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

Wayne-Roy Gayle*

February 13, 2015

Abstract

This paper investigates conditional choice probability estimation of dynamic structural discrete and continuous choice models. I extend the concept of finite dependence in a way that accommodates non-stationary, irreducible transition probabilities. I show that under this new definition of finite dependence, one-period dependence is obtainable in any dynamic structural model. This finite dependence property also provides a convenient and computationally cheap representation of the optimality conditions for the continuous choice variables. I allow for a 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. I show the estimator is root-n–asymptotically normal. I develop a new and computationally cheap algorithm to compute the estimator. I apply our method to estimate a model of education and labor supply choices to investigate properties of the distribution of returns to education, using data from the National Longitudinal Survey of Youth 1979.

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

*Department of Resource Economics, University of Massachusetts, 80 Campus Way, Amherst, MA 01003, E-mail: wayneroy@gmail.com. The author is grateful to Victor Aguirregabiria, George-Levi Gayle, Lala Zun Ma, Holger Sieg, and Steven Stern for insightful comments. I also thank the seminar participants at the University of Virginia and the University of North Carolina, Chapel Hill. All errors are my own.

1 Introduction

In this paper, I investigate conditional choice probability (CCP) estimation of dynamic structural discrete/continuous choice models with unobserved individual heterogeneity. I show that an extension to the definition of finite dependence proposed in Altug and Miller (1998) and Arcidiacono and Miller (2011) 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. I propose a generalized method of moments (GMM) estimator for the structural parameters of the model and derive their asymptotic distributions. I also propose a simple algorithm to implement the estimator. I implement this method to estimate a model of education and labor supply choices to investigate the distribution of returns to education, using data from the NLSY79.

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 potential for an immense reduction in computational costs 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 to estimate the structural parameters.

Empirical application of 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 that 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 the CCP and NFXP approaches to estimation is that 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 the 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 produces estimates 5 to 15 times faster than NFXP. The method that Aguirregabiria and Mira (2002) propose 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), and Arcidiacono and Miller (2011). 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 (2011) 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 an approach that allows 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 estimate 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 a significant reduction in the 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 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 (2011). 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 makes three separate, but closely related contributions to the literature on CCP estimation of dynamic structural models. I extend the concept of finite dependence to allow for general non-stationary and irreducible transition probabilities. Although its definition is precise and well understood, the strategy to construct finite dependence in dynamic structural models have been largely ad hoc, and often achieved by 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 markets and degenerate transition probability assumptions to form counterfactual strategies that obtain finite dependence. A key insight of Arcidiacono and Miller (2011) is that: "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. I 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 I am now able to choose from a continuum of finite dependence representations, the question becomes whether a choice of weights exists that obtains one-period finite dependence. Indeed, one-period finite dependence 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. My approach to accounting for continuous choices does 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 (2011), and the generalized finite dependence result of this paper.

The model developed in this paper allows for a general form of discrete-valued, unobserved heterogeneity in the utility functions, production functions, and the transition probabilities. The distribution of the unobserved random variables may be correlated with observable covariates of the model. I provide sufficient conditions for identification of all the parameters of the model: those governing preferences, transitions, production functions, and the distribution of the unobservable heterogeneity. I propose a GMM estimator for these parameters and an iterative algorithm to compute them. The algorithm is similar to the one proposed in Arcidiacono and Miller (2011), except for the method of updating the distribution of correlated unobservable heterogeneity (types). The approach proposed in this paper has two advantages. First, my method does not require specifying initial conditions and is therefore robust to misspecification of initial conditions. Second, my approach to updating the distribution of types is by iterated projections. This approach allows for the probability of a potential type to be zero. Therefore, given prior information on the maximum possible number of types, my method does not just compute the probabilities for a given number of types, but also computes the number of types.

To illustrate my method, I estimate a model of educational attainment and labor supply to investigate properties of the distribution of the returns to education, using data from the National Longitudinal Survey of Youth 1979 (NLSY79). Key features of the model are that: (1) I allow for individuals to choose to simultaneously participate in the labor market and enroll in school, (2) I treat hours worked as a continuous choice variable and allow for it to affect the probability of completing the grade level enrolled in, (3) I allow for psychic costs of school attendance and labor market activities, and (4) I treat returns to education as a first-order discrete Markov process with probabilities that depend on race, education, and labor market experience categories. I find that a moderate amount of hours worked per week while in school improves the chances of advancing the grade level, whereas working more than 16 hours per week reduces the probability of advancing the grade level. My estimates of the psychic costs to schooling is qualitatively large, but imprecisely estimated, whereas my estimates of the psychic cost of labor market participation is large and significant. I find significant heterogeneity in the returns to education: the mean returns to education vary between 0.13 and 0.17. Not only do I find heterogeneity in the returns to education, but I also find that some groups of individuals face significant uncertainty about the future returns to education.

The rest of the paper proceeds as follows. Section 2 outlines the class of dynamic structural models investigated in this paper and presents the new alternative representation of the value functions that I use to obtain finite dependence. Section 3 then defines generalized finite dependence, shows that one period finite dependence can be obtained in my class of models, and defines first-order conditions for optimal choices. Section 4 outlines my approach to incorporating correlated unobserved heterogeneity in the model. I propose sufficient conditions for identification of the parameters of the model in section 5. Section 6 proposes a GMM estimator for the parameters. Section 7 outlines the algorithm I propose to compute the estimator, and section 8 presents the asymptotic properties of my estimator. In section 9, I implement my method to estimate a model of educational attainment and labor supply. Section 10 concludes. The Appendix contains the proofs and tables reporting the estimation results from my empirical application in section 9.

2 Model

2.1 General framework

This section outlines the class of dynamic structural discrete/continuous choice models that I consider and my alternative representation that extends that of Arcidiacono and Miller (2011). In each period, t, an individual chooses among J discrete, mutually exclusive, and exhaustive alternatives. Let d_{tj} be 1 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 $x_t \in \mathcal{X}_t \subseteq \Re^{D_{x_t}}$, 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 that action (j, c_{tj}) is taken in period t is denoted by $f_{jt}(z_{t+1}, e_{t+1} | z_t, c_{tj}, e_t)$. The vector of shocks, e_t, 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 $y_t^0 = \{(d_{tj}^0(z_t, e_t), c_{tj}^0(z_t, e_t)), j = 1, \cdots, J\}$. 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:

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

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

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

If a Markov decision rule governs behaviour, $V_t(z_t, r_t)$ can be written recursively:

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

where

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

is 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

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

for $l_j = 1,\cdots,L_j$. Given the optimal conditional continuous choice, $c_t^0(z_t,r_t) = (c_{tk}^0(z_t,r_{tk}), k = 1,\cdots,J)$, the individual's discrete choice of alternative j is optimal if

$$d_{tj}^0(z_t,c_t^0(z_t,r_t),e_t) = \begin{cases} 1 & \text{if} \quad v_{tj}(z_t,c_{tj}^0(z_t,r_{tj}),r_{tj})+\varepsilon_{tj} > v_{tk}(z_t,c_{tk}^0(z_t,r_{tk}),r_{tk})+\varepsilon_{tk} \quad \forall k\neq j \\ 0 & \text{otherwise.}\end{cases} \tag{2.4}$$

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

$$c_{tj}^*(z_t,e_{tj}) = d_{tj}^0(z_t,c_t^0(z_t,r_t),e_t)c_{tj}^0(z_t,r_{tj}). \tag{2.5}$$

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, $c_t^0 = (c_{t1}^0,\cdots,c_{tJ}^0)$, is given by

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

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 (2011) show a function $\psi : [0,1]^J \mapsto \Re$ exists such that, for $k = 1, \cdots, J$,

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

Equation (2.7) is simply equation (3.5) of Arcidiacono and Miller (2011), modified so the choice probabilities and value functions are also conditional on the i.i.d. shocks associated with the conditional continuous choices. The key insight is: if (2.7) holds for $k = 1, \cdots, J$, then for any J-dimensional vector of real numbers $a_t = (a_{t1}, \cdots, a_{tJ})$ with $\sum_{k=1}^{J} a_{tk} = 1$,

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

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 1. This extension of the results of Arcidiacono and Miller (2011) provides a powerful tool for obtaining finite dependence in any model that can be formulated as the one developed in the previous section.

Clarifying example

To clarify the alternative representation, I provide a "stripped down" example of the model formation. In this example, I abstract away from the conditional continuous choice and consider the case in which $J = 2$. I also assume the individual-time–specific discrete-choice shock, ε_{itj}, is distributed i.i.d., type 1 extreme value. 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

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

and 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 period $t+1$ conditional choice probability of alternative $j = 1,2$ is given by

$$p_{t+1,j}(z_t) = \frac{e^{v_{t+1,j}(z_{t+1})}}{\sum_{k=1}^{2} e^{v_{t+1,k}(z_{t+1})}}. \tag{2.13}$$

From equation (2.13), the following equality holds for $j = 1,2$:

$$\ln \sum_{k=1}^{2} e^{v_{t+1,k}(z_{t+1})} = v_{t+1,j}(z_{t+1}) - \ln p_{t+1,j}(z_{t+1}). \tag{2.14}$$

Notice equation (2.14) is simply equation (2.7) under the assumptions of this example and evaluated at period $t+1$. Also, note the LHS of equation (2.14) is a 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),

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

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

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

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. I 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

$$\begin{aligned} v_{tj}(z_t) = {} & u_{tj}(z_t) + \beta \int \sum_{k=1}^{2} [u_{t+1,k}(z_{t+1}) - \ln p_{t+1,k}(z_{t+1})] a_{t+1,kj} f_{jt}(z_{t+1}|z_t) dz_{t+1} \\ & + \beta^2 \int V_{t+2}(z_{t+2}) \left[\int \sum_{k=1}^{2} a_{t+1,kj} f_{k,t+1}(z_{t+2}|z_{t+1}) f_{jt}(z_{t+1}|z_t) dz_{t+1} \right] dz_{t+2} \\ & + \beta\gamma. \end{aligned} \tag{3.1}$$

Equation (3.1) can be used to write the difference in the choice-specific conditional value function as follows:

$$\begin{aligned} & v_{t2}(z_t) - v_{t1}(z_t) = u_{t2}(z_t) - u_{t1}(z_t) \\ & + \beta \int \sum_{k=1}^{2} [u_{t+1,k}(z_{t+1}) - \ln p_{t+1,k}(z_{t+1})] \\ & \times [a_{t+1,k2} f_{2t}(z_{t+1}|z_t) - a_{t+1,k1} f_{1t}(z_{t+1}|z_t)] dz_{t+1} \\ & + \beta^2 \int V_{t+2}(z_{t+2}) \\ & \times \left[\int \sum_{k=1}^{2} f_{k,t+1}(z_{t+2}|z_{t+1}) [a_{t+1,k2} f_{2t}(z_{t+1}|z_t) - a_{t+1,k1} f_{1t}(z_{t+1}|z_t)] dz_{t+1} \right] dz_{t+2}. \end{aligned} \tag{3.2}$$

