}, r_{itj}; \pi^{[o+1]}, \mathbf{c}^{0,[o]}, \mathbf{p}^{0,[o]}, s^{[o]}, \theta^{[o]})\right], \end{aligned} \quad (5.4)$$

until convergence, where the procedure is initialized by $c^{0,[o]}_{it}$. Denote the fixed point by $c^{0,[o+1]}_{it}$.

The conditional choice probabilities can be updated as follows:

$$p_{itj}^{[o+1]}(x_{it};\theta^{[o]}) = \Psi_j(v_{itj}(x_{it};\pi^{[o+1]},\mathbf{c}^{0,[o+1]},\mathbf{p}^{0,[o]},s^{[o]},\theta^{[o]})). \tag{5.5}$$

Finally,(s,θ) is updated as follows

$$\begin{aligned}&(s^{[o+1]},\theta^{[o+1]}) = (s^{[o]},\theta^{[o]})\\ &-\left[\left(\frac{\partial}{\partial\theta}m(\pi^{[o+1]},\mathbf{c}^{0,[o+1]},\mathbf{p}^{0,[o+1]},s^{[o]},\theta^{[o]}))\right)'\hat{\Omega}\left(\frac{\partial}{\partial\theta}m(\pi^{[o+1]},\mathbf{c}^{0,[o+1]},\mathbf{p}^{0,[o+1]},s^{[o]},\theta^{[o]})\right)\right]^{-1}\\ &\times\left(\frac{\partial}{\partial\theta}m(\pi^{[o+1]},\mathbf{c}^{0,[o+1]},\mathbf{p}^{0,[o+1]},s^{[o]},\theta^{[o]})\right)'\hat{\Omega}\quad m(\pi^{[o+1]},\mathbf{c}^{0,[o+1]},\mathbf{p}^{0,[o+1]},s^{[o]},\theta^{[o]}).\end{aligned} \tag{5.6}$$

The full algorithm is as follows.

Main Algorithm.
1– Initialize $\theta^{[0]}\in\Theta$, $\mathbf{p}^{0,[0]}\in[0,1]^{[n(T-1)(J-1)]}$, and $\mathbf{c}^{0,[0]}\in\Re^{n(T-1)L}$.
2– For $o\geq 1$
 2.1– Update $\pi^{[o]}(q',q|w)$ using (5.1 - 5.3)
 2.2– Update $c^{0,[o]}$ using (5.4)
 2.3– Update $p^{0,[o]}$ using equation (5.5)
 2.4– Update $(s^{[o]},\theta^{[o]})$ using equation (5.6)
Until convergence in (s,θ).

Notice that steps 2.1-2.3 of the main algorithm is simply evaluating the objective function at the new trial values of θ, while step 2.4 is a Gauss-Newton step. Hence, the main algorithm converges, and the convergence rate is at most quadratic.

6 Asymptotic properties and consistent covariance estimation

This section provides additional sufficient conditions for consistency and $\sqrt{n}$-asymptotic normality of the estimator, $\hat{\theta}$, of $\theta_0\in\Theta$.

Assumption 6.1. *As sample of n independent realizations is drawn from $F(d,c,x)$. For each $i = 1,\cdots,n$, $(d_{it}, c_{it}, x_{it}, t = 1,\cdots,T)$ is observed.*

Assumption 6.2. *$\hat{\Omega}$ is symmetric and positive definite with $\|\hat{\Omega} - \Omega\| = o_p(1)$.*

Theorem 6.3. *Suppose (i) Assumption 3.1 holds, (ii) Assumption 4.1 hold, and (iii) Assumptions 6.1, and 6.2 hold. Then $\hat{\theta} \xrightarrow{p} \theta_0$, and*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{p} N(0,V),$$

where $V = (M'\Omega M)^{-1}(M'\Omega\Sigma\Omega M)(M'\Omega M)^{-1}$, $M = \partial m(\theta_0)/\partial\theta$, and $\Sigma = E[m_i(\theta_0)m_i(\theta_0)']$.

The proof of Theorem 6.3 is standard and can be found in Newey and McFadden (1994). To estimate V, let $\hat{M} = \partial m(\hat{\theta})/\partial\theta$, $\hat{\Sigma} = \sum_{i=1}^{n} m_i(\hat{\theta})m_i(\hat{\theta})'/n$, and define

$$\hat{V} = (\hat{M}'\hat{\Omega}\hat{M})^{-1}(\hat{M}'\hat{\Omega}\hat{\Sigma}\hat{\Omega}\hat{M})(\hat{M}'\hat{\Omega}\hat{M})^{-1}.$$

Theorem 6.4. *Suppose the conditions of Theorem 6.3 hold. Then $\|\hat{V} - V\| = o_p(1)$.*

Again, the proof of Theorem 6.4 is standard, and can also be found in Newey and McFadden (1994).

References

AGUIRREGABIRIA, V. AND P. MIRA (2002): "Swapping The Nested Fixed Point Algorithm: A Class of Estimators for Distrete Markov Decision Models," *Econometrica*, 70, 1519–1543.

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

ALTUG, S. AND R. A. MILLER (1998): "Effect of Work Experience of Female Wages and Labour Supply," *The Review of Economic Studies*, 65, 45–85.

ARCIDIACONO, P. AND R. MILLER (2010): "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *MIMEO, Duke University*.

BAJARI, P., C. BENKARD, AND J. LEVIN (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75.

DUBIN, J. A. AND D. L. MCFADDEN (1984): "An Econometric Analysis of Residential Electric Applicance Holdings and Consumption," *Econometrica*, 52, 345–362.

GAYLE, G.-L. AND R. A. MILLER (2003): "Life-Cycle Fertility and Human Capital Accumulation," *Unpublished Manuscript, Carnegie Mellon University*.

GAYLE, W.-R. (2006): "A Dynamic Structural Model of Labor Market Supply and Educational Attainment," *MIMEO: University of Virginia*.

——— (2013): "Identification and Estimation of Semiparametric, Correlated Finite Mixture Models," *MIMEO: University of Virginia*.

HANEMANN, M. W. (1984): "Discrete/Continuous Models of Consumer Demand," *Econometrica*, 52, 541–561.

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

HOTZ, J. V., R. A. MILLER, S. SANDERS, AND J. SMITH (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, 61, 265–289.

NEWEY, W. K. AND D. MCFADDEN (1994): *Large Sample Estimation and Hypothesis Testing*, Elsevier Science Publishers.

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