Semiparametric Discrete/Continuous Outcome Models with Nonseparable Unobserved Heterogeneity

Wayne-Roy Gayle*

June 10, 2011

Abstract

This paper investigates identification and estimation of discrete and continuous outcome models with nonseparable unobserved heterogeneity, without assuming that: (i) the regression function is strictly monotone in the unobservables, (ii) the joint distribution of the unobservables is independent of any of the observed regressors, or (iii) the observed regressors are continuous random variables. We show that a popular form of semiparametric restriction on the regression function satisfies two key properties. These restrictions are used to obtain identification of the regression function and the conditional distribution of the unobservables. We show, by way of examples, how these restrictions may be informed by economic theory. We investigate application of the identification results to models of dynamic panel data. We propose simple kernel-based estimators of all the parameters of interest, as well as the structural derivative. We derive uniform rates of convergence for the estimators and, in the case where the unobserved effect is discrete, we derive the asymptotic distribution of the structural function and the structural derivative. The results of a small Monte-Carlo study are presented.

KEYWORDS: Nonparametric estimation, nonseparable models, structural derivative.
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 Stephen Cosslett, Dukpa Kim, Shakeeb Khan, Author Lewbel, John Pepper, and Steven Stern for discussions and insightful comments. We also thank the seminar participants at Duke University, Ohio State University, and University of Virginia. All errors are my own.

1 Introduction

In this paper we develop methods for identifying and estimating models of discrete and continuous outcomes with nonseparable unobserved heterogeneity of the form

$$E[Y_j|X = x, U = u] = m_j(x,u), \quad j = 1, \cdots, J, \tag{1.1}$$

where X is a K-dimensional vector of observed explanatory variables, U is an L-dimensional vector of unobserved explanatory variables, and Y_j is the j-th outcome variable which may be discrete or continuous. In particular, we are interested in identifying and estimating the functions, $m_j, j = 1, \cdots, J$, and the joint distribution of the the unobserved random variables, U, without assuming: (i) that m_j is strictly monotone in U, (ii) that the joint distribution of U is independent of any of the Xs, or (iii) that X is a vector of continuous random variables.

In both the cross-section and panel data contexts, there is a large and growing literature on nonparametric identification and estimation of features of the model defined in, or similar to equation (1.1).[1] These papers obtain identification by imposing one, or a combination of the following assumptions: strict monotonicity of the structural function m in the unobservables; conditional independence; and index restrictions on the conditional distribution of the unobservables given the observables. In the cross-section setting, recent papers that investigate nonparametric identification of features (such as the average partial effect) of the structural function include Chesher (2003), Blundell and Powell (2004), Graham and Powell (2008), and Hoderlein and Mammen (2007). Studies that consider identification of the structural function and the conditional distribution of the unobservables in the cross-section setting include Matzkin (2003), Newey and Powell (2003), and Matzkin (2008).

The approach taken in this paper is most closely related to the work of Matzkin (2003) in that we consider identification of the function m, as well as the conditional distribution of the unobservables. However, unlike Matzkin (2003), we do not impose strict monotonicity assumptions of the structural function, nor do we assume that the unobservables are independent of a subset of the observed regressors, given the other observed regressors. Similar to Matzkin (2003), we can learn about the structural function m from prior knowledge of a subset of the population of interest. Matzkin (2003) assumes knowledge of structural func-

[1] See Matzkin (2007) for a survey of this literature.

tion for a known subset of the population and leaves the function generally unspecified for others in the population, except for monotonicity and smoothness constraints. In this paper, we impose index restrictions so that the unobservables do not enter the structural function for a subset of the population. We provide examples where the investigator has prior knowledge of the subpopulation, and where the subpopulation is unknown to the investigator. In the latter case, we derive sufficient conditions for identifying the subpopulation.

There are numerous possible restrictions that deliver identification in this paper, and the choice of the restriction will depend on the application of interest. However, in this paper, all possible restrictions that deliver identification have two key properties, which we outline in the next section. The most popular form of this restriction found in the literature is the single, linear index, (partial) random coefficients restriction. While the random coefficients restriction is popular in the literature, it is less understood relative to the fixed and random effects restrictions. Recent papers that investigate identifiability in these models include, among others, Chamberlain (1992), Hsiao and Pesaran (2004), Wooldridge (2005), Hoderlein and Mammen (2007), Graham and Powell (2008), and Arellano and Bonhomme (2009). However, the observation that the single index, partial random coefficient model satisfies the two key properties that are exploited in this paper seem to be novel.

While the method to identify the structural functions is the same whether the outcome variables are discrete or continuous, identification of the conditional distribution of the unobservables depends on whether the outcome variables are discrete or continuous. The identification results derived in this paper show how the dimension of the unobservables are restricted by J, the number of equations. Without further assumptions, the model can accommodate only discrete-valued unobservables when the outcome variables are discrete-valued. Identification of the unobservable random vector's conditional distribution fail in discrete outcome models when the number of support points of the unobserved random vector is too larger relative to the number of alternatives. In these cases, we discuss two additional restrictions that increases the number of unobservables that the model can handle. These restrictions are the conditional independence and conditional exchangeability restrictions.

The results developed in this paper extend directly to nonlinear panel data models. Recent studies that investigate identifiability in panel data models with nonseparable unobserved heterogeneity include Manski (1987), Newey (1994a), Honoré and Lewbel (2002), Arellano and Carrasco (2003), Altonji and Matzkin (2005), Gayle and Viauroux (2007), Lewbel

(2005), Wooldridge (2005), Graham and Powell (2008), Bester and Hansen (2009), Evdokimov (2009), Arellano and Bonhomme (2009), Hoderlein and White (2009), Gayle and Namoro (2010), and Gayle (2010). Among these papers, only a few consider dynamic panel data. Without further restrictions, the discussion in the above paragraph carry over to the dynamic panel data context. In addition, we show that the existence of panel data provides an additional source of identification of the conditional distribution of the unobservables if it is assumed to be stationary.

The rest of the paper is organized as follows: In the next section, we motivate the model and present the identification results. Section 3 presents estimators for the structural function, the conditional distribution of the unobservables, and the structural derivative. Section 4 presents the asymptotic properties of the proposed estimators. Section 5 presents the results of a limited simulation exercise, and Section 6 concludes. The Appendix contains the proofs of the asymptotic results.

2 Model

The basic model of interest is given by

$$\begin{aligned} E[Y_j|X=x,U=u] &= m_j(x,u), \\ E[Y_j|X=x] &= \int m_j(x,u)dF_{U|X=x}(u), \quad j=1,\cdots,J, \end{aligned} \tag{2.1}$$

where $m_j : \mathcal{X} \times \mathcal{U} \to \mathcal{M} \subset \Re$ with $X \in \mathcal{X} \subset \Re^K$ and $U \in \mathcal{U} \subset \Re^L$. Y and X are observable, and U is unobservable with a conditional distribution $F_{U|X}$ that potentially depends on all of X.

3 Identification

Before developing the general identification results, it is useful to consider the problem of identifying $m(x,u)$ in a single-index, (partial) random coefficients framework. Suppose $J =$

1, $X = (Z,V)$, where Z and V are unidimensional, U is unidimensional, and the conditional expectation in equation (2.1) is given by

$$E[Y|X = x, U = u] = m(z + u \cdot v). \tag{3.1}$$

Suppose $v = 0$ is in the support, $\mathcal{V}$, of V, and the support, $\mathcal{Z}$, of Z is such that for some values (z,v,u) in $\mathcal{X} \times \mathcal{U}$, $z' := z + u \cdot v \in \mathcal{Z}$.[2] Then the model given by equation (3.1) has two properties that are critical for what is to come. The first is that, for any value z of Z, the function m does not vary with U when $V = 0$, or equivalently,

Property 1: Y is mean independent of U given $V = 0$ and any value z of Z.

This observation obtains: $E[Y|Z = z, V = 0, U = u] = E[Y|Z = z, V = 0] = m(z)$. Therefore, the structural function m can be recovered from the subpopulation that has $V = 0$. Casual inspection reveals that the multiplicative specification is not the only specification that generates this property. For example, this property holds if instead, $m(x,u) = m(z + v^u)$, with $V = 0$, or $V = 1$. Define the value(s) of V for which Property 1 hold as $\bar{v}$.

Property 2: For $(z,v,u) \in \mathcal{X} \times \mathcal{U}$, with $z' := z + u \cdot v \in \mathcal{Z}$, $E[Y|Z = z, V = v, U = u] = m(z + u \cdot v) = m([z + u \cdot v] + 0) = E[Y|Z = z', V = 0, U = u]$.

This second property is an equivalence or transferability condition. The interpretation is that the expected outcome of a unit with values of (Z,V,U) equal to (z,v,u) is the same as the expected outcome of a unit with values of (Z,V,U) equal to $(z + u \cdot v, 0, u)$. Again, casual inspection reveals that this index restriction is not the only one that generates the second property. Indeed, the restriction $m(x,u) = m(z + v^u)$ has this property with $Z = z' = z + u^v$ and $V = 0$, or with $Z = z'_t = z + u^v - 1$ and $V = 1$.

There are index restrictions for which the second property is maintained, but the first is not, and vice versa. For example, if $m(x,u) = m(z + \alpha v + u)$, then the second property holds with $z' = z + \alpha v$ and $\bar{v} = 0$, but the first does not. Also, if $m(x,u) = m(zvu)$, then the first property holds with $\bar{v} = 0$, but the second does not. However, these two properties in concert

[2] Because the unknown link function, m, is not assumed to be strictly increasing (or strictly decreasing), assuming that the coefficient on z is 1 is innocuous.

obtains:

$$m(x,u) = E[Y|Z = z, V = v, U = u] = E[Y|Z = z', V = 0, U = u] = E[Y|Z = z', V = 0].$$

Therefore, given the restrictions implied by equation (3.1), the regression function $m(x,u)$ can be recovered from the distribution of the observable random variables.

The results from this example can be extended in a variety of directions. Two relevant directions are extensions to other specifications of the index, and to multiple unobserved regressors. Examples of other index restrictions that have the two key properties described above are: $m(x,u) = m(z\exp(u \cdot v))$, with $\bar{v} = 0$ and $z' = z\exp(u \cdot v)$; $m(x,u) = m(z^{1+u \cdot v})$, with $\bar{v} = 0$ and $z' = z^{1+u \cdot v}$; and $m(x,u) = m(z^{v^u} + v^u)$, with $\bar{v} = 1$, and $z' = z^{v^u} + v^u - 1$. An extension to multiple unobserved regressors is:

$$m(x,u) = m(z_1 + u_1 \cdot v_1, z_2 + u_2 \cdot v_2, \cdots, z_k + u_k \cdot v_k),$$

with $z'_k = z_k + u_k \cdot v_k$, and $\bar{v}_k = 0$, $k = 1, \cdots, K$.

To formally state the restrictions imposed on m, let $X = (W, Z, V)$ with $W \in \mathcal{W} \subset \Re^{K_w}$, $Z \in \mathcal{Z} \subset \Re^{K_z}$, and $V \in \mathcal{V} \subset \Re^{K_v}$, so that $K_w + K_z + K_v = K$. Recall that $U \in \mathcal{U} \subset \Re^L$. For $j = 1, \cdots, J$, let $g_j : \mathcal{Z} \times \mathcal{V} \times \mathcal{U} \to \Re$ be a known function. The restrictions imposed on m_j, $j = 1, \cdots, J$ are:

Assumption 3.1. *For all $x \in X$ and $u \in U$, $m_j(x,u) = m_j(w, g_j(z,v,u))$.*

Assumption 3.2. *For some given value $\bar{v}$ of V and all values z of Z, $g_j(z, \bar{v}, u)$ does not vary with u.*

Assumption 3.3. *The function g is such that, for any value (z,v,u) of (Z,V,U), there is a value z'_j depending on (z,v,u) such that $g_j(z,v,u) = g_j(z'_j, \bar{v}, u)$.*

Assumption 3.1 requires that the investigator possess prior information on the way Z, V, and U enter the function $m = (m_1, \cdots, m_J)$. In many cases, the form of the function $g = (g_1, \cdots, g_J)$ may be dictated by economic theory. In other cases, it may be a statistical restriction that is informed by the investigator's knowledge of the problem being studied.

Assumption 3.2 imposes additional restrictions on the functions g. The key implication of Assumption 3.2 is that the outcome Y_j is mean independent of the unobservables U con-

ditional on all values of X, Z, and $V = \bar{v}$, for some value $\bar{v}$. We will assume that $\bar{v}$ is known to the investigator while discussing identification of m and $F_{U|X}$. At the end of this section, we provide conditions under which $\bar{v}$ is identified.

Assumption 3.3 implies a form of substitutability between Z, V and U in the functions g. This condition is implicit in many semiparametric models such as single, linear index, (partial) random coefficient models. The choice of the function g_j imposes restrictions on the subset of $\mathcal{Z} \times \mathcal{V} \times \mathcal{U}$ on which m_j is identifiable. In particular, m_j is identifiable only at values (z, v, u) of (Z, V, U) where $z_j'(z, v, u) \in \mathcal{Z}$.

While Assumptions 3.1-3.3 may be implied by economic theory, it is rarely the case that economic theory imposes restrictions on the distribution of the unobservables, U, conditional on the observable regressors. In these cases, the estimation strategy proposed in this paper may prove to be fruitful. To help elucidate Assumptions 3.1-3.3 and illustrate the results of this paper, we provide some examples where these assumptions hold. The first example is a case where it is likely that the investigator knows $\bar{v}$. The second example is one where it is less clear that $\bar{v}$ is known to the investigator. In this case, identifiability of $\bar{v}$ becomes a more critical issue.