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

$$\int \sum_{k=1}^{2} f_{k,t+1}(z_{t+2}|z_{t+1})[a_{t+1,k2} f_{2t}(z_{t+1}|z_t) - a_{t+1,k1} f_{1t}(z_{t+1}|z_t)] dz_{t+1} = 0, \tag{3.3}$$

$$\sum_{k=1}^{2} a_{t+1,kj} = 1, \text{ and,} \tag{3.4}$$

$$a_{t+1,k^*2} f_{2t}(z_{t+1}|z_t) \neq a_{t+1,k^*1} f_{1t}(z_{t+1}|z_t) \quad \text{for at least one } k^* \in \{1,2\}. \tag{3.5}$$

The following presents an example of how to find $\{a_{t+1,kj}, k, j = 1,2\}$, which satisfies equations (3.3)-(3.5). First, for any $c \in \Re$, setting $a_{t+1,11} = c$, substituting this and equation (3.4) into equation (3.3), 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 (3.5) is satisfied. Indeed, equation (3.5) is satisfied as 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,

$$\begin{aligned} & v_{t2}(z_t) - v_{t1}(z_t) = u_{t2}(z_t) - u_{t1}(z_t) \\ & + \beta \int \Bigg([u_{t+1,1}(z_{t+1}) - \ln p_{t+1,1}(z_{t+1})] \frac{f_{2,t+1}(z_{t+2}|z_{t+1})}{f_{1,t+1}(z_{t+2}|z_{t+1}) + f_{2,t+1}(z_{t+2}|z_{t+1})} \\ & + [u_{t+1,2}(z_{t+1}) - \ln p_{t+1,2}(z_{t+1})] \frac{f_{1,t+1}(z_{t+2}|z_{t+1})}{f_{1,t+1}(z_{t+2}|z_{t+1}) + f_{2,t+1}(z_{t+2}|z_{t+1})} \Bigg) \\ & \times [f_{2t}(z_{t+1}|z_t) - f_{1t}(z_{t+1}|z_t)] dz_{t+1}. \end{aligned} \tag{3.6}$$

Note that equation (3.6) holds for any z_{t+2}, so for any density function, $h(z_{t+2})$, defined on the support of z_{t+2},

$$\begin{aligned}
& v_{t2}(z_t) - v_{t1}(z_t) = u_{t2}(z_t) - u_{t1}(z_t) \\
& + \beta \int \int \Bigg([u_{t+1,1}(z_{t+1}) - \ln p_{t+1,1}(z_{t+1})] \frac{f_{2,t+1}(z_{t+2}|z_{t+1})}{f_{1,t+1}(z_{t+2}|z_{t+1}) + f_{2,t+1}(z_{t+2}|z_{t+1})} \\
& + [u_{t+1,2}(z_{t+1}) - \ln p_{t+1,2}(z_{t+1})] \frac{f_{1,t+1}(z_{t+2}|z_{t+1})}{f_{1,t+1}(z_{t+2}|z_{t+1}) + f_{2,t+1}(z_{t+2}|z_{t+1})} \Bigg) h(z_{t+2}) dz_{t+2} \\
& \times [f_{2t}(z_{t+1}|z_t) - f_{1t}(z_{t+1}|z_t)] dz_{t+1}. \qquad (3.7)
\end{aligned}$$

Also, notice that, at least in this example, although the weights depend on the choice of c, the difference in the conditional value functions does not. Whether this invariance holds for the general case, which we now present, is an open question. In what follows, we suppress the dependence of $c^0_{tj}(z_t, r_{tj})$ on z_t, r_{tj} and reintroduce them when clarity is required.

Define $f_{jt}(z_{t+1}|z_t) = \int f_{jt}(z_{t+1}|z_t, c^0_{tj}, r_{tj}) g_r(r_{tj}) dr_{tj}$. For any initial choice (j, c^0_{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

$$\kappa_{\tau j}(z_{\tau+1}, |z_t, c^0_{tj}, r_{tj}) = \begin{cases} f_{jt}(z_{t+1}|z_t, c^0_{tj}, r_{tj}) \quad \text{for } \tau = t \\ \int \sum_{k=1}^J a_{\tau+1,kj} f_{k\tau}(z_{\tau+1}|z_\tau) \kappa_{\tau-1,j}(z_\tau|z_t, c^0_{tj}, r_{tj}) dz_\tau \\ \quad \text{for } \tau = t+1, \cdots, t+\rho, \end{cases} \qquad (3.8)$$

where $\int \kappa_{\tau j}(z_{\tau+1}, |z_t, c^0_{tj}, r_{tj}) dz_{\tau+1} = 1$, because $\sum_{k=1}^J a_{\tau+1,kj} = 1$. This restriction does not require $a_j \geq 0$. By forward substitution, equations (2.9) and (3.8) obtain

$$\begin{aligned}
& v_{tj}(z_t, c^0_{tj}, r_{tj}) = u_{tj}(z_t, c^0_{tj}, r_{tj}) \\
& + \sum_{\tau=t+1}^{t+\rho} \sum_{k=1}^{J} \int \beta^{\tau-t} [u_{\tau k}(z_\tau, c^0_{\tau k}, r_{\tau k}) + \psi_k[p_\tau(z_\tau, r_\tau)]] \\
& \times a_{\tau kj} g_r(r_\tau) \kappa_{\tau-1,j}(z_\tau|z_t, c^0_{tj}, r_{tj}) dr_\tau dz_\tau \\
& + \beta^{t+\rho+1} \int V_{t+\rho+1}(z_{t+\rho+1}, r_{t+\rho+1}) \\
& \times g_r(r_{t+\rho+1}) \kappa_{t+\rho+1,j}(z_{t+\rho+1}|z_t, c^0_{tj}, r_{tj}) dr_{t+\rho+1} dz_{t+\rho+1}. \qquad (3.9)
\end{aligned}$$

Using equation (3.9), the difference in the conditional value functions associated with two alternative initial choices, j and j' becomes

$$
\begin{aligned}
& v_{tj}(z_t, c_{tj}^0, r_{tj}) - v_{tj'}(z_t, c_{tj'}^0, r_{tj'}) = u_{tj}(z_t, c_{tj}^0, r_{tj}) - u_{tj'}(z_t, c_{tj'}^0, r_{tj'}) \\
& + \sum_{\tau=t+1}^{t+\rho} \sum_{k=1}^{J} \int \beta^{\tau-t} [u_{\tau k}(z_\tau, c_{\tau k}^0, r_{\tau k}) + \psi_k[p_\tau(z_\tau, r_\tau)]] g_r(r_\tau) dr_\tau \\
& \times [a_{\tau kj} \kappa_{\tau-1,j}(z_\tau | z_t, c_{tj}^0, r_{tj}) - a_{\tau kj'} \kappa_{\tau-1,j'}(z_\tau | z_t, c_{tj'}^0, r_{tj'})] dz_\tau \\
& + \beta^{t+\rho+1} \int V_{t+\rho+1}(z_{t+\rho+1}, r_{t+\rho+1}) g_r(r_{t+\rho+1}) dr_{t+\rho+1} \\
& \times [\kappa_{t+\rho,j}(z_{t+\rho+1} | z_t, c_{tj}^0, r_{tj}) - \kappa_{t+\rho,j'}(z_{t+\rho+1} | z_t, c_{tj'}^0, r_{tj'})] dz_{t+\rho+1}. \qquad (3.10)
\end{aligned}
$$

Therefore, a pair of initial choices, (j, c_{tj}^0) and $(j', c_{tj'}^0)$, exhibits **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

$$\kappa_{t+\rho,j}(z_{t+\rho+1} | z_t, c_{tj}^0, r_{tj}) = \kappa_{t+\rho,j'}(z_{t+\rho+1} | z_t, c_{tj'}^0, r_{tj'})$$

almost everywhere,

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

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

$$a_{\tau k^* j} \kappa_{\tau-1,j}(z_\tau | z_t, c_{tj}^0, r_{tj}) \neq a_{\tau k^* j'} \kappa_{\tau-1,j'}(z_\tau | z_t, c_{tj'}^0, r_{tj'}).$$

I 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}),

$$\kappa_{t+1,j}(z_{t+\rho+1} | z_t, c_{tj}^0, r_{tj}) = \int \sum_{k=1}^{J} a_{t+1,kj} f_{kt+1}(z_{t+2} | z_{t+1}) f_{jt}(z_{t+1} | z_t, c_{tj}^0, r_{tj}) dz_{t+1},$$

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

$$\kappa_{t+1,j}(z_{t+\rho+1}|z_t, c^0_{tj}, r_{tj}) - \kappa_{t+1,j'}(z_{t+\rho+1}|z_t, c^0_{tj'}, r_{tj'})$$
$$= \int \sum_{k=1}^{J} f_{kt+1}(z_{t+2}|z_{t+1})[a_{t+1,kj} f_{jt}(z_{t+1}|z_t, c^0_{tj}, r_{tj}) - a_{t+1,kj'} f_{j't}(z_{t+1}|z_t, c^0_{tj'}, r_{tj'})]dz_{t+1}. \tag{3.11}$$

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

$$\int \sum_{k=1}^{J} f_{kt+1}(z_{t+2}|z_{t+1})$$
$$\times [a_{t+1,kj} f_{jt}(z_{t+1}|z_t, c^0_{tj}, r_{tj}) - a_{t+1,kj'} f_{j't}(z_{t+1}|z_t, c^0_{tj'}, r_{tj'})]dz_{t+1} = 0, \quad j,k = 1,2,$$
$$\sum_{k=1}^{J} a_{t+1,kj} = 1, \quad j = 1\cdots, J, \text{ and,}$$
$$a_{t+1,k^*j} f_{jt}(z_{t+1}|z_t, c^0_{tj}, r_{tj}) \neq a_{t+1,k^*j'} f_{j't}(z_{t+1}|z_t, c^0_{tj'}, r_{tj'}) \quad \text{for at least one } k^* \in \{1, \cdots, J\}.$$

Given the volume of alternative choices of weights that obtains one-period finite dependence, I proceed by assuming $\rho = 1$, so equation (3.9) reduces to

$$v_{tj}(z_t, c^0_{tj}, r_{tj}) - v_{tj'}(z_t, c^0_{tj'}, r_{tj'}) = u_{tj}(z_t, c^0_{tj}, r_{tj}) - u_{tj'}(z_t, c^0_{tj'}, r_{tj'})$$
$$+ \beta \int \left(\int \sum_{k=1}^{J} [u_{t+1,k}(z_{t+1}, c^0_{t+1,k}, r_{t+1,k}) + \psi_k[p_{t+1}(z_{t+1}, r_{t+1})]] g_r(r_{t+1}) dr_{t+1} \right)$$
$$\times [a_{t+1kj} f_{tj}(z_{t+1}|z_t, c^0_{tj}, r_{tj}) - a_{t+1kj'} f_{tj'}(z_{t+1}|z_t, c^0_{tj'}, r_{tj'})]dz_{t+1}. \tag{3.12}$$

An example of the weights, $(a_{t+1,j'}, \cdots, a_{t+\rho,j'})$, that satisfy the above three conditions are found as follows. Set $j' = 1$, and $a_{t+1,11} = c$. For $j = 2, \cdots, J$, set $\{a_{t+1,kj} = 0,\ k =$

$1 \cdots J,\ k \neq j\}$. Solving for $a_{t+1,jk},\ k = 1 \cdots J$ and substituting into equation (3.12) obtains

$$
\begin{aligned}
& v_{tj}(z_t, c^0_{tj}, r_{tj}) - v_{t1}(z_t, c^0_{t1}, r_{t1}) = u_{tj}(z_t, c^0_{tj}, r_{tj}) - u_{t1}(z_t, c^0_{t1}, r_{t1}) \\
& + \beta \int \left\{ \int \left([u_{t+1,1}(z_{t+1}, c^0_{t+1,j}, r_{t+1,1}) + \psi_1[p_{t+1}(z_{t+1}, r_{t+1})]] \right. \right. \\
& \times \frac{f_{tj}(z_{t+1}|z_t, c^0_{tj}, r_{tj})}{f_{t1}(z_{t+1}|z_t, c^0_{t1}, r_{t1}) + f_{tj}(z_{t+1}|z_t, c^0_{tj}, r_{tj})} \\
& + [u_{t+1,j}(z_{t+1}, c^0_{t+1,j}, r_{t+1,j}) + \psi_j[p_{t+1}(z_{t+1}, r_{t+1})]] \\
& \left. \left. \times \ \frac{f_{t1}(z_{t+1}|z_t, c^0_{t1}, r_{t1})}{f_{t1}(z_{t+1}|z_t, c^0_{t1}, r_{t1}) + f_{tj}(z_{t+1}|z_t, c^0_{tj}, r_{tj})} \right) g_r(r_{t+1}) dr_{t+1} \right\} \\
& \times [f_{tj}(z_{t+1}|z_t, c^0_{tj}, r_{tj}) - f_{t1}(z_{t+1}|z_t, c^0_{t1}, r_{t1})] dz_{t+1}, \qquad j = 2, \cdots, J.
\end{aligned}
$$

3.1 Optimal continuous choice

The alternative representation of the difference in conditional value functions provides a simple and convenient representation of the condition for optimal conditional continuous choice, c^0_{tj}, given that alternative j is chosen. The key is to note $\partial v_{tj'}(z_t, c^0_{tj'}, r_{tj'})/\partial c_{tl_j} = 0$ for $j' \neq j$ and $l_j = 1, \cdots, L_j$. This equality and equation (3.12) imply $c^0_{tj}(z_t, r_{tj})$ solves

$$
\begin{aligned}
0 = & \frac{\partial}{\partial c_{tl_j}} u_{tj}(z_t, c^0_{tj}, r_{tj}) \\
& + \beta \int \left(\int \sum_{k=1}^{J} \frac{\partial}{\partial c_{tl_j}} [u_{t+1,k}(z_{t+1}, c^0_{t+1,k}, r_{t+1,k}) + \psi_k[p_{t+1}(z_{t+1}, r_{t+1})]] g_r(r_{t+1}) dr_{t+1} \right) \\
& \times [a_{t+1kj} f_{tj}(z_{t+1}|z_t, c^0_{tj}, r_{tj}) - a_{t+1kj'} f_{tl}(z_{t+1}|z_t, c^0_{tj'}, r_{tj'})] dz_{t+1} \\
& + \beta \int \left(\int \sum_{k=1}^{J} [u_{t+1,k}(z_{t+1}, c^0_{t+1,k}, r_{t+1,k}) + \psi_k[p_{t+1}(z_{t+1}, r_{t+1})]] g_r(r_{t+1}) dr_{t+1} \right) \\
& \times \frac{\partial}{\partial c_{tl_j}} [a_{t+1kj} f_{tj}(z_{t+1}|z_t, c^0_{tj}, r_{tj}) - a_{t+1kj'} f_{tl}(z_{t+1}|z_t, c^0_{tj'}, r_{tj'})] dz_{t+1} \qquad (3.13)
\end{aligned}
$$

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 (3.7) implies the optimality condition for $c_{t2}^0(z_t, r_{t2})$ is

$$
\begin{aligned}
0 = &\frac{\partial}{\partial c_{t2}} u_{t2}(z_t, c_{t2}^0, r_{t2}) \\
&+ \beta \int\int [u_{t+1,2}(z_{t+1}, c_{t2}^0, r_{t+1,2}) - \ln p_{t+1,2}(z_{t+1}, r_{t+1})] \\
&\times \left(\frac{f_{1,t+1}(z_{t+2}|z_{t+1}, c_{t+1,1}^0, r_{t+1,1})}{f_{1,t+1}(z_{t+2}|z_{t+1}, c_{t+1,1}^0, r_{t+1,1}) + f_{2,t+1}(z_{t+2}|z_{t+1}, c_{t+1,2}^0, r_{t+1,2})} \right) g_r(r_{t+1}) dr_{t+1} \\
&\times \frac{\partial}{\partial c_{t2}} f_{t2}(z_{t+1}|z_t, c_{t2}^0, r_{tj}) dz_{t+1}. \qquad (3.14)
\end{aligned}
$$

4 Correlated unobserved heterogeneity

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

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

Define $s = [s_1, \cdots, s_Q]$ and assume that s_t is discretely distributed with Q support points, $s_t \in \{s_1, \cdots, s_Q\}$. Let $\pi_{tq'q}(w_t)$ be the probability of being in state q in period t and q' in period $t+1$, conditional on w_t, and let $\pi_{tq'|q}(w_t)$ be the probability of being in state q' in period $t+1$ given being is state q in period t and w_t. Then for $q = 1, \cdots, Q$, equation (3.12) obtains

$$
\begin{aligned}
&v_{tj}(x_t, s_q, c_{tj}^0, r_{tj}) - v_{tj'}(x_t, s_q, c_{tj'}^0, r_{tj'}) = u_{tj}(x_t, s_q, c_{tj}^0, r_{tj}) - u_{tj'}(x_t, s_q, c_{tj'}^0, r_{tj'}) \\
&+ \beta \sum_{q'=1}^{Q} \int \left(\int \sum_{k=1}^{J} [u_{t+1,k}(x_{t+1}, s_{q'}, c_{t+1,k}^0, r_{t+1,k}) + \psi_k[p_{t+1}(x_{t+1}, s_{q'}, r_{t+1})]] g_r(r_{t+1}) dr_{t+1} \right) \\
&\times [a_{t+1kj} f_{tj}(x_{t+1}|x_t, s_q, c_{tj}^0, r_{tj}) - a_{t+1kj'} f_{tj'}(x_{t+1}|x_t, s_q, c_{tj'}^0, r_{tj'})] dx_{t+1} \pi_{tq'|q}(w_t). \qquad (4.1)
\end{aligned}
$$

Observe that the conditional value function also depends on $(s, \pi_{t\cdot|q})$, where $\pi_{t\cdot|q} = (\pi_{t1|1}, \cdots, \pi_{tQ|q})$. We will suppress notation of the dependence until necessity dictates its presence. 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

$$p_{tj}^0(x_t, s_q, r_t) = E[d_{tj}^0(z_t, e_t)|x_t, s_q, r_t] = \int d_{tj}^0(z_t, r_t, \varepsilon_t) g_\varepsilon(\varepsilon_t) d\varepsilon_t. \tag{4.2}$$

The probability of choosing alternative j at time t and the corresponding optimal continuous choice, conditional on s_q and x_t, are

$$p_{tj}^0(x_t, s_q) = \int p_{tj}^0(x_t, s_q, r_t) g_r(r_t) dr_t. \tag{4.3}$$

The probability of choosing alternative j at time t and the corresponding optimal continuous choice,, conditional on x_t, are

$$p_{tj}^0(x_t) = \sum_{q=1}^{Q} p_{tj}^0(x_t, s_q) \pi_{tq}(w_t), \tag{4.4}$$

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

5 Identification

In this section, we discuss sufficient conditions for identification of the parameters of the model. Define

$$\begin{aligned}
s &= (s_1, \cdots, s_Q), \\
\pi_t(w_t) &= (\{\pi_{tq',q}(w_t),\ q', q = 1, \cdots, Q\})', \\
\pi_t &= \{\pi_t(w_t)', w_t \in \mathcal{X}_t\}, \\
\pi &= \{\pi_t,\ t = 2, \cdots, T-1\} \\
\Pi_{t\cdot|q} &= (s, \pi_{t\cdot|q}), \\
\Pi_t &= (s, \pi_t), \quad \text{and} \\
\Pi &= (s, \pi).
\end{aligned}$$

Identification is semiparametric in the sense that we impose parametric restrictions on u_{tj}, g_ε, g_r, and f_{jt}, but we only impose exclusion restrictions on w_t with respect to x_t for the purpose of identifying the conditional distribution of the unobserved heterogeneity. The parametric restrictions are as follows: $u_{tj}(z_t, s_q, c_{tj}, r_{tj}) = u_{tj}(z_t, s_q, c_{tj}, r_{tj}; B_1,)$ is known up to $B_1 \in \Re^{D_{B_1}}$; $g_r(r_t) = g_r(r_t; B_2)$ is known up to $B_2 \in \Re^{D_{B_2}}$; and $f_{tj}(x_{t+1}|x_t, s_t, c_{tj}) = f_{tj}(x_{t+1}|x_t, s_t, c_{tj}; B_3)$ is known up to $B_3 \in \Re^{D_{B_3}}$. Define $B = (B_1, B_2, B_3) \in \mathcal{B} \subseteq \Re^{D_{B_1}+D_{B_2}+D_{B_3}}$.

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

$$u_{tjk}(x_t, s_q, c_t, r_t; B_1) = u_{tj}(x_t, s_q, c_{tj}, r_{tj}; B_1) - u_{tk}(x_t, s_q, c_{tk}, r_{tk}; B_1), \quad \text{and}$$
$$v_{tjk}(x_t, s_q, c_t, r_t; B, \Pi_{t,\cdot|q}) = v_{tj}(x_t, s_q, c_{tj}, r_{tj}; B, \Pi_{t,\cdot|q}) - v_{tk}(x_t, s_q, c_{tk}, r_{tk}; B, \Pi_{t,\cdot|q}). \quad (5.1)$$

Notice that I make explicit the dependence of the alternative-specific conditional value function on $\Pi_{t,\cdot|q}$, which enters through the continuation value function.