EXAMPLE 1: ***Discrete/Continuous Choice of Vehicle Usage***. Engers et al. (2009) establish the importance of perceived vehicle reliability on usage, measured by annual miles driven. In their work, vehicle dependability is proxied by the age of the vehicle. Engers et al. (2009) do not allow for correlation between observed and unobserved household-car specific heterogeneity in perceived reliability. Also, they do not account for the joint decisions of milage and household-car portfolio, which are likely to be correlated. Failure to account for these sources of endogeneity may lead to inconsistent estimates and erroneous conclusions. This example extends the discrete/continuous choice models of Dubin and McFadden (1984) and Hanemann (1984), using the method proposed in this paper to account for both sources of endogeneity.

Consider a typical household choosing among J vehicle brands and usage (miles driven). Let y_j denote usage of brand j, $j = 1, \cdots, J$. Define $d_j = 1\{y_j > 0\}$, where $1\{\cdot\}$ is the indicator function, equal to 1 if its argument is true, and 0 otherwise. The household's utility

(reward) function is given by

$$R = \bar{R}(c, y_1, \cdots, y_J, a_1, \cdots, a_J) + \sum_{j=1}^{J} d_j \varepsilon_{1j},$$

where c is the numeraire good, a_j is the (perceived) quality of car j, and ε_{1j} is an alternative specific taste shifter that is assumed to be independent of $(c, y_1, \cdots, y_j, a_1, \cdots, a_j)$. Denote $y = (y_1, \cdots, y_J)$. The household chooses (c, y) to maximize R subject to the budget constraint $\sum_{j=1}^{J} p_j y_j + c = I$, where p_j denotes the price of car j, and I is household income; and the non-negativity constraints: $y_j \geq 0, j = 1, \cdots, J$, and $c \geq 0$. Assume that $d = (d_1, \cdots, d_J)$ represent mutually exclusive and exhaustive choices so that $\sum_j d_j = 1$, and that the quality of brand j is relevant only if brand j is consumed, so that

$$y_j = 0 \Rightarrow \frac{\partial R}{\partial a_j} = 0.$$

The conditional direct utility of vehicle j is given by

$$\begin{aligned} R_j(c, y_j, a_j, \varepsilon_{1j}) &= \bar{R}(c, 0, \cdots, 0, y_j, 0, \cdots, 0, a1, \cdots, a_J,) + \varepsilon_{1j}, \\ &= \bar{R}_j(c, y_j, a_j) + \varepsilon_{1j}. \end{aligned}$$

Assume that $\bar{R}_j$ is twice continuously differentiable, strictly increasing, strictly concave in (c, y_j). Assume also that both c and y_j are essential with respect to $\bar{R}_j$. Then optimal usage of car j is obtained by maximizing $\bar{R}_j(I - p_j y_j,\ y_j,\ a_j)$ over $y_j > 0$, obtaining the conditional ordinary demand for car j miles and the numeraire good:

$$\begin{aligned} y_j^* &= \bar{y}_j(p_j, I, a_j), \\ c_j^* &= I - p_j y_j^*. \end{aligned}$$

The corresponding conditional indirect utility function is

$$\begin{aligned} \nu_j &= \bar{R}_j(I - p_j \bar{y}_j(p_j, I, a_j),\ \bar{y}_j(p_j, I, a_j),\ a_j) + \varepsilon_{1j}, \\ &= \bar{\nu}_j(p_j, I, a_j) + \varepsilon_{1j}. \end{aligned}$$

The optimal discrete choice of car type j is therefore given by

$$\begin{aligned} d_j^* &= 1\{\nu_j \geq \nu_k, \ \ \forall k\}, \\ &= 1\{\overline{\nu}_j(p_j, I, a_j) + \varepsilon_{1j} \geq \overline{\nu}_k(p_k, I, a_k) + \varepsilon_{1k}, \ \ \forall k\}. \end{aligned}$$

Now, suppose that household i's perceived quality of car j depends on miles-per-gallon, z_{ij}, other exogenous vehicle characteristics, b_{ij}, household characteristics, h_i, the age of the car, v_{ij}, and household-car specific heterogeneity in perceived reliability, u_{ij}:

$$a_{ij} = \overline{a}_j(b_{ij}, h_i, z_{ij} + u_{ij}v_{ij}).$$

This specification possesses two economic restrictions. First, given the observed characteristics of the household and car, unobserved household-car specific perceptions in car quality are relevant only for older cars. In other words, observationally equivalent households have the same perception of the reliability of observationally equivalent *new* cars. Second, consumers consider miles per gallon and the age of the car as substitutes, where the degree of substitutability depends on their unobserved household-car specific perception of the reliability of a *used* car.

Assume that observed miles traveled by household i with car j are measured with error, ε_{2ij}, so that $y_{ij} = y_{ij}^* + \varepsilon_{2ij}$. Denote $w_{ij} = (p_{ij}, I_i, b_{ij}, h_i)$, and $x_{ij} = (w_{ij}, z_{ij}, v_{ij})$. Define $\varepsilon_{1i} = (\varepsilon_{1i1}, \cdots, \varepsilon_{1iJ})$, $\varepsilon_{2i} = (\varepsilon_{2i1}, \cdots, \varepsilon_{2iJ})$, and $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i})$. Assume that ε_i has mean zero, and $\varepsilon_i \perp (x_i, u_i)$, where $x_i = (x_{i1}, \cdots, x_{iJ})$, and $u_i = (u_{i1}, \cdots, u_{iJ})$. Then the conditional expectation of observed miles traveled by household i with car j is given by

$$\begin{aligned} E[Y_{ij}|W_j = w_{ij}, Z_j = z_{ij}, V_j = v_{ij}, U_j = u_{ij}] &= \overline{y}_j(p_{ij}, I_i, \overline{a}_j(b_{ij}, h_i, z_{ij} + u_{ij}v_{ij})), \\ &= m_{1j}(w_{ij}, z_{ij} + u_{ij}v_{ij}). \end{aligned}$$

Also, the conditional choice probability that household i chooses car j is given by

$$\begin{aligned} &E[d_{ij}^*|W = w_i, Z = z_i, V = v_i, U = u_i] \\ &= Pr(\overline{\nu}_j(w_{ij}, z_{ij} + u_{ij}v_{ij}) + \varepsilon_{1ij} \geq \overline{\nu}_k(w_{ik}, z_{ik} + u_{ik}v_{ik}) + \varepsilon_{1ik}, \ \ \forall k), \\ &= m_{2j}(w_i, z_i + u_i v_i), \end{aligned}$$

where $w_i = (w_{i1}, \cdots, w_{iJ})$, $z_i = (z_{i1}, \cdots, z_{iJ})$, $v_i = (v_{i1}, \cdots, v_{iJ})$, $u_i = (u_{i1}, \cdots, u_{iJ})$, and $z_i + u_i v_i = (z_{ij} + u_{ij} v_{ij},\ j = 1, \cdots, J)$.

The conditional expectations m_{1j} and m_{2j} satisfy Assumption 3.1 with $g_j(z_j, v_j, u_j) = z_j + v_j u_j$. They satisfy Assumption 3.2 for $\bar{v}_j = 0$, and Assumption 3.3 with $z'(z_j, v_j, u_j) = g(z_j, v_j, u_j)$. Note that this model allows for correlation across the unobserved household-car specific perceptions of car reliability. Also it allows for the unobserved household-car specific effects to be correlated with all the other regressors of the model, including prices and the exogenous observable characteristics of cars. Note also that the conditional choice probabilities are generally not strictly monotonic in the unobserved household-car effects.

EXAMPLE 2: ***Dynamic Employment Decision with Unobserved Heterogeneity***. Let (Y_t, X_t) be observable random variables composed of employment decisions, Y_t, and observed explanatory variables, X_t, including tenure in the current job, V_{1t}, age, V_{2t}, accumulated hours worked, Z_t, and other standard regressors such as education, lagged employment decision, and IQ score, W_t. Suppose employment decision of individual i in period t is determined by:

$$Y_{it} = 1\{h[W_{it}\beta + Z_{it} + U_{1it}V_{1it} + U_{2it}V_{2it}(\bar{v}_2 - V_{2it})] - \varepsilon_t > 0\}, \tag{3.2}$$

where h is an unknown function, and ε_t is an unobserved noise term assumed to be independent of all the other regressors, with a strictly increasing and continuous CDF F_ε. The normalization of the coefficient on Z_t is without loss of generality because the function h is unknown.

U_{1it} captures individual heterogeneity in the returns to tenure. In the case where the individual is deciding whether to work in a new job, tenure in that new job is zero, and the decision does not depend on individual heterogeneity in returns to tenure. Because U_{1it} is allowed to depend on all the other regressors, including tenure, the model allows for returns to tenure to depend on experience in the particular job.

U_{2it} captures individual heterogeneity in the effect of age on employment decisions. It is a stylized fact that, after a certain age, the probability of employment becomes decreasing in age. The specification in equation (3.2) captures this pattern. Because the turning point age, $\bar{v}_2$, will be different depending on the population being investigated, it is not clear that the investigator will know $\bar{v}_2$ before hand. As such, being able to identify $\bar{v}_2$ becomes relevant in this context.

From equation (3.2), we obtain

$$E[Y_{it}|X_{it},U_{it}] = m(W_{it}\beta + Z_{it} + U_{1it}V_{1it} + U_{2it}V_{2it}(\bar{v}_2 - V_{2it})), \tag{3.3}$$

where $m = F_\varepsilon \circ h$. Equation (3.3) satisfies Assumption 3.1 with

$$g(z_t, v_t, u_t) = z_{it} + u_{1it}v_{1it} + u_{2it}v_{2it}(\bar{v}_2 - v_{2it}).$$

Assumption 3.2 is also satisfied for $\bar{v}_1 = 0$ and $\bar{v}_2$. Finally, equation (3.3) satisfies Assumption 3.3 with $z' = g(z_t, v_t, u_t)$. The regression function, m, can be identified only at values of accumulated hours worked, z_t tenure, v_{1t}, age, v_{2t}, and unobserved types (u_{1t}, u_{2t}), for which the subpopulation of individuals with accumulated hours equal to $z' = g(z_t, v_t, u_t)$, starting a new job, and are of age $\bar{v}_2$, has positive density.

The above examples show that Assumptions 3.1 - 3.3 are satisfied for a large class of models. The following theorem states formally that, under Assumptions 3.1 - 3.3, the function $m := (m_1, \cdots, m_J)$ is identified. Note that identification does not smoothness or isotonic constraints on m.

Theorem 3.4. *Let Assumptions 3.1 - 3.3 be satisfied. Suppose that, for fixed $(x,u) = (w,z,v,u)$ of (X,U), $z'(z,v,u) \in \mathcal{Z}$. Then the function m is identified at (x,u), and is given by $m_j(x,u) = E[Y_j|W = w, Z = z'_j, V = \bar{v}]$, $j = 1, \cdots, J$.*

Proof.

$$\begin{aligned} m_j(x,u) &= E[Y_j|W = w, Z = z, V = v, U = u] = m_j(w, g_j(z,v,u)) = m_j(w, g_j(z'_j, \bar{v}, u)) \\ &= E[Y_j|W = w, Z = z'_j, V = \bar{v}, U = u] = E[Y_j|W = w, Z = z'_j, V = \bar{v}], \end{aligned}$$

where the second equality is implied by Assumption 3.1, the third equality is implied by Assumption 3.3 and the fifth by Assumption 3.2. Hence, $m_j(x,u)$ is equal to $E[Y_j|W = w, Z = z'_j, V = \bar{v}]$, which is a function of the distribution of the observables. □

The theorem states that the conditional expectation of Y given the values the partial observables (w,z,v,u) of (W,Z,V,U) is equal to the conditional expectation of Y given the values $(w,z',\bar{v})$ of the observables (Z,X,V). The function $g(Z,V,U)$ dictates which values z'

of Z and $\bar{v}$ of V correspond to (z,v,u) for this equality to hold. If the model under investigation imposes further restrictions on the conditional expectation, then additional assumptions will generally be needed to identify these additional parameters.

EXAMPLE 2 CONTINUED: Theorem 3.4 shows that the "type-specific" probability of employment is identified. Assuming that $\mathbf{0} \in \mathcal{W}$, and that F_ε is a known, continuous, and strictly increasing CDF, the function h is identified as follows:

$$h(z) = F_\varepsilon^{-1}(E[Y_t|W_t = \mathbf{0}, Z_t = z, V_t = \bar{v}]).$$

Sufficient conditions for identification of β are that $E[W_t'W_t]$ is invertible, and that h is not periodic on its domain. Specifically, conditioning on $(W_t = w, Z_t = z, V_t = \bar{v})$ obtains

$$h(z + w\beta) = F_\varepsilon^{-1}(E[Y_t|W_t = w, Z_t = z, V_t = \bar{v}]),$$

which holds for all values of Z_t and any fixed w of W_t. Suppose there is another parameter vector, $\bar{\beta}$, for which this holds. Then by h identified and not periodic on its domain, we have that $w\beta = w\bar{\beta}$, which holds for all values w, so that $W_t\beta = W_t\bar{\beta}$. Finally, invertibility of $E[W_t'W_t)]$ implies that $\beta = \bar{\beta}$. This result generalizes the results of Ichimura (1993) to a setting that accounts for correlated unobserved heterogeneity.