Assumption 5.1. *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 and* g_ε *is twice continuously differentiable and log-concave with support* $\Re$.
4. For each $j \in \{1, \cdots, J\}$, *at least one* $k \in \{1, \cdots, J\} \setminus \{j\}$ *and at least one* $t \in \{2, \cdots, T-1\}$ *exist such that*

$$Rank\left\{\left(E\left[\frac{\partial}{\partial(B, s_q)} v_{tjk}(x_t, s_q, c_t, r_t; B, \Pi_{t,\cdot|q})\right]\right)\left(E\left[\frac{\partial}{\partial(B, s_q)} v_{tjk}(x_t, s_q, c_t, r_t; B, \Pi_{t,\cdot|q})\right]\right)'\right\}$$
$$= \Re^{D_B + D_s}.$$

5. $u_{tj}(x_t, s_q, c_{tj}, r_{tj}; B_1)$ *is strictly increasing, strictly concave, and twice continuously differentiable in* c_{tl_j}, $l_j = 1, \cdots, L_j$, *and* $f_{tj}(x_{t+1}|x_t, s_q, c_{tj}; B_3)$ *is twice continuously differentiable in* $c_{tl_j}, l_j = 1, \cdots, L_j$.
6. For any $w_t \in \mathcal{X}_t$ *for some* $j^* \in \{1, \cdots, J\}$ *and all* $k \in \{1, \cdots, J\} \setminus \{j^*\}$, *a non-empty set of* x_t, $\tilde{\mathcal{X}}_t(w_t)$ *exists for which the following hold:*

i. Let $\tilde{s}^q$ be s with s_q replaced by $\tilde{s}_q$. For any $(B, \pi_{t,\cdot|q})$, either
$v_{tj^*k}(x_t, \tilde{s}_q, c_t, r_t; B, \tilde{s}^q, \pi_{t,\cdot|q}) > v_{tj^*k}(x_t, s_q, c_t, r_t; B, \Pi_{t,\cdot|q})$ *or*
$v_{tj^*k}(x_t, \tilde{s}_q, c_t, r_t; B, \tilde{s}^q, \pi_{t,\cdot|q}) < v_{tj^*k}(x_t, s_q, c_t, r_t; B, \Pi_{t,\cdot|q})$ *whenever* $\tilde{s}_q > s_q$.

ii. For $(\tilde{B}, \tilde{\Pi}_{t,\cdot|q}) \neq (\bar{B}, \bar{\Pi}_{t,\cdot|q})$, for any $(\tilde{c}_t, \bar{c}_t)$, and any r_t, $(\tilde{x}, \bar{x}) \in \tilde{\mathcal{X}}_t^2(w_t)$ exists for which
$v_{tj^*k}(\tilde{x}_t, \tilde{s}_q, \tilde{c}_t, r_t; \tilde{B}, \tilde{\Pi}_{t,\cdot|q}) < v_{tj^*k}(\tilde{x}_t, \bar{s}_q, \bar{c}_t, r_t; \bar{B}, \bar{\Pi}_{t,\cdot|q})$, *and*
$v_{tj^*k}(\bar{x}_t, \tilde{s}_q, \tilde{c}_t, r_t; \tilde{B}, \tilde{\Pi}_{t,\cdot|q}) > v_{tj^*k}(\bar{x}_t, \bar{s}_q, \bar{c}_t, r_t; \bar{B}, \bar{\Pi}_{t,\cdot|q})$.

Define $\mathbf{P}(x; B, \Pi) = (p_{tj}(x_t; B, \Pi), j = 1, \cdots, J, t = 1, \cdots, T,)$, where I make explicit the dependence of $p_{tj}(x_t)$ defined in equation (4.3) on (B, Π_t). Let (B_0, Π_0) be the true parameter vector; that is, the probabilities generated from the model at (B_0, Π_0) coincides with the population probabilities: $\mathbf{P}(x; B_0, \Pi_0) = \mathbf{P}^0(x)$.

Theorem 5.2. *Suppose assumption 5.1 holds. Then (B_0, Π_0) is identified in the sense that any $(\tilde{B}, \tilde{\Pi})$ satisfying $\mathbf{P}(x; \tilde{B}, \tilde{\Pi}) = \mathbf{P}^0(x)$ implies $(\tilde{B}, \tilde{\Pi}) = (B_0, \Pi_0)$.*

The proof of theorem 5.2 is found in Gayle (2013).

6 Estimator

In this section, I propose a GMM estimator for the parameters of the model, B and Π. I choose to propose a GMM estimator instead of the ML estimator for two reasons. First, the definition of the GMM estimator does not require specifying the distribution of measurement errors, which is of particular concern in discrete and continuous choice models because observed continuous choice variables are often measured with errors. Second, the GMM estimator is robust to the initial conditions problem: consistent estimation of the parameters does not require observing the initialization of (x_t, s) or for it to be specified. My objective function is constructed in a way that results in a computationally convenient algorithm to compute the estimator of the parameters with the restriction that the type probabilities satisfy the properties of probability and are not constrained to be strictly positive. I begin by imposing restrictions on w.

Assumption 6.1. *1. The random vector w_t belongs to a discrete set $\mathcal{W}_t \subset \mathcal{X}_t$ with R distinct values.*
2. The maximum possible number of types Q is known to the investigator.

Restricting w_t to a discrete set affords the proposal of a relatively simple implementation in estimation and derivation of the asymptotic distribution of the estimator. Under Assumption 6.1.1, Π_t consists of $Q((Q-1)R+1)$ parameters to be estimated. Assumption 6.1.2 says the investigator has prior knowledge of the maximum possible number of types. The proposed estimator requires this assumption for it to be operational. One can set the maximum possible number of types to be large, at the cost of increased computational burden.

From here on, the unobserved state vector s and the deep parameters B are treated symmetrically, and grouping them is convenient for what is to come. To that end, define $\theta = (B,s) \in \Theta = \mathcal{B} \times \Re^Q$. The estimator for the parameters of the model is best framed as the solution to a constrained optimization problem, where the primary parameter of interest is θ, and the constraint imposes restrictions on π. I first define the objective function for θ, and then define the constraint set.

Suppose n observations of the random vectors $y_i = \{(d_{it}, c_{it}), t = 1, \cdots, T\}$ and $x_i = \{x'_{it}, t = 1, \cdots, T\}$ are accessible to the investigator. For each i, and for $t = 2, \cdots, T-1$, define the residuals

$$
\begin{aligned}
\rho_{1itj}(y_{it}, x_{it}; \theta, \pi_t) &= d_{itj} - p^0_{tj}(x_{it}; \theta, \pi_t), \\
\rho_{2itj}(y_{it}, x_{it}; \theta, \pi_t) &= d_{itj}(c_{itj} - c^0_{tj}(x_{it}; \theta, \pi_t)),
\end{aligned}
$$

where

$$
\begin{aligned}
p^0_{tj}(x_{it}; \theta, \pi_t) &= \sum_{w_t \in \mathcal{W}_t} \sum_{q=1}^{Q} p^0_{tj}(x_{it}; \theta, \pi_{t\cdot|q}(w_t))\ \pi_{tq}(w_t) 1\{w_{it} = w_t\}, \quad \text{and} \\
c^0_{tj}(x_{it}; \theta, \pi_t)) &= \sum_{w_t \in \mathcal{W}_t} \sum_{q=1}^{Q} c^0_{tj}(x_{it}; \theta, \pi_{t\cdot|q}(w_t))\ \pi_{tq}(w_t) 1\{w_{it} = w_t\}.
\end{aligned}
$$

Define

$$
\begin{aligned}
\rho_{1it}(y_{it}, x_{it}; \theta, \pi_t) &= \{\rho_{1itj}(y_{it}, x_{it}; \theta, \pi_t), j = 2, \cdots, J\}, \quad \text{and} \\
\rho_{2it}(y_{it}, x_{it}; \theta, \pi_t) &= \{\rho_{2itj}(y_{it}, x_{it}; \theta, \pi_t), j = 1, \cdots, J\}.
\end{aligned}
$$

Define the $L+J-1$-dimensional residual vector

$$\rho_{it}(y_{it},x_{it};\theta,\pi_t)=(\rho_{1it}(y_{it},x_{it};\theta,\pi_t),\rho_{2it}(y_{it},x_{it};\theta,\pi_t))'.$$

Let X_{it} be a $(L+J-1)\times N^{X_t}$ matrix of instruments, and define the N^{X_t}-dimensional vector.

$$m_{it}(\theta;\pi_t)=X_{it}'\rho_{it}(y_{it},x_{it};\theta,\pi_t). \tag{6.1}$$

Define the N^X-dimensional vector $m_i(\theta;\pi)=(m_{i2}(\theta;\pi_2)',\cdots,m_{iT-1}(\theta;\pi_{T-1})')'$, where $N^X=\sum_{t=2}^{T-1}N^{X_t}$, 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. The population objective function is defined as follows.

$$S^\theta(\theta;\pi)=m(\theta;\pi)'\Omega m(\theta;\pi). \tag{6.2}$$

To implement the estimator, let

$$\hat{m}(\theta;\pi)=\frac{1}{n}\sum_{i=1}^{n}m_i(\theta;\pi). \tag{6.3}$$

Then the sample objective function is defined as

$$\hat{S}^\theta(\theta;\pi)=\hat{m}(\theta;\pi)'\hat{\Omega}\hat{m}(\theta;\pi), \tag{6.4}$$

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

To define the constraint set, for each $i=1,\cdots,n$ and $t=2,\cdots,T-1$, let

$$p^0_{itt+1,jk}(x_{it+1},x_{it},s_{q'},s_q;\theta,\pi)=p^0_{it+1,j}(x_{it+1},s_{q'};\theta,\pi_{t+1,\cdot|q'})p^0_{it,k}(x_{it},s_q;\theta,\pi_{t,\cdot|q}), \tag{6.5}$$

and

$$p^0_{itt+1,jk}(x_{it+1},x_{it};\theta,\pi)=\sum_{q'=1}^{Q}\sum_{q=1}^{Q}p^0_{itt+1,jk}(x_{it+1},x_{it},s_{q'},s_q;\theta,\pi)\pi_{tq'q}(w_{it}). \tag{6.6}$$

By defining

$$\mathbf{p}_{tt+1,jk}(x_{it+1},x_{it};\theta,\pi) = (\{p^0_{tt+1,jk}(x_{it+1},x_{it},s_{q'},s_q;\theta,\pi),\ q'=1,\cdots,Q,\ q=1,\cdots,Q\}),$$

I can write

$$p^0_{tt+1,jk}(x_{it+1},x_{it};\theta,\pi) = \mathbf{p}_{tt+1,jk}(x_{it+1},x_{it};\theta,\pi)\pi_t(w_{it}). \tag{6.7}$$

For any $\theta \in \Theta$ and any π, define $\pi_{t0}(w_{it};\theta,\pi)$ as satisfying

$$E[d_{it+1,j}d_{itk}|x_{it+1},x_{it}] = \mathbf{p}_{tt+1,jk}(x_{it+1},x_{it};\theta,\pi)\pi_{t0}(w_{it};\theta,\pi). \tag{6.8}$$

The identification conditions in Assumption 5.1 imply $\bar{\pi}_{t0}(w_{it};\theta_0,\pi_0) = \pi_{t0}(w_{it})$. By defining

$$\nu^p_{itt+1,jk}(\theta,\pi) = d_{it+1,j}d_{itk} - p^0_{tt+1,jk}(x_{it+1},x_{it};\theta,\pi),$$

equation (6.8) can be written as

$$d_{it+1,j}d_{itk} = \mathbf{p}_{tt+1,jk}(x_{it+1},x_{it};\theta,\pi)\pi_{t0}(w_{it};\theta,\pi) + \nu^p_{itt+1,jk}(\theta,\pi), \tag{6.9}$$

where by the law of iterated expectations, $E[\nu^p_{it,t+1,jk}(\theta,\pi)|\mathbf{p}_{tt+1,jk}(x_{it+1},x_{it};\theta,\pi)] = 0$. Similarly, define

$$\begin{aligned}
c_{itt+1,jk} &= d_{it+1,k}c_{it+1,k} \otimes d_{itj}c_{itj}, \\
c^*_{itt+1,jk}(x_{it+1},x_{it},s_{q'},s_q;\theta,\pi) &= d_{it+1,k}c^0_{it+1,k}(x_{it+1},s_{q'};\theta,\pi_{t+1,\cdot|q'}) \otimes d_{itj}c^0_{itj}(x_{it},s_q;\theta,\pi_{t,\cdot|q}), \\
c^*_{itt+1,jk}(x_{it+1},x_{it};\theta,\pi) &= \sum_{q'=1}^{Q}\sum_{q=1}^{Q} c^*_{itt+1,jk}(x_{it+1},x_{it},s_{q'},s_q;\theta,\pi)\pi_{tq'q}(w_{it}) \\
&= \mathbf{c}^*_{itt+1,jk}(x_{it+1},x_{it};\theta,\pi)\pi_t(w_{it}),
\end{aligned} \tag{6.10}$$

where $\mathbf{c}^*_{tt+1,jk}(x_{it+1},x_{it};\theta,\pi) = (\{c^*_{tt+1,jk}(x_{it+1},x_{it},s_{q'},s_q;\theta,\pi)',\ q'=1,\cdots,Q,\ q=1,\cdots,Q\})$, and $\otimes$ denotes the Kronecker product. For any $\theta \in \Theta$ and π, $\pi_{t0}(w_{it};\theta,\pi)$ satisfies

$$E[c_{itt+1,jk}|x_{it+1},x_{it}] = \mathbf{c}^*_{itt+1,jk}(x_{it+1},x_{it};\theta,\pi)\pi_{t0}(w_{it};\theta,\pi), \tag{6.11}$$

and by letting

$$\nu^c_{itt+1,jk}(\theta,\pi) = c_{itt+1,jk} - c^*_{itt+1,jk}(x_{it+1}, x_{it};\theta,\pi),$$

equation (6.11) implies the following:

$$c_{itt+1,jk} = \mathbf{c}^*_{itt+1,jk}(x_{it+1}, x_{it};\theta,\pi)\pi_{t0}(w_{it};\theta,\pi) + \nu^c_{itt+1,jk}(\theta,\pi), \tag{6.12}$$

where again, under the restrictions of the model and by the law of iterated expectations, $E[\nu^c_{itt+1,jk}(\theta,\pi)|\mathbf{c}^*_{tt+1,jk}(x_{it+1},x_{it};\theta,\pi)] = 0$.

Now, define

$$\begin{aligned}
d_{itt+1} &= (\{d_{it+1,j}d_{itk},\ j=1,\cdots,J,\ k=1,\cdots,J\},\backslash\{d_{it+1,1}d_{it1}\})',\\
\mathbf{p}_{itt+1}(\theta,\pi) &= (\{\mathbf{p}'_{tt+1,jk}(x_{it+1},x_{it};\theta,\pi),\ j=1,\cdots,J,\ k=1,\cdots,J\}\\
&\qquad \backslash\{\mathbf{p}'_{tt+1,11}(x_{it+1},x_{it};\theta,\pi)\})',\\
\nu^p_{itt+1}(\theta,\pi) &= (\{\nu^p_{itt+1,jk}(\theta,\pi),\ j=1,\cdots,J,\ k=1,\cdots,J\}\backslash\{\nu^p_{itt+1,11}(\theta,\pi)\})',\\
c_{itt+1} &= (\{c'_{itt+1,jk}\}\ j=1,\cdots,J,\ k=1,\cdots,J)',\\
\mathbf{c}^*_{itt+1}(\theta,\pi) &= (\{(\mathbf{c}^*_{itt+1,jk}(x_{it+1},x_{it};\theta,\pi))',\ j=1,\cdots,J,\ k=1,\cdots,J\})',\\
\nu^c_{itt+1}(\theta,\pi) &= (\{(\nu^c_{itt+1,jk}(\theta,\pi))',\ j=1,\cdots,J,\ k=1,\cdots,J\})'.
\end{aligned}$$

Let

$$\begin{aligned}
y_{itt+1} &= (d'_{itt+1}, c'_{itt+1}),\\
X_{itt+1}(\theta,\pi) &= (\mathbf{p}'_{itt+1}(\theta,\pi), (\mathbf{c}^*_{itt+1}(\theta,\pi))')', \quad \text{and} \qquad (6.13)\\
\nu_{it}(\theta,\pi) &= ((\nu^p_{itt+1}(\theta,\pi))', (\nu^c_{itt+1}(\theta,\pi))')'. \qquad (6.14)
\end{aligned}$$

Then, stacking equations (6.10) and (6.12) obtains

$$y_{itt+1} = X_{itt+1}(\theta,\pi)\pi_{t0}(w_{it};\theta,\pi) + \nu_{itt+1}(\theta,\pi). \tag{6.15}$$

Define

$$M_t^0(\theta,\pi)=\left\{\xi\in\Re^{Q^2}:\xi=\arg\min_{\tilde{\xi}} S^{\pi}(\tilde{\xi};w_t,\theta,\pi),\xi_q\geq 0,q=1,\cdots,Q^2,\sum_{q=1}^{Q^2}\xi_q=1,w_t\in\mathcal{W}_t\right\},$$

where

$$S^{\pi}(\xi;w_t,\theta,\pi)=E\left[(y_{itt+1}-X_{itt+1}(\theta,\pi)\xi)'(y_{itt+1}-X_{itt+1}(\theta,\pi)\xi)1\{w_{it}=w_t\}\right]. \tag{6.16}$$

Define analogously

$$\hat{M}_t(\theta,\pi)=\left\{\xi\in\Re^{Q^2}:\xi=\arg\min_{\tilde{\xi}} \hat{S}^{\pi}(\tilde{\xi};w_t,\theta,\pi),\xi_q\geq 0,q=1,\cdots,Q^2,\sum_{q=1}^{Q^2}\xi_q=1,w_t\in\mathcal{W}_t\right\},$$

where

$$\hat{S}^{\pi}(\xi;w_t,\theta,\pi)=\frac{1}{n}\sum_{i=1}^{n}(y_{itt+1}-X_{itt+1}(\theta,\pi)\xi)'(y_{itt+1}-X_{itt+1}(\theta,\pi)\xi)1\{w_{it}=w_t\}. \tag{6.17}$$

Define $M^0(\theta,\pi)=\times_{t=2}^{T-1}M_t^0(\theta,\pi)$ and $\hat{M}(\theta,\pi)=\times_{t=2}^{T-1}\hat{M}_t(\theta,\pi)$. Then, under the conditions of Assumption 5.1, (θ_0,π_0) solves

$$\begin{aligned}&\min_{\theta\in\Theta} S^{\theta}(\theta;\pi)\\ &s.t. \qquad \pi\in M^0(\theta,\pi),\end{aligned} \tag{6.18}$$

and the estimator $(\hat{\theta},\hat{\pi})$ for (θ_0,π_0) solves

$$\begin{aligned}&\min_{\theta\in\Theta} \hat{S}^{\theta}(\theta;\pi)\\ &s.t. \qquad \pi\in \hat{M}(\theta,\pi).\end{aligned} \tag{6.19}$$

7 Computing The Estimator

In this section, I present a method for computing the estimator proposed in the previous section. I describe the algorithm at the $o+1$ iteration with $(\pi^{[o]},\mathbf{c}^{*,[o]},\mathbf{p}^{[o]},\theta^{[o]})$ in hand. Each component of the algorithm is specified for a generic individual i at a generic point of time t. Computing the algorithm requires integrating over the alternative-specific conditional continuous choice shocks r. The popular methods for this integration (quadrature and simulation methods) require evaluating the objective function at discrete points on the support of r. To conserve on notation, I will use the subscript $'r'$ to denote the discrete point at which the function is being evaluated, as well as the random variable itself.

7.1 Updating π

The approach to updating π is to perform a two-step iterated projection. The first step projects the weights into $\Re^{Q^2}$, and the second projects the result into the Q^2 - dimensional simplex, thereby obtaining probabilities.

Equation (6.15) implies that, for each $t = 2, \cdots, T-1$ and $w_t \in \mathcal{W}_t$, the first-stage projection obtains

$$\bar{\pi}_t^{[o+1]}(w_t) = \left[\sum_{i=1}^{n} X'_{itt+1}\left(\theta^{[o]}, \pi^{[o]}\right) X_{itt+1}\left(\theta^{[o]}, \pi^{[o]}\right) 1\{w_{it} = w_t\}\right]^{-1} \times \sum_{i=1}^{n} X'_{itt+1}\left(\theta^{[o]}, \pi^{[o]}\right) y_{itt+1} 1\{w_{it} = w_t\}, \tag{7.1}$$

where $X_{itt+1}\left(\theta^{[o]}, \pi^{[o]}\right)$ is defined in equation (6.13). The second step is to project $\bar{\pi}_t^{[o+1]}(w_t)$ into the Q^2- dimensional simplex. To perform the projection, I apply the algorithm developed by Chen and Ye (2011), which operates as follows:

Algorithm 1

1. Sort $\bar{\pi}_t^{[o+1]}(w_t)$ in ascending order as $\bar{\pi}_t^{[o+1]}(w_t)_{(1)} \leq \cdots \leq \bar{\pi}_t^{[o+1]}(w_t)_{(Q^2)}$ and set $l = Q^2 - 1$;
2. Compute $\tau_l = \frac{\sum_{l'=l+1}^{Q^2} \bar{\pi}_t^{[o+1]}(w_t)_{(l')}}{Q^2 - 1}$. If $\tau_l \geq \bar{\pi}_t^{[o+1]}(w_t)_{(l)}$, set $\hat{\tau} = \tau_l$, and go to Step 4; otherwise, set $l \leftarrow l - 1$ and redo Step 2 if $l \geq 1$ or go to Step 3 if $l = 0$;
3. Set $\bar{\tau} = \frac{\sum_{l'=1}^{Q^2} \bar{\pi}_t^{[o+1]}(w_t)_{(l')}}{Q^2}$;

4. Return $\hat{\pi}_t^{[o+1]}(w_t) = (\bar{\pi}_t^{[o+1]}(w_t) - \hat{\tau})_+$ as the projection of $\bar{\pi}_t^{[o+1]}(w_t)$ onto the Q^2-dimensional simplex.

5. Unsort $\hat{\pi}_t^{[o+1]}(w_t)$ and return as the updated type probabilities

Algorithm 1 defines a Q^2-dimensional operator Γ with

$$\hat{\pi}_t(\theta, w_t) = \Gamma(\bar{\pi}_t(\theta, w_t)), \tag{7.2}$$

with $\pi_{t0}(w_t) = \Gamma(\pi_{t0}(w_t))$, because $\pi_{t0}(w_t) \in \Delta^{Q^2}$.