The identification result in Theorem 3.4 is the same for models of discrete and continuous outcomes. However, identification of the distribution of U conditional on X depends on whether Y is continuous or discrete. We will provide sufficient conditions for three cases. The following results may be extended in a variety of directions. We discuss some of these possibilities as we go along.

3.1 Case 1: Discrete Y and Discrete U

Consider the discrete response model with J possible outcomes. Define Y_j to be one if alternative j is chosen, and zero otherwise, with $\sum_{j=1}^{J} Y_j = 1$. We make the following assumption about the support of U.

Assumption 3.5. *For $l = 1, \cdots, L$, the random variable U_l takes values in the known set $\mathcal{U}_l = \{u_l^1, \cdots, u_l^{R_l}\}$ with $u_l^r \neq u_l^{r'}$ for $r \neq r'$.*

Let $R = \prod_{l=1}^{L} R_l$, so that the random vector, U, takes values in the set $\mathcal{U} = \times_{l=1}^{L} \mathcal{U}_l = \{u_1, \cdots, u_R\}$. Let the probability associated with type r conditional on $X = x$ be given by $\pi_r(x)$, so that $\sum_{r=1}^{R} \pi_r(x) = 1$. Then we have that $m_j(x) = \sum_{r=1}^{R} m_j(x, u_r)\pi_r(x)$. By $\sum_{r=1}^{R} \pi_r(x) = 1$, $\pi_R(w) = 1 - \sum_{r=1}^{R-1} \pi_r(x)$, so that $H_j(x) := m_j(x) - m_j(x, u_R) = \sum_{l=1}^{R-1} [m_j(x, u_r) - m_j(x, u_R)]\pi_r(x)$. Further, define the $R-1$ dimensional row vector

$$Q_j(x) = [m_j(x, u_1) - m_j(x, u_R), \cdots, m_j(x, u_{R-1}) - m_j(x, u_R)],$$

the $R-1$ dimensional column vector $\pi(x) = [\pi_1(x), \cdots, \pi_{R-1}(x)]'$, and the $(J-1) \times (R-1)$ matrix $Q(x) = [Q_1(x)', \cdots, Q_{J-1}(x)']'$. Finally, let $H(x) := [H_1(x), \cdots, H_{J-1}(x)]'$. We are now in a position to state the identification result for discrete Y and U.

Theorem 3.6. *Suppose that, for given x of X, the rank of $Q(x)'Q(x)$ is $R-1$. Suppose the conditions of Theorem 3.4 and Assumption 3.5 hold. Then $[\pi_1(x), \cdots, \pi_R(x)]$ is identified and is given by*

$$\pi(x) = [Q(x)'Q(x)]^{-1} Q(x)'H(x), \quad \pi_R(x) = 1 - \sum_{l=1}^{R-1} \pi_r(x).$$

Proof. For given x in X, we have that $H(x) = Q(x)\pi(x)$. Note that $H(x)$ and $Q(x)$ are both identified from Theorem 3.4. Pre-multiplying this equation by $Q(x)'$ obtains

$$Q(x)'H(x) = Q(x)'Q(x)\pi(x).$$

If the rank condition of the Theorem is satisfied at x, then

$$\pi(x) = [Q(x)'Q(x)]^{-1} Q(x)'H(x).$$

Finally, we have $\pi_R(x) = 1 - \sum_{l=1}^{R-1} \pi_r(x)$. □

A necessary condition for $Q(x)'Q(x)$ to be of rank $R-1$ is the order condition that $J \geq R$, implying that at most J type probabilities are identified. This approach to identifying the mixture distribution is used in Bajari et al. (2008) to nonparametrically identify the probabilities of equilibrium selection in static and dynamic games. However, in Bajari et al. (2008), the equilibrium specific CCPs are assumed to be known up to a vector of finite dimensional parameters. The results of Theorem 3.4 and 3.6 provides a way to generalize the results of

Bajari et al. (2008).

An important implication of Theorem 3.6 is that the conditional distribution of U given X is not identified for values of x for which $v = \bar{v}$. To gain further insight to this implication, suppose $L = 1$ and $R = J = 2$, so that from Theorem 3.6

$$\pi_1(x) = \frac{m_1(x) - m_1(x,u_2)}{m_1(x,u_1) - m_1(x,u_2)}, \quad \pi_2(x) = 1 - \pi_1(x).$$

Suppose $g(z,v,u) = z + u \cdot v$. Then in this case,

$$\begin{aligned} m_1(x,u_1) \quad &= E[Y_1|X = x, U = u_1] = E[Y_1|W = w, Z = z + u_1 \cdot v, V = 0] \\ &\neq E[Y_1|W = w, Z = z + u_2 \cdot v, V = 0] = E[Y_1|X = x, U = u_2] = m_1(x,u_2), \end{aligned}$$

where the second and third equalities are obtained from Theorem 3.4. In this example, the rank condition of Theorem 3.6 is equivalent to: for any fixed w of W and for $V = 0$, varying Z from $z + u_1 \cdot v$ to $z + u_2 \cdot v$ must result in a change in the regression function. This condition fails if $v = \bar{v}$. However, for $v \neq \bar{v}$, as long as Z is relevant in the conditional expectation of Y, this condition holds.

3.2 Case 2: Continuous Y

In the case where outcomes are continuous, we assume that Y is generated by the following model:

$$Y_j = m_j(X,U) + \varepsilon_j, \quad j = 1, \cdots, J. \tag{3.4}$$

We make the following assumption on the distribution of $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_J)'$:

Assumption 3.7. *ε is independent of (Z,V,U), conditional on W, with conditional mean zero.*

This conditional independence assumption can be relaxed so that the conditional density of ε depends on (Z,V,U) as: $f_{\varepsilon|X,U} = f_{\varepsilon|W,g(Z,V,U)}$, with $E[\varepsilon|X,U] = 0$. This alternative restriction holds if ε is considered to be reduced form errors. The additive separability assumption may also be relaxed using methods of Matzkin (2003). The strategy of identifying

$F_{U|X}$ depends on whether U is discrete or continuous. We first investigate identification where Y_j is discrete.

3.2.1 Discrete U

In order to prove identification of $F_{U|X}$ when the outcomes are continuous and U is discrete, it is sufficient to consider a single outcome variable Y, that is, $J = 1$. Choose $-\infty < a_1 < \cdots < a_S < \infty$ and define $A_1 = (-\infty, a_1]$, $A_s = (a_s, a_{s+1}], s = 2, \cdots, S-1$, and $A_S = (a_S, \infty)$. Define $d_s = 1\{Y \in A_s\}$ where $1\{\cdot\}$ is the indicator function which is equal to 1 if the its argument is true and 0 otherwise. Define $Y_s = Y \cdot d_s, s = 1, \cdots, S$. Because of the assumptions placed on the distribution of ε, $m_s(x,u) := E[Y_s|X = x, U = u]$ inherits the index restrictions from $m(x,u)$, so that $m_s(x,u) = m_s(w, g(z,v,u))$. Therefore, under the assumptions of Theorem 3.4 $m_s(x,u)$ is identified. Let the support of U take the form presented in Assumption 3.5. As in Case 1, let $m_s(x,u_r) = E[Y_s|X = x, U = u_r]$ and $m_s(x) = E[Y_s|X = x]$. Then we have that $m_s(x) = \sum_{r=1}^{R} m_s(x,u_r)\pi_r(x)$. Define $H_s(x) = m_s(x) - m_s(x,u_R) = \sum_{l=1}^{R-1}[m_s(x,u_r) - m_s(x,u_R)]\pi_s(x)$. Further, define the $R-1$ dimensional row vector $Q_s(x) = [m_s(x,u_1), \cdots, m_s(x,u_{R-1})] - m_s(x,u_R)$, the $R-1$ dimensional column vector $\pi(x) = [\pi_1(x), \cdots, \pi_{R-1}(x)]'$, and the $(S-1)\times(R-1)$ matrix $Q(x) = [Q_1(x)', \cdots, Q_{S-1}(x)']'$. Finally, let $H(x) := [H_1(x), \cdots, H_{S-1}(x)]'$.

Theorem 3.8. *Suppose that, for given x of X, the rank of $Q(x)'Q(x)$ is $R-1$. Suppose the conditions of Theorem 3.4, Assumption 3.5, and Assumption 3.7 hold. Then $[\pi_1(x), \cdots, \pi_R(x)]$ is identified and is given by*

$$\pi(x) = [Q(x)'Q(x)]^{-1}Q(x)'H(x), \quad \pi_R(x) = 1 - \sum_{l=1}^{R-1} \pi_r(x).$$

Again the order condition requires that $S \geq R$. However, unlike the case with discrete outcomes, this condition is easily satisfied since the investigator chooses the partition of $\mathcal{Y}$.

3.2.2 Continuous U

The approach to identifying $f_{U|X}$ when U is continuous requires first identification of $f_{\varepsilon|W}$, where $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_J)$. Identification of $F_{\varepsilon|W=w}$ is obtained from the fact that under Assump-

tion 3.1, for fixed $W = w, Z = z, V = \bar{v}$, variation in Y_j is generated only by variation in ε. Define $Y = (Y_1, \cdots, Y_J)$. Then for fixed $\varepsilon = (e_1, \cdots, e_J)$,

$$\begin{aligned} F_{\varepsilon|W=w}(e) &= Pr(\varepsilon \leq e|W = w) = Pr(\varepsilon \leq e|W = w, Z = z, V = \bar{v}) \\ &= Pr(\varepsilon + m(w,z,\bar{v}) \leq e + m(w,z,\bar{v})|W = w, Z = z, V = \bar{v}) \\ &= Pr(Y \leq e + m(w,z,\bar{v})|W = w, Z = z, V = \bar{v}) \\ &= F_{Y|w,z,\bar{v}}(e + m(w,z,\bar{v})), \end{aligned} \tag{3.5}$$

where the third equality follows from Assumption 3.7. Because the last term in the series of equalities is identified from the observables, so is $F_{\varepsilon|W=w}$. Therefore, we have that

$$f_{\varepsilon|W=w}(e) = f_{Y|w,z,\bar{v}}(e + m(w,z,\bar{v})).$$

For fixed x of X with $v \neq \bar{v}$, define $P = P(U) = m(x,U)$. Next, we establish that the probability density function (pdf), $f_{P|X=x}$, is identified. To do so, additional restrictions are required. For any probability density function $q : \Re^J \to \Re$ let $\mathcal{F}[q](t)$ be the Fourier transform of q given by

$$\mathcal{F}[q](t) = \int e^{it \cdot a} q(a) da.$$

We make the following assumptions that are standard in the deconvolution literature.

Assumption 3.9. $|\mathcal{F}[f_{\varepsilon|W=w}](t)| > 0$, *for all real* t.

Assumption 3.10.

$$\sup_t \left| \frac{\mathcal{F}[f_{Y|X=x}](t)}{\mathcal{F}[f_{\varepsilon|W=w}](t)} \right| < \infty; \qquad \int \left| \frac{\mathcal{F}[f_{Y|X=x}](t)}{\mathcal{F}[f_{\varepsilon|W=w}](t)} \right| dt < \infty.$$

From equation (3.4), the conditional density function Y given $X = x$ is given by

$$f_{Y|X=x}(y) = \int f_{\varepsilon|W=w}(y-p) f_{P|X=x}(p) dp. \tag{3.6}$$

It is important to note that, because equation (3.6) is being evaluated at points where $v \neq \bar{v}$, the function $f_{Y|X=x}$ is not equal to the function $f_{\varepsilon|W=w} = f_{Y|w,z,\bar{v}}$. As such, the integral equation (3.6) is a Fredholm equation of the first kind, and not of the second kind (see

Robinson (1968)). Therefore, the standard convolution property for Fourier transforms can be used to solve equation (3.6) for $f_{P|X}$.

The convolution property of Fourier transforms obtains

$$\mathcal{F}[f_{Y|X=x}](t) = \mathcal{F}[f_{P|X=x}](t) \cdot \mathcal{F}[f_{\varepsilon|W=w}](t).$$

Under Assumptions 3.9 and 3.10, the invertibility of Fourier transforms obtains

$$f_{P|X=x}(p) = \mathcal{F}^{-1}\left[\frac{\mathcal{F}[f_{Y|X=x}]}{\mathcal{F}[f_{\varepsilon|W=w}]}\right](p). \tag{3.7}$$

Finally, we show that the conditional density of U can be recovered from $f_{P|X=x}$. At this stage, we will assume that $L = J$, so that the dimension of the random vector, U, is equal to the number of equations. This is with little loss of generality because if $J < L$, then new random variables can often be constructed so that $J = L$. These new equations however must satisfy the assumptions set out so far. Typical constructed random variables that satisfy the assumptions set out are $Y_{j''} = Y_j + Y_{j'}$ and $Y_{j''} = Y_j - Y_{j'}$ for $(j, j') <= J, j \neq j'$.