7.2 Updating the alternative-specific continuous choices

For each $j = 1, \cdots, J$, and $q = 1, \cdots, Q$, the alternative-specific continuous choice variables are updated as follows (see section 3.1):

$$\begin{aligned} c_{itjqr}^{0,[o+1]} &= c_{itjqr}^{0,[o]} \\ &- \left[\frac{\partial^2}{\partial c_{itjqr}^2} v_{itj1}(x_{it}, c_{itjqr}^{0,[o]}, s_q, r_t; p_{it+1}^{[o]}, \theta^{[o]}, \pi^{[o+1]}) \right]^{-1} \\ &\times \left[\frac{\partial}{\partial c_{itjqr}} v_{itj1}(x_{it}, c_{itjqr}^{0,[o]}, s_q, r_t; p_{it+1}^{[o]}, \theta^{[o]}, \pi^{[o+1]}) \right]. \end{aligned} \tag{7.3}$$

Note two key observations from equation (7.3): (i) I make explicit that the dependence on the (difference in the) period t conditional value function depends only on the $t+1$ conditional choice probabilities, a result obtained from the generalized finite dependency framework (see equation 4.1), and (ii) Although I assume alternative 1 is the normalizing alternative, any other alternative (other that alternative j) obtains the same result.

7.3 Updating the CCPs

Recall from equation (4.2) that

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

Define $v_t(x_t,s_q,c_t^0,r_t) = (v_{t1}(x_t,s_q,c_{t1}^0,r_{t1}),\cdots,v_{tJ}(x_t,s_q,c_{tJ}^0,r_{Jt}))$. From the definition of $d_{tj}^0(z_t,r_t,\varepsilon_t)$ in equation (2.4), we have $d_{tj}^0(z_t,r_t,\varepsilon_t) = d_{tj}^0(v_t(x_t,s_q,c_t^0,r_t),\varepsilon_t)$. We can therefore write

$$p_{tj}^0(x_t,s_q,r_t) = \int d_{tj}^0(z_t,r_t,\varepsilon_t)g_\varepsilon(\varepsilon_t)d\varepsilon_t = \Psi_j(v_t(x_t,s_q,c_t^0,r_t)),$$

where Ψ_j is a function of the distribution of ε. For example, if ε is distributed i.i.d. type one extreme value, then

$$\Psi_j(v_t(x_t,s_q,c_t^0,r_t)) = \frac{e^{v_{tj}(x_t,s_q,c_{tj}^0,r_{tj})}}{\sum_{k=1}^J e^{v_{tk}(x_t,s_q,c_{tk}^0,r_{tk})}}.$$

Given these definitions, the conditional choice probabilities can be updated as follows:

$$p_{itjqr}^{[o+1]} = \Psi_j(v_{it}(x_{it},s_q,c_{itjqr}^{0,[o+1]},r_t;p_{it+1}^{[o]},\theta^{[o]},\pi^{[o+1]}). \tag{7.4}$$

7.4 Updating θ

Finally, θ is updated as follows:

$$\begin{aligned}
&\theta^{[o+1]} = \theta^{[o]} \\
&- \left[\left(\frac{\partial}{\partial\theta}m(\theta^{[o]};\pi^{[o+1]},\mathbf{c}^{0,[o+1]},\mathbf{p}^{0,[o+1]}))\right)'\hat{\Omega}\left(\frac{\partial}{\partial\theta}m(\theta^{[o]};\pi^{[o+1]},\mathbf{c}^{0,[o+1]},\mathbf{p}^{0,[o+1]})\right)\right]^{-1} \\
&\times\left(\frac{\partial}{\partial\theta}m(\theta^{[o]});\pi^{[o+1]},\mathbf{c}^{0,[o+1]},\mathbf{p}^{0,[o+1]}\right)'\hat{\Omega}\quad m(\theta^{[o]};\pi^{[o+1]},\mathbf{c}^{0,[o+1]},\mathbf{p}^{0,[o+1]}).
\end{aligned} \tag{7.5}$$

The full algorithm is as follows.

Main Algorithm.
1– Initialize $\theta^{[0]} \in \Theta$, $\mathbf{p}^{0,[0]} \in [0,1]^{[n(T-2)(J-1)]}$, and $\mathbf{c}^{0,[0]} \in \Re^{n(T-2)L}$.
2– For $o \geq 1$,
    2.1– Update $\pi^{[o]}$ using equation (7.1) and Algorithm 1
    2.2– Update $c^{0,[o]}$ using equation (7.3)
    2.3– Update $p^{0,[o]}$ using equation (7.4)
    2.4– Update $\theta^{[o]}$ using equation (7.5)

Until convergence in θ.

The convergent value of θ is the estimator $\hat{\theta}$, and the corresponding convergent value of π is our estimator $\hat{\pi}$. Notice that steps 2.1-2.3 of the main algorithm is simply evaluating the objective function at the new trial values of θ, whereas step 2.4 is a Gauss-Newton step. Hence, the main algorithm converges, and the convergence rate is at most quadratic.

8 Asymptotic properties of the estmator

To derive the asymptotic properties of the estimator $(\hat{\theta}, \hat{\pi})$, some regularity conditions are needed. I use the following notations in all assumptions, theorems, and proofs: $\sup_\theta = \sup_{\theta\in\Theta}$, $\sup_\pi = \sup_{\pi\in[0,1]^{Q^2}}$, and $\sup_{\theta,\pi} = \sup_{\theta\in\Theta}\sup_{\pi\in[0,1]^{Q^2}}$. The first assumption imposes the typical random-sampling restriction of the sampling process.

Assumption 8.1. *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.*

The next assumption imposes restrictions on the parameter space, and the admissible functional forms of the utility functions.

Assumption 8.2. *1. The sets $\mathcal{X}$ and Θ are compact; and 2.* $\sup_{\theta,\pi}\left\|E\left[c_t(x_{it}, s_q, r_t;\theta,\pi)\right]\right\| < \infty$ *for $t = 2,\cdots,T$.*

Assumption 8.3. *$\hat{\Omega}$ is symmetric and positive definite with $\|\hat{\Omega} - \Omega\| = o_p(1)$.*

The proof of the following consistency theorem is in Appendix A.1.

Theorem 8.4. *Suppose (i) Assumption 5.1 holds, (ii) Assumption 6.1 holds, and (iii) Assumptions 8.1, 8.2, and 8.3 hold. Then $\hat{\pi} \xrightarrow{p} \pi_0$ and $\hat{\theta} \xrightarrow{p} \theta_0$.*

Define $m_i = m_i(\theta_0;\pi_0)$, $M_{\theta i} = \partial m_i(\theta_0;\pi_0)/\partial\theta$, $M_\theta = E[M_{\theta i}]$, $M_{\pi_t i} = \partial m_i(\theta_0;\pi_0)/\partial\pi_t(w_{it})$, $M_{\pi_t} = E[M_{\pi_t i}]$,

$$a_{it} = X_{itt+1}(\theta_0,\pi_0)'\left(y_{itt+1} - X_{itt+1}(\theta_0,\pi_0)\pi_{t0}(w_{it})\right).$$

Theorem 8.5. *Suppose the conditions of theorem 8.4 hold, and θ_0 is in the interior of Θ. Then,*

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

where $V = (M_\theta' \Omega M_\theta)^{-1}(M_\theta' \Omega \Sigma \Omega M_\theta)(M_\theta' \Omega M_\theta)^{-1}$, *and*

$$\Sigma = E\left[\left\{m_i + \sum_{t=2}^{T-1} M_{\pi_t} a_{it}\right\}\left\{m_i + \sum_{t=2}^{T-1} M_{\pi_t} a_{it}\right\}'\right].$$

The proof of Theorem 8.5 is in Appendix A.2.

In practice, a consistent estimator for the asymptotic variance V is required. One can be obtained via the plug-in approach, where the parameters in V are replaced with their respective estimators, and the expectations are replaced with sample averages. The proof for consistency of this plug-in estimator is standard and can also be found in Newey and McFadden (1994).

9 Education and Labor Market Choices and the Risky Returns to Education

In this section, I implement the method developed in the previous sections to investigate the life-cycle educational and labor market choices of a sample of young men from the 1979 cohort of the National Longitudinal Survey of Youth (NLSY79), and the resulting long-run "ex-ante" returns to education. Key ingredients of the theoretical model are as follows: I allow for the decision to work while enrolled in school, for hours worked to affect the likelihood that an individual will advance a grade level, and for uncertainty in the returns to education, which may be dependent on observed characteristics of the individual.

The inadequacy of the classical Mincer equation to obtain policy-relevant estimates of the returns to education has been well documented over the last two decades (see Heckman et al. (2006) for a review of the relevant literature). Heckman et al. (2008) cite notable extensions to the classical Mincer wage equation that are likely to reduce the biases in estimates of returns to education. These extensions include direct and psychic costs of schooling, nonseparability between experience and schooling, heterogeneity in returns to education, and disentangling marginal and average returns to schooling. Other important factors that may affect estimates of the returns to education include the endogeneity of schooling and work experience choices, uncertainty about the returns to education, and uncertainty about the completed level of education.

9.1 The theoretical model

The structural model specified below incorporates psychic costs of schooling and working. It accounts for direct benefits from working; the income earned from working and the additional years of experience gained. Nonseparability between working and schooling is accounted for by the simultaneity of these choices and dynamic selection. The specification of the log wage equation allows for the returns to education to vary across race, education, and labor market categories. Furthermore, we disentangle ex-ante and ex-post returns to education by allowing the returns to education to be uncertain over the work life, and by allowing for individual-period specific shocks to wages.

The environment that a generic individual faces is modeled as follows. In each period, t, the individual is endowed with a fixed amount of time that is normalized to 1. He is presented with four mutually exclusive and exhaustive alternatives, j: to stay home ($j=1$); to attend school and not work ($j=2$); to work and not attend school ($j=3$); and to both work and attend school ($j=4$). Let d_{tj} be equal to 1 if the individual chooses alternative j in period t, and zero otherwise. If the individual chooses to work in period t, he must decide how to allocate his time endowment between leisure, l_t, and labor supply, h_t, so that $l_t + h_t = 1$. If the individual chooses not to work in period t, then $l_t = 1$ ($h_t = 0$). Define d_t^h to be equal to 1 if the individual chooses to work in period t, and zero otherwise.

If the individual decides to work in period t, he gains an additional year of experience. If he decides to enroll in school, he advances the grade level with probability $F_t(h_t, x_t^a; \theta^a)$, where $x_t^a = (h_t, d_t^h, BLACK, AGE_t * EDU_t, AFQT)$, where $BLACK$ is equal to 1 if the individual is black, and zero otherwise, AGE_t is his age at period t, EDU_t is his years of education at period t, and $AFQT$ is his Armed Forces qualification test score. This specification of the grade-level advancement probability allows for labor market activities to affect the chances the individual advances the grade level at both the intensive and extensive margins, and it is a key (though not only) source of uncertainty about completed level of education. The individual considers the benefits from working while in school, which include income generated and the level of labor market experience earned, to the costs, which include loss of leisure time and the potentially negative impact of working while in school on the probability of advancing the grade level.

In each period, the individual receives a wage offer w_t, which is parameterized as follows:

$$\ln(w_t) = \theta_1^w + \theta_2^w EXPER_t + \theta_3^w EXPER_t^2 + \theta_4^w AFQT + \theta_6^w BLACK + s_t EDU_t + r_t^w + \nu_t^w, \tag{9.1}$$

where $EXPER_t$ is his years of experience as at period t, s_t is period-specific returns to his level of education in period t, r_t^w is the period-specific shock to his log wage offer, and ν_t^w represents measurement errors in log wages. The individual observes s_t and r_t^w at the beginning of period t, but do not observe their future realizations. We assume returns to education, s_t, are discretely distributed with at most Q support points, $s_t \in \{s_1, \cdots, s_Q\}$. Therefore, in this empirical analysis, $\pi_{q'q}(x_t^s)$ is the probability that the individual's returns to education are s_q in period t and $s_{q'}$ in period $t+1$ given characteristics x_t^s. To account for the extensions to the classical Mincer equation noted by Heckman et al. (2008), I select x_t^s to represent four categories of education (less than high school graduate, high school graduate, some college, and college graduate) and two categories of labor market experience (less than 10 years and at least 10 years). Furthermore, I allow the type probabilities and θ^w to differ by race. I assume r_t^w is independently distributed normal over time with zero mean and variance σ^w.

I assume the period specific utility is strictly increasing and strictly concave in consumption and leisure. I allow for psychic costs of schooling and working. Specifically, the contemporaneous utility function is given by

$$u_t(z_t) = \exp(\theta_1^u x_t^c)(w_t h_t - w_t h_t \ln(h_t)) + \exp(\theta_2^u x_t^l)(l_t - l_t \ln(l_t)) + \theta_3^u d_t^E x_t^E + \theta_4^u d_t^h x_t^h + d_t \varepsilon_t, \tag{9.2}$$

where x_t^c, x_t^l, x_t^E, and x_t^h are observed demographic characteristics that capture individual variation in consumption, leisure, the psychic costs of schooling, and the psychic costs of working, respectively. d_t^E is equal to 1 if the individual enrolls in school at period t, and zero otherwise, $d_t = (\{d_{tj}\},\ j = 1, \cdots, 4)$, where d_{tj} is equal to 1 if the individual chooses alternative j in period t, and $\varepsilon_t = (\{\varepsilon_{tj}\}, j = 1, \cdots, 4)'$, where ε_{tj} is the alternative j specific shock to utility. To see that the period-specific utility is strictly increasing and strictly concave in consumption, $c_t = w_t h_t$, rewrite the utility of consumption as $w_t h_t - w_t h_t \ln(h_t) = w_t h_t (1 + \ln(w_t)) - w_t h_t \ln(w_t h_t)$, and recall that $h_t \in [0,\ 1]$. Another attractive feature of this specification of the utility function is that it ensures that for any choice of the model parameters, the optimal hours that the individual chooses to work lies in the interior of the unit interval. This property of the period-specific utility function is expected to improve the numerical stability of the algorithm.

9.2 Data

The data are taken from the 1979 youth cohort of the National Longitudinal Survey of Labor Market Experience (NLSY79), a comprehensive panel data set that follows individuals over the period 1979 to 2000, who were 14 to 21 years of age as of January 1, 1979. The data set initially consisted of 12,686 individuals: a representative sample of 6,111 individuals, a supplemental sample of 5,295 Hispanics, non-Hispanic blacks, and economically disadvantaged, non-black, non-Hispanics, and a supplemental sample of 1,280 military youth. Interviews were conducted on an annual basis though 1994, after which they adopted a biennial interview schedule. This study makes use of the first 27 years of interviews, from 1979 to 2006. The data are restricted to include non-Hispanic males and respondents with missing observations on the highest grade level completed that cannot be recovered with high confidence from other data information. For further details on the sample restrictions, see Gayle (2006).

9.3 Estimation

The estimator implemented is the one developed in the previous sections, with a few caveats. First, as is typically done in estimation of dynamic structural models, I estimate the grade-transition probability outside of the model. Second, I preset the support of the distribution returns to education. I choose five equally spaced grid points from zero to two, $\{0, 0.5, 1.0, 1.5, 2.0\}$. One can reasonably assume the returns to education are non-negative and not greater than 2. This restriction also assumes that each observable category, x_t^s, has at most five types of individuals in the population with respect to their returns to education. I say at most five types because, unlike the EM likelihood method of Arcidiacono and Miller (2011), the method developed to estimate the conditional type probabilities does not restrict the type transition matrix to be saturated.

9.4 Results

9.4.1 Grade-promotion probability

Table 1 presents the results from estimation of the grade-transition probability under the assumption of a logit probability of advancing to the next grade level given enrollment. The results indicate that given employment, additional hours worked reduces the probability of advancing to the next grade level. However, a positive and statistically significant extensive margin of employment also exists, which is captured by the coefficient on d_t^h. These two results capture the crowding-out and congruence hypothesis. The former hypothesis states that working while enrolled in school crowds out time spent on school activities, thus reducing the chances of completing the grade level. The latter hypothesis states that working moderate hours while enrolled in school improves organization skills, which results in more efficient school activities and higher chances of completing the grade level. The results imply that, ceteris paribus, working more than 16 hours per week during the school year has a negative total impact on the probability of completing the grade level. In the data, approximately 23% of black males work more than 16 hours per week while enrolled in school, whereas 77% of white males work for more 16 hours per week while enrolled in school. However, for any given level of hours worked, the probability of completing the grade level is higher for whites than for blacks because of the higher level of the AFQT score, which has a positive and significant coefficient. This result suggests that white males take advantage of their higher abilities or better preparedness, as measured by the AFQT score, by working more hours while enrolled in school, knowing that the net effect of their labor market activities on the chances of them advancing the grade level is mitigated by their higher average abilities/preparedness.

9.4.2 Period-specific utility

Table 2 presents the estimates of the parameters governing the period-specific utility. The key quantity of interest is the psychic costs of school attendance and labor market participation. As Heckman et al. (2008) discuss, the existence of psychic costs of schooling drives a wedge between the Mincer rate of return and the internal rate of return to education. Indeed, Heckman et al. (2008) show that if psychic costs of schooling are significant and ignored, the Mincer coefficient is expected to be larger than the internal rate of return. Although my point estimates of the psychic cost of school attendance is large, they are estimated imprecisely.

Performing the same analysis as done in Heckman et al. (2008) where the psychic cost of working is included shows this cost also drives a wedge between the Mincer coefficient and the internal rate of return to education. However, the size of the Mincer coefficient relative to the internal rate of return is ambiguous. The estimates in Table 2 suggest the existence of significant psychic costs to labor market participation for individuals with more than nine years of experience.

9.4.3 Returns to education

Table 3 reports the estimates of the mean and variance of the returns to education. The mean returns to education vary from 0.13 to 0.17, implying my estimates of the returns to education lie within the higher end of the estimates from other studies that implement instrumental variables methods and data from similar time periods (see Card (1999) for a review of these studies), but also at the lower end of estimates produced in Heckman et al. (2008).

Table 3 indicates heterogeneity in the expected returns to education across race and labor market experience. Specifically, I find that the expected returns to education for individuals with 10 years of labor market experience or less is greater for whites than blacks. On the other hand, the expected returns to education for individuals with more than 10 years of experience is greater for blacks than for whites. I also find that expected the returns to education is greater for whites with at most 10 years of experience than whites with more than 10 years of experience. This pattern reverses for blacks, for whom the expected returns to education for at most 10 years of experience is less than returns for more than 10 years of experience.

I do not find any significant patterns in the expected returns to education across education levels for whites. However, I find the expected returns to education for black college graduates is greater than for blacks with any other level of educational attainment. Indeed, the expected returns to education for black college graduates is greater than the expected returns to education for any other level of education for both whites and blacks. These results run contrary to the option-values argument of Heckman et al. (2008) that returns to education for high school graduates should be greater than for college graduates, because high school graduates possess the option of going to college, which they would act on if they found it beneficial. However, the model of Heckman et al. (2008) has features that are not included here. In particular, this model does not contain direct costs of schooling such as tuition costs.

Further analysis of the returns to education can be done by directly investigating its conditional distribution. One issue is the distinction between unobserved heterogeneity and uncertainty in the returns to education. Recall the estimator jointly estimates the number of types of individuals in the population, along with their probabilities. We find that, except for college graduates with more than 10 years of experience, two types of white males are in the sample: those with returns to education equal to 0.1 and others with returns to education equal to 0.2. The probabilities of all other types are equal to zero. Furthermore, these transition matrices are diagonal, implying these males do not face uncertainty in their returns to education, and their returns to education can be modeled as permanent unobserved heterogeneity. The transition matrix for white college graduates is not diagonal. Indeed, the chance that the returns to education of a white male college graduate will transition from 0.2 in any year to 0.15 in the following yearn is 11%. Symmetrically, the chance that a white college graduate will transition from having returns to education of 0.15 in any year to 0.2 in the following year is 11%. These results indicate both the possibility of skill appreciation and depreciation for white male college graduates.

The story somewhat reverses for black males: the type-transition matrices for black males with at least some college education and black high school graduates with at most 10 years of experience are diagonal with nonzero probabilities only for returns of education equal to 0.1 and 0.2. Therefore, the results do not provide evidence of uncertainty in the returns to education for these individuals. However, the type-transition matrices for the other groups of black males are significantly more populated, implying significant uncertainty about their returns to education. However, more stayers than movers exist, meaning the main diagonal elements of these type-transition matrices are large relative to the off-diagonals.

This empirical application indicates the existence of both unobserved heterogeneity in and uncertainty about the returns to education. Although the former has been extensively investigates, the latter has received significantly less attention. However, the economic implications of uncertainty in the returns to education are important. If individuals are risk averse, large uncertainty about their returns to education may have the consequence of them choosing lower levels of education than otherwise. This intuition my depend on the levels of education that are relatively risky. The potential importance of uncertainty in returns warrants further investigation and is left for future research.

10 Conclusion

CCP estimation of dynamic structural models has flourished over the last 20 years, largely because of the potential to dramatically reduce computational costs. The current state of art shows 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. Any future choice sequence chosen for a given initial choice generates a corresponding sequence of distributions of states. The term ρ-period finite dependence is obtained if two distinct initial choices with two corresponding future choice sequences can be constructed so that their respective distributions of states are the same after ρ-periods in the future. The smaller ρ is, the better, in that the computational cost reduces and, in many cases, the accuracy of the estimator improves. Though innovative, this property is restrictive and application of the CCP method often requires strong assumptions about the choice sequence, or the state transition probabilities, or both.

I generalize the concept of ρ-period dependence in a way that overcomes these two issues. I show that my generalization obtains one-period dependence ($\rho = 1$) in a large class of dynamic structural models that includes most models that have been estimated using the CCP method. Moreover, our method imposes no restrictions on the state transition probabilities, greatly increasing the scope of models the CCP method can estimate.

The class of models I consider is not restricted to models of discrete choices, but also includes models of both discrete and continuous choices. The representation developed in this model allows for a simple and elegant optimality condition for the continuous choices.