Partition $\mathcal{U} = \{\mathcal{U}^1, \cdots, \mathcal{U}^R\}$ so that $P_j = m_j(x,U), j = 1, \cdots, J$ defines a one-to-one transformation of each $\mathcal{U}^r$ onto $\mathcal{M}$. Let $P^r = m^r(x,U)$ denote the transformation from $\mathcal{U}^r$ onto $\mathcal{M}$. Then, by the standard transformation of random variables (see Theorem 14 of Mood et al. (1974)) we have that for $u \in \mathcal{U}$,

$$f_{U|X=x}(u) = \sum_{r=1}^{R} |J_r| \, f_{P|X=x}(P^r(u)), \tag{3.8}$$

where J_r is the Jacobian of the transformation given by $J_r = Det[\nabla_u m^r(x,u)]$.

3.3 Restrictions on $F_{U|X}$

This section investigates what is gained by imposing restrictions on $F_{U|X}$. The two restrictions considered are conditional independence and conditional exchangeability restrictions. We will consider only the case where the outcomes, Y, and the unobserved effects, U, are discrete valued, because this is the case that imposes the most restrictions on the dimension

and cardinality of U.

To investigate identifiability under conditional independence, partition $X = (X_1, X_2)$, where X_1 is a discrete valued, unidimensional random variable with K_1 points of support. We make the following assumption:

Assumption 3.11. *U is independent of X_1 conditional on X_2.*

Let $x_1^k,\ k = 1 \cdots K_1$, be the support points of X_1. For fixed x_1^k of X_1 and x_2 of X_2, define $H(x_1^k, x_2)$ and $Q(x_1^k, x_2)$ as in Section 3.1. Vertically stacking the $(J-1) \times (R-1)$ matrices $Q(x_1^k, x_2),\ k = 1, \cdots, K_1$ obtains the $(J-1)K_1 \times (R-1)$ matrix $Q(x_2)$. Vertically stacking $H(x_1^k, x_2)$ also obtains the $(J-1)K_1 \times 1$ matrix $H(x_2)$. By Assumption 3.11, $\pi(x)$ specializes to $\pi(x_2)$.

Theorem 3.12. *Suppose the rank of $Q(X_2)'Q(X_2)$ is $R-1$. Suppose the conditions of Theorem 3.4, Assumption 3.5, and Assumption 3.11 hold. Then $[\pi_1(X_2), \cdots, \pi_R(X_2)]$ is identified.*

The proof of Theorem 3.12 is the same as that of Theorem 3.6. Theorem 3.12 shows that at most $(J-1)K_1 + 1$ type probabilities can be identified under the conditions of the theorem. The number of type probabilities that can be identified increases linearly in J and K_1.

Another restriction that can be imposed on $F_{U|X}$ to accommodate identification of a larger number of type probabilities is the conditional exchangeability condition used in Kyriazidou (1997) and Altonji and Matzkin (2005), among others. To expound, consider again the case where X_1 is discrete, and further partition $X = (X_1, X_2, X_3)$, where X_2 is also unidimensional and discrete with the same support points as X_1. For any two fixed values (x, x') in the support of X_1 assume that $\pi_r(x, x', x_3) = \pi_r(x', x, x_3),\ r = 1, \cdots, R$, for all values x_3 of X_3. If $m_j(x, x', x_3, u) \neq m_j(x', x, x_3, u)$ for $x \neq x'$ and any (x_3, u) of (X_3, U), then for these events, at most $2J-1$ type probabilities can be identified. The proof of this statement is exactly as the proof of Theorem 3.6 and is therefore omitted.

3.4 Panel Data

In this section, we investigate identification of the parameters of the model in the panel data context. We focus on the case where the outcome and unobserved variables are discrete

valued, with the model given by

$$E_t[Y_t|X_t = x_t, U_t = u_t] = m_t(x_t, u_t), \quad t = 1, \cdots, T, \tag{3.9}$$

where Y_t is a J dimensional binary random vector. In the dynamic panel data framework, lagged values of Y_t are included in X_t. Note that m_t is identified under the restrictions of Assumptions 3.1-3.3. However, the rank condition of Theorem 3.6 may not be satisfied in many applications. To see why, consider Example 2. Theorem 3.6 shows that without further restrictions, the distribution $F_{U_t|X_t=x_t}$ is not identified since $J = 2$ and, even if U_1 and U_2 are binary random variables, $R = 4$.

The question of interest is if the panel data structure provides any information to help identify the distribution of the unobserved effects. Without any additional restriction, the answer is no. This is because equation (3.9) places no restrictions on m_t or $f_{U_t|X_t=x}$ over time.

In cases where the rank condition fails, the conditional independence and conditional exchangeability restrictions discussed in Section 3.3 may be sufficient satisfy the rank condition needed to identify $f_{U_t|X_t}$. However, there are some restrictions familiar in the panel data literature that increases the number of type probabilities that can be identified. One such restriction is to assume that the conditional distribution of the unobservables is stationary over time, that is $F_{U_t|X_t=x} = F_{U_{t'}|X_{t'}=x}$ for all (t, t'). This assumption is also employed in Arellano and Carrasco (2003) and Graham and Powell (2008). It is restrictive in that it rules out time-specific heteroskedasticity. Under this restriction, Theorem 3.4 can be adopted to show that at most $(J-1)T+1$ type probabilities can be identified as long as the intersection of the time-specific supports $\mathcal{X}_t$ is non-empty.

EXAMPLE 2 CONTINUED: Suppose the model given in equation (3.2) contains q lags of the dependent variable. Then the discussion above suggests that if the conditional distribution of the unobservables is stationary, then at most $T+1-q$ type probabilities are identified. Indeed, this requires that the support the of X_ts to at least overlap. An example where this restriction may be violated is in cases where the data set only contains a single cohort so that the support of age, education and experience changes over time.

3.4.1 Identifying transition probabilities

In the panel data context, it is often of interest to identify the transition probability of the unobservables. To investigate identification of transition probabilities in our framework, we continue to focus on the discrete outcome model. Also, for parsimony, we focus on the single lagged dependent variable model and, separate Y_{t-1} from X_t. The transition probability of interest is the first-order Markov transition probability given by $P(X_t, U_t|Y_{t-1}, X_{t-1}, U_{t-1}) = F_{X_t,U_t|Y_{t-1},X_{t-1},U_{t-1}}$. We make the following limited feedback assumptions.

Assumption 3.13.

$$f_{Y_t|Y_{t-1},Y_{t-2},X_t,X_{t-1},U_t,U_{t-1}} = f_{Y_t|Y_{t-1},X_t,U_t},$$

and

$$f_{U_t|Y_{t-1},Y_{t-2},X_t,X_{t-1},U_{t-1}} = f_{U_t|Y_{t-1},X_t,U_{t-1}}.$$

Under Assumption 3.13, we have

$$f_{Y_{jt},Y_{j't-1}|Y_{t-2},X_{t-1}} = \tag{3.10}$$
$$\int f_{Y_{jt}|Y_{j't-1},X_t,U_t} \cdot f_{Y_{j't-1}|Y_{t-2},X_{t-1},U_{t-1}} \cdot f_{U_{t-1}|Y_{t-2},X_{t-1}} \cdot f_{X_t,U_t|Y_{j't-1},X_{t-1},U_{t-1}} \cdot d(X_t, U_t, U_{t-1}).$$

First consider identification of $f_{X_t,U_t|Y_{j't-1},X_{t-1},U_{t-1}}$ under the assumption that X_t is discrete valued with cardinality $|\mathcal{X}|$, and maintaining Assumption 3.5. Let $\Lambda := \mathcal{U} \times \mathcal{U} \times \mathcal{X}$, with typical element $\lambda^d = (u^{r'_1}, u^{r'_2}, x^{l'})'$. $D = R^2|\mathcal{X}|$ is the number of elements in Λ. For $\lambda^d = (u^{r'_1}, u^{r'_2}, x^{l'})'$ let

$$\varphi_{jd}(Y_{j't-1}, Y_{t-2}, X_{t-1}) :=$$
$$f_{Y_{jt}|Y_{j't-1},X_t=x^l,U_t=u^{r_1}} \cdot f_{Y_{j't-1}|Y_{t-2},X_{t-1},U_{t-1}=u^{r_2}} \cdot f_{u^{r_2}|Y_{t-2},X_{t-1}} / f_{Y_{j't-1}|Y_{t-2},X_{t-1}}$$

and

$$\pi_d(Y_{j't-1}, X_{t-1}) := f_{x^l,u^{r_1}|Y_{j't-1},X_{t-1},U_{t-1}=u^{r_2}}.$$

Then equation (3.10) and $\sum_{d=1}^{D} \pi_{jd}(Y_{j't-1}, X_{t-1}) = 1$ obtains

$$f_{Y_{jt}|Y_{j't-1},Y_{t-2},X_{t-1}} - \varphi_{jD}(Y_{j't-1}, Y_{t-2}, X_{t-1}) = \sum_{d=1}^{D-1} [\varphi_{jd}(Y_{j't-1}, Y_{t-2}, X_{t-1}) - \varphi_{jD}(Y_{j't-1}, Y_{t-2}, X_{t-1})]\pi_d(Y_{j't-1}, X_{t-1}). \quad (3.11)$$

Define the $D-1$ dimensional row vector

$$Q_j(Y_{j't-1}, Y_{t-2}, X_{t-1}) = [\varphi_{j1}(Y_{j't-1}, Y_{t-2}, X_{t-1}), \cdots, \varphi_{j,D-1}(Y_{j't-1}, Y_{t-2}, X_{t-1})] - \varphi_{jD}(Y_{j't-1}, Y_{t-2}, X_{t-1}),$$

and the $(J-1) \times (D-1)$ matrix

$$Q(Y_{j't-1}, Y_{t-2}, X_{t-1}) = [Q_1', \cdots, Q_{J-1}']',$$

where we use the shorthand notation $Q_j = Q_j(Y_{j't-1}, Y_{t-2}, X_{t-1})$. Finally, define the $J(J-1) \times (D-1)$ matrix

$$Q(Y_{j't-1}, X_{t-1}) = [Q(Y_{j't-1}, \{Y_{1,t-2} = 1\}, X_{t-1})', \cdots, Q(Y_{j't-1}, \{Y_{J,t-2} = 1\}, X_{t-1})']'.$$

Theorem 3.14. *For fixed values $y_{j,t-1}$ of $Y_{j,t-1}$ and x_{t-1} of X_{t-1}, suppose the rank of*

$$Q(y_{j,t-1}, x_{t-1})' Q(y_{j,t-1}, x_{t-1})$$

is $D-1$. Suppose the conditions of Theorem 3.1, Assumptions 3.5, and 3.13 hold. Then $\pi_d(y_{j,t-1}, x_{t-1})$, $d = 1, \cdots, D$ are identified.

Proof. Under Assumptions 3.1-3.3, $Q(y_{j,t-1}, x_{t-1})$ and $\varphi_{jD}(Y_{j't-1}, Y_{t-2}, X_{t-1})$ are identified. The rest of proof is the same as that of Theorem 3.6. □

The necessary condition for the rank condition in Theorem 3.14 is that $J(J-1) \geq R^2|\mathcal{X}| - 1$, and $T \geq 3$. This condition will almost certainly not hold in practice, particularly because J is typically small relative to $|\mathcal{X}|$. Therefore, one will often need to impose more structure in order to achieve identification. We consider two additional restrictions. The first is:

Assumption 3.15.

$$F_{X_t,U_t|Y_{t-1},X_{t-1},U_{t-1}} = F_{X_t|Y_{t-1},X_{t-1},U_t} \cdot F_{U_t|Y_{t-1},X_{t-1},U_{t-1}}$$

and

$$F_{X_t|Y_{t-1},X_{t-1},U_t} = P(X_t|Y_{t-1},X_{t-1},U_t) = P(X_t|Y_{t-1},W_{t-1},g_x(Z_{t-1},V_{t-1},U_t)),$$

where $g_x(Z_{t-1},V_{t-1},U_t)$ satisfies Assumptions 3.2 and 3.3.

Under the second part of Assumption 3.15, $F_{X_t|Y_{t-1},X_{t-1},U_t}$ is identified using the strategy outlined in this paper. Note that we no longer require that X_t be discrete-valued. Following the strategy outlined in this section, it can be shown that, with the addition of Assumption 3.15, the necessary condition for identification of the type probabilities is that $J(J-1) \geq R^2 - 1$ with $T \geq 3$, which is easier to satisfy in practice. For example, in panel data, discrete outcome models with a single binary type, identification requires that $J \geq 3$ and $T \geq 3$.

Second, we consider adding the following conditional stationarity assumption.

Assumption 3.16. *For all t,s, and values (y,x,u) of $(Y_{t-1},X_{t-1},U_{t-1})$,*

$$F_{U_t|Y_{t-1}=y,X_{t-1}=x,U_{t-1}=u} = F_{U_s|Y_{s-1}=y,X_{s-1}=x,U_{s-1}=u}.$$

It can be shown that by adding Assumptions 3.15 and 3.16 to Theorem 3.14, the necessary condition for identification of the type probabilities is that $(T-2)J(J-1) \geq R^2 - 1$, which is relatively easy to satisfy in practice. For example, for panel data, binary outcomes models with binary types, identification requires $T \geq 4$.