I show how to include discrete-valued unobserved heterogeneity into my model. I allow for the distribution of the unobserved heterogeneity to depend on observed covariates of the model. I provide sufficient conditions for identification of all the parameters of the model, including the conditional distribution of unobserved heterogeneity, and I propose a GMM estimator for these parameters. I propose an algorithm to compute the estimator. A key feature of the algorithm is how the distribution of unobserved heterogeneity is updated. I propose an iterated projection method, which jointly estimates the number of types. I am not aware of other methods with this property.

I implement my method to estimate a dynamic structural model of educational attainment of labor supply using data from the NLSY79 with the ultimate goal of estimating the distribution of the returns to education. My results show that, indeed, significant heterogeneity exists in the returns to education, with a range that aligns with studies of IV estimation of the returns to education, using the same data set as well as other data sets with comparable cohorts. The distribution depends on categories of race, education, and labor market experience. Particulary, I find returns to education vary between 0.13 and 0.17. I also find evidence of both (permanent) unobserved heterogeneity and uncertainty in the returns to education. Although the unobserved heterogeneity has been extensively investigated and accepted in the literature on the returns to education, the existence, significance, and economic implications of uncertainty regarding the returns to education are relatively unexplored and warrant further investigation.

A LEMMA AND THEOREMS

A.1 Proof of Theorem 8.4

Proof. Consistent with equation (7.1), for any $\theta \in \Theta$, π, and $w_t \in \mathcal{W}_t$, $\bar{\pi}(w_t, \theta, \pi)$ solves

$$\begin{aligned} 0 &= \hat{S}_1^{\pi}(\bar{\pi}(w_t, \theta, \pi); w_t, \theta, \pi) \\ &= \frac{1}{n}\sum_{i=1}^{n} X'_{itt+1}(\theta, \pi)\left(y_{itt+1} - X_{itt+1}(\theta, \pi)\bar{\pi}(w_t, \theta, \pi)\right) I_{it}(w_t) \end{aligned} \tag{A.1}$$

identically over θ over Θ, where $I_{it}(w_t)$ is equal to 1 if $w_{it} = w_t$, and zero otherwise. Standard linear regression calculations obtain.

By the mean value theorem, equation (A.1) obtains

$$\begin{aligned} &\bar{\pi}_t(w_t, \theta, \pi) - \pi_{t0}(w_t, \theta, \pi) \\ &= \left[\frac{1}{n}\sum_{i=1}^{n} X'_{itt+1}(\theta, \pi) X_{itt+1}(\theta, \pi) I_{it}(w_t)\right]^{-1} \frac{1}{n}\sum_{i=1}^{n} X'_{itt+1}(\theta, \pi)\nu_{it,t+1,jk}(\theta, \pi) I_{it}(w_t) \\ &= \frac{1}{n}\sum_{i=1}^{n} a_t(y_i, x_i; w_t, \theta, \pi), \end{aligned} \tag{A.2}$$

where $\nu_{it,t+1,jk}(\theta, \pi)$ is defined in equation (6.14), and

$$\begin{aligned} a_t(y_i, x_i; w_t, \theta, \pi) &= \left[\frac{1}{n}\sum_{i=1}^{n} X'_{itt+1}(\theta, \pi) X_{itt+1}(\theta, \pi) I_{it}(w_t)\right]^{-1} \\ &\quad \times \frac{1}{n}\sum_{i=1}^{n} X'_{itt+1}(\theta, \pi)\nu_{it,t+1,jk}(\theta, \pi) I_{it}(w_t), \end{aligned} \tag{A.3}$$

By Assumption 5.1.4 (continuity of the conditional valuation function for each $\theta \in \Theta$), Assumption 5.1.5 (strict concavity of the period-specific utility function), and Assumption 8.2, $a_t(y_i, x_i; w_t, \theta, \pi)$ is continuous at each $\theta \in \Theta$, two positive and finite constants, $\mathcal{T}_1$ and $\mathcal{T}_2$, exist such that

$$\|a_t(y_i, x_i; w_t, \theta, \pi)\| < \mathcal{T}_1 + \mathcal{T}_2 y_{itt+1}, \quad \text{and} \quad E[y_{itt+1}] < \infty. \tag{A.4}$$

By the law of iterated expectations, $E[a_t(y_i, x_i; w_t, \theta, \pi)] = 0$. Therefore, $a_t(y_i, x_i; w_t, \theta, \pi)$ satisfies the conditions of Theorem 2.4 of Newey and McFadden (1994), implying

$$\sup_{\theta,\pi} \|\bar{\pi}(w_t, \theta, \pi) - \pi_0(w_t, \theta, \pi)\| \xrightarrow{p} 0 \tag{A.5}$$

for each $w_t \in \mathcal{W}_t$. Also, by the continuous mapping theorem,

$$\sup_{\theta,\pi} \|\hat{\pi}(w_t, \theta, \pi) - \pi_0(w_t, \theta, \pi)\| \xrightarrow{p} 0 \tag{A.6}$$

for each $w_t \in \mathcal{W}_t$.

Now, by continuity of $\hat{S}^{\theta}(\theta;\pi)$ in π and θ, and by Assumption 8.2, a positive and finite constant $\mathcal{T}_3 < \infty$ exists for which

$$\left\|\hat{S}^{\theta}(\theta;\hat{\pi}) - S^{\theta}(\theta;\pi_0)\right\| < \mathcal{T}_3 \sup_{\theta,\pi} \left\|\frac{\partial}{\partial\pi}\tilde{S}^{\theta}(\theta;\pi)\right\| \max_{w_t \in \mathcal{W}} \sup_{\theta,\pi} \|\hat{\pi}(w_t, \theta, \pi) - \pi_0(w_t, \theta, \pi)\| = o_p(1), \tag{A.7}$$

where the last inequality follows from equation (A.6). It is straightforward to show that $\Omega m_i(\theta;\pi)$ satisfies the conditions of Theorem 2.6 of Newey and McFadden (1994), from which we conclude

$$\hat{\theta} \xrightarrow{p} \theta_0. \tag{A.8}$$

Finally, by $a_t(y_i, x_i; w_t, \theta, \pi)$, differentiable at each (θ, π) with $\sup_{\theta,\pi} \|\partial a_t(y_i, x_i; \theta, w_t)/\partial(\theta, \pi)\| < \infty$, equations (A.6) and (A.8) imply

$$\hat{\pi}(w_t) = \hat{\pi}(w_t, \hat{\theta}, \hat{\pi}) \xrightarrow{p} \pi_0(w_t) = \pi_0(w_t, \theta_0, \pi_0)$$

for each $w_t \in \mathcal{W}_t$. □

A.2 Proof of Theorem 8.5

Proof. Define $\hat{M}_\theta(\theta;\pi) = \partial\hat{m}(\theta;\pi)/\partial\theta$, where $\hat{m}(\theta;\pi)$ is defined in equation (6.3). Then $\hat{\theta}$ solves

$$0 = \frac{\partial}{\partial\theta}\hat{S}^\theta(\hat{\theta};\hat{\pi}) = \hat{M}_\theta(\hat{\theta};\hat{\pi})'\hat{\Omega}\hat{m}(\hat{\theta};\hat{\pi}). \tag{A.9}$$

The standard linearization procedure of equation (A.9), first in θ and then in π, obtains

$$\begin{aligned}\sqrt{n}(\hat{\theta}-\theta_0) = {}& \left[\hat{M}_\theta(\tilde{\theta};\hat{\pi})'\hat{\Omega}\hat{M}_\theta(\tilde{\theta};\hat{\pi})'\right]^{-1}\hat{M}_\theta(\theta_0;\hat{\pi})'\hat{\Omega} \\ & \times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{m}(\theta_0;\pi_0) + \sum_{t=2}^{T-1}\sum_{w_t\in\mathcal{W}_t}\hat{M}_{\pi_t}(\theta_0,\tilde{\pi})a_t(y_i,x_i;w_t,\theta_0,\pi_0)I_{it}(w_t)\right\},\end{aligned}$$

where $\tilde{\theta}$ and $\tilde{\pi}$ are mean values, $\hat{M}_{\pi_t}(\theta,\pi) = \partial\hat{m}_{\pi_t}(\theta,\pi)/\partial\pi_t(w_t)$, and $a_t(y_i,x_i;w_t,\theta,\pi)$ is defined in equation (A.3). Under the conditions of the theorem, it is straightforward to verify $\hat{m}(\theta_0;\pi_0)$ satisfies the conditions of Theorem 3.4 of Newey and McFadden (1994). Therefore, Theorem 6.1 of Newey and McFadden (1994) obtains the result.

□

B Tables

Table 1: Probability of Grade Promotion

Variable	Estimate	Std. Err.
Constant	1.1802	0.1932
Hours worked (h_t)	-2.6194	0.5173
Employment (d_t^h)	0.1740	0.0734
Education (EDU_t)	-0.0979	0.0345
Age $\times$ Education ($AGE_t EDU_t$)	-0.0144	0.0072
Black	1.3192	0.0917
AFQT	0.0692	0.0193

Table 2: Period-specific Utility

Variable	Estimate	Std. Err.
Utility of consumption $\exp(\theta_1^u x_t^c)(w_t h_t - w_t h_t \ln(h_t))$		
Age	0.1067	0.0006
Utility of leisure $\exp(\theta_2^u x_t^l)(l_t - l_t \ln(l_t))$		
Constant	3.9049	0.0487
Black	4.6626	5.8436
Age	0.0919	0.0010
Psychic cost of school attendance $\theta_3^u d_t^E x_t^E$		
Constant	-8.5048	11.2848
Education	0.7346	0.8167
Psychic cost of working $\theta_4^u d_t^h x_t^h$		
Constant	-5.0435	2.2737
Experience	0.5250	0.1894

Table 3: Mean and variance of returns to education

	White		Black	
	Mean	Variance	Mean	Variance
	At most ten years of experience			
Some high School	0.1481	0.0025	0.1334	0.0038
High school graduate	0.1495	0.0025	0.1381	0.0035
Some college	0.1465	0.0025	0.1389	0.0036
College graduate	0.1443	0.0025	0.1550	0.0025
	More than ten years of experience			
Some high School	0.1438	0.0025	0.1469	0.0028
High school graduate	0.1366	0.0023	0.1579	0.0024
Some college	0.1332	0.0022	0.1539	0.0025
College graduate	0.1412	0.0026	0.1654	0.0023

Table 4: Wage equation

Variable	Estimate	Std. Err.
Constant	-0.1751	0.0281
Experience ($EXPER_t$)	0.0082	0.0002
Squared experience ($EXPER_t^2$)	-0.0005	0.00001
Black	0.5753	1.7772
AFQT	0.0075	0.00001
Wage shock St.Dev	0.2118	0.0013

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. A. MILLER (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 79, 1823–1867.

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

CARD, D. (1999): *The Causal Effect of Education on Earnings*, Elsevier Science Publishers.

CHEN, Y. AND X. YE (2011): "Projection onto a Simplex," *Unpublished Manuscript*.

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.

HECKMAN, J. J., L. J. LOCHNER, AND P. E. TODD (2006): *Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond*, vol. 1, Elsevier Science B.V.

——— (2008): "Earnings Functions and Rates of Return," *Journal of Human Capital*, 2, 1–31.

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.