EXAMPLE 3: ***Conditional Choice Probability (CCP) Representation of Dynamic Discrete Choice Models with Correlated Unobserved Heterogeneity***. Accounting for unobserved heterogeneity when estimating CCP models is known to be difficult, and until recently, considered to be impossible. In a recent paper, Kasahara and Shimotsu (2009) proposed a method of identifying the relevant CCPs jointly with the distribution of the unobserved heterogeneity in a finite mixture framework. The identification strategy employed by Kasahara and Shimotsu (2009) requires that the unobserved effects are time-invariant, and independent of all the observed regressors, except for the initial conditions. Arcidiacono and Miller (2010) propose a method of estimating the type specific CCPs and the preference parameters

by employing a suitable adaptation of the Expectation-Maximization (EM) algorithm. This method allows for time-varying unobservable covariates with transition probabilities that are independent of the observed covariates. They also allow for correlated initial conditions. The algorithm presented in Arcidiacono and Miller (2010) is computationally intensive relative to the case estimating the same model without unobservable heterogeneity. The results presented in this section provides a computationally simple methodology for accounting for correlated unobserved heterogeneity in the CCP framework. It requires no additional technology to what is proposed in Altug and Miller (1998). The reason for the simplicity of the method proposed in this paper is that the type-specific CCPs are directly identified from the distribution of the observables if Assumptions 3.1 - 3.3 are satisfied. Therefore, the goal of this example is to derive sufficient conditions for which the type specific CCPs satisfy Assumptions 3.1 - 3.3. Further notation is required to set up the environment.

Keeping the notation of Section 3.1, define Y_{jt} to be one if alternative j is chosen in period $t \in 0, \cdots, T$, and zero otherwise with $\sum_{j=1}^{J} Y_{jt} = 1$. Let $Y_t = (Y_{1t}, \cdots, Y_{Jt})$. The individual chooses $\{Y_t, t = 0, \cdots, T\}$ to maximize the expected lifetime utility

$$E_0 \left\{ \sum_{t=0}^{T} \sum_{j=1}^{J} \beta^t Y_{jt} [U_{jt}(X_t, U) + \varepsilon_{jt}] \right\},$$

where $U_{jt}(X_t, U) + \varepsilon_{jt}$ is the alternative-period specific utility, $\beta \in (0,\ 1)$ is the time discount factor, X_t is the observed state vector, and U is the permanent unobserved heterogeneity. Assume that $\varepsilon_t = (\varepsilon_{1t}, \cdots, \varepsilon_{Jt})$ is independently and identically distributed over time. Let Y_t^o be the optimal decision rule at time t. Then the type-r specific expected value of lifetime utility at period t, given the state X_t is

$$\Gamma_t(x_t, u^r) = E_t \left\{ \sum_{s=t}^{T} \sum_{j=1}^{J} \beta^{s-t} Y_{js}^o [U_{jt}(X_s, U) + \varepsilon_{js}] | X_t = x_t, U = u^r \right\}.$$

The type-r specific conditional valuation function is therefore given by

$$\nu_{jt}(x_t, u^r) = U_{jt}(x_t, u^r) + \beta \int_{\mathcal{X}} \Gamma_{t+1}(X_{t+1}, U) dF_{jt}(X_{t+1} | x_t, u^r),$$

where $dF_{jt}(X_{t+1} | x_t, u^r)$ is the alternative-period-type-r specific transitional probability of the state vector X_t.

Given the form of the type specific conditional valuation function, sufficient conditions for $\nu_{jt}(X_t,U) = \nu_{jt}(W_t, g(Z_t,V_t,U))$ are that the state dependent, alternative-period specific utility $U_{jt}(X_t,U) = U_{jt}(W_t, g(Z_t,V_t,U))$, and $F_{jt}(X_{t+1}|X_t,U) = F_{jt}(X_{t+1}|W_t, g(Z_t,V_t,U))$. Then assuming that the distribution of ε_t has continuous support, the type-r specific CCP has the form

$$
\begin{aligned}
&P_{jt}(X_t,u^r) \\
&= E\left[1\{\varepsilon_{jt} - \varepsilon_{kt} \geq \nu_{kt}(W_t, g(Z_t,V_t,u^r)) - \nu_{jt}(W_t, g(Z_t,V_t,u^r)) \ \ \forall \ \ k = 1,\cdots,J\}|W_t, g(Z_t,V_t,u^r)\right] \\
&= P_{jt}(W_t, g(Z_t,V_t,u^r)).
\end{aligned}
$$

Therefore, the type-specific CCPs inherit the restrictions imposed on the utility function and the transition probability, and if g satisfies Assumptions 3.1-3.3, then the type specific CCPs are identified. Furthermore, one can use the results of Section 3.1 or 3.2 to identify the distribution on types. Notice that this approach allows for the type specific CCPs and the type distribution to be time specific. Also, notice that even though assuming that $U_{jt}(X_t,U)$ and $F_{jt}(X_{t+1}|X_t,U)$ are strictly increasing in U implies that $\nu_{jt}(X_t,U)$ is strictly increasing in U, it does not imply that $P_{jt}(X_t,U)$ is strictly increasing in U. Therefore, methods that rely on strict monotonicity are generally not available in this framework. Extending the framework presented here to allow for continuous choices is straightforward. Furthermore, assuming that the disturbances, ε, are drawn from the generalized extreme values distribution, and that $U_{J,t}(X_t,U) = 0$, the utilities associated with the other alternatives are identified without further substantive restrictions. This result is derived directly from the inversion theorem of Hotz and Miller (1993).

3.5 Identifying $\bar{v}$

Perhaps the most restrictive assumption made so far is that $\bar{v}$ is known to the investigator. This assumption may limit the scope for application of the results obtained so far. Recall that in Example 2, we argued that the investigator may not have prior information about the turning point age, $\bar{v}_2$, in the model of employment decisions. For models such as Example 2 to be useful, it is important that $\bar{v}$ is identifiable. This section presents sufficient conditions under which $\bar{v}$ is identified.

For any square integrable vector of random variables, A, let $\|A\|^2_{X,U} := E[A'A|X,U]$. For

$j = 1, \cdots, J$, define

$$s_j^2(X,U) = \|Y_j - m_j(X,U)\|_{X,U}^2.$$

Assumption 3.17. *For $j = 1, \cdots, J$, $s_j^2(X,U) = s_j^2(W, g_j(Z,V,U))$.*

Assumption 3.18. *There is a subset $\bar{\mathcal{X}}$ of $\mathcal{X}$ so that $\bar{v} \in \bar{\mathcal{V}}$ and for any u in $\mathcal{U}$ and z of $\bar{\mathcal{Z}}$, $z'(z,v,u) \in \mathcal{Z}$ for any $v \in \bar{\mathcal{V}}$.*

Assumption 3.17 imposes the same index restriction on the conditional variance of Y_j given (X,U) as that imposed on the conditional mean of Y_j given (X,U). An alternative restriction that obtains identification of $\bar{v}$ is that $s_j^2(X,U) = s_j^2(W)$, that is, the conditional variance of Y_j is independent of (Z,V,U) given W. This alternative assumption may be more appealing in cases where Y is a continuous vector. Indeed, this condition is implied by Assumption 3.7. However, this assumption is violated if Y is a vector of binary random variables. In binary choice models such as Example 2, Assumption 3.17 is implied by the model specification. Assumption 3.18 is used to ensure that, to find $\bar{v}$, one needs only to search over the subset $\bar{\mathcal{X}}$ that does not result in $z'(z,v,u) \notin \mathcal{Z}$.

The key result to identifying $\bar{v}$ is the following lemma.

Lemma 3.19. *Suppose Assumptions 3.1 - 3.3, 3.17, and 3.18 hold. Then, for any $v \in \bar{\mathcal{V}}$,*

$$\|Y - E[Y|W,Z,V=v,U]\|_v^2 = \|Y - E[Y|W,Z,V=\bar{v}]\|_{\bar{v}}^2$$

Proof. For any (x,u) of (X,U) satisfying Assumption 3.18, we have

$$\begin{aligned}
&\|Y - E[Y|X=x,U=u]\|_{x,u}^2 = \\
&\sum_{j=1}^{J} \|Y_j - E[Y_j|W=w,Z=z,V=v,U=u]\|_{x,u}^2 = \\
&\sum_{j=1}^{J} \|Y_j - E[Y_j|W=w,Z=z_j'(z,v,u),V=\bar{v},U=u]\|_{w,z_j'(z,v,u),u}^2 = \\
&\sum_{j=1}^{J} \|Y_j - E[Y_j|W=w,Z=z_j',V=\bar{v}]\|_{w,z_j',\bar{v}}^2
\end{aligned}$$

Therefore, we have that

$$
\begin{aligned}
&\|Y - E[Y|W,Z,V=v,U]\|_v^2 = \\
&E[\|Y - E[Y|X=x,U=u]\|_{x,u}^2|V=v] = \\
&\sum_{j=1}^{J} E[\|Y_j - E[Y_j|W=w,Z=z_j',V=\bar{v}]\|_{w,z_j',\bar{v}}^2|V=\bar{v}] = \\
&\sum_{j=1}^{J} \|(Y_j - E[Y_j|W,Z,V=\bar{v}]\|_{\bar{v}}^2 = \\
&\|Y - E[Y|W,Z,V=\bar{v}]\|_{\bar{v}}^2
\end{aligned}
$$

□

Theorem 3.20. *Let Assumptions 3.1 - 3.3, 3.17, and 3.18 hold. The function*

$$\|Y - E[Y|W,Z,V=v]\|_v^2 \tag{3.12}$$

obtains a unique minimum with $V=\bar{v}$.

Proof. For any $v \in \bar{\mathcal{V}}$, $v \neq \bar{v}$, we have

$$
\begin{aligned}
\|Y - E[Y|W,Z,V=v]\|_v^2 &= \|Y - E[Y|W,Z,V=v,U]\|_v^2 \\
&\quad + \|E[Y|W,Z,V=v,U] - E[Y|W,Z,V=v]\|_v^2 \\
&= \|Y - E[Y|W,Z,V=\bar{v}]\|_{\bar{v}}^2 \\
&\quad + \|E[Y|W,Z,V=v,U] - E[Y|W,Z,V=v]\|_v^2 \\
&> \|Y - E[Y|W,Z,V=\bar{v}]\|_{\bar{v}}^2 \\
&= \|Y - E[Y|W,Z,V=\bar{v}]\|_{\bar{v}}^2 \\
&\quad + \|E[Y|W,Z,V=\bar{v},U] - E[Y|W,Z,V=\bar{v}]\|_{\bar{v}}^2 \\
&= \|Y - E[Y|W,Z,V=\bar{v}]\|_{\bar{v}}^2.
\end{aligned}
$$

The first equality follows from the Pythagorean theorem (see Mammen et al. (2001)), and the second equality follows from Lemma 3.19. For the strict inequality, note that $\|E[Y|W,Z,V=v,U] - E[Y|W,Z,V=v]\|_v^2 > 0$ for any $v \neq \bar{v}$. Equality holds if, and only if $v=\bar{v}$ since by assumption, it is in, and only in this case, $E[Y|W=w,Z=z,V=\bar{v},U] = E[Y|W=w,Z=z,V=$

$\bar{v}]$. This obtains the second-to-last equality, and the last equality is again by the Pythagorean theorem.

$\square$

The intuition behind Theorem 3.20 is as follows. Conditional of any v of V, variation in Y comes from variation in the other observed explanatory variables, the residual, and variation in U. This holds for every subpopulation except for those with $V = \bar{v}$, in which case Y does not vary with U. This argument is similar to the arguments presented in Ichimura (1993) to motivate the semiparametric least squares estimator.

Because it is possible that multiple values of v satisfy Assumptions 3.1 and 3.2, the solution to equation (3.12) may be set valued. This does not create a problem, because any element of the solution set may be used to identify m and $f_{U|X}$. Earlier, one such example was given, where $m(x,u) = m(z+u^v)$. In this case, the solution to the minimization problem in Theorem 3.20 is $\bar{v} = \{0,1\}$.

4 The Estimator

This section presents estimators of the $m(X,U)$ and $f_{U|X=x}$ for the case where Y and U are both discrete. Estimation under the case where Y is continuous and U is discrete is a simple modification of the estimators presented here. We can also propose an estimator of $f_{U|X}$ for the case where both Y and U are continuous. This would amount to plugging sample analogs for the unknown quantities in the definition of $f_{U|X}$. However, the asymptotic properties of such an estimator are non-standard given that we would have the empirical Fourier transform of an estimated density in the denominator of an integral equation. Given that finite mixture models are most widely employed specification in empirical work, we choose to concentrate of the case where U is discrete, and leave estimation of $f_{U|X}$ when U is continuous for future work. An estimator for $\bar{v}$ is proposed at the end of this section.

Throughout the remainder of this paper, we focus on the case where $\bar{v}$ is in the interior of $\mathcal{V}$. Estimation of the parameters of interest for cases where $\bar{v}$ is on the boundary of $\mathcal{V}$ can be defined using boundary kernel methods, reflection methods, transformation methods, or a

hybrid of these alternatives.[3] Addressing the additional complications that these alternative present is beyond the scope of this paper.

Suppose the investigator has data $\{y_i, x_i\}_{i=1}^n$. Let $\mathcal{X}$ be a compact, connected set such that the density function f_X is bounded away from zero on $\mathcal{X}$. Define the function

$$K_\sigma(a - a_i) := \prod_{k=1}^{d_a} \sigma_k^{-1} \mathfrak{K}((a_k - a_{ik})/\sigma_k),$$

where d_a is the dimension of a, $\mathfrak{K} : \mathfrak{R} \to \mathfrak{R}$ is a kernel, and σ is the bandwidth, so that $K_\sigma : \mathfrak{R}^{d_a} \to \mathfrak{R}$. The following estimated density of $f_X(w,z,v)$ will be used repeatedly to define different estimators.

$$\hat{f}_X(w,z,v) = \frac{1}{n} \sum_{i=1}^n K_\sigma(w - w_i) K_\sigma(z - z_i) K_\sigma(v - v_i).$$

4.1 Estimation of $m(X,U)$

For fixed $x = (w,z,v)$ of X and fixed u of U, let $z'(z,v,u)$ and $\bar{v}$ satisfy the conditions of Theorem 3.1. Then, the estimator for $m_j(x,u)$ is given by

$$\hat{m}_j(x,u) = \frac{1}{n} \sum_{i=1}^n y_{ij} K_\sigma(w - w_i) K_\sigma(z'_j - z_i) K_\sigma(\bar{v} - v_i) / \hat{f}_X(w, z'_j, \bar{v}). \tag{4.1}$$

The estimation of $m_j(x)$ is given by

$$\hat{m}_j(x) = \sum_{i=1}^n y_{ij} K_\sigma(x - x_i) / \hat{f}_X(x). \tag{4.2}$$

4.2 Estimation of $F_{U|X=x}$ under Case 1

Recall that Case 1 is the case where both the outcome variable Y, and the unobservables, U, are discrete with $\mathcal{U} = \{u_1, \cdots, u_R\}$. Recall also that $\pi_r(x) := Pr(U = u_r | X = x)$, $m_j(x, u_r) =$

[3] See Zhang et al. (1999) for references.

$E[Y_j|X=x,U=u_r]$, and $m_j(x)=E[Y_j|W=x]$. In the proof of Theorem 3.6, it is shown that

$$\pi(x)=[Q(x)'Q(x)]^{-1}Q(x)'H(x),$$

where $H(x)=[H_1(x),\cdots,H_{J-1}(x)]'$, $H_j(x)=m_j(x)-m_j(x,u_R)$, $Q(x)=[Q_1(x)',\cdots,Q_{J-1}(x)']'$, $Q_j(x)=[m_j(x,u_1),\cdots,m_j(x,u_{R-1})]-m_j(x,u_R)$, and $\pi(x)=[\pi_1(x),\cdots,\pi_{R-1}(x)]'$. Kernel estimates of $H(x)$ and $Q(x)$, denoted by $\hat{H}(x)$ and $\hat{Q}(x)$ respectively, can be constructed from $\hat{m}(x,u)$ and $\hat{m}(x)$ obtained from equations (4.1) and (4.2). Then, for given x of X with $v\neq\bar{v}$, the estimator $\hat{\pi}(x)$ is given by

$$\hat{\pi}(x)=[\hat{Q}(x)'\hat{Q}(x)]^{+}\hat{Q}(x)'\hat{H}(x),$$

where $[\hat{Q}(x)'\hat{Q}(x)]^{+}$ is the generalized inverse of $[\hat{Q}(x)'\hat{Q}(x)]$.

4.3 Estimation of the Structural Derivative

A statistic of particular interest is the structural derivative, that is, the derivative of the structural function $m_j(x,u)$ with respect to a particular explanatory variable x_k (denoted by $m_{jk}(x,u)$), averaged over the distribution of the unobservables:

$$\beta_{jk}(x)=\sum_{r=1}^{R}m_{jk}(x,u_r)\pi_r(x),\quad \text{where}\quad m_{jk}(x,u)=\frac{\partial}{\partial x_k}m_j(x,u).$$

To present an estimator of $\beta_k(x)$, further notation is required. For a given (x,u) of (X,U), let $\bar{q}=\bar{q}(x)=(w,z'(z,v,u))$. Using this notation, we have

$$m_{jk}(x,u)=\frac{\partial}{\partial\bar{q}}m_j(\bar{q},\bar{v})\cdot\frac{\partial}{\partial x_k}\bar{q}(x).$$

In what follows, we will suppress the j index and use the notations: $\bar{x}=(\bar{q},\bar{v})$; $m_k(\bar{x})=m_k(x,u)$. Let

$$K_{\sigma,k}(x-x_i)=\sigma_k^{-2}\mathfrak{K}'((x_k-x_{ik})/\sigma_k)K((x_{-k}-x_{i,-k})/\sigma),$$

where $\mathfrak{K}'$ denotes the derivative of $\mathfrak{K}$ and x_{-k} denotes the vector x with the k-th element deleted. Let

$$\begin{aligned}
\hat{h}(x) &= \frac{1}{n}\sum_{i=1}^{n} y_i K_\sigma(x - x_i), \\
\hat{h}_k(x) &= \sum_{i=1}^{n} y_i \frac{\partial \bar{x}_r}{\partial x_k} K_{\sigma,k}(x - x_i), \\
\hat{f}_k(x) &= \sum_{i=1}^{n} y_i \frac{\partial \bar{x}_r}{\partial x_k} K_{\sigma,k}(x - x_i).
\end{aligned}$$

Then an estimator for $m_k(x)$ is given by

$$\hat{m}_k(\bar{x}) = \hat{f}_X^{-1}(\bar{x}) \left[\hat{h}_k(\bar{x}) - \hat{f}_k(\bar{x})\hat{h}(\bar{x})\right].$$

Finally, an estimator for $\beta_k(x)$ is given by

$$\hat{\beta}_k(x) = \sum_{r=1}^{R} \hat{\pi}_r(x)\hat{m}_k(\bar{x}_r).$$

4.4 Estimation of $\bar{v}$

Assume that $\hat{f}_X$ is bounded away from zero on the set $\bar{\mathcal{X}}$, and define $\tau_i = 1\{x_i \in \bar{\mathcal{X}}\}$. Theorem 3.20 suggests that an estimator for $\bar{v}$, denoted by $\hat{v}$, can be defined as minimizing

$$\frac{1}{n}\sum_{i=1}^{n} \tau_i \left[y_i - \hat{m}(w_i, z_i, v)\right]' \left[y_i - \hat{m}(w_i, z_i, v)\right] K_\sigma(v - v_i)$$

over $\bar{\mathcal{X}}$. Because $\bar{v} \in \bar{\mathcal{V}}$ by assumption, it is sufficient to search over the set $\bar{\mathcal{X}}$ for $\bar{v}$, as long as it is known before hand that $\bar{v} \in \bar{V}$. Because of this flexibility the choice of $\bar{\mathcal{X}}$, it can be chosen so the boundedness assumption on $\hat{f}_X$ is satisfied.

5 Asymptotics

In this section, we derive asymptotic results for the estimators proposed in the previous section. In order to do so, additional regularity conditions are required. The regularity conditions that are presented below are stronger than what are needed if the estimators are the final quantities of interest. However, these conditions are typically required if the estimators are intermediate (plug-in) estimators used in a second stage estimation strategy (see Newey and McFadden (1994) and Newey (1994a) for examples). Example 3 is one such case, where the parameters of the utility functions are the ultimate quantities of interest. The regularity conditions are taken primarily from Newey and McFadden (1994) and Newey (1994a). In keeping with their notation, let $h_j(x) = E[Y_j|X = x]\, f_X(x)$, and $\gamma_j(x) = (f_X(x), h_j(x))'$.

Assumption 5.1. *1. $\mathfrak{K}(\cdot)$ is differentiable of order $d \geq 2$, the derivatives of order d are bounded, $\mathfrak{K}(\cdot)$ is zero outside a bounded set, $\int \mathfrak{K}(a)da = 1$, and there is a positive integer s such that for all $l < s$, $\int \mathfrak{K}(a)a^l da = 0$.*
2. $f_X(x)$ is bounded away from zero on $\mathcal{X}$, and for $j = 1, \cdots, J$, there is a version of $\gamma_j(x)$ that is continuously differentiable of order d with bounded derivatives on an open set containing $\mathcal{X}$.
3. There is $p \geq 4$ such that $E[\|Y\|^p] < \infty$ and $[\|Y\|^p|X = x]\, f_X(x) < \infty$.
4. For $k = 1, \cdots, K$, the bandwidth $\sigma_k = c_k\sigma$, where $\sigma = \sigma(n)$ satisfies $n^{1-2/p}\sigma^K/\ln n \to \infty$.

Assumption 5.2. *For $j = 1, \cdots, J$, $g_j(z, v, u)$ is so that $z_j'(z, u, v)$ is continuously differentiable of order $d \geq 2$ with bounded derivatives on an open set containing $\mathcal{Z} \times \mathcal{V} \times \mathcal{U}$.*

Theorem 5.3. *Suppose Assumptions 3.1-3.3, 5.1 and 5.2 hold. Then for $d \geq l + s$ and fixed values (x, u) of X, U, and for $j = 1, \cdots, J$,*

$$\sup_{x,u} \left\| \nabla^l \hat{m}_j(x, u) - \nabla^l m_j(x, u) \right\| = O_p\left(\ln(n)^{1/2} (n\sigma^{2l+K})^{-1/2} + \sigma^s \right).$$

The proof is given in Appendix A.

Theorem 5.4. *Suppose the conditions of Theorem 5.3 and Assumption 3.5 hold. Then for*

$r = 1, \cdots, R$,

$$\sup_x \left\| \nabla^l \hat{\pi}_r(x) - \nabla^l \pi_r(x) \right\| = O_p\left(\ln(n)^{1/2} (n\sigma^{2l+K})^{-1/2} + \sigma^s \right).$$

For fixed (x,u) of (X,U), and recall $\bar{x} = \bar{x}(x) = (w, z'(z,v,u), \bar{v})$. Let $\bar{c} = \prod_{k=1}^K c_k$, $\mu_2(\mathfrak{K}) = \int a^2 \mathfrak{K}(a) da$, $R(\mathfrak{K}) = \int \mathfrak{K}^2(a) da$, and let $s_j^2(\bar{x}) = Var(Y_j - m_j(\bar{x}) | X = \bar{x})$ for $j = 1, \cdots, J$.

Theorem 5.5. *Suppose Assumptions 3.1-3.3, and 5.2 hold. Suppose Assumption 5.1 hold for $d \geq s$, $s = 2$. Let $\sigma^2 \to 0$, $\ln n/(n\sigma^K) \to \infty$, and $(n\sigma^K)^{1/2}\sigma^2 \to \kappa$. Then for fixed values (x,u) of (X,U), and for $j = 1, \cdots, J$,*

$$(n\sigma^K)^{1/2}[\hat{m}_j(x,u) - m_j(x,u)] \xrightarrow{d} N\left(b_j(\bar{x}), \frac{R(\mathfrak{K})^K s_j^2(\bar{x})}{\bar{c} f_X(\bar{x})} \right),$$

where

$$b_j(\bar{x}) = \kappa \mu_2(\mathfrak{K}) \sum_{k=1}^K c_k^2 \left(f_X^{-1}(\bar{x}) \frac{\partial}{\partial x_k} m_j(\bar{x}) \frac{\partial}{\partial x_k} f_X(\bar{x}) + \frac{1}{2} \frac{\partial^2}{\partial x_k^2} m_j(\bar{x}) \right).$$

The form of the asymptotic bias and variance in Theorem 5.5 are precisely those obtained by Ruppert and Wand (1994), specialized to the choice of kernel and bandwidth of this paper. As is standard, the asymptotic bias can be eliminated by the choice of $\mathfrak{K}$ that satisfies part 1 of Assumption 5.1 with $s > 2$ and $d \geq s$, or by undersmoothing. However, the form of the bias given in Theorem 5.5 is relatively simple to compute. Unfortunately, the asymptotic bias of the estimator of the structural derivative is quite complicated. In this case, we will impose assumptions that ensure that the limiting bias is zero.

Theorem 5.6. *Suppose Assumptions 3.1-3.3,and 5.2 hold. Let either: (i) Assumption 5.1 hold for $d \geq 3$, $s = 2$, $\sigma^2 \to 0$, $\ln n/(n\sigma^{K+2}) \to \infty$, and $(n\sigma^{K+2})^{1/2}\sigma^2 \to 0$; or (ii) Assumption 5.1 hold for $d > 3$, $s = 3$, $\sigma^3 \to 0$, $\ln n/(n\sigma^{K+2}) \to \infty$, and $(n\sigma^{K+2})^{1/2}\sigma^3 \to \kappa_1$; Then for fixed values (x,u) of (X,U), and for $j = 1, \cdots, J$,*

$$(n\sigma^{K+2})^{1/2}[\hat{\beta}_j(x) - \beta_j(x)] \xrightarrow{d} N\left(0, R_1(\mathfrak{K}) R(\mathfrak{K})^{K-1} \sum_{r=1}^R \left(\frac{\partial \bar{x}_r}{\partial x_k} \right)^2 \frac{\pi_r^2(x) s_j^2(\bar{x}_r)}{c_k^2 \bar{c} f_X(\bar{x}_r)} \right),$$

where $R_1(\mathfrak{K}) = \int (\mathfrak{K}'(a))^2 da$.

The proof of Theorem 5.6 is in Appendix C.

Remark 5.7. Because the estimation of $\bar{v}$ is an L-dimensional problem, it is easier to estimate than the other parameters, since they are of dimension greater than L. Indeed, the results of Newey (1994b) can be adopted to show that, under a suitable choice of the smoothing parameter, the estimator $\hat{v}$ converges to $\bar{v}$ at a fast enough rate so that its estimation does not affect the results derived above.

6 Monte Carlo Study

In this section, we investigate the small sample performance of the proposed estimator via a small Monte Carlo experiment. The model investigated here is the one where the outcome and unobserved variables are discrete valued. The model design is as follows.

$$
\begin{aligned}
&Y_j^* = h_j(Z + U \cdot V) + \varepsilon_j, \quad j = 1, \cdots, 4, \\
&Y_j = 1\{Y_j^* > Y_k^*,\ k \neq j\}, \\
&h_1(a) = \frac{1}{2}a^2 - 1, \quad h_2(a) = 1 - \frac{1}{2}a^2, \quad h_3(a) = a, \quad h_4(a) = 0, \\
&Z \sim U[-3, 3], \quad V \sim U[-1, 1], \\
&U \in \{0,\ 0.5,\ 1\}, \quad Pr(U = u_r | Z, V) = \frac{\exp(V + u_r Z)}{\sum_{l=1}^{R} \exp(V + u_l Z)}, \\
&F_{\varepsilon_j}(e) = \exp(-\exp(-e)), \quad \varepsilon_j \perp \varepsilon_k, j \neq k.
\end{aligned}
$$

The functions are chosen so that the conditional expectations are highly nonlinear in U, while allowing us to calculate them without the need for simulations. We perform 500 simulations of 500 and 1000 individuals. For each simulation the alternative-type specific conditional choice probabilities, $m_j(\cdot, u_r) = m_{jr}$, are calculated on a fixed grid of 100 points over $[-2.5,\ 2.5] \times [-1,\ 1]$. This ensures that for each point in the grid, the index is in the support of Z. To estimate the functions, we use the normal kernel and the Silverman's rule of thumb bandwidth.

Table 1 presents the mean absolute bias, |bias|, the variance, var, and the mean squared error, mse, of the conditional expectations $m_j(\cdot, u_r)$ and the type probabilities π_r. Figure

1 presents plots of the mean, median, 5th and 95th percentile of selected alternative-type specific choice probabilities.

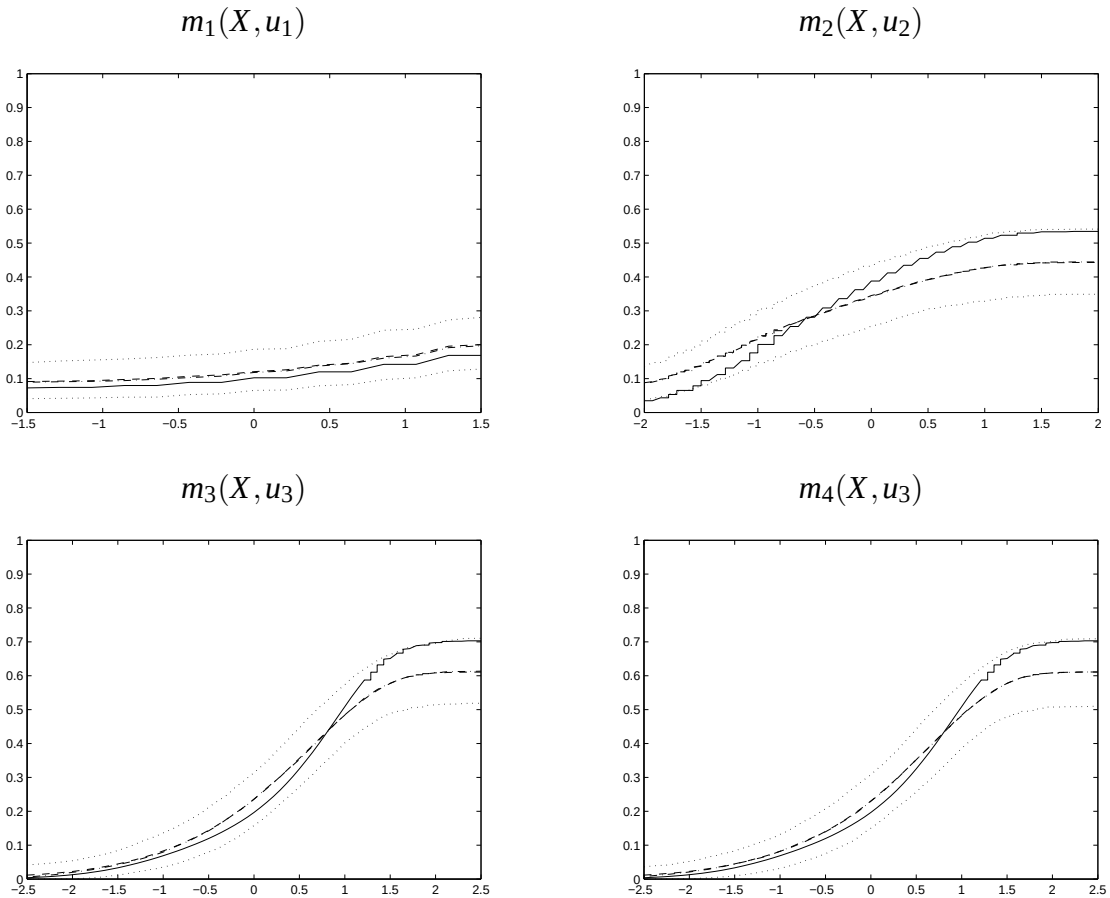
Table 1: Estimation of the Euler equation with habit formation[1]

	U=0			U=0.5			U=1		
	|bias|	var	mse	|bias|	var	mse	|bias|	var	mse
					N=500				
m_1	0.0348	0.0015	0.0020	0.0351	0.0016	0.0021	0.0364	0.0017	0.0023
m_2	0.0635	0.0028	0.0062	0.0639	0.0027	0.0062	0.0639	0.0026	0.0062
m_3	0.0453	0.0020	0.0036	0.0455	0.0022	0.0037	0.0453	0.0020	0.0037
m_4	0.0453	0.0022	0.0036	0.0454	0.0022	0.0037	0.0453	0.0022	0.0037
π	0.2064	0.0397	0.0642	0.1587	0.0273	0.0377	0.2446	0.0586	0.0913
					N=1000				
m_1	0.0274	0.0010	0.0013	0.0278	0.0010	0.0013	0.0230	0.0010	0.0014
m_2	0.0509	0.0018	0.0039	0.0510	0.0018	0.0039	0.0507	0.0017	0.0039
m_3	0.0371	0.0014	0.0024	0.0374	0.0014	0.0025	0.0374	0.0014	0.0025
m_4	0.0349	0.0013	0.0022	0.0351	0.0013	0.0022	0.0350	0.0013	0.0022
π	0.2030	0.0382	0.0625	0.1606	0.0288	0.0386	0.2370	0.0550	0.0859

7 Conclusion

In this paper, we investigated identification and estimation of discrete and continuous outcome models with nonseparable unobserved heterogeneity without assuming: (i) that the regression function is strictly monotone in the unobservables; (ii) that the joint distribution of the unobservables is independent of any of the observed regressors; or (iii) that the observed regressors are continuous random variables. In the case where the unobservable are discrete, we derived estimators for the identified parameters, and show that these estimators are consistent and asymptotically normal. The results from a small simulation exercise indicate that the proposed estimator works well in small samples.

Figure 1: Plot of selected estimates of type-specific conditional choice probabilities. True ——–, Mean – –, Median – · –, 5th and 95th percentile · · ·.



A Proof of Theorem 5.3

Proof. Under Assumption 5.1, Newey (1994b) shows that for $d \geq r+s$,

$$\sup_x \left\| \frac{\partial^r}{\partial x^r} \hat{m}_j(x) - \frac{\partial^r}{\partial x^r} m_j(x) \right\| = O_p\left(\ln(n)^{1/2} (n\sigma^{2r+K})^{-1/2} + \sigma^s \right).$$

Assumptions 3.1 - 3.3, and the definition of $\hat{m}_j(x,u)$ obtain

$$m_j(x,u) = m_j(w, z_j'(z,v,u), \overline{v}) \quad \text{and} \quad \hat{m}_j(x,u) = \hat{m}_j(w, z_j'(z,v,u), \overline{v}).$$

Therefore

$$\begin{aligned} \|\hat{m}_j(x,u) - m_j(x,u)\| &= \|\hat{m}_j(w, z_j'(z,v,u), \overline{v}) - m_j(w, z_j'(z,v,u), \overline{v})\| \\ &\leq \sup_x \|\hat{m}_j(x) - m_j(x)\|, \end{aligned}$$

so that

$$\begin{aligned} \sup_{x,u} \|\hat{m}_j(x,u) - m_j(x,u)\| &\leq \sup_x \|\hat{m}_j(x) - m_j(x)\| \\ &= O_p\left(\ln(n)^{1/2} (n\sigma^K)^{-1/2} + \sigma^s \right). \end{aligned}$$

Now

$$\frac{\partial}{\partial w} m_j(x,u) = \frac{\partial}{\partial w} m_j(w, z_j'(z,v,u), \overline{v}),$$

and, for $a = (z,v,u)$,

$$\frac{\partial}{\partial a} m_j(x,u) = \frac{\partial}{\partial z_j'} m_j(w, z_j'(z,v,u), \overline{v}) \frac{\partial}{\partial a} z_j'(a).$$

The same holds for the derivative of $\hat{m}_j(x,u)$. Therefore,

$$\begin{aligned} \left\| \nabla \hat{m}_j(x,u) - \nabla m_j(x,u) \right\| &\leq \sup_x \left\| \frac{\partial}{\partial x} \hat{m}_j(x) - \frac{\partial}{\partial x} m_j(x) \right\| \sup_a \left\| \frac{\partial}{\partial a} z_j'(a) \right\| \\ &\leq C \sup_x \left\| \frac{\partial}{\partial x} \hat{m}_j(x) - \frac{\partial}{\partial x} m_j(x) \right\|, \end{aligned}$$

so that

$$\begin{aligned}\sup_{x,u}\left\|\nabla\hat{m}_j(x,u)-\nabla m_j(x,u)\right\| &\leq C\sup_x\left\|\frac{\partial}{\partial x}\hat{m}_j(x)-\frac{\partial}{\partial x}m_j(x)\right\| \\ &= O_p\left(\ln(n)^{1/2}(n\sigma^{2+K})^{-1/2}+\sigma^s\right).\end{aligned}$$

The proof for $r>1$ follows the same arguments. □

B Proof of Theorem 5.4

Proof. For the results of Theorem 5.3 we have that

$$\sup_x\left\|\nabla^r\hat{Q}(x)-\nabla^r Q(x)\right\| = O_p\left(\ln(n)^{1/2}(n\sigma^{2r+K})^{-1/2}+\sigma^s\right),$$

and

$$\sup_x\left\|\nabla^r\hat{H}(x)-\nabla^r H(x)\right\| = O_p\left(\ln(n)^{1/2}(n\sigma^{2r+K})^{-1/2}+\sigma^s\right).$$

Now,

$$\begin{aligned}\hat{Q}(x)'\hat{H}(x)-Q(x)'H(x) &= (\hat{Q}(x)-Q(x))'(\hat{H}(x)-H(x)) \\ &+ (\hat{Q}(x)-Q(x))'H(x)+Q(x)'(\hat{H}(x)-H(x)),\end{aligned}$$

so that

$$\begin{aligned}\sup_x\left\|\hat{Q}(x)'\hat{H}(x)-Q(x)'H(x)\right\| &\leq \sup_x\left\|\hat{Q}(x)-Q(x)\right\|\sup_x\left\|\hat{H}(x)-H(x)\right\| \\ &+ \sup_x\left\|\hat{Q}(x)-Q(x)\right\|\sup_x\|H(x)\| \\ &+ \sup_x\|Q(x)\|\sup_x\left\|\hat{H}(x)-H(x)\right\|, \\ &= O_p\left(\ln(n)^{1/2}(n\sigma^K)^{-1/2}+\sigma^s\right). \qquad \text{(B.1)}\end{aligned}$$

The same operation, and continuity of the inverse obtains

$$\sup_x\left\|[\hat{Q}(x)'\hat{Q}(x)]^{-1}-[Q(x)'Q(x)]^{-1}\right\| = O_p\left(\ln(n)^{1/2}(n\sigma^K)^{-1/2}+\sigma^s\right). \qquad \text{(B.2)}$$

Again, the same operation, along with equations (B.1) and (B.2) obtain

$$\sup_x \|\hat{\pi}(x) - \pi(x)\| = O_p\left(\ln(n)^{1/2}(n\sigma^K)^{-1/2} + \sigma^s\right).$$

From this we have that

$$\sup_x \|\hat{\pi}_R(x) - \pi_R(x)\| \leq \sum_{r=1}^{R-1} \sup_x \|\hat{\pi}_r(x) - \pi_r(x)\| = O_p\left(\ln(n)^{1/2}(n\sigma^K)^{-1/2} + \sigma^s\right).$$

The proof for the r-th derivative follows the same lines. □

C Proof of Theorem 5.6

Proof. The standard decomposition into the stochastic and expectations parts is as follows

$$\hat{\beta}_k(x) - \beta_k(x) = \hat{\beta}_k(x) - E[\hat{\beta}_k(x)|X] + E[\hat{\beta}_k(x)|X] - \beta_k(x).$$

Consider first $\tilde{\beta}_k(x) - E[\tilde{\beta}_k(x)|X]$, where $\tilde{\beta}_k(x) = \sum_{r=1}^{R} \pi_r(x)\tilde{m}_k(\bar{x_r})$, and

$$\tilde{m}_k(\bar{x}) = f_X^{-1}(\bar{x})\left[\hat{h}_k(\bar{x}) - f_k(\bar{x})\hat{h}(\bar{x})\right].$$

By $y_i = \bar{m}(x_i) + \varepsilon_i$, with $\bar{m}(x_i) = E[m(X,U)|X = x_i]$ and $E[\varepsilon_i|X] = 0$, we have

$$\begin{aligned}
\tilde{\beta}_k(x) - E[\tilde{\beta}_k(x)|X] &= \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \bar{x_r}}{\partial x_k}\pi_r(x) f_X^{-1}(\bar{x_r})\varepsilon_i K_{\sigma,k}(\bar{x_r} - x_i) \\
&- \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \bar{x_r}}{\partial x_k}\pi_r(x) f_X^{-1}(\bar{x_r}) f_k(\bar{x_r})\varepsilon_i K_{\sigma}(\bar{x_r} - x_i),
\end{aligned}$$

so that

$$
\begin{aligned}
&(n\sigma^{K+2})^{1/2}\left[\tilde{\beta}_k(x)-E[\tilde{\beta}_k(x)|X]\right]= \\
&(c_k\bar{c})^{-1}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{R}\frac{\partial\bar{x}_r}{\partial x_k}\pi_r(x)f_X^{-1}(\bar{x}_r)\varepsilon_i\sigma^{K/2}\mathfrak{K}((\bar{x}_{rk}-x_{ik})/\sigma_k)K((\bar{x}_{r,-k}-x_{i,-k})/\sigma)\right)- \\
&\bar{c}^{-1}\sigma\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{R}\frac{\partial\bar{x}_r}{\partial x_k}\pi_r(x)f_X^{-1}(\bar{x}_r)f_k(\bar{x}_r)\varepsilon_i\sigma^{-K/2}K((\bar{x}_r-x_i)/\sigma)\right)= \\
&(c_k\bar{c})^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_{1i}+\bar{c}^{-1}\sigma\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_{2i}.
\end{aligned}
$$

By $E[\varepsilon_i|X]=0$, the Law of iterated expectations obtains $E[H_{1i}]=0$. For the variance,

$$
\begin{aligned}
&Var(H_{1i})= \\
&E\left[\varepsilon_i^2\left(\sum_{r=1}^{R}\frac{\partial\bar{x}_r}{\partial x_k}\pi_r(x)f_X^{-1}(\bar{x}_r)\mathfrak{K}'\left(\frac{\bar{x}_{rk}-x_{ik}}{\sigma_k}\right)K\left(\frac{\bar{x}_{r,-k}-x_{i,-k}}{\sigma}\right)\right)^2\sigma^{-K}\right]= \\
&\bar{c}\int\left(\sum_{r=1}^{R}\frac{\partial\bar{x}_r}{\partial x_k}\pi_r(x)f_X^{-1}(\bar{x}_r)s(\bar{x}_r-\sigma u_r)f_X^{1/2}(\bar{x}_r-\sigma u_r)\mathfrak{K}'(u_{rk})K(u_{r,-k})\right)^2du= \\
&\bar{c}\sum_{r=1}^{R}\int\left(\frac{\partial\bar{x}_r}{\partial x_k}\right)^2\pi_r^2(x)f_X^{-2}(\bar{x}_r)s^2(\bar{x}_r-\sigma u_r)f_X(\bar{x}_r-\sigma u_r)(\mathfrak{K}'(u_{rk}))^2(K(u_{r,-k}))^2du+ \\
&2\bar{c}\sum_{r=1}^{R-1}\sum_{l=r+1}^{R}\int\frac{\partial\bar{x}_r}{\partial x_k}\frac{\partial\bar{x}_r}{\partial x_l}\pi_r\pi_l f_X^{-1}(\bar{x}_r)f_X^{-1}(\bar{x}_l)s(\bar{x}_r-\sigma u_r)s(\bar{x}_l-\sigma u_l)\times \\
&f_X^{1/2}(\bar{x}_r-\sigma u_r)f_X^{1/2}(\bar{x}_l-\sigma u_l)\mathfrak{K}'(u_{rk})\mathfrak{K}'(u_{rl})K(u_{r,-k})K(u_{r,-l}). \qquad \text{(C.1)}
\end{aligned}
$$

By the uniform boundedness conditions, $\int\mathfrak{K}(a)da=1$, $\int a\mathfrak{K}(a)da=0$, $\int\mathfrak{K}'(a)da=0$, and $\int\mathfrak{K}'(a)da=-\int a^2\mathfrak{K}(a)da$, obtain that the second term on the RHS of equation (C.1) is $O(\sigma)$, while the first is

$$
\bar{c}R_1(\mathfrak{K})R(\mathfrak{K})^{K-1}\sum_{r=1}^{R}\left(\frac{\partial\bar{x}_r}{\partial x_k}\right)^2\pi_r^2(x)f^{-1}(\bar{x}_r)s^2(\bar{x}_r)+O(\sigma).
$$

Therefore

$$
Var(H_{1i})\to\bar{c}R_1(\mathfrak{K})R(\mathfrak{K})^{K-1}\sum_{r=1}^{R}\left(\frac{\partial\bar{x}_r}{\partial x_k}\right)^2\pi_r^2(x)f^{-1}(\bar{x}_r)s^2(\bar{x}_r)
$$

as $n \to \infty$. The same conditions are implemented to show that

$$\frac{1}{n}\sum_{i=1}^{n}|H_{1i}|^{2+\delta} \to 0 \quad \text{as} \quad n \to \infty,$$

so that, by the Liapunov central limit theorem,

$$(c_k\bar{c})^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_{1i} \xrightarrow{d} N\left(0, R_1(\mathfrak{K})R(\mathfrak{K})^{K-1}\sum_{r=1}^{R}\left(\frac{\partial \bar{x}_r}{\partial x_k}\right)^2 \frac{\pi_r^2(x)s^2(\bar{x}_r)}{c_k^2\bar{c}f_X(\bar{x}_r)}\right).$$

The same arguments can be used to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_{2i} = O_p(1),$$

so that $\bar{c}\sigma n^{-1/2}\sum_i H_{2i} \to 0$. By the arguments of Pagan and Ullah (1999), the asymptotic distribution of

$$(n\sigma^{K+2})^{1/2}\left[\tilde{\beta}_k(x) - E[\tilde{\beta}_k(x)|X]\right]$$

is the same as that of $(c_k\bar{c})^{-1}n^{-1/2}\sum_{i=1}H_{1i}$, obtaining

$$(n\sigma^{K+2})^{1/2}\left[\tilde{\beta}_k(x) - E[\tilde{\beta}_k(x)|X]\right] \xrightarrow{d} N\left(0, R_1(\mathfrak{K})R(\mathfrak{K})^{K-1}\sum_{r=1}^{R}\left(\frac{\partial \bar{x}_r}{\partial x_k}\right)^2 \frac{\pi_r^2(x)s^2(\bar{x}_r)}{c_k^2\bar{c}f_X(\bar{x}_r)}\right).$$

Under the conditions of the theorem, the following hold uniformly in x

$$\begin{aligned}
\hat{f}_X(x) &= f_X(x) + O_p(\ln(n)^{1/2}(n\sigma^K)^{-1/2} + \sigma^2), \\
\hat{f}_k(x) &= f_k(x) + O_p(\ln(n)^{1/2}(n\sigma^{K+2})^{-1/2} + \sigma^2), \\
\hat{\pi}_r(x) &= \pi_r(x) + O_p(\ln(n)^{1/2}(n\sigma^K)^{-1/2} + \sigma^2).
\end{aligned}$$

From this, it is straightforward to show that, under the conditions of the theorem,

$$\frac{\hat{\pi}_r(x)\,\hat{f}_X^{-1}(\bar{x}_r)}{\pi_r(x)\,f_X^{-1}(\bar{x}_r)} \xrightarrow{p} 1,$$

$$\frac{\hat{\pi}_r(x)\,\hat{f}_X^{-1}(\bar{x}_r)\,\hat{f}_k(\bar{x}_r)}{\pi_r(x)\,f_X^{-1}(\bar{x}_r)\,f_k(\bar{x}_r)} \xrightarrow{p} 1.$$

Therefore, application of the Slutsky theorem obtains

$$(n\sigma^{K+2})^{1/2}\left[\hat{\beta}_k(x) - E[\hat{\beta}_k(x)|X]\right] \xrightarrow{d} N\left(0, R_1(\mathfrak{K})R(\mathfrak{K})^{K-1}\sum_{r=1}^{R}\left(\frac{\partial \bar{x}_r}{\partial x_k}\right)^2 \frac{\pi_r^2(x)s^2(\bar{x}_r)}{c_k^2\bar{c}f_X(\bar{x}_r)}\right).$$

While standard, the proof that

$$(n\sigma^{K+2})^{1/2}[E[\hat{\beta}_k(x)|X] - \beta_k(x)] \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty$$

under the conditions of the theorem is messy, and is therefore omitted. Putting these together, conclude that

$$(n\sigma^{K+2})^{1/2}[\hat{\beta}_k(x) - \beta_k(x)] \xrightarrow{d} N\left(0, R_1(\mathfrak{K})R(\mathfrak{K})^{K-1}\sum_{r=1}^{R}\left(\frac{\partial \bar{x}_r}{\partial x_k}\right)^2 \frac{\pi_r^2(x)s_j^2(\bar{x}_r)}{c_k^2\bar{c}f_X(\bar{x}_r)}\right).$$

□

References

ALTONJI, J. AND R. MATZKIN (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053–1102.

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*.

ARELLANO, M. AND S. BONHOMME (2009): "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," *MIMEO, CEMFI, Madrid*.

ARELLANO, M. AND R. CARRASCO (2003): "Binary Choice Panel Data Models with Predetermined Variables," *Journal of Econometrics*, 115, 125–157.

BAJARI, P., V. CHERNOZHUKOV, H. HONG, AND D. NEKIPELOV (2008): "Nonparametric and Semiparametric Analysis of a Dynamic Game Model," *MIMEO, University of Minnisota*.

BESTER, A. AND C. HANSEN (2009): "Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model," *Journal of Business and Economic Statistics*, 27, 235–250.

BLUNDELL, R. AND J. POWELL (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 665–679.

CHAMBERLAIN, G. (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567–596.

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

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

ENGERS, M., M. HARTMANN, AND S. S. (2009): "Annual Miles Drive Used Car Prices," *Journal of Applied Econometrics*, 24, 1–33.

EVDOKIMOV, K. (2009): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," *MIMEO, Yale University*.

GAYLE, G. AND C. VIAUROUX (2007): "Root-N Consistent Semiparametric Estimators of a Dynamic Panel Data Sample Selection Model," *Journal of Econometrics*, 141, 179–212.

GAYLE, W. (2010): "Root-N Consistent Estimation of a Dynamic Panel Data Model," *MIMEO: Univeristy of Virginia*.

GAYLE, W.-R. AND S. D. NAMORO (2010): "Estimation of a Nonlinear Panel Data Model with Semiparametric Individual Effects," *MIMEO: University of Virginia*.

GRAHAM, B. AND J. POWELL (2008): "Identification and estimation of irregular correlated random coefficient models," *MIMEO, University of California - Berkeley*.

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

HODERLEIN, S. AND E. MAMMEN (2007): "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica*, 75, 1513–1518.

HODERLEIN, S. AND H. WHITE (2009): "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects," *MIMEO, Brown University*.

HONORÉ, B. AND A. LEWBEL (2002): "Semiparametric Binary Choice Panel Data Models Without Strictly Exogeneous Regressors," *Econometrica*, 70, 2053–2063.

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

HSIAO, C. AND M. PESARAN (2004): "Random Coefficient Panel Data Models," *MIMEO, University of Southern California*.

ICHIMURA, H. (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models," *Journal of Econometrics*, 58, 71–120.

KASAHARA, H. AND K. SHIMOTSU (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77, 135–175.

KYRIAZIDOU, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65, 1335–1364.

LEWBEL, A. (2005): "Simple Endogenous Binary Choice and Selection Panel Data Model Estimators," *MIMEO, Boston College*.

MAMMEN, E., J. MARRON, B. TURLACH, AND M. WAND (2001): "A General Projection Framework for Constrained Smoothing," *Statistical Science*, 16, 232–248.

MANSKI, C. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–362.

MATZKIN, R. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

——— (2007): *Nonparametric Identification*, vol. 6, Part 2, Elsevier Science B.V.

——— (2008): "Identification in Nonparametric Simultaneous Equations Models," *Econometrica*, 76, 945–978.

MOOD, A., F. GRAYBILL, AND D. BOES (1974): *Introduction to the Theory of Statistics*, McGraw Hill.

NEWEY, W. (1994a): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

——— (1994b): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253.

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

NEWEY, W. K. AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578, available at http://ideas.repec.org/a/ecm/emetrp/v71y2003i5p1565-1578.html.

PAGAN, A. AND A. ULLAH (1999): "Review: Nonparametric Econometrics by A. Pagan and A. Ullah, Cambridge University Press: Cambridge, UK. xviii+424 pages. 1999," .

ROBINSON, P. D. (1968): *Fourier and Laplace Transforms*, Dover Publications Inc.

RUPPERT, D. AND M. WAND (1994): "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 22, 1346–1370.

WOOLDRIDGE, J. (2005): "Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models," *Review of Economics and Statistics*, 87, 385–390.

ZHANG, S., R. J. KARUNAMUNI, AND M. C. JONES (1999): "An Improved Estimator of the Density Function at the Boundary," *Journal of the American Statistical Association*, 94, 1231–1241.